universität wien

DISSERTATION

Titel der Dissertation

„Microscopic analysis of the effects of post-translational modifications on protein structure, dynamics and aggregation by molecular dynamics simulations“

verfasst von

Drazen Petrov

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 094 490
Dissertationsgebiet lt. Studienblatt:	Molekulare Biologie
Betreut von:	Dr. Bojan Zagrovic, A.B.

This cumulative dissertation is based on the following publications:

Petrov, D.,* Margreitter, C.,* Grandits, M., Oostenbrink, C. & Zagrovic, B. (2013). A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* **9** (7), e1003154. (Chapter I) **joint first authors*

Margreitter, C.,* Petrov, D.* & Zagrovic, B. (2013). Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* **41** (Web Server issue), W422-6. (Chapter II) **joint first authors*

Petrov, D. & Zagrovic, B. (2011). Microscopic analysis of protein oxidative damage: effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133** (18), 7016-7024. (Chapter III)

Petrov, D. & Zagrovic, B. (2013). Are current atomistic force fields accurate enough to study proteins in crowded environments? To be submitted to *Proc. Natl. Acad. Sci. U.S.A.* (Chapter IV)

Petrov, D., Daura, X. & Zagrovic, B. (2013). Effect of oxidative damage on the stability and dimerization of superoxide dismutase 1. To be submitted to *PLoS Comput. Biol.* (Chapter V)

The author's contribution to each publication is clearly stated at the beginning of each corresponding chapter.

ABSTRACT

From enzymatic activation and cell-cycle control to transcription and translation regulation, post-translational modifications (PTMs) of proteins play a key role in numerous cellular processes by directly affecting protein structure, dynamics and interaction networks. However, despite the importance of understanding how PTMs affect protein properties and behaviour at the atomic level, suitable molecular dynamics (MD) simulation tools and parameters for treating PTMs have never been developed in a systematic fashion. Here, we generate and validate force field parameters (GROMOS 45a3 and 54a7) needed to run MD simulations of PTMs and develop an automated online tool for introducing PTMs into native proteins for over 250 different types of modifications. What is more, we characterize PTMs and compare them against the 20 canonical amino acids in terms of their physico-chemical properties by using thermodynamic integration (TI), molecular hydrophobicity potential and other computational approaches.

As an example of a non-enzymatic PTM, we study metal-catalyzed carbonylation, one of the most dominant mechanisms of oxidative damage to proteins. Notably, highly carbonylated proteins have been observed in potentially cytotoxic aggregates present in late-onset diseases including neurodegenerative diseases, cancer and diabetes. However, the effects of carbonylation on protein stability, dynamics and aggregation at the microscopic level are still poorly understood. Here, we perform extensive MD simulations of carbonylated villin headpiece domain accompanied by bioinformatic analysis of its aggregation propensity. Our results suggest that high concentrations of carbonylation, far above typical cellular levels are required for protein destabilization and unfolding, while aggregation propensity can drastically increase even upon a single “carbonylation mutation”. Additionally, we obtain high-resolution insight into the aggregation process of native and carbonylated villin headpiece and discover that several widely-used classical MD force fields appear to be significantly biased towards protein aggregation. Finally, we use MD and TI to study the impact of oxidative stress on protein stability and protein-protein interactions in the case of SOD1 homo-dimer whose destabilization is believed to be involved in amyotrophic lateral sclerosis.

ZUSAMMENFASSUNG

Eine Vielzahl von physiologischen Prozessen wird von posttranslationalen Modifikationen (PTMs) gesteuert, welche ihren Einfluss über eine Veränderung der Proteinstruktur, -dynamik und -interaktionsnetzwerken ausüben. Darunter fallen unter anderem die Kontrolle des Zellzyklus, die Regulierung von Transkription und Translation sowie die Steuerung der Aktivität vieler Enzyme. Trotz der vielen Bereiche in denen PTMs ein wichtige Rolle spielen fehlen angemessene Werkzeuge und Parameter für molekulardynamische (MD) Simulationen um ihre Auswirkungen auf Proteine genauer verstehen zu können. In dieser Arbeit generieren und validieren wir die notwendigen "Force-Field" Parameter (GROMOS 45a3 und 54a7) um MD Simulationen mit PTMs durchführen zu können. Zusätzlich präsentieren wir ein automatisiertes „online-tool" welches erlaubt native Proteine mit über 250 verschiedenen PTMs zu versehen. Außerdem werden die PTMs charakterisiert und anschließend mit den 20 kanonischen Aminosäuren mittels Thermodynamischer Integration (TI), molekularem hydrophobischen Potential und anderen Computer-basierten Methoden verglichen.

Als ein Beispiel für nicht-enzymatische posttranslationale Modifikationen, wir untersuchen metall-katalysierte Carbonylierung, eine der häufigsten Formen von oxidativem Schaden von Proteinen. Proteine mit hohem Carbonylierungsgrad bilden zellschädigende Aggregate die häufig in Spätformen von neurodegenerativen Krankheiten, Krebs und Diabetes beobachtet werden. Dennoch sind die Effekte der Carbonylierung auf Proteinstabilität, -dynamik und Aggregation auf mikroskopischem Level noch immer schlecht verstanden. Hier führen wir umfangreiche MD Simulationen mit Carbonylierter „villin headpiece domain" durch und analysieren anschließend das Aggregationsverhalten. Unsere Daten weisen darauf hin, dass erst ein überdurchschnittlich hoher, weit über zellulären Levels liegender, Carbonylierungsgrad zu einer Destabilisierung des Proteins und seiner Entfaltung führt, wohingegen die Aggregationswahrscheinlichkeit schon mit einer einzigen Carbonylierungsmutation drastisch steigen kann. Außerdem bekommen wir detaillierte Einsicht in den Aggregationsprozess von nativer und carbonylierter „villin headpiece domain" und entdecken, dass einige weit verbreitete klassische MD-Kraftfelder scheinen in Richtung zur Aggregation erheblich beeinflußt zu sein. Schlussendlich verwenden wir MD und TI um den Einfluss von oxidativem Stress auf Proteinstabilität und Protein-Protein

Interaktion im Fall des SOD1 Homo-Dimers dessen Destabilisierung wahrscheinlich eine entscheidende Rolle in amytropischer Lateralsklerose zu spielt, zu untersuchen.

TABLE OF CONTENTS

Preamble

Proteins carry out a vast array of vital biological functions in living cells ranging from DNA replication and transcription to translation and cell-cycle control to catalysis of metabolic reactions and signal transduction. Importantly, proteomes, i.e., inventories of synthesized proteins, are typically a few orders of magnitude more complex than their cognate genomes.[1-3] For example, there are more than 1 million distinct protein species in humans, significantly outnumbering the 20-25 thousand human genes known.[1,4,5] The expansion of genome coding capabilities is facilitated by two principal mechanisms: 1) alternative splicing at the transcriptional level generating multiple mRNAs from a single DNA template,[6-8] and 2) protein post-translational modifications (PTMs),[3,5,9] the focus of this thesis. More than 400 chemically different types of modifications of amino-acid side chains and the backbone have been identified to date with over 70000 individual PTMs occurring in more than 20000 proteins.[10,11] Importantly, these chemical alterations of polypeptides play a key role in different cellular processes, ranging from enzymatic activation to transcription and translation regulation to disease development and aging.[5,12-17] Notably, enzymes dedicated to such modifications account for approximately 5% of cellular protein content in higher eukaryotes.[5] For example, kinases, which predominantly catalyze phosphorylation of serine, threonine and tyrosine residues and are one of the largest and most important classes of PTM enzymes, actively participate in various regulatory mechanisms with a direct impact on enzyme activation and inhibition, signal transduction pathways and networks, cellular response to external stimuli, cell cycle, tumor suppression, etc.[18-21] Other prominent examples of addition of small functional groups to proteins such as acetylation, methylation, S-nitrosylation and hydroxylation control a large number of biological functions and processes, including cellular metabolism and signaling, gene expression and epigenetic regulation, transport and translocation of biomolecules, protein life-time and turnover and stabilization of structural proteins.[22-29] Remarkably, hydroxyproline, mainly implicated in the formation of collagen fibrils, is more abundant than seven individual canonical amino acids, comprising roughly 4% of all amino acids found in animal tissue.[26] Moreover, enzyme-mediated transfer of larger moieties such as peptides (e.g., ubiquitin and SUMO), cofactors (e.g., biotin, flavins and hemes), lipid anchors (e.g., prenylation, palmitoylation and myristoylation) and carbohydrate chains (glycosylation), targets proteins to degradation by proteasome machinery and relocation to cell membranes, modulates their function and

activity, and is implicated in quality control by aiding and regulating protein folding/refolding and secretion.[5,30-34] Furthermore, a large body of proteins is subjected to simultaneous modifications at multiple sites by different types of reactions, often affecting biological outcome in a combinatorial fashion.[5,17,35] Most notably, interplay of tandem modifications, methylations, acetylations and phosphorylations among others, from a pool of more than 50 experimentally identified PTM sites (generating $>10^{20}$ possible states) in human histones influences both activation and silencing of gene translation, including different counteracting modifications at the same residue (e.g., acetylation and methylation of LYS9 in the histone H3 protein are prevalently associated with opposite effects on gene expression).[11,35,36] What is more, a battery of enzymes designated for removal of introduced modifications control intensity and duration of propagated signals and effects, with only a small subset of PTMs being irreversible and lacking such regulatory mechanisms.[5,37,38]

In addition to enzymatic PTMs, proteins undergo uncontrolled covalent modifications triggered by reactive oxygen species (ROS) and reactive nitrogen species (RNS), including metal-catalyzed carbonylation, oxidation and nitration of aromatic amino-acid residues, oxidation of sulfur-containing residues and the protein backbone, or even protein fragmentation due to backbone breakage.[12,39,40] Protein non-enzymatic modifications may lead to cytotoxicity via impaired biological function and interactions, aggregation and misfolding.[40-43] Importantly, oxidative damage to proteins is associated with aging and age-related disorders such as neurodegenerative diseases, cancer, and diabetes.[12,40,44,45] While oxidized proteins are targeted for degradation, they can escape the ubiquitin-proteasome or other proteolytic pathways and accumulate in cells.[43,46,47] Markedly, late-onset pathologies are characterized by formation of insoluble aggregates typically comprised of highly oxidized proteins.[48-51] Furthermore, antioxidant enzymes (e.g., superoxide dismutase family) and small molecules (e.g., glutathione) scavenge free radicals (ROS and RNS), thus protecting the proteasome integrity, an important factor in preventing cellular senescence and aging.[52-55] Notably, irreversible metal-catalyzed carbonylation is one of the most prominent and the most studied biomarkers of oxidative damage and aging, whose level exponentially increases in the last third of the life-span in different species, ranging from houseflies to humans.[56-58] However, it is still largely unclear whether oxidative modifications to proteins

are a direct cause or only a consequence and useful reporter of aging.[45,53,59] Moreover, in addition to proteins, different biomolecules and cellular components are also affected by ROS and RNS, including DNA and lipid membranes.[52,53,60] In general, mounting evidence has linked oxidative stress and aging through different mechanisms: malfunctioning mitochondria coupled with increased production of ROS and RNS, impaired antioxidant protection, DNA damage, defective protein function and interactions, and protein aggregation in combination with compromised proteolytic machinery.[44,52,53,60-63] However, despite the widely-accepted notion that oxidative stress is a source of damage that accumulates with time and facilitates aging, several beneficial aspects of reactions involving free radicals in biological processes have also been reported, including cell signaling and even life-span prolongation.[53,64-66]

By directly affecting physico-chemical properties of target amino-acid, PTMs carry out their functions through modulation of protein structure, dynamics and interaction networks. However, despite their utmost relevance in different biological contexts, effects of PTMs at the microscopic level remain poorly understood. While experimental methods typically obtain time- and ensemble-averaged data and are limited in the number of measurable observables, molecular dynamics (MD) simulations provide high-resolution insight into biomolecular properties and behavior, and are in principle an ideally suited tool for addressing this problem.[67-69] In particular, MD simulations involve numerical solution of Newton's equation of motion for a system of particles, given an empirical potential energy function defining interactions in the system. Importantly, in MD, positions and velocities of simulated particles, i.e., microscopic states of the system, can be determined with femtosecond temporal and sub-femtometer spatial resolution. In this way, such simulations can access distributions and time series of any definable physical quantity, which can then be linked to macroscopic thermodynamic and kinetic properties of the system through ensemble averaging. Even though the interactions that govern atomic and molecular motions are of quantum nature, they are in MD approximated by a force field, i.e., a set of classical potential energy terms in combination with associated parameters, typically derived by fitting atomic or molecular properties of small molecules against calculated quantum-mechanical or experimentally measured data. For instance, the functional form (exemplified in equation 1) of a widely-used GROMOS force field,[70-72] also employed in this

thesis, is separated into bonded terms related to interactions between covalently linked atoms and non-bonded (non-covalent) terms describing van der Waals and electrostatic interactions. Specifically, bonded interactions are divided into contributions from covalent bond stretching, bond-angle bending, and improper and torsional dihedral angle terms, mimicked by different types of harmonic-like potentials and trigonometric functions, and corresponding parameters (highlighted in red, equation 1): the force constants K_b, K_Θ and K_ξ, and the equilibrium values b_0, Θ_0 and ξ_0 of the bond, the bond angle and the improper dihedral angle terms, respectively, with K_ϕ, δ and m being the force constant, the phase shift and the multiplicity of the torsional dihedral angle term. On the other hand, interactions of non-bonded atom pairs in the GROMOS force field are described by the Lennard-Jones potential (van der Waals contribution) defined by the depth of the potential well ϵ (highlighted in red) and the distance at which the potential is zero σ (related to the distance at which the potential reaches its minimum, highlighted in red), and by the Coulomb potential (electrostatic contribution from charged species) defined by the atomic partial charges q (highlighted in red). In the equation, b, Θ, ξ, ϕ and r represent the actual values of the bond length, the bond angle, the improper and torsional dihedral angles, and the distance between non-bonded atom pair, respectively, i and j are the summation indices, ϵ_0 is the dielectric permittivity of vacuum and ϵ_1 is the relative permittivity of the solvent, and RF stands for a reaction field contribution to electrostatic interactions.

$$E_{pot} = \sum_{bonds} \frac{K_{b,i}}{4}\left(b_i^2 - b_{i,0}^2\right)^2 + \sum_{angles} \frac{K_{\Theta,i}}{2}\left(\cos\Theta_i - \cos\Theta_{i,0}\right)^2 + \sum_{impropers} \frac{K_{\xi,i}}{2}\left(\xi_i - \xi_{i,0}\right)^2 +$$

$$\sum_{dihedrals} \frac{K_{\phi,i}}{2}\left[1 + \cos\delta_i \cos(m_i\phi_i)\right] + \sum_{atom\ pairs}\left(4\epsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{i,j}} + RF_{i,j}\right) \quad (1)$$

In general, atomic force fields (e.g., GROMOS,[72] AMBER,[73] CHARMM[74] and OPLS[75]) have a very similar form of the interaction function, yet, despite their aim to accurately describe biomolecular systems, do markedly differ in parameter values due to alternative parameterization strategies.[72-75] These approximations and discrepancies notwithstanding, MD simulations have been successfully applied to study biomolecular systems, proteins in particular.[67,68,76,77] Various kinetic and thermodynamic aspects of protein folding have been examined in detail, including folding rates, pathways and intermediate states,

conformational entropy, protein stability and unfolding-refolding under denaturing conditions.[77-89] Additionally, dynamics, partial folding and conformational states of intrinsically disordered proteins have been evaluated through extensive sampling of their free energy landscape by MD simulations.[90-92] Furthermore, protein interactions with different biomolecules such as small chemicals, other peptides and proteins, nucleic acids and lipid membranes have been extensively investigated. In particular, MD has been used to study protein-ligand interactions and binding affinities in the context of allosteric regulation, enzymatic functions and underlying mechanisms, and even rational drug and enzyme design,[93-101] together with protein aggregation, DNA-protein complex formation, and membrane transport and permeation.[102-108] What is more, MD simulations are often combined with experimental methods to achieve better understanding of specific phenomena, improve structure refinement (e.g., X-ray crystallography), explore artifacts and limitations of experimental methods, but also to test and validate force field accuracy.[109-115] Taken together, these studies highlight the power of the MD method to probe microscopic-level dynamics of a wide range of biological processes that are not directly accessible to experimentalists. Importantly, with rapid progress in computer capabilities and advances in software development,[76,116,117] the range of applicability of MD simulations is expected to grow in the years to come, with ever-increasing system sizes and time scales.

As explained above, a large number of different PTMs have been discovered and linked to various essential biological processes. Most studies have focused on identification of modified proteins and target sites or impact of PTMs on gross cellular function, often providing very little detail about the underlying molecular mechanisms. On the other hand, microscopic effects of several PTM-mediated processes have been examined, including MD studies typically focusing on a single modification or in the best case a few modification types for a small subset of proteins.[118-123] However, despite the great importance of understanding how PTMs affect proteins at the atomistic level and the power of the MD method to provide such high-resolution insight, simulations of post-translationally modified proteins lag significantly behind the studies of unmodified, native proteins. This is, arguably, first and foremost due to a general deficiency of suitable computational tools and force field parameters for treating PTMs. To fill this gap, the first principal aim of this thesis has been

to provide a comprehensive, user-friendly platform for studying PTMs using MD simulations by systematic parameterization of a large set of PTMs and development of supporting MD-related modeling tools. Second, using these tools, we have explored the effects of non-enzymatic oxidative modifications on protein structure, dynamics and aggregation, and their potential implications in disease development and aging.

OUTLINE

Despite the wide-spread role and biological importance of PTMs, a systematic framework for treating post-translationally modified proteins by MD simulations has never been developed. In **Chapter I**, we present novel force field parameters for more than 250 different types of enzymatic and non-enzymatic PTMs that have been derived in the context of GROMOS 45a3 and 54a7 parameter sets and cover nearly the complete space of known PTMs. Moreover, these parameters have been validated by reproducing experimentally measured hydration free energies, an important thermodynamic property related to hydrophobicity, for a set of small molecules chemically related to PTMs. Using the reported parameters, physico-chemical properties of modified amino acids were quantified and compared with their canonical counterparts. As a complement to this work, in **Chapter II** we present the Vienna-PTM web server, a publically available tool for automated introduction of desired PTMs to protein 3D structures with subsequent structure optimization via energy-minimization. Vienna-PTM also serves as a repository of the developed parameters described in Chapter I, providing a comprehensive platform for preparing, running and analyzing MD simulations of modified proteins.

Metal-catalyzed carbonylation is one of the most important oxidative modifications of proteins associated with late-onset diseases and aging. In **Chapter III**, we discuss MD simulations used to study the effects of carbonylation on structure and dynamics of the villin headpiece domain, a structurally well-characterized actin-binding polypeptide. Moreover, changes in local hydrophobicity upon carbonylation and their impact on protein aggregation propensity are examined. The relationship between protein aggregation and carbonylation is further investigated in **Chapter IV**. In particular, we performed and analyzed an exhaustive set of MD simulations of the native and carbonylated villin

headpiece domain under various conditions, including different protein and ion concentrations, electrostatics treatments and force fields. In addition to providing direct insight into villin self-association mechanism, these simulations were used to test the limitations of current force fields to accurately describe protein-protein interactions and protein behavior in biologically relevant crowded environments.

Finally, in **Chapter V**, by utilizing MD simulations in combination with the thermodynamic integration approach, we explore how different oxidative modifications affect folding and dimerization free energies of superoxide dismutase 1, a key homo-dimeric antioxidant enzyme in humans linked with amyotrophic lateral sclerosis, a devastating age-related disease. We also discuss potential advances and applications of free energy calculation methods.

REFERENCES

1. Jensen, O. N. Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**, 33-41 (2004).
2. Walsh, C. T. Posttranslational modification of proteins: expanding nature's inventory. Roberts and Company Publishers, Englewood, Colorado, 2006.
3. Kamath, K. S., Vasavada, M. S. & Srivastava, S. Proteomic databases and tools to decipher post-translational modifications. *J Proteomics* **75**, 127-144 (2011).
4. Collins, F. S., Lander, E. S., Rogers, J., Waterston, R. H. & Int Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
5. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342-7372 (2005).
6. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243 (2002).
7. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291-336 (2003).
8. Matlin, A. J., Clark, F. & Smith, C. W. J. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386-398 (2005).
9. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255-261 (2003).
10. Lee, T. Y. *et al.* dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res.* **34**, D622-D627 (2006).
11. UniProt, C. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71-D75 (2012).
12. Levine, R. L. & Stadtman, E. R. Oxidative modification of proteins during aging. *Exp. Gerontol.* **36**, 1495-1502 (2001).
13. Kwon, S. J., Choi, E. Y., Choi, Y. J., Ahn, J. H. & Park, O. K. Proteomics studies of post-translational modifications in plants. *J. Exp. Bot.* **57**, 1547-1551 (2006).
14. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat. Biotechnol.* **24**, 1241-1252 (2006).
15. Sims, R. J., III & Reinberg, D. Is there a code embedded in proteins that is based on post-translational modifications? *Nat. Rev. Mol. Cell Biol.* **9**, 815-820 (2008).
16. Nussinov, R., Tsai, C.-J., Xin, F. & Radivojac, P. Allosteric post-translational modification codes. *Trends Biochem. Sci.* **37**, 447-455 (2012).
17. Lothrop, A. P., Torres, M. P. & Fuchs, S. M. Deciphering post-translational modification codes. *FEBS Lett.* **587**, 1247-1257 (2013).
18. Seger, R. & Krebs, E. G. The MAPK signaling cascade. *FASEB J.* **9**, 726-735 (1995).
19. Ashcroft, M., Kubbutat, M. H. G. & Vousden, K. H. Regulation of p53 function and stability by phosphorylation. *Mol. Cell. Biol.* **19**, 1751-1758 (1999).
20. Cohen, P. The role of protein phosphorylation in human health and disease. *Eur. J. Biochem.* **268**, 5001-5010 (2001).
21. Ciesla, J., Fraczyk, T. & Rode, W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochim. Pol.* **58**, 137-147 (2011).

22. Kivirikk.Ki & Prockop, D. J. Enzymatic hydroxylation of proline and lysine in protocollagen. *Proc. Natl. Acad. Sci. U.S.A.* **57**, 782-& (1967).
23. Martinez-Ruiz, A. & Lamas, S. S-nitrosylation: a potential new paradigm in signal transduction. *Cardiovasc. Res.* **62**, 43-52 (2004).
24. Paik, W. K., Paik, D. C. & Kim, S. Historical review: the field of protein methylation. *Trends Biochem. Sci.* **32**, 146-152 (2007).
25. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: Who, what, and why. *Mol. Cell* **33**, 1-13 (2009).
26. Gorres, K. L. & Raines, R. T. Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **45**, 106-124 (2010).
27. Hwang, C.-S., Shemorry, A. & Varshavsky, A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973-977 (2010).
28. Zhao, S. *et al.* Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science* **327**, 1000-1004 (2010).
29. Anand, P. & Stamler, J. S. Enzymatic mechanisms regulating protein S-nitrosylation: implications in health and disease. *J. Mol. Med.* **90**, 233-244 (2012).
30. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).
31. Jitrapakdee, S. & Wallace, J. C. The biotin enzyme family: Conserved structural motifs and domain rearrangements. *Curr. Protein Pept. Sci.* **4**, 217-229 (2003).
32. Resh, M. D. Membrane targeting of lipid modified signal transduction proteins. *Subcell. Biochem.* **37**, 217-232 (2004).
33. Ikeda, F. & Dikic, I. Atypical ubiquitin chains: new molecular signals - 'Protein modifications: Beyond the usual suspects' review series. *EMBO Rep.* **9**, 536-542 (2008).
34. Komander, D. The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **37**, 937-953 (2009).
35. Latham, J. A. & Dent, S. Y. Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.* **14**, 1017-1024 (2007).
36. van Rossum, B., Fischle, W. & Selenko, P. Asymmetrically modified nucleosomes expand the histone code. *Nat. Struct. Mol. Biol.* **19**, 1064-1066 (2012).
37. Jackson, M. D. & Denu, J. M. Molecular reactions of protein phosphatases - Insights from structure and chemistry. *Chem. Rev.* **101**, 2313-2340 (2001).
38. Grozinger, C. M. & Schreiber, S. L. Deacetylase enzymes: Biological functions and the use of small-molecule inhibitors. *Chem. Biol.* **9**, 3-16 (2002).
39. Davies, K. J. A., Delsignore, M. E. & Lin, S. W. Protein damage and degradation by oxygen radicals. II. Modification of amino-acids. *J. Biol. Chem.* **262**, 9902-9907 (1987).
40. Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**, 20313-20316 (1997).
41. Davies, K. J. A. & Delsignore, M. E. Protein damage and degradation by oxygen radicals. III. Modification of secondary and tertiary structure. *J. Biol. Chem.* **262**, 9908-9913 (1987).
42. Linton, S., Davies, M. J. & Dean, R. T. Protein oxidation and ageing. *Exp. Gerontol.* **36**, 1503-1518 (2001).
43. Squier, T. C. Oxidative stress and protein aggregation during biological aging. *Exp. Gerontol.* **36**, 1539-1550 (2001).
44. Sohal, R. S. Role of oxidative stress and protein oxidation in the aging process. *Free Radic. Biol. Med.* **33**, 37-44 (2002).

45. Garcia-Garcia, A. *et al.* Biomarkers of Protein Oxidation in Human Disease. *Curr. Mol. Med.* **12**, 681-697 (2012).
46. Davies, K. J. Degradation of oxidized proteins by the 20S proteasome. *Biochimie* **83**, 301-310 (2001).
47. Grune, T., Jung, T., Merker, K. & Davies, K. J. Decreased proteolysis caused by protein aggregates, inclusion bodies, plaques, lipofuscin, ceroid, and 'aggresomes' during oxidative stress, aging, and disease. *Int. J. Biochem. Cell Biol.* **36**, 2519-2530 (2004).
48. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10-S17 (2004).
49. Choi, J. *et al.* Oxidative modifications and aggregation of Cu,Zn-superoxide dismutase associated with Alzheimer and Parkinson diseases. *J. Biol. Chem.* **280**, 11648-11655 (2005).
50. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
51. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. & Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **52**, 601-623 (2006).
52. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239-247 (2000).
53. Valko, M. *et al.* Free radicals and antioxidants in normal physiological functions and human disease. *Int. J. Biochem. Cell Biol.* **39**, 44-84 (2007).
54. Perez, V. I. *et al.* Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3059-3064 (2009).
55. Krisko, A. & Radman, M. Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14373-14377 (2010).
56. Levine, R. L. Carbonyl modified proteins in cellular regulation, aging, and disease. *Free Radic. Biol. Med.* **32**, 790-796 (2002).
57. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* **24**, 1311-1317 (2005).
58. Dalle-Donne, I. *et al.* Protein carbonylation, cellular dysfunction, and disease progression. *J. Cell. Mol. Med.* **10**, 389-406 (2006).
59. Andersen, J. K. Oxidative stress in neurodegeneration: cause or consequence? *Nat. Med.* **10**, S18-S25 (2004).
60. Kregel, K. C. & Zhang, H. J. An integrated view of oxidative stress in aging: Basic mechanisms, functional effects, and pathological considerations. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **292**, R18-R36 (2007).
61. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195-1214 (2003).
62. Kirkwood, T. B. L. Understanding the odd science of aging. *Cell* **120**, 437-447 (2005).
63. Lin, M. T. & Beal, M. F. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* **443**, 787-795 (2006).
64. Sohal, R. S., Agarwal, S., Dubey, A. & Orr, W. C. Protein oxidative damage is associated with life expectancy of houseflies. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7255-7259 (1993).
65. Gems, D. & Partridge, L. Stress-response hormesis and aging: "That which does not kill us makes us stronger". *Cell Metab.* **7**, 200-203 (2008).
66. Cai, Z. & Yan, L. J. Protein oxidative modifications: Beneficial roles in disease and health. *Journal of biochemical and pharmacological research* **1**, 15-26 (2013).

67. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646-652 (2002).
68. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
69. Kruschel, D. & Zagrovic, B. Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol Biosyst* **5**, 1606-1616 (2009).
70. Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
71. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676 (2004).
72. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
73. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
74. Chen, J. H., Im, W. P. & Brooks, C. L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **128**, 3728-3736 (2006).
75. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* **105**, 6474-6487 (2001).
76. Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S. & Xiao, X. Biomolecular modeling and simulation: a field coming of age. *Q. Rev. Biophys.* **44**, 191-228 (2011).
77. Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**, 52-61 (2012).
78. Duan, Y. & Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740-744 (1998).
79. Daura, X., van Gunsteren, W. F. & Mark, A. E. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. *Proteins* **34**, 269-280 (1999).
80. Schafer, H., Mark, A. E. & van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. *J Chem Phys* **113**, 7809-7817 (2000).
81. Snow, C. D., Nguyen, N., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102-106 (2002).
82. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927-937 (2002).
83. Pande, V. S. *et al.* Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **68**, 91-109 (2003).
84. Muff, S. & Caflisch, A. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein. *Proteins* **70**, 1185-1195 (2008).
85. Baron, R., Huenenberger, P. H. & McCammon, J. A. Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: Correction terms and convergence properties. *J. Chem. Theory. Comput.* **5**, 3150-3160 (2009).
86. Lin, Z., Kornfeld, J., Maechler, M. & van Gunsteren, W. F. Prediction of folding equilibria of differently substituted peptides using one-step perturbation. *J. Am. Chem. Soc.* **132**, 7276-7278 (2010).
87. Seeliger, D. & de Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **98**, 2309-2316 (2010).

88. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341-346 (2010).
89. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517-520 (2011).
90. Cino, E. A., Wong-ekkabut, J., Karttunen, M. & Choy, W.-Y. Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response. *PLoS ONE* **6**, e27371 (2011).
91. Phillips, J. L., Colvin, M. E. & Newsam, S. Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinformatics* **12** (2011).
92. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134**, 3787-3791 (2012).
93. Piana, S., Carloni, P. & Parrinello, M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.* **319**, 567-583 (2002).
94. Liang, S. *et al.* Exploring the molecular design of protein interaction sites with molecular dynamics simulations and free energy calculations. *Biochemistry* **48**, 399-414 (2009).
95. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* **21**, 150-160 (2011).
96. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9** (2011).
97. Grant, B. J. *et al.* Novel allosteric sites on Ras for lead generation. *PLoS ONE* **6** (2011).
98. Huang, D. & Caflisch, A. The free energy landscape of small molecule unbinding. *PLoS Comp. Biol.* **7** (2011).
99. de Ruiter, A. & Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **15**, 547-552 (2011).
100. Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3790-3795 (2012).
101. Shuai Liu *et al.* Lead Optimization Mapper: Automating free energy calculations for lead optimization. *In preparation*.
102. Gsponer, J., Haberthur, U. & Caflisch, A. The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5154-5159 (2003).
103. Noskov, S. Y., Berneche, S. & Roux, B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* **431**, 830-834 (2004).
104. Ma, B. & Nussinov, R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol.* **10**, 445-452 (2006).
105. Pan, Y. & Nussinov, R. Structural basis for p53 binding-induced DNA bending. *J. Biol. Chem.* **282**, 691-699 (2007).
106. MacKerell, A. D., Jr. & Nilsson, L. Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.* **18**, 194-199 (2008).
107. Khalili-Araghi, F. *et al.* Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **19**, 128-137 (2009).
108. Chen, R. & Mark, A. E. The effect of membrane curvature on the conformation of antimicrobial peptides: Implications for binding and the mechanism of action. *Eur. Biophys. J.* **40**, 545-553 (2011).

109. Gros, P., van Gunsteren, W. F. & Hol, W. G. J. Inclusion of thermal motion in crystallographic structures by restrained molecular-dynamics. *Science* **249**, 1149-1152 (1990).
110. Soares, T. A., Daura, X., Oostenbrink, C., Smith, L. J. & van Gunsteren, W. F. Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme. *J. Biomol. NMR* **30**, 407-422 (2004).
111. Zagrovic, B. & van Gunsteren, W. F. Comparing atomistic simulation data with the NMR experiment: How much can NOEs actually tell us? *Proteins* **63**, 210-218 (2006).
112. van Gunsteren, W. F., Dolenc, J. & Mark, A. E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **18**, 149-153 (2008).
113. Kuzmanic, A., Kruschel, D., van Gunsteren, W. F., Pannu, N. S. & Zagrovic, B. Dynamics may significantly influence the estimation of interatomic distances in biomolecular X-ray structures. *J. Mol. Biol.* **411**, 286-297 (2011).
114. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS ONE* **7** (2012).
115. Nygaard, R. *et al.* The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell* **152**, 532-542 (2013).
116. Shaw, D. E. *et al.* Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* **51**, 91-97 (2008).
117. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854 (2013).
118. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429-2433 (2001).
119. Bottomley, M. J. Structures of protein domains that create or recognize histone modifications. *EMBO Rep.* **5**, 464-469 (2004).
120. Grauffel, C., Stote, R. H. & Dejaegere, A. Force field parameters for the simulation of modified histone tails. *J. Comput. Chem.* **31**, 2434-2451 (2010).
121. Polyansky, A. A. & Zagrovic, B. Protein electrostatic properties predefining the level of surface hydrophobicity change upon phosphorylation. *J Phys Chem Lett* **3**, 973-976 (2012).
122. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via lys-16 acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17857-17862 (2012).
123. Seeliger, D. *et al.* Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem. Biol.* **7**, 150-154 (2012).

Chapter I

A systematic framework for molecular dynamics simulations of protein post-translational modifications

Petrov, D.,* Margreitter, C.,* Grandits, M., Oostenbrink, C. & Zagrovic, B. (2013). *PLoS Comput. Biol.* **9** (7), e1003154.

*The authors contribute equally to this work.

DP, CM and BZ conceived and designed the experiments. DP, CM, MG, CO and BZ performed the experiments. DP, CO and BZ analyzed the data. DP, CM, MG, CO and BZ wrote the paper.

ABSTRACT

By directly affecting structure, dynamics and interaction networks of their targets, post-translational modifications (PTMs) of proteins play a key role in different cellular processes ranging from enzymatic activation to regulation of signal transduction to cell-cycle control. Despite the great importance of understanding how PTMs affect proteins at the atomistic level, a systematic framework for treating post-translationally modified amino acids by molecular dynamics (MD) simulations, a premier high-resolution computational biology tool, has never been developed. Here, we report and validate force field parameters (GROMOS 45a3 and 54a7) required to run and analyze MD simulations of more than 250 different types of enzymatic and non-enzymatic PTMs. The newly developed GROMOS 54a7 parameters in particular exhibit near chemical accuracy in matching experimentally measured hydration free energies (RMSE = 4.2 kJ/mol over the validation set). Using this tool, we quantitatively show that the majority of PTMs greatly alter the hydrophobicity and other physico-chemical properties of target amino acids, with the extent of change in many cases being comparable to the complete range spanned by native amino acids.

AUTHOR SUMMARY

Post-translational modifications, i.e., chemical changes of protein amino acids, play a key role in different cellular processes, ranging from enzymatic activation to transcription and translation regulation to disease development and aging. However, our understanding of their effects on protein structure, dynamics and interaction networks at the atomistic level is still largely incomplete. In particular, molecular dynamics simulations, despite their power to provide a high-resolution insight into biomolecular function and underlying mechanisms, have been limited to unmodified, native proteins due to a surprising deficiency of suitable tools and systematically developed parameters for treating modified proteins. To fill this gap, we develop and validate force field parameters, an essential part of the molecular dynamics method, for more than 250 different types of enzymatic and non-enzymatic post-translational modifications. Additionally, using this tool, we quantitatively show that microscopic properties of target amino acids, such as hydrophobicity, are greatly affected by the majority of modifications. The parameters presented in this study greatly expand the range of applicability of computational methods, and in particular molecular dynamics

simulations, to a large set of new systems with utmost biological and biomedical importance.

INTRODUCTION

Proteins in the cell continually get covalently modified in different post-translational, enzyme-controlled reactions.[1-3] Additionally, protein modifications frequently arise in a non-controlled fashion as well, mainly as a consequence of oxidative stress.[4] While enzymatic post-translational modifications (PTMs) play important regulatory roles in a large number of different cellular processes, non-enzymatic PTMs are predominantly linked with protein damage and are involved in age-related diseases such as neurodegenerative disorders, diabetes and cancer.[2,4-7] Despite the general importance of PTMs in different biological contexts, their effect on protein structure, dynamics and interaction networks at the atomistic level remains poorly understood. In particular, molecular dynamics (MD) simulations, a widely used high-resolution computational method for studying biomolecular properties and behavior,[8-10] have been limited to unmodified, native proteins due to a surprising deficiency of suitable tools and systematically developed parameters for treating PTMs, with only sporadic exceptions.[11-16]

MD simulations capture atomic and molecular motions based on Newton's equation of motion and an empirical potential energy function that defines interactions between simulated particles. The latter is defined by a force field, i.e., a self-consistent set of physically realistic equations and semi-empirical parameters describing all interactions in a given system. Force-field parameters are typically obtained by fitting atomic or molecular properties of small molecules against calculated quantum-mechanical or experimentally measured data. As the applied parameterization strategies often differ from each other, considerably different parameter values have been derived in many cases.[17-20] Here, we develop force field parameters for over 250 different types of enzymatic and non-enzymatic modifications of amino-acid side chains as well as protein termini within the context of GROMOS 45a3[19] and 54a7[21,22] force fields (Table S1). We choose GROMOS force fields because of their widespread usage, high accuracy in reproducing experimental results and general transferability of parameters when it comes to identical chemical groups in different compounds[21] (e.g., from the hydroxyl group of tyrosine to the hydroxyl group of 7-hydroxytryptophan). The functional form of a typical force field is exemplified in equation 1 for GROMOS class force fields,

$$E_{pot} = \sum_{bonds} \frac{\boldsymbol{K_{b,i}}}{4}\left(b_i^2 - \boldsymbol{b_{i,0}^2}\right)^2 + \sum_{angles} \frac{\boldsymbol{K_{\Theta,i}}}{2}\left(\cos\Theta_i - \cos\boldsymbol{\Theta_{i,0}}\right)^2 + \sum_{impropers} \frac{\boldsymbol{K_{\xi,i}}}{2}\left(\xi_i - \boldsymbol{\xi_{i,0}}\right)^2 +$$

$$\sum_{dihedrals} \frac{\boldsymbol{K_{\phi,i}}}{2}[1 + \cos\boldsymbol{\delta_i}\cos(\boldsymbol{m_i}\phi_i)] + \sum_{atom\ pairs}\left(4\boldsymbol{\epsilon_{i,j}}\left[\left(\frac{\boldsymbol{\sigma_{i,j}}}{r_{i,j}}\right)^{12} - \left(\frac{\boldsymbol{\sigma_{i,j}}}{r_{i,j}}\right)^{6}\right] + \frac{\boldsymbol{q_i q_j}}{4\pi\epsilon_0\epsilon_1 r_{i,j}} + RF_{i,j}\right), \quad (1)$$

with parameters highlighted using boldface letters and RF representing a reaction field contribution to the electrostatic interactions. The non-bonded interaction terms in the GROMOS force field are primarily parameterized against thermodynamic data of small molecules, either in the pure liquid state, or in aqueous or nonpolar solution. Therefore, we validate the obtained parameters by reproducing experimental hydration free energies (HFEs), a measure of hydrophobicity and arguably one of the most important amino-acid properties with implications in protein folding, ligand binding or protein-lipid interactions. Finally, we analyze physico-chemical properties related to hydrophobicity of all parameterized PTMs according to their type and compare them against the 20 canonical amino acids.

RESULTS

Parameterization of PTMs

One of the principal objectives in our parameterization has been the coverage of experimentally known PTMs, which is as complete as possible. Following an exhaustive literature search and analysis of an online PTM database PTMdb,[23] we have compiled a diverse list of enzymatic and non-enzymatic PTMs, including phosphorylation, methylation, acetylation, hydroxylation, carboxylation, carbonylation, nitration, deamidation and many others (Figure 1a, Table S1), covering a total of 259 distinct PTM reactions or 110 non-redundant post-translationally modified amino acids and protein termini. The lower number in the latter case reflects the fact that different PTM reactions can lead to the same modified product (e.g., glutamic semialdehyde is a product of both arginine and proline carbonylation). We have generated GROMOS 45a3 (Dataset S1) and 54a7 (Dataset S2) force field parameters for the non-redundant set of compounds by either direct transfer or analogy to already parameterized compounds including amino acids, nitrogenous bases and other small molecules or completely novel parameterization (see Methods for more details).

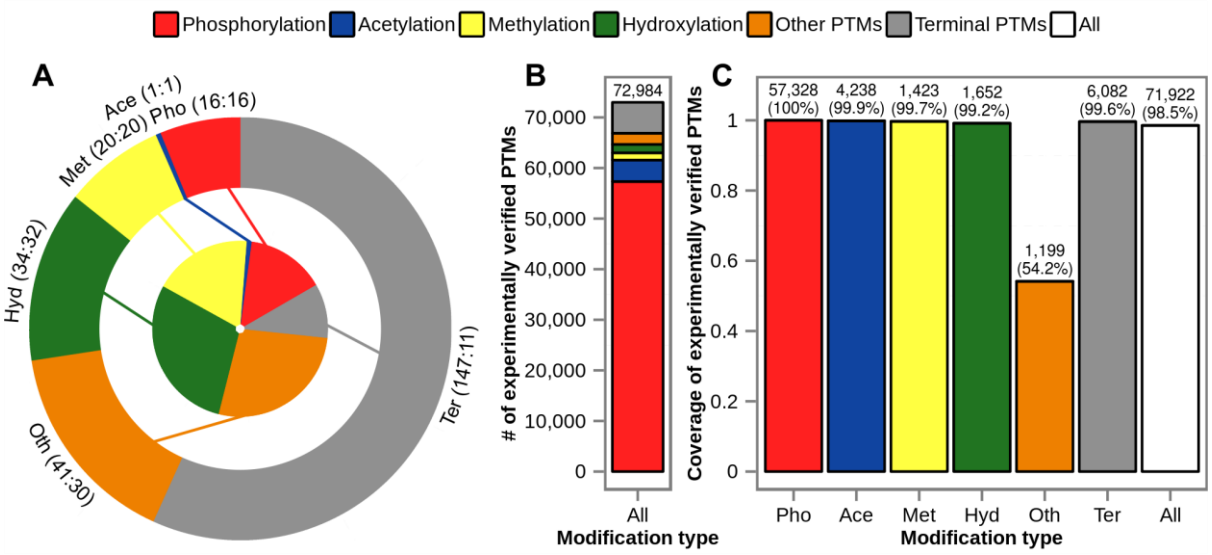


Figure 1. Summary of the number and coverage of parameterized PTMs. a) the number of parameterized PTMs by type (outer annulus) together with the number of parameterized non-redundant compounds by type (inner circle), labeled accordingly (number of PTMs: number of compounds); b) the number of experimentally verified PTMs by type annotated in the UniProt database (total of 72,984); c) coverage of experimentally verified PTMs shown as percentages with the values and the number of covered modifications displayed (top of bars). Color code: phosphorylation-red, acetylation-blue, methylation-yellow, hydroxylation-green, other PTMs-orange, terminal PTMs-gray and all-white.

How well do the obtained parameters cover the space of biologically relevant PTMs? To address this question, we have analyzed PTMs that have been experimentally verified (72,984) and annotated as such in the UniProt database[24] (21,411 protein entries, Dataset S3). Phosphorylation is by-far the most abundant modification type in the UniProt database (78.5% of all UniProt PTMs), followed by acetylation, hydroxylation and methylation (Figure 1b). Note that terminal PTMs account for a sizable fraction of all annotated modification at 8.3%. Strikingly, the parameterized compounds reported herein match every annotated phosphorylation modification, 99.9% of acetylation, 99.2% of hydroxylation and 99.7% of methylation modifications, for a grand-total coverage of 98.5% of all PTMs reported in UniProt (Figure 1c). Concerning PTMs that are not covered by our parameters, they are all extremely rare, each accounting for less than 0.5% of all UniProt PTMs. Finally, we provide parameters for 33 PTMs (Table S1), mostly non-enzymatic ones, that have to date not been reported in UniProt.

Validation against experimental HFEs

HFE, a free energy difference between a compound solvated in water and the same compound in the gas phase, is an experimentally measurable property related to hydrophobicity, and it has been originally used to re-parameterize the GROMOS force field in 2004.[21] A proper description of functional groups in the hydrated phase is of crucial importance for virtually all relevant biomolecular processes, so we have used the same thermodynamic quantity to validate the parameters obtained in the present study. To the best of our knowledge, experimental HFEs are available for the exact side chain analogs of 13 parameterized PTMs only and we have therefore in the validation set also included compounds, which are chemically related to PTM side chains for which no experimental HFEs were available, for a total of 26 different molecules (only a single representative compound was included for each group of PTMs involving the same chemical moiety, Table 1). Note that the additional compounds related to PTM side chains have been parameterized in the same way as the relevant PTMs.

Table 1. HFEs of the molecules in the validation set: comparison between experimental and calculated values using the GROMOS 54a7 parameter set.

Compound	HFE (kJ/mol)	
	experimental	ffG54a7
Validation set. PTM-side-chain analogs		
N-butylacetamide	-39.0	-37.9
o-cresol	-24.6	-24.8
m-cresol	-23.0	-28.2
2-methyl-2-propanol	-18.7	-15.1
2-methyl-1-propanol	-18.8	-18.6
propan-2-ol	-19.8	-16.1
N-methylacetamine	-41.9	-38.0
methylpropanoate	-12.3	-5.8
methylacetate	-13.1	-8.7
dimethylsulfide	-6.7	-9.1
butanal	-13.3	-10.8
propanal	-14.4	-11.8

butane	8.7	9.5
Validation set. Compounds similar to PTM-side-chain analogs		
diethylamine	-17.0	-11.7
trimethylamine	-13.4	-4.7
ethene	5.4	13.8
bromobenzene	-6.1	-8.6
aniline	-23.0	-25.8
acetophenone	-19.2	-16.3
N-methylformamide	-41.9	-39.5
chlorophenol	-19.0	-22.8
*2-nitrophenol	-19.2	-33.8
nitrobenzene	-17.2	-15.6
acetone	-16.1	-8.5
dimethylsulfoxide	-42.3	-39.4
methylsulfonylmethane	-42.2	-41.9
RMSE	**-**	**4.2**
Additional compounds		
#2-nitrophenol	-19.2	-17.4
3-nitrophenol	-40.3	-43.0
4-nitrophenol	-44.5	-44.2
4-methylimidazole (Nδ-H)	-42.9	-46.7
4-methylimidazole (Nε-H)		-62.0
1-methylimidazole (Nδ-H)	-35.2	-25.2
1-methylimidazole (Nε-H)		-34.5

The outlier 2-nitrophenol described by: *parameters used for other nitro-containing compounds, and #parameters derived to match the experimental HFE.

Experimental HFEs are taken from refs.[21,48-50]

We have used MD simulations and the thermodynamic integration (TI) approach[25] (see *Methods* for more details) to calculate the HFEs for neutral forms small-molecule analogs of the canonical amino-acid side chains and for the compounds in the validation set using both the 45a3 (Table S2) and 54a7 (Table 1) parameter sets of the GROMOS force field. As a consequence of the parameterization strategy behind them, the canonical amino acids

exhibit an excellent agreement with experimental HFEs when it comes to the 54a7 parameter set, with a root-mean-square error (RMSE) of 3.3 kJ/mol (RT = 2.5 kJ/mol at room temperature) and an almost perfect correlation with experimental HFEs (correlation coefficient R^2 = 0.98) (Figure 2). Remarkably, the newly generated GROMOS 54a7 force field parameters of PTM-related compounds exhibit a nearly equal level of matching of experimental HFEs with an RMSE of 4.2 kJ/mol (Table 1) and a correlation coefficient R^2 of 0.94 (Figure 2) over 25 different compounds, excluding a single outlier, 2-nitrophenol (Figure 2, red X symbol). This compound, containing nitro and hydroxyl groups attached to a benzene ring, deviates from the experimental value by 14.6 kJ/mol.

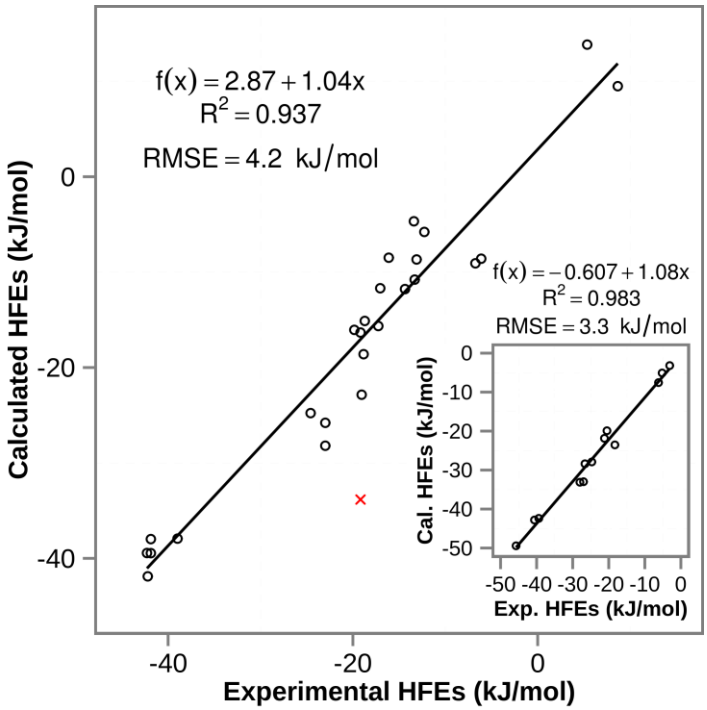


Figure 2. Experimental vs. calculated HFEs of compounds from the validation set (GROMOS 54a7). Correlation is captured by the regression line, its parameters, Pearson correlation coefficient and overall RMSE, with the outlier 2-nitrophenol in red (X symbol). The same comparison for canonical amino acids is shown in the inset. Note that error bars of calculated HFEs are comparable to the size of the symbols, with the average standard error of 0.4 kJ/mol.

Considering the outlier 2-nitrophenol in more detail, additional calculations have shown that *p*-cresol (a tyrosine side-chain analog), *o*-cresol, *m*-cresol and nitrobenzene, compounds containing either a hydroxyl group or a nitro group attached to a benzene ring, agree well with experimental HFEs with an overall RMSE of 2.7 kJ/mol only. This suggests

that, although parameters of individual groups do reproduce experimental HFEs, the agreement with experiment may significantly worsen if they appear in combination. In order to test this, we have calculated HFEs of 3- and 4-nitrophenol and compared them against experimental values. Interestingly, the calculated HFEs of both compounds match experimental values (Table 1) suggesting either that these groups exert a specific influence on each other only in 2-nitrophenol or that the experimentally measured HFE may simply not be reliable for this compound. To account for the former possibility, we have derived a set of parameters *de novo* for 2-nitrophenol that closely match its experimental HFE with an absolute value of the deviation of 1.8 kJ/mol (Table 1). Note that we report both versions of nitrotyrosine (Table S1), a cognate PTM to 2-nitrophenol.

Finally, we have also excluded 4-methylimidazole (a histidine side-chain analog) and 1-methylimidazole from the HFE analysis of the canonical amino acids and PTMs, respectively, even though experimental HFEs are available for both compounds. Since histidine exists in two tautomeric states, described by different parameters, the calculated HFE depends on the choice of the state used for calculations, with one matching the experimental HFE and the other varying by approximately 20 kJ/mol (Table 1). Consequently, the same problem exists for 1'- and 3'-methylhistidine, whose parameters are based on those of histidine, where one tautomer matches while the other deviates from the experimental HFE (Table 1).

In contrast to GROMOS 54a7, the 45a3 parameter set does not reproduce experimental HFEs well (Table S2 and Figure S1). Namely, the slope of 0.79 and the offset of 3.8 kJ/mol of the regression line suggest that the calculated HFEs are largely overestimated (RMSE = 10.8 kJ/mol) for the amino-acid side chain analogs, as observed previously.[21] The same effect persists for the PTM compounds, with a RMSE from experimental HFEs of 15 kJ/mol (Figure S1). As the GROMOS 45a3 parameter set was not parameterized to match experimental HFEs for polar compounds, such level of deviation was to be expected.

Due to a lack of pertinent experimental data, seven parameterized PTMs (carboxylysine, homocitrulline, citrulline, S-carbamoyl-cysteine, S-nitrosocysteine, 2-oxo-histidine and pyruvic acid) have remained unrepresented in the validation set, and therefore unverified in terms of reproducing experimental HFEs. To further assess the quality of the parameters for these compounds, we have compared them to those obtained by the Automated Topology

Builder,[26] a widely used online service for automated parameterization of small molecules compatible with the GROMOS 54a7 force field. While manually curated approaches are arguably superior to automated ones, it is reassuring to see that the two sets of parameters match closely. For example, we have observed close agreement between the sets of partial charges obtained using the two methods for these seven compounds, with a Pearson correlation coefficient *R* of 0.93 and an overall RMSD of 0.2 e^-.

Comparison of physico-chemical properties of PTMs and canonical amino acids

As an application of the newly developed PTM parameters, we focus on the changes in several key physico-chemical properties of amino acids introduced by PTMs. Interestingly, the majority of post-translationally modified amino acids are larger in size than their native counterparts, with more than 85% of PTMs increasing the molecular weight and more than 80% of PTMs increasing the solvent accessible surface area (SASA) of the affected residues (Table S3) as calculated on energy-minimized (using the GROMOS 54a7 parameter set) configurations of PTMs and canonical amino acids. What is more, PTMs introduce significant changes in the electrostatic properties of target residues as illustrated in the case of net charge and dipole moment (Table S3). For example, 42% of all PTMs studied here undergo a charge change of 1 e^- or more in absolute value, with 88% of such changes resulting in a more negatively charged species. Moreover, the average absolute value of the change in dipole moment upon PTM equals 1.7 Debye, which is comparable in magnitude to the average dipole moment of 2.7 Debye or its standard deviation of 1.9 Debye as calculated in both cases over all unmodified residues using GROMOS 54a7 parameters and energy-minimized configurations. Finally, given the general importance of hydrophobicity in various biological processes, it is critical to understand in a quantitative manner how PTMs modulate the hydrophobicity of target amino acids. To address this question, we have used TI and GROMOS 54a7 parameters to calculate HFEs of all parameterized PTMs in neutral protonation states, since the available experimental data is insufficient for such an analysis. Our results show that methylation and carbonylation modifications increase HFEs on average by 18.6 kJ/mol and 20.5 kJ/mol, respectively, while hydroxylation modifications exhibit an opposite effect and decrease HFEs by on average 25.1 kJ/mol (Figure 3a). These changes are extremely relevant if one considers the fact that the two central quartiles of the

distribution of HFEs for canonical amino acids span the range from approximately -40 kJ/mol to -20 kJ/mol (Figure 3a). Furthermore, the most extreme cases, i.e., symmetric di-methylation of arginine (ΔHFE = 46.2 kJ/mol) and di-hydroxylation of phenylalanine (ΔHFE = -60.3 kJ/mol) are comparable in absolute values to the total span of the canonical amino acid HFEs (-49.4 kJ/mol to -3.2 kJ/mol, Figure 3a). In other words, the effect of some PTMs on the HFEs of target amino acids is as large as the difference which would arise by mutating the most hydrophobic to the most hydrophilic canonical amino acid or vice versa. While some of these effects agree well with what one would qualitatively expect, for a number of PTMs our results are the first to provide a quantitative framework for such an analysis.

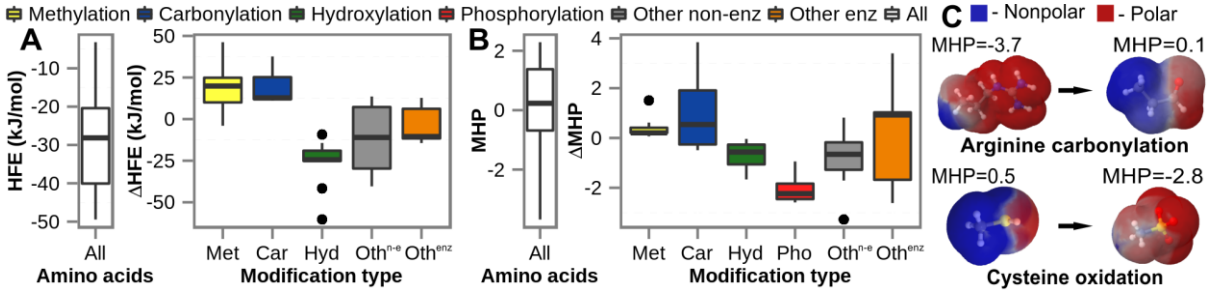


Figure 3. Hydrophobicity-related properties of PTMs compared to canonical amino acids. a) hydration free energies (HFEs) and b) molecular hydrophobicity potentials (MHPs). Distributions calculated of HFEs and MHPs of the canonical amino acids are captured using white boxes on the left side of both a) and b) panels. The distributions of HFE and MHP changes upon different types of PTMs are shown in colored boxes sorted according to the median of the underlying distributions. The distributions are shown using the box-and-whisker plotting method. Color code: methylation-yellow, carbonylation-blue, hydroxylation-green, phosphorylation-red, other enzymatic modifications-gray, other non-enzymatic modification-orange and all-white; c) change in surface MHP upon arginine carbonylation and cysteine oxidation, modifications with the most positive and the most negative MHP change, respectively. Note that we have not taken N-acetylglucosamine into account for the HFE and MHP analysis, since glycosylation modifications predominantly result in carbohydrate chains attached to target residues, while we provide parameters for this carbohydrate only as the first one in a typical chain.

As both calculation and experimental measurement of HFEs are limited to neutral compounds only, the above analysis does not take into account charged modifications such as phosphorylation. To address this, we have used the molecular hydrophobicity potential (MHP)[27] approach to estimate hydrophobicity of all parameterized PTMs using their protonation states at physiological pH. MHP values are semi-empirical estimates of *logP*, a

given compound's partition coefficient between water and the non-polar solvent octanol and are widely used in computational drug design.[28,29] Similarly to the HFEs analysis, MHP calculations show that carbonylation and methylation are hydrophobicity-increasing modifications (Figure 3b), in contrast to phosphorylation and hydroxylation, which are hydrophilicity-increasing modifications. Finally, this analysis shows that PTMs can drastically change hydrophobic/hydrophilic properties of affected residues, e.g., arginine carbonylation shifts a highly hydrophilic to a highly hydrophobic residue, while cysteine oxidation does exactly the opposite (Figure 3c). By changing the chemical nature of affected residues, PTMs frequently completely alter their physico-chemical properties such as hydrophobicity, a feature with potentially far-reaching biological implications.[11,12,30]

DISCUSSION

Despite the importance of understanding PTMs at the molecular level, MD simulations of post-translationally modified proteins lag significantly behind the studies of unmodified proteins, and this seems primarily due to a general lack of suitable computational tools and simulation parameters for treating PTMs. This study is to the best of our knowledge the first-ever effort to develop force-field parameters for the large majority of known PTMs in a systematic fashion. We have generated GROMOS force field (45a3 and 54a7) parameters for over 250 different enzymatic and non-enzymatic PTMs, spanning a wide range of modification types with a close to complete coverage of experimentally verified PTMs (Figure 1). Since GROMOS 54a7 force field parameters were fitted to reproduce experimental HFEs, we have tested the quality of the PTM parameters, obtained by manually curating the parameters of different groups mostly in analogy to canonical amino acids, by comparing the calculated HFEs against the experimental values. The newly generated parameters compatible with the GROMOS 54a7 parameter set reproduce experimental HFEs almost equally well as the original ones (Table 1 and Figure 2). Overall, only a few parameterized PTMs have not been directly validated against experimental HFEs due to a lack of experimentally available data. In those cases, however, good matching with the parameters obtained using an orthogonal, fully automated approach[26] lends support to the general validity of the reported parameters. However, one should emphasize that the full range of validity of the presented parameters could and should be delineated only by

directly comparing MD simulations of different post-translationally modified proteins in biologically relevant contexts with relevant experimental data.

To date, PTMs in MD simulations have been treated in separate studies using different procedures and force fields, typically focusing on a single modification at a time.[11,13,16] Additionally, there are some available tools for automated generation of parameters (e.g., the AMBER[31] feature *antechamber* and online tools SwissParam[32], PRODRG[33], ATB[26] and *q4md-forcefieldtools*[34]), however, envisioned for small molecules rather than protein PTMs. The parameters reported herein have comparative advantage over these sources along three principal directions. First, we provide exclusively human curated and validated PTM force-field parameters, which are mutually fully consistent as well as being consistent with canonical amino acids. Second, we provide PTM parameters in both GROMOS[35] and GROMACS[36] format, widely used MD simulation packages (supporting GROMOS version 11 and GROMACS versions 3.x and newer), suitable for immediate simulation of modified proteins without any additional work required. This should be contrasted with the above tools that provide parameters for isolated compounds only. Finally, in combination with a publicly available online tool for introducing PTMs of choice to a user-supplied protein 3D structure (Vienna-PTM server, http://vienna-ptm.univie.ac.at),[37] we provide a comprehensive, user-friendly toolkit for studying PTMs using MD simulations.

During their lifecycle in the cell, almost all proteins undergo one or more different PTMs affecting their structure, dynamics and interaction networks and, subsequently, their function through direct alteration of chemical and physico-chemical properties of target residues (Figure 3). The force field parameters presented here, together with the Vienna-PTM webserver, provide a systematic framework required to study the effects of PTMs using MD simulations. As a first step in this direction, we have here compared the hydrophobicity-related variables (HFEs and MFP values) of native and modified amino acids and quantitatively showed that PTMs can have an extremely strong, biologically significant effect in this context. It has already been documented that some PTMs exert their biological effect through a general modification of the hydrophobicity of their targets. For example, lysine trimethylation is known to directly affect the binding of retinoic acid receptors, which regulate genes involved in growth, differentiation and apoptosis, to their partners via an

increase in site-specific hydrophobicity.[38] Moreover, acetylated and methylated lysine residues in histones, i.e., some of the key components of the histone code, are recognized by the hydrophobic binding pockets of bromo- and chromo-domains based on the difference in hydrophobicity between the modified and unmodified lysines.[39] Furthermore, we have recently shown that carbonylation, which affects lysine, arginine, proline and threonine residues, drastically increases local propensity for aggregation in proteins by affecting the hydrophobicity of the modified sites.[11] While other, more specific effects of PTMs on the structure, dynamics and interaction profile of target proteins are certainly important, a major change in hydrophobicity, net charge, isoelectric point or any other general physico-chemical property caused by a PTM at a given site could certainly have major biological repercussions. We believe that our present study will provide a solid foundation for exploring such timely and important issues in the future. However, this is only one possible application of the PTM force-field parameters reported herein. From direct MD simulations to biomolecular structure refinement to computational free energy estimation and drug design, these parameters expand the range of MD methodology to a large class of biomolecular systems of paramount importance. It is our hope that this advance will play a catalytic role in bringing together realistic cell biology, dominated by PTMs, and the quantitative, reductionist power of structural biology and chemistry, as embodied in the MD method, and help shed light on a broad spectrum of important biological questions at the microscopic level.

METHODS

Parameterization of PTMs

One of the aims of the GROMOS force fields is to allow for the transfer of parameters between chemically similar groups in different compounds. Accordingly, we have derived GROMOS 45a3 and 54a7 force field parameters describing 110 post-translationally modified amino acids and protein termini (Table S1) by either novel parameterization or direct transfer from or analogy to already parameterized compounds including amino acids, nitrogenous bases and other small molecules according to the following principles and rationales.

General principles:

- Parameters were directly transferred from chemically identical groups (e.g., from the hydroxyl group of tyrosine to the hydroxyl group of 7-hydroxytryptophan) if such exist among parameterized compounds. If not, parameters were either directly transferred or inferred by analogy to the chemically most similar parameterized compound.
- Partial charges were assigned to add up to an integer net charge for every charge group, primarily by adjusting partial charges of less exposed atoms (e.g., the phosphorus atom of phospho-residues), while keeping them intact for terminal, more exposed atoms to affect interactions with other compounds as little as possible.

Modification type-specific principles:

1) PHOSPHORYLATION: Parameters directly transferred from phosphate and hydroxyl groups of nucleotides (e.g., ATP). The partial charge on the phosphorus atom fixed to get an integral net charge of a parameterized compound (dependent on the protonation state). The rest of a parameterized compound left unchanged. Additionally, analogy to the ester group reported by Chandrasekhar and others[40] used for phosphoaspartate.
2) METHYLATION: Parameters directly transferred or derived by analogy to different methyl-containing groups depending on the net charge and chemical context as follows:
    a. directly transferred or derived by analogy from amines reported by Oostenbrink and others[41] for methylated lysine and histidine residues,
    b. directly transferred or derived by analogy from nucleotides (e.g., ATP), arginine and amines reported by Oostenbrink and others[41] for methylated arginine residues,
    c. derived by analogy to the peptide bond and the cognate native residues for methyl-asparagine and methyl-glutamine,
    d. directly transferred from the ester group reported by Chandrasekhar and others[40] for aspartate methyl ester and glutamate methyl ester,

e. directly transferred from methionine for S-methylcysteine.

3) ACETYLATION: Parameters derived by analogy to the peptide bond and the carboxamide group (e.g., glutamine).
4) HYDROXYLATION: Parameters directly transferred from the hydroxyl group of threonine or tyrosine, if attached to an aliphatic or aromatic carbon atom, respectively.
5) CARBOXYLATION: Parameters directly transferred from the carboxyl group (e.g., glutamate).
6) SULFATION: Parameters derived by analogy to the phosphate group of nucleotides (e.g., adenosine).
7) DEHYDRATION: Parameters directly transferred from aliphatic carbon atoms using a bond type with a shorter equilibrium distance to mimic the properties of the double bond.
8) BROMIDATION: Parameters directly transferred from 8-bromo-guanosine triphosphate reported by Hritz and Oostenbrink.[42]
9) S-NYTROSILATION: The oxygen atom parameters directly transferred for the carbonyl group (e.g., the peptide bond), with the nitrogen and sulfur atom partial charges fixed to add up to 0 net charge.
10) CITRULLINATION: Parameters derived by analogy to the peptide bond and the carboxamide group (e.g., glutamine).
11) ALLYSINE FORMATION: The oxygen atom parameters directly transferred for the carbonyl group (e.g., glutamine), with the carbon and hydrogen atom derived by analogy to the aldehyde group reported by Dolenc and others.[43]
12) GLYCOSYLATION: Parameters directly transferred from the peptide bond and monosaccharide molecules (e.g., glucose).
13) CARBONYLATION: The oxygen atom parameters directly transferred for the carbonyl group (e.g., glutamine), with the carbon and hydrogen atom derived by analogy to the aldehyde group reported by Dolenc and others.[43]
14) OXIDATION: Parameters directly transferred from different oxygen-containing groups depending on the net charge and chemical context:

a. from the carbonyl group (e.g., glutamine) and the phosphate group of nucleotides (e.g., adenosine) for methionine sulfoxide and methionine sulfone, respectively, with the partial charge of the sulphur atom fixed to get 0 net charge for oxidative modifications of methionine,
b. from the carboxyl group (e.g., glutamate) for cysteine oxidation modifications,
c. from the carbonyl group (e.g., glutamine) for the remaining oxidation modifications.

15) NITRATION: The oxygen atom parameters directly transferred from different oxygen-containing groups, with the nitrogen and carbon atoms partial charges adjusted to add up to an integer net charge, depending on the protonation state and chemical context; or derived *de novo* to match the experimental HFE of 2-nitrophenol:
    a. from the base-linked oxygen atom of the phosphate group of nucleotides (e.g., adenosine) for the protonated forms of nitrotyrosine and nitrotryptophan,
    b. derived *de novo* to match the HFE of 2-nitrophenol for the protonated form of nitrotyrosine,
    c. from the base-linked oxygen atom of the phosphate group of nucleotides (e.g., adenosine) and the carboxyl group (e.g., glutamate) for the nitro and carboxyl groups, respectively, of the deprotonated form of nitrotyrosine.
16) KYNURENINE FORMATION: Parameters directly transferred from the carbonyl group (e.g., glutamine), the peptide bond and the amine group of the deprotonated arginine, with the carbon and hydrogen atom derived by analogy to the aldehyde group reported by Dolenc and others[43] for formyl-kynurenine.
17) CHLORINATION: Parameters directly transferred from chloroform.
18) CARBAMYLATION: Parameters directly transferred from the peptide bond, carboxyl group (e.g., glutamate) and the carboxamide group (e.g., glutamine).
19) NORLEUCINE: Parameters directly transferred from aliphatic carbon atoms.
20) N-TERMINAL METHYLATION: Parameters directly transferred from lysine methylation.
21) N-TERMINAL ACETYLATION: Parameters directly transferred from lysine acetylation.

22) N-TERMINAL PYRROLIDONE FORMATION: Parameters directly transferred from proline oxidation.
23) N-TERMINAL FORMYLATION: Parameters directly transferred from the peptide bond with the carbon and hydrogen atoms derived by analogy to the aldehyde group reported by Dolenc and others.[43]
24) N-TERMINAL PYRUVATE FORMATION: Parameters directly transferred from the carbonyl group (e.g., glutamine), with a bond type of a shorter equilibrium distance used between the carbonyl carbon atoms to account for the double bond effect.
25) C-TERMINAL AMIDATION: Parameters directly transferred from the carboxamide group of e.g., glutamine.
26) C-TERMINAL METHYLATION: Parameters directly transferred from the ester group reported by Chandrasekhar and others.[40]

We include detailed descriptions of parameter choices as comments in Dataset S1 and Dataset S2.

Molecular dynamics simulations and thermodynamic integration setup

We have used the thermodynamic integration approach,[25] a widely used computational method based on MD simulations, to calculate hydration free energies (HFEs) of neutral forms of small-molecule analogs of 14 amino-acid side chains (the same set as in Oostenbrink et al.[21]), compounds from the validation set and side chain analogs of all parameterized PTMs with a charge neutral protonation state. Non-bonded (van der Waals and Coulomb) interactions, coupled to a parameter λ, were scaled down to zero in a stepwise manner in vacuum and water. Free energy changes of these processes were calculated as the integral of the ensemble average of the derivative of the total Hamiltonian of the system with respect to λ, over the interval from λ = 0 to λ = 1. For vacuum calculations, three independent simulations, each 5 ns long, were run at 21 equally spaced λ-points with the temperature kept at 500 K and additional random kicks introduced by Langevin dynamics integration method,[44] in order to avoid convergence problems due to inefficient sampling of the conformational space. Water simulations were run in five independent copies, each 0.5 ns long, at 21 equally spaced λ-points, together with 10 additional λ-points placed in the regions of the least smoothness of the integrated curve,

using SPC explicit water,[45] a reaction field electrostatic scheme with a cutoff of r_c = 1.4 nm and the dielectric constant of ε_{rf} = 65 and a Berendsen thermostat and barostat keeping the temperature and pressure at 300 K (τ_T = 0.05 ps) and 1 bar (τ_p = 1 ps and compressibility = 4.5×10^{-5} bar^{-1}).[46] A soft-core formalism[47] was used to avoid singularities of the potential energy . The aforementioned integrals were evaluated by the generalized Simpson's rule for non-equidistant nodes using the averages over the independent simulations at each λ-point. HFEs were calculated as the difference between the change in free energy upon the removal of non-bonded interactions calculated in vacuum and calculated in water.

FUNDING

This work was supported in part by the Austrian Science Fund FWF (START grant Y 514-B11 to BZ, http://www.fwf.ac.at/), the Vienna Science and Technology fund (WWFF grant number LS08-QM03 to CO, http://www.wwtf.at/), and European Research Council (ERC Starting Independent grants 279408 to BZ and 260408 to CO http://erc.europa.eu/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGEMENTS

We thank A. Polyansky for help with MHP calculations and members of the Laboratory of Computational Biophysics at MFPL for useful advice and critical reading of the manuscript.

REFERENCES

1. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255-261 (2003).
2. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342-7372 (2005).
3. Bartova, E., Krejci, J., Harnicarova, A., Galiova, G. & Kozubek, S. Histone modifications and nuclear architecture: A review. *J. Histochem. Cytochem.* **56**, 711-721 (2008).
4. Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**, 20313-20316 (1997).
5. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* **24**, 1311-1317 (2005).
6. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **17**, 666-672 (2010).
7. van Rossum, B., Fischle, W. & Selenko, P. Asymmetrically modified nucleosomes expand the histone code. *Nat. Struct. Mol. Biol.* **19**, 1064-1066 (2012).
8. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
9. Kruschel, D. & Zagrovic, B. Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol Biosyst* **5**, 1606-1616 (2009).
10. Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**, 52-61 (2012).
11. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).
12. Polyansky, A. A. & Zagrovic, B. Protein Electrostatic Properties Predefining the Level of Surface Hydrophobicity Change upon Phosphorylation. *J Phys Chem Lett* **3**, 973-976 (2012).
13. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17857-17862 (2012).
14. Olausson, B. E. S. *et al.* Molecular dynamics simulations reveal specific interactions of post-translational palmitoyl modifications with rhodopsin in membranes. *J. Am. Chem. Soc.* **134**, 4324-4331 (2012).
15. Marlowe, A. E., Singh, A. & Yingling, Y. G. The effect of point mutations on structure and mechanical properties of collagen-like fibril: A molecular dynamics study. *Mater Sci Eng C Mater Biol Appl* **32**, 2583-2588 (2012).
16. Seeliger, D. *et al.* Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem. Biol.* **7**, 150-154 (2012).
17. Jorgensen, W. L., Maxwell, D. S. & TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225-11236 (1996).
18. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616 (1998).
19. Schuler, L. D., Daura, X. & van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
20. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012 (2003).

21. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676 (2004).
22. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
23. Lee, T. Y. *et al.* dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res.* **34**, D622-D627 (2006).
24. UniProt, C. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71-D75 (2012).
25. Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431-492 (1989).
26. Malde, A. K. *et al.* An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory. Comput.* **7**, 4026-4037 (2011).
27. Efremov, R. G. *et al.* Molecular lipophilicity in protein modeling and drug design. *Curr. Med. Chem.* **14**, 393-415 (2007).
28. Polyansky, A. A., Chugunov, A. O., Vassilevski, A. A., Grishin, E. V. & Efremov, R. G. Recent Advances in Computational Modeling of alpha-Helical Membrane-Active Peptides. *Curr. Protein Peptide Sci.* **13**, 644-657 (2012).
29. Vistoli, G. *et al.* Predicting the physicochemical profile of diastereoisomeric histidine-containing dipeptides by property space analysis. *Chirality* **24**, 566-576 (2012).
30. Hlevnjak, M., Zitkovic, G. & Zagrovic, B. Hydrophilicity matching - A potential prerequisite for the formation of protein-protein complexes in the cell. *PLoS ONE* **5(6)** (2010).
31. Case, D. A. *et al.* AMBER 12. University of California, San Francisco, 2012.
32. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359-2368 (2011).
33. Schuttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 1355-1363 (2004).
34. Vanquelef, E. *et al.* RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511-W517 (2011).
35. Schmid, N., Christ, C. D., Christen, M., Eichenberger, A. P. & van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **183**, 890-903 (2012).
36. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **7**, 306-317 (2001).
37. Margreitter, C., Petrov, D. & Zagrovic, B. Vienna-PTM web server: A toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* (2013).
38. Huq, M. D. M., Tsai, N. P., Khan, S. A. & Wei, L. N. Lysine trimethylation of retinoic acid receptor-alpha - A novel means to regulate receptor function. *Mol Cell Proteomics* **6**, 677-688 (2007).
39. Bottomley, M. J. Structures of protein domains that create or recognize histone modifications. *EMBO Rep.* **5**, 464-469 (2004).
40. Chandrasekhar, I., Oostenbrink, C. & van Gunsteren, W. F. Simulating the physiological phase of hydrated DPPC bilayers: The ester moiety. *Soft Mater* **2**, 27-45 (2004).
41. Oostenbrink, C., Juchli, D. & van Gunsteren, W. F. Amine hydration: A united-atom force-field solution. *Chemphyschem* **6**, 1800-1804 (2005).

42. Hritz, J. & Oostenbrink, C. Efficient free energy calculations for compounds with multiple stable conformations separated by high energy barriers. *J Phys Chem B* **113**, 12711-12720 (2009).
43. Dolenc, J., Oostenbrink, C., Koller, J. & van Gunsteren, W. F. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site. *Nucleic Acids Res.* **33**, 725-733 (2005).
44. Brunger, A., Brooks, C. L. & Karplus, M. Stochastic boundary-conditions for molecular-dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105**, 495-500 (1984).
45. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed B. Pullman) 331-342 (Reidel, Dordrecht, 1981).
46. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
47. Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529-539 (1994).
48. Gallicchio, E., Zhang, L. Y. & Levy, R. M. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* **23**, 517-529 (2002).
49. Rizzo, R. C., Aynechi, T., Case, D. A. & Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial, charge models and optimization of nonpolar contributions. *J. Chem. Theory. Comput.* **2**, 128-139 (2006).
50. Sulea, T., Wanapun, D., Dennis, S. & Purisima, E. O. Prediction of SAMPL-1 hydration free energies using a continuum electrostatics-dispersion model. *J Phys Chem B* **113**, 4511-4520 (2009).

Appendices to Chapter I

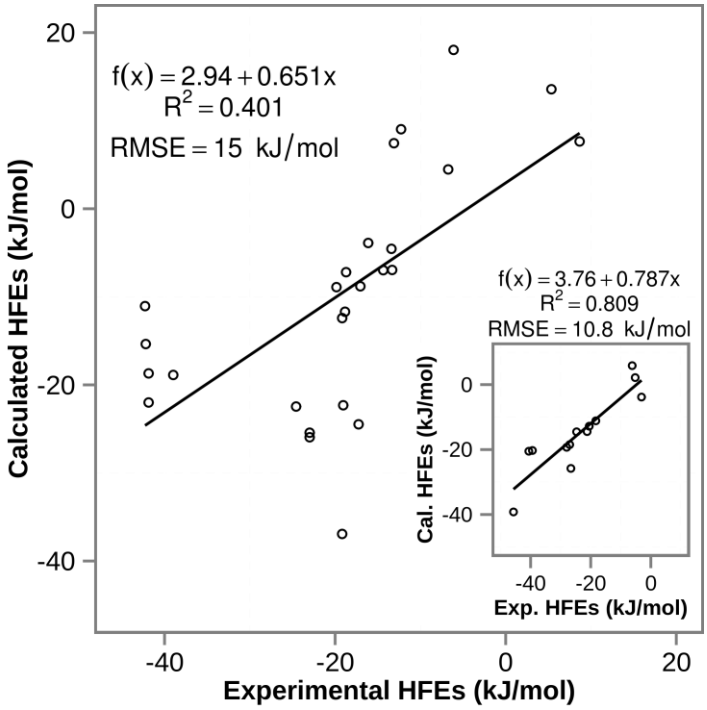
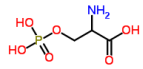
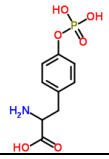
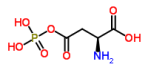
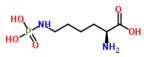
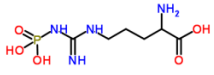
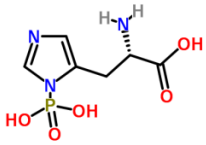
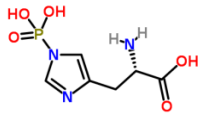


Figure S1. Experimental vs. calculated HFEs of compounds from the validation set (GROMOS 45a3). Correlation is captured by the regression line, its parameters, Pearson correlation coefficient and overall RMSE. The same comparison for canonical amino acids is shown in the inset. Note that error bars of calculated HFEs are comparable to the size of the symbols, with the average standard error of 0.4 kJ/mol.

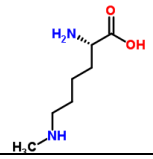
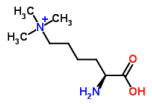
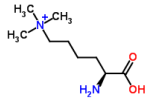
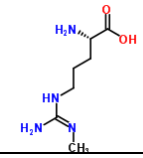
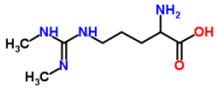
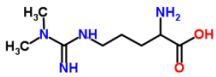
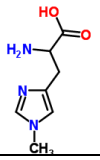
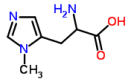
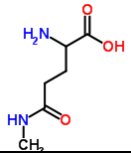
Table S1. Parameterized post-translational modifications with the 3-letter code, chemical names and structures. If two protonation states are possible, the one with higher occupancy at the physiologic pH is highlighted in bold. Note that modifications marked with: 1) ***** were already parameterized in GROMOS force field, 2) [#] have to date not been reported in UNIPROT, 3) [+] no prolines included and 4) [HFE] parameters derived to match the experimental HFE.

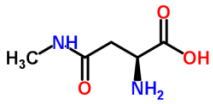
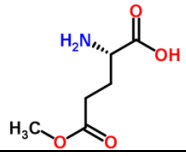
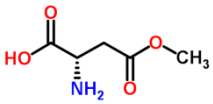
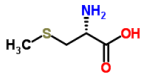
ENZYMATIC

# PTM	AA	code	chemical	structure
Phosphorylation				
1	1	SER	S1P	phosphoserine (-1)	
2	2		**S2P**	phosphoserine (-2)	
3	3	THR	T1P	phosphothreonine (-1)	
4	4		**T2P**	phosphothreonine (-2)	
5	5	TYR	Y1P	phosphotyrosine (-1)	
6	6		**Y2P**	phosphotyrosine (-2)	
7	7	ASP	D1P	phosphoaspartate (-1)	
8	8		**D2P**	phosphoaspartate (-2)	
9	9	LYS	K1P[#]	phospholysine (-1)	
10	10		**K2P[#]**	phospholysine (-2)	
11	11	ARG	R0P[#]	phosphoarginine (0)	
12	12		**R1P[#]**	phosphoarginine (-1)	
13	13	HIS	H11	1-phosphohistidine (-1)	
14	14		**H12**	1-phosphohistidine (-2)	

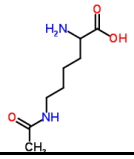
15	15	HIS	H31	3-phosphohistidine (-1)	
16	16		**H32**	3-phosphohistidine (-2)	

Methylation

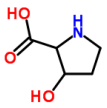
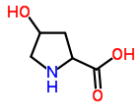
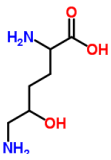
17	17	LYS	KMN	N6-methyllysine (0)	
18	18		**KMC**	N6-methyllysine (+1)	
19	19	LYS	K2M	N6,N6-dimethyllysine (0)	
20	20		**K2C**	N6,N6-dimethyllysine (+1)	
21	21	LYS	K3C	N6,N6,N6-trimethyllysine	
22	22	ARG	RMN	omega-N-methylarginine (0)	
23	23		**RMC**	omega-N-methylarginine (+1)	
24	24	ARG	RSM	symmetric-dimethylarginine (0)	
25	25		**RMS**	symmetric-dimethylarginine (+1)	
26	26	ARG	RAM	asymmetric-dimethylarginine (0)	
27	27		**RMA**	asymmetric-dimethylarginine (+1)	
28	28	HIS	**H1M**	1-methylhistidine (0)	
29	29		H1C	1-methylhistidine (+1)	
30	30	HIS	**H3M**	3-methylhistidine (0)	
31	31		H3C	3-methylhistidine (+1)	
32	32	GLN	QME	N5-methylglutamine	

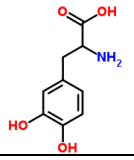
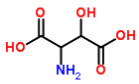
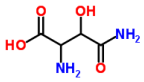
33	33	ASN	NME	N4-methylasparagine	
34	34	GLU	EME	glutamate methyl ester	
35	35	ASP	DMA[#]	aspartate methyl ester	
36	36	CYS	CYM	S-methylcysteine	

Acetylation

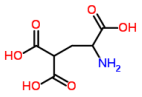
37	37	LYS	KAC	N6-acetyllysine	

Hydroxylation

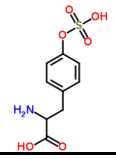
38	38	PRO	PH3	3-hydroxyproline (R)	
39	39		P3H	3-hydroxyproline (S)	
40	**1***	PRO	HYP	4-hydroxyproline (R)	
41	40		HY2	4-hydroxyproline (S)	
42	41	PRO	PHH	3,4-dihydroxyproline	
43	42	LYS	KH5	5-hydroxylysine (0,R)	
44	43		K5H	5-hydroxylysine (0,S)	
45	44		**KPH**	5-hydroxylysine (+1,R)	
46	45		**KHP**	5-hydroxylysine (+1,S)	

47	46	TYR	HTY	3,4-dihydroxyphenylalanine	
48	47	TRP	W7H	7-hydroxytryptophan	
49	48	ASP	**DH3**	3-hydroxyaspartate (-1,R)	
50	49		**D3H**	3-hydroxyaspartate (-1,S)	
51	50		DN3	3-hydroxyaspartate (0,R)	
52	51		D3N	3-hydroxyaspartate (0,S)	
53	52	ASN	N3H	3-hydroxyasparagine (R)	
54	53		NH3	3-hydroxyasparagine (S)	

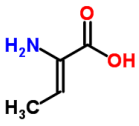
Carboxylation

55	54	GLU	**ECA**	4-carboxyglutamate (-2)	
56	55		ECN	4-carboxyglutamate (-1)	

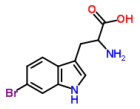
Sulfation

57	56	TYR	YSU	sulfotyrosine	

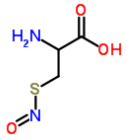
Dehydration

58	57	SER	SDH	dehydroalanine	
59	58	THR	TDH	2,3-didehydrobutyrine	

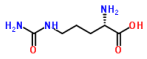
Bromidation

60	59	TRP	WBR	6-bromotryptophan	

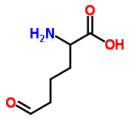
S-nitrosylation

61	60	CYS	CSN	S-nitrosocysteine	

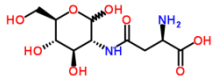
Citrullination

62	61	ARG	RCI	citrulline	

Allysine formation (the same as carbonylation)

63	62	LYS	KAL	allysine (aminoadipic semialdehyde)	

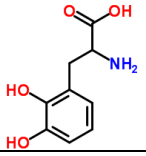
Glycosylation

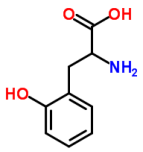
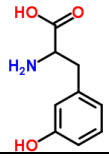
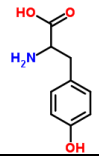
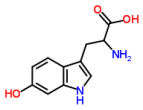
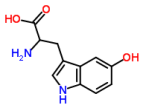
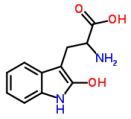
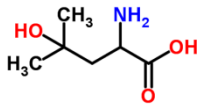
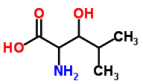
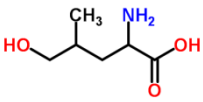
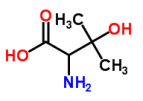
64	63	ASN	NNG	N-acetylglucosamine (N4-linked to ASN)	

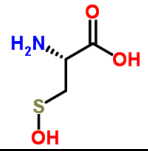
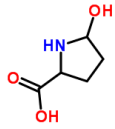
NONENZYMATIC

# PTM	AA	Code	chemical	structure

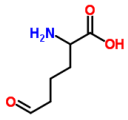
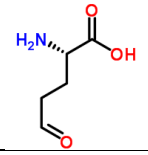
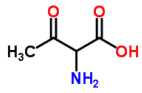
Hydroxylation

65	64	PHE	F23[#]	2,3-dihydroxyphenylalanine	

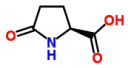
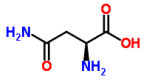
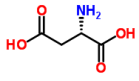
66	65	PHE	F2H#	2-hydroxyphenylalanine	
67	66	PHE	F3H#	3-hydroxyphenylalanine	
68	**2***	PHE	TYR#	tyrosine	
69	67	TRP	W6H#	6-hydroxytryptophan	
70	68	TRP	W5H#	5-hydroxytryptophan	
71	69	TRP	W4H#	4-hydroxytryptophan	
72	70	TRP	W2H#	2-hydroxytryptophan	
73	71	LEU	L3H#	3-hydroxyleucine (R)	
74	72		LH3#	3-hydroxyleucine (S)	
75	73	LEU	L4H#	4-hydroxyleucine	
76	74	LEU	L5H#	5-hydroxyleucine (R)	
77	75		LH5#	5-hydroxyleucine (S)	
78	76	VAL	V3H	3-hyroxyvaline	

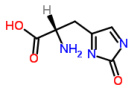
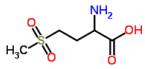
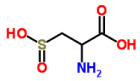
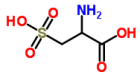
79	77	CYS	CYH	cysteine sulfenic acid	
80	78	PRO	PH5#	5-hydroxyproline (R)	
81	79		P5H#	5-hydroxyproline (S)	

Carbonylation

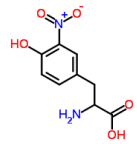
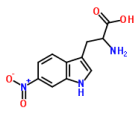
63	**62**	LYS	KAL	allysine (aminoadipic semialdehyde)	
82	80	PRO	GSA#	glutamic semialdehyde	
83		ARG			
84	81	THR	TOX#	2-amino-3-ketobutyric acid	

Oxidation

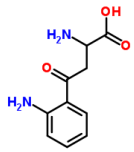
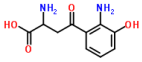
85	82	PRO	PGA#	pyroglutamic acid	
86	**3***	HIS	ASN	asparagine	
87	**4***	HIS	**ASP**	aspartic acid (-1)	
88	**5***		ASPH	aspartic acid (0)	

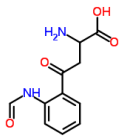
89	83	HIS	H2X[#]	2-oxo-histidine	
90	84	MET	MSX	methionine sulfoxide (R)	
91	85		MXS	methionine sulfoxide (S)	
92	86	MET	MES	methionine sulfone	
93	87	CYS	CSA	cysteine sulfinic acid	
94	88	CYS	CSE[#]	cysteic acid	

Nitration

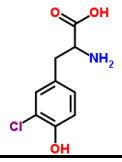
95	89	TYR	**YNI**	3-nitrotyrosine (-1)	
96	90		YNN[HFE]	3-nitrotyrosine (0)	
			YNB	3-nitrotyrosine (0)	
97	91	TRP	WNI[#]	6-nitrotryptophan	

Kynurenine formation

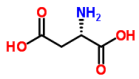
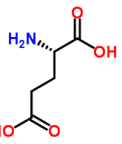
98	92	TRP	WKY[#]	kynurenine	
99	93	TRP	WKH[#]	3-hydroxykynurenine	

100	94	TRP	WKF[#]	formylkynurenine	

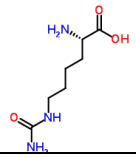
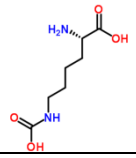
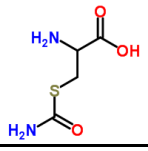
Chlorination

101	95	TYR	YCH[#]	chlorotyrosine	

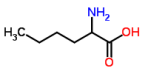
Deamidation

102	**4***	ASN	**ASP**	aspartic acid (-1)	
103	**5***		ASPH	aspartic acid (0)	
104	**6***	GLN	**GLU**	glutamic acid (-1)	
105	**7***		GLUH	glutamic acid (0)	

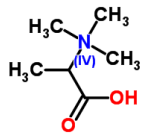
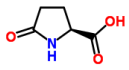
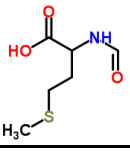
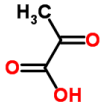
Carbamylation

106	96	LYS	KAM[#]	homocitruline	
107	97	LYS	**KCA**	carboxylysine (+1)	
108	98		KCN	carboxylysine (0)	
109	99	CYS	CAM[#]	S-carbamoylcysteine	

Norleucine

110	100	LEU	LNO[#]	norleucine	
111		LYS			
112		MET			

N-TERMINAL

# PTM		AA	Code	chemical	structure
Methylation					
113-132	101	all	1NM	N-methyl-AA (0)	
133-152	102		**1NM+**	N-methyl-AA (+1)	
153-171	103	**all**$^{+}$	2NM	N,N-dimethyl-AA (0)	
172-191	104		**2NM+**	N,N-dimethyl-AA (+1)	
193-210	105	**all**$^{+}$	3NM+	N,N,N-trimethyl-AA	
Acetylation					
211-230	106	all	NAC	N-acetyl-AA	
Pyrrolidone formation					
231	**81**	**GLN**	PGA	pyroglutamic acid	
232		GLU			
Formylation					
233	107	MET	FOR	N-formylmethionine	
Pyruvate formation					
234	108	**SER**	PYA	pyruvic acid	
235		CYS			
236		VAL			

C-TERMINAL

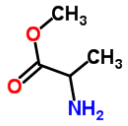
# PTM	AA	Code	chemical	structure
Amidation				
237-256	109	all	AMD	AA-amide 
Methylation				
257	110	**CYS**	CME	AA-methyl ester 
258		LEU		
259		LYS		

Table S2. HFEs of the molecules in the validation set, comparison between the experimental and calculated values using the GROMOS 45a3 parameter set.

Compound	HFE (kJ/mol)	
	experimental	ffG45a3
Validation set. PTM-side-chain analogs		
N-butylacetamide	-39.0	-18.9
o-cresol	-24.6	-22.5
m-cresol	-23.0	-25.9
2-methyl-2-propanol	-18.7	-7.2
2-methyl-1-propanol	-18.8	-11.7
propan-2-ol	-19.8	-8.9
N-methylacetamine	-41.9	-18.7
methylpropanoate	-12.3	9.0
methylacetate	-13.1	7.4
dimethylsulfide	-6.7	4.5
butanal	-13.3	-7.0
propanal	-14.4	-7.0
butane	8.7	7.6
Validation set. Compounds similar to PTM-side-chain analogs		
diethylamine	-17.0	-8.8
trimethylamine	-13.4	-4.6
ethene	5.4	13.6
bromobenzene	-6.1	18.0
aniline	-23.0	-25.4
acetophenone	-19.2	-12.4
N-methylformamide	-41.9	-22.0
chlorophenol	-19.0	-22.3
2-nitrophenol	-19.2	-36.9
nitrobenzene	-17.2	-24.5
acetone	-16.1	-3.9
dimethylsulfoxide	-42.3	-11.1
methylsulfonylmethane	-42.2	-15.4
RMSE	**-**	**15.0**

Table S3. Comparison of physico-chemical properties of PTMs and canonical amino acids. Molecular weight (MW), solvent accessible surface area (SASA), net charge and dipole moment (DM) shown for PTMs and cognate amino acids in parentheses.

PTM (AA)	chemical	MW (u)	SASA (nm^2)	net charge	DM (D)
S1P (SER)	phosphoserine (-1)	110.0 (31.0)	1.88 (1.20)	-1 (0)	N/A (2.30)
S2P (SER)	phosphoserine (-2)	109.0 (31.0)	1.93 (1.20)	-2 (0)	N/A (2.30)
T1P (THR)	phosphothreonine (-1)	124.0 (45.1)	2.12 (1.41)	-1 (0)	N/A (2.23)
T2P (THR)	phosphothreonine (-2)	123.0 (45.1)	2.15 (1.41)	-2 (0)	N/A (2.23)
Y1P (TYR)	phosphotyrosine (-1)	186.1 (107.1)	2.81 (2.41)	-1 (0)	N/A (2.06)
Y2P (TYR)	phosphotyrosine (-2)	185.1 (107.1)	2.73 (2.41)	-2 (0)	N/A (2.06)
D1P (ASP)	phosphoaspartate (-1)	138.0 (58.0)	2.23 (1.57)	-1 (-1)	N/A (N/A)
D2P (ASP)	phosphoaspartate (-2)	137.0 (58.0)	2.17 (1.57)	-2 (-1)	N/A (N/A)
K1P (LYS)	phospholysine (-1)	151.1 (73.1)	2.59 (1.92)	-1 (1)	N/A (N/A)
K2P (LYS)	phospholysine (-2)	150.1 (73.1)	2.54 (1.92)	-2 (1)	N/A (N/A)
R0P (ARG)	phosphoarginine (0)	180.1 (101.2)	2.82 (2.21)	0 (1)	14.65 (N/A)
R1P (ARG)	phosphoarginine (-1)	179.1 (101.2)	2.73 (2.21)	-1 (1)	N/A (N/A)
H11 (HIS)	1-phosphohistidine (-1)	160.1 (81.1)	2.36 (1.91)	-1 (0)	N/A (6.43)
H12 (HIS)	1-phosphohistidine (-2)	159.1 (81.1)	2.39 (1.91)	-2 (0)	N/A (6.43)
H31 (HIS)	3-phosphohistidine (-1)	160.1 (81.1)	2.51 (1.91)	-1 (0)	N/A (6.43)
H32 (HIS)	3-phosphohistidine (-2)	159.1 (81.1)	2.53 (1.91)	-2 (0)	N/A (6.43)
KMN (LYS)	N6-methyllysine (0)	86.2 (73.1)	2.40 (1.92)	0 (1)	1.84 (N/A)
KMC (LYS)	N6-methyllysine (+1)	87.2 (73.1)	2.26 (1.92)	1 (1)	N/A (N/A)
K2M (LYS)	N6,N6-dimethyllysine (0)	100.2 (73.1)	2.72 (1.92)	0 (1)	1.58 (N/A)
K2C (LYS)	N6,N6-dimethyllysine (+1)	101.2 (73.1)	2.45 (1.92)	1 (1)	N/A (N/A)
K3C (LYS)	N6,N6,N6-trimethyllysine	115.2 (73.1)	2.68 (1.92)	1 (1)	N/A (N/A)
RMN (ARG)	omega-N-methylarginine (0)	114.2 (101.2)	2.38 (2.21)	0 (1)	2.78 (N/A)
RMC (ARG)	omega-N-methylarginine (+1)	115.2 (101.2)	2.43 (2.21)	1 (1)	N/A (N/A)
RSM (ARG)	symmetric-dimethylarginine (0)	128.2 (101.2)	2.80 (2.21)	0 (1)	0.94 (N/A)
RMS (ARG)	symmetric-dimethylarginine (+1)	129.2 (101.2)	2.69 (2.21)	1 (1)	N/A (N/A)
RAM (ARG)	asymmetric-dimethylarginine (0)	128.2 (101.2)	2.62 (2.21)	0 (1)	3.55 (N/A)
RMA (ARG)	asymmetric-dimethylarginine (+1)	129.2 (101.2)	2.69 (2.21)	1 (1)	N/A (N/A)
H1M (HIS)	1-methylhistidine (0)	95.1 (81.1)	2.24 (1.91)	0 (0)	6.10 (6.43)
H1C (HIS)	1-methylhistidine (+1)	96.1 (81.1)	2.19 (1.91)	1 (0)	N/A (6.43)
H3M (HIS)	3-methylhistidine (0)	95.1 (81.1)	2.32 (1.91)	0 (0)	4.33 (6.43)
H3C (HIS)	3-methylhistidine (+1)	96.1 (81.1)	2.01 (1.91)	1 (0)	N/A (6.43)
QME (GLN)	N5-methylglutamine	86.1 (72.1)	2.07 (1.80)	0 (0)	2.28 (4.94)
NME (ASN)	N4-methylasparagine	72.1 (58.1)	1.97 (1.56)	0 (0)	4.14 (4.90)
EME (GLU)	glutamate methyl ester	87.1 (72.1)	1.99 (1.79)	0 (-1)	5.27 (N/A)
DMA (ASP)	aspartate methyl ester	73.1 (58.0)	1.80 (1.57)	0 (-1)	5.31 (N/A)
CYM (CYS)	S-methylcysteine	61.1 (47.1)	1.65 (1.41)	0 (0)	2.74 (1.84)
KAC (LYS)	N6-acetyllysine	114.2 (73.1)	2.67 (1.92)	0 (1)	4.12 (N/A)
PH3 (PRO)	3-hydroxyproline (R)	58.1 (42.1)	1.61 (1.59)	0 (0)	2.32 (N/A)
P3H (PRO)	3-hydroxyproline (S)	58.1 (42.1)	1.64 (1.59)	0 (0)	2.27 (N/A)

HYP (PRO)	4-hydroxyproline (R)	58.1 (42.1)	1.77 (1.59)	0 (0)	2.14 (N/A)
HY2 (PRO)	4-hydroxyproline (S)	58.1 (42.1)	1.67 (1.59)	0 (0)	2.20 (N/A)
PHH (PRO)	3,4-dihydroxyproline	74.1 (42.1)	1.76 (1.59)	0 (0)	3.66 (N/A)
KH5 (LYS)	5-hydroxylysine (0,R)	88.1 (73.1)	2.03 (1.92)	0 (1)	3.62 (N/A)
K5H (LYS)	5-hydroxylysine (0,S)	88.1 (73.1)	2.17 (1.92)	0 (1)	5.86 (N/A)
KPH (LYS)	5-hydroxylysine (+1,R)	89.1 (73.1)	2.05 (1.92)	1 (1)	N/A (N/A)
KHP (LYS)	5-hydroxylysine (+1,S)	89.1 (73.1)	2.03 (1.92)	1 (1)	N/A (N/A)
HTY (TYR)	3,4-dihydroxyphenylalanine	123.1 (107.1)	2.29 (2.41)	0 (0)	3.97 (2.06)
W7H (TRP)	7-hydroxytryptophan	146.2 (130.2)	2.70 (2.63)	0 (0)	3.19 (3.65)
DH3 (ASP)	3-hydroxyaspartate (-1,R)	74.0 (58.0)	1.52 (1.57)	-1 (-1)	N/A (N/A)
D3H (ASP)	3-hydroxyaspartate (-1,S)	74.0 (58.0)	1.57 (1.57)	-1 (-1)	N/A (N/A)
DN3 (ASP)	3-hydroxyaspartate (0,R)	75.0 (58.0)	1.61 (1.57)	0 (-1)	0.58 (N/A)
D3N (ASP)	3-hydroxyaspartate (0,S)	75.0 (58.0)	1.56 (1.57)	0 (-1)	3.79 (N/A)
N3H (ASN)	3-hydroxyasparagine (R)	74.1 (58.1)	1.71 (1.56)	0 (0)	3.34 (4.90)
NH3 (ASN)	3-hydroxyasparagine (S)	74.1 (58.1)	1.84 (1.56)	0 (0)	6.30 (4.90)
ECA (GLU)	4-carboxyglutamate (-2)	115.1 (72.1)	2.06 (1.79)	-2 (-1)	N/A (N/A)
ECN (GLU)	4-carboxyglutamate (-1)	116.1 (72.1)	2.20 (1.79)	-1 (-1)	N/A (N/A)
YSU (TYR)	sulfotyrosine	186.2 (107.1)	2.66 (2.41)	-1 (0)	N/A (2.06)
SDH (SER)	dehydroalanine	14.0 (31.0)	1.06 (1.20)	0 (0)	0.00 (2.30)
TDH (THR)	2,3-didehydrobutyrine	28.1 (45.1)	1.39 (1.41)	0 (0)	0.00 (2.23)
WBR (TRP)	6-bromotryptophan	209.1 (130.2)	2.88 (2.63)	0 (0)	3.16 (3.65)
CSN (CYS)	S-nitrosocysteine	76.1 (47.1)	1.71 (1.41)	0 (0)	3.15 (1.84)
RCI (ARG)	citrulline	101.1 (101.2)	2.22 (2.21)	0 (1)	5.21 (N/A)
KAL (LYS)	allysine (aminoadipic semialdehyde)	71.1 (73.1)	2.04 (1.92)	0 (1)	2.96 (N/A)
NNG (ASN)	N-acetylglucosamine	261.3 (58.1)	3.73 (1.56)	0 (0)	9.10 (4.90)
F23 (PHE)	2,3-dihydroxyphenylalanine	123.1 (91.1)	2.29 (2.28)	0 (0)	4.55 (0.74)
F2H (PHE)	2-hydroxyphenylalanine	107.1 (91.1)	2.13 (2.28)	0 (0)	3.11 (0.74)
F3H (PHE)	3-hydroxyphenylalanine	107.1 (91.1)	2.08 (2.28)	0 (0)	2.73 (0.74)
TYR (PHE)	tyrosine	107.1 (91.1)	2.41 (2.28)	0 (0)	2.06 (0.74)
W6H (TRP)	6-hydroxytryptophan	146.2 (130.2)	2.75 (2.63)	0 (0)	1.32 (3.65)
W5H (TRP)	5-hydroxytryptophan	146.2 (130.2)	2.71 (2.63)	0 (0)	3.46 (3.65)
W4H (TRP)	4-hydroxytryptophan	146.2 (130.2)	2.71 (2.63)	0 (0)	4.88 (3.65)
W2H (TRP)	2-hydroxytryptophan	146.2 (130.2)	2.73 (2.63)	0 (0)	5.19 (3.65)
L3H (LEU)	3-hydroxyleucine (R)	73.1 (57.1)	1.96 (1.88)	0 (0)	2.23 (0.00)
LH3 (LEU)	3-hydroxyleucine (S)	73.1 (57.1)	1.86 (1.88)	0 (0)	2.25 (0.00)
L4H (LEU)	4-hydroxyleucine	73.1 (57.1)	1.92 (1.88)	0 (0)	2.19 (0.00)
L5H (LEU)	5-hydroxyleucine (R)	73.1 (57.1)	1.98 (1.88)	0 (0)	2.17 (0.00)
LH5 (LEU)	5-hydroxyleucine (S)	73.1 (57.1)	2.01 (1.88)	0 (0)	2.19 (0.00)
V3H (VAL)	3-hyroxyvaline	59.1 (43.1)	1.70 (1.55)	0 (0)	2.30 (0.00)
CYH (CYS)	cysteine sulfenic acid	63.1 (47.1)	1.62 (1.41)	0 (0)	2.38 (1.84)
PH5 (PRO)	5-hydroxyproline (R)	58.1 (42.1)	1.69 (1.59)	0 (0)	2.21 (N/A)
P5H (PRO)	5-hydroxyproline (S)	58.1 (42.1)	1.79 (1.59)	0 (0)	2.17 (N/A)
GSA (PRO)	glutamic semialdehyde	57.1 (42.1)	1.64 (1.59)	0 (0)	2.95 (N/A)
GSA (ARG)	glutamic semialdehyde	57.1 (101.2)	1.64 (2.21)	0 (1)	2.95 (N/A)

TOX (THR)	2-amino-3-ketobutyric acid	43.0 (45.1)	1.40 (1.41)	0 (0)	2.66 (2.23)
PGA (PRO)	pyroglutamic acid	56.1 (42.1)	1.74 (1.59)	0 (0)	2.65 (N/A)
ASN (HIS)	asparagine	58.1 (81.1)	1.56 (1.91)	0 (0)	4.90 (6.43)
ASP (HIS)	aspartic acid (-1)	58.0 (81.1)	1.57 (1.91)	-1 (0)	N/A (6.43)
H2X (HIS)	2-oxo-histidine	95.1 (81.1)	2.00 (1.91)	0 (0)	9.64 (6.43)
MSX (MET)	methionine sulfoxide (R)	91.2 (75.2)	2.02 (1.94)	0 (0)	4.27 (2.62)
MXS (MET)	methionine sulfoxide (S)	91.2 (75.2)	2.11 (1.94)	0 (0)	4.23 (2.62)
MES (MET)	methionine sulfone	107.2 (75.2)	2.14 (1.94)	0 (0)	6.80 (2.62)
CSA (CYS)	cysteine sulfinic acid	78.1 (47.1)	1.69 (1.41)	-1 (0)	N/A (1.84)
CSE (CYS)	cysteic acid	94.1 (47.1)	1.80 (1.41)	-1 (0)	N/A (1.84)
YNI (TYR)	3-nitrotyrosine (-1)	151.1 (107.1)	2.61 (2.41)	-1 (0)	N/A (2.06)
YNN (TYR)	3-nitrotyrosine (0)	152.1 (107.1)	2.62 (2.41)	0 (0)	5.44 (2.06)
YNB (TYR)	3-nitrotyrosine (0)	152.1 (107.1)	2.62 (2.41)	0 (0)	3.83 (2.06)
WNI (TRP)	6-nitrotryptophan	175.2 (130.2)	3.06 (2.63)	0 (0)	3.08 (3.65)
WKY (TRP)	kynurenine	134.2 (130.2)	2.60 (2.63)	0 (0)	4.16 (3.65)
WKH (TRP)	3-hydroxykynurenine	150.2 (130.2)	2.66 (2.63)	0 (0)	2.78 (3.65)
WKF (TRP)	formylkynurenine	162.2 (130.2)	2.98 (2.63)	0 (0)	3.03 (3.65)
YCH (TYR)	chlorotyrosine	141.6 (107.1)	2.44 (2.41)	0 (0)	3.43 (2.06)
ASP (ASN)	aspartic acid (-1)	58.0 (58.1)	1.57 (1.56)	-1 (0)	N/A (4.90)
GLU (GLN)	glutamic acid (-1)	72.1 (72.1)	1.79 (1.80)	-1 (0)	N/A (4.94)
KAM (LYS)	homocitruline	115.2 (73.1)	2.46 (1.92)	0 (1)	5.24 (N/A)
KCA (LYS)	carboxylysine (+1)	115.1 (73.1)	2.48 (1.92)	-1 (1)	N/A (N/A)
KCN (LYS)	carboxylysine (0)	116.1 (73.1)	2.49 (1.92)	0 (1)	3.63 (N/A)
CAM (CYS)	S-carbamoylcysteine	90.1 (47.1)	1.99 (1.41)	0 (0)	4.81 (1.84)
LNO (LEU)	norleucine	57.1 (57.1)	1.88 (1.88)	0 (0)	0.00 (0.00)
LNO (LYS)	norleucine	57.1 (73.1)	1.88 (1.92)	0 (1)	0.00 (N/A)
LNO (MET)	norleucine	57.1 (75.2)	1.88 (1.94)	0 (0)	0.00 (2.62)

Dataset S1. Force field parameters for the GROMOS force field 45A3 parameter set

```
; This file contains extended force field parameters for the GROMOS force field 45A3
parameters set
; GROMACS 4.5.x format - files: aminoacids.rtp, aminoacids.n.tdb, aminoacids.c.tdb
and aminoacids.hdb
; Authors: Drazen Petrov, Christian Margreitter, Melanie Grandits, Chris Oostenbrink
& Bojan Zagrovic
; Parameter files in GROMACS 4.3.x and 4.4.x, and GROMOS formats available at
http://vienna-ptm.univie.ac.at/


; aminoacids.rtp file (backbone and side chain parameters)

[ bondedtypes ]
; bonds  angles  dihedrals  impropers
   2      2         1         2

; phosphoserine (-1)
[ S1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OG   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.63000    2 ; to add up to -1 net charge
  OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE3   OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HE3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   OG   gb_17
  OG   PD   gb_27
  PD   OE1   gb_23
  PD   OE2   gb_23
  PD   OE3   gb_27
 OE3  HE3   gb_1
 [ exclusions ]
;  ai    aj
  OE1  HE3
  OE2  HE3
  OG   HE3
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N     H    ga_31
    H    N    CA    ga_17
   -C    N    CA    ga_30
    N   CA     C    ga_12
   CA    C    +N    ga_18
   CA    C     O    ga_29
    O    C    +N    ga_32
    N   CA    CB    ga_12
    C   CA    CB    ga_12
   CA   CB    OG    ga_12
   CB   OG    PD    ga_25
   OG   PD   OE1    ga_13
   OG   PD   OE2    ga_13
   OG   PD   OE3    ga_4
  OE1   PD   OE2    ga_28
  OE1   PD   OE3    ga_13
  OE2   PD   OE3    ga_13
   PD  OE3  HE3     ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
    C    CA    +N    O    gi_1
   CA    N     C    CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C     N    CA    gd_4
  -C    N     CA    C     gd_19
   N    CA    C    +N     gd_20
   N    CA    CB   OG     gd_17
  CA    CB    OG   PD     gd_12
  CB    OG    PD   OE3    gd_9
  CB    OG    PD   OE3    gd_11
  OG    PD   OE3  HE3     gd_9
  OG    PD   OE3  HE3     gd_11

; phosphoserine (-2)
[ S2P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OG   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.11500    2 ; to add up to -2 net charge
  OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   OG   gb_17
  OG   PD   gb_27
  PD   OE1   gb_23
  PD   OE2   gb_23
  PD   OE3   gb_23
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N     H    ga_31
    H    N    CA    ga_17
   -C    N    CA    ga_30
    N   CA     C    ga_12
   CA    C    +N    ga_18
   CA    C     O    ga_29
    O    C    +N    ga_32
    N   CA    CB    ga_12
    C   CA    CB    ga_12
   CA   CB    OG    ga_12
   CB   OG    PD    ga_25
   OG   PD   OE1    ga_13
   OG   PD   OE2    ga_13
   OG   PD   OE3    ga_13
  OE1   PD   OE2    ga_13
  OE1   PD   OE3    ga_13
  OE2   PD   OE3    ga_13
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
    C    CA    +N    O    gi_1
   CA    N     C    CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C     N    CA    gd_4
  -C    N     CA    C     gd_19
   N    CA    C    +N     gd_20
   N    CA    CB   OG     gd_17
  CA    CB    OG   PD     gd_12
  CB    OG    PD   OE1    gd_9
  CB    OG    PD   OE1    gd_11

; phosphothreonine (-1)
[ T1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH1    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OG1   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.63000    2 ; to add up to -1 net charge
  OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OE3   OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HE3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
  CG2  CH3    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1   PD   gb_27
  PD  OE1   gb_23
  PD  OE2   gb_23
  PD  OE3   gb_27
```

```
 OE3  HE3  gb_1
[ exclusions ]
;  ai   aj
 OE1  HE3
 OE2  HE3
 OG1  HE3
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_14
 OG1   CB  CG2    ga_14
  CB  OG1   PD    ga_25
 OG1   PD  OE1    ga_13
 OG1   PD  OE2    ga_13
 OG1   PD  OE3    ga_4
 OE1   PD  OE2    ga_28
 OE1   PD  OE3    ga_13
 OE2   PD  OE3    ga_13
  PD  OE3  HE3    ga_11
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  OG1    gd_17
  CA   CB  OG1   PD    gd_12
  CB  OG1   PD  OE3    gd_9
  CB  OG1   PD  OE3    gd_11
 OG1   PD  OE3  HE3    gd_9
 OG1   PD  OE3  HE3    gd_11

; phosphothreonine (-2)
[ T2P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
 OG1   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.11500    2 ; to add up to -2 net charge
 OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 CG2  CH3    0.00000    3
   C    C     0.380    4
   O    O    -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1   PD   gb_27
  PD  OE1   gb_23
  PD  OE2   gb_23
  PD  OE3   gb_23
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_14
 OG1   CB  CG2    ga_14
  CB  OG1   PD    ga_25
 OG1   PD  OE1    ga_13
 OG1   PD  OE2    ga_13
 OG1   PD  OE3    ga_13
 OE1   PD  OE2    ga_13
 OE1   PD  OE3    ga_13
 OE2   PD  OE3    ga_13
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  OG1    gd_17
  CA   CB  OG1   PD    gd_12
  CB  OG1   PD  OE1    gd_9
  CB  OG1   PD  OE1    gd_11

; phosphotyrosine (-1)
[ Y1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C   -0.10000    3
 HD2   HC    0.10000    3
 CE1    C   -0.10000    4
 HE1   HC    0.10000    4
 CE2    C   -0.10000    5
 HE2   HC    0.10000    5
  CZ    C    0.15000    6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OH   OA   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
  PT    P    0.63000    6 ; to add up to -1 net charge
 OI1   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OA   -0.54800    6 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HI3    H    0.39800    6 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C     0.380    7
   O    O    -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  CE1   gb_15
 CD2  HD2   gb_3
 CD2  CE2   gb_15
 CE1  HE1   gb_3
 CE1   CZ   gb_15
 CE2  HE2   gb_3
 CE2   CZ   gb_15
  CZ   OH   gb_12
  OH   PT   gb_27
  PT  OI1   gb_23
  PT  OI2   gb_23
  PT  OI3   gb_27
 OI3  HI3   gb_1
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE1
  CG  HE2
  CG   CZ
 CD1  HD2
 CD1  CE2
 CD1   OH
 HD1  CD2
 HD1  HE1
 HD1   CZ
 CD2  CE1
 CD2   OH
 HD2  HE2
 HD2   CZ
 CE1  HE2
 HE1  CE2
 HE1   OH
 HE2   OH
  OH  HI3
```

```
  OI1   HI3
  OI2   HI3
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_12
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_12
   C    CA    CB   ga_12
  CA    CB    CG   ga_14
  CB    CG    CD1  ga_26
  CB    CG    CD2  ga_26
  CD1   CG    CD2  ga_26
  CG    CD1   HD1  ga_24
  HD1   CD1   CE1  ga_24
  CG    CD1   CE1  ga_26
  CG    CD2   HD2  ga_24
  HD2   CD2   CE2  ga_24
  CG    CD2   CE2  ga_26
  CD1   CE1   HE1  ga_24
  HE1   CE1   CZ   ga_24
  CD1   CE1   CZ   ga_26
  CD2   CE2   HE2  ga_24
  HE2   CE2   CZ   ga_24
  CD2   CE2   CZ   ga_26
  CE1   CZ    CE2  ga_26
  CE1   CZ    OH   ga_26
  CE2   CZ    OH   ga_26
  CZ    OH    PT   ga_25
  OH    PT    OI1  ga_13
  OH    PT    OI2  ga_13
  OH    PT    OI3  ga_4
  OI1   PT    OI2  ga_28
  OI1   PT    OI3  ga_13
  OI2   PT    OI3  ga_13
 HI3    OI3   PT   ga_11
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
   C    CA    +N   O    gi_1
  CA    N     C    CB   gi_2
  CG    CD1   CD2  CB   gi_1
  CD2   CG    CD1  CE1  gi_1
  CD1   CG    CD2  CE2  gi_1
  CG    CD1   CE1  CZ   gi_1
  CG    CD2   CE2  CZ   gi_1
  CD1   CE1   CZ   CE2  gi_1
  CD2   CE2   CZ   CE1  gi_1
  CD1   CG    CE1  HD1  gi_1
  CD2   CG    CE2  HD2  gi_1
  CE1   CZ    CD1  HE1  gi_1
  CE2   CZ    CD2  HE2  gi_1
  CZ    CE1   CE2  OH   gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_4
  -C    N     CA   C    gd_19
   N    CA    C    +N   gd_20
   N    CA    CB   CG   gd_17
  CA    CB    CG   CD1  gd_20
  CE1   CZ    OH   PT   gd_2
  CZ    OH    PT   OI3  gd_9
  CZ    OH    PT   OI3  gd_11
  OH    PT    OI3  HI3  gd_9
  OH    PT    OI3  HI3  gd_11

; phosphotyrosine (-2)
[ Y2P ]
 [ atoms ]
   N    N    -0.28000    0
   H    H     0.28000    0
  CA    CH1   0.00000    1
  CB    CH2   0.00000    1
  CG    C     0.00000    1
  CD1   C    -0.10000    2
  HD1   HC    0.10000    2
  CD2   C    -0.10000    3
  HD2   HC    0.10000    3
  CE1   C    -0.10000    4
  HE1   HC    0.10000    4
  CE2   C    -0.10000    5
  HE2   HC    0.10000    5
  CZ    C     0.15000    6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OH    OA   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
  PT    P     0.11500    6 ; to add up to -2 net charge
  OI1   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
  OI2   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
  OI3   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C     0.380    7
   O    O    -0.380    7
 [ bonds ]
   N    H    gb_2
   N    CA   gb_20
  CA    C    gb_26
   C    O    gb_4
   C    +N   gb_9
  CA    CB   gb_26
  CB    CG   gb_26
  CG    CD1  gb_15
  CG    CD2  gb_15
 CD1    HD1  gb_3
 CD1    CE1  gb_15
 CD2    HD2  gb_3
 CD2    CE2  gb_15
 CE1    HE1  gb_3
 CE1    CZ   gb_15
 CE2    HE2  gb_3
 CE2    CZ   gb_15
  CZ    OH   gb_12
  OH    PT   gb_27
  PT    OI1  gb_23
  PT    OI2  gb_23
  PT    OI3  gb_23
 [ exclusions ]
; ai    aj
  CB    HD1
  CB    HD2
  CB    CE1
  CB    CE2
  CG    HE1
  CG    HE2
  CG    CZ
 CD1    HD2
 CD1    CE2
 CD1    OH
 HD1    CD2
 HD1    HE1
 HD1    CZ
 CD2    CE1
 CD2    OH
 HD2    HE2
 HD2    CZ
 CE1    HE2
 HE1    CE2
 HE1    OH
 HE2    OH
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_12
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_12
   C    CA    CB   ga_12
  CA    CB    CG   ga_14
  CB    CG    CD1  ga_26
  CB    CG    CD2  ga_26
 CD1    CG    CD2  ga_26
  CG    CD1   HD1  ga_24
 HD1    CD1   CE1  ga_24
  CG    CD1   CE1  ga_26
  CG    CD2   HD2  ga_24
 HD2    CD2   CE2  ga_24
  CG    CD2   CE2  ga_26
 CD1    CE1   HE1  ga_24
 HE1    CE1   CZ   ga_24
 CD1    CE1   CZ   ga_26
 CD2    CE2   HE2  ga_24
 HE2    CE2   CZ   ga_24
 CD2    CE2   CZ   ga_26
 CE1    CZ    CE2  ga_26
 CE1    CZ    OH   ga_26
 CE2    CZ    OH   ga_26
  CZ    OH    PT   ga_25
  OH    PT    OI1  ga_13
  OH    PT    OI2  ga_13
  OH    PT    OI3  ga_13
 OI1    PT    OI2  ga_13
 OI1    PT    OI3  ga_13
 OI2    PT    OI3  ga_13
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
   C    CA    +N   O    gi_1
  CA    N     C    CB   gi_2
  CG    CD1   CD2  CB   gi_1
 CD2    CG    CD1  CE1  gi_1
 CD1    CG    CD2  CE2  gi_1
```

```
  CG   CD1  CE1   CZ    gi_1
  CG   CD2  CE2   CZ    gi_1
  CD1  CE1  CZ   CE2    gi_1
  CD2  CE2  CZ   CE1    gi_1
  CD1  CG   CE1  HD1    gi_1
  CD2  CG   CE2  HD2    gi_1
  CE1  CZ   CD1  HE1    gi_1
  CE2  CZ   CD2  HE2    gi_1
  CZ  CE1   CE2  OH     gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA   C     gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  CD1    gd_20
 CE1   CZ   OH   PT    gd_2
  CZ   OH   PT  OI1    gd_9
  CZ   OH   PT  OI1    gd_11

; phosphoaspartate (-1)
[ D1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.53000    2 ; to add up to -1 net charge
 OD1    O   -0.38000    2 ; from the carbonyl oxygen (of e.g., the peptide bond)
 OD2   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PE    P    0.63000    2 ; from the phosphate group (of e.g., S1P)
 OZ1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ3   OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HZ3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  OD2   gb_12
 OD2   PE   gb_27
  PE  OZ1   gb_23
  PE  OZ2   gb_23
  PE  OZ3   gb_27
 OZ3  HZ3   gb_1
 [ exclusions ]
; ai   aj
 OZ1  HZ3
 OZ2  HZ3
 OD2  HZ3
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  OD1    ga_21
  CB   CG  OD2    ga_18
 OD1   CG  OD2    ga_32
  CG  OD2   PE    ga_25
 OD2   PE  OZ1    ga_13
 OD2   PE  OZ2    ga_13
 OD2   PE  OZ3    ga_4
 OZ1   PE  OZ2    ga_28
 OZ1   PE  OZ3    ga_13
 OZ2   PE  OZ3    ga_13
  PE  OZ3  HZ3    ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  OD1  OD2   CB    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA   C     gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  OD2    gd_20
  CB   CG  OD2   PE    gd_3
  CG  OD2   PE  OZ3    gd_9
  CG  OD2   PE  OZ3    gd_11
 OD2   PE  OZ3  HZ3    gd_9
 OD2   PE  OZ3  HZ3    gd_11

; phosphoaspartate (-2)
[ D2P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.53000    2 ; to add up to -2 net charge
 OD1    O   -0.38000    2 ; from the carbonyl oxygen (of e.g., the peptide bond)
 OD2   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PE    P    0.11500    2 ; from the phosphate group (of e.g., S2P)
 OZ1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  OD2   gb_12
 OD2   PE   gb_27
  PE  OZ1   gb_23
  PE  OZ2   gb_23
  PE  OZ3   gb_23
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  OD1    ga_21
  CB   CG  OD2    ga_18
 OD1   CG  OD2    ga_32
  CG  OD2   PE    ga_25
 OD2   PE  OZ1    ga_13
 OD2   PE  OZ2    ga_13
 OD2   PE  OZ3    ga_13
 OZ1   PE  OZ2    ga_13
 OZ1   PE  OZ3    ga_13
 OZ2   PE  OZ3    ga_13
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  OD1  OD2   CB    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA   C     gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  OD2    gd_20
  CB   CG  OD2   PE    gd_3
  CG  OD2   PE  OZ1    gd_9
  CG  OD2   PE  OZ1    gd_11

; phospholysine (-1)
[ K1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ   NE   -0.28000    4 ; from NE of ARGN
  HZ    H    0.28000    4 ; from HE of ARGN
  PH    P    0.42000    5 ; to add up to -1 net charge
 OI1   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OA   -0.54800    5 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HI3    H    0.39800    5 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C      0.380    6
```

```
  O    O     -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   PH   gb_23
  PH  OI1   gb_23
  PH  OI2   gb_23
  PH  OI3   gb_27
 OI3  HI3   gb_1
 [ exclusions ]
 ;  ai    aj
 OI1  HI3
 OI2  HI3
  NZ  HI3
  HZ  OI1
  HZ  OI2
  HZ  OI3
 [ angles ]
 ;  ai    aj    ak   gromos type
   -C    N     H    ga_31
    H    N    CA    ga_17
   -C    N    CA    ga_30
    N   CA     C    ga_12
   CA    C    +N    ga_18
   CA    C     O    ga_29
    O    C    +N    ga_32
    N   CA    CB    ga_12
    C   CA    CB    ga_12
   CA   CB    CG    ga_14
   CB   CG    CD    ga_14
   CG   CD    CE    ga_14
   CD   CE    NZ    ga_14
   CE   NZ    HZ    ga_19
   HZ   NZ    PH    ga_22
   CE   NZ    PH    ga_32
   NZ   PH   OI1    ga_13
   NZ   PH   OI2    ga_13
   NZ   PH   OI3    ga_4
  OI1   PH   OI2    ga_28
  OI1   PH   OI3    ga_13
  OI2   PH   OI3    ga_13
  HI3  OI3    PH    ga_11
 [ impropers ]
 ;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
    C    CA    +N    O    gi_1
   CA    N     C    CB    gi_2
   NZ    CE    PH   HZ    gi_1
 [ dihedrals ]
 ;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA    gd_4
   -C     N    CA     C    gd_19
    N    CA     C    +N    gd_20
    N    CA    CB    CG    gd_17
   CA    CB    CG    CD    gd_17
   CB    CG    CD    CE    gd_17
   CG    CD    CE    NZ    gd_17
   CD    CE    NZ    PH    gd_19
   CE    NZ    PH   OI3    gd_19
   NZ    PH   OI3   HI3    gd_9
   NZ    PH   OI3   HI3    gd_11

 ; phospholysine (-2)
 [ K2P ]
  [ atoms ]
    N    N    -0.28000    0
    H    H     0.28000    0
   CA   CH1     0.00000    1
   CB   CH2     0.00000    1
   CG   CH2     0.00000    2
   CD   CH2     0.00000    2
   CE   CH2     0.00000    3
   NZ    NE    -0.28000    4 ; from NE of ARGN
   HZ    H      0.28000    4 ; from HE of ARGN
   PH    P     -0.09500    5 ; to add up to -2 net charge
  OI1   OM     -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI2   OM     -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI3   OM     -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
    C    C      0.380     6
    O    O     -0.380     6
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG   CD   gb_26
   CD   CE   gb_26
   CE   NZ   gb_20
   NZ   HZ   gb_2
   NZ   PH   gb_23
   PH  OI1   gb_23
   PH  OI2   gb_23
   PH  OI3   gb_23
  [ exclusions ]
 ;  ai    aj
   HZ  OI1
   HZ  OI2
   HZ  OI3
  [ angles ]
 ;  ai    aj    ak   gromos type
   -C    N     H    ga_31
    H    N    CA    ga_17
   -C    N    CA    ga_30
    N   CA     C    ga_12
   CA    C    +N    ga_18
   CA    C     O    ga_29
    O    C    +N    ga_32
    N   CA    CB    ga_12
    C   CA    CB    ga_12
   CA   CB    CG    ga_14
   CB   CG    CD    ga_14
   CG   CD    CE    ga_14
   CD   CE    NZ    ga_14
   CE   NZ    HZ    ga_19
   HZ   NZ    PH    ga_22
   CE   NZ    PH    ga_32
   NZ   PH   OI1    ga_13
   NZ   PH   OI2    ga_13
   NZ   PH   OI3    ga_13
  OI1   PH   OI2    ga_13
  OI1   PH   OI3    ga_13
  OI2   PH   OI3    ga_13
  [ impropers ]
 ;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
    C    CA    +N    O    gi_1
   CA     N     C    CB    gi_2
   NZ    CE    PH   HZ    gi_1
  [ dihedrals ]
 ;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA    gd_4
   -C     N    CA     C    gd_19
    N    CA     C    +N    gd_20
    N    CA    CB    CG    gd_17
   CA    CB    CG    CD    gd_17
   CB    CG    CD    CE    gd_17
   CG    CD    CE    NZ    gd_17
   CD    CE    NZ    PH    gd_19
   CE    NZ    PH   OI1    gd_19

 ; phosphoarginine (0)
 [ R0P ]
  [ atoms ]
    N    N    -0.28000    0
    H    H     0.28000    0
   CA   CH1     0.00000    1
   CB   CH2     0.00000    1
   CG   CH2     0.00000    1
   CD   CH2     0.09000    2
   NE    NE    -0.11000    2
   HE     H     0.24000    2
   CZ     C     0.43000    2 ; to add up to 1 net charge
  NH1    NZ    -0.26000    2
  HH11    H     0.24000    2
  HH12    H     0.24000    2
  NH2    NE    -0.11000    2 ; from NE nitrogen atom of ARG
  HH2     H     0.24000    2 ; from terminal hydrogen atoms of ARG
   PT     P     0.42000    3 ; to add up to -1 net charge
  OI1    OM    -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
  OI2    OM    -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
  OI3    OA    -0.54800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HI3     H     0.39800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
    C     C      0.380     4
    O     O     -0.380     4
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG   CD   gb_26
   CD   NE   gb_20
```

```
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  HH11   gb_2
 NH1  HH12   gb_2
 NH2  HH2   gb_2
 NH2   PT   gb_23
  PT  OI1   gb_23
  PT  OI2   gb_23
  PT  OI3   gb_27
 OI3  HI3   gb_1
 [ exclusions ]
; ai   aj
 HH2  OI1
 HH2  OI2
 HH2  OI3
 HI3  OI1
 HI3  OI2
 HI3  NH2
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H    ga_31
   H   N   CA    ga_17
  -C   N   CA    ga_30
   N  CA    C    ga_12
  CA   C   +N    ga_18
  CA   C    O    ga_29
   O   C   +N    ga_32
   N  CA   CB    ga_12
   C  CA   CB    ga_12
  CA  CB   CG    ga_14
  CB  CG   CD    ga_14
  CG  CD   NE    ga_12
  CD  NE   HE    ga_19
  HE  NE   CZ    ga_22
  CD  NE   CZ    ga_32
  NE  CZ  NH1    ga_27
  NE  CZ  NH2    ga_27
 NH1  CZ  NH2    ga_27
  CZ NH1 HH11    ga_22
  CZ NH1 HH12    ga_22
HH11 NH1 HH12    ga_23
  CZ NH2  HH2    ga_22
 HH2 NH2   PT    ga_19
  CZ NH2   PT    ga_32
 NH2   PT  OI1   ga_13
 NH2   PT  OI2   ga_13
 NH2   PT  OI3   ga_4
 OI1   PT  OI2   ga_28
 OI1   PT  OI3   ga_13
 OI2   PT  OI3   ga_13
  PT  OI3  HI3   ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1 HH11 HH12   CZ    gi_1
 NH2   CZ   PT  HH2    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   NE    gd_17
  CG   CD   NE   CZ    gd_19
  CD   NE   CZ  NH2    gd_4
  NE   CZ  NH1 HH11    gd_4
  NE   CZ  NH2   PT    gd_4
  CZ  NH2   PT  OI3    gd_19
 NH2   PT  OI3  HI3    gd_9
 NH2   PT  OI3  HI3    gd_11

; phosphoarginine (-1)
[ R1P ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.43000    2 ; to add up to 1 net charge
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NE   -0.11000    2 ; from NE nitrogen atom of ARG
 HH2    H    0.24000    2 ; from terminal hydrogen atoms of ARG
  PT    P   -0.09500    3 ; to add up to -2 net charge
 OI1   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
  C    C     0.380    4
  O    O    -0.380    4
[ bonds ]
  N    H   gb_2
  N   CA   gb_20
 CA    C   gb_26
  C    O   gb_4
  C   +N   gb_9
 CA   CB   gb_26
 CB   CG   gb_26
 CG   CD   gb_26
 CD   NE   gb_20
 NE   HE   gb_2
 NE   CZ   gb_10
 CZ  NH1   gb_10
 CZ  NH2   gb_10
NH1 HH11   gb_2
NH1 HH12   gb_2
NH2  HH2   gb_2
NH2   PT   gb_23
 PT  OI1   gb_23
 PT  OI2   gb_23
 PT  OI3   gb_23
[ exclusions ]
; ai   aj
 HH2  OI1
 HH2  OI2
 HH2  OI3
[ angles ]
; ai   aj   ak  gromos type
  -C   N    H    ga_31
   H   N   CA    ga_17
  -C   N   CA    ga_30
   N  CA    C    ga_12
  CA   C   +N    ga_18
  CA   C    O    ga_29
   O   C   +N    ga_32
   N  CA   CB    ga_12
   C  CA   CB    ga_12
  CA  CB   CG    ga_14
  CB  CG   CD    ga_14
  CG  CD   NE    ga_12
  CD  NE   HE    ga_19
  HE  NE   CZ    ga_22
  CD  NE   CZ    ga_32
  NE  CZ  NH1    ga_27
  NE  CZ  NH2    ga_27
 NH1  CZ  NH2    ga_27
  CZ NH1 HH11    ga_22
  CZ NH1 HH12    ga_22
HH11 NH1 HH12    ga_23
  CZ NH2  HH2    ga_22
 HH2 NH2   PT    ga_19
  CZ NH2   PT    ga_32
 NH2   PT  OI1   ga_13
 NH2   PT  OI2   ga_13
 NH2   PT  OI3   ga_13
 OI1   PT  OI2   ga_13
 OI1   PT  OI3   ga_13
 OI2   PT  OI3   ga_13
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1 HH11 HH12   CZ    gi_1
 NH2   CZ   PT  HH2    gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   NE    gd_17
  CG   CD   NE   CZ    gd_19
  CD   NE   CZ  NH2    gd_4
  NE   CZ  NH1 HH11    gd_4
  NE   CZ  NH2   PT    gd_4
  CZ  NH2   PT  OI1    gd_19

; 1-phosphohistidine (-1)
[ H11 ]
 [ atoms ]
   N    N   -0.28000    0
```

```
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C    0.29000    2
  ND1   NR   -0.58000    2
  CE1   CR1    0.29000    2
  CD2   CR1    0.00000    3
  NE2   NR    0.00000    3
  PZ    P    0.42000    4 ; to add up to -1 net charge
  OH1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OH2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OH3   OA   -0.54800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HH3    H    0.39800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C     0.380    5
   O    O    -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
  ND1  CE1   gb_9
  CD2  NE2   gb_9
  CE1  NE2   gb_9
  NE2   PZ   gb_23
  PZ  OH1   gb_23
  PZ  OH2   gb_23
  PZ  OH3   gb_27
  OH3  HH3   gb_1
 [ exclusions ]
 ;  ai    aj
  CB   CE1
  CB   NE2
  PZ   CD2
  PZ   NE2
  HH3   NE2
  HH3   OH1
  HH3   OH2
 [ angles ]
 ;  ai    aj    ak   gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB    ga_12
    C   CA   CB    ga_12
   CA   CB   CG    ga_14
   CB   CG  ND1    ga_36
   CB   CG  CD2    ga_36
  ND1   CG  CD2    ga_6
   CG  ND1  CE1    ga_6
   CG  CD2  NE2    ga_6
  ND1  CE1  NE2    ga_6
  CD2  NE2  CE1    ga_6
  CD2  NE2   PZ    ga_36
  CE1  NE2   PZ    ga_36
  NE2   PZ  OH1    ga_13
  NE2   PZ  OH2    ga_13
  NE2   PZ  OH3    ga_4
  OH1   PZ  OH2    ga_28
  OH1   PZ  OH3    ga_13
  OH2   PZ  OH3    ga_13
   PZ  OH3  HH3    ga_11
 [ impropers ]
 ;  ai    aj    ak    al   gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CG  ND1  CD2   CB    gi_1
  CD2   CG  ND1  CE1    gi_1
  ND1   CG  CD2  NE2    gi_1
   CG  ND1  CE1  NE2    gi_1
   CG  CD2  NE2  CE1    gi_1
  CD2  NE2  CE1  ND1    gi_1
  NE2  CD2  CE1   PZ    gi_1
 [ dihedrals ]
 ;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG  ND1    gd_20
  CD2  NE2   PZ  OH3    gd_19
  NE2   PZ  OH3  HH3    gd_9
  NE2   PZ  OH3  HH3    gd_11

 ; 1-phosphohistidine (-2)
 [ H12 ]
  [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG    C    0.29000    2
  ND1   NR   -0.58000    2
  CE1   CR1    0.29000    2
  CD2   CR1    0.00000    3
  NE2   NR    0.00000    3
   PZ    P   -0.09500    4 ; to add up to -2 net charge
  OH1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OH2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OH3   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C     0.380    5
   O    O    -0.380    5
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG  ND1   gb_9
   CG  CD2   gb_9
  ND1  CE1   gb_9
  CD2  NE2   gb_9
  CE1  NE2   gb_9
  NE2   PZ   gb_23
   PZ  OH1   gb_23
   PZ  OH2   gb_23
   PZ  OH3   gb_23
  [ exclusions ]
 ;  ai    aj
   CB   CE1
   CB   NE2
   PZ   CD2
   PZ   NE2
  [ angles ]
 ;  ai    aj    ak   gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB    ga_12
    C   CA   CB    ga_12
   CA   CB   CG    ga_14
   CB   CG  ND1    ga_36
   CB   CG  CD2    ga_36
  ND1   CG  CD2    ga_6
   CG  ND1  CE1    ga_6
   CG  CD2  NE2    ga_6
  ND1  CE1  NE2    ga_6
  CD2  NE2  CE1    ga_6
  CD2  NE2   PZ    ga_36
  CE1  NE2   PZ    ga_36
  NE2   PZ  OH1    ga_13
  NE2   PZ  OH2    ga_13
  NE2   PZ  OH3    ga_13
  OH1   PZ  OH2    ga_13
  OH1   PZ  OH3    ga_13
  OH2   PZ  OH3    ga_13
  [ impropers ]
 ;  ai    aj    ak    al   gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CG  ND1  CD2   CB    gi_1
  CD2   CG  ND1  CE1    gi_1
  ND1   CG  CD2  NE2    gi_1
   CG  ND1  CE1  NE2    gi_1
   CG  CD2  NE2  CE1    gi_1
  CD2  NE2  CE1  ND1    gi_1
  NE2  CD2  CE1   PZ    gi_1
  [ dihedrals ]
 ;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG  ND1    gd_20
  CD2  NE2   PZ  OH1    gd_19

 ; 3-phosphohistidine (-1)
 [ H31 ]
  [ atoms ]
    N    N   -0.28000    0
```

```
  H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C    0.00000    2
  ND1   NR    0.00000    2
  CD2   CR1    0.29000    3
  CE1   CR1    0.29000    3
  NE2   NR   -0.58000    3
  PE3    P    0.42000    4 ; to add up to -1 net charge
  OZ1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ3   OA   -0.54800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HZ3    H    0.39800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
  C    C     0.380    5
  O    O    -0.380    5
 [ bonds ]
  N    H   gb_2
  N   CA   gb_20
  CA    C   gb_26
  C    O   gb_4
  C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
  ND1  CE1   gb_9
  ND1  PE3   gb_23
  CD2  NE2   gb_9
  CE1  NE2   gb_9
  PE3  OZ1   gb_23
  PE3  OZ2   gb_23
  PE3  OZ3   gb_27
  OZ3  HZ3   gb_1
 [ exclusions ]
; ai   aj
  CB  CE1
  CB  NE2
  CB  PE3
  PE3  CD2
  PE3  NE2
  HZ3  ND1
  HZ3  OZ1
  HZ3  OZ2
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  ND1    ga_36
  CB   CG  CD2    ga_36
 ND1   CG  CD2    ga_6
  CG  ND1  CE1    ga_6
  CG  ND1  PE3    ga_36
 CE1  ND1  PE3    ga_36
  CG  CD2  NE2    ga_6
 ND1  CE1  NE2    ga_6
 CD2  NE2  CE1    ga_6
 ND1  PE3  OZ1    ga_13
 ND1  PE3  OZ2    ga_13
 ND1  PE3  OZ3    ga_4
 OZ1  PE3  OZ2    ga_28
 OZ1  PE3  OZ3    ga_13
 OZ2  PE3  OZ3    ga_13
 PE3  OZ3  HZ3    ga_11
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
 CD2   CG  ND1  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 CD2  NE2  CE1  ND1    gi_1
 ND1   CG  CE1  PE3    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND1    gd_20
  CG  ND1  PE3  OZ3    gd_19
 ND1  PE3  OZ3  HZ3    gd_9
 ND1  PE3  OZ3  HZ3    gd_11
```

```
; 3-phosphohistidine (-2)
[ H32 ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR    0.00000    2
 CD2   CR1    0.29000    3
 CE1   CR1    0.29000    3
 NE2   NR   -0.58000    3
 PE3    P   -0.09500    4 ; to add up to -2 net charge
 OZ1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ3   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  C    C     0.380    5
  O    O    -0.380    5
 [ bonds ]
  N    H   gb_2
  N   CA   gb_20
  CA    C   gb_26
  C    O   gb_4
  C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
 ND1  CE1   gb_9
 ND1  PE3   gb_23
 CD2  NE2   gb_9
 CE1  NE2   gb_9
 PE3  OZ1   gb_23
 PE3  OZ2   gb_23
 PE3  OZ3   gb_23
 [ exclusions ]
; ai   aj
  CB  CE1
  CB  NE2
  CB  PE3
  PE3  CD2
  PE3  NE2
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  ND1    ga_36
  CB   CG  CD2    ga_36
 ND1   CG  CD2    ga_6
  CG  ND1  CE1    ga_6
  CG  CD2  NE2    ga_6
 ND1  CE1  NE2    ga_6
 CD2  NE2  CE1    ga_6
  CG  ND1  PE3    ga_36
 CE1  ND1  PE3    ga_36
 ND1  PE3  OZ1    ga_13
 ND1  PE3  OZ2    ga_13
 ND1  PE3  OZ3    ga_13
 OZ1  PE3  OZ2    ga_13
 OZ2  PE3  OZ3    ga_13
 OZ1  PE3  OZ3    ga_13
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
 CD2   CG  ND1  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 CD2  NE2  CE1  ND1    gi_1
 ND1   CG  CE1  PE3    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND1    gd_20
  CG  ND1  PE3  OZ1    gd_19

; N6-methyllysine (0)
[ KMN ]
```

```
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
  NZ   NT   -0.83000    3 ; from NZ of LYS
  HZ    H    0.41000    3 ; derived from terminal hydrogen atoms of LYS
  CH  CH3    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   CH   gb_20
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ   HZ    ga_10
  HZ   NZ   CH    ga_10
  CE   NZ   CH    ga_12
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N    -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17
  CG   CD   CE   NZ    gd_17
  CD   CE   NZ   CH    gd_14

; N6-methyllysine (+1)
[ KMC ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  NZ   NL    0.10400    3 ; to add up to +1 net charge
 HZ1    H    0.24800    3
 HZ2    H    0.24800    3
  CH  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ  HZ1   gb_2
  NZ  HZ2   gb_2
  NZ   CH   gb_20
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ  HZ1    ga_10
  CE   NZ  HZ2    ga_10
 HZ1   NZ  HZ2     ga_9
 HZ1   NZ   CH    ga_10
 HZ2   NZ   CH    ga_10
  CE   NZ   CH    ga_12
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N    -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17
  CG   CD   CE   NZ    gd_17
  CD   CE   NZ   CH    gd_14

; N6,N6-dimethyllysine (0)
[ K2M ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
  NZ   NT   -0.63000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
 CH1  CH3    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
 CH2  CH3    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ  CH1   gb_20
  NZ  CH2   gb_20
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ  CH1    ga_12
  CE   NZ  CH2    ga_12
 CH1   NZ  CH2    ga_12
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N    -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
```

```
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_4
   -C     N    CA     C     gd_19
    N    CA     C    +N     gd_20
    N    CA    CB    CG     gd_17
   CA    CB    CG    CD     gd_17
   CB    CG    CD    CE     gd_17
   CG    CD    CE    NZ     gd_17
   CD    CE    NZ   CH1     gd_14

; N6,N6-dimethyllysine (+1)
[ K2C ]
 [ atoms ]
    N     N   -0.28000     0
    H     H    0.28000     0
   CA   CH1    0.00000     1
   CB   CH2    0.00000     1
   CG   CH2    0.00000     2
   CD   CH2    0.00000     2
   CE   CH2    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   NZ    NL    0.15200     3 ; to add up to +1 net charge
   HZ     H    0.24800     3
  CH1   CH3    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  CH2   CH3    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
    C     C      0.380     4
    O     O     -0.380     4
 [ bonds ]
    N     H    gb_2
    N    CA    gb_20
   CA     C    gb_26
    C     O    gb_4
    C    +N    gb_9
   CA    CB    gb_26
   CB    CG    gb_26
   CG    CD    gb_26
   CD    CE    gb_26
   CE    NZ    gb_20
   NZ    HZ    gb_2
   NZ   CH1    gb_20
   NZ   CH2    gb_20
 [ angles ]
;  ai    aj    ak   gromos type
   -C     N     H    ga_31
    H     N    CA    ga_17
   -C     N    CA    ga_30
    N    CA     C    ga_12
   CA     C    +N    ga_18
   CA     C     O    ga_29
    O     C    +N    ga_32
    N    CA    CB    ga_12
    C    CA    CB    ga_12
   CA    CB    CG    ga_14
   CB    CG    CD    ga_14
   CG    CD    CE    ga_14
   CD    CE    NZ    ga_14
   CE    NZ    HZ    ga_10
   HZ    NZ   CH1    ga_10
   HZ    NZ   CH2    ga_10
   CE    NZ   CH1    ga_12
   CE    NZ   CH2    ga_12
  CH1    NZ   CH2    ga_12
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA     H    gi_1
    C    CA    +N     O    gi_1
   CA     N     C    CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_4
   -C     N    CA     C     gd_19
    N    CA     C    +N     gd_20
    N    CA    CB    CG     gd_17
   CA    CB    CG    CD     gd_17
   CB    CG    CD    CE     gd_17
   CG    CD    CE    NZ     gd_17
   CD    CE    NZ   CH1     gd_14

; N6,N6,N6-trimethyllysine
[ K3C ]
 [ atoms ]
    N     N   -0.28000     0
    H     H    0.28000     0
   CA   CH1    0.00000     1
   CB   CH2    0.00000     1
   CG   CH2    0.00000     2
   CD   CH2    0.00000     2
   CE   CH2    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   NZ    NL    0.20000     3 ; to add up to +1 net charge
  CH1   CH3    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  CH2   CH3    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  CH3   CH3    0.20000     3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
    C     C      0.380     4
    O     O     -0.380     4
 [ bonds ]
    N     H    gb_2
    N    CA    gb_20
   CA     C    gb_26
    C     O    gb_4
    C    +N    gb_9
   CA    CB    gb_26
   CB    CG    gb_26
   CG    CD    gb_26
   CD    CE    gb_26
   CE    NZ    gb_20
   NZ   CH1    gb_20
   NZ   CH2    gb_20
   NZ   CH3    gb_20
 [ angles ]
;  ai    aj    ak   gromos type
   -C     N     H    ga_31
    H     N    CA    ga_17
   -C     N    CA    ga_30
    N    CA     C    ga_12
   CA     C    +N    ga_18
   CA     C     O    ga_29
    O     C    +N    ga_32
    N    CA    CB    ga_12
    C    CA    CB    ga_12
   CA    CB    CG    ga_14
   CB    CG    CD    ga_14
   CG    CD    CE    ga_14
   CD    CE    NZ    ga_14
   CE    NZ   CH1    ga_12
   CE    NZ   CH2    ga_12
   CE    NZ   CH3    ga_12
  CH1    NZ   CH2    ga_12
  CH1    NZ   CH3    ga_12
  CH2    NZ   CH3    ga_12
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA     H    gi_1
    C    CA    +N     O    gi_1
   CA     N     C    CB    gi_2
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_4
   -C     N    CA     C     gd_19
    N    CA     C    +N     gd_20
    N    CA    CB    CG     gd_17
   CA    CB    CG    CD     gd_17
   CB    CG    CD    CE     gd_17
   CG    CD    CE    NZ     gd_17
   CD    CE    NZ   CH1     gd_14

; omega-N-methylarginine (0)
[ RMN ]
 [ atoms ]
    N     N   -0.28000     0
    H     H    0.28000     0
   CA   CH1    0.00000     1
   CB   CH2    0.00000     1
   CG   CH2    0.00000     2
   CD   CH2    0.00000     2
   NE    NE   -0.28000     3
   HE     H    0.28000     3
   CZ     C    0.18000     4 ; 0.36 charge divided between CZ and CT to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
  NH1    NE   -0.36000     4 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
   CT   CH3    0.18000     4 ; 0.36 charge divided between CZ and CT to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
  NH2    NZ   -0.83000     5
 HH21     H    0.41500     5
 HH22     H    0.41500     5
    C     C      0.380     6
    O     O     -0.380     6
 [ bonds ]
    N     H    gb_2
    N    CA    gb_20
   CA     C    gb_26
    C     O    gb_4
    C    +N    gb_9
   CA    CB    gb_26
   CB    CG    gb_26
   CG    CD    gb_26
   CD    NE    gb_20
   NE    HE    gb_2
```

```
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1   CT   gb_20
 NH2  HH21  gb_2
 NH2  HH22  gb_2
[ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG   CD   ga_14
  CG   CD   NE   ga_12
  CD   NE   HE   ga_19
  HE   NE   CZ   ga_22
  CD   NE   CZ   ga_32
  NE   CZ  NH1   ga_27
  NE   CZ  NH2   ga_27
 NH1   CZ  NH2   ga_27
  CZ  NH1   CT   ga_26
  CZ  NH2  HH21  ga_22
  CZ  NH2  HH22  ga_22
HH21  NH2  HH22  ga_23
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NE   CD   CZ   HE   gi_1
  CZ  NH1  NH2   NE   gi_1
 NH2  HH21 HH22   CZ   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1   gd_4
  NE   CZ  NH1   CT   gd_4
  NE   CZ  NH2  HH21  gd_4

; omega-N-methylarginine (+1)
[ RMC ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.34000    2
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NZ   -0.11000    2 ; from NE atom of ARG
 HH2    H    0.24000    2
 CT  CH3    0.09000    2 ; from CD atom of ARG
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   NE   gb_20
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  HH11  gb_2
 NH1  HH12  gb_2
 NH2  HH2   gb_2
 NH2   CT   gb_20
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG   CD   ga_14
  CG   CD   NE   ga_12
  CD   NE   HE   ga_19
  HE   NE   CZ   ga_22
  CD   NE   CZ   ga_32
  NE   CZ  NH1   ga_27
  NE   CZ  NH2   ga_27
 NH1   CZ  NH2   ga_27
  CZ  NH1  HH11  ga_22
  CZ  NH1  HH12  ga_22
HH11  NH1  HH12  ga_23
  CZ  NH2  HH2   ga_22
 HH2  NH2   CT   ga_19
  CZ  NH2   CT   ga_32
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NE   CD   CZ   HE   gi_1
  CZ  NH1  NH2   NE   gi_1
 NH1  HH11 HH12   CZ   gi_1
 NH2  HH2   CT   CZ   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1   gd_4
  NE   CZ  NH1  HH11  gd_4
  NE   CZ  NH2   CT   gd_4

; symmetric-dimethylarginine (0)
[ RSM ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  NE   NE   -0.28000    3
  HE    H    0.28000    3
  CZ    C    0.18000    4 ; 0.36 charge divided between CZ and CT1 to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
 NH1   NE   -0.36000    4 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
 CT1  CH3    0.18000    4 ; 0.36 charge divided between CZ and CT1 to add up to 0
net charge, similar to methyl groups of other methylation residues and also used in
other building blocks (e.g., TMP)
 NH2   NE   -0.28000    5 ; from NE atom of ARGN
 HH2    H    0.28000    5 ; from HE atom of ARGN
 CT2  CH3    0.00000    6 ; from CD atom of ARGN
   C    C      0.380    7
   O    O     -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   NE   gb_20
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  CT1   gb_20
 NH2  HH2   gb_2
 NH2  CT2   gb_20
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
```

```
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   NE    ga_12
  CD   NE   HE    ga_19
  HE   NE   CZ    ga_22
  CD   NE   CZ    ga_32
  NE   CZ  NH1    ga_27
  NE   CZ  NH2    ga_27
 NH1   CZ  NH2    ga_27
  CZ  NH1  CT1    ga_26
  CZ  NH2  HH2    ga_22
 HH2  NH2  CT2    ga_19
  CZ  NH2  CT2    ga_32
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH2  HH2  CT2   CZ    gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1    gd_4
  NE   CZ  NH1  CT1    gd_4
  NE   CZ  NH2  CT2    gd_4

; symmetric-dimethylarginine (+1)
[ RMS ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.34000    2
 NH1   NZ   -0.11000    2 ; from NE atom of ARG
 HH1    H    0.24000    2
 CT1  CH3    0.09000    2 ; from CD atom of ARG
 NH2   NZ   -0.11000    2 ; from NE atom of ARG
 HH2    H    0.24000    2
 CT2  CH3    0.09000    2 ; from CD atom of ARG
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   NE   gb_20
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  HH1   gb_2
 NH1  CT1   gb_20
 NH2  HH2   gb_2
 NH2  CT2   gb_20
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   NE    ga_12
  CD   NE   HE    ga_19
  HE   NE   CZ    ga_22
  CD   NE   CZ    ga_32
  NE   CZ  NH1    ga_27
  NE   CZ  NH2    ga_27
 NH1   CZ  NH2    ga_27
  CZ  NH1  HH1    ga_22
 HH1  NH1  CT1    ga_19
  CZ  NH1  CT1    ga_32
  CZ  NH2  HH2    ga_22
 HH2  NH2  CT2    ga_19
  CZ  NH2  CT2    ga_32
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1  HH1  CT1   CZ    gi_1
 NH2  HH2  CT2   CZ    gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1    gd_4
  NE   CZ  NH1  CT1    gd_4
  NE   CZ  NH2  CT2    gd_4

; asymmetric-dimethylarginine (0)
[ RAM ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  NE   NE   -0.28000    3
  HE    H    0.28000    3
  CZ    C    0.15000    4
 NH1   NE   -0.54800    4
 HH1    H    0.39800    4
 NH2   NZ   -0.20000    5 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
 CT1  CH3    0.10000    5 ; 0.2 charge divided between CT1 and CT2 to add up to 0
net charge, also similar to methyl groups of methyl-arginine modifications and used
for ring carbons in nucleotides (e.g., DCYT)
 CT2  CH3    0.10000    5 ; 0.2 charge divided between CT1 and CT2 to add up to 0
net charge, also similar to methyl groups of methyl-arginine modifications and used
for ring carbons in nucleotides (e.g., DCYT)
   C    C      0.380    6
   O    O     -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   NE   gb_20
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  HH1   gb_2
 NH2  CT1   gb_20
 NH2  CT2   gb_20
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   NE    ga_12
  CD   NE   HE    ga_19
  HE   NE   CZ    ga_22
  CD   NE   CZ    ga_32
  NE   CZ  NH1    ga_27
  NE   CZ  NH2    ga_27
 NH1   CZ  NH2    ga_27
  CZ  NH1  HH1    ga_22
  CZ  NH2  CT1    ga_27
  CZ  NH2  CT2    ga_27
 CT1  NH2  CT2    ga_27
```

```
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NE   CD   CZ   HE   gi_1
  CZ  NH1  NH2   NE   gi_1
 NH2  CT1  CT2   CZ   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1   gd_4
  NE   CZ  NH1  HH1   gd_4
  NE   CZ  NH2  CT1   gd_4

; asymmetric-dimethylarginine (+1)
[ RMA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.43000    2 ; from CZ of R1P
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NZ   -0.05000    2 ; to add up to +1 net charge, also used for nitrogen atoms
in histidine building blocks
 CT1  CH3    0.09000    2 ; from CD atom of ARG
 CT2  CH3    0.09000    2 ; from CD atom of ARG
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   NE   gb_20
  NE   HE   gb_2
  NE   CZ   gb_10
  CZ  NH1   gb_10
  CZ  NH2   gb_10
 NH1  HH11   gb_2
 NH1  HH12   gb_2
 NH2  CT1   gb_20
 NH2  CT2   gb_20
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG   CD   ga_14
  CG   CD   NE   ga_12
  CD   NE   HE   ga_19
  HE   NE   CZ   ga_22
  CD   NE   CZ   ga_32
  NE   CZ  NH1   ga_27
  NE   CZ  NH2   ga_27
 NH1   CZ  NH2   ga_27
  CZ  NH1  HH11   ga_22
  CZ  NH1  HH12   ga_22
HH11  NH1  HH12   ga_23
  CZ  NH2  CT1   ga_27
  CZ  NH2  CT2   ga_27
 CT1  NH2  CT2   ga_27
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NE   CD   CZ   HE   gi_1
  CZ  NH1  NH2   NE   gi_1
 NH1  HH11  HH12   CZ   gi_1
 NH2  CT1  CT2   CZ   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   NE   gd_17
  CG   CD   NE   CZ   gd_19
  CD   NE   CZ  NH1   gd_4
  NE   CZ  NH1  HH11   gd_4
  NE   CZ  NH2  CT1   gd_4

; 1-methylhistidine (0)
[ H1M ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.13000    2
 ND1   NR   -0.58000    2
 CD2  CR1    0.00000    2
 CE1  CR1    0.26000    2
 NE2   NR    0.00000    2
  CZ  CH3    0.19000    2 ; from HE2 in HISB
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
 ND1  CE1   gb_9
 CD2  NE2   gb_9
 CE1  NE2   gb_9
 NE2   CZ   gb_21
 [ exclusions ]
;  ai   aj
  CB  CE1
  CB  NE2
  CG   CZ
 ND1   CZ
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG  ND1   ga_36
  CB   CG  CD2   ga_36
 ND1   CG  CD2    ga_6
  CG  ND1  CE1    ga_6
  CG  CD2  NE2    ga_6
 ND1  CE1  NE2    ga_6
 CD2  NE2  CE1    ga_6
 CD2  NE2   CZ   ga_36
 CE1  NE2   CZ   ga_36
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
 CD2   CG  ND1  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 CD2  NE2  CE1  ND1   gi_1
 NE2  CD2  CE1   CZ   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  ND1   gd_20

; 1-methylhistidine (+1)
[ H1C ]
```

```
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.05000    2
 ND1   NR    0.38000    2
 HD1    H    0.30000    2
 CD2  CR1    0.00000    2
 CE1  CR1   -0.14000    2
 NE2   NR    0.31000    2
  CZ  CH3    0.20000    2 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
 ND1  HD1   gb_2
 ND1  CE1   gb_9
 CD2  NE2   gb_9
 CE1  NE2   gb_9
 NE2   CZ   gb_21
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  CE1
  CB  NE2
  CG   CZ
 ND1   CZ
 HD1  CD2
 HD1  NE2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  ND1    ga_36
  CB   CG  CD2    ga_36
 ND1   CG  CD2    ga_6
  CG  ND1  HD1    ga_35
  CG  ND1  CE1    ga_6
 HD1  ND1  CE1    ga_35
  CG  CD2  NE2    ga_6
 ND1  CE1  NE2    ga_6
 CD2  NE2  CE1    ga_6
 CD2  NE2   CZ    ga_36
 CE1  NE2   CZ    ga_36
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
 CD2   CG  ND1  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 CD2  NE2  CE1  ND1    gi_1
 ND1   CG  CE1  HD1    gi_1
 NE2  CD2  CE1   CZ    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND1    gd_20

; 3-methylhistidine (0)
[ H3M ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR    0.00000    2
 CD2  CR1    0.13000    2
 CE1  CR1    0.26000    2
 CE3  CH3    0.19000    2 ; from HD1 in HISA
 NE2   NR   -0.58000    2
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
  CG  CD2   gb_9
 ND1  CE1   gb_9
 ND1  CE3   gb_21
 CD2  NE2   gb_9
 CE1  NE2   gb_9
 [ exclusions ]
;  ai   aj
  CB  CE3
  CB  CE1
  CB  NE2
 CE3  CD2
 CE3  NE2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  ND1    ga_36
  CB   CG  CD2    ga_36
 ND1   CG  CD2    ga_6
  CG  ND1  CE3    ga_36
  CG  ND1  CE1    ga_6
 CE3  ND1  CE1    ga_36
  CG  CD2  NE2    ga_6
 ND1  CE1  NE2    ga_6
 CD2  NE2  CE1    ga_6
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
 CD2   CG  ND1  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 CD2  NE2  CE1  ND1    gi_1
 ND1   CG  CE1  CE3    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND1    gd_20

; 3-methylhistidine (+1)
[ H3C ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.05000    2
 ND1   NR    0.38000    2
 CD2  CR1    0.00000    2
 CE1  CR1   -0.14000    2
 CE3  CH3    0.20000    2 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
 NE2   NR    0.31000    2
 HE2    H    0.30000    2
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  ND1   gb_9
```

```
  CG   CD2   gb_9
  ND1   CE3   gb_21
  ND1   CE1   gb_9
  CD2   NE2   gb_9
  CE1   NE2   gb_9
  NE2   HE2   gb_2
 [ exclusions ]
;  ai    aj
  CB   CE3
  CB   CE1
  CB   NE2
  CG   HE2
  ND1   HE2
  CE3   CD2
  CE3   NE2
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  ND1    ga_36
  CB   CG  CD2    ga_36
  ND1   CG  CD2    ga_6
  CG  ND1  CE3    ga_36
  CG  ND1  CE1    ga_6
  CE3  ND1  CE1    ga_36
  CG  CD2  NE2    ga_6
  ND1  CE1  NE2    ga_6
  CD2  NE2  CE1    ga_6
  CD2  NE2  HE2    ga_35
  CE1  NE2  HE2    ga_35
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
  CD2   CG  ND1  CE1    gi_1
  ND1   CG  CD2  NE2    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
  CD2  NE2  CE1  ND1    gi_1
  ND1   CG  CE1  CE3    gi_1
  NE2  CD2  CE1  HE2    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND1    gd_20

; N5-methylglutamine
[ QME ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  CD    C    0.28000    2 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
  OE1    O   -0.38000    2
  NE2    N   -0.28000    3 ; from the peptide bond
  HE2    H    0.28000    3 ; from the peptide bond
  CZ  CH3    0.00000    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD  OE1   gb_4
  CD  NE2   gb_9
  NE2  HE2   gb_2
  NE2   CZ   gb_20
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD  OE1    ga_29
  CG   CD  NE2    ga_18
 OE1   CD  NE2    ga_32
  CD  NE2  HE2    ga_31
 HE2  NE2   CZ    ga_17
  CD  NE2   CZ    ga_30
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD  OE1  NE2   CG    gi_1
 NE2   CZ   CD  HE2    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD  NE2    gd_20
  CG   CD  NE2   CZ    gd_4

; N4-methylasparagine
[ NME ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  CG    C    0.28000    2 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
  OD1    O   -0.38000    2
  ND2    N   -0.28000    3 ; from the peptide bond
  HD2    H    0.28000    3 ; from the peptide bond
  CE  CH3    0.00000    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  ND2   gb_9
 ND2  HD2   gb_2
 ND2   CE   gb_20
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  OD1    ga_29
  CB   CG  ND2    ga_18
 OD1   CG  ND2    ga_32
  CG  ND2  HD2    ga_31
 HD2  ND2   CE    ga_17
  CG  ND2   CE    ga_30
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  OD1  ND2   CB    gi_1
 ND2   CE   CG  HD2    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  ND2    gd_20
  CB   CG  ND2   CE    gd_4
```

```
; glutamate methyl ester
[ EME ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD    C    0.56000    2 ; from the carbonyl group (of e.g., GLU) with the partial
charge increased by 0.18 - see the comment of CZ atom
 OE1    O   -0.38000    2 ; from the carbonyl group (of e.g., GLU)
 OE2   OA   -0.36000    2 ; from the ring oxygen atom of nucleotides (e.g., ATP)
  CZ  CH3    0.18000    2 ; 0.36 charge divided between CD and CZ to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD  OE1   gb_4
  CD  OE2   gb_12
 OE2   CZ   gb_17
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD  OE1    ga_29
  CG   CD  OE2    ga_18
 OE1   CD  OE2    ga_32
  CD  OE2   CZ    ga_11
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CD  OE1  OE2   CG   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD  OE2   gd_20
  CG   CD  OE2   CZ   gd_3

; aspartate methyl ester
[ DMA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.56000    2 ; from the carbonyl group (of e.g., GLU) with the partial
charge increased by 0.18 - see the comment of CE atom
 OD1    O   -0.38000    2 ; from the carbonyl group (of e.g., GLU)
 OD2   OA   -0.36000    2 ; from the ring oxygen atom of nucleotides (e.g., ATP)
  CE  CH3    0.18000    2 ; 0.36 charge divided between CG and CE to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  OD2   gb_12
 OD2   CE   gb_17
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  OD1    ga_29
  CB   CG  OD2    ga_18
 OD1   CG  OD2    ga_32
  CG  OD2   CE    ga_11
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  OD1  OD2   CB   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  OD2   gd_20
  CB   CG  OD2   CE   gd_3

; S-methylcysteine
[ CYM ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1 ; from MET
  SG    S    0.00000    2 ; from MET
  CD  CH3    0.00000    2 ; from MET
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   SG   gb_30
  SG   CD   gb_29
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   SG    ga_15
  CB   SG   CD    ga_3
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   SG   gd_17
  CA   CB   SG   CD   gd_13

; N6-acetyllysine
[ KAC ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ    N   -0.28000    4 ; from the peptide bond
  HZ    H    0.28000    4 ; from the peptide bond
  CH    C    0.28000    5 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
 OI2    O   -0.38000    5 ; from the carbonyl group (of e.g., GLN)
 CI1  CH3    0.10000    5 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
   C    C      0.380    6
   O    O     -0.380    6
```

```
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   CH   gb_9
  CH  OI2   gb_4
  CH  CI1   gb_26
 [ angles ]
 ; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ   HZ    ga_17
  HZ   NZ   CH    ga_31
  CE   NZ   CH    ga_30
  NZ   CH  OI2    ga_32
  NZ   CH  CI1    ga_18
 OI2   CH  CI1    ga_29
 [ impropers ]
 ; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NZ   CH   CE   HZ   gi_1
  CH  CI1   NZ  OI2   gi_1
 [ dihedrals ]
 ; ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   CE   gd_17
  CG   CD   CE   NZ   gd_17
  CD   CE   NZ   CH   gd_19
  CE   NZ   CH  CI1   gd_4

 ; 3-hydroxyproline (R)
 [ PH3 ]
  [ atoms ]
    N    N    0.00000    0
   CA  CH1    0.00000    1
   CB  CH1    0.15000    2 ; from the hydroxyl group of THR
  OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
  HG1    H    0.39800    2 ; from the hydroxyl group of THR
  CG2  CH2    0.00000    3
  CD   CH2    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
  [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
 OG1  HG1   gb_1
  CB  CG2   gb_26
 CG2   CD   gb_26
  CD    N   gb_20
 [ angles ]
 ; ai   aj   ak  gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
 OG1   CB  CG2    ga_12
  CB  CG2   CD    ga_12
 CG2   CD    N    ga_12
  CD    N   CA    ga_20
  -C    N   CD    ga_30
 [ impropers ]
 ; ai   aj   ak   al  gromos type
   N   -C   CA   CD   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CB  CG2   CA  OG1   gi_2
 [ dihedrals ]
 ; ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB  CG2   gd_17
  CA   CB  OG1  HG1   gd_12
  CA   CB  CG2   CD   gd_17
  CB  CG2   CD    N   gd_17
 CG2   CD    N   CA   gd_19

 ; 3-hydroxyproline (S)
 [ P3H ]
  [ atoms ]
    N    N    0.00000    0
   CA  CH1    0.00000    1
   CB  CH1    0.15000    2 ; from the hydroxyl group of THR
  OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
  HG1    H    0.39800    2 ; from the hydroxyl group of THR
  CG2  CH2    0.00000    3
  CD   CH2    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
  [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
 OG1  HG1   gb_1
  CB  CG2   gb_26
 CG2   CD   gb_26
  CD    N   gb_20
 [ angles ]
 ; ai   aj   ak  gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
 OG1   CB  CG2    ga_12
  CB  CG2   CD    ga_12
 CG2   CD    N    ga_12
  CD    N   CA    ga_20
  -C    N   CD    ga_30
 [ impropers ]
 ; ai   aj   ak   al  gromos type
   N   -C   CA   CD   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CB  CA   CG2  OG1   gi_2
 [ dihedrals ]
 ; ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB  CG2   gd_17
  CA   CB  OG1  HG1   gd_12
  CA   CB  CG2   CD   gd_17
  CB  CG2   CD    N   gd_17
 CG2   CD    N   CA   gd_19

 ; 4-hydroxyproline (S)
 [ HY2 ]
  [ atoms ]
    N    N    0.00000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH1    0.15000    2 ; from the hydroxyl group of THR
  OD1   OA   -0.54800    2 ; from the hydroxyl group of THR
  HD1    H    0.39800    2 ; from the hydroxyl group of THR
  CD2  CH2    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
  [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
```

```
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_17
 OD1  HD1   gb_1
  CG  CD2   gb_26
 CD2    N   gb_20
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_12
  CB   CG  OD1    ga_12
  CB   CG  CD2    ga_12
  CG  OD1  HD1    ga_11
 OD1   CG  CD2    ga_12
  CG  CD2    N    ga_12
 CD2    N   CA    ga_20
  -C    N  CD2    ga_30
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA  CD2    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG   CB  CD2  OD1    gi_2
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD2   gd_17
  CB   CG  OD1  HD1   gd_12
  CB   CG  CD2    N   gd_17
  CG  CD2    N   CA   gd_19

; 3,4-dihydroxyproline
[ PHH ]
 [ atoms ]
   N    N   0.00000    0
  CA  CH1   0.00000    1
  CB  CH1   0.15000    2 ; from the hydroxyl group of THR
  OG1   OA  -0.54800    2 ; from the hydroxyl group of THR
  HG1    H   0.39800    2 ; from the hydroxyl group of THR
  CG2  CH1   0.15000    3 ; from the hydroxyl group of THR
  OD1   OA  -0.54800    3 ; from the hydroxyl group of THR
  HD1    H   0.39800    3 ; from the hydroxyl group of THR
  CD2  CH2   0.00000    4
   C    C    0.380    5
   O    O   -0.380    5
 [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
 OG1  HG1   gb_1
  CB  CG2   gb_26
 CG2  OD1   gb_17
 OD1  HD1   gb_1
 CG2  CD2   gb_26
 CD2    N   gb_20
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
 OG1   CB  CG2    ga_12
  CB  CG2  OD1    ga_12
  CB  CG2  CD2    ga_12
 CG2  OD1  HD1    ga_11
 OD1  CG2  CD2    ga_12
 CG2  CD2    N    ga_12
 CD2    N   CA    ga_20
  -C    N  CD2    ga_30
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA  CD2    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB   CA  CG2  OG1    gi_2
 CG2   CB  CD2  OD1    gi_2
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C   gd_19
    N   CA    C   +N   gd_20
    N   CA   CB  CG2   gd_17
   CA   CB  OG1  HG1   gd_12
   CA   CB  CG2  CD2   gd_17
   CB  CG2  OD1  HD1   gd_12
   CB  CG2  CD2    N   gd_17
  CG2  CD2    N   CA   gd_19

; 5-hydroxylysine (0,R)
[ KH5 ]
 [ atoms ]
   N    N  -0.28000    0
   H    H   0.28000    0
  CA  CH1   0.00000    1
  CB  CH2   0.00000    1
  CG  CH2   0.00000    2
  CD  CH1   0.15000    3 ; from the hydroxyl group of THR
 OE1   OA  -0.54800    3 ; from the hydroxyl group of THR
 HE1    H   0.39800    3 ; from the hydroxyl group of THR
 CE2  CH2   0.00000    4
  NZ   NT  -0.83000    5
 HZ1    H   0.41500    5
 HZ2    H   0.41500    5
   C    C    0.380    6
   O    O   -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD  OE1   gb_17
  CD  CE2   gb_26
 OE1  HE1   gb_1
 CE2   NZ   gb_20
  NZ  HZ1   gb_2
  NZ  HZ2   gb_2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG   CD   ga_14
  CG   CD  OE1   ga_12
  CG   CD  CE2   ga_12
 OE1   CD  CE2   ga_12
  CD  OE1  HE1   ga_11
  CD  CE2   NZ   ga_14
 CE2   NZ  HZ1   ga_10
 CE2   NZ  HZ2   ga_10
 HZ1   NZ  HZ2   ga_9
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CD  CE2  OE1   CG   gi_2
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD  CE2   gd_17
  CG   CD  OE1  HE1   gd_12
  CG   CD  CE2   NZ   gd_17
  CD  CE2   NZ  HZ1   gd_14

; 5-hydroxylysine (0,S)
[ K5H ]
 [ atoms ]
   N    N  -0.28000    0
   H    H   0.28000    0
  CA  CH1   0.00000    1
  CB  CH2   0.00000    1
  CG  CH2   0.00000    2
  CD  CH1   0.15000    3 ; from the hydroxyl group of THR
```

```
  OE1   OA   -0.54800    3 ; from the hydroxyl group of THR
  HE1   H    0.39800    3 ; from the hydroxyl group of THR
  CE2   CH2    0.00000    4
  NZ   NT   -0.83000    5
  HZ1   H    0.41500    5
  HZ2   H    0.41500    5
   C    C      0.380    6
   O    O     -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   OE1   gb_17
  CD   CE2   gb_26
  OE1   HE1   gb_1
  CE2   NZ   gb_20
  NZ   HZ1   gb_2
  NZ   HZ2   gb_2
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   OE1    ga_12
  CG   CD   CE2    ga_12
  OE1   CD   CE2    ga_12
  CD   OE1   HE1    ga_11
  CD   CE2   NZ    ga_14
  CE2   NZ   HZ1    ga_10
  CE2   NZ   HZ2    ga_10
  HZ1   NZ   HZ2    ga_9
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD   OE1   CE2   CG    gi_2
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE2    gd_17
  CG   CD   OE1   HE1    gd_12
  CG   CD   CE2   NZ    gd_17
  CD   CE2   NZ   HZ1    gd_14

; 5-hydroxylysine (+1,R)
[ KPH ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    2
  CD   CH1    0.15000    3 ; from the hydroxyl group of THR
  OE1   OA   -0.54800    3 ; from the hydroxyl group of THR
  HE1   H    0.39800    3 ; from the hydroxyl group of THR
  CE2   CH2    0.12700    4
  NZ   NL    0.12900    4
  HZ1   H    0.24800    4
  HZ2   H    0.24800    4
  HZ3   H    0.24800    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   OE1   gb_17
  CD   CE2   gb_26
  OE1   HE1   gb_1
  CE2   NZ   gb_20
  NZ   HZ1   gb_2
  NZ   HZ2   gb_2
  NZ   HZ3   gb_2
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   OE1    ga_12
  CG   CD   CE2    ga_12
  OE1   CD   CE2    ga_12
  CD   OE1   HE1    ga_11
  CD   CE2   NZ    ga_14
  CE2   NZ   HZ1    ga_10
  CE2   NZ   HZ2    ga_10
  CE2   NZ   HZ3    ga_10
  HZ1   NZ   HZ2    ga_9
  HZ1   NZ   HZ3    ga_9
  HZ2   NZ   HZ3    ga_9
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD   CE2   OE1   CG    gi_2
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE2    gd_17
  CG   CD   OE1   HE1    gd_12
  CG   CD   CE2   NZ    gd_17
  CD   CE2   NZ   HZ1    gd_14

; 5-hydroxylysine (+1,S)
[ KHP ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    2
  CD   CH1    0.15000    3 ; from the hydroxyl group of THR
  OE1   OA   -0.54800    3 ; from the hydroxyl group of THR
  HE1   H    0.39800    3 ; from the hydroxyl group of THR
  CE2   CH2    0.12700    4
  NZ   NL    0.12900    4
  HZ1   H    0.24800    4
  HZ2   H    0.24800    4
  HZ3   H    0.24800    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   OE1   gb_17
  CD   CE2   gb_26
  OE1   HE1   gb_1
  CE2   NZ   gb_20
  NZ   HZ1   gb_2
  NZ   HZ2   gb_2
  NZ   HZ3   gb_2
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   OE1    ga_12
```

```
  CG   CD  CE2    ga_12
  OE1   CD  CE2    ga_12
  CD  OE1  HE1    ga_11
  CD  CE2   NZ    ga_14
  CE2   NZ  HZ1    ga_10
  CE2   NZ  HZ2    ga_10
  CE2   NZ  HZ3    ga_10
  HZ1   NZ  HZ2    ga_9
  HZ1   NZ  HZ3    ga_9
  HZ2   NZ  HZ3    ga_9
 [ impropers ]
 ; ai   aj   ak   al   gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CD  OE1  CE2   CG    gi_2
 [ dihedrals ]
 ; ai   aj   ak   al   gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG   CD    gd_17
   CB   CG   CD  CE2    gd_17
   CG   CD  OE1  HE1    gd_12
   CG   CD  CE2   NZ    gd_17
   CD  CE2   NZ  HZ1    gd_14

 ; 3,4-dihydroxyphenylalanine
 [ HTY ]
  [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.00000    1
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C    0.15000    4 ; from the hydroxyl group of TYR
  OZ1   OA   -0.54800    4 ; from the hydroxyl group of TYR
  HZ1    H    0.39800    4 ; from the hydroxyl group of TYR
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
  CZ2    C    0.15000    6 ; from the hydroxyl group of TYR
   OH   OA   -0.54800    6 ; from the hydroxyl group of TYR
   HH    H    0.39800    6 ; from the hydroxyl group of TYR
    C    C      0.380    7
    O    O     -0.380    7
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG  CD1   gb_15
   CG  CD2   gb_15
  CD1  HD1   gb_3
  CD1  CE1   gb_15
  CD2  HD2   gb_3
  CD2  CE2   gb_15
  CE1  CZ2   gb_15
  CE1  OZ1   gb_12
  OZ1  HZ1   gb_1
  CE2  HE2   gb_3
  CE2  CZ2   gb_15
  CZ2   OH   gb_12
   OH   HH   gb_1
  [ exclusions ]
  ; ai   aj
   CB  HD1
   CB  HD2
   CB  CE1
   CB  CE2
   CG  OZ1
   CG  HE2
   CG  CZ2
  CD1  HD2
  CD1  CE2
  CD1   OH
  HD1  CD2
  HD1  OZ1
  HD1  CZ2
  CD2  CE1
  CD2   OH
  HD2  HE2
  HD2  CZ2
  CE1  HE2
  OZ1  CE2
  OZ1   OH
  HE2   OH
  HH   OZ1
  HH   HZ1
  HZ1   OH
  [ angles ]
  ; ai   aj   ak  gromos type
   -C    N    H   ga_31
    H    N   CA   ga_17
   -C    N   CA   ga_30
    N   CA    C   ga_12
   CA    C   +N   ga_18
   CA    C    O   ga_29
    O    C   +N   ga_32
    N   CA   CB   ga_12
    C   CA   CB   ga_12
   CA   CB   CG   ga_14
   CB   CG  CD1   ga_26
   CB   CG  CD2   ga_26
  CD1   CG  CD2   ga_26
   CG  CD1  HD1   ga_24
  HD1  CD1  CE1   ga_24
   CG  CD1  CE1   ga_26
   CG  CD2  HD2   ga_24
  HD2  CD2  CE2   ga_24
   CG  CD2  CE2   ga_26
  CD1  CE1  OZ1   ga_26
  CD1  CE1  CZ2   ga_26
  OZ1  CE1  CZ2   ga_26
  CE1  OZ1  HZ1   ga_11
  CD2  CE2  HE2   ga_24
  HE2  CE2  CZ2   ga_24
  CD2  CE2  CZ2   ga_26
  CE1  CZ2  CE2   ga_26
  CE1  CZ2   OH   ga_26
  CE2  CZ2   OH   ga_26
  CZ2   OH   HH   ga_11
  [ impropers ]
  ; ai   aj   ak   al   gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CG  CD1  CD2   CB    gi_1
  CD2   CG  CD1  CE1    gi_1
  CD1   CG  CD2  CE2    gi_1
   CG  CD1  CE1  CZ2    gi_1
   CG  CD2  CE2  CZ2    gi_1
  CD1  CE1  CZ2  CE2    gi_1
  CD2  CE2  CZ2  CE1    gi_1
  CD1   CG  CE1  HD1    gi_1
  CD2   CG  CE2  HD2    gi_1
  CE1  CZ2  CD1  OZ1    gi_1
  CE2  CZ2  CD2  HE2    gi_1
  CZ2  CE1  CE2   OH    gi_1
  [ dihedrals ]
  ; ai   aj   ak   al   gromos type
   -CA   -C    N   CA    gd_4
    -C    N   CA    C    gd_19
     N   CA    C   +N    gd_20
     N   CA   CB   CG    gd_17
    CA   CB   CG  CD1    gd_20
   CD1  CE1  OZ1  HZ1    gd_2
   CE1  CZ2   OH   HH    gd_2

 ; 7-hydroxytryptophan
 [ W7H ]
  [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C   -0.14000    2
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C    0.00000    2
  NE1   NR   -0.05000    2
  HE1    H    0.19000    2
  CE2    C    0.00000    2
  CE3    C   -0.10000    3
  HE3   HC    0.10000    3
  CZ2    C    0.15000    4 ; from the hydroxyl group of TYR
  OH2   OA   -0.54800    4 ; from the hydroxyl group of TYR
  HH2    H    0.39800    4 ; from the hydroxyl group of TYR
  CZ3    C   -0.10000    5
  HZ3   HC    0.10000    5
  CH3    C   -0.10000    6
  HH3   HC    0.10000    6
    C    C      0.380    7
    O    O     -0.380    7
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
```

```
  C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_9
  CG  CD2   gb_15
  CD1  HD1   gb_3
  CD1  NE1   gb_9
  CD2  CE2   gb_15
  CD2  CE3   gb_15
  NE1  HE1   gb_2
  NE1  CE2   gb_9
  CE2  CZ2   gb_15
  CE3  HE3   gb_3
  CE3  CZ3   gb_15
  CZ2  OH2   gb_12
  CZ2  CH3   gb_15
  CZ3  HZ3   gb_3
  CZ3  CH3   gb_15
  OH2  HH2   gb_1
  CH3  HH3   gb_3
 [ exclusions ]
 ;  ai    aj
  CB   HD1
  CB   NE1
  CB   CE2
  CB   CE3
  CG   HE1
  CG   HE3
  CG   CZ2
  CG   CZ3
  CD1  CE3
  CD1  CZ2
  HD1  CD2
  HD1  HE1
  HD1  CE2
  CD2  HE1
  CD2  HZ3
  CD2  OH2
  CD2  CH3
  NE1  CE3
  NE1  OH2
  NE1  CH3
  HE1  CZ2
  CE2  HE3
  CE2  CZ3
  CE2  HH3
  CE3  CZ2
  CE3  HH3
  HE3  HZ3
  HE3  CH3
  CZ2  HZ3
  CZ3  OH2
  HZ3  HH3
  OH2  HH3
 [ angles ]
 ;  ai    aj    ak  gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB    ga_12
    C   CA   CB    ga_12
   CA   CB   CG    ga_14
   CB   CG  CD1    ga_36
   CB   CG  CD2    ga_36
  CD1   CG  CD2    ga_6
   CG  CD1  HD1    ga_35
  HD1  CD1  NE1    ga_35
   CG  CD1  NE1    ga_6
   CG  CD2  CE2    ga_6
  CD1  NE1  CE2    ga_6
  CD1  NE1  HE1    ga_35
  HE1  NE1  CE2    ga_35
  NE1  CE2  CD2    ga_6
   CG  CD2  CE3    ga_38
  NE1  CE2  CZ2    ga_38
  CD2  CE2  CZ2    ga_26
  CE2  CD2  CE3    ga_26
  CD2  CE3  HE3    ga_24
  HE3  CE3  CZ3    ga_24
  CD2  CE3  CZ3    ga_26
  CE2  CZ2  OH2    ga_26
  CE2  CZ2  CH3    ga_26
  OH2  CZ2  CH3    ga_26
  CE3  CZ3  HZ3    ga_24
  HZ3  CZ3  CH3    ga_24
  CE3  CZ3  CH3    ga_26
  CZ2  OH2  HH2    ga_11
  CZ2  CH3  HH3    ga_24
  HH3  CH3  CZ3    ga_24
  CZ2  CH3  CZ3    ga_26
 [ impropers ]
 ;  ai    aj    ak    al  gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CG  CD1  CD2   CB    gi_1
  CD2   CG  CD1  NE1    gi_1
  CD1   CG  CD2  CE2    gi_1
   CG  CD1  NE1  CE2    gi_1
   CG  CD2  CE2  NE1    gi_1
  CD1  NE1  CE2  CD2    gi_1
  CD1   CG  NE1  HD1    gi_1
  NE1  CD1  CE2  HE1    gi_1
  CD2  CE2  CE3   CG    gi_1
  CE2  CD2  CZ2  NE1    gi_1
  CE3  CD2  CE2  CZ2    gi_1
  CD2  CE2  CZ2  CH3    gi_1
  CE2  CD2  CE3  CZ3    gi_1
  CE2  CZ2  CH3  CZ3    gi_1
  CD2  CE3  CZ3  CH3    gi_1
  CE3  CZ3  CH3  CZ2    gi_1
  CE3  CD2  CZ3  HE3    gi_1
  CZ2  CE2  CH3  OH2    gi_1
  CZ3  CE3  CH3  HZ3    gi_1
  CH3  CZ2  CZ3  HH3    gi_1
 [ dihedrals ]
 ;  ai    aj    ak    al  gromos type
  -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  CD2    gd_20
  CE2  CZ2  OH2  HH2    gd_2

; 3-hydroxyaspartate (-1,R)
[ DH3 ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.27000    3
 OD1   OM   -0.63500    3
 OD2   OM   -0.63500    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_5
 CG2  OD2   gb_5
 [ angles ]
 ;  ai    aj    ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_21
  CB  CG2  OD2    ga_21
 OD1  CG2  OD2    ga_37
 [ impropers ]
 ;  ai    aj    ak    al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  CG2  OG1   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
 [ dihedrals ]
 ;  ai    aj    ak    al  gromos type
  -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  CG2    gd_17
  CA   CB  OG1  HG1    gd_12
```

```
  CA   CB  CG2  OD1    gd_20

; 3-hydroxyaspartate (-1,S)
[ D3H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.27000    3
 OD1   OM   -0.63500    3
 OD2   OM   -0.63500    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_5
 CG2  OD2   gb_5
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_21
  CB  CG2  OD2    ga_21
 OD1  CG2  OD2    ga_37
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  CG2    gd_17
  CA   CB  OG1  HG1    gd_12
  CA   CB  CG2  OD1    gd_20

; 3-hydroxyaspartate (0,R)
[ DN3 ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.53000    3
 OD1    O   -0.38000    3
 OD2   OA   -0.54800    3
 HD2    H    0.39800    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_4
 CG2  OD2   gb_12
 OD2  HD2   gb_1
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_29
  CB  CG2  OD2    ga_18
 OD1  CG2  OD2    ga_32
 CG2  OD2  HD2    ga_11
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  CG2    gd_17
  CA   CB  OG1  HG1    gd_12
  CA   CB  CG2  OD2    gd_20
  CB  CG2  OD2  HD2    gd_3

; 3-hydroxyaspartate (0,S)
[ D3N ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.53000    3
 OD1    O   -0.38000    3
 OD2   OA   -0.54800    3
 HD2    H    0.39800    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_4
 CG2  OD2   gb_12
 OD2  HD2   gb_1
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  OG1    ga_12
  CA   CB  CG2    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_29
  CB  CG2  OD2    ga_18
 OD1  CG2  OD2    ga_32
 CG2  OD2  HD2    ga_11
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB  CG2    gd_17
```

```
  CA   CB   OG1  HG1   gd_12
  CA   CB   CG2  OD2   gd_20
  CB   CG2  OD2  HD2   gd_3

; 3-hydroxyasparagine (R)
[ N3H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.38000    3
 OD1    O   -0.38000    3
 ND2   NT   -0.83000    4
HD21    H    0.41500    4
HD22    H    0.41500    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_4
 CG2  ND2   gb_8
 ND2 HD21   gb_2
 ND2 HD22   gb_2
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  CG2    ga_12
  CA   CB  OG1    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_29
  CB  CG2  ND2    ga_18
 OD1  CG2  ND2    ga_32
 CG2  ND2 HD21    ga_22
 CG2  ND2 HD22    ga_22
HD21  ND2 HD22    ga_23
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  CG2  OG1   CA    gi_2
 CG2  OD1  ND2   CB    gi_1
 ND2 HD21 HD22  CG2    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB  CG2   gd_17
  CA   CB  OG1  HG1   gd_12
  CA   CB  CG2  ND2   gd_20
  CB  CG2  ND2 HD21    gd_4

; 3-hydroxyasparagine (S)
[ NH3 ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1   OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1    H    0.39800    2 ; from the hydroxyl group of THR
 CG2    C    0.38000    3
 OD1    O   -0.38000    3
 ND2   NT   -0.83000    4
HD21    H    0.41500    4
HD22    H    0.41500    4
   C    C      0.380    5
   O    O     -0.380    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_17
  CB  CG2   gb_26
 OG1  HG1   gb_1
 CG2  OD1   gb_4
 CG2  ND2   gb_8
 ND2 HD21   gb_2
 ND2 HD22   gb_2
[ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB  CG2    ga_12
  CA   CB  OG1    ga_12
 OG1   CB  CG2    ga_12
  CB  OG1  HG1    ga_11
  CB  CG2  OD1    ga_29
  CB  CG2  ND2    ga_18
 OD1  CG2  ND2    ga_32
 CG2  ND2 HD21    ga_22
 CG2  ND2 HD22    ga_22
HD21  ND2 HD22    ga_23
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
 CG2  OD1  ND2   CB    gi_1
 ND2 HD21 HD22  CG2    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB  CG2   gd_17
  CA   CB  OG1  HG1   gd_12
  CA   CB  CG2  ND2   gd_20
  CB  CG2  ND2 HD21    gd_4

; 4-carboxyglutamate (-2)
[ ECA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH1    0.00000    1
 CD1    C    0.27000    2
 OE1   OM   -0.63500    2
 OE2   OM   -0.63500    2
 CD2    C    0.27000    3 ; from GLU
 OE3   OM   -0.63500    3 ; from GLU
 OE4   OM   -0.63500    3 ; from GLU
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_26
 CD1  OE1   gb_5
 CD1  OE2   gb_5
  CG  CD2   gb_26
 CD2  OE3   gb_5
 CD2  OE4   gb_5
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_12
  CB   CG  CD2    ga_12
```

```
  CD1   CG   CD2    ga_12
  CG   CD1   OE1    ga_21
  CG   CD1   OE2    ga_21
 OE1   CD1   OE2    ga_37
  CG   CD2   OE3    ga_21
  CG   CD2   OE4    ga_21
 OE3   CD2   OE4    ga_37
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N    -C    CA    H     gi_1
   C    CA    +N    O     gi_1
  CA    N     C     CB    gi_2
  CG   CD1   CD2   CB    gi_2
 CD1   OE1   OE2   CG    gi_1
 CD2   OE3   OE4   CG    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N     CA    gd_4
  -C    N     CA    C     gd_19
   N    CA    C     +N    gd_20
   N    CA    CB    CG    gd_17
  CA    CB    CG    CD1   gd_17
  CB    CG    CD1   OE2   gd_20
  CB    CG    CD2   OE4   gd_20

; 4-carboxyglutamate (-1)
[ ECN ]
 [ atoms ]
   N     N    -0.28000    0
   H     H     0.28000    0
  CA   CH1     0.00000    1
  CB   CH2     0.00000    1
  CG   CH1     0.00000    1
 CD1     C     0.53000    2 ; from GLUH
 OE1     O    -0.38000    2 ; from GLUH
 OE2    OA    -0.54800    2 ; from GLUH
 HE2     H     0.39800    2 ; from GLUH
 CD2     C     0.27000    3
 OE3    OM    -0.63500    3
 OE4    OM    -0.63500    3
   C     C      0.380    4
   O     O     -0.380    4
 [ bonds ]
   N    H    gb_2
   N   CA    gb_20
  CA    C    gb_26
   C    O    gb_4
   C   +N    gb_9
  CA   CB    gb_26
  CB   CG    gb_26
  CG  CD1    gb_26
 CD1  OE1    gb_4
 CD1  OE2    gb_12
 OE2  HE2    gb_1
  CG  CD2    gb_26
 CD2  OE3    gb_5
 CD2  OE4    gb_5
 [ angles ]
;  ai    aj    ak   gromos type
  -C     N     H     ga_31
   H     N     CA    ga_17
  -C     N     CA    ga_30
   N     CA    C     ga_12
  CA     C     +N    ga_18
  CA     C     O     ga_29
   O     C     +N    ga_32
   N     CA    CB    ga_12
   C     CA    CB    ga_12
  CA    CB    CG    ga_14
  CB    CG    CD1   ga_12
  CB    CG    CD2   ga_12
 CD1    CG    CD2   ga_12
  CG   CD1    OE1   ga_29
  CG   CD1    OE2   ga_18
 OE1   CD1    OE2   ga_32
 CD1   OE2    HE2   ga_11
  CG   CD2    OE3   ga_21
  CG   CD2    OE4   ga_21
 OE3   CD2    OE4   ga_37
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N    -C    CA    H     gi_1
   C    CA    +N    O     gi_1
  CA    N     C     CB    gi_2
  CG   CD1   CD2   CB    gi_2
 CD1   OE1   OE2   CG    gi_1
 CD2   OE3   OE4   CG    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N     CA    gd_4
  -C    N     CA    C     gd_19
   N    CA    C     +N    gd_20
   N    CA    CB    CG    gd_17
  CA    CB    CG    CD1   gd_17
  CB    CG    CD1   OE2   gd_20
  CG   CD1   OE2   HE2   gd_3
  CB    CG   CD2   OE4   gd_20

; sulfotyrosine
[ YSU ]
 [ atoms ]
   N     N    -0.28000    0
   H     H     0.28000    0
  CA   CH1     0.00000    1
  CB   CH2     0.00000    1
  CG     C     0.00000    1
 CD1     C    -0.10000    2
 HD1    HC     0.10000    2
 CD2     C    -0.10000    3
 HD2    HC     0.10000    3
 CE1     C    -0.10000    4
 HE1    HC     0.10000    4
 CE2     C    -0.10000    5
 HE2    HC     0.10000    5
  CZ     C     0.15000    6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OH    OA    -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
  ST     S     1.11500    6 ; to add up to -1 net charge
 OI1    OM    -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2    OM    -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3    OM    -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
   C     C      0.380    7
   O     O     -0.380    7
 [ bonds ]
   N    H    gb_2
   N   CA    gb_20
  CA    C    gb_26
   C    O    gb_4
   C   +N    gb_9
  CA   CB    gb_26
  CB   CG    gb_26
  CG  CD1    gb_15
  CG  CD2    gb_15
 CD1  HD1    gb_3
 CD1  CE1    gb_15
 CD2  HD2    gb_3
 CD2  CE2    gb_15
 CE1  HE1    gb_3
 CE1   CZ    gb_15
 CE2  HE2    gb_3
 CE2   CZ    gb_15
  CZ   OH    gb_12
  OH   ST    gb_24
  ST  OI1    gb_24
  ST  OI2    gb_24
  ST  OI3    gb_24
 [ exclusions ]
;  ai    aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE1
  CG  HE2
  CG   CZ
 CD1  HD2
 CD1  CE2
 CD1   OH
 HD1  CD2
 HD1  HE1
 HD1   CZ
 CD2  CE1
 CD2   OH
 HD2  HE2
 HD2   CZ
 CE1  HE2
 HE1  CE2
 HE1   OH
 HE2   OH
 [ angles ]
;  ai    aj    ak   gromos type
  -C     N     H     ga_31
   H     N     CA    ga_17
  -C     N     CA    ga_30
   N     CA    C     ga_12
  CA     C     +N    ga_18
  CA     C     O     ga_29
   O     C     +N    ga_32
   N     CA    CB    ga_12
   C     CA    CB    ga_12
  CA    CB    CG    ga_14
  CB    CG    CD1   ga_26
  CB    CG    CD2   ga_26
 CD1    CG    CD2   ga_26
  CG   CD1    HD1   ga_24
 HD1   CD1    CE1   ga_24
```

```
  CG   CD1   CE1    ga_26
  CG   CD2   HD2    ga_24
  HD2  CD2   CE2    ga_24
  CG   CD2   CE2    ga_26
  CD1  CE1   HE1    ga_24
  HE1  CE1   CZ     ga_24
  CD1  CE1   CZ     ga_26
  CD2  CE2   HE2    ga_24
  HE2  CE2   CZ     ga_24
  CD2  CE2   CZ     ga_26
  CE1  CZ    CE2    ga_26
  CE1  CZ    OH     ga_26
  CE2  CZ    OH     ga_26
  CZ   OH    ST     ga_11
  OH   ST    OI1    ga_12
  OH   ST    OI2    ga_12
  OH   ST    OI3    ga_12
  OI1  ST    OI2    ga_12
  OI1  ST    OI3    ga_12
  OI2  ST    OI3    ga_12
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA    H    gi_1
   C    CA    +N    O    gi_1
  CA    N     C     CB   gi_2
  CG    CD1   CD2   CB   gi_1
  CD2   CG    CD1   CE1  gi_1
  CD1   CG    CD2   CE2  gi_1
  CG    CD1   CE1   CZ   gi_1
  CG    CD2   CE2   CZ   gi_1
  CD1   CE1   CZ    CE2  gi_1
  CD2   CE2   CZ    CE1  gi_1
  CD1   CG    CE1   HD1  gi_1
  CD2   CG    CE2   HD2  gi_1
  CE1   CZ    CD1   HE1  gi_1
  CE2   CZ    CD2   HE2  gi_1
  CZ    CE1   CE2   OH   gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
  -CA   -C    N     CA   gd_4
  -C    N     CA    C    gd_19
   N    CA    C     +N   gd_20
   N    CA    CB    CG   gd_17
  CA    CB    CG    CD1  gd_20
  CE1   CZ    OH    ST   gd_2
  CZ    OH    ST    OI1  gd_20

; dehydroalanine
[ SDH ]
 [ atoms ]
   N    N    -0.28000    0
   H    H     0.28000    0
  CA   CH0     0.00000    1 ; from aliphatic carbon atoms
  CB   CH2     0.00000    1 ; from aliphatic carbon atoms
   C    C       0.380    2
   O    O      -0.380    2
 [ bonds ]
   N    H    gb_2
   N    CA   gb_20
  CA    C    gb_26
   C    O    gb_4
   C    +N   gb_9
  CA    CB   gb_9   ; a shorter bond type to mimic double bond properties
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_26
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_26
   C    CA    CB   ga_26
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA    H    gi_1
   C    CA    +N    O    gi_1
  CA    N     C     CB   gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
  -CA   -C    N     CA   gd_4
  -C    N     CA    C    gd_19
   N    CA    C     +N   gd_20

; 2,3-didehydrobutyrine
[ TDH ]
 [ atoms ]
   N    N    -0.28000    0
   H    H     0.28000    0
  CA   CH0     0.00000    1 ; from aliphatic carbon atoms
  CB   CH1     0.00000    2 ; from aliphatic carbon atoms
  CG   CH3     0.00000    2
   C    C       0.380    3
   O    O      -0.380    3
 [ bonds ]
   N    H    gb_2
   N    CA   gb_20
  CA    C    gb_26
   C    O    gb_4
   C    +N   gb_9
  CA    CB   gb_9   ; a shorter bond type to mimic double bond properties
  CB    CG   gb_26
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_26
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_26
   C    CA    CB   ga_26
  CA    CB    CG   ga_26
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA    H    gi_1
   C    CA    +N    O    gi_1
  CA    N     C     CB   gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
  -CA   -C    N     CA   gd_4
  -C    N     CA    C    gd_19
   N    CA    C     +N   gd_20
   N    CA    CB    CG   gd_4

; 6-bromotryptophan
[ WBR ]
 [ atoms ]
   N    N    -0.28000    0
   H    H     0.28000    0
  CA   CH1     0.00000    1
  CB   CH2     0.00000    1
  CG    C     -0.14000    2
  CD1   C     -0.10000    2
  HD1   HC     0.10000    2
  CD2   C      0.00000    2
  NE1   NR    -0.05000    2
  HE1   H      0.19000    2
  CE2   C      0.00000    2
  CE3   C     -0.10000    3
  HE3   HC     0.10000    3
  CZ2   C     -0.10000    4
  HZ2   HC     0.10000    4
  CZ3   C     -0.10000    5
  HZ3   HC     0.10000    5
  CH2   C      0.05500    6 ; from 8-bromo-guanosine reported by Hritz and
Oostenbrink. DOI:10.1021/jp902968m
  BRT   BR    -0.05500    6 ; from 8-bromo-guanosine reported by Hritz and
Oostenbrink. DOI:10.1021/jp902968m
   C    C       0.380    7
   O    O      -0.380    7
 [ bonds ]
   N    H    gb_2
   N    CA   gb_20
  CA    C    gb_26
   C    O    gb_4
   C    +N   gb_9
  CA    CB   gb_26
  CB    CG   gb_26
  CG    CD1  gb_9
  CG    CD2  gb_15
  CD1   HD1  gb_3
  CD1   NE1  gb_9
  CD2   CE2  gb_15
  CD2   CE3  gb_15
  NE1   HE1  gb_2
  NE1   CE2  gb_9
  CE2   CZ2  gb_15
  CE3   HE3  gb_3
  CE3   CZ3  gb_15
  CZ2   HZ2  gb_3
  CZ2   CH2  gb_15
  CZ3   HZ3  gb_3
  CZ3   CH2  gb_15
  CH2   BRT  gb_39
 [ exclusions ]
; ai    aj
  CB    HD1
  CB    NE1
  CB    CE2
  CB    CE3
  CG    HE1
  CG    HE3
  CG    CZ2
```

```
  CG   CZ3
  CD1  CE3
  CD1  CZ2
  HD1  CD2
  HD1  HE1
  HD1  CE2
  CD2  HE1
  CD2  HZ2
  CD2  HZ3
  CD2  CH2
  NE1  CE3
  NE1  HZ2
  NE1  CH2
  HE1  CZ2
  CE2  HE3
  CE2  CZ3
  CE2  BRT
  CE3  CZ2
  CE3  BRT
  HE3  HZ3
  HE3  CH2
  CZ2  HZ3
  HZ2  CZ3
  HZ2  BRT
  HZ3  BRT
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_31
   H   N    CA  ga_17
  -C   N    CA  ga_30
   N   CA   C   ga_12
  CA   C    +N  ga_18
  CA   C    O   ga_29
   O   C    +N  ga_32
   N   CA   CB  ga_12
   C   CA   CB  ga_12
  CA   CB   CG  ga_14
  CB   CG   CD1 ga_36
  CB   CG   CD2 ga_36
  CD1  CG   CD2 ga_6
  CG   CD1  HD1 ga_35
  HD1  CD1  NE1 ga_35
  CG   CD1  NE1 ga_6
  CG   CD2  CE2 ga_6
  CD1  NE1  CE2 ga_6
  CD1  NE1  HE1 ga_35
  HE1  NE1  CE2 ga_35
  NE1  CE2  CD2 ga_6
  CG   CD2  CE3 ga_38
  NE1  CE2  CZ2 ga_38
  CD2  CE2  CZ2 ga_26
  CE2  CD2  CE3 ga_26
  CD2  CE3  HE3 ga_24
  HE3  CE3  CZ3 ga_24
  CD2  CE3  CZ3 ga_26
  CE2  CZ2  HZ2 ga_24
  HZ2  CZ2  CH2 ga_24
  CE2  CZ2  CH2 ga_26
  CE3  CZ3  HZ3 ga_24
  HZ3  CZ3  CH2 ga_24
  CE3  CZ3  CH2 ga_26
  CZ2  CH2  BRT ga_26
  BRT  CH2  CZ3 ga_26
  CZ2  CH2  CZ3 ga_26
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
   C   CA   +N   O   gi_1
  CA   N    C    CB  gi_2
  CG   CD1  CD2  CB  gi_1
  CD2  CG   CD1  NE1 gi_1
  CD1  CG   CD2  CE2 gi_1
  CG   CD1  NE1  CE2 gi_1
  CG   CD2  CE2  NE1 gi_1
  CD1  NE1  CE2  CD2 gi_1
  CD1  CG   NE1  HD1 gi_1
  NE1  CD1  CE2  HE1 gi_1
  CD2  CE2  CE3  CG  gi_1
  CE2  CD2  CZ2  NE1 gi_1
  CE3  CD2  CE2  CZ2 gi_1
  CD2  CE2  CZ2  CH2 gi_1
  CE2  CD2  CE3  CZ3 gi_1
  CE2  CZ2  CH2  CZ3 gi_1
  CD2  CE3  CZ3  CH2 gi_1
  CE3  CZ3  CH2  CZ2 gi_1
  CE3  CD2  CZ3  HE3 gi_1
  CZ2  CE2  CH2  HZ2 gi_1
  CZ3  CE3  CH2  HZ3 gi_1
  CH2  CZ2  CZ3  BRT gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C   N    CA  gd_4
  -C   N    CA   C   gd_19
   N   CA   C    +N  gd_20
   N   CA   CB   CG  gd_17
  CA   CB   CG   CD2 gd_20

; S-nitrosocysteine
[ CSN ]
 [ atoms ]
   N   N   -0.28000   0
   H   H    0.28000   0
  CA  CH1   0.00000   1
  CB  CH2   0.00000   1
  SG   S    0.10000   2 ; to add up to 0 net charge
  ND  NR    0.28000   2 ; to add up to 0 net charge
  OE   O   -0.38000   2 ; from the carbonyl group (of e.g., the peptide bond)
   C   C    0.380     3
   O   O   -0.380     3
 [ bonds ]
   N   H   gb_2
   N   CA  gb_20
  CA   C   gb_26
   C   O   gb_4
   C   +N  gb_9
  CA   CB  gb_26
  CB   SG  gb_30
  SG   ND  gb_29 ; from MET
  ND   OE  gb_4  ; from the carbonyl group (of e.g., the peptide bond)
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_31
   H   N    CA  ga_17
  -C   N    CA  ga_30
   N   CA   C   ga_12
  CA   C    +N  ga_18
  CA   C    O   ga_29
   O   C    +N  ga_32
   N   CA   CB  ga_12
   C   CA   CB  ga_12
  CA   CB   SG  ga_15
  CB   SG   ND  ga_5
  SG   ND   OE  ga_26
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
   C   CA   +N   O   gi_1
  CA   N    C    CB  gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C   N    CA  gd_4
  -C   N    CA   C   gd_19
   N   CA   C    +N  gd_20
   N   CA   CB   SG  gd_17
  CA   CB   SG   ND  gd_13
  CB   SG   ND   OE  gd_10

; citrulline
[ RCI ]
 [ atoms ]
   N   N   -0.28000   0
   H   H    0.28000   0
  CA  CH1   0.00000   1
  CB  CH2   0.00000   1
  CG  CH2   0.00000   2
  CD  CH2   0.00000   2
  NE   N   -0.28000   3
  HE   H    0.28000   3
  CZ   C    0.38000   4 ; from ASN/GLN
 OH1   O   -0.38000   4 ; from ASN/GLN
 NH2  NT   -0.83000   5 ; from ASN/GLN
HH21   H    0.41500   5 ; from ASN/GLN
HH22   H    0.41500   5 ; from ASN/GLN
   C   C    0.380     6
   O   O   -0.380     6
 [ bonds ]
   N   H   gb_2
   N   CA  gb_20
  CA   C   gb_26
   C   O   gb_4
   C   +N  gb_9
  CA   CB  gb_26
  CB   CG  gb_26
  CG   CD  gb_26
  CD   NE  gb_20
  NE   HE  gb_2
  NE   CZ  gb_9
  CZ  OH1  gb_4
  CZ  NH2  gb_8
 NH2 HH21  gb_2
 NH2 HH22  gb_2
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_31
   H   N    CA  ga_17
  -C   N    CA  ga_30
```

```
   N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB    ga_12
    C   CA   CB    ga_12
   CA   CB   CG    ga_14
   CB   CG   CD    ga_14
   CG   CD   NE    ga_12
   CD   NE   HE    ga_17
   HE   NE   CZ    ga_31
   CD   NE   CZ    ga_30
   NE   CZ  OH1    ga_21
   NE   CZ  NH2    ga_29
  OH1   CZ  NH2    ga_21
   CZ  NH2 HH21    ga_22
   CZ  NH2 HH22    ga_22
 HH21  NH2 HH22    ga_23
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   NE   CD   CZ   HE    gi_1
   CZ  OH1  NH2   NE    gi_1
  NH2 HH21 HH22   CZ    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
   CA   CB   CG   CD    gd_17
   CB   CG   CD   NE    gd_17
   CG   CD   NE   CZ    gd_19
   CD   NE   CZ  NH2    gd_4
   NE   CZ  NH2 HH21    gd_4

; allysine (aminoadipic semialdehyde)
[ KAL ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE    C    0.28000    3 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
   OZ    O   -0.38000    3 ; from the carbonyl group (of e.g., GLN)
   HE   HC    0.10000    3 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
    C    C      0.380    4
    O    O     -0.380    4
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG   CD   gb_26
   CD   CE   gb_26
   CE   HE   gb_3
   CE   OZ   gb_4
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H   ga_31
    H    N   CA   ga_17
   -C    N   CA   ga_30
    N   CA    C   ga_12
   CA    C   +N   ga_18
   CA    C    O   ga_29
    O    C   +N   ga_32
    N   CA   CB   ga_12
    C   CA   CB   ga_12
   CA   CB   CG   ga_14
   CB   CG   CD   ga_14
   CG   CD   CE   ga_14
   CD   CE   HE   ga_24
   HE   CE   OZ   ga_24
   CD   CE   OZ   ga_26
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CE   OZ   CD   HE    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
   CA   CB   CG   CD    gd_17
   CB   CG   CD   CE    gd_17
   CG   CD   CE   OZ    gd_20

; N-acetylglucosamine (N4-linked to ASN)
[ NNG ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.38000    2
  OD1    O   -0.38000    2
  ND2    N   -0.28000    3
  HD2    H    0.28000    3
   C1  CH1    0.18000    4 ; 0.36 charge divided between C1 and C5 to add up to 0 net
charge, also used in other building blocks (e.g., TMP)
   O5   OA   -0.36000    4 ; from carbohydrates (e.g., GALB)
   C5  CH1    0.18000    4 ; 0.36 charge divided between C1 and C5 to add up to 0 net
charge, also used in other building blocks (e.g., TMP)
   C6  CH2    0.15000    5 ; from carbohydrates (e.g., GALB)
   O6   OA   -0.54800    5 ; from carbohydrates (e.g., GALB)
  HO6    H    0.39800    5 ; from carbohydrates (e.g., GALB)
   C2  CH1    0.00000    6 ; from aliphatic carbon atoms
   C4  CH1    0.15000    7 ; from carbohydrates (e.g., GALB)
   O4   OA   -0.54800    7 ; from carbohydrates (e.g., GALB)
  HO4    H    0.39800    7 ; from carbohydrates (e.g., GALB)
   C3  CH1    0.15000    8 ; from carbohydrates (e.g., GALB)
   O3   OA   -0.54800    8 ; from carbohydrates (e.g., GALB)
  HO3    H    0.39800    8 ; from carbohydrates (e.g., GALB)
   N2    N   -0.28000    9 ; from the peptide bond
  HN2    H    0.28000    9 ; from the peptide bond
   C7    C    0.28000   10 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
   O7    O   -0.38000   10 ; from the carbonyl group (of e.g., GLN)
   C8  CH3    0.10000   10 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
    C    C      0.380   11
    O    O     -0.380   11
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG  OD1   gb_4
   CG  ND2   gb_9
  ND2  HD2   gb_2
  ND2   C1   gb_20
   C1   O5   gb_19
   O5   C5   gb_19
   C5   C6   gb_25
   C6   O6   gb_19
   O6  HO6   gb_1
   C5   C4   gb_25
   C4   O4   gb_19
   O4  HO4   gb_1
   C4   C3   gb_25
   C3   O3   gb_19
   O3  HO3   gb_1
   C3   C2   gb_25
   C2   C1   gb_25
   C2   N2   gb_20
   N2  HN2   gb_2
   N2   C7   gb_9
   C7   O7   gb_4
   C7   C8   gb_26
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H   ga_31
    H    N   CA   ga_17
   -C    N   CA   ga_30
    N   CA    C   ga_12
   CA    C   +N   ga_18
   CA    C    O   ga_29
    O    C   +N   ga_32
    N   CA   CB   ga_12
    C   CA   CB   ga_12
   CA   CB   CG   ga_14
   CB   CG  OD1   ga_29
   CB   CG  ND2   ga_18
  OD1   CG  ND2   ga_32
   CG  ND2  HD2   ga_31
   CG  ND2   C1   ga_30
   C1  ND2  HD2   ga_17
  ND2   C1   O5   ga_8
  ND2   C1   C2   ga_8
   O5   C1   C2   ga_8
   C5   O5   C1   ga_9
```

```
  O5   C5   C6    ga_8
  O5   C5   C4    ga_8
  C6   C5   C4    ga_7
  C5   C6   O6    ga_8
  C6   O6   HO6   ga_11
  C5   C4   O4    ga_8
  C5   C4   C3    ga_7
  O4   C4   C3    ga_8
  C4   O4   HO4   ga_11
  C4   C3   O3    ga_8
  C4   C3   C2    ga_7
  C2   C3   O3    ga_8
  C3   O3   HO3   ga_11
  C3   C2   C1    ga_7
  C3   C2   N2    ga_8
  C1   C2   N2    ga_8
  C2   N2   HN2   ga_17
  C2   N2   C7    ga_30
 HN2   N2   C7    ga_31
  N2   C7   O7    ga_32
  N2   C7   C8    ga_18
  O7   C7   C8    ga_29
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  OD1  ND2   CB   gi_1
 ND2   CG   C1  HD2   gi_1
  C1   O5  ND2   C2   gi_2
  C5   O5   C6   C4   gi_2
  C2   N2   C3   C1   gi_2
  C3   O3   C2   C4   gi_2
  C4   C3   O4   C5   gi_2
  N2   C7   C2  HN2   gi_1
  C7   C8   N2   O7   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  ND2   gd_20
  CB   CG  ND2   C1   gd_4
  CG  ND2   C1   O5   gd_19
 ND2   C1   C2   C3   gd_17
 ND2   C1   C2   C3   gd_7
  O5   C1   C2   C3   gd_7
  O5   C1   C2   N2   gd_8
 ND2   C1   C2   N2   gd_8
  C1   C2   N2   C7   gd_19
  C2   N2   C7   C8   gd_4
  C1   C2   C3   C4   gd_17
  C1   C2   C3   O3   gd_7
  N2   C2   C3   C4   gd_7
  N2   C2   C3   O3   gd_8
  C2   C3   O3  HO3   gd_12
  C2   C3   C4   C5   gd_17
  C2   C3   C4   O4   gd_7
  O3   C3   C4   C5   gd_7
  O3   C3   C4   O4   gd_8
  C3   C4   O4  HO4   gd_12
  C2   C1   O5   C5   gd_14
  C1   O5   C5   C4   gd_14
  C4   C5   C6   O6   gd_17
  C4   C5   C6   O6   gd_7
  O5   C5   C6   O6   gd_8
  C5   C6   O6  HO6   gd_12
  C6   C5   C4   C3   gd_17
  O5   C5   C4   C3   gd_7
  C6   C5   C4   O4   gd_7
  O5   C5   C4   O4   gd_8

; 2,3-dihydroxyphenylalanine
[ F23 ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C    0.15000    2 ; from the hydroxyl group of TYR
 OE3   OA   -0.54800    2 ; from the hydroxyl group of TYR
 HE3    H    0.39800    2 ; from the hydroxyl group of TYR
 CD2    C   -0.10000    3
 HD2   HC    0.10000    3
 CE1    C    0.15000    4 ; from the hydroxyl group of TYR
 OZ1   OA   -0.54800    4 ; from the hydroxyl group of TYR
 HZ1    H    0.39800    4 ; from the hydroxyl group of TYR
 CE2    C   -0.10000    5
 HE2   HC    0.10000    5
 CZ2    C   -0.10000    6
 HZ2   HC    0.10000    6
   C    C    0.380    7
   O    O   -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
 CD1  CE1   gb_15
 CD1  OE3   gb_12
 OE3  HE3   gb_1
 CD2  HD2   gb_3
 CD2  CE2   gb_15
 CE1  CZ2   gb_15
 CE1  OZ1   gb_12
 OZ1  HZ1   gb_1
 CE2  HE2   gb_3
 CE2  CZ2   gb_15
 CZ2  HZ2   gb_3
 [ exclusions ]
; ai   aj
  CB  HD2
  CB  CE1
  CB  CE2
  CB  OE3
  CG  OZ1
  CG  HE2
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  HZ2
 CD2  CE1
 CD2  OE3
 CD2  HZ2
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 CE2  OZ1
 HE2  HZ2
 OE3  CZ2
 OE3  OZ1
 HZ2  OZ1
 OE3  HZ1
 HE3  OZ1
 HE3  HZ1
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG  CD1   ga_26
  CB   CG  CD2   ga_26
 CD1   CG  CD2   ga_26
  CG  CD1  CE1   ga_26
  CG  CD1  OE3   ga_26
 CE1  CD1  OE3   ga_26
  CG  CD2  HD2   ga_24
 HD2  CD2  CE2   ga_24
  CG  CD2  CE2   ga_26
 CD1  CE1  CZ2   ga_26
 CD1  CE1  OZ1   ga_26
 CZ2  CE1  OZ1   ga_26
 CD2  CE2  HE2   ga_24
 HE2  CE2  CZ2   ga_24
 CD2  CE2  CZ2   ga_26
 CD1  OE3  HE3   ga_11
 CE1  CZ2  HZ2   ga_24
 CE2  CZ2  HZ2   ga_24
 CE1  CZ2  CE2   ga_26
 CE1  OZ1  HZ1   ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD1   CG  CD2  CE2   gi_1
  CG  CD1  CE1  CZ2   gi_1
  CG  CD2  CE2  CZ2   gi_1
 CD1  CE1  CZ2  CE2   gi_1
 CD2  CE2  CZ2  CE1   gi_1
```

```
 CD1   CG  CE1  OE3   gi_1
 CD2   CG  CE2  HD2   gi_1
 CE1   CZ2  CD1  OZ1   gi_1
 CE2   CZ2  CD2  HE2   gi_1
 CZ2   CE1  CE2  HZ2   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD1   gd_20
  CG  CD1  OE3  HE3   gd_2
  CD1  CE1  OZ1  HZ1   gd_2

; 2-hydroxyphenylalanine
[ F2H ]
[ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
  CD1    C    0.15000    2 ; from the hydroxyl group of TYR
  OE3   OA   -0.54800    2 ; from the hydroxyl group of TYR
  HE3    H    0.39800    2 ; from the hydroxyl group of TYR
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C   -0.10000    4
  HE1   HC    0.10000    4
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
   CZ    C   -0.10000    6
   HZ   HC    0.10000    6
    C    C     0.380    7
    O    O    -0.380    7
[ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
  CD1  OE3   gb_12
  CD1  CE1   gb_15
  CD2  HD2   gb_3
  CD2  CE2   gb_15
  CE1  HE1   gb_3
  CE1   CZ   gb_15
  CE2  HE2   gb_3
  CE2   CZ   gb_15
  OE3  HE3   gb_1
   CZ   HZ   gb_3
[ exclusions ]
;  ai   aj
   CB  HD2
   CB  CE1
   CB  CE2
   CB  OE3
   CG  HE1
   CG  HE2
   CG   CZ
  CD1  HD2
  CD1  CE2
  CD1   HZ
  CD2  OE3
  CD2  CE1
  CD2   HZ
  HD2  HE2
  HD2   CZ
  CE1  HE2
  HE1  CE2
  HE1  OE3
  HE1   HZ
  HE2   HZ
  OE3   CZ
[ angles ]
;  ai   aj   ak  gromos type
   -C    N    H   ga_31
    H    N   CA   ga_17
   -C    N   CA   ga_30
    N   CA    C   ga_12
   CA    C   +N   ga_18
   CA    C    O   ga_29
    O    C   +N   ga_32
    N   CA   CB   ga_12
    C   CA   CB   ga_12
   CA   CB   CG   ga_14
   CB   CG  CD1   ga_26
   CB   CG  CD2   ga_26
  CD1   CG  CD2   ga_26
   CG  CD1  CE1   ga_26
   CG  CD1  OE3   ga_26
  CE1  CD1  OE3   ga_26
   CG  CD2  HD2   ga_24
  HD2  CD2  CE2   ga_24
   CG  CD2  CE2   ga_26
  CD1  CE1  HE1   ga_24
  HE1  CE1   CZ   ga_24
  CD1  CE1   CZ   ga_26
  CD2  CE2  HE2   ga_24
  HE2  CE2   CZ   ga_24
  CD2  CE2   CZ   ga_26
  CD1  OE3  HE3   ga_11
  CE1   CZ   HZ   ga_24
  CE2   CZ   HZ   ga_24
  CE1   CZ  CE2   ga_26
[ impropers ]
;  ai   aj   ak   al  gromos type
    N   -C   CA    H   gi_1
    C   CA   +N    O   gi_1
   CA    N    C   CB   gi_2
   CG  CD1  CD2   CB   gi_1
  CD2   CG  CD1  CE1   gi_1
  CD1   CG  CD2  CE2   gi_1
   CG  CD1  CE1   CZ   gi_1
   CG  CD2  CE2   CZ   gi_1
  CD1  CE1   CZ  CE2   gi_1
  CD2  CE2   CZ  CE1   gi_1
  CD1   CG  CE1  OE3   gi_1
  CD2   CG  CE2  HD2   gi_1
  CE1   CZ  CD1  HE1   gi_1
  CE2   CZ  CD2  HE2   gi_1
   CZ  CE1  CE2   HZ   gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_4
   -C    N   CA    C   gd_19
    N   CA    C   +N   gd_20
    N   CA   CB   CG   gd_17
   CA   CB   CG  CD1   gd_20
   CG  CD1  OE3  HE3   gd_2

; 3-hydroxyphenylalanine
[ F3H ]
[ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C    0.15000    4 ; from the hydroxyl group of TYR
  OZ1   OA   -0.54800    4 ; from the hydroxyl group of TYR
  HZ1    H    0.39800    4 ; from the hydroxyl group of TYR
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
  CZ2    C   -0.10000    6
  HZ2   HC    0.10000    6
    C    C     0.380    7
    O    O    -0.380    7
[ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
  CD1  HD1   gb_3
  CD1  CE1   gb_15
  CD2  HD2   gb_3
  CD2  CE2   gb_15
  CE1  CZ2   gb_15
  CE1  OZ1   gb_12
  CE2  HE2   gb_3
  CE2  CZ2   gb_15
  CZ2  HZ2   gb_3
  OZ1  HZ1   gb_1
[ exclusions ]
;  ai   aj
   CB  HD1
   CB  HD2
   CB  CE1
   CB  CE2
   CG  HE2
   CG  CZ2
   CG  OZ1
```

```
 CD1  HD2
 CD1  CE2
 CD1  HZ2
 HD1  CD2
 HD1  CZ2
 HD1  OZ1
 CD2  CE1
 CD2  HZ2
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 CE2  OZ1
 HE2  HZ2
 HZ2  OZ1
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_26
  CB   CG  CD2    ga_26
 CD1   CG  CD2    ga_26
  CG  CD1  HD1    ga_24
 HD1  CD1  CE1    ga_24
  CG  CD1  CE1    ga_26
  CG  CD2  HD2    ga_24
 HD2  CD2  CE2    ga_24
  CG  CD2  CE2    ga_26
 CD1  CE1  CZ2    ga_26
 CD1  CE1  OZ1    ga_26
 CZ2  CE1  OZ1    ga_26
 CD2  CE2  HE2    ga_24
 HE2  CE2  CZ2    ga_24
 CD2  CE2  CZ2    ga_26
 CE1  CZ2  HZ2    ga_24
 CE2  CZ2  HZ2    ga_24
 CE1  CZ2  CE2    ga_26
 CE1  OZ1  HZ1    ga_11
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
 CD2   CG  CD1  CE1    gi_1
 CD1   CG  CD2  CE2    gi_1
  CG  CD1  CE1  CZ2    gi_1
  CG  CD2  CE2  CZ2    gi_1
 CD1  CE1  CZ2  CE2    gi_1
 CD2  CE2  CZ2  CE1    gi_1
 CD1   CG  CE1  HD1    gi_1
 CD2   CG  CE2  HD2    gi_1
 CE1  CZ2  CD1  OZ1    gi_1
 CE2  CZ2  CD2  HE2    gi_1
 CZ2  CE1  CE2  HZ2    gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG  CD1    gd_20
 CD1  CE1  OZ1  HZ1    gd_2

; 6-hydroxytryptophan
[ W6H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.14000    2
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C    0.00000    2
 NE1   NR   -0.05000    2
 HE1    H    0.19000    2
 CE2    C    0.00000    2
 CE3    C   -0.10000    3
 HE3   HC    0.10000    3
 CZ2    C   -0.10000    4
 HZ2   HC    0.10000    4
 CZ3    C   -0.10000    5
 HZ3   HC    0.10000    5
 CH2    C    0.15000    6 ; from the hydroxyl group of TYR
  OI   OA   -0.54800    6 ; from the hydroxyl group of TYR
  HI    H    0.39800    6 ; from the hydroxyl group of TYR
   C    C    0.380    7
   O    O   -0.380    7
[ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_9
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  NE1   gb_9
 CD2  CE2   gb_15
 CD2  CE3   gb_15
 NE1  HE1   gb_2
 NE1  CE2   gb_9
 CE2  CZ2   gb_15
 CE3  HE3   gb_3
 CE3  CZ3   gb_15
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_15
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_15
 CH2   OI   gb_12
  OI   HI   gb_1
[ exclusions ]
;  ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2   OI
 CE3  CZ2
 CE3   OI
 HE3  HZ3
 HE3  CH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2   OI
 HZ3   OI
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_36
  CB   CG  CD2    ga_36
 CD1   CG  CD2    ga_6
  CG  CD1  HD1    ga_35
 HD1  CD1  NE1    ga_35
  CG  CD1  NE1    ga_6
  CG  CD2  CE2    ga_6
 CD1  NE1  CE2    ga_6
 CD1  NE1  HE1    ga_35
 HE1  NE1  CE2    ga_35
 NE1  CE2  CD2    ga_6
  CG  CD2  CE3    ga_38
 NE1  CE2  CZ2    ga_38
 CD2  CE2  CZ2    ga_26
 CE2  CD2  CE3    ga_26
 CD2  CE3  HE3    ga_24
 HE3  CE3  CZ3    ga_24
 CD2  CE3  CZ3    ga_26
 CE2  CZ2  HZ2    ga_24
```

```
 HZ2  CZ2  CH2   ga_24
 CE2  CZ2  CH2   ga_26
 CE3  CZ3  HZ3   ga_24
 HZ3  CZ3  CH2   ga_24
 CE3  CZ3  CH2   ga_26
 CZ2  CH2  CZ3   ga_26
 CZ2  CH2  OI    ga_26
 CZ3  CH2  OI    ga_26
 CH2  OI   HI    ga_11
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
 CD2   CG  CD1  NE1   gi_1
 CD1   CG  CD2  CE2   gi_1
  CG  CD1  NE1  CE2   gi_1
  CG  CD2  CE2  NE1   gi_1
 CD1  NE1  CE2  CD2   gi_1
 CD1   CG  NE1  HD1   gi_1
 NE1  CD1  CE2  HE1   gi_1
 CD2  CE2  CE3   CG   gi_1
 CE2  CD2  CZ2  NE1   gi_1
 CE3  CD2  CE2  CZ2   gi_1
 CD2  CE2  CZ2  CH2   gi_1
 CE2  CD2  CE3  CZ3   gi_1
 CE2  CZ2  CH2  CZ3   gi_1
 CD2  CE3  CZ3  CH2   gi_1
 CE3  CZ3  CH2  CZ2   gi_1
 CE3  CD2  CZ3  HE3   gi_1
 CZ2  CE2  CH2  HZ2   gi_1
 CZ3  CE3  CH2  HZ3   gi_1
 CH2  CZ2  CZ3   OI   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD2   gd_20
 CZ2  CH2   OI   HI   gd_2

; 5-hydroxytryptophan
[ W5H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.14000    2
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C    0.00000    2
 NE1   NR   -0.05000    2
 HE1    H    0.19000    2
 CE2    C    0.00000    2
 CE3    C   -0.10000    3
 HE3   HC    0.10000    3
 CZ2    C   -0.10000    4
 HZ2   HC    0.10000    4
 CZ3    C    0.15000    5 ; from the hydroxyl group of TYR
 OH3   OA   -0.54800    5 ; from the hydroxyl group of TYR
 HH3    H    0.39800    5 ; from the hydroxyl group of TYR
 CH2    C   -0.10000    6
 HH2   HC    0.10000    6
   C    C      0.380    7
   O    O     -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_9
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  NE1   gb_9
 CD2  CE2   gb_15
 CD2  CE3   gb_15
 NE1  HE1   gb_2
 NE1  CE2   gb_9
 CE2  CZ2   gb_15
 CE3  HE3   gb_3
 CE3  CZ3   gb_15
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_15
 CZ3  CH2   gb_15
 CZ3  OH3   gb_12
 CH2  HH2   gb_3
 OH3  HH3   gb_1
[ exclusions ]
; ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  OH3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2  HH2
 CE3  CZ2
 CE3  HH2
 HE3  OH3
 HE3  CH2
 CZ2  OH3
 HZ2  CZ3
 HZ2  HH2
 OH3  HH2
[ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG  CD1   ga_36
  CB   CG  CD2   ga_36
 CD1   CG  CD2   ga_6
  CG  CD1  HD1   ga_35
 HD1  CD1  NE1   ga_35
  CG  CD1  NE1   ga_6
  CG  CD2  CE2   ga_6
 CD1  NE1  CE2   ga_6
 CD1  NE1  HE1   ga_35
 HE1  NE1  CE2   ga_35
 NE1  CE2  CD2   ga_6
  CG  CD2  CE3   ga_38
 NE1  CE2  CZ2   ga_38
 CD2  CE2  CZ2   ga_26
 CE2  CD2  CE3   ga_26
 CD2  CE3  HE3   ga_24
 HE3  CE3  CZ3   ga_24
 CD2  CE3  CZ3   ga_26
 CE2  CZ2  HZ2   ga_24
 HZ2  CZ2  CH2   ga_24
 CE2  CZ2  CH2   ga_26
 CE3  CZ3  OH3   ga_26
 CH2  CZ3  OH3   ga_26
 CZ3  OH3  HH3   ga_11
 CE3  CZ3  CH2   ga_26
 CZ2  CH2  HH2   ga_24
 HH2  CH2  CZ3   ga_24
 CZ2  CH2  CZ3   ga_26
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
 CD2   CG  CD1  NE1   gi_1
 CD1   CG  CD2  CE2   gi_1
  CG  CD1  NE1  CE2   gi_1
  CG  CD2  CE2  NE1   gi_1
 CD1  NE1  CE2  CD2   gi_1
 CD1   CG  NE1  HD1   gi_1
 NE1  CD1  CE2  HE1   gi_1
 CD2  CE2  CE3   CG   gi_1
 CE2  CD2  CZ2  NE1   gi_1
 CE3  CD2  CE2  CZ2   gi_1
 CD2  CE2  CZ2  CH2   gi_1
 CE2  CD2  CE3  CZ3   gi_1
 CE2  CZ2  CH2  CZ3   gi_1
```

```
 CD2  CE3  CZ3  CH2   gi_1
 CE3  CZ3  CH2  CZ2   gi_1
 CE3  CD2  CZ3  HE3   gi_1
 CZ2  CE2  CH2  HZ2   gi_1
 CZ3  CE3  CH2  OH3   gi_1
 CH2  CZ2  CZ3  HH2   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD2   gd_20
 CE3  CZ3  OH3  HH3   gd_2

; 4-hydroxytryptophan
[ W4H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.14000    2
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C    0.00000    2
 NE1   NR   -0.05000    2
 HE1    H    0.19000    2
 CE2    C    0.00000    2
 CE3    C    0.15000    3 ; from the hydroxyl group of TYR
 OZ4   OA   -0.54800    3 ; from the hydroxyl group of TYR
 HZ4    H    0.39800    3 ; from the hydroxyl group of TYR
 CZ2    C   -0.10000    4
 HZ2   HC    0.10000    4
 CZ3    C   -0.10000    5
 HZ3   HC    0.10000    5
 CH2    C   -0.10000    6
 HH2   HC    0.10000    6
   C    C      0.380    7
   O    O     -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_9
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  NE1   gb_9
 CD2  CE2   gb_15
 CD2  CE3   gb_15
 NE1  HE1   gb_2
 NE1  CE2   gb_9
 CE2  CZ2   gb_15
 CE3  OZ4   gb_12
 CE3  CZ3   gb_15
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_15
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_15
 OZ4  HZ4   gb_1
 CH2  HH2   gb_3
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  CZ2
  CG  CZ3
  CG  OZ4
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  CZ3
 CE2  HH2
 CE2  OZ4
 CE3  CZ2
 CE3  HH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2  HH2
 HZ3  OZ4
 HZ3  HH2
 OZ4  CH2
[ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_31
   H    N   CA   ga_17
  -C    N   CA   ga_30
   N   CA    C   ga_12
  CA    C   +N   ga_18
  CA    C    O   ga_29
   O    C   +N   ga_32
   N   CA   CB   ga_12
   C   CA   CB   ga_12
  CA   CB   CG   ga_14
  CB   CG  CD1   ga_36
  CB   CG  CD2   ga_36
 CD1   CG  CD2   ga_6
  CG  CD1  HD1   ga_35
 HD1  CD1  NE1   ga_35
  CG  CD1  NE1   ga_6
  CG  CD2  CE2   ga_6
 CD1  NE1  CE2   ga_6
 CD1  NE1  HE1   ga_35
 HE1  NE1  CE2   ga_35
 NE1  CE2  CD2   ga_6
  CG  CD2  CE3   ga_38
 NE1  CE2  CZ2   ga_38
 CD2  CE2  CZ2   ga_26
 CE2  CD2  CE3   ga_26
 CD2  CE3  CZ3   ga_26
 CD2  CE3  OZ4   ga_26
 CZ3  CE3  OZ4   ga_26
 CE2  CZ2  HZ2   ga_24
 HZ2  CZ2  CH2   ga_24
 CE2  CZ2  CH2   ga_26
 CE3  CZ3  HZ3   ga_24
 HZ3  CZ3  CH2   ga_24
 CE3  CZ3  CH2   ga_26
 CE3  OZ4  HZ4   ga_11
 CZ2  CH2  HH2   ga_24
 HH2  CH2  CZ3   ga_24
 CZ2  CH2  CZ3   ga_26
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
 CD2   CG  CD1  NE1   gi_1
 CD1   CG  CD2  CE2   gi_1
  CG  CD1  NE1  CE2   gi_1
  CG  CD2  CE2  NE1   gi_1
 CD1  NE1  CE2  CD2   gi_1
 CD1   CG  NE1  HD1   gi_1
 NE1  CD1  CE2  HE1   gi_1
 CD2  CE2  CE3   CG   gi_1
 CE2  CD2  CZ2  NE1   gi_1
 CE3  CD2  CE2  CZ2   gi_1
 CD2  CE2  CZ2  CH2   gi_1
 CE2  CD2  CE3  CZ3   gi_1
 CE2  CZ2  CH2  CZ3   gi_1
 CD2  CE3  CZ3  CH2   gi_1
 CE3  CZ3  CH2  CZ2   gi_1
 CE3  CD2  CZ3  OZ4   gi_1
 CZ2  CE2  CH2  HZ2   gi_1
 CZ3  CE3  CH2  HZ3   gi_1
 CH2  CZ2  CZ3  HH2   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD2   gd_20
 CD2  CE3  OZ4  HZ4   gd_2

; 2-hydroxytryptophan
[ W2H ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.14000    2
 CD1    C    0.15000    2 ; from the hydroxyl group of TYR
 OE4   OA   -0.54800    2 ; from the hydroxyl group of TYR
 HE4    H    0.39800    2 ; from the hydroxyl group of TYR
 CD2    C    0.00000    2
 NE1   NR   -0.05000    2
```

```
  HE1    H     0.19000    2
  CE2    C     0.00000    2
  CE3    C    -0.10000    3
  HE3   HC     0.10000    3
  CZ2    C    -0.10000    4
  HZ2   HC     0.10000    4
  CZ3    C    -0.10000    5
  HZ3   HC     0.10000    5
  CH2    C    -0.10000    6
  HH2   HC     0.10000    6
   C     C       0.380    7
   O     O      -0.380    7
 [ bonds ]
   N    H    gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O    gb_4
   C   +N    gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1    gb_9
  CG  CD2   gb_15
 CD1  OE4   gb_12
 CD1  NE1    gb_9
 CD2  CE2   gb_15
 CD2  CE3   gb_15
 NE1  HE1    gb_2
 NE1  CE2    gb_9
 CE2  CZ2   gb_15
 CE3  HE3    gb_3
 CE3  CZ3   gb_15
 OE4  HE4    gb_1
 CZ2  HZ2    gb_3
 CZ2  CH2   gb_15
 CZ3  HZ3    gb_3
 CZ3  CH2   gb_15
 CH2  HH2    gb_3
 [ exclusions ]
 ;  ai    aj
   CB   NE1
   CB   CE2
   CB   CE3
   CB   OE4
   CG   HE1
   CG   HE3
   CG   CZ2
   CG   CZ3
  CD1   CE3
  CD1   CZ2
  CD2   HE1
  CD2   OE4
  CD2   HZ2
  CD2   HZ3
  CD2   CH2
  NE1   CE3
  NE1   HE4
  NE1   HZ2
  NE1   CH2
  HE1   OE4
  HE1   CZ2
  CE2   HE3
  CE2   OE4
  CE2   CZ3
  CE2   HH2
  CE3   CZ2
  CE3   HH2
  HE3   HZ3
  HE3   CH2
  CZ2   HZ3
  HZ2   CZ3
  HZ2   HH2
  HZ3   HH2
 [ angles ]
 ;  ai    aj    ak  gromos type
   -C     N     H    ga_31
    H     N    CA    ga_17
   -C     N    CA    ga_30
    N    CA     C    ga_12
   CA     C    +N    ga_18
   CA     C     O    ga_29
    O     C    +N    ga_32
    N    CA    CB    ga_12
    C    CA    CB    ga_12
   CA    CB    CG    ga_14
   CB    CG   CD1    ga_36
   CB    CG   CD2    ga_36
  CD1    CG   CD2     ga_6
   CG   CD1   NE1     ga_6
   CG   CD1   OE4    ga_36
  NE1   CD1   OE4    ga_36
   CG   CD2   CE2     ga_6
  CD1   NE1   CE2     ga_6
  CD1   NE1   HE1    ga_35
  HE1   NE1   CE2    ga_35
  NE1   CE2   CD2     ga_6
   CG   CD2   CE3    ga_38
  NE1   CE2   CZ2    ga_38
  CD2   CE2   CZ2    ga_26
  CE2   CD2   CE3    ga_26
  CD2   CE3   HE3    ga_24
  HE3   CE3   CZ3    ga_24
  CD2   CE3   CZ3    ga_26
  CD1   OE4   HE4    ga_11
  CE2   CZ2   HZ2    ga_24
  HZ2   CZ2   CH2    ga_24
  CE2   CZ2   CH2    ga_26
  CE3   CZ3   HZ3    ga_24
  HZ3   CZ3   CH2    ga_24
  CE3   CZ3   CH2    ga_26
  CZ2   CH2   HH2    ga_24
  HH2   CH2   CZ3    ga_24
  CZ2   CH2   CZ3    ga_26
 [ impropers ]
 ;  ai    aj    ak    al  gromos type
    N    -C    CA     H    gi_1
    C    CA    +N     O    gi_1
   CA     N     C    CB    gi_2
   CG   CD1   CD2    CB    gi_1
  CD2    CG   CD1   NE1    gi_1
  CD1    CG   CD2   CE2    gi_1
   CG   CD1   NE1   CE2    gi_1
   CG   CD2   CE2   NE1    gi_1
  CD1   NE1   CE2   CD2    gi_1
  CD1    CG   NE1   OE4    gi_1
  NE1   CD1   CE2   HE1    gi_1
  CD2   CE2   CE3    CG    gi_1
  CE2   CD2   CZ2   NE1    gi_1
  CE3   CD2   CE2   CZ2    gi_1
  CD2   CE2   CZ2   CH2    gi_1
  CE2   CD2   CE3   CZ3    gi_1
  CE2   CZ2   CH2   CZ3    gi_1
  CD2   CE3   CZ3   CH2    gi_1
  CE3   CZ3   CH2   CZ2    gi_1
  CE3   CD2   CZ3   HE3    gi_1
  CZ2   CE2   CH2   HZ2    gi_1
  CZ3   CE3   CH2   HZ3    gi_1
  CH2   CZ2   CZ3   HH2    gi_1
 [ dihedrals ]
 ;  ai    aj    ak    al  gromos type
  -CA    -C     N    CA    gd_4
   -C     N    CA     C    gd_19
    N    CA     C    +N    gd_20
    N    CA    CB    CG    gd_17
   CA    CB    CG   CD2    gd_20
   CG   CD1   OE4   HE4    gd_2

 ; 3-hydroxyleucine (R)
 [ L3H ]
 [ atoms ]
   N     N   -0.28000    0
   H     H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH1    0.15000    2 ; from the hydroxyl group of THR
 OG1    OA   -0.54800    2 ; from the hydroxyl group of THR
 HG1     H    0.39800    2 ; from the hydroxyl group of THR
 CG2   CH1    0.00000    3
 CD1   CH3    0.00000    3
 CD2   CH3    0.00000    3
   C     C      0.380    4
   O     O     -0.380    4
 [ bonds ]
   N    H    gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O    gb_4
   C   +N    gb_9
  CA   CB   gb_26
  CB  CG2   gb_26
  CB  OG1   gb_17
 OG1  HG1    gb_1
 CG2  CD1   gb_26
 CG2  CD2   gb_26
 [ angles ]
 ;  ai    aj    ak  gromos type
   -C     N     H    ga_31
    H     N    CA    ga_17
   -C     N    CA    ga_30
    N    CA     C    ga_12
   CA     C    +N    ga_18
   CA     C     O    ga_29
    O     C    +N    ga_32
    N    CA    CB    ga_12
    C    CA    CB    ga_12
   CA    CB   OG1    ga_12
   CA    CB  CG2    ga_12
  OG1    CB  CG2    ga_12
```

```
  CB   CG2   CD1    ga_14
  CB   CG2   CD2    ga_14
 CD1   CG2   CD2    ga_14
  CB   OG1   HG1    ga_11
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA     H    gi_1
   C    CA    +N     O    gi_1
  CA     N     C    CB    gi_2
  CB   OG1   CG2    CA    gi_2
 CG2   CD2   CD1    CB    gi_2
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA    -C     N    CA    gd_4
  -C     N    CA     C    gd_19
   N    CA     C    +N    gd_20
   N    CA    CB   CG2    gd_17
  CA    CB   OG1   HG1    gd_12
  CA    CB   CG2   CD1    gd_17

; 3-hydroxyleucine (S)
[ LH3 ]
 [ atoms ]
   N     N   -0.28000     0
   H     H    0.28000     0
  CA   CH1    0.00000     1
  CB   CH1    0.15000     2 ; from the hydroxyl group of THR
 OG1    OA   -0.54800     2 ; from the hydroxyl group of THR
 HG1     H    0.39800     2 ; from the hydroxyl group of THR
 CG2   CH1    0.00000     3
 CD1   CH3    0.00000     3
 CD2   CH3    0.00000     3
   C     C      0.380     4
   O     O     -0.380     4
 [ bonds ]
   N     H    gb_2
   N    CA    gb_20
  CA     C    gb_26
   C     O    gb_4
   C    +N    gb_9
  CA    CB    gb_26
  CB   CG2    gb_26
  CB   OG1    gb_17
 OG1   HG1    gb_1
 CG2   CD1    gb_26
 CG2   CD2    gb_26
 [ angles ]
; ai    aj    ak   gromos type
  -C     N     H    ga_31
   H     N    CA    ga_17
  -C     N    CA    ga_30
   N    CA     C    ga_12
  CA     C    +N    ga_18
  CA     C     O    ga_29
   O     C    +N    ga_32
   N    CA    CB    ga_12
   C    CA    CB    ga_12
  CA    CB   OG1    ga_12
  CA    CB   CG2    ga_12
 OG1    CB   CG2    ga_12
  CB   CG2   CD1    ga_14
  CB   CG2   CD2    ga_14
 CD1   CG2   CD2    ga_14
  CB   OG1   HG1    ga_11
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA     H    gi_1
   C    CA    +N     O    gi_1
  CA     N     C    CB    gi_2
  CB   CG2   OG1    CA    gi_2
 CG2   CD2   CD1    CB    gi_2
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA    -C     N    CA    gd_4
  -C     N    CA     C    gd_19
   N    CA     C    +N    gd_20
   N    CA    CB   CG2    gd_17
  CA    CB   OG1   HG1    gd_12
  CA    CB   CG2   CD1    gd_17

; 4-hydroxyleucine
[ L4H ]
 [ atoms ]
   N     N   -0.28000     0
   H     H    0.28000     0
  CA   CH1    0.00000     1
  CB   CH2    0.00000     1
  CG   CH0    0.15000     2 ; from the hydroxyl group of THR
 OD3    OA   -0.54800     2 ; from the hydroxyl group of THR
 HD3     H    0.39800     2 ; from the hydroxyl group of THR
 CD1   CH3    0.00000     3
 CD2   CH3    0.00000     4
   C     C      0.380     5
   O     O     -0.380     5
 [ bonds ]
   N     H    gb_2
   N    CA    gb_20
  CA     C    gb_26
   C     O    gb_4
   C    +N    gb_9
  CA    CB    gb_26
  CB   CG    gb_26
  CG   CD1    gb_26
  CG   CD2    gb_26
  CG   OD3    gb_17
  OD3   HD3    gb_1
 [ angles ]
; ai    aj    ak   gromos type
  -C     N     H    ga_31
   H     N    CA    ga_17
  -C     N    CA    ga_30
   N    CA     C    ga_12
  CA     C    +N    ga_18
  CA     C     O    ga_29
   O     C    +N    ga_32
   N    CA    CB    ga_12
   C    CA    CB    ga_12
  CA    CB    CG    ga_14
  CB   CG    CD1    ga_12
  CB   CG    CD2    ga_12
  CB   CG    OD3    ga_12
 CD1   CG    CD2    ga_12
 CD1   CG    OD3    ga_12
 CD2   CG    OD3    ga_12
  CG   OD3   HD3    ga_11
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA     H    gi_1
   C    CA    +N     O    gi_1
  CA     N     C    CB    gi_2
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA    -C     N    CA    gd_4
  -C     N    CA     C    gd_19
   N    CA     C    +N    gd_20
   N    CA    CB    CG    gd_17
  CA    CB    CG   OD3    gd_17
  CB    CG   OD3   HD3    gd_12

; 5-hydroxyleucine (R)
[ L5H ]
 [ atoms ]
   N     N   -0.28000     0
   H     H    0.28000     0
  CA   CH1    0.00000     1
  CB   CH2    0.00000     1
  CG   CH1    0.00000     2
 CD1   CH3    0.00000     2
 CD2   CH2    0.15000     3 ; from the hydroxyl group of THR
  OE    OA   -0.54800     3 ; from the hydroxyl group of THR
  HE     H    0.39800     3 ; from the hydroxyl group of THR
   C     C      0.380     4
   O     O     -0.380     4
 [ bonds ]
   N     H    gb_2
   N    CA    gb_20
  CA     C    gb_26
   C     O    gb_4
   C    +N    gb_9
  CA    CB    gb_26
  CB   CG    gb_26
  CG   CD1    gb_26
  CG   CD2    gb_26
 CD2    OE    gb_17
  OE    HE    gb_1
 [ angles ]
; ai    aj    ak   gromos type
  -C     N     H    ga_31
   H     N    CA    ga_17
  -C     N    CA    ga_30
   N    CA     C    ga_12
  CA     C    +N    ga_18
  CA     C     O    ga_29
   O     C    +N    ga_32
   N    CA    CB    ga_12
   C    CA    CB    ga_12
  CA    CB    CG    ga_14
  CB   CG    CD1    ga_14
  CB   CG    CD2    ga_14
 CD1   CG    CD2    ga_14
  CG   CD2    OE    ga_12
 CD2    OE    HE    ga_11
 [ impropers ]
; ai    aj    ak    al   gromos type
   N    -C    CA     H    gi_1
   C    CA    +N     O    gi_1
```

```
  CA    N    C   CB    gi_2
  CG  CD2  CD1   CB    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG  CD2    gd_17
   CB   CG  CD2   OE    gd_17
   CG  CD2   OE   HE    gd_12

; 5-hydroxyleucine (S)
[ LH5 ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH1    0.00000    2
  CD1  CH3    0.00000    2
  CD2  CH2    0.15000    3 ; from the hydroxyl group of THR
   OE   OA   -0.54800    3 ; from the hydroxyl group of THR
   HE    H    0.39800    3 ; from the hydroxyl group of THR
    C    C      0.380    4
    O    O     -0.380    4
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG  CD1   gb_26
   CG  CD2   gb_26
  CD2   OE   gb_17
   OE   HE   gb_1
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB     ga_12
    C   CA   CB     ga_12
   CA   CB   CG     ga_14
   CB   CG  CD1    ga_14
   CB   CG  CD2    ga_14
  CD1   CG  CD2    ga_14
   CG  CD2   OE    ga_12
  CD2   OE   HE    ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
   CG  CD1  CD2   CB    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG  CD2    gd_17
   CB   CG  CD2   OE    gd_17
   CG  CD2   OE   HE    gd_12

; 3-hyroxyvaline
[ V3H ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH0    0.15000    2 ; from the hydroxyl group of THR
  OG3   OA   -0.54800    2 ; from the hydroxyl group of THR
  HG3    H    0.39800    2 ; from the hydroxyl group of THR
  CG1  CH3    0.00000    3
  CG2  CH3    0.00000    4
    C    C      0.380    5
    O    O     -0.380    5
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB  CG1   gb_26
   CB  CG2   gb_26
   CB  OG3   gb_17
  OG3  HG3   gb_1
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB     ga_12
    C   CA   CB     ga_12
   CA   CB  CG1    ga_12
   CA   CB  CG2    ga_12
   CA   CB  OG3    ga_12
  CG1   CB  CG2    ga_12
  CG1   CB  OG3    ga_12
  CG2   CB  OG3    ga_12
   CB  OG3  HG3    ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   OG3    gd_17
   CA   CB  OG3  HG3    gd_12

; cysteine sulfenic acid
[ CYH ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   SG    S    0.15000    2 ; from the hydroxyl group of THR
   OD   OA   -0.54800    2 ; from the hydroxyl group of THR
   HD    H    0.39800    2 ; from the hydroxyl group of THR
    C    C      0.380    3
    O    O     -0.380    3
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   SG   gb_30
   SG   OD   gb_27
   OD   HD   gb_1
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB     ga_12
    C   CA   CB     ga_12
   CA   CB   SG     ga_15
   CB   SG   OD     ga_12
   SG   OD   HD     ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
    N   -C   CA    H    gi_1
    C   CA   +N    O    gi_1
   CA    N    C   CB    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   SG    gd_17
   CA   CB   SG   OD    gd_13
   CB   SG   OD   HD    gd_12

; 5-hydroxyproline (R)
[ PH5 ]
 [ atoms ]
    N    N    0.00000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH1    0.15000    3 ; from the hydroxyl group of THR
   OE   OA   -0.54800    3 ; from the hydroxyl group of THR
   HE    H    0.39800    3 ; from the hydroxyl group of THR
```

```
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   OE   gb_17
  OE   HE   gb_1
  CD    N   gb_20
 [ exclusions ]
; ai   aj
   N   HE
 [ angles ]
; ai   aj   ak   gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_12
  CB   CG   CD    ga_12
  CG   CD   OE    ga_12
  CG   CD    N    ga_12
  OE   CD    N    ga_12
  CD   OE   HE    ga_11
  CD    N   CA    ga_20
  -C    N   CD    ga_30
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   CD    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD   CG    N   OE    gi_2
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB    CG    gd_17
  CA   CB   CG    CD    gd_17
  CB   CG   CD    N    gd_17
  CG   CD   OE    HE    gd_12
  CG   CD    N   CA    gd_19

; 5-hydroxyproline (S)
[ P5H ]
 [ atoms ]
   N    N    0.00000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH1    0.15000    3 ; from the hydroxyl group of THR
  OE   OA   -0.54800    3 ; from the hydroxyl group of THR
  HE    H    0.39800    3 ; from the hydroxyl group of THR
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   OE   gb_17
  OE   HE   gb_1
  CD    N   gb_20
 [ exclusions ]
; ai   aj
   N   HE
 [ angles ]
; ai   aj   ak   gromos type
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_12
  CB   CG   CD    ga_12
  CG   CD   OE    ga_12
  CG   CD    N    ga_12
  OE   CD    N    ga_12
  CD   OE   HE    ga_11
  CD    N   CA    ga_20
  -C    N   CD    ga_30
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   CD    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD    N   CG   OE    gi_2
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB    CG    gd_17
  CA   CB   CG    CD    gd_17
  CB   CG   CD    N    gd_17
  CG   CD   OE    HE    gd_12
  CG   CD    N   CA    gd_19

; glutamic semialdehyde
[ GSA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD    C    0.28000    2 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
  HD   HC    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  OE    O   -0.38000    2 ; from the carbonyl group (of e.g., GLN)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   HD   gb_3
  CD   OE   gb_4
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   HD    ga_24
  HD   CD   OE    ga_24
  CG   CD   OE    ga_26
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD   OE   CG   HD    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB    CG    gd_17
  CA   CB   CG    CD    gd_17
  CB   CG   CD    OE    gd_20

; 2-amino-3-ketobutyric acid
[ TOX ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB    C    0.38000    2 ; from the carbonyl group (of e.g., GLN)
 OG1    O   -0.38000    2 ; from the carbonyl group (of e.g., GLN)
 CG2  CH3    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB  OG1   gb_4
```

```
  CB   CG2    gb_26
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_12
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_12
   C    CA    CB   ga_12
  CA    CB    OG1  ga_26
  CA    CB    CG2  ga_26
 OG1    CB    CG2  ga_26
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
   C    CA    +N   O    gi_1
  CA    N     C    CB   gi_2
  CB    OG1   CG2  CA   gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_4
  -C    N     CA   C    gd_19
   N    CA    C    +N   gd_20
   N    CA    CB   OG1  gd_20

; pyroglutamic acid
[ PGA ]
 [ atoms ]
   N    N     0.00000    0
  CA    CH1   0.00000    1
  CB    CH2   0.00000    1
  CG    CH2   0.00000    2
  CD    C     0.38000    3 ; from the carbonyl group (of e.g., GLN)
  OE    O     -0.38000   3 ; from the carbonyl group (of e.g., GLN)
   C    C     0.380      4
   O    O     -0.380     4
 [ bonds ]
   N    CA    gb_20
  CA    C     gb_26
   C    O     gb_4
   C    +N    gb_9
  CA    CB    gb_26
  CB    CG    gb_26
  CG    CD    gb_26
  CD    OE    gb_4
  CD    N     gb_9
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     CA   ga_30
   N    CA    C    ga_12
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_12
   C    CA    CB   ga_12
  CA    CB    CG   ga_12
  CB    CG    CD   ga_12
  CG    CD    N    ga_18
  CG    CD    OE   ga_29
   N    CD    OE   ga_32
  CD    N     CA   ga_20
  -C    N     CD   ga_30
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   CD   gi_1
   C    CA    +N   O    gi_1
  CA    N     C    CB   gi_2
  CD    CG    N    OE   gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_4
  -C    N     CA   C    gd_19
   N    CA    C    +N   gd_20
   N    CA    CB   CG   gd_17
  CA    CB    CG   CD   gd_17
  CB    CG    CD   N    gd_20
  CG    CD    N    CA   gd_4

; 2-oxo-histidine
[ H2X ]
 [ atoms ]
   N    N     -0.28000   0
   H    H     0.28000    0
  CA    CH1   0.00000    1
  CB    CH2   0.00000    1
  CG    C     0.00000    2
 ND1    NR    -0.28000   3
 HD1    H     0.28000    3
 CD2    CR1   0.00000    4
 NE2    NR    -0.28000   5
 HE2    H     0.28000    5
 CE1    C     0.38000    6 ; from the carbonyl group (of e.g., GLN)
 OZ     O     -0.38000   6 ; from the carbonyl group (of e.g., GLN)
  C     C     0.380      7
  O     O     -0.380     7
 [ bonds ]
   N    H     gb_2
   N    CA    gb_20
  CA    C     gb_26
   C    O     gb_4
   C    +N    gb_9
  CA    CB    gb_26
  CB    CG    gb_26
  CG    ND1   gb_9
  CG    CD2   gb_9
 ND1    HD1   gb_2
 ND1    CE1   gb_9
 CD2    NE2   gb_9
 CE1    NE2   gb_9
 CE1    OZ    gb_4
 NE2    HE2   gb_2
 [ exclusions ]
; ai    aj
  CB    HD1
  CB    CE1
  CB    NE2
  CG    HE2
  CG    OZ
 HD1    CD2
 HD1    NE2
 HD1    OZ
 ND1    HE2
 CD2    OZ
 HE2    OZ
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_31
   H    N     CA   ga_17
  -C    N     CA   ga_30
   N    CA    C    ga_12
  CA    C     +N   ga_18
  CA    C     O    ga_29
   O    C     +N   ga_32
   N    CA    CB   ga_12
   C    CA    CB   ga_12
  CA    CB    CG   ga_14
  CB    CG    ND1  ga_36
  CB    CG    CD2  ga_36
 ND1    CG    CD2  ga_6
  CG    ND1   HD1  ga_35
  CG    ND1   CE1  ga_6
 HD1    ND1   CE1  ga_35
  CG    CD2   NE2  ga_6
 ND1    CE1   NE2  ga_6
 ND1    CE1   OZ   ga_36
 NE2    CE1   OZ   ga_36
 CD2    NE2   CE1  ga_6
 CD2    NE2   HE2  ga_35
 CE1    NE2   HE2  ga_35
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
   C    CA    +N   O    gi_1
  CA    N     C    CB   gi_2
  CG    ND1   CD2  CB   gi_1
 CD2    CG    ND1  CE1  gi_1
 ND1    CG    CD2  NE2  gi_1
 ND1    CG    CE1  HD1  gi_1
  CG    ND1   CE1  NE2  gi_1
  CG    CD2   NE2  CE1  gi_1
 CD2    NE2   CE1  ND1  gi_1
 CE1    OZ    NE2  ND1  gi_1
 NE2    CD2   CE1  HE2  gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_4
  -C    N     CA   C    gd_19
   N    CA    C    +N   gd_20
   N    CA    CB   CG   gd_17
  CA    CB    CG   ND1  gd_20

; methionine sulfoxide (R)
[ MSX ]
 [ atoms ]
   N    N     -0.28000   0
   H    H     0.28000    0
  CA    CH1   0.00000    1
  CB    CH2   0.00000    1
  CG    CH2   0.00000    1
  SD    S     0.38000    2 ; to add up to 0 net charge
 OE2    O     -0.38000   2 ; from the carbonyl group (e.g., GLN)
 CE1    CH3   0.00000    3
  C     C     0.380      4
```

```
   O    O    -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   SD   gb_30
  SD  CE1   gb_30
  SD  OE2   gb_24
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   SD    ga_15
  CG   SD  CE1    ga_3
  CG   SD  OE2    ga_5
 CE1   SD  OE2    ga_5
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  SD  OE2  CE1   CG   gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   SD   gd_17
  CB   CG   SD  CE1   gd_13

; methionine sulfoxide (S)
[ MXS ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  SD    S    0.38000    2 ; to add up to 0 net charge
 OE2    O   -0.38000    2 ; from the carbonyl group (e.g., GLN)
 CE1  CH3    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   SD   gb_30
  SD  CE1   gb_30
  SD  OE2   gb_24
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   SD    ga_15
  CG   SD  CE1    ga_3
  CG   SD  OE2    ga_5
 CE1   SD  OE2    ga_5
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  SD  CE1  OE2   CG   gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   SD   gd_17
  CB   CG   SD  CE1   gd_13

; methionine sulfone
[ MES ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  SD    S    0.72000    2 ; to add up to 0 net charge
 OE2    O   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3    O   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
 CE1  CH3    0.00000    3
   C    C      0.380    4
   O    O     -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   SD   gb_30
  SD  CE1   gb_29
  SD  OE2   gb_24
  SD  OE3   gb_24
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   SD    ga_15
  CG   SD  CE1    ga_12
  CG   SD  OE2    ga_12
  CG   SD  OE3    ga_12
 OE2   SD  OE3    ga_12
 CE1   SD  OE2    ga_12
 CE1   SD  OE3    ga_12
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   SD   gd_17
  CB   CG   SD  CE1   gd_13

; cysteine sulfinic acid
[ CSA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  SG    S    0.12000    2 ; to add up to -1 net charge
 OD1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OD2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   SG   gb_30
  SG  OD1   gb_24
  SG  OD2   gb_24
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
```

```
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   SG    ga_15
  CB   SG  OD1    ga_5
 OD1   SG  OD2    ga_5
  CB   SG  OD2    ga_5
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  SG   CB  OD1  OD2    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   SG    gd_17
  CA   CB   SG  OD1    gd_13

; cysteic acid
[ CSE ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  SG    S    0.75500    2 ; to add up to -1 net charge
 OD1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OD2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OD3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C     0.380    3
   O    O    -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   SG   gb_30
  SG  OD1   gb_24
  SG  OD2   gb_24
  SG  OD3   gb_24
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   SG    ga_15
  CB   SG  OD1    ga_12
 OD1   SG  OD2    ga_12
  CB   SG  OD2    ga_12
  CB   SG  OD3    ga_12
 OD1   SG  OD3    ga_12
 OD2   SG  OD3    ga_12
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   SG    gd_17
  CA   CB   SG  OD1    gd_13

; 3-nitrotyrosine (-1)
[ YNI ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C   -0.10000    3
 HD2   HC    0.10000    3
 CE1    C    0.10000    4 ; to add up to -1 net charge
 NZ1   NR    0.05500    4 ; to add up to -1 net charge
 OH1    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OH2    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 CZ2    C    0.20000    4 ; to add up to -1 net charge
 OH3   OM   -0.63500    4 ; from the carboxyl group (of e.g., GLU)
 CE2    C   -0.10000    5
 HE2   HC    0.10000    5
   C    C     0.380    6
   O    O    -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  CE1   gb_15
 CD2  HD2   gb_3
 CD2  CE2   gb_15
 CE1  NZ1   gb_11
 NZ1  OH1   gb_5
 NZ1  OH2   gb_5
 CE1  CZ2   gb_15
 CE2  HE2   gb_3
 CE2  CZ2   gb_15
 CZ2  OH3   gb_12
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  OH3
 HD1  CD2
 HD1  NZ1
 HD1  CZ2
 CD2  CE1
 CD2  OH3
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 HE2  OH3
 NZ1  CE2
 NZ1  OH3
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_26
  CB   CG  CD2    ga_26
 CD1   CG  CD2    ga_26
  CG  CD1  HD1    ga_24
 HD1  CD1  CE1    ga_24
  CG  CD1  CE1    ga_26
  CG  CD2  HD2    ga_24
 HD2  CD2  CE2    ga_24
  CG  CD2  CE2    ga_26
 CD1  CE1  NZ1    ga_26
 CD1  CE1  CZ2    ga_26
 NZ1  CE1  CZ2    ga_26
 CE1  NZ1  OH1    ga_26
 CE1  NZ1  OH2    ga_26
 OH1  NZ1  OH2    ga_26
 CD2  CE2  HE2    ga_24
 HE2  CE2  CZ2    ga_24
 CD2  CE2  CZ2    ga_26
 CE1  CZ2  CE2    ga_26
 CE1  CZ2  OH3    ga_26
 CE2  CZ2  OH3    ga_26
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
 CD2   CG  CD1  CE1    gi_1
 CD1   CG  CD2  CE2    gi_1
```

```
  CG   CD1   CE1   CZ2    gi_1
  CG   CD2   CE2   CZ2    gi_1
  CD1   CE1   CZ2   CE2    gi_1
  CD2   CE2   CZ2   CE1    gi_1
  CD1   CG   CE1   HD1    gi_1
  CD2   CG   CE2   HD2    gi_1
  CE1   CZ2   CD1   NZ1    gi_1
  CE2   CZ2   CD2   HE2    gi_1
  CZ2   CE1   CE2   OH3    gi_1
  NZ1   CE1   OH1   OH2    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N    CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG   CD1    gd_20
  CE1   CZ2   OH3   HH3    gd_2
  CD1   CE1   NZ1   OH1    gd_4

; 3-nitrotyrosine (0)
[ YNN ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG    C    0.00000    1
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C    0.10000    4 ; newly developed parameters tp match the experimental
HFE
  NZ1   NR    0.30000    4 ; newly developed parameters tp match the experimental
HFE
  OH1    O   -0.20000    4 ; newly developed parameters tp match the experimental
HFE
  OH2    O   -0.20000    4 ; newly developed parameters tp match the experimental
HFE
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
  CZ2    C    0.05000    6 ; newly developed parameters tp match the experimental
HFE
  OH3   OA   -0.36000    6 ; newly developed parameters tp match the experimental
HFE
  HH3    H    0.31000    6 ; newly developed parameters tp match the experimental
HFE
    C    C      0.380    7
    O    O     -0.380    7
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG   CD1   gb_15
   CG   CD2   gb_15
  CD1   HD1   gb_3
  CD1   CE1   gb_15
  CD2   HD2   gb_3
  CD2   CE2   gb_15
  CE1   NZ1   gb_11
  NZ1   OH1   gb_5
  NZ1   OH2   gb_5
  CE1   CZ2   gb_15
  CE2   HE2   gb_3
  CE2   CZ2   gb_15
  CZ2   OH3   gb_12
  OH3   HH3   gb_1
 [ exclusions ]
;  ai    aj
   CB   HD1
   CB   HD2
   CB   CE1
   CB   CE2
   CG   HE2
   CG   NZ1
   CG   CZ2
  CD1   HD2
  CD1   CE2
  CD1   OH3
  HD1   CD2
  HD1   NZ1
  HD1   CZ2
  CD2   CE1
  CD2   OH3
  HD2   HE2
  HD2   CZ2
  CE1   HE2
  HE2   OH3
  NZ1   CE2
  NZ1   OH3
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_31
    H    N   CA    ga_17
   -C    N   CA    ga_30
    N   CA    C    ga_12
   CA    C   +N    ga_18
   CA    C    O    ga_29
    O    C   +N    ga_32
    N   CA   CB    ga_12
    C   CA   CB    ga_12
   CA   CB   CG    ga_14
   CB   CG   CD1    ga_26
   CB   CG   CD2    ga_26
  CD1   CG   CD2    ga_26
   CG   CD1   HD1    ga_24
  HD1   CD1   CE1    ga_24
   CG   CD1   CE1    ga_26
   CG   CD2   HD2    ga_24
  HD2   CD2   CE2    ga_24
   CG   CD2   CE2    ga_26
  CD1   CE1   NZ1    ga_26
  CD1   CE1   CZ2    ga_26
  NZ1   CE1   CZ2    ga_26
  CE1   NZ1   OH1    ga_26
  CE1   NZ1   OH2    ga_26
  OH1   NZ1   OH2    ga_26
  CD2   CE2   HE2    ga_24
  HE2   CE2   CZ2    ga_24
  CD2   CE2   CZ2    ga_26
  CE1   CZ2   CE2    ga_26
  CE1   CZ2   OH3    ga_26
  CE2   CZ2   OH3    ga_26
  CZ2   OH3   HH3    ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H    gi_1
    C   CA    +N    O    gi_1
   CA    N    C   CB    gi_2
   CG   CD1   CD2   CB    gi_1
  CD2   CG   CD1   CE1    gi_1
  CD1   CG   CD2   CE2    gi_1
   CG   CD1   CE1   CZ2    gi_1
   CG   CD2   CE2   CZ2    gi_1
  CD1   CE1   CZ2   CE2    gi_1
  CD2   CE2   CZ2   CE1    gi_1
  CD1   CG   CE1   HD1    gi_1
  CD2   CG   CE2   HD2    gi_1
  CE1   CZ2   CD1   NZ1    gi_1
  CE2   CZ2   CD2   HE2    gi_1
  CZ2   CE1   CE2   OH3    gi_1
  NZ1   CE1   OH1   OH2    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N    CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG   CD1    gd_20
  CE1   CZ2   OH3   HH3    gd_2
  CD1   CE1   NZ1   OH1    gd_4

; 3-nitrotyrosine (0)
[ YNB ]
 [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG    C    0.00000    1
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C    0.10000    4 ; to add up to 0 net charge
  NZ1   NR    0.62000    4 ; to add up to 0 net charge
  OH1    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OH2    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
  CZ2    C    0.15000    6
  OH3   OA   -0.54800    6
  HH3    H    0.39800    6
    C    C      0.380    7
    O    O     -0.380    7
 [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
```

```
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_15
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  CE1   gb_15
 CD2  HD2   gb_3
 CD2  CE2   gb_15
 CE1  NZ1   gb_11
 NZ1  OH1   gb_5
 NZ1  OH2   gb_5
 CE1  CZ2   gb_15
 CE2  HE2   gb_3
 CE2  CZ2   gb_15
 CZ2  OH3   gb_12
 OH3  HH3   gb_1
[ exclusions ]
;  ai    aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  OH3
 HD1  CD2
 HD1  NZ1
 HD1  CZ2
 CD2  CE1
 CD2  OH3
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 HE2  OH3
 NZ1  CE2
 NZ1  OH3
[ angles ]
;  ai    aj    ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_26
  CB   CG  CD2    ga_26
 CD1   CG  CD2    ga_26
  CG  CD1  HD1    ga_24
 HD1  CD1  CE1    ga_24
  CG  CD1  CE1    ga_26
  CG  CD2  HD2    ga_24
 HD2  CD2  CE2    ga_24
  CG  CD2  CE2    ga_26
 CD1  CE1  NZ1    ga_26
 CD1  CE1  CZ2    ga_26
 NZ1  CE1  CZ2    ga_26
 CE1  NZ1  OH1    ga_26
 CE1  NZ1  OH2    ga_26
 OH1  NZ1  OH2    ga_26
 CD2  CE2  HE2    ga_24
 HE2  CE2  CZ2    ga_24
 CD2  CE2  CZ2    ga_26
 CE1  CZ2  CE2    ga_26
 CE1  CZ2  OH3    ga_26
 CE2  CZ2  OH3    ga_26
 CZ2  OH3  HH3    ga_11
[ impropers ]
;  ai    aj    ak    al  gromos type
   N   -C    CA    H    gi_1
   C   CA    +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
 CD2   CG  CD1  CE1    gi_1
 CD1   CG  CD2  CE2    gi_1
  CG  CD1  CE1  CZ2    gi_1
  CG  CD2  CE2  CZ2    gi_1
 CD1  CE1  CZ2  CE2    gi_1
 CD2  CE2  CZ2  CE1    gi_1
 CD1   CG  CE1  HD1    gi_1
 CD2   CG  CE2  HD2    gi_1
 CE1  CZ2  CD1  NZ1    gi_1
 CE2  CZ2  CD2  HE2    gi_1
 CZ2  CE1  CE2  OH3    gi_1
 NZ1  CE1  OH1  OH2    gi_1
[ dihedrals ]
;  ai    aj    ak    al  gromos type
 -CA   -C    N   CA    gd_4
 -C    N   CA    C    gd_19
  N   CA    C   +N    gd_20
  N   CA   CB   CG    gd_17
 CA   CB   CG  CD1    gd_20
 CE1  CZ2  OH3  HH3    gd_2
 CD1  CE1  NZ1  OH1    gd_4

; 6-nitrotryptophan
[ WNI ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.14000    2
 CD1    C   -0.10000    2
 HD1   HC    0.10000    2
 CD2    C    0.00000    2
 NE1   NR   -0.05000    2
 HE1    H    0.19000    2
 CE2    C    0.00000    2
 CE3    C   -0.10000    3
 HE3   HC    0.10000    3
 CZ2    C   -0.10000    4
 HZ2   HC    0.10000    4
 CZ3    C   -0.10000    5
 HZ3   HC    0.10000    5
 CH2    C    0.10000    6 ; to add up to 0 net charge
  NT   NR    0.62000    6 ; to add up to 0 net charge
 OI1    O   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2    O   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C     0.380    7
   O    O    -0.380    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  CD1   gb_9
  CG  CD2   gb_15
 CD1  HD1   gb_3
 CD1  NE1   gb_9
 CD2  CE2   gb_15
 CD2  CE3   gb_15
 NE1  HE1   gb_2
 NE1  CE2   gb_9
 CE2  CZ2   gb_15
 CE3  HE3   gb_3
 CE3  CZ3   gb_15
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_15
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_15
 CH2   NT   gb_11
  NT  OI1   gb_5
  NT  OI2   gb_5
 [ exclusions ]
 ;  ai    aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2   NT
 CE3  CZ2
 CE3   NT
 HE3  HZ3
 HE3  CH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2   NT
```

```
 HZ3   NT
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H    ga_31
   H   N   CA    ga_17
  -C   N   CA    ga_30
   N   CA   C    ga_12
  CA   C   +N    ga_18
  CA   C   O    ga_29
   O   C   +N    ga_32
   N   CA  CB    ga_12
   C   CA  CB    ga_12
  CA   CB  CG    ga_14
  CB   CG  CD1   ga_36
  CB   CG  CD2   ga_36
 CD1   CG  CD2   ga_6
  CG   CD1 HD1   ga_35
 HD1   CD1 NE1   ga_35
  CG   CD1 NE1   ga_6
  CG   CD2 CE2   ga_6
 CD1   NE1 CE2   ga_6
 CD1   NE1 HE1   ga_35
 HE1   NE1 CE2   ga_35
 NE1   CE2 CD2   ga_6
  CG   CD2 CE3   ga_38
 NE1   CE2 CZ2   ga_38
 CD2   CE2 CZ2   ga_26
 CE2   CD2 CE3   ga_26
 CD2   CE3 HE3   ga_24
 HE3   CE3 CZ3   ga_24
 CD2   CE3 CZ3   ga_26
 CE2   CZ2 HZ2   ga_24
 HZ2   CZ2 CH2   ga_24
 CE2   CZ2 CH2   ga_26
 CE3   CZ3 HZ3   ga_24
 HZ3   CZ3 CH2   ga_24
 CE3   CZ3 CH2   ga_26
 CZ2   CH2 CZ3   ga_26
 CZ2   CH2 NT    ga_26
 CZ3   CH2 NT    ga_26
 CH2   NT  OI1   ga_26
 CH2   NT  OI2   ga_26
 OI1   NT  OI2   ga_26
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H    gi_1
   C   CA   +N   O    gi_1
  CA   N    C   CB    gi_2
  CG   CD1  CD2  CB   gi_1
 CD2   CG   CD1  NE1  gi_1
 CD1   CG   CD2  CE2  gi_1
  CG   CD1  NE1  CE2  gi_1
  CG   CD2  CE2  NE1  gi_1
 CD1   NE1  CE2  CD2   gi_1
 CD1   CG   NE1  HD1   gi_1
 NE1   CD1  CE2  HE1   gi_1
 CD2   CE2  CE3  CG    gi_1
 CE2   CD2  CZ2  NE1   gi_1
 CE3   CD2  CE2  CZ2   gi_1
 CD2   CE2  CZ2  CH2   gi_1
 CE2   CD2  CE3  CZ3   gi_1
 CE2   CZ2  CH2  CZ3   gi_1
 CD2   CE3  CZ3  CH2   gi_1
 CE3   CZ3  CH2  CZ2   gi_1
 CE3   CD2  CZ3  HE3   gi_1
 CZ2   CE2  CH2  HZ2   gi_1
 CZ3   CE3  CH2  HZ3   gi_1
 CH2   CZ2  CZ3  NT    gi_1
  NT   OI1  OI2  CH2   gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C   N    CA   gd_4
  -C   N    CA   C    gd_19
   N   CA   C    +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD2  gd_20
 CZ2   CH2  NT   OI1  gd_4

; kynurenine
[ WKY ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.38000    2
 OD1    O   -0.38000    2
 CD2    C    0.00000    3
 CE1    C    0.00000    3
 NZ1   NT   -0.83000    4 ; from ARGN
HZ11    H    0.41500    4 ; from ARGN
HZ12    H    0.41500    4 ; from ARGN
 CE2    C   -0.10000    5
 HE2   HC    0.10000    5
 CZ2    C   -0.10000    6
 HZ2   HC    0.10000    6
 CZ3    C   -0.10000    7
 HZ3   HC    0.10000    7
  CH    C   -0.10000    8
  HH   HC    0.10000    8
   C    C     0.380    9
   O    O    -0.380    9
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  CD2   gb_22
 CD2  CE1   gb_15
 CD2  CE2   gb_15
 CE1  NZ1   gb_8
 CE1  CZ2   gb_15
 CE2  HE2   gb_3
 CE2  CZ3   gb_15
 NZ1 HZ11   gb_2
 NZ1 HZ12   gb_2
 CZ2  HZ2   gb_3
 CZ2   CH   gb_15
 CZ3  HZ3   gb_3
 CZ3   CH   gb_15
  CH   HH   gb_3
 [ exclusions ]
; ai   aj
  CB  CE1
  CB  CE2
  CG  NZ1
  CG  HE2
  CG  CZ2
  CG  CZ3
 CD2 HZ11
 CD2 HZ12
 CD2  HZ2
 CD2  HZ3
 CD2   CH
 CE1  HE2
 CE1  CZ3
 CE1   HH
 CE2  CZ2
 CE2   HH
 HE2  HZ3
 HE2   CH
 NZ1  CE2
 NZ1  HZ2
 NZ1   CH
HZ11  CZ2
HZ12  CZ2
 CZ2  HZ3
 HZ2  CZ3
 HZ2   HH
 HZ3   HH
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H    ga_31
   H   N   CA    ga_17
  -C   N   CA    ga_30
   N   CA   C    ga_12
  CA   C   +N    ga_18
  CA   C   O    ga_29
   O   C   +N    ga_32
   N   CA  CB    ga_12
   C   CA  CB    ga_12
  CA   CB  CG    ga_14
  CB   CG  OD1   ga_26
  CB   CG  CD2   ga_26
 OD1   CG  CD2   ga_26
  CG   CD2 CE1   ga_26
  CG   CD2 CE2   ga_26
 CE1   CD2 CE2   ga_26
 CD2   CE1 NZ1   ga_26
 CD2   CE1 CZ2   ga_26
 NZ1   CE1 CZ2   ga_26
 CD2   CE2 HE2   ga_24
 HE2   CE2 CZ3   ga_24
 CD2   CE2 CZ3   ga_26
 CE1   NZ1 HZ11   ga_22
 CE1   NZ1 HZ12   ga_22
HZ11   NZ1 HZ12   ga_23
 CE1   CZ2 HZ2   ga_24
 HZ2   CZ2  CH   ga_24
 CE1   CZ2  CH   ga_26
 CE2   CZ3 HZ3   ga_24
 HZ3   CZ3  CH   ga_24
```

```
  CE2   CZ3    CH    ga_26
  CZ2   CH    HH    ga_24
  CZ3   CH    HH    ga_24
  CZ2   CH  CZ3    ga_26
 [ impropers ]
; ai    aj    ak    al  gromos type
    N    -C    CA    H    gi_1
    C    CA   +N    O    gi_1
   CA     N     C   CB    gi_2
   CG   CD2   OD1   CB    gi_1
  CD2   CE1   CE2   CG    gi_1
  CE2   CD2   CE1  CZ2    gi_1
  CE1   CD2   CE2  CZ3    gi_1
  CD2   CE1   CZ2   CH    gi_1
  CD2   CE2   CZ3   CH    gi_1
  CE1   CZ2    CH  CZ3    gi_1
  CE2   CZ3    CH  CZ2    gi_1
  CE1   CD2   CZ2  NZ1    gi_1
  CE2   CD2   CZ3  HE2    gi_1
  NZ1  HZ11  HZ12  CE1    gi_1
  CZ2    CH   CE1  HZ2    gi_1
  CZ3    CH   CE2  HZ3    gi_1
   CH   CZ2   CZ3   HH    gi_1
 [ dihedrals ]
; ai    aj    ak    al  gromos type
  -CA    -C     N    CA    gd_4
   -C     N    CA     C   gd_19
    N    CA     C   +N    gd_20
    N    CA    CB    CG    gd_17
   CA    CB    CG  CD2    gd_20
   CB    CG  CD2   CE1    gd_1
  CD2   CE1   NZ1  HZ11    gd_4

; 3-hydroxykynurenine
[ WKH ]
 [ atoms ]
    N     N   -0.28000     0
    H     H    0.28000     0
   CA   CH1    0.00000     1
   CB   CH2    0.00000     1
   CG     C    0.38000     2
  OD1     O   -0.38000     2
  CD2     C    0.00000     3
  CE1     C    0.00000     3
  NZ1    NT   -0.83000     4 ; from ARGN
 HZ11     H    0.41500     4 ; from ARGN
 HZ12     H    0.41500     4 ; from ARGN
  CE2     C   -0.10000     5
  HE2    HC    0.10000     5
  CZ2     C    0.15000     6 ; from the hydroxyl group of TYR
  OH1    OA   -0.54800     6 ; from the hydroxyl group of TYR
  HH1     H    0.39800     6 ; from the hydroxyl group of TYR
  CZ3     C   -0.10000     7
  HZ3    HC    0.10000     7
  CH2     C   -0.10000     8
  HH2    HC    0.10000     8
    C     C      0.380     9
    O     O     -0.380     9
 [ bonds ]
    N     H   gb_2
    N    CA   gb_20
   CA     C   gb_26
    C     O   gb_4
    C    +N   gb_9
   CA    CB   gb_26
   CB    CG   gb_26
   CG   OD1   gb_4
   CG   CD2   gb_22
  CD2   CE1   gb_15
  CD2   CE2   gb_15
  CE1   NZ1   gb_8
  CE1   CZ2   gb_15
  CE2   HE2   gb_3
  CE2   CZ3   gb_15
  NZ1  HZ11   gb_2
  NZ1  HZ12   gb_2
  CZ2   OH1   gb_12
  CZ2   CH2   gb_15
  CZ3   HZ3   gb_3
  CZ3   CH2   gb_15
  OH1   HH1   gb_1
  CH2   HH2   gb_3
 [ exclusions ]
; ai    aj
   CB   CE1
   CB   CE2
   CG   HE2
   CG   NZ1
   CG   CZ2
   CG   CZ3
  CD2  HZ11
  CD2  HZ12
  CD2   HZ3
  CD2   OH1
  CD2   CH2
  CE1   HE2
  CE1   CZ3
  CE1   HH1
  CE1   HH2
  CE2   CZ2
  CE2   HH2
  HE2   HZ3
  HE2   CH2
  NZ1   CE2
  NZ1   OH1
  NZ1   HH1
  NZ1   CH2
 HZ11   CZ2
 HZ11   OH1
 HZ12   CZ2
 HZ12   OH1
  CZ2   HZ3
  HZ3   HH2
  OH1   CZ3
  OH1   HH2
  HH1   CH2
 [ angles ]
; ai    aj    ak  gromos type
   -C     N     H   ga_31
    H     N    CA   ga_17
   -C     N    CA   ga_30
    N    CA     C   ga_12
   CA     C    +N   ga_18
   CA     C     O   ga_29
    O     C    +N   ga_32
    N    CA    CB   ga_12
    C    CA    CB   ga_12
   CA    CB    CG   ga_14
   CB    CG   OD1   ga_26
   CB    CG   CD2   ga_26
  OD1    CG   CD2   ga_26
   CG   CD2   CE1   ga_26
   CG   CD2   CE2   ga_26
  CE1   CD2   CE2   ga_26
  CD2   CE1   NZ1   ga_26
  CD2   CE1   CZ2   ga_26
  NZ1   CE1   CZ2   ga_26
  CD2   CE2   HE2   ga_24
  HE2   CE2   CZ3   ga_24
  CD2   CE2   CZ3   ga_26
  CE1   NZ1  HZ11   ga_22
  CE1   NZ1  HZ12   ga_22
 HZ11   NZ1  HZ12   ga_23
  CE1   CZ2   OH1   ga_26
  CE1   CZ2   CH2   ga_26
  OH1   CZ2   CH2   ga_26
  CE2   CZ3   HZ3   ga_24
  HZ3   CZ3   CH2   ga_24
  CE2   CZ3   CH2   ga_26
  CZ2   OH1   HH1   ga_11
  CZ2   CH2   HH2   ga_24
  CZ3   CH2   HH2   ga_24
  CZ2   CH2   CZ3   ga_26
 [ impropers ]
; ai    aj    ak    al  gromos type
    N    -C    CA    H    gi_1
    C    CA   +N    O    gi_1
   CA     N     C   CB    gi_2
   CG   CD2   OD1   CB    gi_1
  CD2   CE1   CE2   CG    gi_1
  CE2   CD2   CE1  CZ2    gi_1
  CE1   CD2   CE2  CZ3    gi_1
  CD2   CE1   CZ2  CH2    gi_1
  CD2   CE2   CZ3  CH2    gi_1
  CE1   CZ2   CH2  CZ3    gi_1
  CE2   CZ3   CH2  CZ2    gi_1
  CE1   CD2   CZ2  NZ1    gi_1
  CE2   CD2   CZ3  HE2    gi_1
  NZ1  HZ11  HZ12  CE1    gi_1
  CZ2   CH2   CE1  OH1    gi_1
  CZ3   CH2   CE2  HZ3    gi_1
  CH2   CZ2   CZ3  HH2    gi_1
 [ dihedrals ]
; ai    aj    ak    al  gromos type
  -CA    -C     N    CA    gd_4
   -C     N    CA     C   gd_19
    N    CA     C   +N    gd_20
    N    CA    CB    CG    gd_17
   CA    CB    CG  CD2    gd_20
   CB    CG  CD2   CE1    gd_1
  CD2   CE1   NZ1  HZ11    gd_4
  CE1   CZ2   OH1   HH1    gd_2

; formyl-kynurenine
[ WKF ]
 [ atoms ]
```

```
   N    N   -0.28000    0
   H    H    0.28000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C    0.38000    2
  OD1   O   -0.38000    2
  CD2   C    0.00000    3
  CE1   C    0.00000    3
  NZ1   N   -0.28000    4 ; from the peptide bond
  HZ1   H    0.28000    4 ; from the peptide bond
  CH1   C    0.28000    5 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
  HH1   HC   0.10000    5 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  OI    O   -0.38000    5 ; from the carbonyl group (of e.g., GLN)
  CE2   C   -0.10000    6
  HE2   HC   0.10000    6
  CZ2   C   -0.10000    7
  HZ2   HC   0.10000    7
  CZ3   C   -0.10000    8
  HZ3   HC   0.10000    8
  CH2   C   -0.10000    9
  HH2   HC   0.10000    9
   C    C     0.380   10
   O    O    -0.380   10
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG  OD1   gb_4
  CG  CD2   gb_22
  CD2  CE1   gb_15
  CD2  CE2   gb_15
  CE1  NZ1   gb_8
  CE1  CZ2   gb_15
  CE2  HE2   gb_3
  CE2  CZ3   gb_15
  NZ1  HZ1   gb_2
  NZ1  CH1   gb_9
  CZ2  HZ2   gb_3
  CZ2  CH2   gb_15
  CZ3  HZ3   gb_3
  CZ3  CH2   gb_15
  CH1  HH1   gb_3
  CH1   OI   gb_4
  CH2  HH2   gb_3
 [ exclusions ]
;  ai   aj
   CB  CE1
   CB  CE2
   CG  NZ1
   CG  HE2
   CG  CZ2
   CG  CZ3
  CD2  HZ1
  CD2  HZ2
  CD2  HZ3
  CD2  CH2
  CE1  HE2
  CE1  CZ3
  CE1  HH2
  CE2  CZ2
  CE2  HH2
  HE2  HZ3
  HE2  CH2
  NZ1  CE2
  NZ1  HZ2
  NZ1  CH2
  HZ1  CZ2
  CZ2  HZ3
  HZ2  CZ3
  HZ2  HH2
  HZ3  HH2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  OD1    ga_26
  CB   CG  CD2    ga_26
 OD1   CG  CD2    ga_26
  CG  CD2  CE1    ga_26
  CG  CD2  CE2    ga_26
 CE1  CD2  CE2    ga_26
 CD2  CE1  NZ1    ga_26
 CD2  CE1  CZ2    ga_26
 NZ1  CE1  CZ2    ga_26
 CD2  CE2  HE2    ga_24
 HE2  CE2  CZ3    ga_24
 CD2  CE2  CZ3    ga_26
 CE1  NZ1  HZ1    ga_17
 HZ1  NZ1  CH1    ga_31
 CE1  NZ1  CH1    ga_30
 CE1  CZ2  HZ2    ga_24
 HZ2  CZ2  CH2    ga_24
 CE1  CZ2  CH2    ga_26
 CE2  CZ3  HZ3    ga_24
 HZ3  CZ3  CH2    ga_24
 CE2  CZ3  CH2    ga_26
 NZ1  CH1  HH1    ga_18
 HH1  CH1   OI    ga_29
 NZ1  CH1   OI    ga_32
 CZ2  CH2  HH2    ga_24
 CZ3  CH2  HH2    ga_24
 CZ2  CH2  CZ3    ga_26
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CG  CD2  OD1   CB    gi_1
 CD2  CE1  CE2   CG    gi_1
 CE2  CD2  CE1  CZ2    gi_1
 CE1  CD2  CE2  CZ3    gi_1
 CD2  CE1  CZ2  CH2    gi_1
 CD2  CE2  CZ3  CH2    gi_1
 CE1  CZ2  CH2  CZ3    gi_1
 CE2  CZ3  CH2  CZ2    gi_1
 CE1  CD2  CZ2  NZ1    gi_1
 CE2  CD2  CZ3  HE2    gi_1
 NZ1  CH1  CE1  HZ1    gi_1
 CZ2  CH2  CE1  HZ2    gi_1
 CZ3  CH2  CE2  HZ3    gi_1
 CH1   OI  NZ1  HH1    gi_1
 CH2  CZ2  CZ3  HH2    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_4
   -C    N   CA    C    gd_19
    N   CA    C   +N    gd_20
    N   CA   CB   CG    gd_17
   CA   CB   CG  CD2    gd_20
   CB   CG  CD2  CE1    gd_1
  CD2  CE1  NZ1  CH1    gd_4
  CE1  NZ1  CH1   OI    gd_4

 ; chlorotyrosine
 [ YCH ]
  [ atoms ]
    N    N   -0.28000    0
    H    H    0.28000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.00000    1
  CD1    C   -0.10000    2
  HD1   HC    0.10000    2
  CD2    C   -0.10000    3
  HD2   HC    0.10000    3
  CE1    C    0.08700    4 ; from CHCL3
 CLZ1   CL   -0.08700    4 ; from CHCL3
  CE2    C   -0.10000    5
  HE2   HC    0.10000    5
  CZ2    C    0.15000    6
   OH   OA   -0.54800    6
   HH    H    0.39800    6
    C    C     0.380    7
    O    O    -0.380    7
  [ bonds ]
    N    H   gb_2
    N   CA   gb_20
   CA    C   gb_26
    C    O   gb_4
    C   +N   gb_9
   CA   CB   gb_26
   CB   CG   gb_26
   CG  CD1   gb_15
   CG  CD2   gb_15
  CD1  HD1   gb_3
  CD1  CE1   gb_15
  CD2  HD2   gb_3
  CD2  CE2   gb_15
 CE1 CLZ1   gb_37
  CE1  CZ2   gb_15
  CE2  HE2   gb_3
  CE2  CZ2   gb_15
```

```
  CZ2   OH   gb_12
  OH   HH   gb_1
 [ exclusions ]
; ai   aj
  CB   HD1
  CB   HD2
  CB   CE1
  CB   CE2
  CG   HE2
  CG  CLZ1
  CG   CZ2
 CD1   HD2
 CD1   CE2
 CD1    OH
 HD1   CD2
 HD1  CLZ1
 HD1   CZ2
 CD2   CE1
 CD2    OH
 HD2   HE2
 HD2   CZ2
 CE1   HE2
 HE2    OH
CLZ1   CE2
CLZ1    OH
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG  CD1    ga_26
  CB   CG  CD2    ga_26
 CD1   CG  CD2    ga_26
  CG  CD1  HD1    ga_24
 HD1  CD1  CE1    ga_24
  CG  CD1  CE1    ga_26
  CG  CD2  HD2    ga_24
 HD2  CD2  CE2    ga_24
  CG  CD2  CE2    ga_26
 CD1  CE1 CLZ1    ga_26
 CD1  CE1  CZ2    ga_26
CLZ1  CE1  CZ2    ga_26
 CD2  CE2  HE2    ga_24
 HE2  CE2  CZ2    ga_24
 CD2  CE2  CZ2    ga_26
 CE1  CZ2  CE2    ga_26
 CE1  CZ2   OH    ga_26
 CE2  CZ2   OH    ga_26
 CZ2   OH   HH    ga_11
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD1   CG  CD2  CE2   gi_1
  CG  CD1  CE1  CZ2   gi_1
  CG  CD2  CE2  CZ2   gi_1
 CD1  CE1  CZ2  CE2   gi_1
 CD2  CE2  CZ2  CE1   gi_1
 CD1   CG  CE1  HD1   gi_1
 CD2   CG  CE2  HD2   gi_1
 CE1  CZ2  CD1 CLZ1   gi_1
 CE2  CZ2  CD2  HE2   gi_1
 CZ2  CE1  CE2   OH   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG  CD1   gd_20
 CE1  CZ2   OH   HH   gd_2

; homocitrulline
[ KAM ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    2
  NZ    N   -0.28000    3 ; from the peptide bond
  HZ    H    0.28000    3 ; from the peptide bond
  CH    C    0.38000    4 ; from ASN/GLN
 OI1    O   -0.38000    4 ; from ASN/GLN
 NI2   NT   -0.83000    5 ; from ASN/GLN
HI21    H    0.41500    5 ; from ASN/GLN
HI22    H    0.41500    5 ; from ASN/GLN
   C    C      0.380    6
   O    O     -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   CH   gb_9
  CH  OI1   gb_4
  CH  NI2   gb_8
 NI2 HI21   gb_2
 NI2 HI22   gb_2
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ   HZ    ga_17
  CE   NZ   CH    ga_30
  HZ   NZ   CH    ga_31
  NZ   CH  OI1    ga_29
  NZ   CH  NI2    ga_21
 OI1   CH  NI2    ga_29
  CH  NI2 HI21    ga_22
  CH  NI2 HI22    ga_22
HI21  NI2 HI22    ga_23
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
  NZ   CH   CE   HZ   gi_1
  CH  NI2  OI1   NZ   gi_1
 NI2 HI21 HI22   CH   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG   gd_17
  CA   CB   CG   CD   gd_17
  CB   CG   CD   CE   gd_17
  CG   CD   CE   NZ   gd_17
  CD   CE   NZ   CH   gd_19
  CE   NZ   CH  NI2   gd_4
  NZ   CH  NI2 HI21   gd_4

; carboxylysine (-1)
[ KCA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ    N   -0.28000    4 ; from the peptide bond
  HZ    H    0.28000    4 ; from the peptide bond
  CH    C    0.27000    5 ; from the carboxyl group (of e.g., GLU)
 OI1   OM   -0.63500    5 ; from the carboxyl group (of e.g., GLU)
 OI2   OM   -0.63500    5 ; from the carboxyl group (of e.g., GLU)
   C    C      0.380    6
   O    O     -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
```

```
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   CH   gb_9
  CH  OI1   gb_5
  CH  OI2   gb_5
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ   HZ    ga_17
  CE   NZ   CH    ga_30
  HZ   NZ   CH    ga_31
  NZ   CH  OI1    ga_21
  NZ   CH  OI2    ga_21
 OI1   CH  OI2    ga_37
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NZ   CH   CE   HZ    gi_1
  CH   NZ  OI1  OI2    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17
  CG   CD   CE   NZ    gd_17
  CD   CE   NZ   CH    gd_19
  CE   NZ   CH  OI1    gd_4

; carboxylysine (0)
[ KCN ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ    N   -0.28000    4 ; from the peptide bond
  HZ    H    0.28000    4 ; from the peptide bond
  CH    C    0.53000    5 ; from the carboxyl group (of e.g., GLUH)
 OI1    O   -0.38000    5 ; from the carboxyl group (of e.g., GLUH)
 OI2   OA   -0.54800    5 ; from the carboxyl group (of e.g., GLUH)
 HI2    H    0.39800    5 ; from the carboxyl group (of e.g., GLUH)
   C    C    0.380    6
   O    O   -0.380    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE   NZ   gb_20
  NZ   HZ   gb_2
  NZ   CH   gb_9
  CH  OI1   gb_4
  CH  OI2   gb_12
 OI2  HI2   gb_1
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE   NZ    ga_14
  CE   NZ   HZ    ga_17
  CE   NZ   CH    ga_30
  HZ   NZ   CH    ga_31
  NZ   CH  OI1    ga_30
  NZ   CH  OI2    ga_20
 OI1   CH  OI2    ga_30
  CH  OI2  HI2    ga_11
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  NZ   CH   HZ   CE    gi_1
  CH   NZ  OI1  OI2    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17
  CG   CD   CE   NZ    gd_17
  CD   CE   NZ   CH    gd_19
  CE   NZ   CH  OI2    gd_4
  NZ   CH  OI2  HI2    gd_3

; S-carbamoyl-cysteine
[ CAM ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  SG    S    0.00000    1
  CD    C    0.38000    2 ; from ASN/GLN
 OE1    O   -0.38000    2 ; from ASN/GLN
 NE2   NT   -0.83000    3 ; from ASN/GLN
HE21    H    0.41500    3 ; from ASN/GLN
HE22    H    0.41500    3 ; from ASN/GLN
   C    C    0.380    4
   O    O   -0.380    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   SG   gb_30
  SG   CD   gb_30
  CD  OE1   gb_4
  CD  NE2   gb_8
 NE2 HE21   gb_2
 NE2 HE22   gb_2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   SG    ga_15
  CB   SG   CD    ga_3
  SG   CD  OE1    ga_29
  SG   CD  NE2    ga_18
 OE1   CD  NE2    ga_32
  CD  NE2 HE21    ga_22
  CD  NE2 HE22    ga_22
HE21  NE2 HE22    ga_23
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
  CD  NE2  OE1   SG    gi_1
 NE2   CD HE21 HE22    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   SG    gd_17
  CA   CB   SG   CD    gd_13
```

```
  CB   SG   CD  NE2    gd_20
  SG   CD  NE2  HE21    gd_4

; norleucine
[ LNO ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1 ; from aliphatic carbon atoms
  CG  CH2    0.00000    1 ; from aliphatic carbon atoms
  CD  CH2    0.00000    2 ; from aliphatic carbon atoms
  CE  CH3    0.00000    2 ; from aliphatic carbon atoms
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
   C   CA   +N    O   gi_1
  CA    N    C   CB   gi_2
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C   gd_19
   N   CA    C   +N   gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17

; pyruvic acid (N-terminal modification)
[ PYA ]
 [ atoms ]
  CA    C      0.380    0 ; from the carbonyl group (of e.g., GLN)
 OB2    O     -0.380    0 ; from the carbonyl group (of e.g., GLN)
 CB1  CH3    0.00000    1 ; from aliphatic carbon atoms
   C    C      0.380    2
   O    O     -0.380    2
 [ bonds ]
  CA    C   gb_22 ; shorter bond type to account for double bond effect
  CA  OB2   gb_4
   C    O   gb_4
   C   +N   gb_9
  CA  CB1   gb_26
 [ angles ]
; ai   aj   ak  gromos type
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   C   CA  CB1    ga_26
   C   CA  OB2    ga_26
 OB2   CA  CB1    ga_26
 [ impropers ]
; ai   aj   ak   al  gromos type
   C   CA   +N    O   gi_1
  CA  OB2    C  CB1   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 CB1   CA    C   +N    gd_4


; aminoacids.n.tdb file (N-terminal parameters)

; N-methyl-AA (0)
[ 1NM ]
[ replace ]
N                       NT              14.0067         -0.83
                ; from NZ of LYS
CA                      CH1             13.019          0.210           0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1               1               H1              N               CA              C
                H               1.008           0.410
                ; derived from terminal hydrogen atoms of LYS
1               4               CN1             N               CA              C
                CH3             15.035          0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N               H1              gb_2
N               CN1             gb_20
[ angles ]
CN1             N               H1              ga_10
CA              N               H1              ga_10
CA              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N-methyl-glycine (0)
[ GLY-1NM ]
[ replace ]
N                       NT              14.0067         -0.83
                ; from NZ of LYS
CA                      CH2             14.027          0.210           0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1               1               H1              N               CA              C
                H               1.008           0.410
                ; derived from terminal hydrogen atoms of LYS
1               4               CN1             N               CA              C
                CH3             15.035          0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N               H1              gb_2
N               CN1             gb_20
[ angles ]
CN1             N               H1              ga_10
CA              N               H1              ga_10
CA              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N-methyl-proline (0)
[ PRO-1NM ]
[ replace ]
N                       NT              14.0067         -0.63
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA                      CH1             13.019          0.210           0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD                      CH2             14.027          0.210           0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1               4               CN1             N               CA              C
                CH3             15.035          0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N               CN1             gb_20
[ angles ]
CA              N               CN1             ga_12
CD              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N-methyl-AA (+1)
[ 1NM+ ]
[ replace ]
N                       NL              14.0067         0.104
                ; to add up to +1 net charge
CA                      CH1             13.019          0.200           0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
2               4               H               N               CA              C
                H               1.008           0.248
1               1               CN1             N               CA              C
                CH3             15.035          0.200
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
```

```
[ bonds ]
N               H1              gb_2
N               H2              gb_2
N               CN1             gb_20
[ angles ]
H1              N               H2              ga_9
H2              N               CN1             ga_10
CN1             N               H1              ga_10
CA              N               H1              ga_10
CA              N               H2              ga_10
CA              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N-methyl-glycine (+1)
[ GLY-1NM+ ]
[ replace ]
N           NL     14.0067 0.104                                ; to add up to +1 net
charge
CA          CH2    14.027  0.200           0                    ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2     4     H     N     CA    C
      H     1.008  0.248
1     1     CN1   N     CA    C
      CH3   15.035  0.200                                       ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N     H1    gb_2
N     H2    gb_2
N     CN1   gb_20
[ angles ]
H1    N     H2    ga_9
H2    N     CN1   ga_10
CN1   N     H1    ga_10
CA    N     H1    ga_10
CA    N     H2    ga_10
CA    N     CN1   ga_12
[ dihedrals ]
CN1   N     CA    C     gd_14

; N-methyl-proline (+1)
[ PRO-1NM+ ]
[ replace ]
N           NL     14.0067 0.152                                ; to add up to +1 net
charge
CA          CH1    13.019  0.200           0                    ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD          CH2    14.027  0.200           0                    ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1     4     H     N     CA    C
      H     1.008  0.248
1     1     CN1   N     CA    C
      CH3   15.035  0.200                                       ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N     H     gb_2
N     CN1   gb_20
[ angles ]
H     N     CD    ga_10
CD    N     CN1   ga_12
CN1   N     H     ga_10
CA    N     H     ga_10
CA    N     CD    ga_12
CA    N     CN1   ga_12
[ dihedrals ]
CN1   N     CA    C     gd_14

; N,N-dimethyl-AA (0)
[ 2NM ]
[ replace ]
N                           NT              14.0067         -0.63
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA                          CH1             13.019          0.210           0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2               4               CN              N               CA              C
                CH3             15.035          0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N               CN1             gb_20
N               CN2             gb_20
[ angles ]
CN1             N               CN2             ga_12
CA              N               CN2             ga_12
CA              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N,N-dimethyl-glycine (0)
[ GLY-2NM ]
[ replace ]
N           NT     14.0067 -0.63                                ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
CA          CH2    14.027  0.210           0                    ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
2     4     CN    N     CA    C
      CH3   15.035  0.210                                       ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N     CN1   gb_20
N     CN2   gb_20
[ angles ]
CN1   N     CN2   ga_12
CA    N     CN2   ga_12
CA    N     CN1   ga_12
[ dihedrals ]
CN1   N     CA    C     gd_14

; N,N-dimethyl-AA (+1)
[ 2NM+ ]
[ replace ]
N                           NL              14.0067         0.152
                ; to add up to +1 net charge
CA                          CH1             13.019          0.200           0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
1               1               H1              N               CA              C
                H               1.008           0.248
2               4               CN              N               CA              C
                CH3             15.035          0.200
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N               H1              gb_2
N               CN1             gb_20
N               CN2             gb_20
[ angles ]
CN1             N               H1              ga_10
CN1             N               CN2             ga_12
CN2             N               H1              ga_10
CA              N               H1              ga_10
CA              N               CN2             ga_12
CA              N               CN1             ga_12
[ dihedrals ]
CN1             N               CA              C               gd_14

; N,N-dimethyl-glycine (+1)
[ GLY-2NM+ ]
[ replace ]
N           NL     14.0067 0.152                                ; to add up to +1 net
charge
CA          CH2    14.027  0.200           0                    ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1     1     H1    N     CA    C
      H     1.008  0.248
2     4     CN    N     CA    C
      CH3   15.035  0.200                                       ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N     H1    gb_2
N     CN1   gb_20
N     CN2   gb_20
[ angles ]
CN1   N     H1    ga_10
CN1   N     CN2   ga_12
CN2   N     H1    ga_10
CA    N     H1    ga_10
CA    N     CN2   ga_12
CA    N     CN1   ga_12
[ dihedrals ]
CN1   N     CA    C     gd_14
```

```
; N,N-dimethyl-proline (+1)
[ PRO-2NM+ ]
[ replace ]
N       NL    14.0067 0.200                          ; to add up to +1 net
charge
CA      CH1   13.019  0.200          0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD      CH2   14.027  0.200          0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2   4   CN   N   CA   C
    CH3   15.035  0.200                             ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N   CN1   gb_20
N   CN2   gb_20
[ angles ]
CN1   N   CD   ga_12
CN1   N   CN2   ga_12
CN2   N   CD   ga_12
CA   N   CN2   ga_12
CA   N   CN1   ga_12
[ dihedrals ]
CN1   N   CA   C   gd_14

; N,N,N-trimethyl-AA (+1)
[ 3NM+ ]
[ replace ]
N                       NL          14.0067     0.200
                ; to add up to +1 net charge
CA                      CH1         13.019      0.200       0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
3           4           CN          N           CA          C
            CH3         15.035      0.200
            ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N           CN1         gb_20
N           CN2         gb_20
N           CN3         gb_20
[ angles ]
CN1         N           CN3         ga_12
CN1         N           CN2         ga_12
CN2         N           CN3         ga_12
CA          N           CN3         ga_12
CA          N           CN2         ga_12
CA          N           CN1         ga_12
[ dihedrals ]
CN1         N           CA          C           gd_14

; N,N,N-trimethyl-glycine (+1)
[ GLY-3NM+ ]
[ replace ]
N       NL    14.0067 0.200                          ; to add up to +1 net
charge
CA      CH2   14.027  0.200          0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
3   4   CN   N   CA   C
    CH3   15.035  0.200                             ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N   CN1   gb_20
N   CN2   gb_20
N   CN3   gb_20
[ angles ]
CN1   N   CN3   ga_12
CN1   N   CN2   ga_12
CN2   N   CN3   ga_12
CA   N   CN3   ga_12
CA   N   CN2   ga_12
CA   N   CN1   ga_12
[ dihedrals ]
CN1   N   CA   C   gd_14

; N-acetyl-AA
[ NAC ]
[ add ]
1           2           H           N           C           CA
            H           1.008       0.2800
            ; from the peptide bond
1           1           CN1         N           C           CA
            C           12.011      0.2800
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
1           2           ON2         N           CA          C
            O           15.9994     -0.3800
            ; from the carbonyl group (of e.g., GLU)
1           2           CN2         N           C           CA
            CH3         15.035      0.1000
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N           H           gb_2
N           CN1         gb_9
CN1         ON2         gb_4
CN1         CN2         gb_26
[ angles ]
CA          N           H           ga_17
CN1         N           H           ga_31
CN1         N           CA          ga_30
N           CN1         ON2         ga_32
N           CN1         CN2         ga_18
CN2         CN1         ON2         ga_29
[ impropers ]
N           CN1         CA          H           gi_1
CN1         CN2         N           ON2         gi_1
[ dihedrals ]
CN2         CN1         N           CA          gd_4
CN1         N           CA          C           gd_19

; N-acetyl-proline
[ PRO-NAC ]
[ add ]
1           1           CN1         N           C           CA
            C           12.011      0.3800
            ; from the carbonyl group (of e.g., GLU)
1           2           ON2         N           CA          C
            O           15.9994     -0.2800
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
1           2           CN2         N           C           CA
            CH3         15.035      0.1000
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N           CN1         gb_9
CN1         ON2         gb_4
CN1         CN2         gb_26
[ angles ]
CN1         N           CD          ga_30
CN1         N           CA          ga_30
N           CN1         ON2         ga_32
N           CN1         CN2         ga_18
CN2         CN1         ON2         ga_29
[ impropers ]
N           CN1         CA          CD          gi_1
CN1         CN2         N           ON2         gi_1
[ dihedrals ]
CN2         CN1         N           CA          gd_4
CN1         N           CA          C           gd_19

; pyroglutamic acid
[ PGA-NH ]
[ replace ]
N                       N           14.0067     -0.28
[ add ]
1           1           H           N           CA          CD
            H           1.008       0.28
[ bonds ]
N           H           gb_2
[ angles ]
CA          N           H           ga_30
CD          N           H           ga_30
[ impropers ]
N           CA          CD          H           gi_1
[ dihedrals ]
CD          N           CA          C           gd_19

; N-formylmethionine
[ MET-FOR ]
[ add ]
1           2           H           N           C           CA
            H           1.008       0.28
            ; from the peptide bond
1           1           CN1         N           C           CA
            C           12.011      0.2800
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
1           2           ON2         N           CA          C
            O           15.9994     -0.3800
            ; from the carbonyl group (of e.g., GLU)
1           2           HN1         N           C           CA
```

```
            H           1.008       0.1
            ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N           H           gb_2
N           CN1         gb_9
CN1         ON2         gb_4
CN1         HN1         gb_3
[ angles ]
CA          N           H           ga_17
CA          N           CN1         ga_30
H           N           CN1         ga_31
N           CN1         ON2         ga_32
N           CN1         HN1         ga_18
ON2         CN1         HN1         ga_29
[ impropers ]
N           CN1         CA          H           gi_1
CN1         N           ON2         HN1         gi_1
[ dihedrals ]
ON2         CN1         N           CA          gd_4
CN1         N           CA          C           gd_19


; aminoacids.c.tdb file (C-terminal parameters)

; AA-amide
[ AMD ]
[ add ]
1           2           NC1         C           CA          N
            NT          14.00670    -0.8300     ; from ASN/GLN
1           1           HC11        C           O           N
            H           1.008       0.4150
            ; from ASN/GLN
1           1           HC12        C           N           O
            H           1.008       0.4150
            ; from ASN/GLN
[ bonds ]
C           NC1         gb_8
NC1         HC11        gb_2
NC1         HC12        gb_2
[ angles ]
O           C           NC1         ga_32
CA          C           O           ga_29
CA          C           NC1         ga_18
C           NC1         HC11        ga_22
C           NC1         HC12        ga_22
HC11        NC1         HC12        ga_23
[ impropers ]
C           CA          O           NC1         gi_1
NC1         C           HC11        HC12        gi_1
[ dihedrals ]
N           CA          C           NC1         gd_20
CA          C           NC1         HC11        gd_4

; AA-methyl ester
[ CME ]
[ replace ]
C                       C           12.011      0.56
            ; from the carbonyl group (of e.g., GLU) with the partial charge
increased by 0.18 - see the comment of CC2 atom
O                       OA          15.9994     -0.36
            ; from the ring oxygen atom of nucleotides (e.g., ATP)
[ add ]
1           2           OC1         C           CA          N
            O           15.9994     -0.38
            ; from the carbonyl group (of e.g., GLU)
1           2           CC2         O           C           CA
            CH3         15.035      0.18
            ; 0.36 charge divided between CG and CE to add up to 0 net charge,
similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
[ bonds ]
C           O           gb_12
C           OC1         gb_4
O           CC2         gb_17
[ angles ]
O           C           OC1         ga_32
C           O           CC2         ga_11
CA          C           O           ga_18
CA          C           OC1         ga_29
[ dihedrals ]
N           CA          C           O           gd_20
CA          C           O           CC2         gd_3
[ impropers ]
C           CA          OC1         O           gi_1


; aminoacids.hdb file (hydrogen databese)

; phosphoserine (-1)
S1P    2
1           1           H           N           -C          CA
1           2           HE3         OE3         PD          OG
; phosphoserine (-2)
S2P    1
1           1           H           N           -C          CA
; phosphothreonine (-1)
T1P    2
1           1           H           N           -C          CA
1           2           HE3         OE3         PD          OG1
; phosphothreonine (-2)
T2P    1
1           1           H           N           -C          CA
; phosphotyrosine (-1)
Y1P    6
1           1           H           N           -C          CA
1           1           HD1         CD1         CG          CE1
1           1           HD2         CD2         CG          CE2
1           1           HE1         CE1         CD1         CZ
1           1           HE2         CE2         CD2         CZ
1           2           HI3         OI3         PT          OH
; phosphotyrosine (-2)
Y2P    5
1           1           H           N           -C          CA
1           1           HD1         CD1         CG          CE1
1           1           HD2         CD2         CG          CE2
1           1           HE1         CE1         CD1         CZ
1           1           HE2         CE2         CD2         CZ
; phosphoaspartate (-1)
D1P    2
1           1           H           N           -C          CA
1           2           HZ3         OZ3         PE          OD2
; phosphoaspartate (-2)
D2P    1
1           1           H           N           -C          CA
; phospholysine (-1)
K1P    3
1           1           H           N           -C          CA
1           1           HZ          NZ          PH          CE
1           2           HI3         OI3         PH          NZ
; phospholysine (-2)
K2P    2
1           1           H           N           -C          CA
1           1           HZ          NZ          PH          CE
; phosphoarginine (0)
R0P    5
1           1           H           N           -C          CA
1           1           HE          NE          CD          CZ
2           3           HH1         NH1         CZ          NE
1           1           HH2         NH2         PT          CZ
1           2           HI3         OI3         PT          NH2
; phosphoarginine (-1)
R1P    4
1           1           H           N           -C          CA
1           1           HE          NE          CD          CZ
2           3           HH1         NH1         CZ          NE
1           1           HH2         NH2         PT          CZ
; 1-phosphohistidine (-1)
H11    2
1           1           H           N           -C          CA
1           2           HH3         OH3         PZ          NE2
; 1-phosphohistidine (-2)
H12    1
1           1           H           N           -C          CA
; 3-phosphohistidine (-1)
H31    2
1           1           H           N           -C          CA
1           2           HZ3         OZ3         PE3         ND1
; 3-phosphohistidine (-2)
H32    1
1           1           H           N           -C          CA
; N6-methyllysine (0)
KMN    2
1           1           H           N           -C          CA
1           2           HZ          NZ          CE          CD
; N6-methyllysine (+1)
KMC    2
1           1           H           N           -C          CA
2           4           HZ          NZ          CE          CD
; N6,N6-dimethyllysine (0)
K2M    1
1           1           H           N           -C          CA
; N6,N6-dimethyllysine (+1)
K2C    2
1           1           H           N           -C          CA
1           2           HZ          NZ          CE          CD
; N6,N6,N6-trimethyllysine
K3C    1
1           1           H           N           -C          CA
; omega-N-methylarginine (0)
RMN    3
1           1           H           N           -C          CA
1           1           HE          NE          CD          CZ
2           3           HH2         NH2         CZ          NE
; omega-N-methylarginine (+1)
RMC    4
```

1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH1 NH1 CZ NE
1 1 HH2 NH2 CT CZ
; symmetric-dimethylarginine (0)
RSM 3
1 1 H N -C CA
1 1 HE NE CD CZ
1 1 HH2 NH2 CT2 CZ
; symmetric-dimethylarginine (+1)
RMS 4
1 1 H N -C CA
1 1 HE NE CD CZ
1 1 HH1 NH1 CT1 CZ
1 1 HH2 NH2 CT2 CZ
; asymmetric-dimethylarginine (0)
RAM 3
1 1 H N -C CA
1 1 HE NE CD CZ
1 2 HH1 NH1 CZ NE
; asymmetric-dimethylarginine (+1)
RMA 3
1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH1 NH1 CZ NE
; 1-methylhistidine (0)
H1M 1
1 1 H N -C CA
; 1-methylhistidine (+1)
H1C 2
1 1 H N -C CA
1 1 HD1 ND1 CG CE1
; 3-methylhistidine (0)
H3M 1
1 1 H N -C CA
; 3-methylhistidine (+1)
H3C 2
1 1 H N -C CA
1 1 HE2 NE2 CE1 CD2
; N5-methylglutamine
QME 2
1 1 H N -C CA
1 1 HE2 NE2 CZ CD
; N4-methylasparagine
NME 2
1 1 H N -C CA
1 1 HD2 ND2 CE CG
; glutamate methyl ester
EME 1
1 1 H N -C CA
; aspartate methyl ester
DMA 1
1 1 H N -C CA
; S-methylcysteine
CYM 1
1 1 H N -C CA
; N6-acetyllysine
KAC 2
1 1 H N -C CA
1 1 HZ NZ CE CH
; 3-hydroxyproline (R)
PH3 1
1 2 HG1 OG1 CB CA
; 3-hydroxyproline (S)
P3H 1
1 2 HG1 OG1 CB CA
; 4-hydroxyproline (S)
HY2 1
1 2 HD1 OD1 CG CB
; 3,4-dihydroxyproline
PHH 2
1 2 HG1 OG1 CB CA
1 2 HD1 OD1 CG2 CB
; 5-hydroxylysine (0,R)
KH5 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
2 4 HZ NZ CE2 CD
; 5-hydroxylysine (0,S)
K5H 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
2 4 HZ NZ CE2 CD
; 5-hydroxylysine (+1,R)
KPH 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
3 4 HZ NZ CE2 CD
; 5-hydroxylysine (+1,S)
KHP 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
3 4 HZ NZ CE2 CD
; 3,4-dihydroxyphenylalanine
HTY 6
1 1 H N -C CA
1 1 HD1 CD1 CG CE1
1 1 HD2 CD2 CG CE2
1 2 HZ1 OZ1 CE1 CD1
1 1 HE2 CE2 CD2 CZ2
1 2 HH OH CZ2 CE1
; 7-hydroxytryptophan
W7H 7
1 1 H N -C CA
1 1 HD1 CD1 CG NE1
1 1 HE1 NE1 CD1 CE2
1 1 HE3 CE3 CD2 CZ3
1 1 HZ3 CZ3 CE3 CH3
1 1 HH3 CH3 CZ3 CZ2
1 2 HH2 OH2 CZ2 CE2
; 3-hydroxyaspartate (-1,R)
DH3 2
1 1 H N -C CA
1 2 HG1 OG1 CB CA
; 3-hydroxyaspartate (-1,S)
D3H 2
1 1 H N -C CA
1 2 HG1 OG1 CB CA
; 3-hydroxyaspartate (0,R)
DN3 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
1 2 HD2 OD2 CG2 CB
; 3-hydroxyaspartate (0,S)
D3N 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
1 2 HD2 OD2 CG2 CB
; 3-hydroxyasparagine (R)
N3H 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
2 3 HD2 ND2 CG2 CB
; 3-hydroxyasparagine (S)
NH3 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
2 3 HD2 ND2 CG2 CB
; 4-carboxyglutamate (-2)
ECA 1
1 1 H N -C CA
; 4-carboxyglutamate (-1)
ECN 2
1 1 H N -C CA
1 2 HE2 OE2 CD1 CG
; sulfotyrosine
YSU 5
1 1 H N -C CA
1 1 HD1 CD1 CG CE1
1 1 HD2 CD2 CG CE2
1 1 HE1 CE1 CD1 CZ
1 1 HE2 CE2 CD2 CZ
; dehydroalanine
SDH 1
1 1 H N -C CA
; 2,3-didehydrobutyrine
TDH 1
1 1 H N -C CA
; 6-bromotryptophan
WBR 6
1 1 H N -C CA
1 1 HD1 CD1 CG NE1
1 1 HE1 NE1 CD1 CE2
1 1 HE3 CE3 CD2 CZ3
1 1 HZ3 CZ3 CE3 CH2
1 1 HZ2 CZ2 CE2 CH2
; S-nitrosocysteine
CSN 1
1 1 H N -C CA
; citrulline
RCI 3
1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH2 NH2 CZ NE
; allysine (aminoadipic semialdehyde)
KAL 2
1 1 H N -C CA
1 1 HE CE OZ CD
; N-acetylglucosamine (N4-linked to ASN)
NNG 6
1 1 H N -C CA
1 1 HD2 ND2 CG C1
1 1 HN2 N2 C7 C2
1 2 HO3 O3 C3 C2
1 2 HO4 O4 C4 C3
1 2 HO6 O6 C6 C5

```
; 2,3-dihydroxyphenylalanine
F23    6
1      1      H      N      -C     CA
1      2      HE3    OE3    CD1    CG
1      1      HD2    CD2    CG     CE2
1      2      HZ1    OZ1    CE1    CD1
1      1      HE2    CE2    CD2    CZ2
1      1      HZ2    CZ2    CE1    CE2
; 2-hydroxyphenylalanine
F2H    6
1      1      H      N      -C     CA
1      2      HE3    OE3    CD1    CG
1      1      HD2    CD2    CG     CE2
1      1      HE1    CE1    CD1    CZ
1      1      HE2    CE2    CD2    CZ
1      1      HZ     CZ     CE1    CE2
; 3-hydroxyphenylalanine
F3H    6
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      2      HZ1    OZ1    CE1    CD1
1      1      HE2    CE2    CD2    CZ2
1      1      HZ2    CZ2    CE1    CE2
; 6-hydroxytryptophan
W6H    7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      2      HI     OI     CH2    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 5-hydroxytryptophan
W5H    7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      2      HH3    OH3    CZ3    CE2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 4-hydroxytryptophan
W4H    7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      2      HZ4    OZ4    CE3    CD2
1      1      HZ3    CZ3    CE3    CH2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 2-hydroxytryptophan
W2H    7
1      1      H      N      -C     CA
1      2      HE4    OE4    CD1    CG
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 3-hydroxyleucine (R)
L3H    2
1      1      H      N      -C     CA
1      2      HG1    OG1    CB     CA
; 3-hydroxyleucine (S)
LH3    2
1      1      H      N      -C     CA
1      2      HG1    OG1    CB     CA
; 4-hydroxyleucine
L4H    2
1      1      H      N      -C     CA
1      2      HD3    OD3    CG     CB
; 5-hydroxyleucine (R)
L5H    2
1      1      H      N      -C     CA
1      2      HE     OE     CD2    CG
; 5-hydroxyleucine (S)
LH5    2
1      1      H      N      -C     CA
1      2      HE     OE     CD2    CG
; 3-hyroxyvaline
V3H    2
1      1      H      N      -C     CA
1      2      HG3    OG3    CB     CA
; cysteine sulfenic acid
CYH    2
1      1      H      N      -C     CA
1      2      HD     OD     SG     CB
; 5-hydroxyproline (R)
PH5    1
1      2      HE     OE     CD     CG
; 5-hydroxyproline (S)
P5H    1
1      2      HE     OE     CD     CG
; glutamic semialdehyde
GSA    2
1      1      H      N      -C     CA
1      1      HD     CD     OE     CG
; 2-amino-3-ketobutyric acid
TOX    1
1      1      H      N      -C     CA
; 2-oxo-histidine
H2X    3
1      1      H      N      -C     CA
1      1      HD1    ND1    CG     CE1
1      1      HE2    NE2    CE1    CD2
; methionine sulfoxide (R)
MSX    1
1      1      H      N      -C     CA
; methionine sulfoxide (S)
MXS    1
1      1      H      N      -C     CA
; methionine sulfone
MES    1
1      1      H      N      -C     CA
; cysteine sulfinic acid
CSA    1
1      1      H      N      -C     CA
; cysteic acid
CSE    1
1      1      H      N      -C     CA
; 3-nitrotyrosine (-1)
YNI    4
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
; 3-nitrotyrosine (0)
YNN    5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
1      2      HH3    OH3    CZ2    CE1
; 3-nitrotyrosine (0)
YNB    5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
1      2      HH3    OH3    CZ2    CE1
; 6-nitrotryptophan
WNI    6
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      1      HZ2    CZ2    CE2    CH2
; kynurenine
WKY    6
1      1      H      N      -C     CA
1      1      HE2    CE2    CD2    CZ3
2      3      HZ1    NZ1    CE1    CD2
1      1      HZ2    CZ2    CE1    CH
1      1      HZ3    CZ3    CE2    CH
1      1      HH     CH     CZ2    CZ3
; 3-hydroxykynurenine
WKH    6
1      1      H      N      -C     CA
1      1      HE2    CE2    CD2    CZ3
2      3      HZ1    NZ1    CE1    CD2
1      1      HZ3    CZ3    CE2    CH2
1      2      HH1    OH1    CZ2    CE1
1      1      HH2    CH2    CZ2    CZ3
; formyl-kynurenine
WKF    7
1      1      H      N      -C     CA
1      1      HE2    CE2    CD2    CZ3
1      1      HZ1    NZ1    CE1    CH1
1      1      HZ2    CZ2    CE1    CH2
1      1      HZ3    CZ3    CE2    CH2
1      1      HH1    CH1    NZ1    OI
1      1      HH2    CH2    CZ2    CZ3
; chlorotyrosine
YCH    5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
1      2      HH     OH     CZ2    CE1
; homocitrulline
KAM    3
1      1      H      N      -C     CA
1      1      HZ     NZ     CH     CE
2      4      HI2    NI2    CH     NZ
```

```
; carboxylysine (-1)
KCA    2
1              1              H              N              -C             CA
1              1              HZ             NZ             CH             CE
; carboxylysine (0)
KCN    3
1              1              H              N              -C             CA
1              1              HZ             NZ             CH             CE
1              2              HI2            OI2            CH             NZ
; S-carbamoyl-cysteine
CAM    2
1              1              H              N              -C             CA
2              3              HE2            NE2            CD             SG
; norleucine
LNO    1
1              1              H              N              -C             CA
```

Dataset S2. Force field parameters for the GROMOS force field 54A7 parameter set.

```
; This file contains extended force field parameters for the GROMOS force field 54A7
parameters set
; GROMACS 4.5.x format - files: aminoacids.rtp, aminoacids.n.tdb, aminoacids.c.tdb
and aminoacids.hdb
; Authors: Drazen Petrov, Christian Margreitter, Melanie Grandits, Chris Oostenbrink
& Bojan Zagrovic
; Parameter files in GROMACS 4.3.x and 4.4.x, and GROMOS formats available at
http://vienna-ptm.univie.ac.at/


; aminoacids.rtp file (backbone and side chain parameters)

[ bondedtypes ]
; bonds  angles  dihedrals  impropers
   2      2        1        2

; phosphoserine (-1)
[ S1P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OG    OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.63000    2 ; to add up to -1 net charge
 OE1    OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE2    OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3    OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HE3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C    0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   OG   gb_18
  OG   PD   gb_28
  PD   OE1   gb_24
  PD   OE2   gb_24
  PD   OE3   gb_28
 OE3   HE3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai    aj
 OE1   HE3
 OE2   HE3
  OG   HE3
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N     H    ga_32
  -C    N    CA    ga_31
   H    N    CA    ga_18
   N   CA    CB    ga_13
   N   CA     C    ga_13
  CB   CA     C    ga_13
  CA   CB    OG    ga_13
  CB   OG    PD    ga_26
  OG   PD   OE1    ga_14
  OG   PD   OE2    ga_14
  OG   PD   OE3    ga_5
 OE1   PD   OE2    ga_29
 OE1   PD   OE3    ga_14
 OE2   PD   OE3    ga_14
  PD   OE3  HE3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C    CA    H    gi_1
  CA    N    C    CB    gi_2
   C   CA    +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA   -C     N    CA    gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
   N   CA    CB    OG    gd_34
   N   CA    C     +N    gd_45
   N   CA    C     +N    gd_42
  CA   CB    OG    PD    gd_23
  CB   OG    PD   OE3    gd_19
  CB   OG    PD   OE3    gd_22
  OG   PD   OE3  HE3    gd_19
  OG   PD   OE3  HE3    gd_22

; phosphoserine (-2)
[ S2P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OG    OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.11500    2 ; to add up to -2 net charge
 OE1    OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE2    OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3    OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C    0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   OG   gb_18
  OG   PD   gb_28
  PD   OE1   gb_24
  PD   OE2   gb_24
  PD   OE3   gb_24
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N     H    ga_32
  -C    N    CA    ga_31
   H    N    CA    ga_18
   N   CA    CB    ga_13
   N   CA     C    ga_13
  CB   CA     C    ga_13
  CA   CB    OG    ga_13
  CB   OG    PD    ga_26
  OG   PD   OE1    ga_14
  OG   PD   OE2    ga_14
  OG   PD   OE3    ga_14
 OE1   PD   OE2    ga_14
 OE1   PD   OE3    ga_14
 OE2   PD   OE3    ga_14
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C    CA    H    gi_1
  CA    N    C    CB    gi_2
   C   CA    +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA   -C     N    CA    gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
   N   CA    CB    OG    gd_34
   N   CA    C     +N    gd_45
   N   CA    C     +N    gd_42
  CA   CB    OG    PD    gd_23
  CB   OG    PD   OE1    gd_19
  CB   OG    PD   OE1    gd_22

; phosphothreonine (-1)
[ T1P ]
```

```
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
 OG1   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.63000    2 ; to add up to -1 net charge
 OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3   OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HE3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
 CG2  CH3    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1   PD   gb_28
  PD  OE1   gb_24
  PD  OE2   gb_24
  PD  OE3   gb_28
 OE3  HE3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
 OE1  HE3
 OE2  HE3
 OG1  HE3
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_15
 OG1   CB  CG2    ga_15
  CB  OG1   PD    ga_26
 OG1   PD  OE1    ga_14
 OG1   PD  OE2    ga_14
 OG1   PD  OE3    ga_5
 OE1   PD  OE2    ga_29
 OE1   PD  OE3    ga_14
 OE2   PD  OE3    ga_14
  PD  OE3  HE3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB  OG1    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB  OG1   PD    gd_23
  CB  OG1   PD  OE3    gd_19
  CB  OG1   PD  OE3    gd_22
 OG1   PD  OE3  HE3    gd_19
 OG1   PD  OE3  HE3    gd_22

; phosphothreonine (-2)
[ T2P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH1    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
 OG1   OA   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
  PD    P    0.11500    2 ; to add up to -2 net charge
 OE1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 CG2  CH3    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1   PD   gb_28
  PD  OE1   gb_24
  PD  OE2   gb_24
  PD  OE3   gb_24
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_15
 OG1   CB  CG2    ga_15
  CB  OG1   PD    ga_26
 OG1   PD  OE1    ga_14
 OG1   PD  OE2    ga_14
 OG1   PD  OE3    ga_14
 OE1   PD  OE2    ga_14
 OE1   PD  OE3    ga_14
 OE2   PD  OE3    ga_14
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB  OG1    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB  OG1   PD    gd_23
  CB  OG1   PD  OE1    gd_19
  CB  OG1   PD  OE1    gd_22

; phosphotyrosine (-1)
[ Y1P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C   -0.14000    4
 HE1   HC    0.14000    4
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
  CZ    C    0.15000    6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OH   OA   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
  PT    P    0.63000    6 ; to add up to -1 net charge
 OI1   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OA   -0.54800    6 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HI3    H    0.39800    6 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  HE1   gb_3
 CE1   CZ   gb_16
 CE2  HE2   gb_3
 CE2   CZ   gb_16
```

```
  CZ   OH   gb_13
  OH   PT   gb_28
  PT   OI1  gb_24
  PT   OI2  gb_24
  PT   OI3  gb_28
  OI3  HI3  gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB   HD1
  CB   HD2
  CB   CE1
  CB   CE2
  CG   HE1
  CG   HE2
  CG   CZ
  CD1  HD2
  CD1  CE2
  CD1  OH
  HD1  CD2
  HD1  HE1
  HD1  CZ
  CD2  CE1
  CD2  OH
  HD2  HE2
  HD2  CZ
  CE1  HE2
  HE1  CE2
  HE1  OH
  HE2  OH
  OH   HI3
  OI1  HI3
  OI2  HI3
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_32
  -C   N   CA   ga_31
   H   N   CA   ga_18
   N   CA  CB   ga_13
   N   CA  C    ga_13
   CB  CA  C    ga_13
   CA  CB  CG   ga_15
   CB  CG  CD1  ga_27
   CB  CG  CD2  ga_27
   CD1 CG  CD2  ga_27
   CG  CD1 HD1  ga_25
   CG  CD1 CE1  ga_27
   HD1 CD1 CE1  ga_25
   CG  CD2 HD2  ga_25
   CG  CD2 CE2  ga_27
   HD2 CD2 CE2  ga_25
   CD1 CE1 HE1  ga_25
   CD1 CE1 CZ   ga_27
   HE1 CE1 CZ   ga_25
   CD2 CE2 HE2  ga_25
   CD2 CE2 CZ   ga_27
   HE2 CE2 CZ   ga_25
   CE1 CZ  CE2  ga_27
   CE1 CZ  OH   ga_27
   CE2 CZ  OH   ga_27
   CZ  OH  PT   ga_26
   OH  PT  OI1  ga_14
   OH  PT  OI2  ga_14
   OH  PT  OI3  ga_5
   OI1 PT  OI2  ga_29
   OI1 PT  OI3  ga_14
   OI2 PT  OI3  ga_14
   PT  OI3 HI3  ga_12
   CA  C   O    ga_30
   CA  C   +N   ga_19
   O   C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
   CA  N    C    CB  gi_2
   CG  CD1  CD2  CB  gi_1
   CG  CD1  CE1  CZ  gi_1
   CG  CD2  CE2  CZ  gi_1
   CD1 CG   CD2  CE2 gi_1
   CD1 CG   CE1  HD1 gi_1
   CD1 CE1  CZ   CE2 gi_1
   CD2 CG   CD1  CE1 gi_1
   CD2 CG   CE2  HD2 gi_1
   CD2 CE2  CZ   CE1 gi_1
   HE1 CD1  CZ   CE1 gi_1
   HE2 CD2  CZ   CE2 gi_1
   CZ  CE1  CE2  OH  gi_1
   C   CA   +N   O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA  -C   N    CA  gd_14
  -C   N    CA   C   gd_44
  -C   N    CA   C   gd_43
   N   CA   CB   CG  gd_34
   N   CA   C    +N  gd_45
   N   CA   C    +N  gd_42
   CA  CB   CG   CD1 gd_40
   CE1 CZ   OH   PT  gd_11
   CZ  OH   PT   OI3 gd_19
   CZ  OH   PT   OI3 gd_22
   OH  PT   OI3  HI3 gd_19
   OH  PT   OI3  HI3 gd_22

; phosphotyrosine (-2)
[ Y2P ]
 [ atoms ]
   N    N   -0.31000   0
   H    H    0.31000   0
   CA   CH1   0.00000   1
   CB   CH2   0.00000   1
   CG   C     0.00000   1
   CD1  C    -0.14000   2
   HD1  HC    0.14000   2
   CD2  C    -0.14000   3
   HD2  HC    0.14000   3
   CE1  C    -0.14000   4
   HE1  HC    0.14000   4
   CE2  C    -0.14000   5
   HE2  HC    0.14000   5
   CZ   C     0.15000   6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
   OH   OA   -0.36000   6 ; from the phosphate group of nucleotides (e.g., ATP)
   PT   P     0.11500   6 ; to add up to -2 net charge
   OI1  OM   -0.63500   6 ; from the phosphate group of nucleotides (e.g., ATP)
   OI2  OM   -0.63500   6 ; from the phosphate group of nucleotides (e.g., ATP)
   OI3  OM   -0.63500   6 ; from the phosphate group of nucleotides (e.g., ATP)
    C   C      0.450   7
    O   O     -0.450   7
 [ bonds ]
   N   H   gb_2
   N   CA  gb_21
   CA  CB  gb_27
   CA  C   gb_27
   CB  CG  gb_27
   CG  CD1 gb_16
   CG  CD2 gb_16
   CD1 HD1 gb_3
   CD1 CE1 gb_16
   CD2 HD2 gb_3
   CD2 CE2 gb_16
   CE1 HE1 gb_3
   CE1 CZ  gb_16
   CE2 HE2 gb_3
   CE2 CZ  gb_16
   CZ  OH  gb_13
   OH  PT  gb_28
   PT  OI1 gb_24
   PT  OI2 gb_24
   PT  OI3 gb_24
   C   O   gb_5
   C   +N  gb_10
 [ exclusions ]
; ai   aj
  CB   HD1
  CB   HD2
  CB   CE1
  CB   CE2
  CG   HE1
  CG   HE2
  CG   CZ
  CD1  HD2
  CD1  CE2
  CD1  OH
  HD1  CD2
  HD1  HE1
  HD1  CZ
  CD2  CE1
  CD2  OH
  HD2  HE2
  HD2  CZ
  CE1  HE2
  HE1  CE2
  HE1  OH
  HE2  OH
 [ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_32
  -C   N   CA   ga_31
   H   N   CA   ga_18
   N   CA  CB   ga_13
   N   CA  C    ga_13
   CB  CA  C    ga_13
   CA  CB  CG   ga_15
   CB  CG  CD1  ga_27
   CB  CG  CD2  ga_27
```

```
  CD1   CG  CD2    ga_27
  CG  CD1  HD1    ga_25
  CG  CD1  CE1    ga_27
  HD1  CD1  CE1    ga_25
  CG  CD2  HD2    ga_25
  CG  CD2  CE2    ga_27
  HD2  CD2  CE2    ga_25
  CD1  CE1  HE1    ga_25
  CD1  CE1   CZ    ga_27
  HE1  CE1   CZ    ga_25
  CD2  CE2  HE2    ga_25
  CD2  CE2   CZ    ga_27
  HE2  CE2   CZ    ga_25
  CE1   CZ  CE2    ga_27
  CE1   CZ   OH    ga_27
  CE2   CZ   OH    ga_27
   CZ   OH   PT    ga_26
   OH   PT  OI1    ga_14
   OH   PT  OI2    ga_14
   OH   PT  OI3    ga_14
  OI1   PT  OI2    ga_14
  OI1   PT  OI3    ga_14
  OI2   PT  OI3    ga_14
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H    gi_1
   CA    N    C   CB    gi_2
   CG  CD1  CD2   CB    gi_1
   CG  CD1  CE1   CZ    gi_1
   CG  CD2  CE2   CZ    gi_1
  CD1   CG  CD2  CE2    gi_1
  CD1   CG  CE1  HD1    gi_1
  CD1  CE1   CZ  CE2    gi_1
  CD2   CG  CD1  CE1    gi_1
  CD2   CG  CE2  HD2    gi_1
  CD2  CE2   CZ  CE1    gi_1
  HE1  CD1   CZ  CE1    gi_1
  HE2  CD2   CZ  CE2    gi_1
   CZ  CE1  CE2   OH    gi_1
    C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N    CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   CG  CD1    gd_40
  CE1   CZ   OH   PT    gd_11
   CZ   OH   PT  OI1    gd_19
   CZ   OH   PT  OI1    gd_22

; phosphoaspartate (-1)
[ D1P ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.30900    2 ; to add up to -1 net charge
  OD1    O   -0.45000    2 ; from the carbonyl oxygen (of e.g., the peptide bond)
  OD2   OE   -0.06900    2 ; from the ester oxygen reported by Horta et al. DOI:
10.1021/ct1006407
   PE    P    0.63000    2 ; from the phosphate group (of e.g., S1P)
  OZ1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ3   OA   -0.54800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HZ3    H    0.39800    2 ; from the hydroxyl group of nucleotides (e.g., ATP)
    C    C      0.450    3
    O    O     -0.450    3
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG  OD1   gb_5
   CG  OD2   gb_13
  OD2   PE   gb_28
   PE  OZ1   gb_24
   PE  OZ2   gb_24
   PE  OZ3   gb_28
  OZ3  HZ3   gb_1
    C    O   gb_5
    C   +N   gb_10
 [ exclusions ]
;  ai    aj
  OD2  HZ3
  OZ1  HZ3
  OZ2  HZ3
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG  OD1    ga_30
   CB   CG  OD2    ga_19
  OD1   CG  OD2    ga_33
   CG  OD2   PE    ga_26
  OD2   PE  OZ1    ga_14
  OD2   PE  OZ2    ga_14
  OD2   PE  OZ3    ga_5
  OZ1   PE  OZ2    ga_29
  OZ1   PE  OZ3    ga_14
  OZ2   PE  OZ3    ga_14
   PE  OZ3  HZ3    ga_12
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H    gi_1
   CA    N    C   CB    gi_2
   CG  OD1  OD2   CB    gi_1
    C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C    N    CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   CG  OD2    gd_40
   CB   CG  OD2   PE    gd_12
   CG  OD2   PE  OZ3    gd_19
   CG  OD2   PE  OZ3    gd_22
  OD2   PE  OZ3  HZ3    gd_19
  OD2   PE  OZ3  HZ3    gd_22

; phosphoaspartate (-2)
[ D2P ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.30900    2 ; to add up to -2 net charge
  OD1    O   -0.45000    2 ; from the carbonyl oxygen (of e.g., the peptide bond)
  OD2   OE   -0.06900    2 ; from the ester oxygen reported by Horta et al. DOI:
10.1021/ct1006407
   PE    P    0.11500    2 ; from the phosphate group (of e.g., S2P)
  OZ1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
    C    C      0.450    3
    O    O     -0.450    3
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG  OD1   gb_5
   CG  OD2   gb_13
  OD2   PE   gb_28
   PE  OZ1   gb_24
   PE  OZ2   gb_24
   PE  OZ3   gb_24
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG  OD1    ga_30
   CB   CG  OD2    ga_19
  OD1   CG  OD2    ga_33
   CG  OD2   PE    ga_26
  OD2   PE  OZ1    ga_14
  OD2   PE  OZ2    ga_14
  OD2   PE  OZ3    ga_14
  OZ1   PE  OZ2    ga_14
```

```
  OZ1   PE   OZ3    ga_14
  OZ2   PE   OZ3    ga_14
  CA    C    O     ga_30
  CA    C   +N     ga_19
   O    C   +N     ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H     gi_1
   CA    N    C    CB     gi_2
   CG   OD1   OD2   CB     gi_1
    C   CA   +N    O     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C     N    CA     gd_14
   -C    N    CA    C     gd_44
   -C    N    CA    C     gd_43
    N   CA    CB    CG     gd_34
    N   CA    C    +N     gd_45
    N   CA    C    +N     gd_42
   CA   CB    CG   OD2     gd_40
   CB   CG   OD2   PE     gd_12
   CG   OD2   PE   OZ1     gd_19
   CG   OD2   PE   OZ1     gd_22

; phospholysine (-1)
[ K1P ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG   CH2    0.00000    2
   CD   CH2    0.00000    2
   CE   CH2    0.00000    3
   NZ    NE   -0.31000    4 ; from NE of ARGN
   HZ    H    0.31000    4 ; from HE of ARGN
   PH    P    0.42000    5 ; to add up to -1 net charge
  OI1   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI2   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI3   OA   -0.54800    5 ; from the hydroxyl group of nucleotides (e.g., ATP)
  HI3    H    0.39800    5 ; from the hydroxyl group of nucleotides (e.g., ATP)
    C    C     0.450    6
    O    O    -0.450    6
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ   HZ   gb_2
   NZ   PH   gb_24
   PH  OI1   gb_24
   PH  OI2   gb_24
   PH  OI3   gb_28
  OI3  HI3   gb_1
    C    O   gb_5
    C   +N   gb_10
 [ exclusions ]
;  ai    aj
   HZ  OI1
   HZ  OI2
   HZ  OI3
   NZ  HI3
  OI1  HI3
  OI2  HI3
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
   CG   CD   CE    ga_15
   CD   CE   NZ    ga_15
   CE   NZ   HZ    ga_20
   HZ   NZ   PH    ga_23
   CE   NZ   PH    ga_33
   NZ   PH  OI1    ga_14
   NZ   PH  OI2    ga_14
   NZ   PH  OI3    ga_5
  OI1   PH  OI2    ga_29
  OI1   PH  OI3    ga_14
  OI2   PH  OI3    ga_14
  HI3  OI3   PH    ga_12
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H     gi_1
   CA    N    C    CB     gi_2
    C   CA   +N    O     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C     N    CA     gd_14
   -C    N    CA    C     gd_44
   -C    N    CA    C     gd_43
    N   CA    CB    CG     gd_34
    N   CA    C    +N     gd_45
    N   CA    C    +N     gd_42
   CA   CB    CG   CD     gd_34
   CB   CG    CD   CE     gd_34
   CG   CD    CE   NZ     gd_34
   CD   CE    NZ   PH     gd_39
   CE   NZ    PH  OI3     gd_39
   NZ   PH   OI3  HI3     gd_19
   NZ   PH   OI3  HI3     gd_22

; phospholysine (-2)
[ K2P ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG   CH2    0.00000    2
   CD   CH2    0.00000    2
   CE   CH2    0.00000    3
   NZ    NE   -0.31000    4 ; from NE of ARGN
   HZ    H    0.31000    4 ; from HE of ARGN
   PH    P   -0.09500    5 ; to add up to -2 net charge
  OI1   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI2   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
  OI3   OM   -0.63500    5 ; from the phosphate group of nucleotides (e.g., ATP)
    C    C     0.450    6
    O    O    -0.450    6
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ   HZ   gb_2
   NZ   PH   gb_24
   PH  OI1   gb_24
   PH  OI2   gb_24
   PH  OI3   gb_24
    C    O   gb_5
    C   +N   gb_10
 [ exclusions ]
;  ai    aj
   HZ  OI1
   HZ  OI2
   HZ  OI3
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
   CG   CD   CE    ga_15
   CD   CE   NZ    ga_15
   CE   NZ   HZ    ga_20
   HZ   NZ   PH    ga_23
   CE   NZ   PH    ga_33
   NZ   PH  OI1    ga_14
   NZ   PH  OI2    ga_14
   NZ   PH  OI3    ga_14
  OI1   PH  OI2    ga_14
  OI1   PH  OI3    ga_14
  OI2   PH  OI3    ga_14
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA    H     gi_1
   CA    N    C    CB     gi_2
    C   CA   +N    O     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C     N    CA     gd_14
   -C    N    CA    C     gd_44
```

```
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ   PH    gd_39
  CE   NZ   PH  OI1    gd_39

; phosphoarginine (0)
[ R0P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.43000    2 ; to add up to 1 net charge
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NE   -0.11000    2 ; from NE nitrogen atom of ARG
 HH2    H    0.24000    2 ; from terminal hydrogen atoms of ARG
  PT    P    0.42000    3 ; to add up to -1 net charge
 OI1   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OA   -0.54800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HI3    H    0.39800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C    0.450    4
   O    O   -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   NE   gb_21
  NE   HE   gb_2
  NE   CZ   gb_11
  CZ  NH1   gb_11
  CZ  NH2   gb_11
 NH1  HH11   gb_2
 NH1  HH12   gb_2
 NH2   HH2   gb_2
 NH2   PT   gb_24
  PT  OI1   gb_24
  PT  OI2   gb_24
  PT  OI3   gb_28
 OI3  HI3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
 HH2  OI1
 HH2  OI2
 HH2  OI3
 NH2  HI3
 OI1  HI3
 OI2  HI3
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   NE    ga_13
  CD   NE   HE    ga_20
  CD   NE   CZ    ga_33
  HE   NE   CZ    ga_23
  NE   CZ  NH1    ga_28
  NE   CZ  NH2    ga_28
 NH1   CZ  NH2    ga_28
  CZ  NH1 HH11    ga_23
  CZ  NH1 HH12    ga_23
HH11  NH1 HH12    ga_24
  CZ  NH2  HH2    ga_23
 HH2  NH2   PT    ga_20
  CZ  NH2   PT    ga_33
 NH2   PT  OI1    ga_14
 NH2   PT  OI2    ga_14
 NH2   PT  OI3    ga_5
 OI1   PT  OI2    ga_29
 OI1   PT  OI3    ga_14
 OI2   PT  OI3    ga_14
  PT  OI3  HI3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1 HH11 HH12   CZ    gi_1
 NH2   PT   CZ  HH2    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   NE    gd_34
  CG   CD   NE   CZ    gd_39
  CD   NE   CZ  NH2    gd_14
  NE   CZ  NH1 HH11    gd_14
  NE   CZ  NH2   PT    gd_14
  CZ  NH2   PT  OI3    gd_39
 NH2   PT  OI3  HI3    gd_19
 NH2   PT  OI3  HI3    gd_22

; phosphoarginine (-1)
[ R1P ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.43000    2 ; to add up to 1 net charge
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NE   -0.11000    2 ; from NE nitrogen atom of ARG
 HH2    H    0.24000    2 ; from terminal hydrogen atoms of ARG
  PT    P   -0.09500    3 ; to add up to -2 net charge
 OI1   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C    0.450    4
   O    O   -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   NE   gb_21
  NE   HE   gb_2
  NE   CZ   gb_11
  CZ  NH1   gb_11
  CZ  NH2   gb_11
 NH1  HH11   gb_2
 NH1  HH12   gb_2
 NH2   HH2   gb_2
 NH2   PT   gb_24
  PT  OI1   gb_24
  PT  OI2   gb_24
  PT  OI3   gb_24
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
 HH2  OI1
 HH2  OI2
 HH2  OI3
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   NE    ga_13
  CD   NE   HE    ga_20
  CD   NE   CZ    ga_33
```

```
  HE   NE   CZ    ga_23
  NE   CZ  NH1    ga_28
  NE   CZ  NH2    ga_28
 NH1   CZ  NH2    ga_28
  CZ  NH1 HH11    ga_23
  CZ  NH1 HH12    ga_23
HH11  NH1 HH12    ga_24
  CZ  NH2  HH2    ga_23
 HH2  NH2   PT    ga_20
  CZ  NH2   PT    ga_33
 NH2   PT  OI1    ga_14
 NH2   PT  OI2    ga_14
 NH2   PT  OI3    ga_14
 OI1   PT  OI2    ga_14
 OI1   PT  OI3    ga_14
 OI2   PT  OI3    ga_14
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1 HH11 HH12   CZ    gi_1
 NH2   PT   CZ  HH2    gi_1
   C   CA   +N    O    gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   NE    gd_34
  CG   CD   NE   CZ    gd_39
  CD   NE   CZ  NH2    gd_14
  NE   CZ  NH1 HH11    gd_14
  NE   CZ  NH2   PT    gd_14
  CZ  NH2   PT  OI1    gd_39

; 1-phosphohistidine (-1)
[ H11 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR   -0.54000    2
 CD2    C    0.13000    2
 HD2   HC    0.14000    2
 CE1    C    0.13000    2
 HE1   HC    0.14000    2
 NE2   NR    0.00000    2
  PZ    P    0.42000    3 ; to add up to -1 net charge
 OH1   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OH2   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OH3   OA   -0.54800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HH3    H    0.39800    3 ; from the hydroxyl group of nucleotides (e.g., ATP)
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  CE1   gb_10
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 CE1  HE1   gb_3
 CE1  NE2   gb_10
 NE2   PZ   gb_24
  PZ  OH1   gb_24
  PZ  OH2   gb_24
  PZ  OH3   gb_28
 OH3  HH3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE1
  CG   PZ
 ND1  HD2
 ND1   PZ
 CD2  HE1
 HD2  CE1
 HD2   PZ
 HE1   PZ
 HH3  NE2
 HH3  OH1
 HH3  OH2
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  ND1    ga_37
  CB   CG  CD2    ga_37
 ND1   CG  CD2     ga_7
  CG  ND1  CE1     ga_7
  CG  CD2  HD2    ga_36
  CG  CD2  NE2     ga_7
 HD2  CD2  NE2    ga_36
 ND1  CE1  HE1    ga_36
 ND1  CE1  NE2     ga_7
 HE1  CE1  NE2    ga_36
 CD2  NE2  CE1     ga_7
 CD2  NE2   PZ    ga_37
 CE1  NE2   PZ    ga_37
 NE2   PZ  OH1    ga_14
 NE2   PZ  OH2    ga_14
 NE2   PZ  OH3     ga_5
 OH1   PZ  OH2    ga_29
 OH1   PZ  OH3    ga_14
 OH2   PZ  OH3    ga_14
  PZ  OH3  HH3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
 ND1  CE1  NE2  CD2    gi_1
 CD2   CG  ND1  CE1    gi_1
 CD2   CG  NE2  HD2    gi_1
 CE1  ND1  NE2  HE1    gi_1
 NE2  CD2  CE1   PZ    gi_1
   C   CA   +N    O    gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  ND1    gd_40
 CD2  NE2   PZ  OH3    gd_19
 CD2  NE2   PZ  OH3    gd_22
 NE2   PZ  OH3  HH3    gd_19
 NE2   PZ  OH3  HH3    gd_22

; 1-phosphohistidine (-2)
[ H12 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR   -0.54000    2
 CD2    C    0.13000    2
 HD2   HC    0.14000    2
 CE1    C    0.13000    2
 HE1   HC    0.14000    2
 NE2   NR    0.00000    2
  PZ    P   -0.09500    3 ; to add up to -2 net charge
 OH1   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OH2   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
 OH3   OM   -0.63500    3 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
```

```
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  CE1   gb_10
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 CE1  HE1   gb_3
 CE1  NE2   gb_10
 NE2   PZ   gb_24
  PZ  OH1   gb_24
  PZ  OH2   gb_24
  PZ  OH3   gb_24
   C    O   gb_5
   C   +N   gb_10
[ exclusions ]
;  ai   aj
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE1
  CG   PZ
 ND1  HD2
 ND1   PZ
 CD2  HE1
 HD2  CE1
 HD2   PZ
 HE1   PZ
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  ND1    ga_37
  CB   CG  CD2    ga_37
 ND1   CG  CD2    ga_7
  CG  ND1  CE1    ga_7
  CG  CD2  HD2    ga_36
  CG  CD2  NE2    ga_7
 HD2  CD2  NE2    ga_36
 ND1  CE1  HE1    ga_36
 ND1  CE1  NE2    ga_7
 HE1  CE1  NE2    ga_36
 CD2  NE2  CE1    ga_7
 CD2  NE2   PZ    ga_37
 CE1  NE2   PZ    ga_37
 NE2   PZ  OH1    ga_14
 NE2   PZ  OH2    ga_14
 NE2   PZ  OH3    ga_14
 OH1   PZ  OH2    ga_14
 OH1   PZ  OH3    ga_14
 OH2   PZ  OH3    ga_14
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
 ND1  CE1  NE2  CD2   gi_1
 CD2   CG  ND1  CE1   gi_1
 CD2   CG  NE2  HD2   gi_1
 CE1  ND1  NE2  HE1   gi_1
 NE2  CD2  CE1   PZ   gi_1
   C   CA   +N    O   gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  ND1   gd_40
 CD2  NE2   PZ  OH1   gd_19
 CD2  NE2   PZ  OH1   gd_22

; 3-phosphohistidine (-1)
[ H31 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR    0.00000    2
 CD2    C    0.13000    3
 HD2   HC    0.14000    3
 CE1    C    0.13000    3
 HE1   HC    0.14000    3
 NE2   NR   -0.54000    3
 PE3    P    0.42000    4 ; to add up to -1 net charge
 OZ1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OZ3   OA   -0.54800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
 HZ3    H    0.39800    4 ; from the hydroxyl group of nucleotides (e.g., ATP)
  C    C     0.450    5
  O    O    -0.450    5
[ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG  ND1   gb_10
 CG  CD2   gb_10
ND1  CE1   gb_10
ND1  PE3   gb_24
CD2  HD2   gb_3
CD2  NE2   gb_10
CE1  HE1   gb_3
CE1  NE2   gb_10
PE3  OZ1   gb_24
PE3  OZ2   gb_24
PE3  OZ3   gb_28
OZ3  HZ3   gb_1
 C    O   gb_5
 C   +N   gb_10
[ exclusions ]
;  ai   aj
 CB  PE3
 CB  HD2
 CB  CE1
 CB  NE2
 CG  HE1
ND1  HD2
PE3  CD2
PE3  HE1
PE3  NE2
CD2  HE1
HD2  CE1
HZ3  ND1
HZ3  OZ1
HZ3  OZ2
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  ND1    ga_37
  CB   CG  CD2    ga_37
 ND1   CG  CD2    ga_7
  CG  ND1  CE1    ga_7
  CG  ND1  PE3    ga_37
 CE1  ND1  PE3    ga_37
  CG  CD2  HD2    ga_36
  CG  CD2  NE2    ga_7
 HD2  CD2  NE2    ga_36
 ND1  CE1  HE1    ga_36
 ND1  CE1  NE2    ga_7
 HE1  CE1  NE2    ga_36
 CD2  NE2  CE1    ga_7
 ND1  PE3  OZ1    ga_14
 ND1  PE3  OZ2    ga_14
 ND1  PE3  OZ3    ga_5
 OZ1  PE3  OZ2    ga_29
 OZ1  PE3  OZ3    ga_14
 OZ2  PE3  OZ3    ga_14
 PE3  OZ3  HZ3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
 ND1  CE1  NE2  CD2   gi_1
 CD2   CG  ND1  CE1   gi_1
 CD2   CG  NE2  HD2   gi_1
 CE1  ND1  NE2  HE1   gi_1
 ND1   CG  CE1  PE3   gi_1
```

```
  C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   CG  ND1    gd_40
   CG  ND1  PE3  OZ3    gd_19
   CG  ND1  PE3  OZ3    gd_22
  ND1  PE3  OZ3  HZ3    gd_19
  ND1  PE3  OZ3  HZ3    gd_22

; 3-phosphohistidine (-2)
[ H32 ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.00000    2
  ND1   NR    0.00000    2
  CD2    C    0.13000    3
  HD2   HC    0.14000    3
  CE1    C    0.13000    3
  HE1   HC    0.14000    3
  NE2   NR   -0.54000    3
  PE3    P   -0.09500    4 ; to add up to -2 net charge
  OZ1   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ2   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
  OZ3   OM   -0.63500    4 ; from the phosphate group of nucleotides (e.g., ATP)
    C    C     0.450    5
    O    O    -0.450    5
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG  ND1   gb_10
   CG  CD2   gb_10
  ND1  CE1   gb_10
  ND1  PE3   gb_24
  CD2  HD2   gb_3
  CD2  NE2   gb_10
  CE1  HE1   gb_3
  CE1  NE2   gb_10
  PE3  OZ1   gb_24
  PE3  OZ2   gb_24
  PE3  OZ3   gb_24
    C    O   gb_5
    C   +N   gb_10
 [ exclusions ]
; ai   aj
   CB  PE3
   CB  HD2
   CB  CE1
   CB  NE2
   CG  HE1
  ND1  HD2
  PE3  CD2
  PE3  HE1
  PE3  NE2
  CD2  HE1
  HD2  CE1
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H   ga_32
   -C    N   CA   ga_31
    H    N   CA   ga_18
    N   CA   CB   ga_13
    N   CA    C   ga_13
   CB   CA    C   ga_13
   CA   CB   CG   ga_15
   CB   CG  ND1   ga_37
   CB   CG  CD2   ga_37
  ND1   CG  CD2    ga_7
   CG  ND1  CE1    ga_7
   CG  ND1  PE3   ga_37
  CE1  ND1  PE3   ga_37
   CG  CD2  HD2   ga_36
   CG  CD2  NE2    ga_7
  HD2  CD2  NE2   ga_36
  ND1  CE1  HE1   ga_36
  ND1  CE1  NE2    ga_7
  HE1  CE1  NE2   ga_36
  CD2  NE2  CE1    ga_7
  ND1  PE3  OZ1   ga_14
  ND1  PE3  OZ2   ga_14
  ND1  PE3  OZ3   ga_14
  OZ1  PE3  OZ2   ga_14
  OZ1  PE3  OZ3   ga_14
  OZ2  PE3  OZ3   ga_14
   CA    C    O   ga_30
   CA    C   +N   ga_19
    O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
    N   -C   CA    H   gi_1
   CA    N    C   CB   gi_2
   CG  ND1  CD2   CB   gi_1
   CG  ND1  CE1  NE2   gi_1
   CG  CD2  NE2  CE1   gi_1
  ND1   CG  CD2  NE2   gi_1
  ND1  CE1  NE2  CD2   gi_1
  CD2   CG  ND1  CE1   gi_1
  CD2   CG  NE2  HD2   gi_1
  CE1  ND1  NE2  HE1   gi_1
  ND1   CG  CE1  PE3   gi_1
    C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   CG  ND1    gd_40
   CG  ND1  PE3  OZ1    gd_19
   CG  ND1  PE3  OZ1    gd_22

; N6-methyllysine (0)
[ KMN ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE  CH2    0.22000    3 ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   NZ   NT   -0.88000    3 ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   HZ    H    0.44000    3 ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   CH  CH3    0.22000    3 ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
    C    C     0.450    4
    O    O    -0.450    4
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ   HZ   gb_2
   NZ   CH   gb_21
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
   -C    N    H   ga_32
   -C    N   CA   ga_31
    H    N   CA   ga_18
    N   CA   CB   ga_13
    N   CA    C   ga_13
   CB   CA    C   ga_13
   CA   CB   CG   ga_15
   CB   CG   CD   ga_15
   CG   CD   CE   ga_15
   CD   CE   NZ   ga_15
   CE   NZ   HZ   ga_11
   HZ   NZ   CH   ga_11
   CE   NZ   CH   ga_13
   CA    C    O   ga_30
   CA    C   +N   ga_19
    O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
    N   -C   CA    H   gi_1
   CA    N    C   CB   gi_2
    C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
```

```
   N   CA    C   +N    gd_42
   CA   CB   CG   CD    gd_34
   CB   CG   CD   CE    gd_34
   CG   CD   CE   NZ    gd_34
   CD   CE   NZ   CH    gd_29

; N6-methyllysine (+1)
[ KMC ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE  CH2    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   NZ   NL    0.10400    3 ; to add up to +1 net charge
  HZ1    H    0.24800    3
  HZ2    H    0.24800    3
   CH  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ  HZ1   gb_2
   NZ  HZ2   gb_2
   NZ   CH   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
 ;  ai   aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
   CG   CD   CE    ga_15
   CD   CE   NZ    ga_15
   CE   NZ  HZ1    ga_11
   CE   NZ  HZ2    ga_11
  HZ1   NZ  HZ2    ga_10
  HZ1   NZ   CH    ga_11
  HZ2   NZ   CH    ga_11
   CE   NZ   CH    ga_13
   CA    C    O    ga_30
   CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
 ;  ai   aj    ak   al   gromos type
   N   -C   CA    H    gi_1
   CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
 ;  ai   aj    ak   al   gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
   CA   CB   CG   CD    gd_34
   CB   CG   CD   CE    gd_34
   CG   CD   CE   NZ    gd_34
   CD   CE   NZ   CH    gd_29

; N6,N6-dimethyllysine (0)
[ K2M ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE  CH2    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   NZ   NT   -0.63000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
  CH1  CH3    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
  CH2  CH3    0.21000    3 ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ  CH1   gb_21
   NZ  CH2   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
 ;  ai   aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
   CG   CD   CE    ga_15
   CD   CE   NZ    ga_15
   CE   NZ  CH1    ga_13
   CE   NZ  CH2    ga_13
  CH1   NZ  CH2    ga_13
   CA    C    O    ga_30
   CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
 ;  ai   aj    ak   al   gromos type
   N   -C   CA    H    gi_1
   CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
 ;  ai   aj    ak   al   gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C    gd_44
   -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
   CA   CB   CG   CD    gd_34
   CB   CG   CD   CE    gd_34
   CG   CD   CE   NZ    gd_34
   CD   CE   NZ  CH1    gd_29

; N6,N6-dimethyllysine (+1)
[ K2C ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE  CH2    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   NZ   NL    0.15200    3 ; to add up to +1 net charge
   HZ    H    0.24800    3
  CH1  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  CH2  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   NZ   gb_21
   NZ   HZ   gb_2
   NZ  CH1   gb_21
   NZ  CH2   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
 ;  ai   aj    ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
```

```
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CD   CE   NZ    ga_15
  CE   NZ   HZ    ga_11
 CH1   NZ   HZ    ga_11
 CH2   NZ   HZ    ga_11
  CE   NZ  CH1    ga_13
  CE   NZ  CH2    ga_13
 CH1   NZ  CH2    ga_13
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ  CH1    gd_29

; N6,N6,N6-trimethyllysine
[ K3C ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
  NZ   NL    0.20000    3 ; to add up to +1 net charge
 CH1  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
 CH2  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
 CH3  CH3    0.20000    3 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C    0.450    4
   O    O   -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   CE   gb_27
  CE   NZ   gb_21
  NZ  CH1   gb_21
  NZ  CH2   gb_21
  NZ  CH3   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CD   CE   NZ    ga_15
  CE   NZ  CH1    ga_13
  CE   NZ  CH2    ga_13
  CE   NZ  CH3    ga_13
 CH1   NZ  CH2    ga_13
 CH1   NZ  CH3    ga_13
 CH2   NZ  CH3    ga_13
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ  CH1    gd_29

; omega-N-methylarginine (0)
[ RMN ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  NE   NE   -0.31000    3
  HE    H    0.31000    3
  CZ    C    0.18000    4 ; 0.36 charge divided between CZ and CT to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
 NH1   NE   -0.36000    4 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
  CT  CH3    0.18000    4 ; 0.36 charge divided between CZ and CT to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
 NH2   NZ   -0.88000    5
HH21    H    0.44000    5
HH22    H    0.44000    5
   C    C      0.450    6
   O    O     -0.450    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   NE   gb_21
  NE   HE   gb_2
  NE   CZ   gb_11
  CZ  NH1   gb_11
  CZ  NH2   gb_11
 NH1   CT   gb_21
 NH2 HH21   gb_2
 NH2 HH22   gb_2
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   NE    ga_13
  CD   NE   HE    ga_20
  CD   NE   CZ    ga_33
  HE   NE   CZ    ga_23
  NE   CZ  NH1    ga_28
  NE   CZ  NH2    ga_28
 NH1   CZ  NH2    ga_28
  CZ  NH1   CT    ga_27
  CZ  NH2 HH21    ga_23
  CZ  NH2 HH22    ga_23
HH21  NH2 HH22    ga_24
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH2 HH21 HH22   CZ    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
```

```
  CB   CG   CD   NE    gd_34
  CG   CD   NE   CZ    gd_39
  CD   NE   CZ  NH1    gd_14
  NE   CZ  NH1   CT    gd_14
  NE   CZ  NH2 HH21    gd_14

; omega-N-methylarginine (+1)
[ RMC ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.09000    2
  NE   NE   -0.11000    2
  HE    H    0.24000    2
  CZ    C    0.34000    2
 NH1   NZ   -0.26000    2
HH11    H    0.24000    2
HH12    H    0.24000    2
 NH2   NZ   -0.11000    2 ; from NE atom of ARG
 HH2    H    0.24000    2
  CT  CH3    0.09000    2 ; from CD atom of ARG
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   NE   gb_21
  NE   HE   gb_2
  NE   CZ   gb_11
  CZ  NH1   gb_11
  CZ  NH2   gb_11
 NH1 HH11   gb_2
 NH1 HH12   gb_2
 NH2  HH2   gb_2
 NH2   CT   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   NE    ga_13
  CD   NE   HE    ga_20
  CD   NE   CZ    ga_33
  HE   NE   CZ    ga_23
  NE   CZ  NH1    ga_28
  NE   CZ  NH2    ga_28
 NH1   CZ  NH2    ga_28
  CZ  NH1 HH11    ga_23
  CZ  NH1 HH12    ga_23
HH11  NH1 HH12    ga_24
  CZ  NH2  HH2    ga_23
 HH2  NH2   CT    ga_20
  CZ  NH2   CT    ga_33
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH1 HH11 HH12   CZ    gi_1
 NH2  HH2   CT   CZ    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   NE    gd_34
  CG   CD   NE   CZ    gd_39
  CD   NE   CZ  NH1    gd_14
  NE   CZ  NH1 HH11    gd_14
  NE   CZ  NH2   CT    gd_14

; symmetric-dimethylarginine (0)
[ RSM ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  NE   NE   -0.31000    3
  HE    H    0.31000    3
  CZ    C    0.18000    4 ; 0.36 charge divided between CZ and CT1 to add up to 0 net
charge, similar to methyl groups of other methylation residues and also used in other
building blocks (e.g., TMP)
 NH1   NE   -0.36000    4 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
 CT1  CH3    0.18000    4 ; 0.36 charge divided between CZ and CT1 to add up to 0
net charge, similar to methyl groups of other methylation residues and also used in
other building blocks (e.g., TMP)
 NH2   NE   -0.31000    5 ; from NE atom of ARGN
 HH2    H    0.31000    5 ; from HE atom of ARGN
 CT2  CH3    0.00000    6 ; from CD atom of ARGN
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   NE   gb_21
  NE   HE   gb_2
  NE   CZ   gb_11
  CZ  NH1   gb_11
  CZ  NH2   gb_11
 NH1  CT1   gb_21
 NH2  HH2   gb_2
 NH2  CT2   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   NE    ga_13
  CD   NE   HE    ga_20
  CD   NE   CZ    ga_33
  HE   NE   CZ    ga_23
  NE   CZ  NH1    ga_28
  NE   CZ  NH2    ga_28
 NH1   CZ  NH2    ga_28
  CZ  NH1  CT1    ga_27
  CZ  NH2  HH2    ga_23
 HH2  NH2  CT2    ga_20
  CZ  NH2  CT2    ga_33
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  NE   CD   CZ   HE    gi_1
  CZ  NH1  NH2   NE    gi_1
 NH2  HH2  CT2   CZ    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   NE    gd_34
  CG   CD   NE   CZ    gd_39
  CD   NE   CZ  NH1    gd_14
  NE   CZ  NH1  CT1    gd_14
  NE   CZ  NH2  CT2    gd_14

; symmetric-dimethylarginine (+1)
[ RMS ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
```

```
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    1
  CD   CH2    0.09000    2
  NE   NE    -0.11000    2
  HE    H     0.24000    2
  CZ    C     0.34000    2
 NH1   NZ    -0.11000    2 ; from NE atom of ARG
 HH1    H     0.24000    2
 CT1   CH3    0.09000    2 ; from CD atom of ARG
 NH2   NZ    -0.11000    2 ; from NE atom of ARG
 HH2    H     0.24000    2
 CT2   CH3    0.09000    2 ; from CD atom of ARG
   C    C      0.450    3
   O    O     -0.450    3
[ bonds ]
   N    H    gb_2
   N   CA    gb_21
  CA   CB    gb_27
  CA    C    gb_27
  CB   CG    gb_27
  CG   CD    gb_27
  CD   NE    gb_21
  NE   HE    gb_2
  NE   CZ    gb_11
  CZ  NH1    gb_11
  CZ  NH2    gb_11
 NH1  HH1    gb_2
 NH1  CT1    gb_21
 NH2  HH2    gb_2
 NH2  CT2    gb_21
   C    O    gb_5
   C   +N    gb_10
[ angles ]
;  ai    aj    ak   gromos type
  -C    N     H     ga_32
  -C    N    CA     ga_31
   H    N    CA     ga_18
   N   CA    CB     ga_13
   N   CA     C     ga_13
  CB   CA     C     ga_13
  CA   CB    CG     ga_15
  CB   CG    CD     ga_15
  CG   CD    NE     ga_13
  CD   NE    HE     ga_20
  CD   NE    CZ     ga_33
  HE   NE    CZ     ga_23
  NE   CZ   NH1     ga_28
  NE   CZ   NH2     ga_28
 NH1   CZ   NH2     ga_28
  CZ  NH1   HH1     ga_23
 HH1  NH1   CT1     ga_20
  CZ  NH1   CT1     ga_33
  CZ  NH2   HH2     ga_23
 HH2  NH2   CT2     ga_20
  CZ  NH2   CT2     ga_33
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
[ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C    CA     H     gi_1
  CA    N     C    CB     gi_2
  NE   CD    CZ    HE     gi_1
  CZ  NH1   NH2    NE     gi_1
 NH1  HH1   CT1    CZ     gi_1
 NH2  HH2   CT2    CZ     gi_1
   C   CA    +N     O     gi_1
[ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C     N    CA     gd_14
  -C    N     CA     C     gd_44
  -C    N     CA     C     gd_43
   N   CA     CB    CG     gd_34
   N   CA     C    +N     gd_45
   N   CA     C    +N     gd_42
  CA   CB     CG    CD     gd_34
  CB   CG     CD    NE     gd_34
  CG   CD     NE    CZ     gd_39
  CD   NE     CZ   NH1     gd_14
  NE   CZ    NH1   CT1     gd_14
  NE   CZ    NH2   CT2     gd_14

; asymmetric-dimethylarginine (0)
[ RAM ]
 [ atoms ]
   N    N    -0.31000    0
   H    H     0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    2
  CD   CH2    0.00000    2
  NE   NE    -0.31000    3
  HE    H     0.31000    3
  CZ    C     0.26600    4
 NH1   NE    -0.67400    4
 HH1    H     0.40800    4
 NH2   NZ    -0.20000    5 ; from ring nitrogen atoms of nucleotides (e.g., ATP)
 CT1   CH3    0.10000    5 ; 0.2 charge divided between CT1 and CT2 to add up to 0
net charge, also similar to methyl groups of methyl-arginine modifications and used
for ring carbons in nucleotides (e.g., DCYT)
 CT2   CH3    0.10000    5 ; 0.2 charge divided between CT1 and CT2 to add up to 0
net charge, also similar to methyl groups of methyl-arginine modifications and used
for ring carbons in nucleotides (e.g., DCYT)
   C    C      0.450    6
   O    O     -0.450    6
[ bonds ]
   N    H    gb_2
   N   CA    gb_21
  CA   CB    gb_27
  CA    C    gb_27
  CB   CG    gb_27
  CG   CD    gb_27
  CD   NE    gb_21
  NE   HE    gb_2
  NE   CZ    gb_11
  CZ  NH1    gb_11
  CZ  NH2    gb_11
 NH1  HH1    gb_2
 NH2  CT1    gb_21
 NH2  CT2    gb_21
   C    O    gb_5
   C   +N    gb_10
[ angles ]
;  ai    aj    ak   gromos type
  -C    N     H     ga_32
  -C    N    CA     ga_31
   H    N    CA     ga_18
   N   CA    CB     ga_13
   N   CA     C     ga_13
  CB   CA     C     ga_13
  CA   CB    CG     ga_15
  CB   CG    CD     ga_15
  CG   CD    NE     ga_13
  CD   NE    HE     ga_20
  CD   NE    CZ     ga_33
  HE   NE    CZ     ga_23
  NE   CZ   NH1     ga_28
  NE   CZ   NH2     ga_28
 NH1   CZ   NH2     ga_28
  CZ  NH1   HH1     ga_23
  CZ  NH2   CT1     ga_28
  CZ  NH2   CT2     ga_28
 CT1  NH2   CT2     ga_28
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
[ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C    CA     H     gi_1
  CA    N     C    CB     gi_2
  NE   CD    CZ    HE     gi_1
  CZ  NH1   NH2    NE     gi_1
 NH2  CT1   CT2    CZ     gi_1
   C   CA    +N     O     gi_1
[ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA    -C     N    CA     gd_14
  -C    N     CA     C     gd_44
  -C    N     CA     C     gd_43
   N   CA     CB    CG     gd_34
   N   CA     C    +N     gd_45
   N   CA     C    +N     gd_42
  CA   CB     CG    CD     gd_34
  CB   CG     CD    NE     gd_34
  CG   CD     NE    CZ     gd_39
  CD   NE     CZ   NH1     gd_14
  NE   CZ    NH1   HH1     gd_14
  NE   CZ    NH2   CT1     gd_14

; asymmetric-dimethylarginine (+1)
[ RMA ]
 [ atoms ]
   N    N    -0.31000    0
   H    H     0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    1
  CD   CH2    0.09000    2
  NE   NE    -0.11000    2
  HE    H     0.24000    2
  CZ    C     0.43000    2 ; from CZ of R1P
 NH1   NZ    -0.26000    2
HH11    H     0.24000    2
HH12    H     0.24000    2
```

```
  NH2   NZ   -0.05000    2 ; to add up to +1 net charge, also used for nitrogen atoms
in hisitidine building blocks
  CT1   CH3    0.09000    2 ; from CD atom of ARG
  CT2   CH3    0.09000    2 ; from CD atom of ARG
    C    C      0.450    3
    O    O     -0.450    3
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   NE   gb_21
   NE   HE   gb_2
   NE   CZ   gb_11
   CZ  NH1   gb_11
   CZ  NH2   gb_11
  NH1 HH11   gb_2
  NH1 HH12   gb_2
  NH2  CT1   gb_21
  NH2  CT2   gb_21
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N     H     ga_32
   -C    N    CA     ga_31
    H    N    CA     ga_18
    N   CA    CB     ga_13
    N   CA     C     ga_13
   CB   CA     C     ga_13
   CA   CB    CG     ga_15
   CB   CG    CD     ga_15
   CG   CD    NE     ga_13
   CD   NE    HE     ga_20
   CD   NE    CZ     ga_33
   HE   NE    CZ     ga_23
   NE   CZ   NH1     ga_28
   NE   CZ   NH2     ga_28
  NH1   CZ   NH2     ga_28
   CZ  NH1  HH11     ga_23
   CZ  NH1  HH12     ga_23
 HH11  NH1  HH12     ga_24
   CZ  NH2   CT1     ga_28
   CZ  NH2   CT2     ga_28
  CT1  NH2   CT2     ga_28
   CA    C     O     ga_30
   CA    C    +N     ga_19
    O    C    +N     ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA     H     gi_1
   CA    N     C    CB     gi_2
   NE   CD    CZ    HE     gi_1
   CZ  NH1   NH2    NE     gi_1
  NH1 HH11  HH12    CZ     gi_1
  NH2  CT1   CT2    CZ     gi_1
    C   CA    +N     O     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_14
   -C     N    CA     C     gd_44
   -C     N    CA     C     gd_43
    N    CA    CB    CG     gd_34
    N    CA     C    +N     gd_45
    N    CA     C    +N     gd_42
   CA    CB    CG    CD     gd_34
   CB    CG    CD    NE     gd_34
   CG    CD    NE    CZ     gd_39
   CD    NE    CZ   NH1     gd_14
   NE    CZ   NH1  HH11     gd_14
   NE    CZ   NH2   CT1     gd_14

; 1-methylhistidine (0)
[ H1M ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.00000    2
  ND1   NR   -0.54000    2
  CD2    C    0.00000    2
  HD2   HC    0.14000    2
  CE1    C    0.09000    2 ; to add up to 0 net charge, also from CD atom of ARG
  HE1   HC    0.14000    2
  NE2   NR   -0.05000    2
   CZ  CH3    0.22000    2 ; from the methyl groups of dimethylamine reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
    C    C      0.450    3
    O    O     -0.450    3
 [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG  ND1   gb_10
   CG  CD2   gb_10
  ND1  CE1   gb_10
  CD2  HD2   gb_3
  CD2  NE2   gb_10
  CE1  HE1   gb_3
  CE1  NE2   gb_10
  NE2   CZ   gb_22
    C    O   gb_5
    C   +N   gb_10
 [ exclusions ]
;  ai    aj
   CB  HD2
   CB  CE1
   CB  NE2
   CG  HE1
   CG   CZ
  ND1  HD2
  ND1   CZ
  CD2  HE1
  HD2  CE1
  HD2   CZ
  HE1   CZ
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N     H     ga_32
   -C    N    CA     ga_31
    H    N    CA     ga_18
    N   CA    CB     ga_13
    N   CA     C     ga_13
   CB   CA     C     ga_13
   CA   CB    CG     ga_15
   CB   CG   ND1     ga_37
   CB   CG   CD2     ga_37
  ND1   CG   CD2     ga_7
   CG  ND1   CE1     ga_7
   CG  CD2   HD2     ga_36
   CG  CD2   NE2     ga_7
  HD2  CD2   NE2     ga_36
  ND1  CE1   HE1     ga_36
  ND1  CE1   NE2     ga_7
  HE1  CE1   NE2     ga_36
  CD2  NE2   CE1     ga_7
  CD2  NE2    CZ     ga_37
  CE1  NE2    CZ     ga_37
   CA    C     O     ga_30
   CA    C    +N     ga_19
    O    C    +N     ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N   -C    CA     H     gi_1
   CA    N     C    CB     gi_2
   CG  ND1   CD2    CB     gi_1
   CG  ND1   CE1   NE2     gi_1
   CG  CD2   NE2   CE1     gi_1
  ND1   CG   CD2   NE2     gi_1
  ND1  CE1   NE2   CD2     gi_1
  CD2   CG   ND1   CE1     gi_1
  CD2   CG   NE2   HD2     gi_1
  CE1  ND1   NE2   HE1     gi_1
  NE2  CD2   CE1    CZ     gi_1
    C   CA    +N     O     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_14
   -C     N    CA     C     gd_44
   -C     N    CA     C     gd_43
    N    CA    CB    CG     gd_34
    N    CA     C    +N     gd_45
    N    CA     C    +N     gd_42
   CA    CB    CG   ND1     gd_40

; 1-methylhistidine (+1)
[ H1C ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C   -0.05000    2
  ND1   NR    0.38000    2
  HD1    H    0.30000    2
  CD2    C   -0.10000    2
  HD2   HC    0.10000    2
  CE1    C   -0.24000    2 ; to add up to +1 net charge
  HE1   HC    0.10000    2
  NE2   NR    0.31000    2
```

```
  CZ  CH3    0.20000    2 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  HD1   gb_2
 ND1  CE1   gb_10
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 CE1  HE1   gb_3
 CE1  NE2   gb_10
 NE2   CZ   gb_22
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE1
  CG   CZ
 ND1  HD2
 ND1   CZ
 HD1  CD2
 HD1  HE1
 HD1  NE2
 CD2  HE1
 HD2  CE1
 HD2   CZ
 HE1   CZ
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  ND1   ga_37
  CB   CG  CD2   ga_37
 ND1   CG  CD2    ga_7
  CG  ND1  HD1   ga_36
  CG  ND1  CE1    ga_7
 HD1  ND1  CE1   ga_36
  CG  CD2  HD2   ga_36
  CG  CD2  NE2    ga_7
 HD2  CD2  NE2   ga_36
 ND1  CE1  HE1   ga_36
 ND1  CE1  NE2    ga_7
 HE1  CE1  NE2   ga_36
 CD2  NE2  CE1    ga_7
 CD2  NE2   CZ   ga_37
 CE1  NE2   CZ   ga_37
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
 ND1   CG  CE1  HD1   gi_1
 ND1  CE1  NE2  CD2   gi_1
 CD2   CG  ND1  CE1   gi_1
 CD2   CG  NE2  HD2   gi_1
 CE1  ND1  NE2  HE1   gi_1
 NE2  CD2  CE1   CZ   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  ND1   gd_40

; 3-methylhistidine (0)
[ H3M ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR   -0.05000    2
 CD2    C    0.00000    2
 HD2   HC    0.14000    2
 CE1    C    0.09000    2 ; to add up to 0 net charge
 CE3  CH3    0.22000    2 ; from the methyl groups of dimethylamine reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
 HE1   HC    0.14000    2
 NE2   NR   -0.54000    2
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  CE3   gb_22
 ND1  CE1   gb_10
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 CE1  HE1   gb_3
 CE1  NE2   gb_10
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  CE3
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE1
 ND1  HD2
 CE3  CD2
 CE3  HE1
 CE3  NE2
 CD2  HE1
 HD2  CE1
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  ND1   ga_37
  CB   CG  CD2   ga_37
 ND1   CG  CD2    ga_7
  CG  ND1  CE3   ga_37
  CG  ND1  CE1    ga_7
 CE3  ND1  CE1   ga_37
  CG  CD2  HD2   ga_36
  CG  CD2  NE2    ga_7
 HD2  CD2  NE2   ga_36
 ND1  CE1  HE1   ga_36
 ND1  CE1  NE2    ga_7
 HE1  CE1  NE2   ga_36
 CD2  NE2  CE1    ga_7
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
 ND1   CG  CE1  CE3   gi_1
 ND1  CE1  NE2  CD2   gi_1
 CD2   CG  ND1  CE1   gi_1
 CD2   CG  NE2  HD2   gi_1
 CE1  ND1  NE2  HE1   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  ND1   gd_40
```

```
; 3-methylhistidine (+1)
[ H3C ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.05000    2
 ND1   NR    0.38000    2
 CE3  CH3    0.20000    2 ; derived by analogy to methyl groups of amines reported
by Oostenbrink et al. DOI: 10.1002/cphc.200400542
 CD2    C   -0.10000    2
 HD2   HC    0.10000    2
 CE1    C   -0.24000    2 ; to add up to +1 net charge
 HE1   HC    0.10000    2
 NE2   NR    0.31000    2
 HE2    H    0.30000    2
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  CE3   gb_22
 ND1  CE1   gb_10
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 CE1  HE1   gb_3
 CE1  NE2   gb_10
 NE2  HE2   gb_2
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  CE3
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE1
  CG  HE2
 ND1  HD2
 ND1  HE2
 CE3  CD2
 CE3  HE1
 CE3  NE2
 CD2  HE1
 HD2  CE1
 HD2  HE2
 HE1  HE2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  ND1   ga_37
  CB   CG  CD2   ga_37
 ND1   CG  CD2   ga_7
  CG  ND1  CE3   ga_37
  CG  ND1  CE1   ga_7
 CE3  ND1  CE1   ga_37
  CG  CD2  HD2   ga_36
  CG  CD2  NE2   ga_7
 HD2  CD2  NE2   ga_36
 ND1  CE1  HE1   ga_36
 ND1  CE1  NE2   ga_7
 HE1  CE1  NE2   ga_36
 CD2  NE2  CE1   ga_7
 CD2  NE2  HE2   ga_36
 CE1  NE2  HE2   ga_36
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  ND1  CD2   CB   gi_1
  CG  ND1  CE1  NE2   gi_1
  CG  CD2  NE2  CE1   gi_1
 ND1   CG  CD2  NE2   gi_1
 ND1   CG  CE1  CE3   gi_1
 ND1  CE1  NE2  CD2   gi_1
 CD2   CG  ND1  CE1   gi_1
 CD2   CG  NE2  HD2   gi_1
 CE1  ND1  NE2  HE1   gi_1
 NE2  CD2  CE1  HE2   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  ND1   gd_40

; N5-methylglutamine
[ QME ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  CD    C    0.19000    2 ; by analogy to GLN, the peptide bond and the aldehyde
group reported by Dolenc et al. DOI: 10.1093/nar/gki195
 OE1    O   -0.45000    2
 NE2    N   -0.15000    2 ; by analogy to GLN and the peptide bond
 HE2    H    0.31000    2 ; from the hydrogen atom of the peptide bond
  CZ  CH3    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD  OE1   gb_5
  CD  NE2   gb_9
 NE2  HE2   gb_2
 NE2   CZ   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG   CD   ga_15
  CG   CD  OE1   ga_30
  CG   CD  NE2   ga_19
 OE1   CD  NE2   ga_33
  CD  NE2  HE2   ga_32
 HE2  NE2   CZ   ga_18
  CD  NE2   CZ   ga_31
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CD  OE1  NE2   CG   gi_1
 NE2   CZ   CD  HE2   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG   CD   gd_34
  CB   CG   CD  NE2   gd_40
  CG   CD  NE2   CZ   gd_14

; N4-methylasparagine
[ NME ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  CG    C    0.19000    2 ; by analogy to ASN, the peptide bond and the aldehyde
group reported by Dolenc et al. DOI: 10.1093/nar/gki195
 OD1    O   -0.45000    2
 ND2    N   -0.15000    2 ; by analogy to ASN and the peptide bond
 HD2    H    0.31000    2 ; from the hydrogen of the peptide bond
```

```
  CE   CH3    0.00000    3
   C    C     0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  OD1   gb_5
  CG  ND2   gb_9
 ND2  HD2   gb_2
 ND2   CE   gb_21
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_30
  CB   CG  ND2   ga_19
 OD1   CG  ND2   ga_33
  CG  ND2  HD2   ga_32
 HD2  ND2   CE   ga_18
  CG  ND2   CE   ga_31
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  OD1  ND2   CB   gi_1
 ND2   CE   CG  HD2   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  ND2   gd_40
  CB   CG  ND2   CE   gd_14

; glutamate methyl ester
[ EME ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD    C    0.25300    2 ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
  OE1   O   -0.45000    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
  OE2  OE   -0.06900    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
  CZ  CH3    0.26600    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
   C    C     0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD  OE1   gb_5
  CD  OE2   gb_13
 OE2   CZ   gb_18
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG   CD   ga_15
  CG   CD  OE1   ga_30
  CG   CD  OE2   ga_19
 OE1   CD  OE2   ga_33
  CD  OE2   CZ   ga_22
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CD  OE1  OE2   CG   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG   CD   gd_34
  CB   CG   CD  OE2   gd_40
  CG   CD  OE2   CZ   gd_12

; aspartate methyl ester
[ DMA ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.25300    2 ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
  OD1   O   -0.45000    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
  OD2  OE   -0.06900    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
  CE  CH3    0.26600    2 ; from the ester group reported by Chandrasekhar et al.
DOI: 10.1081/smts-120030764
   C    C     0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  OD1   gb_5
  CG  OD2   gb_13
 OD2   CE   gb_18
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_30
  CB   CG  OD2   ga_19
 OD1   CG  OD2   ga_33
  CG  OD2   CE   ga_22
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  OD1  OD2   CB   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  OD2   gd_40
  CB   CG  OD2   CE   gd_12

; S-methylcysteine
[ CYM ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.24100    2 ; from MET
  SG    S   -0.48200    2 ; from MET
  CD  CH3    0.24100    2 ; from MET
```

```
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   SG   gb_32
  SG   CD   gb_31
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   SG    ga_16
  CB   SG   CD    ga_4
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   SG    gd_34
  CA   CB   SG   CD    gd_26
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42

; N6-acetyllysine
[ KAC ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ    N   -0.15000    4 ; by analogy to ASN/GLN and the peptide bond
  HZ    H    0.31000    4 ; from the hydrogen of the peptide bond
  CH    C    0.19000    4 ; by analogy to ASN/GLN, the peptide bond and the aldehyde
group reported by Dolenc et al. DOI: 10.1093/nar/gki195
 OI2    O   -0.45000    4 ; from the carbonyl group (of e.g., GLN)
 CI1  CH3    0.10000    4 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
   C    C      0.450    5
   O    O     -0.450    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   CE   gb_27
  CE   NZ   gb_21
  NZ   HZ   gb_2
  NZ   CH   gb_10
  CH  OI2   gb_5
  CH  CI1   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CD   CE   NZ    ga_15
  CE   NZ   HZ    ga_18
  HZ   NZ   CH    ga_32
  CE   NZ   CH    ga_31
  NZ   CH  OI2    ga_33
  NZ   CH  CI1    ga_19
 OI2   CH  CI1    ga_30
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  NZ   CH   CE   HZ    gi_1
  CH  CI1   NZ  OI2    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ   CH    gd_39
  CE   NZ   CH  CI1    gd_14

; 3-hydroxyproline (R)
[ PH3 ]
 [ atoms ]
   N    N    0.00000    0
  CA  CH1    0.00000    1
  CB  CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1   OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1    H    0.40800    2 ; from the hydroxyl group of THR
 CG2 CH2R    0.00000    3
  CD CH2R    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N   CA   gb_21
   N   CD   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1  HG1   gb_1
 CG2   CD   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N   CA    ga_31
  -C    N   CD    ga_31
  CA    N   CD    ga_21
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_13
 OG1   CB  CG2    ga_13
  CB  OG1  HG1    ga_12
  CB  CG2   CD    ga_13
   N   CD  CG2    ga_13
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   CD    gi_1
  CA    N    C   CB    gi_2
  CB  CG2   CA  OG1    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
  CA    N   CD  CG2    gd_39
   N   CA   CB  CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB  OG1  HG1    gd_23
  CA   CB  CG2   CD    gd_34
  CB  CG2   CD    N    gd_34

; 3-hydroxyproline (S)
[ P3H ]
 [ atoms ]
   N    N    0.00000    0
  CA  CH1    0.00000    1
  CB  CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1   OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1    H    0.40800    2 ; from the hydroxyl group of THR
 CG2 CH2R    0.00000    3
  CD CH2R    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
```

```
 [ bonds ]
   N    CA   gb_21
   N    CD   gb_21
  CA    CB   gb_27
  CA    C    gb_27
  CB   OG1   gb_18
  CB   CG2   gb_27
 OG1   HG1   gb_1
 CG2   CD    gb_27
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
; ai    aj    ak   gromos type
  -C    N    CA    ga_31
  -C    N    CD    ga_31
  CA    N    CD    ga_21
   N    CA   CB    ga_13
   N    CA   C     ga_13
  CB    CA   C     ga_13
  CA    CB   OG1   ga_13
  CA    CB   CG2   ga_13
 OG1    CB   CG2   ga_13
  CB   OG1   HG1   ga_12
  CB   CG2   CD    ga_13
   N    CD   CG2   ga_13
  CA    C    O     ga_30
  CA    C   +N     ga_19
   O    C   +N     ga_33
 [ impropers ]
; ai    aj    ak   al   gromos type
   N   -C    CA   CD    gi_1
  CA    N    C    CB    gi_2
  CB   CA   CG2   OG1    gi_2
   C   CA   +N    O     gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA   -C    N    CA    gd_14
  -C    N    CA   C     gd_44
  -C    N    CA   C     gd_43
  CA    N    CD   CG2   gd_39
   N    CA   CB   CG2   gd_34
   N    CA   C    +N    gd_45
   N    CA   C    +N    gd_42
  CA    CB   OG1  HG1   gd_23
  CA    CB   CG2  CD    gd_34
  CB    CG2  CD   N     gd_34

; 4-hydroxyproline (S)
[ HY2 ]
 [ atoms ]
   N    N    0.00000    0
  CA   CH1    0.00000    1
  CB  CH2R    0.00000    1
  CG   CH1    0.26600    2 ; from the hydroxyl group of THR
 OD1    OA   -0.67400    2 ; from the hydroxyl group of THR
 HD1     H    0.40800    2 ; from the hydroxyl group of THR
 CD2  CH2R    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    CA   gb_21
   N   CD2   gb_21
  CA    CB   gb_27
  CA    C    gb_27
  CB    CG   gb_27
  CG   OD1   gb_18
  CG   CD2   gb_27
 OD1   HD1   gb_1
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    CA    ga_31
  -C    N   CD2    ga_31
  CA    N   CD2    ga_21
   N    CA   CB    ga_13
   N    CA   C     ga_13
  CB    CA   C     ga_13
  CA    CB   CG    ga_13
  CB    CG   OD1   ga_13
  CB    CG   CD2   ga_13
 OD1    CG   CD2   ga_13
  CG   OD1   HD1   ga_12
   N   CD2    CG   ga_13
  CA    C    O     ga_30
  CA    C   +N     ga_19
   O    C   +N     ga_33
 [ impropers ]
;  ai    aj    ak   al   gromos type
   N   -C    CA   CD2    gi_1
  CA    N    C    CB    gi_2
  CG   CB   CD2   OD1    gi_2
   C   CA   +N    O     gi_1
 [ dihedrals ]
;  ai    aj    ak   al   gromos type
 -CA   -C    N    CA    gd_14
  -C    N    CA   C     gd_44
  -C    N    CA   C     gd_43
  CA    N   CD2    CG    gd_39
   N    CA   CB   CG    gd_34
   N    CA   C    +N    gd_45
   N    CA   C    +N    gd_42
  CA    CB   CG  CD2    gd_34
  CB    CG  OD1  HD1    gd_23
  CB    CG  CD2   N     gd_34

; 3,4-dihydroxyproline
[ PHH ]
 [ atoms ]
   N    N    0.00000    0
  CA   CH1    0.00000    1
  CB   CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1    OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1     H    0.40800    2 ; from the hydroxyl group of THR
 CG2   CH1    0.26600    3 ; from the hydroxyl group of THR
 OD1    OA   -0.67400    3 ; from the hydroxyl group of THR
 HD1     H    0.40800    3 ; from the hydroxyl group of THR
 CD2  CH2R    0.00000    4
   C    C      0.450    5
   O    O     -0.450    5
 [ bonds ]
   N    CA   gb_21
   N   CD2   gb_21
  CA    CB   gb_27
  CA    C    gb_27
  CB   OG1   gb_18
  CB   CG2   gb_27
 OG1   HG1   gb_1
 CG2   OD1   gb_18
 CG2   CD2   gb_27
 OD1   HD1   gb_1
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    CA    ga_31
  -C    N   CD2    ga_31
  CA    N   CD2    ga_21
   N    CA   CB    ga_13
   N    CA   C     ga_13
  CB    CA   C     ga_13
  CA    CB   OG1   ga_13
  CA    CB   CG2   ga_13
 OG1    CB   CG2   ga_13
  CB   OG1   HG1   ga_12
  CB   CG2   OD1   ga_13
  CB   CG2   CD2   ga_13
 OD1   CG2   CD2   ga_13
 CG2   OD1   HD1   ga_12
   N   CD2   CG2   ga_13
  CA    C    O     ga_30
  CA    C   +N     ga_19
   O    C   +N     ga_33
 [ impropers ]
;  ai    aj    ak   al   gromos type
   N   -C    CA   CD2    gi_1
  CA    N    C    CB    gi_2
  CB   CA   CG2   OG1    gi_2
 CG2    CB   CD2   OD1    gi_2
   C   CA   +N    O     gi_1
 [ dihedrals ]
;  ai    aj    ak   al   gromos type
 -CA   -C    N    CA    gd_14
  -C    N    CA   C     gd_44
  -C    N    CA   C     gd_43
  CA    N   CD2   CG2    gd_39
   N    CA   CB   CG2    gd_34
   N    CA   C    +N    gd_45
   N    CA   C    +N    gd_42
  CA    CB   OG1  HG1    gd_23
  CA    CB   CG2  CD2    gd_34
  CB   CG2   OD1  HD1    gd_23
  CB   CG2   CD2   N     gd_34

; 5-hydroxylysine (0,R)
[ KH5 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    2
  CD   CH1    0.26600    3 ; from the hydroxyl group of THR
 OE1    OA   -0.67400    3 ; from the hydroxyl group of THR
 HE1     H    0.40800    3 ; from the hydroxyl group of THR
 CE2   CH2   -0.24000    4
```

```
  NZ   NT   -0.64000    4
 HZ1    H    0.44000    4
 HZ2    H    0.44000    4
  C    C     0.450     5
  O    O    -0.450     5
 [ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG   CD   gb_27
 CD  OE1   gb_18
 CD  CE2   gb_27
OE1  HE1   gb_1
CE2   NZ   gb_21
 NZ  HZ1   gb_2
 NZ  HZ2   gb_2
  C    O   gb_5
  C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD  OE1    ga_13
  CG   CD  CE2    ga_13
 OE1   CD  CE2    ga_13
  CD  OE1  HE1    ga_12
  CD  CE2   NZ    ga_15
 CE2   NZ  HZ1    ga_11
 CE2   NZ  HZ2    ga_11
 HZ1   NZ  HZ2    ga_10
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CD  CE2  OE1   CG    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD  CE2    gd_34
  CG   CD  OE1  HE1    gd_23
  CG   CD  CE2   NZ    gd_34
  CD  CE2   NZ  HZ1    gd_29

; 5-hydroxylysine (0,S)
[ K5H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH1    0.26600    3 ; from the hydroxyl group of THR
 OE1   OA   -0.67400    3 ; from the hydroxyl group of THR
 HE1    H    0.40800    3 ; from the hydroxyl group of THR
 CE2  CH2   -0.24000    4
  NZ   NT   -0.64000    4
 HZ1    H    0.44000    4
 HZ2    H    0.44000    4
  C    C     0.450     5
  O    O    -0.450     5
 [ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG   CD   gb_27
 CD  OE1   gb_18
 CD  CE2   gb_27
OE1  HE1   gb_1
CE2   NZ   gb_21
 NZ  HZ1   gb_2
 NZ  HZ2   gb_2
  C    O   gb_5
  C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD  OE1    ga_13
  CG   CD  CE2    ga_13
 OE1   CD  CE2    ga_13
  CD  OE1  HE1    ga_12
  CD  CE2   NZ    ga_15
 CE2   NZ  HZ1    ga_11
 CE2   NZ  HZ2    ga_11
 HZ1   NZ  HZ2    ga_10
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CD  OE1  CE2   CG    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD  CE2    gd_34
  CG   CD  OE1  HE1    gd_23
  CG   CD  CE2   NZ    gd_34
  CD  CE2   NZ  HZ1    gd_29

; 5-hydroxylysine (+1,R)
[ KPH ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH1    0.26600    3 ; from the hydroxyl group of THR
 OE1   OA   -0.67400    3 ; from the hydroxyl group of THR
 HE1    H    0.40800    3 ; from the hydroxyl group of THR
 CE2  CH2    0.12700    4
  NZ   NL    0.12900    4
 HZ1    H    0.24800    4
 HZ2    H    0.24800    4
 HZ3    H    0.24800    4
  C    C     0.450     5
  O    O    -0.450     5
 [ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG   CD   gb_27
 CD  OE1   gb_18
 CD  CE2   gb_27
OE1  HE1   gb_1
CE2   NZ   gb_21
 NZ  HZ1   gb_2
 NZ  HZ2   gb_2
 NZ  HZ3   gb_2
  C    O   gb_5
  C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD  OE1    ga_13
  CG   CD  CE2    ga_13
 OE1   CD  CE2    ga_13
  CD  OE1  HE1    ga_12
  CD  CE2   NZ    ga_15
 CE2   NZ  HZ1    ga_11
 CE2   NZ  HZ2    ga_11
 CE2   NZ  HZ3    ga_11
```

```
  HZ1   NZ   HZ2    ga_10
  HZ1   NZ   HZ3    ga_10
  HZ2   NZ   HZ3    ga_10
  CA    C    O      ga_30
  CA    C    +N     ga_19
  O     C    +N     ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   H    gi_1
  CA    N    C   CB   gi_2
  CD   CE2  OE1  CG   gi_2
   C   CA   +N   O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG   gd_34
   N   CA   C    +N   gd_45
   N   CA   C    +N   gd_42
  CA   CB   CG   CD   gd_34
  CB   CG   CD   CE2  gd_34
  CG   CD   OE1  HE1  gd_23
  CG   CD   CE2  NZ   gd_34
  CD   CE2  NZ   HZ1  gd_29

; 5-hydroxylysine (+1,S)
[ KHP ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH2    0.00000    2
  CD   CH1    0.26600    3 ; from the hydroxyl group of THR
 OE1    OA   -0.67400    3 ; from the hydroxyl group of THR
 HE1    H    0.40800    3 ; from the hydroxyl group of THR
 CE2   CH2    0.12700    4
  NZ    NL    0.12900    4
 HZ1    H    0.24800    4
 HZ2    H    0.24800    4
 HZ3    H    0.24800    4
   C    C     0.450    5
   O    O    -0.450    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD  OE1   gb_18
  CD  CE2   gb_27
 OE1  HE1   gb_1
 CE2   NZ   gb_21
  NZ  HZ1   gb_2
  NZ  HZ2   gb_2
  NZ  HZ3   gb_2
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD  OE1   ga_13
  CG   CD  CE2   ga_13
 OE1   CD  CE2   ga_13
  CD  OE1  HE1   ga_12
  CD  CE2   NZ   ga_15
 CE2   NZ  HZ1   ga_11
 CE2   NZ  HZ2   ga_11
 CE2   NZ  HZ3   ga_11
 HZ1   NZ  HZ2   ga_10
 HZ1   NZ  HZ3   ga_10
 HZ2   NZ  HZ3   ga_10
  CA    C   O     ga_30
  CA    C   +N    ga_19
  O     C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
  CA    N    C   CB  gi_2
  CD  OE1  CE2   CG   gi_2
   C   CA   +N   O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG   gd_34
   N   CA   C    +N   gd_45
   N   CA   C    +N   gd_42
  CA   CB   CG   CD   gd_34
  CB   CG   CD  CE2   gd_34
  CG   CD  OE1  HE1   gd_23
  CG   CD  CE2   NZ   gd_34
  CD  CE2   NZ  HZ1   gd_29

; 3,4-dihydroxyphenylalanine
[ HTY ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C     0.00000    1
 CD1    C    -0.14000    2
 HD1   HC     0.14000    2
 CD2    C    -0.14000    3
 HD2   HC     0.14000    3
 CE1    C     0.20300    4 ; from the hydroxyl group of TYR
 OZ1   OA    -0.61100    4 ; from the hydroxyl group of TYR
 HZ1    H     0.40800    4 ; from the hydroxyl group of TYR
 CE2    C    -0.14000    5
 HE2   HC     0.14000    5
 CZ2    C     0.20300    6 ; from the hydroxyl group of TYR
  OH   OA    -0.61100    6 ; from the hydroxyl group of TYR
  HH    H     0.40800    6 ; from the hydroxyl group of TYR
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  OZ1   gb_13
 CE1  CZ2   gb_16
 OZ1  HZ1   gb_1
 CE2  HE2   gb_3
 CE2  CZ2   gb_16
 CZ2   OH   gb_13
  OH   HH   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  OZ1
  CG  HE2
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1   OH
 HD1  CD2
 HD1  OZ1
 HD1  CZ2
 CD2  CE1
 CD2   OH
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 OZ1  CE2
 OZ1   OH
 HE2   OH
  HH  OZ1
  HH  HZ1
 HZ1   OH
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1   ga_27
  CB   CG  CD2   ga_27
 CD1   CG  CD2   ga_27
  CG  CD1  HD1   ga_25
```

```
  CG   CD1   CE1    ga_27
  HD1  CD1   CE1    ga_25
  CG   CD2   HD2    ga_25
  CG   CD2   CE2    ga_27
  HD2  CD2   CE2    ga_25
  CD1  CE1   OZ1    ga_27
  CD1  CE1   CZ2    ga_27
  OZ1  CE1   CZ2    ga_27
  CE1  OZ1   HZ1    ga_12
  CD2  CE2   HE2    ga_25
  CD2  CE2   CZ2    ga_27
  HE2  CE2   CZ2    ga_25
  CE1  CZ2   CE2    ga_27
  CE1  CZ2   OH     ga_27
  CE2  CZ2   OH     ga_27
  CZ2  OH    HH     ga_12
  CA   C     O      ga_30
  CA   C     +N     ga_19
   O   C     +N     ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C    CA    H    gi_1
  CA    N    C     CB   gi_2
  CG   CD1   CD2   CB    gi_1
  CG   CD1   CE1   CZ2   gi_1
  CG   CD2   CE2   CZ2   gi_1
  CD1   CG   CD2   CE2   gi_1
  CD1   CG   CE1   HD1   gi_1
  CD1  CE1   CZ2   CE2   gi_1
  CD2   CG   CD1   CE1   gi_1
  CD2   CG   CE2   HD2   gi_1
  CD2  CE2   CZ2   CE1   gi_1
  CE1  CZ2   CD1   OZ1   gi_1
  HE2  CD2   CZ2   CE2   gi_1
  CZ2  CE1   CE2   OH    gi_1
   C   CA    +N     O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N    CA    gd_14
  -C    N    CA   C     gd_44
  -C    N    CA   C     gd_43
   N   CA    CB   CG    gd_34
   N   CA    C    +N    gd_45
   N   CA    C    +N    gd_42
  CA   CB    CG   CD1   gd_40
  CD1  CE1   OZ1  HZ1   gd_11
  CE1  CZ2   OH   HH    gd_11

; 7-hydroxytryptophan
[ W7H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG    C   -0.21000    2
  CD1   C   -0.14000    2
  HD1   HC   0.14000    2
  CD2   C    0.00000    2
  NE1   NR  -0.10000    2
  HE1   H    0.31000    2
  CE2   C    0.00000    2
  CE3   C   -0.14000    3
  HE3   HC   0.14000    3
  CZ2   C    0.20300    4 ; from the hydroxyl group of TYR
  OH2   OA  -0.61100    4 ; from the hydroxyl group of TYR
  HH2   H    0.40800    4 ; from the hydroxyl group of TYR
  CZ3   C   -0.14000    5
  HZ3   HC   0.14000    5
  CH3   C   -0.14000    6
  HH3   HC   0.14000    6
   C    C     0.450    7
   O    O    -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD1  gb_10
  CG   CD2  gb_16
  CD1  HD1  gb_3
  CD1  NE1  gb_10
  CD2  CE2  gb_16
  CD2  CE3  gb_16
  NE1  HE1  gb_2
  NE1  CE2  gb_10
  CE2  CZ2  gb_16
  CE3  HE3  gb_3
  CE3  CZ3  gb_16
  CZ2  OH2  gb_13
  OH2  HH2  gb_1
  CZ2  CH3  gb_16
  CZ3  HZ3  gb_3
  CZ3  CH3  gb_16
  CH3  HH3  gb_3
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB   HD1
  CB   NE1
  CB   CE2
  CB   CE3
  CG   HE1
  CG   HE3
  CG   CZ2
  CG   CZ3
  CD1  CE3
  CD1  CZ2
  HD1  CD2
  HD1  HE1
  HD1  CE2
  CD2  HE1
  CD2  OH2
  CD2  HZ3
  CD2  CH3
  NE1  CE3
  NE1  OH2
  NE1  CH3
  HE1  CZ2
  CE2  HE3
  CE2  CZ3
  CE2  HH3
  CE3  CZ2
  CE3  HH3
  HE3  HZ3
  HE3  CH3
  CZ2  HZ3
  OH2  CZ3
  OH2  HH3
  HZ3  HH3
 [ angles ]
;  ai   aj   ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG  CD1    ga_37
   CB   CG  CD2    ga_37
  CD1   CG  CD2    ga_7
   CG  CD1  HD1    ga_36
   CG  CD1  NE1    ga_7
  HD1  CD1  NE1    ga_36
   CG  CD2  CE2    ga_7
   CG  CD2  CE3    ga_39
  CE2  CD2  CE3    ga_27
  CD1  NE1  HE1    ga_36
  CD1  NE1  CE2    ga_7
  HE1  NE1  CE2    ga_36
  CD2  CE2  NE1    ga_7
  CD2  CE2  CZ2    ga_27
  NE1  CE2  CZ2    ga_39
  CD2  CE3  HE3    ga_25
  CD2  CE3  CZ3    ga_27
  HE3  CE3  CZ3    ga_25
  CE2  CZ2  OH2    ga_27
  CE2  CZ2  CH3    ga_27
  OH2  CZ2  CH3    ga_27
  HH2  OH2  CZ2    ga_12
  CE3  CZ3  HZ3    ga_25
  CE3  CZ3  CH3    ga_27
  HZ3  CZ3  CH3    ga_25
  CZ2  CH3  CZ3    ga_27
  CZ2  CH3  HH3    ga_25
  CZ3  CH3  HH3    ga_25
   CA   C    O     ga_30
   CA   C   +N     ga_19
    O   C   +N     ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C    CA    H    gi_1
  CA    N    C     CB   gi_2
  CG   CD1   CD2   CB    gi_1
  CG   CD1   NE1   CE2   gi_1
  CG   CD2   CE2   NE1   gi_1
  CD1   CG   CD2   CE2   gi_1
  CD1   CG   NE1   HD1   gi_1
  CD1  NE1   CE2   CD2   gi_1
  CD2   CG   CD1   NE1   gi_1
  CD2  CE2   CE3   CG    gi_1
  CD2  CE2   CZ2   CH3   gi_1
  CD2  CE3   CZ3   CH3   gi_1
```

```
 NE1  CD1  CE2  HE1   gi_1
 CE2  CD2  CE3  CZ3   gi_1
 CE2  CD2  CZ2  NE1   gi_1
 CE2  CZ2  CH3  CZ3   gi_1
 CE3  CD2  CE2  CZ2   gi_1
 CE3  CD2  CZ3  HE3   gi_1
 CE3  CZ3  CH3  CZ2   gi_1
 CZ2  CE2  CH3  OH2   gi_1
 CZ3  CE3  CH3  HZ3   gi_1
 CH3  CZ2  CZ3  HH3   gi_1
  C   CA   +N   O    gi_1
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG    gd_34
   N   CA   C   +N    gd_45
   N   CA   C   +N    gd_42
  CA   CB   CG  CD2    gd_40
 CE2  CZ2  OH2  HH2    gd_11

; 3-hydroxyaspartate (-1,R)
[ DH3 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1   OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1    H    0.40800    2 ; from the hydroxyl group of THR
 CG2    C    0.27000    3
 OD1   OM   -0.63500    3
 OD2   OM   -0.63500    3
   C    C      0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1  HG1   gb_1
 CG2  OD1   gb_6
 CG2  OD2   gb_6
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_13
 OG1   CB  CG2    ga_13
  CB  OG1  HG1    ga_12
  CB  CG2  OD1    ga_22
  CB  CG2  OD2    ga_22
 OD1  CG2  OD2    ga_38
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  CG2  OG1   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB  CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB  OG1  HG1    gd_23
  CA   CB  CG2  OD1    gd_40

; 3-hydroxyaspartate (-1,S)
[ D3H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1   OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1    H    0.40800    2 ; from the hydroxyl group of THR
 CG2    C    0.27000    3
 OD1   OM   -0.63500    3
 OD2   OM   -0.63500    3
   C    C      0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1  HG1   gb_1
 CG2  OD1   gb_6
 CG2  OD2   gb_6
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_13
 OG1   CB  CG2    ga_13
  CB  OG1  HG1    ga_12
  CB  CG2  OD1    ga_22
  CB  CG2  OD2    ga_22
 OD1  CG2  OD2    ga_38
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_2
 CG2  OD1  OD2   CB    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB  CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB  OG1  HG1    gd_23
  CA   CB  CG2  OD1    gd_40

; 3-hydroxyaspartate (0,R)
[ DN3 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH1    0.26600    2 ; from the hydroxyl group of THR
 OG1   OA   -0.67400    2 ; from the hydroxyl group of THR
 HG1    H    0.40800    2 ; from the hydroxyl group of THR
 CG2    C    0.33000    3
 OD1    O   -0.45000    3
 OD2   OA   -0.28800    3
 HD2    H    0.40800    3
   C    C      0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_18
  CB  CG2   gb_27
 OG1  HG1   gb_1
 CG2  OD1   gb_5
 CG2  OD2   gb_13
 OD2  HD2   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_13
  CA   CB  CG2    ga_13
 OG1   CB  CG2    ga_13
  CB  OG1  HG1    ga_12
```

```
  CB   CG2   OD1    ga_30
  CB   CG2   OD2    ga_19
  OD1   CG2   OD2    ga_33
  CG2   OD2   HD2    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB   CG2   OG1   CA    gi_2
  CG2   OD1   OD2   CB    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   OG1   HG1    gd_23
  CA   CB   CG2   OD2    gd_40
  CB   CG2   OD2   HD2    gd_12


; 3-hydroxyaspartate (0,S)
[ D3N ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH1    0.26600    2 ; from the hydroxyl group of THR
  OG1    OA   -0.67400    2 ; from the hydroxyl group of THR
  HG1    H    0.40800    2 ; from the hydroxyl group of THR
  CG2    C    0.33000    3
  OD1    O   -0.45000    3
  OD2    OA   -0.28800    3
  HD2    H    0.40800    3
   C    C     0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   OG1   gb_18
  CB   CG2   gb_27
  OG1   HG1   gb_1
  CG2   OD1   gb_5
  CG2   OD2   gb_13
  OD2   HD2   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   OG1    ga_13
  CA   CB   CG2    ga_13
  OG1   CB   CG2    ga_13
  CB   OG1   HG1    ga_12
  CB   CG2   OD1    ga_30
  CB   CG2   OD2    ga_19
  OD1   CG2   OD2    ga_33
  CG2   OD2   HD2    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB   OG1   CG2   CA    gi_2
  CG2   OD1   OD2   CB    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   OG1   HG1    gd_23
  CA   CB   CG2   OD2    gd_40
  CB   CG2   OD2   HD2    gd_12

; 3-hydroxyasparagine (R)
[ N3H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH1    0.26600    2 ; from the hydroxyl group of THR
  OG1    OA   -0.67400    2 ; from the hydroxyl group of THR
  HG1    H    0.40800    2 ; from the hydroxyl group of THR
  CG2    C    0.29000    3
  OD1    O   -0.45000    3
  ND2   NT   -0.72000    3
 HD21    H    0.44000    3
 HD22    H    0.44000    3
   C    C     0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   OG1   gb_18
  CB   CG2   gb_27
  OG1   HG1   gb_1
  CG2   OD1   gb_5
  CG2   ND2   gb_9
  ND2  HD21   gb_2
  ND2  HD22   gb_2
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   OG1    ga_13
  CA   CB   CG2    ga_13
  OG1   CB   CG2    ga_13
  CB   OG1   HG1    ga_12
  CB   CG2   OD1    ga_30
  CB   CG2   ND2    ga_19
  OD1   CG2   ND2    ga_33
  CG2   ND2  HD21    ga_23
  CG2   ND2  HD22    ga_23
 HD21   ND2  HD22    ga_24
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB   CG2   OG1   CA    gi_2
  CG2   OD1   ND2   CB    gi_1
  ND2  HD21  HD22   CG2    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   OG1   HG1    gd_23
  CA   CB   CG2   ND2    gd_40
  CB   CG2   ND2  HD21    gd_14

; 3-hydroxyasparagine (S)
[ NH3 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH1    0.26600    2 ; from the hydroxyl group of THR
  OG1    OA   -0.67400    2 ; from the hydroxyl group of THR
  HG1    H    0.40800    2 ; from the hydroxyl group of THR
  CG2    C    0.29000    3
  OD1    O   -0.45000    3
  ND2   NT   -0.72000    3
 HD21    H    0.44000    3
 HD22    H    0.44000    3
   C    C     0.450    4
   O    O    -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   OG1   gb_18
```

```
  CB   CG2   gb_27
  OG1  HG1   gb_1
  CG2  OD1   gb_5
  CG2  ND2   gb_9
  ND2  HD21  gb_2
  ND2  HD22  gb_2
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB  OG1   ga_13
  CA   CB  CG2   ga_13
  OG1  CB  CG2   ga_13
  CB  OG1  HG1   ga_12
  CB  CG2  OD1   ga_30
  CB  CG2  ND2   ga_19
  OD1 CG2  ND2   ga_33
  CG2 ND2  HD21  ga_23
  CG2 ND2  HD22  ga_23
 HD21 ND2  HD22  ga_24
  CA   C    O    ga_30
  CA   C   +N    ga_19
   O   C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
  CA    N   C   CB   gi_2
  CB  OG1  CG2  CA   gi_2
  CG2 OD1  ND2  CB   gi_1
  ND2 HD21 HD22 CG2  gi_1
   C   CA   +N   O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA  -C    N   CA   gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG2  gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB  OG1  HG1  gd_23
  CA   CB  CG2  ND2  gd_40
  CB   CG2 ND2  HD21 gd_14

; 4-carboxyglutamate (-2)
[ ECA ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH1    0.00000    1
  CD1   C     0.27000    2
  OE1   OM   -0.63500    2
  OE2   OM   -0.63500    2
  CD2   C     0.27000    3 ; from GLU
  OE3   OM   -0.63500    3 ; from GLU
  OE4   OM   -0.63500    3 ; from GLU
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_27
  CD1 OE1   gb_6
  CD1 OE2   gb_6
  CG  CD2   gb_27
  CD2 OE3   gb_6
  CD2 OE4   gb_6
   C   O    gb_5
   C  +N    gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  CD1   ga_13
  CB   CG  CD2   ga_13
  CD1  CG  CD2   ga_13
  CG  CD1  OE1   ga_22
  CG  CD1  OE2   ga_22
  OE1 CD1  OE2   ga_38
  CG  CD2  OE3   ga_22
  CG  CD2  OE4   ga_22
  OE3 CD2  OE4   ga_38
  CA   C   O     ga_30
  CA   C  +N     ga_19
   O   C  +N     ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
  CA    N   C   CB   gi_2
  CG  CD1  CD2  CB   gi_2
  CD1 OE1  OE2  CG   gi_1
  CD2 OE3  OE4  CG   gi_1
   C   CA   +N   O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA  -C    N   CA   gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  CD1   gd_34
  CB   CG  CD1  OE2   gd_40
  CB   CG  CD2  OE4   gd_40

; 4-carboxyglutamate (-1)
[ ECN ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  CG   CH1    0.00000    1
  CD1   C     0.33000    2 ; from GLUH
  OE1   O    -0.45000    2 ; from GLUH
  OE2   OA   -0.28800    2 ; from GLUH
  HE2   H     0.40800    2 ; from GLUH
  CD2   C     0.27000    3
  OE3   OM   -0.63500    3
  OE4   OM   -0.63500    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_27
  CD1 OE1   gb_5
  CD1 OE2   gb_13
  OE2 HE2   gb_1
  CG  CD2   gb_27
  CD2 OE3   gb_6
  CD2 OE4   gb_6
   C   O    gb_5
   C  +N    gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  CD1   ga_13
  CB   CG  CD2   ga_13
  CD1  CG  CD2   ga_13
  CG  CD1  OE1   ga_30
  CG  CD1  OE2   ga_19
  OE1 CD1  OE2   ga_33
  CD1 OE2  HE2   ga_12
  CG  CD2  OE3   ga_22
  CG  CD2  OE4   ga_22
  OE3 CD2  OE4   ga_38
  CA   C   O     ga_30
  CA   C  +N     ga_19
   O   C  +N     ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
  CA    N   C   CB   gi_2
  CG  CD1  CD2  CB   gi_2
  CD1 OE1  OE2  CG   gi_1
  CD2 OE3  OE4  CG   gi_1
   C   CA   +N   O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
  -CA  -C    N   CA   gd_14
  -C    N   CA   C    gd_44
```

```
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_34
  CB   CG  CD1  OE2    gd_40
  CG  CD1  OE2  HE2    gd_12
  CB   CG  CD2  OE4    gd_40

; sulfotyrosine
[ YSU ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C   -0.14000    4
 HE1   HC    0.14000    4
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
  CZ    C    0.15000    6 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  OH   OA   -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
  ST    S    1.11500    6 ; to add up to -1 net charge
 OI1   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI3   OM   -0.63500    6 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C    0.450    7
   O    O   -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  HE1   gb_3
 CE1   CZ   gb_16
 CE2  HE2   gb_3
 CE2   CZ   gb_16
  CZ   OH   gb_13
  OH   ST   gb_25
  ST  OI1   gb_25
  ST  OI2   gb_25
  ST  OI3   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE1
  CG  HE2
  CG   CZ
 CD1  HD2
 CD1  CE2
 CD1   OH
 HD1  CD2
 HD1  HE1
 HD1   CZ
 CD2  CE1
 CD2   OH
 HD2  HE2
 HD2   CZ
 CE1  HE2
 HE1  CE2
 HE1   OH
 HE2   OH
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_27
  CB   CG  CD2    ga_27
 CD1   CG  CD2    ga_27
  CG  CD1  HD1    ga_25
  CG  CD1  CE1    ga_27
 HD1  CD1  CE1    ga_25
  CG  CD2  HD2    ga_25
  CG  CD2  CE2    ga_27
 HD2  CD2  CE2    ga_25
 CD1  CE1  HE1    ga_25
 CD1  CE1   CZ    ga_27
 HE1  CE1   CZ    ga_25
 CD2  CE2  HE2    ga_25
 CD2  CE2   CZ    ga_27
 HE2  CE2   CZ    ga_25
 CE1   CZ  CE2    ga_27
 CE1   CZ   OH    ga_27
 CE2   CZ   OH    ga_27
  CZ   OH   ST    ga_12
  OH   ST  OI1    ga_13
  OH   ST  OI2    ga_13
  OH   ST  OI3    ga_13
 OI1   ST  OI2    ga_13
 OI1   ST  OI3    ga_13
 OI2   ST  OI3    ga_13
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
  CG  CD1  CE1   CZ    gi_1
  CG  CD2  CE2   CZ    gi_1
 CD1   CG  CD2  CE2    gi_1
 CD1   CG  CE1  HD1    gi_1
 CD1  CE1   CZ  CE2    gi_1
 CD2   CG  CD1  CE1    gi_1
 CD2   CG  CE2  HD2    gi_1
 CD2  CE2   CZ  CE1    gi_1
 HE1  CD1   CZ  CE1    gi_1
 HE2  CD2   CZ  CE2    gi_1
  CZ  CE1  CE2   OH    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_40
 CE1   CZ   OH   ST    gd_11
  CZ   OH   ST  OI1    gd_40


; dehydroalanine
[ SDH ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH0    0.00000    1 ; from aliphatic carbon atoms
  CB  CH2    0.00000    1 ; from aliphatic carbon atoms
   C    C     0.450    2
   O    O    -0.450    2
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_10  ; a shorter bond type to mimic double bond properties
  CA    C   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_27
   N   CA    C    ga_27
  CB   CA    C    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
```

```
; 2,3-didehydrobutyrine
[ TDH ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH0    0.00000    1 ; from aliphatic carbon atoms
  CB   CH1    0.00000    1 ; from aliphatic carbon atoms
  CG   CH3    0.00000    2
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_10  ; a shorter bond type to mimic double bond properties
  CA    C   gb_27
  CB   CG   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_27
   N   CA    C    ga_27
  CB   CA    C    ga_27
  CA   CB   CG    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_14
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42

; 6-bromotryptophan
[ WBR ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.21000    2
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C    0.00000    2
 NE1   NR   -0.10000    2
 HE1    H    0.31000    2
 CE2    C    0.00000    2
 CE3    C   -0.14000    3
 HE3   HC    0.14000    3
 CZ2    C   -0.14000    4
 HZ2   HC    0.14000    4
 CZ3    C   -0.14000    5
 HZ3   HC    0.14000    5
 CH2    C    0.05500    6 ; from 8-bromo-guanosine reported by Hritz and
Oostenbrink. DOI:10.1021/jp902968m
 BRT   BR   -0.05500    6 ; from 8-bromo-guanosine reported by Hritz and
Oostenbrink. DOI:10.1021/jp902968m
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_10
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  NE1   gb_10
 CD2  CE2   gb_16
 CD2  CE3   gb_16
 NE1  HE1   gb_2
 NE1  CE2   gb_10
 CE2  CZ2   gb_16
 CE3  HE3   gb_3
 CE3  CZ3   gb_16
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_16
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_16
 CH2  BRT   gb_42
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2  BRT
 CE3  CZ2
 CE3  BRT
 HE3  HZ3
 HE3  CH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2  BRT
 HZ3  BRT
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_37
  CB   CG  CD2    ga_37
 CD1   CG  CD2    ga_7
  CG  CD1  HD1    ga_36
  CG  CD1  NE1    ga_7
 HD1  CD1  NE1    ga_36
  CG  CD2  CE2    ga_7
  CG  CD2  CE3    ga_39
 CE2  CD2  CE3    ga_27
 CD1  NE1  HE1    ga_36
 CD1  NE1  CE2    ga_7
 HE1  NE1  CE2    ga_36
 CD2  CE2  NE1    ga_7
 CD2  CE2  CZ2    ga_27
 NE1  CE2  CZ2    ga_39
 CD2  CE3  HE3    ga_25
 CD2  CE3  CZ3    ga_27
 HE3  CE3  CZ3    ga_25
 CE2  CZ2  HZ2    ga_25
 CE2  CZ2  CH2    ga_27
 HZ2  CZ2  CH2    ga_25
 CE3  CZ3  HZ3    ga_25
 CE3  CZ3  CH2    ga_27
 HZ3  CZ3  CH2    ga_25
 CZ2  CH2  CZ3    ga_27
 CZ2  CH2  BRT    ga_27
 CZ3  CH2  BRT    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
  CG  CD1  NE1  CE2   gi_1
  CG  CD2  CE2  NE1   gi_1
 CD1   CG  CD2  CE2   gi_1
 CD1   CG  NE1  HD1   gi_1
 CD1  NE1  CE2  CD2   gi_1
 CD2   CG  CD1  NE1   gi_1
 CD2  CE2  CE3   CG   gi_1
 CD2  CE2  CZ2  CH2   gi_1
 CD2  CE3  CZ3  CH2   gi_1
 NE1  CD1  CE2  HE1   gi_1
 CE2  CD2  CE3  CZ3   gi_1
 CE2  CD2  CZ2  NE1   gi_1
 CE2  CZ2  CH2  CZ3   gi_1
```

```
  CE3   CD2   CE2   CZ2    gi_1
  CE3   CD2   CZ3   HE3    gi_1
  CE3   CZ3   CH2   CZ2    gi_1
  CZ2   CE2   CH2   HZ2    gi_1
  CZ3   CE3   CH2   HZ3    gi_1
  CH2   CZ2   CZ3   BRT    gi_1
   C    CA    +N    O     gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
  -CA   -C    N    CA    gd_14
   -C    N   CA    C     gd_44
   -C    N   CA    C     gd_43
    N   CA   CB    CG    gd_34
    N   CA    C    +N    gd_45
    N   CA    C    +N    gd_42
   CA   CB   CG   CD2    gd_40

; S-nitrosocysteine
[ CSN ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   SG    S    0.10000    2 ; to add up to 0 net charge
   ND   NR    0.35000    2 ; to add up to 0 net charge
   OE    O   -0.45000    2 ; from the carbonyl group (of e.g., the peptide bond)
    C    C      0.450    3
    O    O     -0.450    3
 [ bonds ]
    N    H    gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   SG   gb_32
   SG   ND   gb_31   ; from MET
   ND   OE   gb_5    ; from the carbonyl group (of e.g., the peptide bond)
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   SG    ga_16
   CB   SG   ND    ga_6
   SG   ND   OE    ga_26
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
    N   -C   CA    H    gi_1
   CA    N    C   CB    gi_2
    C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C     gd_44
   -C    N   CA    C     gd_43
    N   CA   CB   SG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   SG   ND    gd_26
   CB   SG   ND   OE    gd_21

; citrulline
[ RCI ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   NE    N   -0.31000    3
   HE    H    0.31000    3
   CZ    C    0.29000    4 ; from ASN/GLN
  OH1    O   -0.45000    4 ; from ASN/GLN
  NH2   NT   -0.72000    4 ; from ASN/GLN
 HH21    H    0.44000    4 ; from ASN/GLN
 HH22    H    0.44000    4 ; from ASN/GLN
    C    C      0.450    5
    O    O     -0.450    5
 [ bonds ]
    N    H    gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   NE   gb_21
   NE   HE   gb_2
   NE   CZ   gb_10
   CZ  OH1   gb_5
   CZ  NH2   gb_9
  NH2 HH21   gb_2
  NH2 HH22   gb_2
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
   CG   CD   NE    ga_13
   CD   NE   HE    ga_18
   CD   NE   CZ    ga_31
   HE   NE   CZ    ga_32
   NE   CZ  OH1    ga_30
   NE   CZ  NH2    ga_22
  OH1   CZ  NH2    ga_30
   CZ  NH2 HH21    ga_23
   CZ  NH2 HH22    ga_23
 HH21  NH2 HH22    ga_24
   CA    C    O    ga_30
   CA    C   +N    ga_19
    O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
    N   -C   CA    H    gi_1
   CA    N    C   CB    gi_2
   NE   CD   CZ   HE    gi_1
   CZ  OH1  NH2   NE    gi_1
  NH2 HH21 HH22   CZ    gi_1
    C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
  -CA   -C    N   CA    gd_14
   -C    N   CA    C     gd_44
   -C    N   CA    C     gd_43
    N   CA   CB   CG    gd_34
    N   CA    C   +N    gd_45
    N   CA    C   +N    gd_42
   CA   CB   CG   CD    gd_34
   CB   CG   CD   NE    gd_34
   CG   CD   NE   CZ    gd_39
   CD   NE   CZ  NH2    gd_14
   NE   CZ  NH2 HH21    gd_14

; allysine (aminoadipic semialdehyde)
[ KAL ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG  CH2    0.00000    2
   CD  CH2    0.00000    2
   CE    C    0.35000    3 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
   HE   HC    0.10000    3 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
   OZ    O   -0.45000    3 ; from the carbonyl group (of e.g., GLN)
    C    C      0.450    4
    O    O     -0.450    4
 [ bonds ]
    N    H    gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG   CD   gb_27
   CD   CE   gb_27
   CE   HE   gb_3
   CE   OZ   gb_5
    C    O   gb_5
    C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
   -C    N    H    ga_32
   -C    N   CA    ga_31
    H    N   CA    ga_18
    N   CA   CB    ga_13
    N   CA    C    ga_13
   CB   CA    C    ga_13
   CA   CB   CG    ga_15
   CB   CG   CD    ga_15
```

```
  CG   CD   CE    ga_15
  CD   CE   HE    ga_25
  CD   CE   OZ    ga_27
  HE   CE   OZ    ga_25
  CA   C    O     ga_30
  CA   C    +N    ga_19
  O    C    +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   H    gi_1
  CA   N    C   CB    gi_2
   C   CA   +N   O     gi_1
  CE   OZ   CD   HE    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C   N    CA    gd_14
  -C   N    CA   C     gd_44
  -C   N    CA   C     gd_43
   N   CA   CB   CG    gd_34
   N   CA   C    +N    gd_45
   N   CA   C    +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   OZ    gd_40

; N-acetylglucosamine (N4-linked to ASN)
[ NNG ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.45000    2 ; from the peptide bond
 OD1    O   -0.45000    2 ; from the peptide bond
 ND2    N   -0.31000    3 ; from the peptide bond
 HD2    H    0.31000    3 ; from the peptide bond
  C1  CH1    0.24000    4 ; 0.48 charge divided between C1 and C5 to add up to 0 net
charge
  O5   OA   -0.48000    4 ; from carbohydrates (e.g., GALB)
  C5  CH1    0.24000    4 ; 0.48 charge divided between C1 and C5 to add up to 0 net
charge
  C6  CH2    0.23200    5 ; from carbohydrates (e.g., GALB)
  O6   OA   -0.64200    5 ; from carbohydrates (e.g., GALB)
 HO6    H    0.41000    5 ; from carbohydrates (e.g., GALB)
  C2  CH1    0.00000    6 ; from aliphatic carbon atoms
  C4  CH1    0.23200    7 ; from carbohydrates (e.g., GALB)
  O4   OA   -0.64200    7 ; from carbohydrates (e.g., GALB)
 HO4    H    0.41000    7 ; from carbohydrates (e.g., GALB)
  C3  CH1    0.23200    8 ; from carbohydrates (e.g., GALB)
  O3   OA   -0.64200    8 ; from carbohydrates (e.g., GALB)
 HO3    H    0.41000    8 ; from carbohydrates (e.g., GALB)
  N2    N   -0.15000    9 ; by analogy to ASN/GLN and the peptide bond
 HN2    H    0.31000    9 ; from the hydrogen of the peptide bond
  C7    C    0.19000    9 ; by analogy to ASN/GLN, the peptide bond and the aldehyde
group reported by Dolenc et al. DOI: 10.1093/nar/gki195
  O7    O   -0.45000    9 ; from the carbonyl group (of e.g., GLN)
  C8  CH3    0.10000    9 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
   C    C     0.450   10
   O    O    -0.450   10
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  OD1   gb_5
  CG  ND2   gb_9
 ND2  HD2   gb_2
 ND2   C1   gb_21
  C4   C3   gb_26
  C4   C5   gb_26
  C4   O4   gb_20
  O4  HO4   gb_1
  C3   O3   gb_20
  C3   C2   gb_26
  O3  HO3   gb_1
  C2   C1   gb_26
  C2   N2   gb_21
  N2  HN2   gb_2
  N2   C7   gb_9
  C7   O7   gb_5
  C7   C8   gb_27
  C6   O6   gb_20
  C6   C5   gb_26
  O6  HO6   gb_1
  C5   O5   gb_20
  O5   C1   gb_20
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_30
  CB   CG  ND2   ga_19
 OD1   CG  ND2   ga_33
  CG  ND2  HD2   ga_18
  CG  ND2   C1   ga_31
 HD2  ND2   C1   ga_32
 ND2   C1   C2   ga_9
 ND2   C1   O5   ga_9
  C2   C1   O5   ga_9
  C3   C2   N2   ga_9
  C3   C2   C1   ga_8
  N2   C2   C1   ga_9
  C2   N2  HN2   ga_18
  C2   N2   C7   ga_31
 HN2   N2   C7   ga_32
  N2   C7   O7   ga_33
  N2   C7   C8   ga_19
  O7   C7   C8   ga_30
  C4   C3   O3   ga_9
  C4   C3   C2   ga_8
  O3   C3   C2   ga_9
  C3   O3  HO3   ga_12
  C5   C4   O4   ga_9
  C5   C4   C3   ga_8
  O4   C4   C3   ga_9
  C4   O4  HO4   ga_12
  C4   C5   C6   ga_8
  C4   C5   O5   ga_9
  C6   C5   O5   ga_9
  O6   C6   C5   ga_9
  C6   O6  HO6   ga_12
  C5   O5   C1   ga_10
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  OD1  ND2   CB    gi_1
 ND2   CG   C1  HD2    gi_1
  C1   O5  ND2   C2    gi_2
  C5   O5   C6   C4    gi_2
  C2   N2   C3   C1    gi_2
  C3   O3   C2   C4    gi_2
  C4   C3   O4   C5    gi_2
  N2   C7   C2  HN2    gi_1
  C7   C8   N2   O7    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  ND2    gd_40
  CB   CG  ND2   C1    gd_14
  CG  ND2   C1   O5    gd_39
  C1   C2   N2   C7    gd_39
  C2   N2   C7   C8    gd_14
  C3   C2   C1   O5    gd_17
  C3   C2   C1  ND2    gd_17
  C3   C2   C1  ND2    gd_34
  N2   C2   C1  ND2    gd_18
  N2   C2   C1   O5    gd_18
  C2   C3   O3  HO3    gd_30
  C4   C3   C2   N2    gd_17
  C4   C3   C2   C1    gd_34
  O3   C3   C2   N2    gd_18
  O3   C3   C2   C1    gd_17
  C3   C4   O4  HO4    gd_30
  C5   C4   C3   O3    gd_17
  C5   C4   C3   C2    gd_34
  O4   C4   C3   O2    gd_18
  O4   C4   C3   C2    gd_17
  O4   C4   C5   C6    gd_17
  C3   C4   C5   C6    gd_34
  C3   C4   C5   O5    gd_17
  O4   C4   C5   O5    gd_18
  C5   C6   O6  HO6    gd_30
  O6   C6   C5   C4    gd_1
  O6   C6   C5   O5    gd_3
  O6   C6   C5   O5    gd_35
  C4   C5   O5   C1    gd_29
  C5   O5   C1   C2    gd_29
```

```
; 2,3-dihydroxyphenylalanine
[ F23 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C    0.20300    2 ; from the hydroxyl group of TYR
 OE3   OA   -0.61100    2 ; from the hydroxyl group of TYR
 HE3    H    0.40800    2 ; from the hydroxyl group of TYR
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C    0.20300    4 ; from the hydroxyl group of TYR
 OZ1   OA   -0.61100    4 ; from the hydroxyl group of TYR
 HZ1    H    0.40800    4 ; from the hydroxyl group of TYR
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C   -0.14000    6
 HZ2   HC    0.14000    6
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  OE3   gb_13
 CD1  CE1   gb_16
 CD2  HD2    gb_3
 CD2  CE2   gb_16
 CE1  OZ1   gb_13
 CE1  CZ2   gb_16
 CE2  HE2    gb_3
 CE2  CZ2   gb_16
 OE3  HE3    gb_1
 CZ2  HZ2    gb_3
 OZ1  HZ1    gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD2
  CB  CE1
  CB  CE2
  CB  OE3
  CG  HE2
  CG  OZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  HZ2
 CD2  CE1
 CD2  OE3
 CD2  HZ2
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 CE2  OZ1
 HE2  HZ2
 OE3  CZ2
 OE3  OZ1
 HZ2  OZ1
 OE3  HZ1
 HE3  OZ1
 HE3  HZ1
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_27
  CB   CG  CD2    ga_27
 CD1   CG  CD2    ga_27
  CG  CD1  CE1    ga_27
  CG  CD1  OE3    ga_27
 CE1  CD1  OE3    ga_27
  CG  CD2  HD2    ga_25
  CG  CD2  CE2    ga_27
 HD2  CD2  CE2    ga_25
 CD1  CE1  CZ2    ga_27
 CD1  CE1  OZ1    ga_27
 CZ2  CE1  OZ1    ga_27
 CD2  CE2  HE2    ga_25
 CD2  CE2  CZ2    ga_27
 HE2  CE2  CZ2    ga_25
 CD1  OE3  HE3    ga_12
 CE1  CZ2  CE2    ga_27
 CE1  CZ2  HZ2    ga_25
 CE2  CZ2  HZ2    ga_25
 CE1  OZ1  HZ1    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
  CG  CD1  CE1  CZ2    gi_1
  CG  CD2  CE2  CZ2    gi_1
 CD1   CG  CD2  CE2    gi_1
 CD1   CG  CE1  OE3    gi_1
 CD1  CE1  CZ2  CE2    gi_1
 CD2   CG  CD1  CE1    gi_1
 CD2   CG  CE2  HD2    gi_1
 CD2  CE2  CZ2  CE1    gi_1
 OZ1  CD1  CZ2  CE1    gi_1
 HE2  CD2  CZ2  CE2    gi_1
 CZ2  CE1  CE2  HZ2    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_40
  CG  CD1  OE3  HE3    gd_11
 CD1  CE1  OZ1  HZ1    gd_11

; 2-hydroxyphenylalanine
[ F2H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C    0.20300    2 ; from the hydroxyl group of TYR
 OE3   OA   -0.61100    2 ; from the hydroxyl group of TYR
 HE3    H    0.40800    2 ; from the hydroxyl group of TYR
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C   -0.14000    4
 HE1   HC    0.14000    4
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
  CZ    C   -0.14000    6
  HZ   HC    0.14000    6
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  OE3   gb_13
 CD1  CE1   gb_16
 CD2  HD2    gb_3
 CD2  CE2   gb_16
 CE1  HE1    gb_3
 CE1   CZ   gb_16
 CE2  HE2    gb_3
 CE2   CZ   gb_16
 OE3  HE3    gb_1
  CZ   HZ    gb_3
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE1
  CG  HE2
  CB  OE3
  CG   CZ
 CD1  HD2
 CD1  CE2
 CD1   HZ
 CD2  CE1
 CD2  OE3
```

```
  CD2   HZ
  HD2  HE2
  HD2   CZ
  CE1  HE2
  HE1  CE2
  HE1  OE3
  HE1   HZ
  HE2   HZ
  OE3   CZ
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_27
  CB   CG  CD2    ga_27
 CD1   CG  CD2    ga_27
  CG  CD1  CE1    ga_27
  CG  CD1  OE3    ga_27
 CE1  CD1  OE3    ga_27
  CG  CD2  HD2    ga_25
  CG  CD2  CE2    ga_27
 HD2  CD2  CE2    ga_25
 CD1  CE1  HE1    ga_25
 CD1  CE1   CZ    ga_27
 HE1  CE1   CZ    ga_25
 CD2  CE2  HE2    ga_25
 CD2  CE2   CZ    ga_27
 HE2  CE2   CZ    ga_25
 CD1  OE3  HE3    ga_12
 CE1   CZ  CE2    ga_27
 CE1   CZ   HZ    ga_25
 CE2   CZ   HZ    ga_25
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
  CG  CD1  CE1   CZ   gi_1
  CG  CD2  CE2   CZ   gi_1
 CD1   CG  CD2  CE2   gi_1
 CD1   CG  CE1  OE3   gi_1
 CD1  CE1   CZ  CE2   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD2   CG  CE2  HD2   gi_1
 CD2  CE2   CZ  CE1   gi_1
 HE1  CD1   CZ  CE1   gi_1
 HE2  CD2   CZ  CE2   gi_1
  CZ  CE1  CE2   HZ   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
   N   CA    C   +N   gd_45
   N   CA    C   +N   gd_42
  CA   CB   CG  CD1   gd_40
  CG  CD1  OE3  HE3   gd_11

; 3-hydroxyphenylalanine
[ F3H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C    0.20300    4 ; from the hydroxyl group of TYR
 OZ1   OA   -0.61100    4 ; from the hydroxyl group of TYR
 HZ1    H    0.40800    4 ; from the hydroxyl group of TYR
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C   -0.14000    6
 HZ2   HC    0.14000    6
   C    C     0.450    7
   O    O    -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  OZ1   gb_13
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ2   gb_16
 CZ2  HZ2   gb_3
 OZ1  HZ1   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  CZ2
  CG  OZ1
 CD1  HD2
 CD1  CE2
 CD1  HZ2
 HD1  CD2
 HD1  CZ2
 HD1  OZ1
 CD2  CE1
 CD2  HZ2
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 CE2  OZ1
 HE2  HZ2
 HZ2  OZ1
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_27
  CB   CG  CD2    ga_27
 CD1   CG  CD2    ga_27
  CG  CD1  HD1    ga_25
  CG  CD1  CE1    ga_27
 HD1  CD1  CE1    ga_25
  CG  CD2  HD2    ga_25
  CG  CD2  CE2    ga_27
 HD2  CD2  CE2    ga_25
 CD1  CE1  CZ2    ga_27
 CD1  CE1  OZ1    ga_27
 CZ2  CE1  OZ1    ga_27
 CD2  CE2  HE2    ga_25
 CD2  CE2  CZ2    ga_27
 HE2  CE2  CZ2    ga_25
 CE1  CZ2  CE2    ga_27
 CE1  CZ2  HZ2    ga_25
 CE2  CZ2  HZ2    ga_25
 CE1  OZ1  HZ1    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
  CG  CD1  CE1  CZ2   gi_1
  CG  CD2  CE2  CZ2   gi_1
 CD1   CG  CD2  CE2   gi_1
 CD1   CG  CE1  HD1   gi_1
 CD1  CE1  CZ2  CE2   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD2   CG  CE2  HD2   gi_1
 CD2  CE2  CZ2  CE1   gi_1
 OZ1  CD1  CZ2  CE1   gi_1
 HE2  CD2  CZ2  CE2   gi_1
 CZ2  CE1  CE2  HZ2   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA    C   gd_44
  -C    N   CA    C   gd_43
   N   CA   CB   CG   gd_34
```

```
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_40
 CD1  CE1  OZ1  HZ1    gd_11

; 6-hydroxytryptophan
[ W6H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.21000    2
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C    0.00000    2
 NE1   NR   -0.10000    2
 HE1    H    0.31000    2
 CE2    C    0.00000    2
 CE3    C   -0.14000    3
 HE3   HC    0.14000    3
 CZ2    C   -0.14000    4
 HZ2   HC    0.14000    4
 CZ3    C   -0.14000    5
 HZ3   HC    0.14000    5
 CH2    C    0.20300    6 ; from the hydroxyl group of TYR
  OI   OA   -0.61100    6 ; from the hydroxyl group of TYR
  HI    H    0.40800    6 ; from the hydroxyl group of TYR
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_10
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  NE1   gb_10
 CD2  CE2   gb_16
 CD2  CE3   gb_16
 NE1  HE1   gb_2
 NE1  CE2   gb_10
 CE2  CZ2   gb_16
 CE3  HE3   gb_3
 CE3  CZ3   gb_16
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_16
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_16
 CH2   OI   gb_13
  OI   HI   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2   OI
 CE3  CZ2
 CE3   OI
 HE3  HZ3
 HE3  CH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2   OI
 HZ3   OI
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_37
  CB   CG  CD2    ga_37
 CD1   CG  CD2    ga_7
  CG  CD1  HD1    ga_36
  CG  CD1  NE1    ga_7
 HD1  CD1  NE1    ga_36
  CG  CD2  CE2    ga_7
  CG  CD2  CE3    ga_39
 CE2  CD2  CE3    ga_27
 CD1  NE1  HE1    ga_36
 CD1  NE1  CE2    ga_7
 HE1  NE1  CE2    ga_36
 CD2  CE2  NE1    ga_7
 CD2  CE2  CZ2    ga_27
 NE1  CE2  CZ2    ga_39
 CD2  CE3  HE3    ga_25
 CD2  CE3  CZ3    ga_27
 HE3  CE3  CZ3    ga_25
 CE2  CZ2  HZ2    ga_25
 CE2  CZ2  CH2    ga_27
 HZ2  CZ2  CH2    ga_25
 CE3  CZ3  HZ3    ga_25
 CE3  CZ3  CH2    ga_27
 HZ3  CZ3  CH2    ga_25
 CZ2  CH2  CZ3    ga_27
 CZ2  CH2   OI    ga_27
 CZ3  CH2   OI    ga_27
 CH2   OI   HI    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
  CG  CD1  NE1  CE2    gi_1
  CG  CD2  CE2  NE1    gi_1
 CD1   CG  CD2  CE2    gi_1
 CD1   CG  NE1  HD1    gi_1
 CD1  NE1  CE2  CD2    gi_1
 CD2   CG  CD1  NE1    gi_1
 CD2  CE2  CE3   CG    gi_1
 CD2  CE2  CZ2  CH2    gi_1
 CD2  CE3  CZ3  CH2    gi_1
 NE1  CD1  CE2  HE1    gi_1
 CE2  CD2  CE3  CZ3    gi_1
 CE2  CD2  CZ2  NE1    gi_1
 CE2  CZ2  CH2  CZ3    gi_1
 CE3  CD2  CE2  CZ2    gi_1
 CE3  CD2  CZ3  HE3    gi_1
 CE3  CZ3  CH2  CZ2    gi_1
 CZ2  CE2  CH2  HZ2    gi_1
 CZ3  CE3  CH2  HZ3    gi_1
 CH2  CZ2  CZ3   OI    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_40
 CZ2  CH2   OI   HI    gd_11

; 5-hydroxytryptophan
[ W5H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.21000    2
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C    0.00000    2
 NE1   NR   -0.10000    2
 HE1    H    0.31000    2
 CE2    C    0.00000    2
 CE3    C   -0.14000    3
 HE3   HC    0.14000    3
 CZ2    C   -0.14000    4
 HZ2   HC    0.14000    4
 CZ3    C    0.20300    5 ; from the hydroxyl group of TYR
 OH3   OA   -0.61100    5 ; from the hydroxyl group of TYR
 HH3    H    0.40800    5 ; from the hydroxyl group of TYR
```

```
 CH2    C   -0.14000    6
 HH2   HC    0.14000    6
  C    C      0.450    7
  O    O     -0.450    7
[ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG  CD1   gb_10
 CG  CD2   gb_16
CD1  HD1   gb_3
CD1  NE1   gb_10
CD2  CE2   gb_16
CD2  CE3   gb_16
NE1  HE1   gb_2
NE1  CE2   gb_10
CE2  CZ2   gb_16
CE3  HE3   gb_3
CE3  CZ3   gb_16
CZ2  HZ2   gb_3
CZ2  CH2   gb_16
CZ3  CH2   gb_16
CZ3  OH3   gb_13
CH2  HH2   gb_3
OH3  HH3   gb_1
 C    O   gb_5
 C   +N   gb_10
[ exclusions ]
;  ai    aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  OH3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  HE3
 CE2  CZ3
 CE2  HH2
 CE3  CZ2
 CE3  HH2
 HE3  OH3
 HE3  CH2
 CZ2  OH3
 HZ2  CZ3
 HZ2  HH2
 OH3  HH2
[ angles ]
;  ai    aj    ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_37
  CB   CG  CD2    ga_37
 CD1   CG  CD2    ga_7
  CG  CD1  HD1    ga_36
  CG  CD1  NE1    ga_7
 HD1  CD1  NE1    ga_36
  CG  CD2  CE2    ga_7
  CG  CD2  CE3    ga_39
 CE2  CD2  CE3    ga_27
 CD1  NE1  HE1    ga_36
 CD1  NE1  CE2    ga_7
 HE1  NE1  CE2    ga_36
 CD2  CE2  NE1    ga_7
 CD2  CE2  CZ2    ga_27
 NE1  CE2  CZ2    ga_39
 CD2  CE3  HE3    ga_25
 CD2  CE3  CZ3    ga_27
 HE3  CE3  CZ3    ga_25
 CE2  CZ2  HZ2    ga_25
 CE2  CZ2  CH2    ga_27
 HZ2  CZ2  CH2    ga_25
 CE3  CZ3  CH2    ga_27
 CE3  CZ3  OH3    ga_27
 CH2  CZ3  OH3    ga_27
 CZ3  OH3  HH3    ga_12
 CZ2  CH2  CZ3    ga_27
 CZ2  CH2  HH2    ga_25
 CZ3  CH2  HH2    ga_25
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai    aj    ak    al   gromos type
   N   -C    CA    H    gi_1
  CA    N    C    CB    gi_2
  CG  CD1   CD2   CB    gi_1
  CG  CD1   NE1  CE2    gi_1
  CG  CD2   CE2  NE1    gi_1
 CD1   CG  CD2  CE2    gi_1
 CD1   CG  NE1  HD1    gi_1
 CD1  NE1  CE2  CD2    gi_1
 CD2   CG  CD1  NE1    gi_1
 CD2  CE2  CE3   CG    gi_1
 CD2  CE2  CZ2  CH2    gi_1
 CD2  CE3  CZ3  CH2    gi_1
 NE1  CD1  CE2  HE1    gi_1
 CE2  CD2  CE3  CZ3    gi_1
 CE2  CD2  CZ2  NE1    gi_1
 CE2  CZ2  CH2  CZ3    gi_1
 CE3  CD2  CE2  CZ2    gi_1
 CE3  CD2  CZ3  HE3    gi_1
 CE3  CZ3  CH2  CZ2    gi_1
 CZ2  CE2  CH2  HZ2    gi_1
 CZ3  CE3  CH2  OH3    gi_1
 CH2  CZ2  CZ3  HH2    gi_1
  C   CA   +N    O    gi_1
[ dihedrals ]
;  ai    aj    ak    al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_40
 CE3  CZ3  OH3  HH3    gd_11

; 4-hydroxytryptophan
[ W4H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.21000    2
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C    0.00000    2
 NE1   NR   -0.10000    2
 HE1    H    0.31000    2
 CE2    C    0.00000    2
 CE3    C    0.20300    3 ; from the hydroxyl group of TYR
 OZ4   OA   -0.61100    3 ; from the hydroxyl group of TYR
 HZ4    H    0.40800    3 ; from the hydroxyl group of TYR
 CZ2    C   -0.14000    4
 HZ2   HC    0.14000    4
 CZ3    C   -0.14000    5
 HZ3   HC    0.14000    5
 CH2    C   -0.14000    6
 HH2   HC    0.14000    6
  C    C      0.450    7
  O    O     -0.450    7
 [ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG  CD1   gb_10
 CG  CD2   gb_16
CD1  HD1   gb_3
CD1  NE1   gb_10
CD2  CE2   gb_16
CD2  CE3   gb_16
NE1  HE1   gb_2
NE1  CE2   gb_10
CE2  CZ2   gb_16
CE3  OZ4   gb_13
CE3  CZ3   gb_16
CZ2  HZ2   gb_3
CZ2  CH2   gb_16
CZ3  HZ3   gb_3
CZ3  CH2   gb_16
OZ4  HZ4   gb_1
```

```
 CH2  HH2   gb_3
  C   O    gb_5
  C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  NE1
  CB  CE2
  CB  CE3
  CG  HE1
  CG  CZ2
  CG  CZ3
  CG  OZ4
 CD1  CE3
 CD1  CZ2
 HD1  CD2
 HD1  HE1
 HD1  CE2
 CD2  HE1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HZ2
 NE1  CH2
 HE1  CZ2
 CE2  CZ3
 CE2  OZ4
 CE2  HH2
 CE3  CZ2
 CE3  HH2
 CZ2  HZ3
 HZ2  CZ3
 HZ2  HH2
 HZ3  OZ4
 HZ3  HH2
 OZ4  CH2
 [ angles ]
;  ai   aj   ak   gromos type
  -C   N    H    ga_32
  -C   N    CA   ga_31
   H   N    CA   ga_18
   N   CA   CB   ga_13
   N   CA   C    ga_13
  CB   CA   C    ga_13
  CA   CB   CG   ga_15
  CB   CG   CD1  ga_37
  CB   CG   CD2  ga_37
 CD1   CG   CD2  ga_7
  CG  CD1   HD1  ga_36
  CG  CD1   NE1  ga_7
 HD1  CD1   NE1  ga_36
  CG  CD2   CE2  ga_7
  CG  CD2   CE3  ga_39
 CE2  CD2   CE3  ga_27
 CD1  NE1   HE1  ga_36
 CD1  NE1   CE2  ga_7
 HE1  NE1   CE2  ga_36
 CD2  CE2   NE1  ga_7
 CD2  CE2   CZ2  ga_27
 NE1  CE2   CZ2  ga_39
 CD2  CE3   CZ3  ga_27
 CD2  CE3   OZ4  ga_27
 CZ3  CE3   OZ4  ga_27
 CE2  CZ2   HZ2  ga_25
 CE2  CZ2   CH2  ga_27
 HZ2  CZ2   CH2  ga_25
 CE3  CZ3   HZ3  ga_25
 CE3  CZ3   CH2  ga_27
 HZ3  CZ3   CH2  ga_25
 CE3  OZ4   HZ4  ga_12
 CZ2  CH2   CZ3  ga_27
 CZ2  CH2   HH2  ga_25
 CZ3  CH2   HH2  ga_25
  CA   C    O    ga_30
  CA   C    +N   ga_19
   O   C    +N   ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA   H    gi_1
  CA   N    C    CB   gi_2
  CG  CD1   CD2  CB   gi_1
  CG  CD1   NE1  CE2  gi_1
  CG  CD2   CE2  NE1  gi_1
 CD1   CG   CD2  CE2  gi_1
 CD1   CG   NE1  HD1  gi_1
 CD1  NE1   CE2  CD2  gi_1
 CD2   CG   CD1  NE1  gi_1
 CD2  CE2   CE3  CG   gi_1
 CD2  CE2   CZ2  CH2  gi_1
 CD2  CE3   CZ3  CH2  gi_1
 NE1  CD1   CE2  HE1  gi_1
 CE2  CD2   CE3  CZ3  gi_1
 CE2  CD2   CZ2  NE1  gi_1
 CE2  CZ2   CH2  CZ3  gi_1
 CE3  CD2   CE2  CZ2  gi_1
 CE3  CD2   CZ3  OZ4  gi_1
 CE3  CZ3   CH2  CZ2  gi_1
 CZ2  CE2   CH2  HZ2  gi_1
 CZ3  CE3   CH2  HZ3  gi_1
 CH2  CZ2   CZ3  HH2  gi_1
  C   CA   +N   O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C   N    CA   gd_14
  -C   N    CA   C    gd_44
  -C   N    CA   C    gd_43
   N   CA   CB   CG   gd_34
   N   CA   C    +N   gd_45
   N   CA   C    +N   gd_42
  CA   CB   CG   CD2  gd_40
 CD2  CE3  OZ4  HZ4   gd_11

; 2-hydroxytryptophan
[ W2H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C   -0.21000    2
 CD1    C    0.20300    2 ; from the hydroxyl group of TYR
 OE4   OA   -0.61100    2 ; from the hydroxyl group of TYR
 HE4    H    0.40800    2 ; from the hydroxyl group of TYR
 CD2    C    0.00000    2
 NE1   NR   -0.10000    2
 HE1    H    0.31000    2
 CE2    C    0.00000    2
 CE3    C   -0.14000    3
 HE3   HC    0.14000    3
 CZ2    C   -0.14000    4
 HZ2   HC    0.14000    4
 CZ3    C   -0.14000    5
 HZ3   HC    0.14000    5
 CH2    C   -0.14000    6
 HH2   HC    0.14000    6
  C     C    0.450    7
  O     O   -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_10
  CG  CD2   gb_16
 CD1  OE4   gb_13
 CD1  NE1   gb_10
 CD2  CE2   gb_16
 CD2  CE3   gb_16
 NE1  HE1   gb_2
 NE1  CE2   gb_10
 CE2  CZ2   gb_16
 CE3  HE3   gb_3
 CE3  CZ3   gb_16
 OE4  HE4   gb_1
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_16
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_16
 CH2  HH2   gb_3
  C   O    gb_5
  C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  NE1
  CB  CE2
  CB  CE3
  CB  OE4
  CG  HE1
  CG  HE3
  CG  CZ2
  CG  CZ3
 CD1  CE3
 CD1  CZ2
 CD2  HE1
 CD2  OE4
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 NE1  CE3
 NE1  HE4
 NE1  HZ2
 NE1  CH2
 HE1  OE4
 HE1  CZ2
```

```
  CE2   HE3
  CE2   OE4
  CE2   CZ3
  CE2   HH2
  CE3   CZ2
  CE3   HH2
  HE3   HZ3
  HE3   CH2
  CZ2   HZ3
  HZ2   CZ3
  HZ2   HH2
  HZ3   HH2
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_32
  -C    N     CA   ga_31
   H    N     CA   ga_18
   N    CA    CB   ga_13
   N    CA    C    ga_13
  CB    CA    C    ga_13
  CA    CB    CG   ga_15
  CB    CG    CD1  ga_37
  CB    CG    CD2  ga_37
  CD1   CG    CD2  ga_7
  CG    CD1   NE1  ga_7
  CG    CD1   OE4  ga_37
  NE1   CD1   OE4  ga_37
  CG    CD2   CE2  ga_7
  CG    CD2   CE3  ga_39
  CE2   CD2   CE3  ga_27
  CD1   NE1   HE1  ga_36
  CD1   NE1   CE2  ga_7
  HE1   NE1   CE2  ga_36
  CD2   CE2   NE1  ga_7
  CD2   CE2   CZ2  ga_27
  NE1   CE2   CZ2  ga_39
  CD2   CE3   HE3  ga_25
  CD2   CE3   CZ3  ga_27
  HE3   CE3   CZ3  ga_25
  CD1   OE4   HE4  ga_12
  CE2   CZ2   HZ2  ga_25
  CE2   CZ2   CH2  ga_27
  HZ2   CZ2   CH2  ga_25
  CE3   CZ3   HZ3  ga_25
  CE3   CZ3   CH2  ga_27
  HZ3   CZ3   CH2  ga_25
  CZ2   CH2   CZ3  ga_27
  CZ2   CH2   HH2  ga_25
  CZ3   CH2   HH2  ga_25
  CA    C     O    ga_30
  CA    C     +N   ga_19
   O    C     +N   ga_33
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
  CA    N     C    CB   gi_2
  CG    CD1   CD2  CB   gi_1
  CG    CD1   NE1  CE2  gi_1
  CG    CD2   CE2  NE1  gi_1
  CD1   CG    CD2  CE2  gi_1
  CD1   CG    NE1  OE4  gi_1
  CD1   NE1   CE2  CD2  gi_1
  CD2   CG    CD1  NE1  gi_1
  CD2   CE2   CE3  CG   gi_1
  CD2   CE2   CZ2  CH2  gi_1
  CD2   CE3   CZ3  CH2  gi_1
  NE1   CD1   CE2  HE1  gi_1
  CE2   CD2   CE3  CZ3  gi_1
  CE2   CD2   CZ2  NE1  gi_1
  CE2   CZ2   CH2  CZ3  gi_1
  CE3   CD2   CE2  CZ2  gi_1
  CE3   CD2   CZ3  HE3  gi_1
  CE3   CZ3   CH2  CZ2  gi_1
  CZ2   CE2   CH2  HZ2  gi_1
  CZ3   CE3   CH2  HZ3  gi_1
  CH2   CZ2   CZ3  HH2  gi_1
   C    CA    +N   O    gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_14
  -C    N     CA   C    gd_44
  -C    N     CA   C    gd_43
   N    CA    CB   CG   gd_34
   N    CA    C    +N   gd_45
   N    CA    C    +N   gd_42
  CA    CB    CG   CD2  gd_40
  CG    CD1   OE4  HE4  gd_11

; 3-hydroxyleucine (R)
[ L3H ]
 [ atoms ]
   N    N    -0.31000    0
   H    H     0.31000    0
  CA   CH1     0.00000    1
  CB   CH1     0.26600    2 ; from the hydroxyl group of THR
 OG1    OA    -0.67400    2 ; from the hydroxyl group of THR
 HG1     H     0.40800    2 ; from the hydroxyl group of THR
 CG2   CH1     0.00000    3
 CD1   CH3     0.00000    3
 CD2   CH3     0.00000    3
   C     C       0.450    4
   O     O      -0.450    4
 [ bonds ]
   N     H    gb_2
   N     CA   gb_21
  CA     CB   gb_27
  CA     C    gb_27
  CB     OG1  gb_18
  CB     CG2  gb_27
 OG1     HG1  gb_1
 CG2     CD1  gb_27
 CG2     CD2  gb_27
   C     O    gb_5
   C     +N   gb_10
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_32
  -C    N     CA   ga_31
   H    N     CA   ga_18
   N    CA    CB   ga_13
   N    CA    C    ga_13
  CB    CA    C    ga_13
  CA    CB    OG1  ga_13
  CA    CB    CG2  ga_13
 OG1    CB    CG2  ga_13
  CB    OG1   HG1  ga_12
  CB    CG2   CD1  ga_15
  CB    CG2   CD2  ga_15
 CD1    CG2   CD2  ga_15
  CA    C     O    ga_30
  CA    C     +N   ga_19
   O    C     +N   ga_33
 [ impropers ]
; ai    aj    ak   al   gromos type
   N    -C    CA   H    gi_1
  CA    N     C    CB   gi_2
  CB    OG1   CG2  CA   gi_2
  CB    CD1   CD2  CG2  gi_2
   C    CA    +N   O    gi_1
 [ dihedrals ]
; ai    aj    ak   al   gromos type
 -CA    -C    N    CA   gd_14
  -C    N     CA   C    gd_44
  -C    N     CA   C    gd_43
   N    CA    CB   CG2  gd_34
   N    CA    C    +N   gd_45
   N    CA    C    +N   gd_42
  CA    CB    CG2  CD1  gd_34
  CA    CB    OG1  HG1  gd_23

; 3-hydroxyleucine (S)
[ LH3 ]
 [ atoms ]
   N    N    -0.31000    0
   H    H     0.31000    0
  CA   CH1     0.00000    1
  CB   CH1     0.26600    2 ; from the hydroxyl group of THR
 OG1    OA    -0.67400    2 ; from the hydroxyl group of THR
 HG1     H     0.40800    2 ; from the hydroxyl group of THR
 CG2   CH1     0.00000    3
 CD1   CH3     0.00000    3
 CD2   CH3     0.00000    3
   C     C       0.450    4
   O     O      -0.450    4
 [ bonds ]
   N     H    gb_2
   N     CA   gb_21
  CA     CB   gb_27
  CA     C    gb_27
  CB     OG1  gb_18
  CB     CG2  gb_27
 OG1     HG1  gb_1
 CG2     CD1  gb_27
 CG2     CD2  gb_27
   C     O    gb_5
   C     +N   gb_10
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H    ga_32
  -C    N     CA   ga_31
   H    N     CA   ga_18
   N    CA    CB   ga_13
   N    CA    C    ga_13
  CB    CA    C    ga_13
  CA    CB    OG1  ga_13
  CA    CB    CG2  ga_13
```

```
  OG1   CB   CG2    ga_13
  CB   OG1   HG1    ga_12
  CB   CG2   CD1    ga_15
  CB   CG2   CD2    ga_15
 CD1   CG2   CD2    ga_15
  CA    C    O     ga_30
  CA    C   +N     ga_19
   O    C   +N     ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB   CG2   OG1   CA    gi_2
  CB   CD1   CD2   CG2    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG2    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG2   CD1    gd_34
  CA   CB   OG1   HG1    gd_23

; 4-hydroxyleucine
[ L4H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH0    0.26600    2 ; from the hydroxyl group of THR
 OD3   OA   -0.67400    2 ; from the hydroxyl group of THR
 HD3    H    0.40800    2 ; from the hydroxyl group of THR
 CD1  CH3    0.00000    3
 CD2  CH3    0.00000    4
   C    C      0.450    5
   O    O     -0.450    5
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_27
  CG  CD2   gb_27
  CG  OD3   gb_18
 OD3  HD3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_13
  CB   CG  CD2    ga_13
  CB   CG  OD3    ga_13
 CD1   CG  CD2    ga_13
 CD1   CG  OD3    ga_13
 CD2   CG  OD3    ga_13
  CG  OD3  HD3    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  OD3    gd_34
  CB   CG  OD3  HD3    gd_23

; 5-hydroxyleucine (R)
[ L5H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH1    0.00000    2
 CD1  CH3    0.00000    2
 CD2  CH2    0.26600    3 ; from the hydroxyl group of THR
  OE   OA   -0.67400    3 ; from the hydroxyl group of THR
  HE    H    0.40800    3 ; from the hydroxyl group of THR
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_27
  CG  CD2   gb_27
 CD2   OE   gb_18
  OE   HE   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_15
  CB   CG  CD2    ga_15
 CD1   CG  CD2    ga_15
 CD1   CG  OE    ga_13
 CD2   OE   HE    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  CD1  CD2   CG    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_34
  CB   CG  CD2   OE    gd_34
  CG  CD2   OE   HE    gd_23

; 5-hydroxyleucine (S)
[ LH5 ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH1    0.00000    2
 CD1  CH3    0.00000    2
 CD2  CH2    0.26600    3 ; from the hydroxyl group of THR
  OE   OA   -0.67400    3 ; from the hydroxyl group of THR
  HE    H    0.40800    3 ; from the hydroxyl group of THR
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_27
  CG  CD2   gb_27
 CD2   OE   gb_18
  OE   HE   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_15
  CB   CG  CD2    ga_15
 CD1   CG  CD2    ga_15
```

```
  CG   CD2    OE    ga_13
  CD2   OE    HE    ga_12
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
 [ impropers ]
; ai    aj    ak    al   gromos type
   N   -C    CA    H     gi_1
  CA    N     C    CB     gi_2
  CB   CD2   CD1   CG     gi_2
   C   CA    +N    O     gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA   -C     N    CA     gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
   N   CA    CB    CG     gd_34
   N   CA    C    +N     gd_45
   N   CA    C    +N     gd_42
  CA   CB    CG   CD2     gd_34
  CB   CG   CD2   OE     gd_34
  CG   CD2   OE   HE     gd_23

; 3-hyroxyvaline
[ V3H ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH0    0.26600    2 ; from the hydroxyl group of THR
  OG3   OA   -0.67400    2 ; from the hydroxyl group of THR
  HG3    H    0.40800    2 ; from the hydroxyl group of THR
  CG1   CH3    0.00000    3
  CG2   CH3    0.00000    4
   C    C      0.450    5
   O    O     -0.450    5
 [ bonds ]
   N    H    gb_2
   N   CA    gb_21
  CA   CB    gb_27
  CA    C    gb_27
  CB   CG1    gb_27
  CB   CG2    gb_27
  CB   OG3    gb_18
 OG3   HG3    gb_1
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H     ga_32
  -C    N    CA     ga_31
   H    N    CA     ga_18
   N   CA    CB     ga_13
   N   CA    C     ga_13
  CB    CA    C     ga_13
  CA   CB   CG1     ga_13
  CA   CB   CG2     ga_13
  CA   CB   OG3     ga_13
 CG1   CB   CG2     ga_13
 CG1   CB   OG3     ga_13
 CG2   CB   OG3     ga_13
  CB   OG3   HG3     ga_12
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
 [ impropers ]
; ai    aj    ak    al   gromos type
   N   -C    CA    H     gi_1
  CA    N     C    CB     gi_2
   C   CA    +N    O     gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA   -C     N    CA     gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
   N   CA    CB   OG3     gd_34
   N   CA    C    +N     gd_45
   N   CA    C    +N     gd_42
  CA   CB   OG3   HG3     gd_23

; cysteine sulfenic acid
[ CYH ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA   CH1    0.00000    1
  CB   CH2    0.00000    1
  SG    S     0.26600    2 ; from the hydroxyl group of THR
  OD   OA    -0.67400    2 ; from the hydroxyl group of THR
  HD    H     0.40800    2 ; from the hydroxyl group of THR
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H    gb_2
   N   CA    gb_21
  CA   CB    gb_27
  CA    C    gb_27
  CB   SG    gb_32
  SG   OD    gb_28
  OD   HD    gb_1
   C    O    gb_5
   C   +N    gb_10
 [ angles ]
; ai    aj    ak   gromos type
  -C    N     H     ga_32
  -C    N    CA     ga_31
   H    N    CA     ga_18
   N   CA    CB     ga_13
   N   CA    C     ga_13
  CB    CA    C     ga_13
  CA   CB   SG     ga_16
  CB   SG   OD     ga_13
  SG   OD   HD     ga_12
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
 [ impropers ]
; ai    aj    ak    al   gromos type
   N   -C    CA    H     gi_1
  CA    N     C    CB     gi_2
   C   CA    +N    O     gi_1
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA   -C     N    CA     gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
   N   CA    CB   SG     gd_34
   N   CA    C    +N     gd_45
   N   CA    C    +N     gd_42
  CA   CB   SG   OD     gd_26
  CB   SG   OD   HD     gd_23

; 5-hydroxyproline (R)
[ PH5 ]
 [ atoms ]
   N    N    0.00000    0
  CA   CH1    0.00000    1
  CB  CH2R    0.00000    1
  CG  CH2R    0.00000    2
  CD   CH1    0.26600    3 ; from the hydroxyl group of THR
  OE    OA   -0.67400    3 ; from the hydroxyl group of THR
  HE     H    0.40800    3 ; from the hydroxyl group of THR
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N   CA    gb_21
   N   CD    gb_21
  CA   CB    gb_27
  CA    C    gb_27
  CB   CG    gb_27
  CG   CD    gb_27
  CD   OE    gb_18
  OE   HE    gb_1
   C    O    gb_5
   C   +N    gb_10
 [ exclusions ]
; ai    aj
   N   HE
 [ angles ]
; ai    aj    ak   gromos type
  -C    N    CA     ga_31
  -C    N    CD     ga_31
  CA    N    CD     ga_21
   N   CA    CB     ga_13
   N   CA    C     ga_13
  CB    CA    C     ga_13
  CA   CB   CG     ga_13
  CB   CG   CD     ga_13
   N   CD   CG     ga_13
   N   CD   OE     ga_13
  CG   CD   OE     ga_13
  CD   OE   HE     ga_12
  CA    C     O     ga_30
  CA    C    +N     ga_19
   O    C    +N     ga_33
 [ impropers ]
; ai    aj    ak    al   gromos type
   N   -C    CA    CD     gi_1
  CA    N     C    CB     gi_2
   C   CA    +N    O     gi_1
  CD   CG     N    OE     gi_2
 [ dihedrals ]
; ai    aj    ak    al   gromos type
 -CA   -C     N    CA     gd_14
  -C    N    CA    C     gd_44
  -C    N    CA    C     gd_43
```

```
  CA    N   CD   CG    gd_39
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD    N    gd_34
  CG   CD   OE   HE    gd_23

; 5-hydroxyproline (S)
[ P5H ]
 [ atoms ]
   N    N    0.00000    0
  CA  CH1    0.00000    1
  CB CH2R    0.00000    1
  CG CH2R    0.00000    2
  CD  CH1    0.26600    3 ; from the hydroxyl group of THR
  OE   OA   -0.67400    3 ; from the hydroxyl group of THR
  HE    H    0.40800    3 ; from the hydroxyl group of THR
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N   CA   gb_21
   N   CD   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   OE   gb_18
  OE   HE   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
   N   HE
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N   CA    ga_31
  -C    N   CD    ga_31
  CA    N   CD    ga_21
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_13
  CB   CG   CD    ga_13
   N   CD   CG    ga_13
   N   CD   OE    ga_13
  CG   CD   OE    ga_13
  CD   OE   HE    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA   CD    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  CD    N   CG   OE    gi_2
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
  CA    N   CD   CG    gd_39
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD    N    gd_34
  CG   CD   OE   HE    gd_23

; glutamic semialdehyde
[ GSA ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD    C    0.35000    2 ; by analogy to the aldehyde group reported by Dolenc et al.
DOI: 10.1093/nar/gki195
  HD   HC    0.10000    2 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  OE    O   -0.45000    2 ; from the carbonyl group (of e.g., GLN)
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   HD   gb_3
  CD   OE   gb_5
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   HD    ga_25
  CG   CD   OE    ga_27
  HD   CD   OE    ga_25
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  CD   OE   CG   HD    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   OE    gd_40

; 2-amino-3-ketobutyric acid
[ TOX ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB    C    0.45000    2 ; from the carbonyl group (of e.g., GLN)
 OG1    O   -0.45000    2 ; from the carbonyl group (of e.g., GLN)
 CG2  CH3    0.00000    3
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB  OG1   gb_5
  CB  CG2   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB  OG1    ga_27
  CA   CB  CG2    ga_27
 OG1   CB  CG2    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CB  OG1  CG2   CA    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB  OG1    gd_40
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42

; pyroglutamic acid
[ PGA ]
 [ atoms ]
   N    N    0.00000    0
  CA  CH1    0.00000    1
  CB CH2R    0.00000    1
  CG CH2R    0.00000    2
  CD    C    0.45000    3 ; from the carbonyl group (of e.g., GLN)
```

```
  OE    O   -0.45000    3 ; from the carbonyl group (of e.g., GLN)
   C    C     0.450    4
   O    O    -0.450    4
 [ bonds ]
   N   CA   gb_21
   N   CD   gb_10
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   OE   gb_5
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N   CA   ga_31
  -C    N   CD   ga_31
  CA    N   CD   ga_21
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG    ga_13
  CB   CG   CD    ga_13
   N   CD   CG    ga_19
   N   CD   OE    ga_33
  CG   CD   OE    ga_30
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA   CD    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  CD   CG    N   OE    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
  CA    N   CD   CG    gd_14
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD    N    gd_40

; 2-oxo-histidine
[ H2X ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    2
 ND1   NR   -0.31000    3
 HD1    H    0.31000    3
 CD2    C   -0.14000    4 ; to add up to 0 net charge
 HD2   HC    0.14000    4
 NE2   NR   -0.31000    5
 HE2    H    0.31000    5
 CE1    C    0.45000    6 ; from the carbonyl group (of e.g., GLN)
  OZ    O   -0.45000    6 ; from the carbonyl group (of e.g., GLN)
   C    C     0.450    7
   O    O    -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  ND1   gb_10
  CG  CD2   gb_10
 ND1  CE1   gb_10
 ND1  HD1   gb_2
 CD2  HD2   gb_3
 CD2  NE2   gb_10
 NE2  HE2   gb_2
 CE1  NE2   gb_10
 CE1   OZ   gb_5
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  NE2
  CG  HE2
  CG   OZ
 ND1  HD2
 ND1  HE2
 HD1  CD2
 HD1  NE2
 HD1   OZ
 CD2   OZ
 HD2  CE1
 HD2  HE2
 HE2   OZ
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  ND1   ga_37
  CB   CG  CD2   ga_37
 ND1   CG  CD2   ga_7
  CG  ND1  HD1   ga_36
  CG  ND1  CE1   ga_7
 HD1  ND1  CE1   ga_36
  CG  CD2  HD2   ga_36
  CG  CD2  NE2   ga_7
 HD2  CD2  NE2   ga_36
 ND1  CE1  NE2   ga_7
 ND1  CE1   OZ   ga_37
 NE2  CE1   OZ   ga_37
 CD2  NE2  CE1   ga_7
 CD2  NE2  HE2   ga_36
 CE1  NE2  HE2   ga_36
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  ND1  CD2   CB    gi_1
  CG  ND1  CE1  NE2    gi_1
  CG  CD2  NE2  CE1    gi_1
 ND1   CG  CD2  NE2    gi_1
 ND1  CE1  NE2  CD2    gi_1
 ND1   CG  CE1  HD1    gi_1
 CD2   CG  ND1  CE1    gi_1
 CD2   CG  NE2  HD2    gi_1
 CE1  ND1  NE2   OZ    gi_1
 NE2  CD2  CE1  HE2    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  ND1    gd_40

; methionine sulfoxide (R)
[ MSX ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.24100    2
  SD    S   -0.03200    2 ; to add up to 0 net charge
 CE1  CH3    0.24100    2
 OE2    O   -0.45000    2 ; from the carbonyl group (of e.g., GLN)
   C    C     0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   SD   gb_32
  SD  CE1   gb_32
  SD  OE2   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak   gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG   SD   ga_16
  CG   SD  CE1   ga_4
```

```
  CG   SD  OE2    ga_6
 CE1   SD  OE2    ga_6
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  SD  OE2  CE1   CG    gi_2
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   SD    gd_34
  CB   CG   SD  CE1    gd_26

; methionine sulfoxide (S)
[ MXS ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.24100    2
  SD    S   -0.03200    2 ; to add up to 0 net charge
 CE1  CH3    0.24100    2
 OE2    O   -0.45000    2 ; from the carbonyl group (of e.g., GLN)
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H    gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   SD   gb_32
  SD  CE1   gb_32
  SD  OE2   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   SD    ga_16
  CG   SD  CE1    ga_4
  CG   SD  OE2    ga_6
 CE1   SD  OE2    ga_6
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  SD  CE1  OE2   CG    gi_2
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   SD    gd_34
  CB   CG   SD  CE1    gd_26

; methionine sulfone
[ MES ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.24100    2
  SD    S    0.23800    2 ; to add up to 0 net charge
 CE1  CH3    0.24100    2
 OE2    O   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OE3    O   -0.36000    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H    gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   SD   gb_32
  SD  CE1   gb_31
  SD  OE2   gb_25
  SD  OE3   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   SD    ga_16
  CG   SD  CE1    ga_13
  CG   SD  OE2    ga_13
  CG   SD  OE3    ga_13
 CE1   SD  OE2    ga_13
 CE1   SD  OE3    ga_13
 OE2   SD  OE3    ga_13
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   SD    gd_34
  CB   CG   SD   CE    gd_26

; cysteine sulfinic acid
[ CSA ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  SG    S    0.12000    2 ; to add up to -1 net charge
 OD1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
 OD2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H    gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   SG   gb_32
  SG  OD1   gb_25
  SG  OD2   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
;  ai   aj   ak   gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   SG    ga_16
  CB   SG  OD1    ga_6
  CB   SG  OD2    ga_6
 OD1   SG  OD2    ga_6
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
;  ai   aj   ak   al   gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  SG   CB  OD1  OD2    gi_2
 [ dihedrals ]
;  ai   aj   ak   al   gromos type
```

```
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   SG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   SG  OD1    gd_26

; cysteic acid
[ CSE ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.15000    2 ; from the carbon atom attached to the phosphate group of
nucleotides (e.g., ATP)
  SG    S    0.75500    2 ; to add up to -1 net charge
  OD1   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OD2   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
  OD3   OM   -0.63500    2 ; from the phosphate group of nucleotides (e.g., ATP)
   C    C     0.450    3
   O    O    -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   SG   gb_32
  SG  OD1   gb_25
  SG  OD2   gb_25
  SG  OD3   gb_25
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   SG   ga_16
  CB   SG  OD1   ga_13
 OD1   SG  OD2   ga_13
  CB   SG  OD2   ga_13
  CB   SG  OD3   ga_13
 OD1   SG  OD3   ga_13
 OD2   SG  OD3   ga_13
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
   C   CA   +N    O   gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   SG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   SG  OD1    gd_26

; 3-nitrotyrosine (-1)
[ YNI ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C    0.10000    4 ; to add up to -1 net charge
 NZ1   NR    0.05500    4 ; to add up to -1 net charge
 OH1    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OH2    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 CZ2    C    0.20000    4 ; to add up to -1 net charge
 OH3   OM   -0.63500    4 ; from the carboxyl group (of e.g., GLU)
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
   C    C     0.450    6
   O    O    -0.450    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  NZ1   gb_12
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ2   gb_16
 NZ1  OH1   gb_6
 NZ1  OH2   gb_6
 CZ2  OH3   gb_13
 OH3  HH3   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  OH3
 HD1  CD2
 HD1  NZ1
 HD1  CZ2
 CD2  CE1
 CD2  OH3
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 HE2  OH3
 NZ1  CE2
 NZ1  OH3
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  CD1   ga_27
  CB   CG  CD2   ga_27
 CD1   CG  CD2   ga_27
  CG  CD1  HD1   ga_25
  CG  CD1  CE1   ga_27
 HD1  CD1  CE1   ga_25
  CG  CD2  HD2   ga_25
  CG  CD2  CE2   ga_27
 HD2  CD2  CE2   ga_25
 CD1  CE1  NZ1   ga_27
 CD1  CE1  CZ2   ga_27
 NZ1  CE1  CZ2   ga_27
 CD2  CE2  HE2   ga_25
 CD2  CE2  CZ2   ga_27
 HE2  CE2  CZ2   ga_25
 CE1  NZ1  OH1   ga_27
 CE1  NZ1  OH2   ga_27
 OH1  NZ1  OH2   ga_27
 CE1  CZ2  CE2   ga_27
 CE1  CZ2  OH3   ga_27
 CE2  CZ2  OH3   ga_27
 CZ2  OH3  HH3   ga_12
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
  CG  CD1  CE1  CZ2   gi_1
  CG  CD2  CE2  CZ2   gi_1
 CD1   CG  CD2  CE2   gi_1
 CD1   CG  CE1  HD1   gi_1
 CD1  CE1  CZ2  CE2   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD2   CG  CE2  HD2   gi_1
 CD2  CE2  CZ2  CE1   gi_1
 CE1  CD1  CZ2  NZ1   gi_1
 HE2  CD2  CZ2  CE2   gi_1
 NZ1  OH1  OH2  CE1   gi_1
 CZ2  CE1  CE2  OH3   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
```

```
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_40
 CD1  CE1  NZ1  OH1    gd_14
 CE1  CZ2  OH3  HH3    gd_11

; 3-nitrotyrosine (0)
[ YNN ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C    0.10000    4 ; newly developed parameters tp match the experimental
HFE
 NZ1   NR    0.30000    4 ; newly developed parameters tp match the experimental
HFE
 OH1    O   -0.20000    4 ; newly developed parameters tp match the experimental
HFE
 OH2    O   -0.20000    4 ; newly developed parameters tp match the experimental
HFE
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C    0.05000    6 ; newly developed parameters tp match the experimental
HFE
 OH3   OA   -0.36000    6 ; newly developed parameters tp match the experimental
HFE
 HH3    H    0.31000    6 ; newly developed parameters tp match the experimental
HFE
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  NZ1   gb_12
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ2   gb_16
 NZ1  OH1   gb_6
 NZ1  OH2   gb_6
 CZ2  OH3   gb_13
 OH3  HH3   gb_1
   C    O   gb_5
   C   +N   gb_10
[ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  OH3
 HD1  CD2
 HD1  NZ1
 HD1  CZ2
 CD2  CE1
 CD2  OH3
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 HE2  OH3
 NZ1  CE2
 NZ1  OH3
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  CD1   ga_27
  CB   CG  CD2   ga_27
 CD1   CG  CD2   ga_27
  CG  CD1  HD1   ga_25
  CG  CD1  CE1   ga_27
 HD1  CD1  CE1   ga_25
  CG  CD2  HD2   ga_25
  CG  CD2  CE2   ga_27
 HD2  CD2  CE2   ga_25
 CD1  CE1  NZ1   ga_27
 CD1  CE1  CZ2   ga_27
 NZ1  CE1  CZ2   ga_27
 CD2  CE2  HE2   ga_25
 CD2  CE2  CZ2   ga_27
 HE2  CE2  CZ2   ga_25
 CE1  NZ1  OH1   ga_27
 CE1  NZ1  OH2   ga_27
 OH1  NZ1  OH2   ga_27
 CE1  CZ2  CE2   ga_27
 CE1  CZ2  OH3   ga_27
 CE2  CZ2  OH3   ga_27
 CZ2  OH3  HH3   ga_12
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD1  CD2   CB   gi_1
  CG  CD1  CE1  CZ2   gi_1
  CG  CD2  CE2  CZ2   gi_1
 CD1   CG  CD2  CE2   gi_1
 CD1   CG  CE1  HD1   gi_1
 CD1  CE1  CZ2  CE2   gi_1
 CD2   CG  CD1  CE1   gi_1
 CD2   CG  CE2  HD2   gi_1
 CD2  CE2  CZ2  CE1   gi_1
 CE1  CD1  CZ2  NZ1   gi_1
 HE2  CD2  CZ2  CE2   gi_1
 NZ1  OH1  OH2  CE1   gi_1
 CZ2  CE1  CE2  OH3   gi_1
   C   CA   +N    O   gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD1    gd_40
 CD1  CE1  NZ1  OH1    gd_14
 CE1  CZ2  OH3  HH3    gd_11

; 3-nitrotyrosine (0)
[ YNB ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.00000    1
 CD1    C   -0.14000    2
 HD1   HC    0.14000    2
 CD2    C   -0.14000    3
 HD2   HC    0.14000    3
 CE1    C    0.10000    4 ; to add up to 0 net charge
 NZ1   NR    0.62000    4 ; to add up to 0 net charge
 OH1    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 OH2    O   -0.36000    4 ; from the phosphate group of nucleotides (e.g., ATP)
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C    0.20300    6
 OH3   OA   -0.61100    6
 HH3    H    0.40800    6
   C    C      0.450    7
   O    O     -0.450    7
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  CD1   gb_16
  CG  CD2   gb_16
 CD1  HD1   gb_3
 CD1  CE1   gb_16
 CD2  HD2   gb_3
 CD2  CE2   gb_16
 CE1  NZ1   gb_12
```

```
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ2   gb_16
 NZ1  OH1   gb_6
 NZ1  OH2   gb_6
 CZ2  OH3   gb_13
 OH3  HH3   gb_1
  C   O    gb_5
  C   +N   gb_10
[ exclusions ]
;  ai   aj
  CB  HD1
  CB  HD2
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
 CD1  HD2
 CD1  CE2
 CD1  OH3
 HD1  CD2
 HD1  NZ1
 HD1  CZ2
 CD2  CE1
 CD2  OH3
 HD2  HE2
 HD2  CZ2
 CE1  HE2
 HE2  OH3
 NZ1  CE2
 NZ1  OH3
[ angles ]
;  ai   aj   ak  gromos type
  -C   N    H    ga_32
  -C   N   CA    ga_31
   H   N   CA    ga_18
   N   CA  CB    ga_13
   N   CA  C     ga_13
  CB   CA  C     ga_13
  CA   CB  CG    ga_15
  CB   CG  CD1   ga_27
  CB   CG  CD2   ga_27
 CD1   CG  CD2   ga_27
  CG   CD1 HD1   ga_25
  CG   CD1 CE1   ga_27
 HD1   CD1 CE1   ga_25
  CG   CD2 HD2   ga_25
  CG   CD2 CE2   ga_27
 HD2   CD2 CE2   ga_25
 CD1   CE1 NZ1   ga_27
 CD1   CE1 CZ2   ga_27
 NZ1   CE1 CZ2   ga_27
 CD2   CE2 HE2   ga_25
 CD2   CE2 CZ2   ga_27
 HE2   CE2 CZ2   ga_25
 CE1   NZ1 OH1   ga_27
 CE1   NZ1 OH2   ga_27
 OH1   NZ1 OH2   ga_27
 CE1   CZ2 CE2   ga_27
 CE1   CZ2 OH3   ga_27
 CE2   CZ2 OH3   ga_27
 CZ2   OH3 HH3   ga_12
  CA   C   O     ga_30
  CA   C   +N    ga_19
   O   C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA   H    gi_1
  CA   N    C    CB   gi_2
  CG   CD1  CD2  CB   gi_1
  CG   CD1  CE1  CZ2  gi_1
  CG   CD2  CE2  CZ2  gi_1
 CD1   CG   CD2  CE2  gi_1
 CD1   CG   CE1  HD1  gi_1
 CD1   CE1  CZ2  CE2  gi_1
 CD2   CG   CD1  CE1  gi_1
 CD2   CG   CE2  HD2  gi_1
 CD2   CE2  CZ2  CE1  gi_1
 CE1   CD1  CZ2  NZ1  gi_1
 HE2   CD2  CZ2  CE2  gi_1
 NZ1   OH1  OH2  CE1  gi_1
 CZ2   CE1  CE2  OH3  gi_1
  C    CA   +N   O    gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA   gd_14
  -C    N   CA   C    gd_44
  -C    N   CA   C    gd_43
   N   CA   CB   CG   gd_34
   N   CA   C    +N   gd_45
   N   CA   C    +N   gd_42
  CA   CB   CG   CD1  gd_40
 CD1   CE1  NZ1  OH1  gd_14
 CE1   CZ2  OH3  HH3  gd_11

; 6-nitrotryptophan
[ WNI ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG   C    -0.21000    2
 CD1   C    -0.14000    2
 HD1  HC     0.14000    2
 CD2   C     0.00000    2
 NE1  NR    -0.10000    2
 HE1   H     0.31000    2
 CE2   C     0.00000    2
 CE3   C    -0.14000    3
 HE3  HC     0.14000    3
 CZ2   C    -0.14000    4
 HZ2  HC     0.14000    4
 CZ3   C    -0.14000    5
 HZ3  HC     0.14000    5
 CH2   C     0.10000    6 ; to add up to 0 net charge
  NT  NR     0.62000    6 ; to add up to 0 net charge
 OI1   O    -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
 OI2   O    -0.36000    6 ; from the phosphate group of nucleotides (e.g., ATP)
   C   C      0.450    7
   O   O     -0.450    7
 [ bonds ]
   N   H    gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA   C    gb_27
  CB   CG   gb_27
  CG   CD1  gb_10
  CG   CD2  gb_16
 CD1   HD1  gb_3
 CD1   NE1  gb_10
 CD2   CE2  gb_16
 CD2   CE3  gb_16
 NE1   HE1  gb_2
 NE1   CE2  gb_10
 CE2   CZ2  gb_16
 CE3   HE3  gb_3
 CE3   CZ3  gb_16
 CZ2   HZ2  gb_3
 CZ2   CH2  gb_16
 CZ3   HZ3  gb_3
 CZ3   CH2  gb_16
 CH2   NT   gb_12
  NT   OI1  gb_6
  NT   OI2  gb_6
   C   O    gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB   HD1
  CB   NE1
  CB   CE2
  CB   CE3
  CG   HE1
  CG   HE3
  CG   CZ2
  CG   CZ3
 CD1   CE3
 CD1   CZ2
 HD1   CD2
 HD1   HE1
 HD1   CE2
 CD2   HE1
 CD2   HZ2
 CD2   HZ3
 CD2   CH2
 NE1   CE3
 NE1   HZ2
 NE1   CH2
 HE1   CZ2
 CE2   HE3
 CE2   CZ3
 CE2   NT
 CE3   CZ2
 CE3   NT
 HE3   HZ3
 HE3   CH2
 CZ2   HZ3
 HZ2   CZ3
 HZ2   NT
 HZ3   NT
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
```

```
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG  CD1    ga_37
  CB   CG  CD2    ga_37
 CD1   CG  CD2    ga_7
  CG  CD1  HD1    ga_36
  CG  CD1  NE1    ga_7
 HD1  CD1  NE1    ga_36
  CG  CD2  CE2    ga_7
  CG  CD2  CE3    ga_39
 CE2  CD2  CE3    ga_27
 CD1  NE1  HE1    ga_36
 CD1  NE1  CE2    ga_7
 HE1  NE1  CE2    ga_36
 CD2  CE2  NE1    ga_7
 CD2  CE2  CZ2    ga_27
 NE1  CE2  CZ2    ga_39
 CD2  CE3  HE3    ga_25
 CD2  CE3  CZ3    ga_27
 HE3  CE3  CZ3    ga_25
 CE2  CZ2  HZ2    ga_25
 CE2  CZ2  CH2    ga_27
 HZ2  CZ2  CH2    ga_25
 CE3  CZ3  HZ3    ga_25
 CE3  CZ3  CH2    ga_27
 HZ3  CZ3  CH2    ga_25
 CZ2  CH2  CZ3    ga_27
 CZ2  CH2   NT    ga_27
 CZ3  CH2   NT    ga_27
 CH2   NT  OI1    ga_27
 CH2   NT  OI2    ga_27
 OI1   NT  OI2    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
[ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD1  CD2   CB    gi_1
  CG  CD1  NE1  CE2    gi_1
  CG  CD2  CE2  NE1    gi_1
 CD1   CG  CD2  CE2    gi_1
 CD1   CG  NE1  HD1    gi_1
 CD1  NE1  CE2  CD2    gi_1
 CD2   CG  CD1  NE1    gi_1
 CD2  CE2  CE3   CG    gi_1
 CD2  CE2  CZ2  CH2    gi_1
 CD2  CE3  CZ3  CH2    gi_1
 NE1  CD1  CE2  HE1    gi_1
 CE2  CD2  CE3  CZ3    gi_1
 CE2  CD2  CZ2  NE1    gi_1
 CE2  CZ2  CH2  CZ3    gi_1
 CE3  CD2  CE2  CZ2    gi_1
 CE3  CD2  CZ3  HE3    gi_1
 CE3  CZ3  CH2  CZ2    gi_1
 CZ2  CE2  CH2  HZ2    gi_1
 CZ3  CE3  CH2  HZ3    gi_1
 CH2  CZ2  CZ3   NT    gi_1
  NT  OI1  OI2  CH2    gi_1
   C   CA   +N    O    gi_1
[ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_40
 CZ2  CH2   NT  OI1    gd_14

; kynurenine
[ WKY ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.45000    2
 OD1    O   -0.45000    2
 CD2    C    0.00000    3
 CE1    C    0.00000    3
 NZ1   NT   -0.88000    4 ; from ARGN
HZ11    H    0.44000    4 ; from ARGN
HZ12    H    0.44000    4 ; from ARGN
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C   -0.14000    6
 HZ2   HC    0.14000    6
 CZ3    C   -0.14000    7
 HZ3   HC    0.14000    7
 CH     C   -0.14000    8
 HH    HC    0.14000    8
  C     C     0.450    9
  O     O    -0.450    9
[ bonds ]
  N    H   gb_2
  N   CA   gb_21
 CA   CB   gb_27
 CA    C   gb_27
 CB   CG   gb_27
 CG  OD1   gb_5
 CG  CD2   gb_23
CD2  CE1   gb_16
CD2  CE2   gb_16
CE1  NZ1   gb_9
CE1  CZ2   gb_16
CE2  HE2   gb_3
CE2  CZ3   gb_16
NZ1 HZ11   gb_2
NZ1 HZ12   gb_2
CZ2  HZ2   gb_3
CZ2   CH   gb_16
CZ3  HZ3   gb_3
CZ3   CH   gb_16
 CH   HH   gb_3
  C    O   gb_5
  C   +N   gb_10
[ exclusions ]
;  ai   aj
 CB  CE1
 CB  CE2
 CG  NZ1
 CG  HE2
 CG  CZ2
 CG  CZ3
CD2 HZ11
CD2 HZ12
CD2  HZ2
CD2  HZ3
CD2   CH
CE1  HE2
CE1  CZ3
CE1   HH
CE2  CZ2
CE2   HH
HE2  HZ3
HE2   CH
NZ1  CE2
NZ1  HZ2
NZ1   CH
HZ11  CZ2
HZ12  CZ2
CZ2  HZ3
HZ2  CZ3
HZ2   HH
HZ3   HH
[ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_27
  CB   CG  CD2   ga_27
 OD1   CG  CD2   ga_27
  CG  CD2  CE1   ga_27
  CG  CD2  CE2   ga_27
 CE1  CD2  CE2   ga_27
 CD2  CE1  NZ1   ga_27
 CD2  CE1  CZ2   ga_27
 NZ1  CE1  CZ2   ga_27
 CD2  CE2  HE2   ga_25
 CD2  CE2  CZ3   ga_27
 HE2  CE2  CZ3   ga_25
 CE1  NZ1 HZ11   ga_23
 CE1  NZ1 HZ12   ga_23
HZ11  NZ1 HZ12   ga_24
 CE1  CZ2  HZ2   ga_25
 CE1  CZ2   CH   ga_27
 HZ2  CZ2   CH   ga_25
 CE2  CZ3  HZ3   ga_25
 CE2  CZ3   CH   ga_27
 HZ3  CZ3   CH   ga_25
 CZ2   CH  CZ3   ga_27
 CZ2   CH   HH   ga_25
 CZ3   CH   HH   ga_25
  CA    C    O   ga_30
  CA    C   +N   ga_19
```

```
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD2  OD1   CB    gi_1
 CD2  CE1  CE2   CG    gi_1
 CD2  CE1  CZ2   CH    gi_1
 CD2  CE2  CZ3   CH    gi_1
 CE1  CD2  CE2  CZ3    gi_1
 CE1  CD2  CZ2  NZ1    gi_1
 CE1  CZ2   CH  CZ3    gi_1
 CE2  CD2  CE1  CZ2    gi_1
 CE2  CD2  CZ3  HE2    gi_1
 CE2  CZ3   CH  CZ2    gi_1
 NZ1 HZ11 HZ12  CE1    gi_1
 HZ2  CE1   CH  CZ2    gi_1
 HZ3  CE2   CH  CZ3    gi_1
  CH  CZ2  CZ3   HH    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_40
  CB   CG  CD2  CE1    gd_10
 CD2  CE1  NZ1 HZ11    gd_14

; 3-hydroxykynurenine
[ WKH ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.45000    2
 OD1    O   -0.45000    2
 CD2    C    0.00000    3
 CE1    C    0.00000    3
 NZ1   NT   -0.88000    4 ; from ARGN
HZ11    H    0.44000    4 ; from ARGN
HZ12    H    0.44000    4 ; from ARGN
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C    0.20300    6 ; from the hydroxyl group of TYR
 OH1   OA   -0.61100    6 ; from the hydroxyl group of TYR
 HH1    H    0.40800    6 ; from the hydroxyl group of TYR
 CZ3    C   -0.14000    7
 HZ3   HC    0.14000    7
 CH2    C   -0.14000    8
 HH2   HC    0.14000    8
   C    C     0.450    9
   O    O    -0.450    9
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  OD1   gb_5
  CG  CD2   gb_23
 CD2  CE1   gb_16
 CD2  CE2   gb_16
 CE1  NZ1   gb_9
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ3   gb_16
 NZ1 HZ11   gb_2
 NZ1 HZ12   gb_2
 CZ2  OH1   gb_13
 CZ2  CH2   gb_16
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_16
 OH1  HH1   gb_1
 CH2  HH2   gb_3
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
; ai   aj
  CB  CE1
  CB  CE2
  CG  HE2
  CG  NZ1
  CG  CZ2
  CG  CZ3
 CD2 HZ11
 CD2 HZ12
 CD2  HZ3
 CD2  OH1
 CD2  CH2
 CE1  HE2
 CE1  CZ3
 CE1  HH1
 CE1  HH2
 CE2  CZ2
 CE2  HH2
 HE2  HZ3
 HE2  CH2
 NZ1  CE2
 NZ1  OH1
 NZ1  HH1
 NZ1  CH2
HZ11  CZ2
HZ11  OH1
HZ12  CZ2
HZ12  OH1
 CZ2  HZ3
 HZ3  HH2
 OH1  CZ3
 OH1  HH2
 HH1  CH2
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_27
  CB   CG  CD2   ga_27
 OD1   CG  CD2   ga_27
  CG  CD2  CE1   ga_27
  CG  CD2  CE2   ga_27
 CE1  CD2  CE2   ga_27
 CD2  CE1  NZ1   ga_27
 CD2  CE1  CZ2   ga_27
 NZ1  CE1  CZ2   ga_27
 CD2  CE2  HE2   ga_25
 CD2  CE2  CZ3   ga_27
 HE2  CE2  CZ3   ga_25
 CE1  NZ1 HZ11   ga_23
 CE1  NZ1 HZ12   ga_23
HZ11  NZ1 HZ12   ga_24
 CE1  CZ2  OH1   ga_27
 CE1  CZ2  CH2   ga_27
 OH1  CZ2  CH2   ga_27
 CE2  CZ3  HZ3   ga_25
 CE2  CZ3  CH2   ga_27
 HZ3  CZ3  CH2   ga_25
 CZ2  OH1  HH1   ga_12
 CZ2  CH2  CZ3   ga_27
 CZ2  CH2  HH2   ga_25
 CZ3  CH2  HH2   ga_25
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
  CG  CD2  OD1   CB    gi_1
 CD2  CE1  CE2   CG    gi_1
 CD2  CE1  CZ2  CH2    gi_1
 CD2  CE2  CZ3  CH2    gi_1
 CE1  CD2  CE2  CZ3    gi_1
 CE1  CD2  CZ2  NZ1    gi_1
 CE1  CZ2  CH2  CZ3    gi_1
 CE2  CD2  CE1  CZ2    gi_1
 CE2  CD2  CZ3  HE2    gi_1
 CE2  CZ3  CH2  CZ2    gi_1
 NZ1 HZ11 HZ12  CE1    gi_1
 OH1  CE1  CH2  CZ2    gi_1
 HZ3  CE2  CH2  CZ3    gi_1
 CH2  CZ2  CZ3   HH    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG  CD2    gd_40
  CB   CG  CD2  CE1    gd_10
 CD2  CE1  NZ1 HZ11    gd_14
 CE1  CZ2  OH1  HH1    gd_11

; formyl-kynurenine
[ WKF ]
```

```
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG    C    0.45000    2
 OD1    O   -0.45000    2
 CD2    C    0.00000    3
 CE1    C    0.00000    3
 NZ1    N   -0.15000    4 ; by analogy to ASN/GLN and the peptide bond
 HZ1    H    0.31000    4 ; from the hydrogen of the peptide bond
 CH1    C    0.19000    4 ; by analogy to ASN/GLN, the peptide bond and the
aldehyde group reported by Dolenc et al. DOI: 10.1093/nar/gki195
 HH1   HC    0.10000    4 ; by analogy to the aldehyde group reported by Dolenc et
al. DOI: 10.1093/nar/gki195
  OI    O   -0.45000    4 ; from the carbonyl group (of e.g., GLN)
 CE2    C   -0.14000    5
 HE2   HC    0.14000    5
 CZ2    C   -0.14000    6
 HZ2   HC    0.14000    6
 CZ3    C   -0.14000    7
 HZ3   HC    0.14000    7
 CH2    C   -0.14000    8
 HH2   HC    0.14000    8
   C    C      0.450    9
   O    O     -0.450    9
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG  OD1   gb_5
  CG  CD2   gb_23
 CD2  CE1   gb_16
 CD2  CE2   gb_16
 CE1  NZ1   gb_9
 CE1  CZ2   gb_16
 CE2  HE2   gb_3
 CE2  CZ3   gb_16
 NZ1  HZ1   gb_2
 NZ1  CH1   gb_10
 CZ2  HZ2   gb_3
 CZ2  CH2   gb_16
 CZ3  HZ3   gb_3
 CZ3  CH2   gb_16
 CH1  HH1   gb_3
 CH1   OI   gb_5
 CH2  HH2   gb_3
   C    O   gb_5
   C   +N   gb_10
 [ exclusions ]
;  ai   aj
  CB  CE1
  CB  CE2
  CG  NZ1
  CG  HE2
  CG  CZ2
  CG  CZ3
 CD2  HZ1
 CD2  HZ2
 CD2  HZ3
 CD2  CH2
 CE1  HE2
 CE1  CZ3
 CE1  HH2
 CE2  CZ2
 CE2  HH2
 HE2  HZ3
 HE2  CH2
 NZ1  CE2
 NZ1  HZ2
 NZ1  CH2
 HZ1  CZ2
 CZ2  HZ3
 HZ2  CZ3
 HZ2  HH2
 HZ3  HH2
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H   ga_32
  -C    N   CA   ga_31
   H    N   CA   ga_18
   N   CA   CB   ga_13
   N   CA    C   ga_13
  CB   CA    C   ga_13
  CA   CB   CG   ga_15
  CB   CG  OD1   ga_27
  CB   CG  CD2   ga_27
 OD1   CG  CD2   ga_27
  CG  CD2  CE1   ga_27
  CG  CD2  CE2   ga_27
 CE1  CD2  CE2   ga_27
 CD2  CE1  NZ1   ga_27
 CD2  CE1  CZ2   ga_27
 NZ1  CE1  CZ2   ga_27
 CD2  CE2  HE2   ga_25
 CD2  CE2  CZ3   ga_27
 HE2  CE2  CZ3   ga_25
 CE1  NZ1  HZ1   ga_18
 CE1  NZ1  CH1   ga_31
 HZ1  NZ1  CH1   ga_32
 CE1  CZ2  HZ2   ga_25
 CE1  CZ2  CH2   ga_27
 HZ2  CZ2  CH2   ga_25
 CE2  CZ3  HZ3   ga_25
 CE2  CZ3  CH2   ga_27
 HZ3  CZ3  CH2   ga_25
 NZ1  CH1  HH1   ga_19
 NZ1  CH1   OI   ga_33
 HH1  CH1   OI   ga_30
 CZ2  CH2  CZ3   ga_27
 CZ2  CH2  HH2   ga_25
 CZ3  CH2  HH2   ga_25
  CA    C    O   ga_30
  CA    C   +N   ga_19
   O    C   +N   ga_33
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H   gi_1
  CA    N    C   CB   gi_2
  CG  CD2  OD1   CB   gi_1
 CD2  CE1  CE2   CG   gi_1
 CD2  CE1  CZ2  CH2   gi_1
 CD2  CE2  CZ3  CH2   gi_1
 CE1  CD2  CE2  CZ3   gi_1
 CE1  CD2  CZ2  NZ1   gi_1
 CE1  CZ2  CH2  CZ3   gi_1
 CE2  CD2  CE1  CZ2   gi_1
 CE2  CD2  CZ3  HE2   gi_1
 CE2  CZ3  CH2  CZ2   gi_1
 NZ1  CH1  CE1  HZ1   gi_1
 HZ2  CE1  CH2  CZ2   gi_1
 HZ3  CE2  CH2  CZ3   gi_1
 CH1   OI  NZ1  HH1   gi_1
 CH2  CZ2  CZ3  HH2   gi_1
   C   CA   +N    O   gi_1
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
  -CA   -C    N   CA   gd_14
   -C    N   CA    C   gd_44
   -C    N   CA    C   gd_43
    N   CA   CB   CG   gd_34
    N   CA    C   +N   gd_45
    N   CA    C   +N   gd_42
   CA   CB   CG  CD2   gd_40
   CB   CG  CD2  CE1   gd_10
  CD2  CE1  NZ1  CH1   gd_14
  CE1  NZ1  CH1   OI   gd_14

 ; chlorotyrosine
 [ YCH ]
  [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA  CH1    0.00000    1
   CB  CH2    0.00000    1
   CG    C    0.00000    1
  CD1    C   -0.14000    2
  HD1   HC    0.14000    2
  CD2    C   -0.14000    3
  HD2   HC    0.14000    3
  CE1    C    0.08700    4 ; from CHCL3
 CLZ1   CL   -0.08700    4 ; from CHCL3
  CE2    C   -0.14000    5
  HE2   HC    0.14000    5
  CZ2    C    0.20300    6
   OH   OA   -0.61100    6
   HH    H    0.40800    6
    C    C      0.450    7
    O    O     -0.450    7
  [ bonds ]
    N    H   gb_2
    N   CA   gb_21
   CA   CB   gb_27
   CA    C   gb_27
   CB   CG   gb_27
   CG  CD1   gb_16
   CG  CD2   gb_16
  CD1  HD1   gb_3
  CD1  CE1   gb_16
  CD2  HD2   gb_3
  CD2  CE2   gb_16
  CE1 CLZ1   gb_40
  CE1  CZ2   gb_16
  CE2  HE2   gb_3
```

```
  CE2   CZ2    gb_16
  CZ2   OH    gb_13
  OH    HH    gb_1
   C    O    gb_5
   C    +N    gb_10
 [ exclusions ]
;  ai    aj
  CB   HD1
  CB   HD2
  CB   CE1
  CB   CE2
  CG   HE2
  CG  CLZ1
  CG   CZ2
  CD1   HD2
  CD1   CE2
  CD1    OH
  HD1   CD2
  HD1  CLZ1
  HD1   CZ2
  CD2   CE1
  CD2    OH
  HD2   HE2
  HD2   CZ2
  CE1   HE2
  HE2    OH
 CLZ1   CE2
 CLZ1    OH
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N    CA    ga_31
    H    N    CA    ga_18
    N    CA    CB    ga_13
    N    CA    C    ga_13
   CB    CA    C    ga_13
   CA    CB    CG    ga_15
   CB    CG   CD1    ga_27
   CB    CG   CD2    ga_27
  CD1    CG   CD2    ga_27
   CG   CD1   HD1    ga_25
   CG   CD1   CE1    ga_27
  HD1   CD1   CE1    ga_25
   CG   CD2   HD2    ga_25
   CG   CD2   CE2    ga_27
  HD2   CD2   CE2    ga_25
  CD1   CE1  CLZ1    ga_27
  CD1   CE1   CZ2    ga_27
 CLZ1   CE1   CZ2    ga_27
  CD2   CE2   HE2    ga_25
  CD2   CE2   CZ2    ga_27
  HE2   CE2   CZ2    ga_25
  CE1   CZ2   CE2    ga_27
  CE1   CZ2    OH    ga_27
  CE2   CZ2    OH    ga_27
  CZ2    OH    HH    ga_12
   CA    C    O    ga_30
   CA    C    +N    ga_19
    O    C    +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
   CA    N    C    CB    gi_2
   CG   CD1   CD2    CB    gi_1
   CG   CD1   CE1   CZ2    gi_1
   CG   CD2   CE2   CZ2    gi_1
  CD1    CG   CD2   CE2    gi_1
  CD1    CG   CE1   HD1    gi_1
  CD1   CE1   CZ2   CE2    gi_1
  CD2    CG   CD1   CE1    gi_1
  CD2    CG   CE2   HD2    gi_1
  CD2   CE2   CZ2   CE1    gi_1
 CLZ1   CD1   CZ2   CE1    gi_1
  HE2   CD2   CZ2   CE2    gi_1
  CZ2   CE1   CE2    OH    gi_1
    C    CA    +N    O    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C    N    CA    gd_14
   -C    N    CA    C    gd_44
   -C    N    CA    C    gd_43
    N    CA    CB    CG    gd_34
    N    CA    C    +N    gd_45
    N    CA    C    +N    gd_42
   CA    CB    CG   CD1    gd_40
  CE1   CZ2    OH    HH    gd_11

; homocitrulline
[ KAM ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG   CH2    0.00000    2
   CD   CH2    0.00000    2
   CE   CH2    0.00000    3
   NZ    N   -0.31000    4 ; from the peptide bond
   HZ    H    0.31000    4 ; from the peptide bond
   CH    C    0.29000    5 ; from ASN/GLN
  OI1    O   -0.45000    5 ; from ASN/GLN
  NI2   NT   -0.72000    5 ; from ASN/GLN
 HI21    H    0.44000    5 ; from ASN/GLN
 HI22    H    0.44000    5 ; from ASN/GLN
    C    C      0.450    6
    O    O     -0.450    6
 [ bonds ]
    N    H    gb_2
    N    CA    gb_21
   CA    CB    gb_27
   CA    C    gb_27
   CB    CG    gb_27
   CG    CD    gb_27
   CD    CE    gb_27
   CE    NZ    gb_21
   NZ    HZ    gb_2
   NZ    CH    gb_10
   CH   OI1    gb_5
   CH   NI2    gb_9
  NI2  HI21    gb_2
  NI2  HI22    gb_2
    C    O    gb_5
    C    +N    gb_10
 [ angles ]
;  ai    aj    ak   gromos type
   -C    N    H    ga_32
   -C    N    CA    ga_31
    H    N    CA    ga_18
    N    CA    CB    ga_13
    N    CA    C    ga_13
   CB    CA    C    ga_13
   CA    CB    CG    ga_15
   CB    CG    CD    ga_15
   CG    CD    CE    ga_15
   CD    CE    NZ    ga_15
   CE    NZ    HZ    ga_18
   CE    NZ    CH    ga_31
   HZ    NZ    CH    ga_32
   NZ    CH   OI1    ga_54
   NZ    CH   NI2    ga_53
  OI1    CH   NI2    ga_54
   CH   NI2  HI21    ga_23
   CH   NI2  HI22    ga_23
 HI21   NI2  HI22    ga_24
   CA    C    O    ga_30
   CA    C    +N    ga_19
    O    C    +N    ga_33
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA    H    gi_1
   CA    N    C    CB    gi_2
    C    CA    +N    O    gi_1
   NZ    CH    CE    HZ    gi_1
   CH   NI2   OI1    NZ    gi_1
  NI2  HI21  HI22    CH    gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C    N    CA    gd_14
   -C    N    CA    C    gd_44
   -C    N    CA    C    gd_43
    N    CA    CB    CG    gd_34
    N    CA    C    +N    gd_45
    N    CA    C    +N    gd_42
   CA    CB    CG    CD    gd_34
   CB    CG    CD    CE    gd_34
   CG    CD    CE    NZ    gd_34
   CD    CE    NZ    CH    gd_39
   CE    NZ    CH   NI2    gd_14
   NZ    CH   NI2  HI21    gd_14

; carboxylysine (-1)
[ KCA ]
 [ atoms ]
    N    N   -0.31000    0
    H    H    0.31000    0
   CA   CH1    0.00000    1
   CB   CH2    0.00000    1
   CG   CH2    0.00000    2
   CD   CH2    0.00000    2
   CE   CH2    0.00000    3
   NZ    N   -0.31000    4 ; from the peptide bond
   HZ    H    0.31000    4 ; from the peptide bond
   CH    C    0.27000    5 ; from the carboxyl group (of e.g., GLU)
  OI1   OM   -0.63500    5 ; from the carboxyl group (of e.g., GLU)
  OI2   OM   -0.63500    5 ; from the carboxyl group (of e.g., GLU)
    C    C      0.450    6
```

```
   O    O    -0.450    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   CE   gb_27
  CE   NZ   gb_21
  NZ   HZ   gb_2
  NZ   CH   gb_10
  CH  OI1   gb_6
  CH  OI2   gb_6
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CD   CE   NZ    ga_15
  CE   NZ   HZ    ga_18
  CE   NZ   CH    ga_31
  HZ   NZ   CH    ga_32
  NZ   CH  OI1    ga_22
  NZ   CH  OI2    ga_22
 OI1   CH  OI2    ga_38
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  NZ   CH   CE   HZ    gi_1
  CH  OI2  OI1   NZ    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ   CH    gd_39
  CE   NZ   CH  OI1    gd_14

; carboxylysine (0)
[ KCN ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    2
  CD  CH2    0.00000    2
  CE  CH2    0.00000    3
  NZ    N   -0.31000    4 ; from the peptide bond
  HZ    H    0.31000    4 ; from the peptide bond
  CH    C    0.33000    5 ; from the carboxyl group (of e.g., GLUH)
 OI1    O   -0.45000    5 ; from the carboxyl group (of e.g., GLUH)
 OI2   OA   -0.28800    5 ; from the carboxyl group (of e.g., GLUH)
 HI2    H    0.40800    5 ; from the carboxyl group (of e.g., GLUH)
   C    C      0.450    6
   O    O     -0.450    6
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   CE   gb_27
  CE   NZ   gb_21
  NZ   HZ   gb_2
  NZ   CH   gb_10
  CH  OI1   gb_5
  CH  OI2   gb_13
 OI2  HI2   gb_1
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CD   CE   NZ    ga_15
  CE   NZ   HZ    ga_18
  CE   NZ   CH    ga_31
  HZ   NZ   CH    ga_32
  NZ   CH  OI1    ga_31
  NZ   CH  OI2    ga_21
 OI1   CH  OI2    ga_31
  CH  OI2  HI2    ga_12
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
  NZ   CH   CE   HZ    gi_1
  CH  OI2  OI1   NZ    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34
  CG   CD   CE   NZ    gd_34
  CD   CE   NZ   CH    gd_39
  CE   NZ   CH  OI2    gd_14
  NZ   CH  OI2  HI2    gd_12

; S-carbamoyl-cysteine
[ CAM ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    2
  SG    S    0.00000    2
  CD    C    0.29000    3 ; from ASN/GLN
 OE1    O   -0.45000    3 ; from ASN/GLN
 NE2   NT   -0.72000    3 ; from ASN/GLN
HE21    H    0.44000    3 ; from ASN/GLN
HE22    H    0.44000    3 ; from ASN/GLN
   C    C      0.450    4
   O    O     -0.450    4
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   SG   gb_32
  SG   CD   gb_32
  CD  OE1   gb_5
  CD  NE2   gb_9
 NE2  HE21  gb_2
 NE2  HE22  gb_2
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   SG    ga_16
  CB   SG   CD    ga_4
  SG   CD  OE1    ga_30
  SG   CD  NE2    ga_19
 OE1   CD  NE2    ga_33
  CD  NE2 HE21    ga_23
  CD  NE2 HE22    ga_23
HE21  NE2 HE22    ga_24
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
```

```
  CA    N    C   CB    gi_2
  CD   OE1   NE2   SG    gi_1
 NE2  HE21  HE22   CD    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   SG    gd_34
  CA   CB   SG   CD    gd_26
  CB   SG   CD  NE2    gd_40
  SG   CD  NE2  HE21    gd_14
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42

; norleucine
[ LNO ]
 [ atoms ]
   N    N   -0.31000    0
   H    H    0.31000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1 ; from aliphatic carbon atoms
  CG  CH2    0.00000    1 ; from aliphatic carbon atoms
  CD  CH2    0.00000    2 ; from aliphatic carbon atoms
  CE  CH3    0.00000    2 ; from aliphatic carbon atoms
   C    C      0.450    3
   O    O     -0.450    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_21
  CA   CB   gb_27
  CA    C   gb_27
  CB   CG   gb_27
  CG   CD   gb_27
  CD   CE   gb_27
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
  -C    N    H    ga_32
  -C    N   CA    ga_31
   H    N   CA    ga_18
   N   CA   CB    ga_13
   N   CA    C    ga_13
  CB   CA    C    ga_13
  CA   CB   CG    ga_15
  CB   CG   CD    ga_15
  CG   CD   CE    ga_15
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
  CA    N    C   CB    gi_2
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_14
  -C    N   CA    C    gd_44
  -C    N   CA    C    gd_43
   N   CA   CB   CG    gd_34
   N   CA    C   +N    gd_45
   N   CA    C   +N    gd_42
  CA   CB   CG   CD    gd_34
  CB   CG   CD   CE    gd_34

; pyruvic acid (N-terminal modification)
[ PYA ]
 [ atoms ]
  CA    C    0.45000    0 ; from the carbonyl group (of e.g., GLN)
 OB2    O   -0.45000    0 ; from the carbonyl group (of e.g., GLN)
 CB1  CH3    0.00000    1 ; from aliphatic carbon atoms
   C    C      0.450    2
   O    O     -0.450    2
 [ bonds ]
  CA  CB1   gb_27
  CA  OB2   gb_5
  CA    C   gb_23   ; shorter bond type to account for double bond effect
   C    O   gb_5
   C   +N   gb_10
 [ angles ]
; ai   aj   ak  gromos type
 CB1   CA  OB2    ga_27
 CB1   CA    C    ga_27
 OB2   CA    C    ga_27
  CA    C    O    ga_30
  CA    C   +N    ga_19
   O    C   +N    ga_33
 [ impropers ]
; ai   aj   ak   al  gromos type
  CA  OB2    C  CB1    gi_1
   C   CA   +N    O    gi_1
 [ dihedrals ]
; ai   aj   ak   al  gromos type
 CB1   CA    C   +N    gd_14


; aminoacids.n.tdb file (N-terminal parameters)

; N-methyl-AA (0)
[ 1NM ]
[ replace ]
N                NT          14.0067      -0.88
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA               CH1         13.019       0.220       0
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1          1          H1         N          CA         C
                H          1.008      0.440
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
1          4          CN1        N          CA         C
                CH3        15.035     0.220
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N          H1         gb_2
N          CN1        gb_21
[ angles ]
CN1        N          H1         ga_11
CA         N          H1         ga_11
CA         N          CN1        ga_13
[ dihedrals ]
CN1        N          CA         C          gd_29

; N-methyl-glycine (0)
[ GLY-1NM ]
[ replace ]
N                NT          14.0067      -0.88
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA               CH2         14.027       0.220       0
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1          1          H1         N          CA         C
                H          1.008      0.440
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
1          4          CN1        N          CA         C
                CH3        15.035     0.220
                ; from dimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N          H1         gb_2
N          CN1        gb_21
[ angles ]
CN1        N          H1         ga_11
CA         N          H1         ga_11
CA         N          CN1        ga_13
[ dihedrals ]
CN1        N          CA         C          gd_29

; N-methyl-proline (0)
[ PRO-1NM ]
[ replace ]
N                NT          14.0067      -0.63
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA               CH1         13.019       0.210       0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD               CH2         14.027       0.210       0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1          4          CN1        N          CA         C
                CH3        15.035     0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N          CN1        gb_21
[ angles ]
CA         N          CN1        ga_13
CD         N          CN1        ga_13
[ dihedrals ]
CN1        N          CA         C          gd_29
```

```
; N-methyl-AA (+1)
[ 1NM+ ]
[ replace ]
N                       NL          14.0067     0.104
                ; to add up to +1 net charge
CA                      CH1         13.019      0.200       0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
2           4           H           N           CA          C
                H           1.008       0.248
1           1           CN1         N           CA          C
                CH3         15.035      0.200
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N           H1          gb_2
N           H2          gb_2
N           CN1         gb_21
[ angles ]
H1          N           H2          ga_10
H2          N           CN1         ga_11
CN1         N           H1          ga_11
CA          N           H1          ga_11
CA          N           H2          ga_11
CA          N           CN1         ga_13
[ dihedrals ]
CN1         N           CA          C           gd_29

; N-methyl-glycine (+1)
[ GLY-1NM+ ]
[ replace ]
N       NL    14.0067 0.104                     ; to add up to +1 net
charge
CA        CH2   14.027  0.200         0         ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2   4   H   N   CA   C
    H   1.008  0.248
1   1   CN1  N   CA   C
    CH3  15.035  0.200                          ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N   H1   gb_2
N   H2   gb_2
N   CN1  gb_21
[ angles ]
H1   N   H2   ga_10
H2   N   CN1  ga_11
CN1  N   H1   ga_11
CA   N   H1   ga_11
CA   N   H2   ga_11
CA   N   CN1  ga_13
[ dihedrals ]
CN1  N   CA   C   gd_29

; N-methyl-proline (+1)
[ PRO-1NM+ ]
[ replace ]
N       NL    14.0067 0.152                     ; to add up to +1 net
charge
CA        CH1   13.019  0.200         0         ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD        CH2   14.027  0.200         0         ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1   4   H   N   CA   C
    H   1.008  0.248
1   1   CN1  N   CA   C
    CH3  15.035  0.200                          ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N   H   gb_2
N   CN1  gb_21
[ angles ]
H   N   CD   ga_11
CD  N   CN1  ga_13
CN1 N   H    ga_11
CA  N   H    ga_11
CA  N   CD   ga_13
CA  N   CN1  ga_13
[ dihedrals ]
CN1  N   CA   C   gd_29

; N,N-dimethyl-AA (0)
[ 2NM ]
[ replace ]
N                       NT          14.0067     -0.63
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CA                      CH1         13.019      0.210       0
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2           4           CN          N           CA          C
                CH3         15.035      0.210
                ; from trimethylamine reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N           CN1         gb_21
N           CN2         gb_21
[ angles ]
CN1         N           CN2         ga_13
CA          N           CN2         ga_13
CA          N           CN1         ga_13
[ dihedrals ]
CN1         N           CA          C           gd_29

; N,N-dimethyl-glycine (0)
[ GLY-2NM ]
[ replace ]
N       NT    14.0067 -0.63                     ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
CA        CH2   14.027  0.210         0         ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
2   4   CN   N   CA   C
    CH3  15.035  0.210                          ; from trimethylamine
reported by Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N   CN1  gb_21
N   CN2  gb_21
[ angles ]
CN1  N   CN2  ga_13
CA   N   CN2  ga_13
CA   N   CN1  ga_13
[ dihedrals ]
CN1  N   CA   C   gd_29

; N,N-dimethyl-AA (+1)
[ 2NM+ ]
[ replace ]
N                       NL          14.0067     0.152
                ; to add up to +1 net charge
CA                      CH1         13.019      0.200       0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
1           1           H1          N           CA          C
                H           1.008       0.248
2           4           CN          N           CA          C
                CH3         15.035      0.200
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N           H1          gb_2
N           CN1         gb_21
N           CN2         gb_21
[ angles ]
CN1         N           H1          ga_11
CN1         N           CN2         ga_13
CN2         N           H1          ga_11
CA          N           H1          ga_11
CA          N           CN2         ga_13
CA          N           CN1         ga_13
[ dihedrals ]
CN1         N           CA          C           gd_29

; N,N-dimethyl-glycine (+1)
[ GLY-2NM+ ]
[ replace ]
N       NL    14.0067 0.152                     ; to add up to +1 net
charge
CA        CH2   14.027  0.200         0         ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
1   1   H1   N   CA   C
    H   1.008  0.248
2   4   CN   N   CA   C
```

```
     CH3    15.035  0.200                                   ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N      H1     gb_2
N      CN1    gb_21
N      CN2    gb_21
[ angles ]
CN1    N      H1     ga_11
CN1    N      CN2    ga_13
CN2    N      H1     ga_11
CA     N      H1     ga_11
CA     N      CN2    ga_13
CA     N      CN1    ga_13
[ dihedrals ]
CN1    N      CA     C      gd_29

; N,N-dimethyl-proline (+1)
[ PRO-2NM+ ]
[ replace ]
N           NL     14.0067 0.200                           ; to add up to +1 net
charge
CA          CH1    13.019  0.200           0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
CD          CH2    14.027  0.200           0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
2      4      CN     N      CA     C
     CH3    15.035  0.200                                   ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ bonds ]
N      CN1    gb_21
N      CN2    gb_21
[ angles ]
CN1    N      CD     ga_13
CN1    N      CN2    ga_13
CN2    N      CD     ga_13
CA     N      CN2    ga_13
CA     N      CN1    ga_13
[ dihedrals ]
CN1    N      CA     C      gd_29

; N,N,N-trimethyl-AA (+1)
[ 3NM+ ]
[ replace ]
N                      NL          14.0067     0.200
                ; to add up to +1 net charge
CA                     CH1         13.019      0.200       0
                ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ add ]
3              4           CN          N           CA          C
               CH3         15.035      0.200
               ; derived by analogy to methyl groups of amines reported by
Oostenbrink et al. DOI: 10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N              CN1         gb_21
N              CN2         gb_21
N              CN3         gb_21
[ angles ]
CN1            N           CN3         ga_13
CN1            N           CN2         ga_13
CN2            N           CN3         ga_13
CA             N           CN3         ga_13
CA             N           CN2         ga_13
CA             N           CN1         ga_13
[ dihedrals ]
CN1            N           CA          C           gd_29

; N,N,N-trimethyl-glycine (+1)
[ GLY-3NM+ ]
[ replace ]
N           NL     14.0067 0.200                           ; to add up to +1 net
charge
CA          CH2    14.027  0.200           0               ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ add ]
3      4      CN     N      CA     C
     CH3    15.035  0.200                                   ; derived by analogy to
methyl groups of amines reported by Oostenbrink et al. DOI:
10.1002/cphc.200400542
[ delete ]
H
[ bonds ]
N      CN1    gb_21
N      CN2    gb_21
N      CN3    gb_21
[ angles ]
CN1    N      CN3    ga_13
CN1    N      CN2    ga_13
CN2    N      CN3    ga_13
CA     N      CN3    ga_13
CA     N      CN2    ga_13
CA     N      CN1    ga_13
[ dihedrals ]
CN1    N      CA     C      gd_29

; N-acetyl-AA
[ NAC ]
[ replace ]
N                      N           14.0067     -0.15
                ; by analogy to ASN/GLN and the peptide bond
[ add ]
1              2           H           N           C           CA
               H           1.008       0.3100
               ; from the hydrogen of the peptide bond
1              1           CN1         N           C           CA
               C           12.011      0.1900
               ; by analogy to ASN/GLN, the peptide bond and the aldehyde group
reported by Dolenc et al. DOI: 10.1093/nar/gki195
1              2           ON2         N           CA          C
               O           15.9994     -0.4500
               ; from the carbonyl group (of e.g., GLN)
1              2           CN2         N           C           CA
               CH3         15.035      0.1000
               ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N              H           gb_2
N              CN1         gb_10
CN1            ON2         gb_5
CN1            CN2         gb_27
[ angles ]
CA             N           H           ga_18
CN1            N           H           ga_32
CN1            N           CA          ga_31
N              CN1         ON2         ga_33
N              CN1         CN2         ga_19
CN2            CN1         ON2         ga_30
[ impropers ]
N              CN1         CA          H           gi_1
CN1            CN2         N           ON2         gi_1
[ dihedrals ]
CN2            CN1         N           CA          gd_14
CN1            N           CA          C           gd_39

; N-acetyl-proline
[ PRO-NAC ]
[ add ]
1              1           CN1         N           C           CA
               C           12.011      0.3500
               ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
1              2           ON2         N           CA          C
               O           15.9994     -0.4500
               ; from the carbonyl group (of e.g., GLN)
1              2           CN2         N           C           CA
               CH3         15.035      0.1000
               ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N              CN1         gb_10
CN1            ON2         gb_5
CN1            CN2         gb_27
[ angles ]
CN1            N           CD          ga_31
CN1            N           CA          ga_31
N              CN1         ON2         ga_33
N              CN1         CN2         ga_19
CN2            CN1         ON2         ga_30
[ impropers ]
N              CN1         CA          CD          gi_1
CN1            CN2         N           ON2         gi_1
[ dihedrals ]
CN2            CN1         N           CA          gd_14
CN1            N           CA          C           gd_39

; pyroglutamic acid
[ PGA-NH ]
[ replace ]
N                      N           14.0067     -0.31
[ add ]
1              1           H           N           CA          CD
               H           1.008       0.31
[ bonds ]
N              H           gb_2
[ angles ]
CA             N           H           ga_31
```

```
CD          N           H           ga_31
[ impropers ]
N           CA          CD          H           gi_1
[ dihedrals ]
CD          N           CA          C           gd_39

; N-formylmethionine
[ MET-FOR ]
[ replace ]
N                       N           14.0067     -0.15
                ; by analogy to ASN/GLN and the peptide bond
[ add ]
1           2           H           N           C           CA
                H       1.008       0.31
                ; from the hydrogen of the peptide bond
1           1           CN1         N           C           CA
                C       12.011      0.1900
                ; by analogy to ASN/GLN, the peptide bond and the aldehyde group
reported by Dolenc et al. DOI: 10.1093/nar/gki195
1           2           ON2         N           CA          C
                O       15.9994     -0.4500
                ; from the carbonyl group (of e.g., GLN)
1           2           HC1         N           C           CA
                H       1.008       0.1
                ; by analogy to the aldehyde group reported by Dolenc et al. DOI:
10.1093/nar/gki195
[ bonds ]
N           H           gb_2
N           CN1         gb_10
CN1         ON2         gb_5
CN1         HC1         gb_3
[ angles ]
CA          N           H           ga_18
CA          N           CN1         ga_31
H           N           CN1         ga_32
N           CN1         ON2         ga_33
N           CN1         HC1         ga_19
ON2         CN1         HC1         ga_30
[ impropers ]
N           CN1         CA          H           gi_1
CN1         N           ON2         HC1         gi_1
[ dihedrals ]
ON2         CN1         N           CA          gd_14
CN1         N           CA          C           gd_39


; aminoacids.c.tdb file (C-terminal parameters)

; AA-amide
[ AMD ]
[ replace ]
C                       C           12.011      0.29
                ; from ASN/GLN
[ add ]
1           1           NC1         C           N           O
                NT      14.00670    -0.7200                 ; from
ASN/GLN
1           1           HC11        C           CA          N
                H       1.008       0.4400
                ; from ASN/GLN
1           2           HC12        C           CA          N
                H       1.008       0.4400
                ; from ASN/GLN
[ bonds ]
C           NC1         gb_9
NC1         HC11        gb_2
NC1         HC12        gb_2
[ angles ]
O           C           NC1         ga_33
CA          C           O           ga_30
CA          C           NC1         ga_19
C           NC1         HC11        ga_23
C           NC1         HC12        ga_23
HC11        NC1         HC12        ga_24
[ impropers ]
C           CA          O           NC1         gi_1
NC1         C           HC11        HC12        gi_1
[ dihedrals ]
N           CA          C           NC1         gd_40
CA          C           NC1         HC11        gd_14

; AA-methyl ester
[ CME ]
[ replace ]
C                       C           12.011      0.253
                ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
O                       OE          15.9994     -0.069
                ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
[ add ]
1           2           OC1         C           CA          N
                O       15.9994     -0.45
                ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
1           2           CC2         O           C           CA
                CH3     15.035      0.266
                ; from the ester group reported by Chandrasekhar et al. DOI:
10.1081/smts-120030764
[ bonds ]
C           O           gb_13
C           OC1         gb_5
O           CC2         gb_18
[ angles ]
O           C           OC1         ga_33
C           O           CC2         ga_22
CA          C           O           ga_19
CA          C           OC1         ga_30
[ impropers ]
C           CA          OC1         O           gi_1
[ dihedrals ]
N           CA          C           O           gd_40
CA          C           O           CC2         gd_12


; aminoacids.hdb file (hydrogen database)

; phosphoserine (-1)
S1P    2
1           1           H           N           -C          CA
1           2           HE3         OE3         PD          OG
; phosphoserine (-2)
S2P    1
1           1           H           N           -C          CA
; phosphothreonine (-1)
T1P    2
1           1           H           N           -C          CA
1           2           HE3         OE3         PD          OG1
; phosphothreonine (-2)
T2P    1
1           1           H           N           -C          CA
; phosphotyrosine (-1)
Y1P    6
1           1           H           N           -C          CA
1           1           HD1         CD1         CG          CE1
1           1           HD2         CD2         CG          CE2
1           1           HE1         CE1         CD1         CZ
1           1           HE2         CE2         CD2         CZ
1           2           HI3         OI3         PT          OH
; phosphotyrosine (-2)
Y2P    5
1           1           H           N           -C          CA
1           1           HD1         CD1         CG          CE1
1           1           HD2         CD2         CG          CE2
1           1           HE1         CE1         CD1         CZ
1           1           HE2         CE2         CD2         CZ
; phosphoaspartate (-1)
D1P    2
1           1           H           N           -C          CA
1           2           HZ3         OZ3         PE          OD2
; phosphoaspartate (-2)
D2P    1
1           1           H           N           -C          CA
; phospholysine (-1)
K1P    3
1           1           H           N           -C          CA
1           1           HZ          NZ          PH          CE
1           2           HI3         OI3         PH          NZ
; phospholysine (-2)
K2P    2
1           1           H           N           -C          CA
1           1           HZ          NZ          PH          CE
; phosphoarginine (0)
R0P    5
1           1           H           N           -C          CA
1           1           HE          NE          CD          CZ
2           3           HH1         NH1         CZ          NE
1           1           HH2         NH2         PT          CZ
1           2           HI3         OI3         PT          NH2
; phosphoarginine (-1)
R1P    4
1           1           H           N           -C          CA
1           1           HE          NE          CD          CZ
2           3           HH1         NH1         CZ          NE
1           1           HH2         NH2         PT          CZ
; 1-phosphohistidine (-1)
H11    4
1           1           H           N           -C          CA
1           1           HD2         CD2         CG          NE2
1           1           HE1         CE1         ND1         NE2
1           2           HH3         OH3         PZ          NE2
; 1-phosphohistidine (-2)
H12    3
1           1           H           N           -C          CA
1           1           HD2         CD2         CG          NE2
```

```
1 1 HE1 CE1 ND1 NE2
; 3-phosphohistidine (-1)
H31 4
1 1 H N -C CA
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
1 2 HZ3 OZ3 PE3 ND1
; 3-phosphohistidine (-2)
H32 3
1 1 H N -C CA
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
; N6-methyllysine (0)
KMN 2
1 1 H N -C CA
1 2 HZ NZ CE CD
; N6-methyllysine (+1)
KMC 2
1 1 H N -C CA
2 4 HZ NZ CE CD
; N6,N6-dimethyllysine (0)
K2M 1
1 1 H N -C CA
; N6,N6-dimethyllysine (+1)
K2C 2
1 1 H N -C CA
1 2 HZ NZ CE CD
; N6,N6,N6-trimethyllysine
K3C 1
1 1 H N -C CA
; omega-N-methylarginine (0)
RMN 3
1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH2 NH2 CZ NE
; omega-N-methylarginine (+1)
RMC 4
1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH1 NH1 CZ NE
1 1 HH2 NH2 CT CZ
; symmetric-dimethylarginine (0)
RSM 3
1 1 H N -C CA
1 1 HE NE CD CZ
1 1 HH2 NH2 CT2 CZ
; symmetric-dimethylarginine (+1)
RMS 4
1 1 H N -C CA
1 1 HE NE CD CZ
1 1 HH1 NH1 CT1 CZ
1 1 HH2 NH2 CT2 CZ
; asymmetric-dimethylarginine (0)
RAM 3
1 1 H N -C CA
1 1 HE NE CD CZ
1 2 HH1 NH1 CZ NE
; asymmetric-dimethylarginine (+1)
RMA 3
1 1 H N -C CA
1 1 HE NE CD CZ
2 3 HH1 NH1 CZ NE
; 1-methylhistidine (0)
H1M 3
1 1 H N -C CA
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
; 1-methylhistidine (+1)
H1C 4
1 1 H N -C CA
1 1 HD1 ND1 CG CE1
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
; 3-methylhistidine (0)
H3M 3
1 1 H N -C CA
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
; 3-methylhistidine (+1)
H3C 4
1 1 H N -C CA
1 1 HD2 CD2 CG NE2
1 1 HE1 CE1 ND1 NE2
1 1 HE2 NE2 CE1 CD2
; N5-methylglutamine
QME 2
1 1 H N -C CA
1 1 HE2 NE2 CZ CD
; N4-methylasparagine
NME 2
1 1 H N -C CA
1 1 HD2 ND2 CE CG
; glutamate methyl ester
EME 1
1 1 H N -C CA
; aspartate methyl ester
DMA 1
1 1 H N -C CA
; S-methylcysteine
CYM 1
1 1 H N -C CA
; N6-acetyllysine
KAC 2
1 1 H N -C CA
1 1 HZ NZ CE CH
; 3-hydroxyproline (R)
PH3 1
1 2 HG1 OG1 CB CA
; 3-hydroxyproline (S)
P3H 1
1 2 HG1 OG1 CB CA
; 4-hydroxyproline (S)
HY2 1
1 2 HD1 OD1 CG CB
; 3,4-dihydroxyproline
PHH 2
1 2 HG1 OG1 CB CA
1 2 HD1 OD1 CG2 CB
; 5-hydroxylysine (0,R)
KH5 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
2 4 HZ NZ CE2 CD
; 5-hydroxylysine (0,S)
K5H 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
2 4 HZ NZ CE2 CD
; 5-hydroxylysine (+1,R)
KPH 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
3 4 HZ NZ CE2 CD
; 5-hydroxylysine (+1,S)
KHP 3
1 1 H N -C CA
1 2 HE1 OE1 CD CB
3 4 HZ NZ CE2 CD
; 3,4-dihydroxyphenylalanine
HTY 6
1 1 H N -C CA
1 1 HD1 CD1 CG CE1
1 1 HD2 CD2 CG CE2
1 2 HZ1 OZ1 CE1 CD1
1 1 HE2 CE2 CD2 CZ2
1 2 HH OH CZ2 CE1
; 7-hydroxytryptophan
W7H 7
1 1 H N -C CA
1 1 HD1 CD1 CG NE1
1 1 HE1 NE1 CD1 CE2
1 1 HE3 CE3 CD2 CZ3
1 1 HZ3 CZ3 CE3 CH3
1 1 HH3 CH3 CZ3 CZ2
1 2 HH2 OH2 CZ2 CE2
; 3-hydroxyaspartate (-1,R)
DH3 2
1 1 H N -C CA
1 2 HG1 OG1 CB CA
; 3-hydroxyaspartate (-1,S)
D3H 2
1 1 H N -C CA
1 2 HG1 OG1 CB CA
; 3-hydroxyaspartate (0,R)
DN3 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
1 2 HD2 OD2 CG2 CB
; 3-hydroxyaspartate (0,S)
D3N 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
1 2 HD2 OD2 CG2 CB
; 3-hydroxyasparagine (R)
N3H 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
2 3 HD2 ND2 CG2 CB
; 3-hydroxyasparagine (S)
NH3 3
1 1 H N -C CA
1 2 HG1 OG1 CB CA
2 3 HD2 ND2 CG2 CB
; 4-carboxyglutamate (-2)
ECA 1
1 1 H N -C CA
```

```
; 4-carboxyglutamate (-1)
ECN   2
1      1      H      N      -C     CA
1      2      HE2    OE2    CD1    CG
; sulfotyrosine
YSU   5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE1    CE1    CD1    CZ
1      1      HE2    CE2    CD2    CZ
; dehydroalanine
SDH   1
1      1      H      N      -C     CA
; 2,3-didehydrobutyrine
TDH   1
1      1      H      N      -C     CA
; 6-bromotryptophan
WBR   6
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      1      HZ2    CZ2    CE2    CH2
; S-nitrosocysteine
CSN   1
1      1      H      N      -C     CA
; citrulline
RCI   3
1      1      H      N      -C     CA
1      1      HE     NE     CD     CZ
2      3      HH2    NH2    CZ     NE
; allysine (aminoadipic semialdehyde)
KAL   2
1      1      H      N      -C     CA
1      1      HE     CE     OZ     CD
; N-acetylglucosamine (N4-linked to ASN)
NNG   6
1      1      H      N      -C     CA
1      1      HD2    ND2    CG     C1
1      1      HN2    N2     C7     C2
1      2      HO3    O3     C3     C2
1      2      HO4    O4     C4     C3
1      2      HO6    O6     C6     C5
; 2,3-dihydroxyphenylalanine
F23   6
1      1      H      N      -C     CA
1      2      HE3    OE3    CD1    CG
1      1      HD2    CD2    CG     CE2
1      2      HZ1    OZ1    CE1    CD1
1      1      HE2    CE2    CD2    CZ2
1      1      HZ2    CZ2    CE1    CE2
; 2-hydroxyphenylalanine
F2H   6
1      1      H      N      -C     CA
1      2      HE3    OE3    CD1    CG
1      1      HD2    CD2    CG     CE2
1      1      HE1    CE1    CD1    CZ
1      1      HE2    CE2    CD2    CZ
1      1      HZ     CZ     CE1    CE2
; 3-hydroxyphenylalanine
F3H   6
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      2      HZ1    OZ1    CE1    CD1
1      1      HE2    CE2    CD2    CZ2
1      1      HZ2    CZ2    CE1    CE2
; 6-hydroxytryptophan
W6H   7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      2      HI     OI     CH2    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 5-hydroxytryptophan
W5H   7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      2      HH3    OH3    CZ3    CE2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 4-hydroxytryptophan
W4H   7
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     NE1
1      1      HE1    NE1    CD1    CE2
1      2      HZ4    OZ4    CE3    CD2
1      1      HZ3    CZ3    CE3    CH2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 2-hydroxytryptophan
W2H   7
1      1      H      N      -C     CA
1      2      HE4    OE4    CD1    CG
1      1      HE1    NE1    CD1    CE2
1      1      HE3    CE3    CD2    CZ3
1      1      HZ3    CZ3    CE3    CH2
1      1      HH2    CH2    CZ3    CZ2
1      1      HZ2    CZ2    CE2    CH2
; 3-hydroxyleucine (R)
L3H   2
1      1      H      N      -C     CA
1      2      HG1    OG1    CB     CA
; 3-hydroxyleucine (S)
LH3   2
1      1      H      N      -C     CA
1      2      HG1    OG1    CB     CA
; 4-hydroxyleucine
L4H   2
1      1      H      N      -C     CA
1      2      HD3    OD3    CG     CB
; 5-hydroxyleucine (R)
L5H   2
1      1      H      N      -C     CA
1      2      HE     OE     CD2    CG
; 5-hydroxyleucine (S)
LH5   2
1      1      H      N      -C     CA
1      2      HE     OE     CD2    CG
; 3-hyroxyvaline
V3H   2
1      1      H      N      -C     CA
1      2      HG3    OG3    CB     CA
; cysteine sulfenic acid
CYH   2
1      1      H      N      -C     CA
1      2      HD     OD     SG     CB
; 5-hydroxyproline (R)
PH5   1
1      2      HE     OE     CD     CG
; 5-hydroxyproline (S)
P5H   1
1      2      HE     OE     CD     CG
; glutamic semialdehyde
GSA   2
1      1      H      N      -C     CA
1      1      HD     CD     OE     CG
; 2-amino-3-ketobutyric acid
TOX   1
1      1      H      N      -C     CA
; 2-oxo-histidine
H2X   4
1      1      H      N      -C     CA
1      1      HD1    ND1    CG     CE1
1      1      HD2    CD2    CG     NE2
1      1      HE2    NE2    CE1    CD2
; methionine sulfoxide (R)
MSX   1
1      1      H      N      -C     CA
; methionine sulfoxide (S)
MXS   1
1      1      H      N      -C     CA
; methionine sulfone
MES   1
1      1      H      N      -C     CA
; cysteine sulfinic acid
CSA   1
1      1      H      N      -C     CA
; cysteic acid
CSE   1
1      1      H      N      -C     CA
; 3-nitrotyrosine (-1)
YNI   4
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
; 3-nitrotyrosine (0)
YNN   5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
1      2      HH3    OH3    CZ2    CE1
; 3-nitrotyrosine (0)
YNB   5
1      1      H      N      -C     CA
1      1      HD1    CD1    CG     CE1
1      1      HD2    CD2    CG     CE2
1      1      HE2    CE2    CD2    CZ2
```

```
1 2 HH3 OH3 CZ2 CE1
; 6-nitrotryptophan
WNI 6
1 1 H N -C CA
1 1 HD1 CD1 CG NE1
1 1 HE1 NE1 CD1 CE2
1 1 HE3 CE3 CD2 CZ3
1 1 HZ3 CZ3 CE3 CH2
1 1 HZ2 CZ2 CE2 CH2
; kynurenine
WKY 6
1 1 H N -C CA
1 1 HE2 CE2 CD2 CZ3
2 3 HZ1 NZ1 CE1 CD2
1 1 HZ2 CZ2 CE1 CH
1 1 HZ3 CZ3 CE2 CH
1 1 HH CH CZ2 CZ3
; 3-hydroxykynurenine
WKH 6
1 1 H N -C CA
1 1 HE2 CE2 CD2 CZ3
2 3 HZ1 NZ1 CE1 CD2
1 1 HZ3 CZ3 CE2 CH2
1 2 HH1 OH1 CZ2 CE1
1 1 HH2 CH2 CZ2 CZ3
; formyl-kynurenine
WKF 7
1 1 H N -C CA
1 1 HE2 CE2 CD2 CZ3
1 1 HZ1 NZ1 CE1 CH1
1 1 HZ2 CZ2 CE1 CH2
1 1 HZ3 CZ3 CE2 CH2
1 1 HH1 CH1 NZ1 OI
1 1 HH2 CH2 CZ2 CZ3
; chlorotyrosine
YCH 5
1 1 H N -C CA
1 1 HD1 CD1 CG CE1
1 1 HD2 CD2 CG CE2
1 1 HE2 CE2 CD2 CZ2
1 2 HH OH CZ2 CE1
; homocitrulline
KAM 3
1 1 H N -C CA
1 1 HZ NZ CH CE
2 4 HI2 NI2 CH NZ
; carboxylysine (-1)
KCA 2
1 1 H N -C CA
1 1 HZ NZ CH CE
; carboxylysine (0)
KCN 3
1 1 H N -C CA
1 1 HZ NZ CH CE
1 2 HI2 OI2 CH NZ
; S-carbamoyl-cysteine
CAM 2
1 1 H N -C CA
2 3 HE2 NE2 CD SG
; norleucine
LNO 1
1 1 H N -C CA
```

Dataset S3. UniProt entries of post-translationally modified proteins. The dataset is excluded from the thesis due to its size, but it is available online as a part of the publication.

Chapter II

Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications

Margreitter, C.,* Petrov, D.* & Zagrovic, B. (2013). *Nucleic Acids Res.* **41** (Web Server issue), W422-6.

*The authors contribute equally to this work.

CM, DP and BZ conceived and designed the study. CM, DP and BZ designed and CM and DP created the back-end module, while CM, DP and BZ designed and CM created the front-end module of the web server. CM, DP and BZ wrote the paper.

ABSTRACT

Post-translational modifications (PTMs) play a key role in numerous cellular processes by directly affecting structure, dynamics and interaction networks of target proteins. Despite their importance, our understanding of protein PTMs at the atomistic level is still largely incomplete. Molecular dynamics (MD) simulations, which provide high-resolution insight into biomolecular function and underlying mechanisms, are in principle ideally suited to tackle this problem. However, because of the challenges associated with the development of novel MD parameters and a general lack of suitable computational tools for incorporating PTMs in target protein structures, MD simulations of post-translationally modified proteins have historically lagged significantly behind the studies of unmodified proteins. Here we present Vienna-PTM web server (http://vienna-ptm.univie.ac.at), a platform for automated introduction of PTMs of choice to protein 3D structures (PDB files) in a user-friendly visual environment. With 256 different enzymatic and non-enzymatic PTMs available, the server performs geometrically realistic introduction of modifications at sites of interests, as well as subsequent energy minimization. Finally, the server makes available force field parameters and input files needed to run MD simulations of modified proteins within the framework of the widely used GROMOS 54A7 and 45A3 force fields and GROMACS simulation package.

INTRODUCTION

Post-translational modifications (PTMs) of proteins, such as phosphorylation, acetylation, methylation, carboxylation or hydroxylation, play a key role in a variety of different cellular processes.[1,2] For example, PTMs have been shown to be important in regulating enzyme activity, ensuring proper localization of biomolecules,[3,4] modifying protein stability[5,6] or directing chromatin remodeling.[6,7] What is more, non-enzymatic PTMs, such as carbonylation or oxidation, frequently arise as a consequence of oxidative stress and are considered to be a ubiquitous mode of non-specific protein damage[8,9] involved in age-related disorders including neurodegenerative diseases, cancer and diabetes. Importantly, amino acids often undergo a significant change in their physico-chemical properties on modification, resulting sometimes in a dramatic alteration of the structure of the affected protein, its dynamics and the way it interacts with the environment.[1,10-13] Of the 20 canonical amino acids, 17 can be modified, thus creating a vast source of proteome diversification. The paramount importance of such modifications is underscored by the fact that ~5% of the human genome encodes enzymes related to PTMs.[1] However, despite their extreme biological relevance, our atomistic-level understanding of PTMs and their effect on protein structure, dynamics and interaction networks is still rudimentary.

Molecular dynamics (MD) computer simulations using semi-empirical atomistic force fields are a powerful way to study biomolecules at a single-molecule level with atomistic spatial resolution and femtosecond temporal resolution.[14,15] In particular, MD simulations allow one to study properties and processes that are not directly accessible through experiment and frequently play an important role in interpreting time- and ensemble-averaged experimental results.[15,16] What is more, the power of MD simulations in particular and computational modeling approaches in general is expected to only increase in the future because of growing computational capabilities and ever-improving models. Despite this inherent potential, simulation studies of PTMs have typically lagged behind both wet-laboratory research and simulation studies of unmodified proteins, focusing even in best cases only on a few modification types for a small subset of proteins.[9,10,12,17-22] (G. A. Khoury, J. P. Thompson and C. A. Floudas, unpublished results). The reasons for this are 2-fold. First, there are currently no computational tools allowing one to quickly and accurately modify

protein structures with PTMs of choice, a necessary prerequisite for any MD simulations of PTMs. Second, there are no self-consistent, comprehensive force field parameters for treating the large majority of protein PTMs in MD simulations. Although there exist several automated or semi-automated tools for generating MD parameters for novel groups, such as ParamChem or SwissParam for CHARMM,[23-26] q4md-forcefieldtools for AMBER/GLYCAM[27,28] or ATB for GROMOS[29] force fields, none of them focuses exclusively on PTMs or provides human-curated and validated PTM parameters.

This article focuses on the first of the aforementioned challenges by presenting Vienna-PTM web server (http://vienna-ptm.univie.ac.at), a web-based platform for introducing PTMs of choice in Protein Data Bank (PDB) structures[30] and GROMACS[31,32] structure files quickly and in a realistic fashion. Practically, adding PTMs to a structure of choice entails altering the chemical composition of select residues including deletion of unnecessary atoms, geometrically and energetically realistic addition of new atoms, renumbering of atomic indices and residue renaming. In particular, addition of new atoms to a structure can take considerable effort, as the appropriate atomic coordinates have to be determined for each individual modified amino acid and any inconsistency with force field definitions may lead to severe problems. To assist with this, Vienna-PTM web server provides an automated protein structure modification procedure including 256 chemically distinct PTM reactions whereby users are able to give their instructions through an intuitive graphical interface, limiting errors to a minimum (the workflow of the server is illustrated in Figure 1). The required time from the initial PDB structure to the energy-minimized altered structure of choice can thus be reduced to several seconds. Finally, as a repository of newly developed PTM parameters (D. Petrov, C. Margreitter, M. Grandits, C. Oostenbrink and B. Zagrovic, under review) for two widely used and extensively tested MD force fields (GROMOS 45A3[34] and 54A7[35,36]), the server also directly addresses the second challenge aforementioned. In particular, in addition to modified PDB files, the output of the server includes all relevant structure and topology files needed to run MD simulations of modified proteins using GROMACS biomolecular simulation package and one of the aforementioned two force fields.

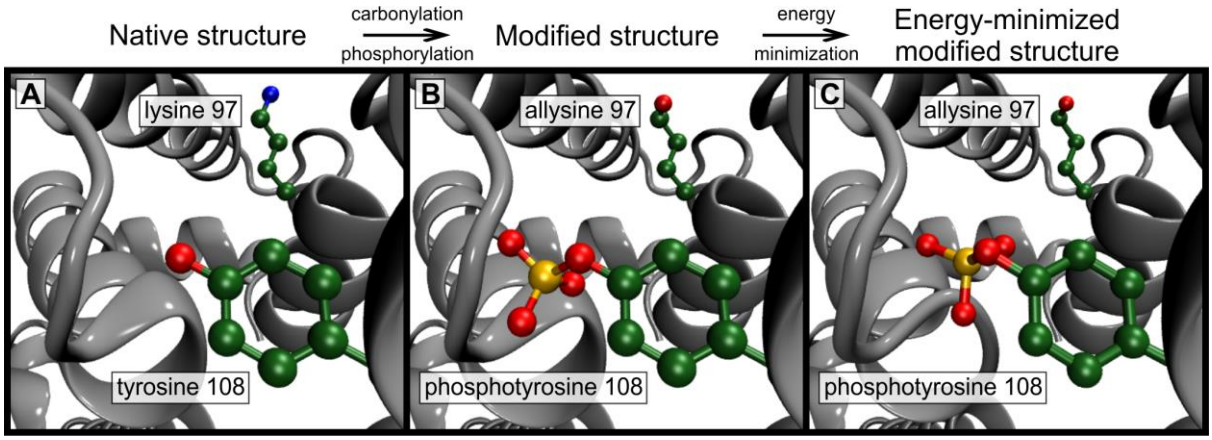


Figure 1. Vienna-PTM web server workflow. The server introduces one or more PMTs of choice to a user-supplied PDB structure followed by optional energy minimization. In the example, carbonylation and phosphorylation modifications are added to two select residues in human serum albumine (PDB code: 1N5U[33]).

MATERIALS AND METHODS

Vienna-PTM web server: input data and overall workflow

The input data that are passed to the server consists of (i) a protein X-ray or nuclear magnetic resonance structure in the form of a code-specified or manually uploaded PDB file together with processing options, such as MD force field and energy-minimization specification, and (ii) a choice of residues to be modified together with desired modifications. On upload of a protein structure in Step 1, the server redirects users to a page with the sequence from the uploaded PDB file interactively displayed either as 'pearls on a string' (Figure 2, graphical interface, JavaScript required) or a list (text-based interface) with available modifications for each residue given in drop-down menus. Depending on the interface type, selected modifications are either collected in a hidden text field or forwarded separately. In the graphical interface, a modified residue is visually labeled with a modification mark (Figure 2). When a job is submitted, the server adds, deletes, renames and renumbers atoms to apply the selected PTM(s), followed by an optional energy minimization/geometry optimization. To maximize input coverage, all statements in the input PDB file except ATOM lines are ignored in the main modification step. This also means that already modified proteins may be uploaded again. Non-canonical residues in the input PDB file are represented as an exclamation mark and cannot be modified. Finally, if one uses

a nuclear magnetic resonance structure with multiple model structures as input, the server modifies just the first model and includes it in the modified PDB file. In addition, a notification is issued on the results page informing the user of this fact. Detailed instructions can be obtained on server webpage (http://vienna-ptm.univie.ac.at/about.php), whereas support requests and reports of problems can be communicated in a user board (http://vienna-ptm.univie.ac.at/wbb).

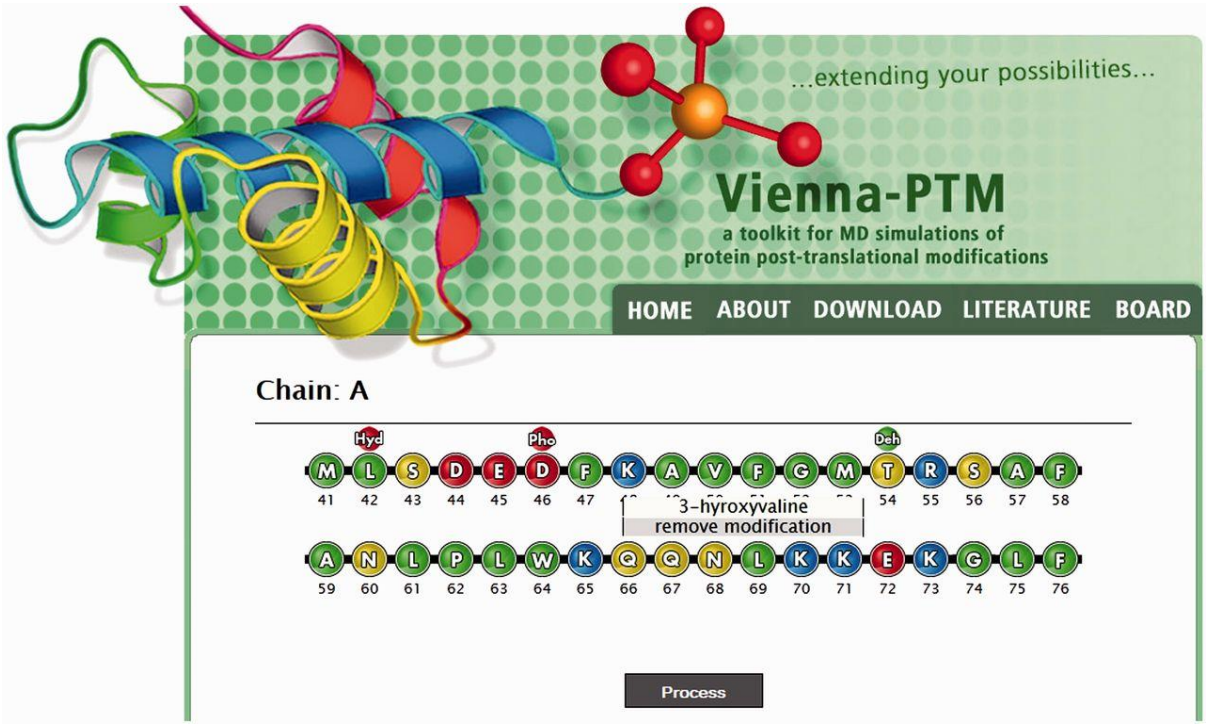


Figure 2. Data input. Modifications of choice are specified via a user-friendly graphical interface (depicted) or an optional text-based interface.

Writing configuration files, calling back-end modules and checking status of current jobs (every few seconds) are carried out by back-end processing scripts. The back-end module provides parameters for each particular modification and force field combination. New atoms are added using relative pre-minimized coordinates for the modified side chain in a coordinate system whose axes are defined taking the last remaining bond, the reference point and the last dihedral orientation into account to avoid unfavorable side-chain conformations. The modification step itself takes ~9 s on average, with minimization up to 3 min for largest systems. On the final result page, job-related information, such as the log file, is displayed to the user together with download links.

Output

The final output of the server includes (i) a three-letter-code sequence of the modified protein, (ii) a PTM-containing PDB file (visualized on the webpage using Jmol[37]), (iii) GROMACS MD simulation input files and force field parameters for simulating the modified protein, including the GROMACS structure file (.gro) and topology file (.top) and (iv) modification and energy minimization log files. In the output PDB file, the modified residues are treated the same as canonical ones, meaning that they are added in the ATOM instead of the HETATM section of the resulting PDB file. Moreover, the original HETATM entries are also included in the output and renumbered properly, together with chain information and MODEL/ENDMDL statements. Depending on user's specifications, the original header information, including REMARK and COMPND fields, is also included in the output file. Note that if energy minimization is not chosen during initialization, a modified PDB file is produced without the associated GROMACS files. Finally, one should emphasize that the server only provides files needed to prepare and run MD simulations, but not computational resources to do so.

Features of the web server

Handling of input files

Input files can be either uploaded from a local hard drive or specified by a PDB ID. In the latter case, the PDB file is automatically downloaded from http://www.pdb.org. On user's request on the initialization page, header information may be copied (not parsed) to the output PDB file. Depending on a modification, some information given in the original header may not be consistent with the modified structure; therefore, this option should be used with caution.

Available modifications

The server currently covers 256 distinct PTM reactions, including phosphorylation, methylation, acetylation, hydroxylation, carboxylation, carbonylation, nitration, deamidation and many others or 110 non-redundant post-translationally modified amino acids and protein termini. The difference between these two figures arises from the fact that a number of PTM reactions result in the same final modification. The complete list of all

available modifications and the associated chemical structures are given in Supplementary Materials, whereas the details of the parameter development and the results of the validation procedure are further discussed in reference (D. Petrov, C. Margreitter, M. Grandits, C. Oostenbrink and B. Zagrovic, under review). In principle, the number of modifications that can be applied simultaneously is limited only by the size of the protein. However, if a large number of simultaneous modifications are requested in combination with energy minimization, the time limit for a particular job may be exceeded in rare cases. If this occurs, the final output contains an unminimized structure. Finally, in all output files, the newly introduced modified residues are represented using a three-letter residue code to match PDB file format definitions (version considered: 3.20).

Energy minimization

All modified residues have been pre-energy-minimized using the GROMOS 45A3[34] and 54A7[35,36] force fields before being incorporated into the target protein. To optimize the geometry and energy of the entire modified protein, energy minimization may be requested during the initialization step. In such cases, an initial test is performed to check whether the uploaded file is suitable as input for minimization. A negative outcome leads automatically to deactivation of minimization. The reasons for failure can be inclusion of non-standard residues or unique ligands, missing atoms or residues, non-standard formatting of PDB file and others. It is the responsibility of the user to provide a suitable PDB file for energy minimization. Note, however, that the initial coordinates of newly added atoms have been pre-minimized, thus ensuring meaningful initial coordinates even if minimization of the whole molecule is disabled or fails. Energy minimization uses GROMACS routines to perform steepest gradient minimization: 1500 minimization steps are performed *in vacuo* with a maximum force convergence threshold of 1.0 kJ/mol/nm. A cut-off range of 1.4 nm is used for both the van-der-Waals and Coulomb interactions. The .mdp files used for minimization are available for download in the 'DOWNLOAD' section of the server (http://viennaptm.univie.ac.at/download.php).

Security

Job files cannot be downloaded or deleted (both uploaded and rendered) without the correct passphrase, which is generated automatically. This key is provided implicitly in links and, in case this is specified, sent to the user by email. Note that no email will be sent in case the job gets aborted. Once the job has been deleted, there is no way to recover data.

Technical details

Vienna-PTM runs on a dedicated web server with sufficient storage capacity for ∼18 000 jobs. At the moment, eight jobs can run in parallel. The job limit is due to the fact that both the server and the modification programs are executed on the same physical machine. The server software is Apache2. The front end is written using MySQL, PHP5, JavaScript, CSS and Jmol plugin,[37] whereas the back end is written in C++ (OO). GROMACS version 4.5.5 is used for energy minimization.

CONCLUSIONS

Vienna-PTM is a freely available tool, which allows rapid and reliable addition of a wide variety of PTMs to protein side chains and termini. The workflow of the server results in an output PDB file, which can be downloaded and used for simulation studies or visualization purposes. By also including molecular dynamics parameters for modified amino acids and relevant input files, Vienna-PTM web server also provides a comprehensive platform to support all key steps in setting up MD simulations of post-translationally modified proteins. The parameters are currently available for GROMOS force fields 45a3[34] and 54a7[35,36] and are provided in GROMACS file formats both for versions <4.5.x and ≥4.5.x.[31,32] Addition of new modification types and even completely new force fields to the server is logistically straightforward because of its flexible structure. Although MD parameters for several different PTMs have been developed and used before,[9,10,12,17-22] (G. A. Khoury, J. P. Thompson and C. A. Floudas, unpublished results) Vienna-PTM is to the best of our knowledge the first publicly available repository containing human-curated and validated parameters for an almost complete set of biologically relevant modifications.

The Vienna-PTM server was launched in June 2012 for testing purposes and is expected to have high visibility in MD and PTM research communities. The focus in web design was on

compatibility, preferably almost independent of the user's operating system and browser settings. In conjunction with extensive beta-testing (altogether, ~3000 test jobs have been performed by the authors and another 1000 by external beta-testers), this ensures maximal stability and user-friendliness. From direct MD simulations to biomolecular structure refinement to computational free-energy estimation and drug design, Vienna-PTM web server greatly expands the range of MD methodology to a large class of biomolecular systems of paramount importance. It is our hope that this advance will further catalyze the usage of analytical, quantitative methods of structural biophysics and chemistry, as embodied in the MD method, in addressing questions concerning realistic, PTM-dominated cell biology.

FUNDING

Austrian Science Fund FWF [START grant Y 514-B11 to B.Z.] (in part) and European Research Council Starting Independent grant [279408 to B.Z.]. Funding for open access charge: European Research Council Starting Independent grant [279408 to B.Z.].

ACKNOWLEDGEMENTS

The authors thank members of the Laboratory of Computational Biophysics at Max F. Perutz Laboratories for useful input and critical reading of the manuscript. They especially thank Alexander Zech for his help in graphics design and all beta-testers for testing the server.

REFERENCES

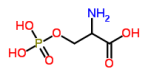
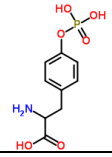
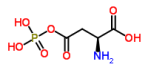
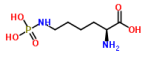
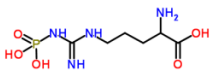
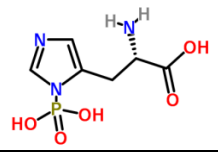
1. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342-7372 (2005).
2. Walsh, C. T. Posttranslational modification of proteins: expanding nature's inventory. Roberts and Company Publishers, Englewood, Colorado, 2006.
3. Seger, R. & Krebs, E. G. The MAPK signaling cascade. *FASEB J.* **9**, 726-735 (1995).
4. Arozarena, I., Calvo, F. & Crespo, P. Ras, an actor on many stages: Posttranslational modifications, localization, and site-specified events. *Genes Cancer* **2**, 182-194 (2011).
5. Feng, L., Lin, T., Uranishi, H., Gu, W. & Xu, Y. Functional analysis of the roles of posttranslational modifications at the p53 C terminus in regulating p53 stability and activity. *Mol. Cell. Biol.* **25**, 5389-5395 (2005).
6. Latham, J. A. & Dent, S. Y. Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.* **14**, 1017-1024 (2007).
7. Paik, W. K., Paik, D. C. & Kim, S. Historical review: the field of protein methylation. *Trends Biochem. Sci.* **32**, 146-152 (2007).
8. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *Embo J.* **24**, 1311-1317 (2005).
9. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).
10. Polyansky, A. A. & Zagrovic, B. Protein electrostatic properties predefining the level of surface hydrophobicity change upon phosphorylation. *J Phys Chem Lett* **3**, 973-976 (2012).
11. Polyansky, A. A., Kuzmanic, A., Hlevnjak, M. & Zagrovic, B. On the contribution of linear correlations to quasi-harmonic conformational entropy in proteins. *J Chem Theory Comput* **8**, 3820-3829 (2012).
12. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via lys-16 acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17857-17862 (2012).
13. Terman, J. R. & Kashina, A. Post-translational modification and regulation of actin. *Curr. Opin. Cell Biol.* (2012).
14. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646-652 (2002).
15. van Gunsteren, W. F. *et al.* Biomolecular modeling: goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
16. Kruschel, D. & Zagrovic, B. Conformational averaging in structural biology: Issues, challenges and computational solutions. *Mol Biosyst* **5**, 1606-1616 (2009).
17. Chen, H. F. Molecular dynamics simulation of phosphorylated KID post-translational modification. *PLoS ONE* **4**, e6516 (2009).
18. Grauffel, C., Stote, R. H. & Dejaegere, A. Force field parameters for the simulation of modified histone tails. *J Comput Chem* **31**, 2434-2451 (2010).
19. Seeliger, D. *et al.* Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem. Biol.* **7**, 150-154 (2012).
20. Olausson, B. E. *et al.* Molecular dynamics simulations reveal specific interactions of post-translational palmitoyl modifications with rhodopsin in membranes. *J. Am. Chem. Soc.* **134**, 4324-4331 (2012).
21. Steinbrecher, T., Latzer, J. & Case, D. A. Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput* **8**, 4405-4412 (2012).

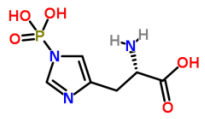
22. Marlowe, A. E., Singh, A. & Yingling, Y. G. The effect of point mutations on structure and mechanical properties of collagen-like fibril: A molecular dynamics study. *Mater Sci Eng C Mater Biol Appl* **32**, 2583 - 2588 (2012).
23. Vanommeslaeghe, K. & MacKerell Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model* **52**, 3144 - 3154 (2012).
24. Vanommeslaeghe, K. & MacKerell Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model* **52**, 3155 - 3168 (2012).
25. Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **33**, 189 - 202 (2012).
26. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam, a fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359 - 2368 (2011).
27. Vanquelef, E. *et al.* RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511-W517 (2011).
28. Dupradeau, F.-Y. *et al.* The R.ED. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* **12**, 7821-7839 (2010).
29. Malde, A. K. *et al.* An Automated force field Topology Builder (ATB) and repository: version 1.0. *J Chem Theory Comput* **7**, 4026 - 4037 (2011).
30. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242 (2000).
31. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 435-447 (2008).
32. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701-1718 (2005).
33. Wardell, M. *et al.* The atomic structure of human methemalbumin at 1.9 A. *Biochem. Biophys. Res. Commun.* **291**, 813-819 (2002).
34. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **25**, 1656-1676 (2004).
35. Schuler, L. D., Daura, X. & van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem* **22**, 1205-1218 (2001).
36. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
37. Steinbeck, C. *et al.* The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **43**, 493-500 (2003).

Appendices to Chapter II

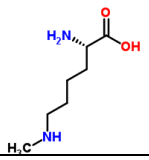
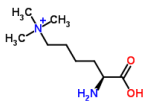
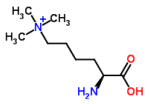
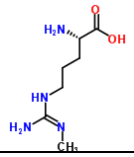
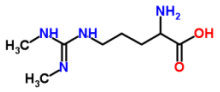
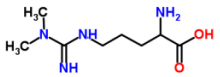
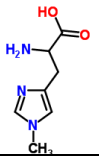
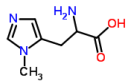
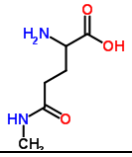
Table 1. Post-translational modifications available in Vienna-PTM with PTM index number (column 1), chemical moiety index number (column 2), amino acids they target given using the canonical 3-letter code (AA, column 3), PTM 3-letter code (column 4), chemical names (column 5) and structures (column 6). If two protonation states are possible, the one with higher occupancy at the physiological pH is highlighted in bold. * no prolines included.

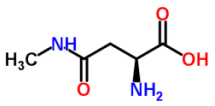
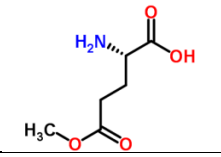
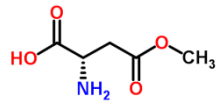
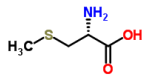
ENZYMATIC

# PTM	AA	code	chemical	structure
Phosphorylation				
1	1	SER	S1P	phosphoserine (-1)	
2	2		**S2P**	phosphoserine (-2)	
3	3	THR	T1P	phosphothreonine (-1)	
4	4		**T2P**	phosphothreonine (-2)	
5	5	TYR	Y1P	phosphotyrosine (-1)	
6	6		**Y2P**	phosphotyrosine (-2)	
7	7	ASP	D1P	phosphoaspartate (-1)	
8	8		**D2P**	phosphoaspartate (-2)	
9	9	LYS	K1P	phospholysine (-1)	
10	10		**K2P**	phospholysine (-2)	
11	11	ARG	R1P	phosphoarginine (0)	
12	12		**R2P**	phosphoarginine (-1)	
13	13	HIS	H11	1-phosphohistidine (-1)	
14	14		**H12**	1-phosphohistidine (-2)	

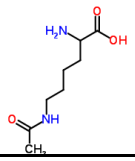
15	15	HIS	H31	3-phosphohistidine (-1)	
16	16		**H32**	3-phosphohistidine (-2)	

Methylation

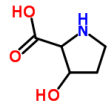
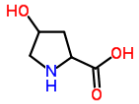
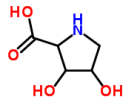
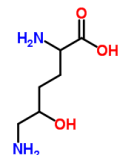
17	17	LYS	KMN	methyllysine (0)	
18	18		**KMC**	N6-methyllysine (+1)	
19	19	LYS	K2M	N6,N6-dimethyllysine (0)	
20	20		**K2C**	N6,N6-dimethyllysine (+1)	
21	21	LYS	K3C	N6,N6,N6-trimethyllysine	
22	22	ARG	RMN	omega-N-methylarginine (0)	
23	23		**RMC**	omega-N-methylarginine (+1)	
24	24	ARG	RSM	symmetric-dimethylarginine (0)	
25	25		**RMS**	symmetric-dimethylarginine (+1)	
26	26	ARG	RAM	asymmetric-dimethylarginine (0)	
27	27		**RMA**	asymmetric-dimethylarginine (+1)	
28	28	HIS	**H1M**	1-methylhistidine (0)	
29	29		H1C	1-methylhistidine (+1)	
30	30	HIS	**H3M**	3-methylhistidine (0)	
31	31		H3C	3-methylhistidine (+1)	
32	32	GLN	QME	N5-methylglutamine	

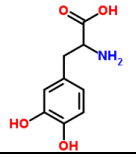
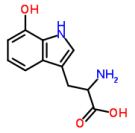
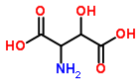
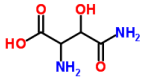
33	33	ASN	NME	N4-methylasparagine	
34	34	GLU	EME	glutamate methyl ester	
35	35	ASP	DME	aspartate methyl ester	
36	36	CYS	CYM	S-methylcysteine	

Acetylation

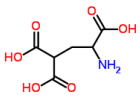
37	37	LYS	KAC	N-acetyllysine	

Hydroxylation

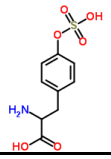
38	38	PRO	PH3	3-hydroxyproline (R)	
39	39		P3H	3-hydroxyproline (S)	
40	40	PRO	HYP	4-hydroxyproline (R)	
41	41		HY2	4-hydroxyproline (S)	
42	42	PRO	PHH	3,4-dihydroxyproline	
43	43	LYS	KH6	5-hydroxylysine (0,R)	
44	44		K6H	5-hydroxylysine (0,S)	
45	45		**KPH**	5-hydroxylysine (+1,R)	
46	46		**KHP**	5-hydroxylysine (+1,S)	

47	47	TYR	HTY	3,4-dihydroxyphenylalanine	
48	48	TRP	W7H	7-hydroxytryptophan	
49	49	ASP	**DH3**	3-hydroxyaspartate (-1,R)	
50	50		**D3H**	3-hydroxyaspartate (-1,S)	
51	51		DN3	3-hydroxyaspartate (0,R)	
52	52		D3N	3-hydroxyaspartate (0,S)	
53	53	ASN	N3H	3-hydroxyasparagine (R)	
54	54		NH3	3-hydroxyasparagine (S)	

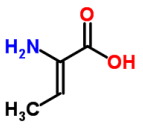
Carboxylation

55	55	GLU	**ECA**	4-carboxyglutamate (-2)	
56	56		ECN	4-carboxyglutamate (-1)	

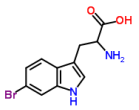
Sulfation

57	57	TYR	YSU	sulfotyrosine	

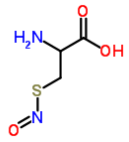
Dehydration

58	58	SER	SDH	dehydroalanine	
59	59	THR	TDH	2,3-didehydrobutyrine	

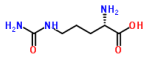
Bromidation

60	60	TRP	WBR	6-bromotryptophan	

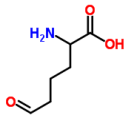
S-nitrosylation

61	61	CYS	CSN	S-nitrosocysteine	

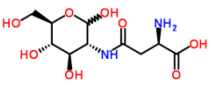
Citrullination

62	62	ARG	RCI	citrulline	

Allysine formation (the same as carbonylation)

63	63	LYS	KAL	allysine (aminoadipic semialdehyde)	

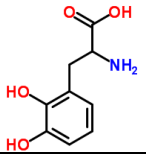
Glycosylation

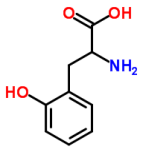
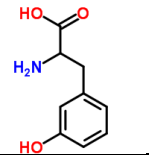
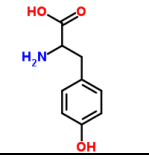
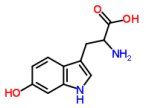
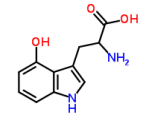
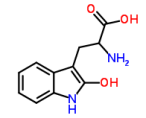
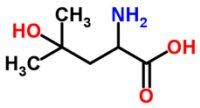
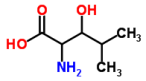
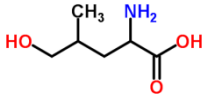
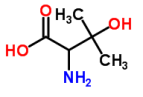
64	64	ASN	NNG	N-acetylglucosamine	

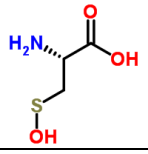
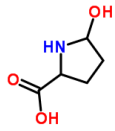
NONENZYMATIC

# PTM	AA	Code	chemical	structure

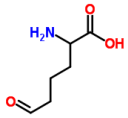
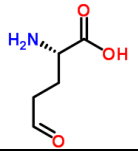
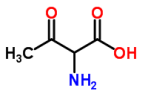
Hydroxylation

65	65	PHE	F23	2,3-dihydroxyphenylalanine	

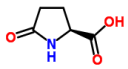
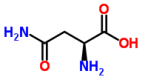
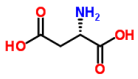
66	66	PHE	F2H	2-hydroxyphenylalanine	
67	67	PHE	F3H	3-hydroxyphenylalanine	
68	68	PHE	TYR	tyrosine	
69	69	TRP	W6H	6-hydroxytryptophan	
70	70	TRP	W5H	5-hydroxytryptophan	
71	71	TRP	W4H	4-hydroxytryptophan	
72	72	TRP	W2H	2-hydroxytryptophan	
73	73	LEU	L3H	3-hydroxyleucine (R)	
74	74		LH3	3-hydroxyleucine (S)	
75	75	LEU	L4H	4-hydroxyleucine	
76	76	LEU	L5H	5-hydroxyleucine (R)	
77	77		LH5	5-hydroxyleucine (S)	
78	78	VAL	V3H	3-hyroxyvaline	

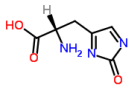
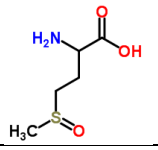
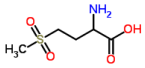
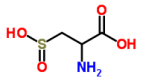
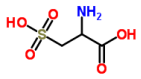
79	79	CYS	CYH	cysteine sulfenic acid	
80	80	PRO	PH5	5-hydroxyproline (R)	
81	81		P5H	5-hydroxyproline (S)	

Carbonylation

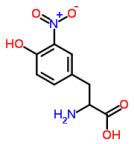
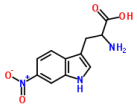
63	**63**	LYS	KAL	allysine (aminoadipic semialdehyde)	
82	82	PRO	GSA	glutamic semialdehyde	
83		ARG			
84	83	THR	TOX	2-amino-3-ketobutyric acid	

Oxidation

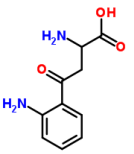
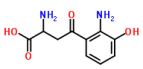
85	84	PRO	PGA	pyroglutamic acid	
86	85	HIS	ASN	asparagine	
87	86	HIS	**ASP**	aspartic acid (-1)	

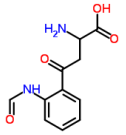
88	87	HIS	H2X	2-oxo-histidine	
89	88	MET	MSX	methionine sulfoxide (R)	
90	89		MXS	methionine sulfoxide (S)	
91	90	MET	MES	methionine sulfone	
92	91	CYS	CSA	cysteine sulfinic acid	
93	92	CYS	CSE	cysteic acid	

Nitration

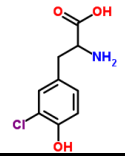
94	93	TYR	**YNI**	3-nitrotyrosine (-1)	
95	94		YNN	3-nitrotyrosine (0)	
96	95	TRP	WNI	6-nitrotryptophan	

Kynurenine formation

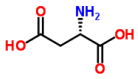
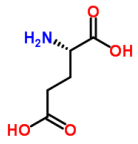
97	96	TRP	WKY	kynurenine	
98	97	TRP	WKH	3-hydroxykynurenine	

99	98	TRP	WKF	formyl-kynurenine	

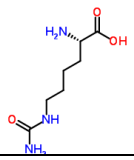
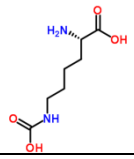
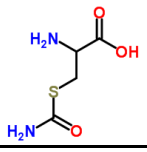
Chlorination

100	99	TYR	YCH	chlorotyrosine	

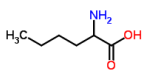
Deamidation

101	86	ASN	**ASP**	aspartic acid (-1)	
102	100	GLN	**GLU**	glutamic acid (-1)	

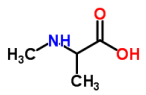
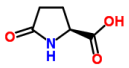
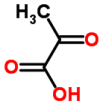
Carbamylation

103	101	LYS	KAM	homocitruline	
104	102	LYS	**KCA**	carboxylysine (+1)	
105	103		KCN	carboxylysine (0)	
106	104	CYS	CAM	S-carbamoyl-cysteine	

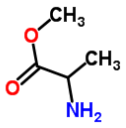
Norleucine

107	105	LEU	LNO	norleucine	
108		LYS			
109		MET			

N-TERMINAL

# PTM	AA	Code	chemical	structure
Methylation				
110-129	106	1NM	N-methyl-AA (0)	
130-149	107	**1NM+**	N-methyl-AA (+1)	
150-168	108	2NM	N,N-dimethyl-AA (0)	
169-188	109	**2NM+**	N,N-dimethyl-AA (+1)	
189-207	110	all*	3NM+	N,N,N-trimethyl-AA	
Acetylation				
208-227	111	all	NAC	N-acetyl-AA	
Pyrrolidone formation				
228	84	**GLN**	PGA	pyroglutamic acid	
229		GLU		
Formylation				
230	112	MET	FOR	N-formylmethionine	
Pyruvate formation				
231	113	**SER**	PYA	pyruvic acid	
232		CYS		
233		VAL		

C-TERMINAL

# PTM	AA	Code	chemical	structure
Amidation				
234-253	114	all	AMD	AA-amide 
Methylation				
254	115	**CYS**	CME	AA-methyl ester 
255		LEU		
256		LYS		

Chapter III

Microscopic analysis of protein oxidative damage: effect of carbonylation on structure, dynamics, and aggregability of villin headpiece

Petrov, D. & Zagrovic, B. (2011). *J. Am. Chem. Soc.* **133** (18), 7016-7024.

DP and BZ conceived and designed the study. DP collected and DP and BZ analyzed the data. DP and BZ wrote the paper.

ABSTRACT

One of the most important irreversible oxidative modifications of proteins is carbonylation, the process of introducing a carbonyl group in reaction with reactive oxygen species. Notably, carbonylation increases with the age of cells and is associated with the formation of intracellular protein aggregates and the pathogenesis of age-related disorders such as neurodegenerative diseases and cancer. However, it is still largely unclear how carbonylation affects protein structure, dynamics, and aggregability at the atomic level. Here, we use classical molecular dynamics simulations to study structure and dynamics of the carbonylated headpiece domain of villin, a key actin-organizing protein. We perform an exhaustive set of molecular dynamics simulations of a native villin headpiece together with every possible combination of carbonylated versions of its seven lysine, arginine, and proline residues, quantitatively the most important carbonylable amino acids. Surprisingly, our results suggest that high levels of carbonylation, far above those associated with cell death *in vivo*, may be required to destabilize and unfold protein structure through the disruption of specific stabilizing elements, such as salt bridges or proline kinks, or tampering with the hydrophobic effect. On the other hand, by using thermodynamic integration and molecular hydrophobicity potential approaches, we quantitatively show that carbonylation of hydrophilic lysine and arginine residues is equivalent to introducing hydrophobic, charge-neutral mutations in their place, and, by comparison with experimental results, we demonstrate that this by itself significantly increases the intrinsic aggregation propensity of both structured, native proteins and their unfolded states. Finally, our results provide a foundation for a novel experimental strategy to study the effects of carbonylation on protein structure, dynamics, and aggregability using site-directed mutagenesis.

INTRODUCTION

Proteins are frequently modified by different reactions involving reactive oxygen species (ROS), including metal-catalyzed carbonylation, oxidation of aromatic and sulfur-containing amino acid residues, oxidation of the protein backbone, or even protein fragmentation due to backbone breakage.[1-3] One of the most important mechanisms of oxidative damage to proteins is metal-catalyzed carbonylation, where ROS are created in a Fenton-type reaction involving transition metals such as iron or copper.[4] Quantitatively the most important products of carbonylation of amino acids are aminoadipic semialdehyde (Asa), derived from lysine, and glutamic semialdehyde (Gsa), derived from arginine and proline residues (Figure 1a).[5,6] Albeit to a much lower extent, carbonyl groups can also be introduced in proteins by direct carbonylation of threonine residues, a secondary reaction with aldehydes produced during lipid peroxidation, or a secondary reaction with carbonyl derivatives generated in reaction of reducing sugars.[1,2]

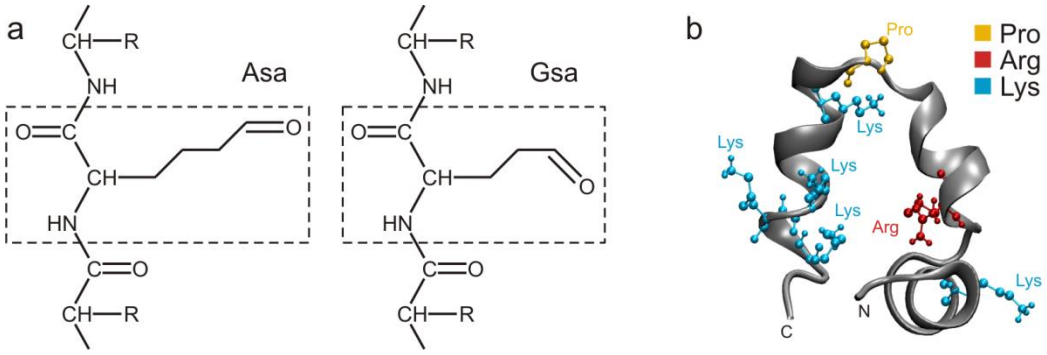


Figure 1. (a) Chemical structures of Asa, derived by carbonylation of Lys, and Gsa, derived by carbonylation of Arg and Pro. (b) Villin structure. The seven quantitatively most important carbonylable sites are colored as follows: Pro, yellow; Arg, red; Lys, blue.

As compared to other modifications induced by ROS, carbonylation is relatively difficult to induce, but once proteins get carbonylated, the change is permanent due to the irreversible nature of the reaction.[7,8] For this reason, protein carbonyl content is by far the most widely used marker of protein oxidation.[9] The cellular carbonylation level increases with age, with the concentration of carbonylated proteins rising exponentially during the last third of the life span in a wide range of organisms.[3,10] Furthermore, carbonylation is associated with age-related disorders such as neurodegenerative diseases, cancer, and diabetes, but it should be emphasized that it is still unclear whether carbonylation is a direct cause of aging or just a

consequence and a useful reporter on aging.[4] Importantly, the presence of highly carbonylated protein aggregates has been observed in many of these diseases, but so far no clear causal relationship between carbonylation and aggregation has been established.[11,12] Finally, the extreme robustness to ionizing radiation and UV light of some extremophile bacteria such as *Deinococcus radiodurans* has been shown to depend on efficient protection of the proteome against basal and radiation-induced protein carbonylation.[13]

The basal level of carbonylation in cells is approximately 2 nmol of carbonyl per milligram of protein, while the level that appears to correlate with cell and organism death is approximately 6 nmol of carbonyl per milligram of protein, corresponding to about one carbonylated residue per 4000 amino acids on average.[3,10,13] Although these average numbers are actually relatively low, it has been speculated that cellular aging is a direct consequence of the loss of structural stability, unfolding, and exposure of hydrophobic residues of select proteins upon carbonylation.[11] However, to the best of our knowledge, only a few high-resolution studies of structural stability of proteins in the face of oxidative stress have been carried out, and not one focusing explicitly on carbonylation. Most studies focused on the gross functional and structural consequences of oxidation, but with typically very little atomistic details.[14-20] A common denominator of most of these studies is that oxidation of amino acids leads to local disruption of tertiary structure of proteins with a concomitant exposure of hydrophobic amino acids and subsequent aggregation.

Here, we use molecular dynamics (MD)[21] simulation to model the carbonylated villin headpiece molecule. Villin is a tissue-specific actin-binding protein involved in different functions such as cell motility, definition of morphology, and cell death,[22] and it carries out these functions by bundling, nucleating, capping, and severing actin filaments.[23] Our interest in villin was motivated by two principal rationales. First, oxidative stress in non-muscle mammalian cells is known to cause major changes in cellular morphology and structure of the actin cytoskeleton.[24,25] Second, the 36-residue C-terminal headpiece domain of villin is one of the most widely studied and best understood proteins when it comes to folding mechanism and stability.[26-33] Here, we study atomistic-level changes in structure and dynamics of the villin headpiece at different carbonylation levels. In addition to specific effects, such as the disruption of a key salt bridge and a proline kink or alteration of the

hydrophobic effect, we ask what overall level of carbonylation can be tolerated without major effects on the molecule's structural and dynamical integrity. Finally, we ask how does carbonylation affects the intrinsic aggregability of the molecule by altering the basic physicochemical properties of the affected amino acids.

MATERIALS AND METHODS

We have used MD to study the stability of the carbonylated villin headpiece domain (sequence: MLSDEDF**K**AVFGMT**R**SAFANL**P**LW**K**QQNL**KK**E**K**GLF). The seven bolded letters mark the quantitatively most important carbonylable amino acids (K, R, and P) in villin headpiece (Figure 1b). The simulations were run using the Gromacs 3.3.3 biomolecular simulation package. United-atom GROMOS 45A3 force field,[34,35] SPC explicit water,[36] and a 2 fs integration step were used. For electrostatics calculations, reaction field was employed with a cutoff of r_c = 1.4 nm and the dielectric constant of ε_{rf} = 65. An NMR structure of the villin headpiece domain (PDB code 1VII) was used as the starting structure.[26] After steepest descent minimization was performed in vacuum (500 steps) and subsequently in water (1500 steps), the system was equilibrated by gradually increasing the temperature (from 100 to 300 K) over 100 ps with gradually decreasing position restraints (from 25 000 to 5000 kJ mol^{-1} nm^{-2}) at constant volume and temperature, and finally additionally equilibrated for 20 ps at constant pressure and temperature of 1 bar and 300 K. The temperature and pressure in all production simulations were kept at 300 K and 1 bar using a Berendsen thermostat (τ_T = 0.05 ps) and barostat (τ_p = 1 ps and compressibility = 4.5 x 10^{-5} bar^{-1}).[37]

Since the villin headpiece domain has seven potential carbonylation sites belonging to the quantitatively most important carbonylable residues K, R, and P (see sequence above), the total number of all possible carbonylation combinations of these residues is 2^7 = 128. Altogether, 136 independent 110-ns-long trajectories were generated for a total of 14.96 µs of simulation time (five copies of the native and completely carbonylated villin headpiece plus one copy of every other combination: 5 + 5 + 126 = 136). Coordinates were output every 100 ps, and for the analysis of equilibrium properties, only the last 25 ns of each trajectory was used. Carbonylation modifications were introduced before energy minimization by changing the villin headpiece coordinate file and introducing parameters for the two carbonylated amino acids, Asa and Gsa, in the force field. Details of

parametrization for Asa and Gsa are given in the Supporting Information (SI). For analysis, the simulations were divided into four sets: (1) simulations in set K include different levels of carbonylation of Lys residues, (2) those in set KR include Arg15 carbonylation together with different levels of Lys carbonylation, (3) those in set KP include carbonylation of Pro22 together with different levels of Lys carbonylation, and (4) those in set KRP include simultaneous Arg15 and Pro22 carbonylation with different levels of Lys carbonylation. The trajectories were analyzed primarily using Gromacs tools,[38] including calculation of root-mean-square deviation (rmsd), solvent-accessible surface area (SASA), molecular volume, and distances between groups of atoms, except for (1) secondary structure analysis, where DSSP[39] was used, (2) conformational entropy calculations, where quasiharmonic approximation for calculating entropy was used as described in refs 40 and 41, and (3) characterization of hydrophobic properties of protein surface, where the molecular hydrophobicity potential (MHP) calculation was employed, as described by Efremov *et al.*[42] Throughout the paper, atom-positional rmsd after rotational-translational fitting was calculated with respect to the native NMR villin headpiece structure (PDB code 1VII, residues 43-74; the two residues at each end of the peptide were excluded in calculation as they tend to be dynamic). The PDB structure 1VII starts with the residue 41 and ends with the residue 76, since the complete villin headpiece is 76 residues long, while the reported structure captures just the 36-residues-long C-terminal domain. We used 1-36 numbering here, using backbone atoms for alignment and all atoms for rmsd calculation, if not stated otherwise. Exposure of residues to water was estimated by calculating solvent accessibility fraction, calculated as a fraction of SASA calculated for a given side chain in the context of the villin headpiece structure as compared with the SASA of the same side chain when completely exposed to solvent. Distance between given residues was calculated as the distance between centers of masses of their side chains.

We used the formula proposed by Chiti *et al*.[43] to calculate the change in aggregability upon carbonylation (for structurally destabilized proteins):

$$ln(v_{mut}/v_{wt}) = A\Delta Hydr + B\left(\Delta\Delta G_{coil\,\alpha} + \Delta\Delta G_{\beta\,coil}\right) + C\Delta Charge \qquad (1)$$

where v_{mut} and v_{wt} are rates of forming aggregates, Δ*Hydr* is the change in hydrophobicity according to the hydrophobicity scale based on water-to-octanol partition,[43] $\Delta\Delta G_{coil\text{-}\alpha}$ +

$\Delta\Delta G_{\beta\text{-coil}}$ is the change in propensity to form an α-helix over a β-sheet, Δ*Charge* is the change in the absolute value of protein net charge, and *A* = 0.633, *B* = 0.198, and *C*=-0.491 are fitted constants. We used only the first term and the last term, with water/octanol partitioning hydrophobicity values for Asa and Gsa obtained from the MHP calculation described by Efremov *et al.*[42] and the correlation given in Figure 5a (below), but also provided upper and lower bounds derived from the standard deviation of α-helix and β-sheet propensities over the 27 protein studied by Chiti *et al.*[43] We used the "Zyggregator"[44] model to study the carbonylation-induced change in intrinsic aggregability directly from the native state of the villin headpiece. For this calculation, Asa and Gsa residues were replaced by amino acids that most closely match them in terms of charge and MHP, i.e., Leu and Val, respectively. We could not use Asa and Gsa directly, as not all parameters (such as α and β propensities) needed for prediction are available at this point.

RESULTS

Effect of Carbonylation on Protein Structural and Dynamical Integrity

The structure of the villin headpiece domain is principally stabilized by (1) three phenylalanines buried in the core of the protein, keeping its α-helices together,[32] and (2) a salt bridge between Asp4 and Arg15 residues.[33] Simultaneous carbonylation of all seven major carbonylable residues in the protein results in the loss of approximately 40% of its α-helical secondary structure and most of its tertiary structure in approximately 100 ns (Figure 2a,b). While the simulated native protein keeps its α-helical content and tertiary structure largely intact over the same period, the carbonylated protein unfolds and starts exploring the accessible areas of the Ramachandran map more freely. Unfolding of the protein upon carbonylation occurs simultaneously with the disintegration of its Phe core (Phe7, Phe11, and Phe18) (Figure 2c,d). Interestingly, the solvent exposure of the Phe core increases multiple-fold as a consequence of carbonylation (Figure 2c), while the total SASA of the protein remains the same although the core residues become exposed to the surface.

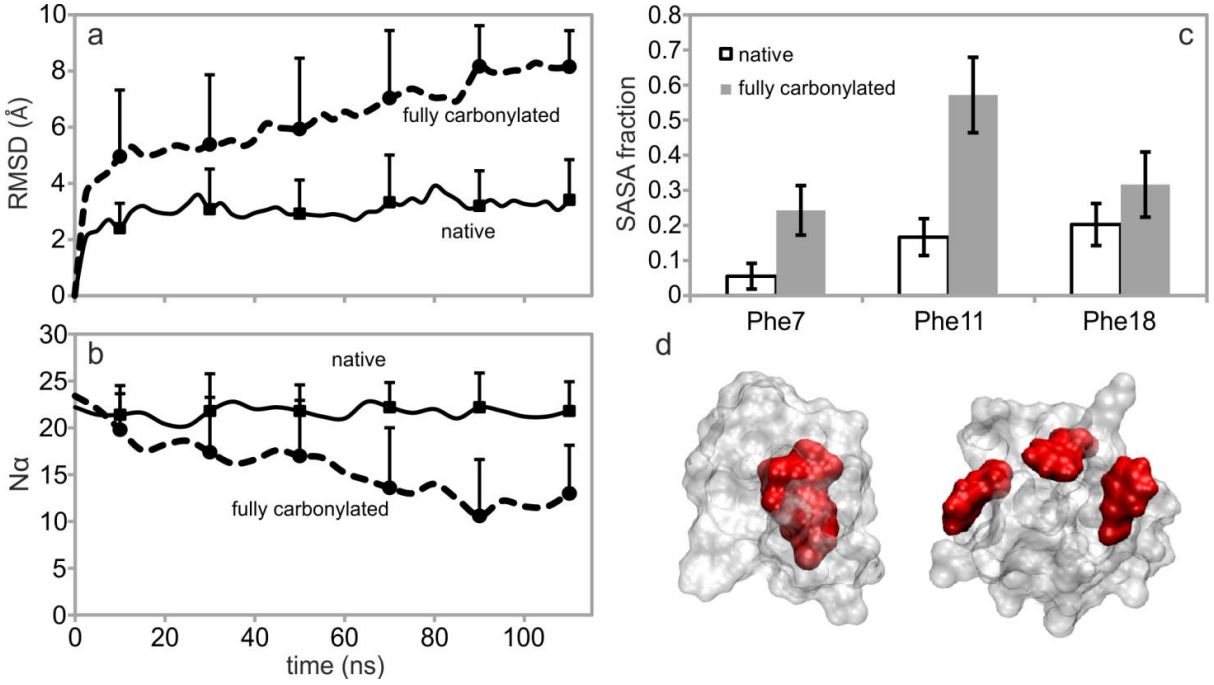


Figure 2. Complete carbonylation leads to unfolding of the villin headpiece. (a) All-atom rmsd from the native villin headpiece structure 1VII (residues 43-74) as a function of time. (b) The number of residues in R-helical conformation as a function of time. In both plots the solid line represents native villin headpiece while the dashed line represents fully carbonylated villin headpiece simulations. (c) Total SASA of the three core Phe residues normalized by the total SASA of fully solvent-exposed Phe. Empty bars represent native villin headpiece while gray bars represent fully carbonylated villin headpiece simulations. All curves and bars in panels a-c were obtained as averages over five independent native/fully carbonylated simulations and are shown with standard deviations. (d) A representative structure of native (left) and completely carbonylated (right) villin headpiece showing the surface of the three core phenylalanines in red and the rest of the protein in white.

Table 1. Types of trajectories studied. Each set contains a given carbonylation event plus every possible combination of carbonylating the 0-5 Lys residues in the villin headpiece. Set K includes five simulations of the native villin headpiece, i.e., a combination with no Lys residues carbonylated. Similarly, set KRP contains five simulations of the completely carbonylated villin headpiece, i.e., a combination with all five Lys residues carbonylated.

SET	# TRAJ.	CARB. RES.	EVALUATED EFFECTS
K	36	0-5 Lys	Δ hydrophobicity
KR	32	0-5 Lys + Arg	salt bridge disruption + Δ hydrophobicity
KP	32	0-5 Lys + Pro	proline kink disruption + Δ hydrophobicity
KRP	36	0-5 Lys + Arg + Pro	all three effects combined

What happens if just a subset of different carbonylable residues in villin headpiece are carbonylated, and how does this depend on the type and position of the residues affected? To address this, we have divided our simulations into four distinct sets which were analyzed separately (Table 1). The four sets were organized in such a way to probe different structural effects of carbonylation: surface hydrophobicity change, Asp4-Arg15 salt bridge removal, proline kink removal, or a combination thereof. Interestingly, the structure remains intact when it comes to tertiary fold (as measured by rmsd from the native structure), core compactness (as measured by the sum of distances between the centers of mass of core phenylalanines, Σ*d*Phe), and secondary structure (as measured by the number of α-helical residues) for all single and double carbonylation hits, regardless of their type or location (Figure 3a-c).

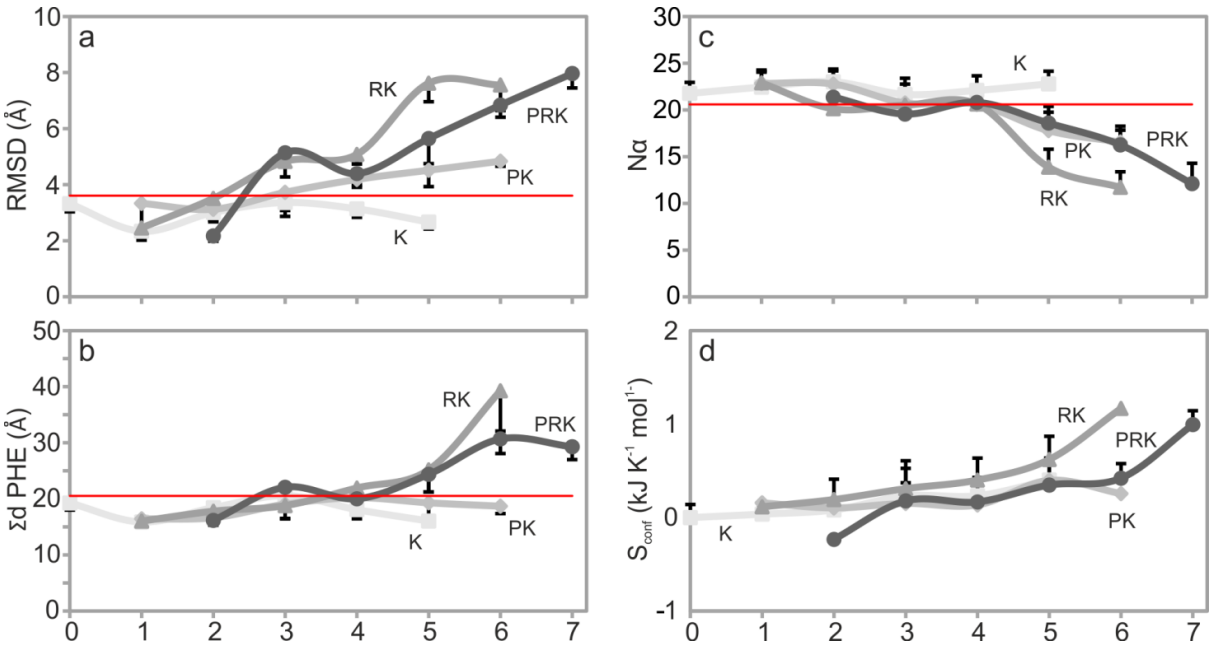


Figure 3. Structural stability of the villin headpiece as a function of carbonylation level: (a) rmsd from the native villin headpiece structure 1VII, (b) sum of the distances between the three core phenylalanines, (c) number of residues in α-helical conformation, and (d) quasi-harmonic conformational entropy normalized by number of atoms, relative to that of native villin headpiece, all shown as a function of the number of carbonylated residues. Each line is obtained from a different subset of simulations: the lightest lines, set K; dark gray lines, set KR; light gray lines, set KP; the darkest lines, set KRP. All points in the plots were calculated as averages over the last 25 ns in each simulation and all simulations in a given subset with the same number of carbonylated residues and are shown with standard deviations. Red lines represent averages of given observables over the five native simulations, increased by the standard deviations, and were used as a criterion for determining whether a given structure is folded or unfolded.

Moreover, the structural features remain largely unchanged with regard to all structural measures, even with all five lysine residues carbonylated (K set) (Figure 3a-c). This is striking as Asa, the product of carbonylation of Lys residues, is significantly less hydrophilic than Lys itself (i.e., its hydration free energy is more positive, SI Figure S1), which could lead to a significant disruption of the stabilizing hydrophobic effect. Our thermodynamic integration (TI) calculations suggest that ΔG_{hydr} between them exceeds 10 kJ mol^{-1} in uncharged form and 230 kJ mol^{-1} in charged form (SI Figure S1), with the carbonylation of Arg having a similar effect.

For all three structural measures of foldedness (rmsd, core compactness, and α-helicity), an ensemble of structures with a given number of carbonylated residues (regardless of their type or position) was defined to be unfolded if the average value of the measure in question, reduced by its standard deviation, was greater that its average over the five native trajectories, increased by its respective standard deviation (e.g., if $\overline{rmsd^{carb}} - \sigma_{rmsd^{carb}} > \overline{rmsd^{nat}} - \sigma_{rmsd^{nat}}$. According to this definition, significant unfolding (i.e., if $\overline{rmsd^{carb}} - \sigma_{rmsd^{carb}}$ > 3.6 Å, $\overline{\Sigma dPhe^{carb}} - \sigma_{\Sigma dPhe^{carb}}$ > 20.5 Å, and $\overline{N\alpha^{carb}} - \sigma_{N\alpha^{carb}}$ < 20.6) was observed for (1) simulations in the KR set with Arg15 and at least four Lys residues carbonylated according to α-helicity and core compactness, and in simulations with Arg15 and at least two Lys residues carbonylated according to rmsd; (2) simulations in the KP set with Pro22 and at least four Lys residues carbonylated according to α-helicity, and simulations with Pro22 and at least three Lys residues carbonylated according to rmsd; and (3) simulations in the KRP set with Arg15, Pro22, and at least three Lys residues carbonylated according to α-helicity and core compactness, and in simulations with Arg15, Pro22, and at least one Lys residue carbonylated according to rmsd. When it comes to conformational entropy, our results suggest that there is a sizable increase in conformational entropy after complete carbonylation (approximately 1 J K^{-1} mol^{-1} per atom). While this increase is consistent with unfolding of the protein, its absolute magnitude could be affected by the sampling employed. On the whole, our results suggest that globular proteins like villin headpiece likely remain structurally stable upon carbonylation, unless the carbonylation level is high. Moreover, these results suggest that the type and the position of the affected residues do make a difference in terms of the extent of structural damage

caused. These findings are further corroborated by the analysis of the average pairwise rmsd between different simulated trajectories exhibiting varying levels and types of carbonylation (Figure 4). According to this analysis, ensembles with five or more carbonylated residues are by-and-large non-native-like, and furthermore their unfolded ensembles are largely mutually different when it comes to structure.

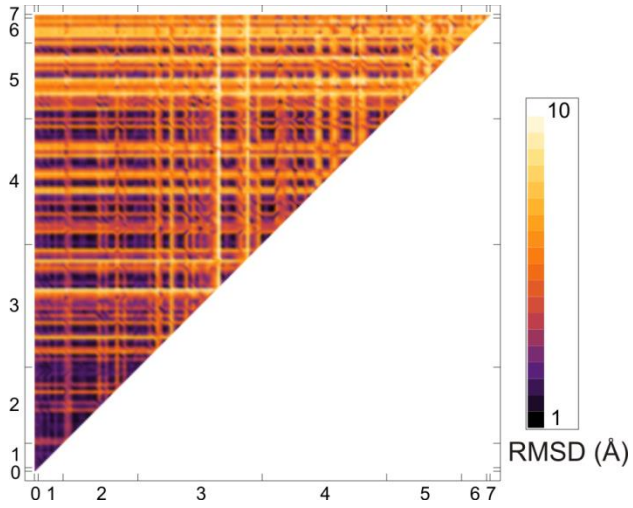


Figure 4. Pairwise rmsd density plot. Each point represents the results of comparison of two ensembles of structures from a given pair of simulations of given combinations of carbonylated residues (linear average of the distribution of all-against-all rmsd evaluations for the two ensembles). Altogether 128 combinations were analyzed, and both x and y axis range from 0 to 7 (number of carbonylated residues), where 0 represents the native ensemble and 7 represents the fully carbonylated ensemble. Backbone atoms were used for alignment and rmsd calculations.

In order to estimate to what extent carbonylation of different residues affects protein stability, we used a simple, two-state model with the difference in free energy between the folded and the unfolded states of villin headpiece represented as $\Delta G_{f\text{-}u}$. We assumed that carbonylation of each residue acts independently and that carbonylation of each Lys residue has the same thermodynamic effect. Using these assumptions, the total change in the free energy of stabilization of villin headpiece upon carbonylation, $\Delta\Delta G_{carb}$, can be expressed as

$$\Delta\Delta G_{carb} = n_K \Delta\Delta G_K + n_R \Delta\Delta G_R + n_P \Delta\Delta G_P \quad (2)$$

where $\Delta\Delta G_K$, $\Delta\Delta G_R$, and $\Delta\Delta G_P$ are free energy changes upon carbonylation of individual Lys, Arg, and Pro residues, respectively, and n_K, n_R, and n_P are the numbers of carbonylated Lys, Arg, and Pro residues, respectively. According to this,

$$\Delta\Delta G_{carb} > \Delta G_{f-u} \quad (3)$$

when the villin headpiece unfolds, and

$$\Delta\Delta G_{carb} < \Delta G_{f-u} \quad (4)$$

when the molecule remains in the native conformation upon carbonylation. We used our simulated data and traces given in Figure 3a-c, together with the definition of foldedness given above, to determine whether a given ensemble of structures is folded or unfolded. For example, one inequality, derived using core compactness as a reporter of the state of the villin headpiece (folded or unfolded), is

$$5\Delta\Delta G_K + \Delta\Delta G_P < \Delta G_{f-u} \quad (5)$$

Using the same approach, we derived 19 additional inequalities using rmsd, core compactness, and α-helicity criteria (see SI for details). This system of inequalities has no unique solution, partly because different structural measures that were used are not necessarily mutually consistent when it comes to the definition of foldedness. However, it was possible to find an approximate solution using numerical approaches (see SI for details). Average values over 1136 variations of $\Delta\Delta G_K$, $\Delta\Delta G_R$, and $\Delta\Delta G_P$ values that fulfilled the maximal 16 out of 20 inequalities are $\Delta\Delta G_K = 0.13\Delta G_{f\text{-}u}$, $\Delta\Delta G_R = 0.54\Delta G_{f\text{-}u}$ and $\Delta\Delta G_P = 0.13\Delta G_{f\text{-}u}$. In other words, carbonylation of Arg15 has by far the largest effect on the protein's stability, while the effect of carbonylating Lys and Pro residues is markedly smaller.

Effect of Carbonylation on Protein Aggregability

Since hydrophobicity is one of the most dominant properties that determines aggregability, we have compared the hydrophobicity (related to aggregability) of native Lys, Arg, and Pro residues and their carbonylated counterparts, Asa and Gsa, using the MHP approach[42] (Figure 5). MHP values correlate well with octanol/water partition factors for amino acid side-chain analogues (Figure 5a), a widely used measure of amino acid hydrophobicity. In agreement with our TI result discussed above, the intrinsic hydrophobicity of Lys and Arg residues increases dramatically upon carbonylation. In terms of MHP weights, converting Lys to Asa or Arg to Gsa is similar to mutating them to highly hydrophobic leucine and valine, respectively (Figure 5b). In fact, the similarity between Asa and Leu, and Gsa and Val, respectively, extends to other basic physicochemical properties as well (SI Table S1). The

effect of carbonylating Pro to Gsa is less dramatic, albeit still resulting in a net increase in hydrophobicity (Figure 5b). In addition to increasing hydrophobicity, carbonylation also results in charge neutralization when it comes to Lys and Arg residues, and both of these effects potentially increase the intrinsic aggregability of the affected proteins.[43-45]

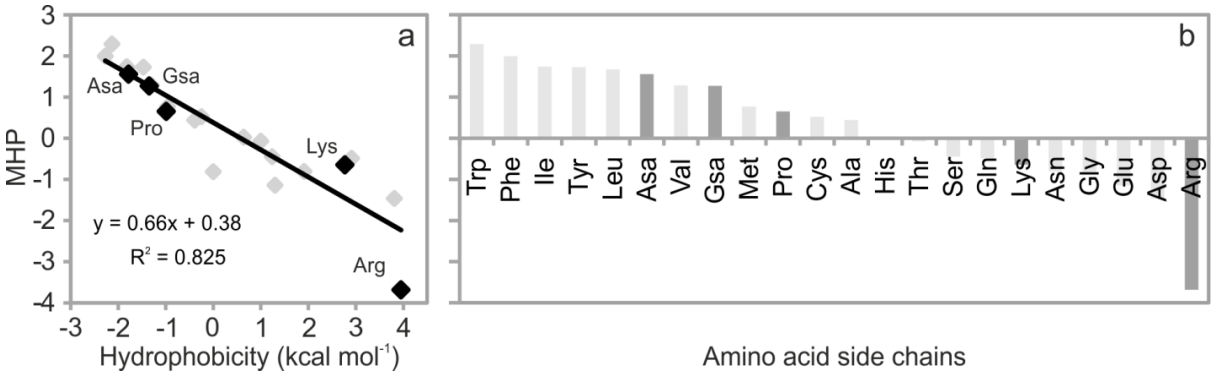


Figure 5. Molecular hydrophobicity potential (MHP) values of native and carbonylated residues. (a) Comparison of the calculated MHP values with the hydrophobicity scale based on free energy of partitioning between aqueous phase and octanol.[43] Large black diamonds represent carbonylable and carbonylated residues Lys, Arg, Pro, Asa, and Gsa, while small gray diamonds represent standard amino acid residues. The regression line (calculated without Asa and Gsa residues) with R^2 = 0.825 shows that the calculated values correlate well with hydrophobicity scale for standard amino acid residues. (b) Calculated MHP values for all native and carbonylated residues (black bars, carbonylable and carbonylated residues; gray bars, other native residues).

To analyze this more closely, we have used the MHP approach to characterize the hydrophobicity of a protein surface and compare it with the degree of the protein's structural integrity. In particular, we have studied the projections of the free energy surface for native and fully carbonylated forms of villin headpiece, as well as for an intermediate including all combinations with four carbonylated residues, using MHP, phenylalanine core compactness, and rmsd from the experimental NMR structure as order parameters (Figure 6 and SI Figure S2). Note that free energy projections using MHP as an order parameter were first used by Polyansky and Zagrovic in the context of protein phosphorylation (manuscript in preparation). These free energy maps clearly demonstrate that surface hydrophobicity and the fraction of partially unfolded, aggregation-prone structures increase as the level of carbonylation increases. However, this analysis also shows that compact and structurally native-like, yet significantly hydrophobic structures are present to a high degree in ensembles with moderate levels of carbonylation (Figure 6b). This is a direct consequence of

the fact that Asa and Gsa are significantly more hydrophobic compared with Lys and Arg, respectively, which at moderate levels of carbonylation may not be enough to unfold the molecules but is enough to increase their surface hydrophobicity. On the other hand, only unfolded structures with high surface hydrophobicity are populated in the fully carbonylated ensemble (Figure 6c and SI Figure S2). Taken together, these results suggest that carbonylation increases surface hydrophobicity through unfolding, but it does so also just by itself by modifying the physicochemical properties of the affected residues, which consequently may lead to an increase in aggregability.

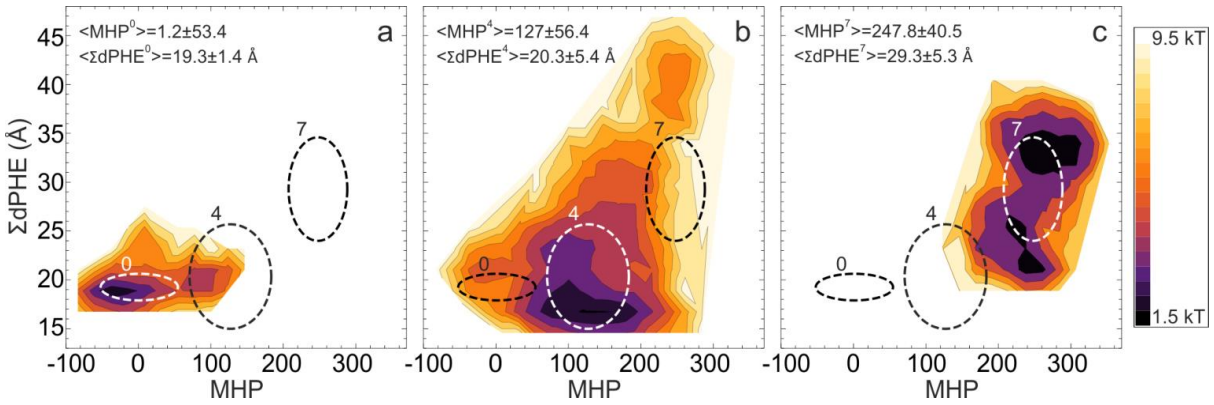


Figure 6. Two-dimensional projections of the free-energy surface as a function of MHP and hydrophobic core compactness (as captured by the sum of the distances between the three core phenylalanines) as calculated for (a) the native ensemble (five simulations), (b) the ensemble containing all combinations with four carbonylations events (35 simulations), and (c) the fully carbonylated ensemble (five simulations). Only the last 25 ns of each simulation were used to calculate the free-energy maps. The ellipses are centered at the average values of MHP and Σ*d*Phe for the three populations, with the major and minor semiaxes equal to the standard deviations of the distributions. The relative free energies were calculated as $F = -kT\ln P$, where P is the probability of occurrence of a given state, as seen in our simulations.

To put this on a quantitative footing, we have used the model of Chiti *et al.*[43] to estimate the expected change in the aggregation rate of an unstructured protein as a consequence of a single point "mutation" induced by carbonylation (i.e., Lys to Asa or Arg to Gsa conversion). Changes in aggregability for these two "mutations" are $\ln(v_{mut}/v_{wt}) = 3.37 \pm 0.73$ (Lys to Asa) and $\ln(v_{mut}/v_{wt}) = 3.85 \pm 0.73$ (Arg to Gsa), i.e., more than a 30- to 40-fold increase in the aggregation rate for a single carbonylation event, with the error bounds capturing the expected effects of the change in the intrinsic α-helix and β-sheet propensity, not modeled here. Note that this model is sequence-independent and that these changes in aggregability

are applicable not just to the villin headpiece but rather to any given positively charged protein. Carbonylation increases aggregability of negatively charged proteins as well, but to a somewhat lower extent ($\ln(v_{mut}/v_{wt}) = 2.39 \pm 0.73$ for Lys to Asa, and $\ln(v_{mut}/v_{wt}) = 2.87 \pm 0.73$ for Arg to Gsa), which is still extremely significant. In Figure 7a, we compare this change in aggregability with the predicted change in aggregation rate for 27 different point mutations in seven different unstructured peptides or intrinsically disordered proteins causing diabetes and different neurodegenerative diseases (amylin, two prion peptides, α-synuclein, amyloid β-peptide, tau protein, and a model protein) examined by Chiti *et al.*[43] Note that the latter predictions correlate well with experimental measurements (Pearson $r = 0.85$). Strikingly, the effect of carbonylation of a single Lys or Arg residue is comparable in magnitude with the most drastic aggregation-inducing mutations in these proteins, with only 2 out of 27 such mutations having a greater effect on protein aggregability than either one of these carbonylation induced conversions.

Finally, to probe the effect of carbonylation on protein aggregability directly from the native state, we have used the model of Vendruscolo *et al.*,[44] replacing Asa and Gsa with amino acids that most closely match them in terms of charge and MHP hydrophobicity, i.e., Leu and Val, respectively. These approximations notwithstanding, it appears that carbonylation of a single residue indeed significantly increases intrinsic aggregability of villin headpiece even in the absence of any major unfolding (Figure 7b). For example, the carbonylation of Pro22 increases its Zagg score, i.e., the aggregability of the region around this residue, from an unfavorable -1 to a highly favorable 0.5. In this model, the Zagg score reflects the combined effect of the intrinsic aggregation propensity of a given sequence and its tendency to be structurally unstable in the native state. Our analysis (Figure 7b and SI Figure S3), showing a dramatic increase in Zagg upon carbonylation throughout the villin sequence, suggests that both of these factors change significantly upon carbonylation in the direction that favors aggregation.

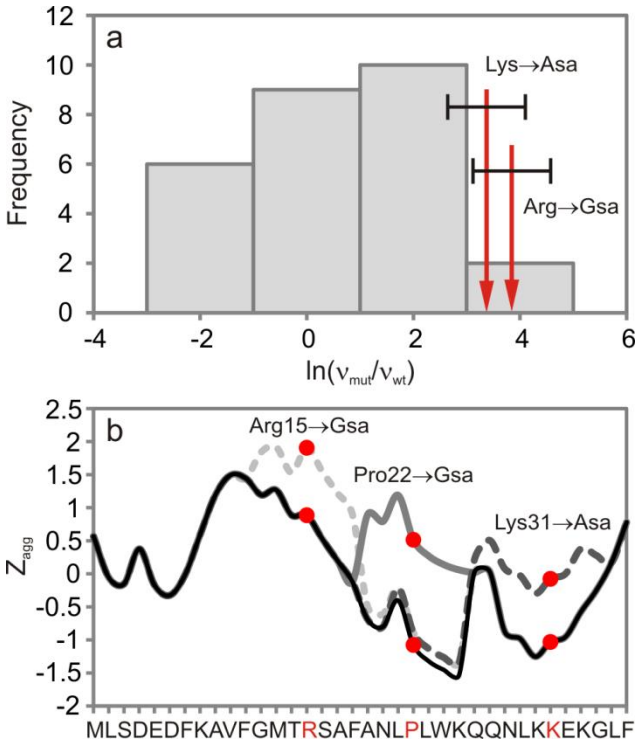


Figure 7. Aggregability versus carbonylation. (a) Distribution of changes in aggregability upon 27 single-point mutations in six known proteins causing amyloidogenic disease and in one model protein. Red arrows represent the estimation of aggregability change upon Lys (left arrow, $\ln(\nu_{mut}/\nu_{wt}) = 3.37 \pm 0.73$) and Arg (right arrow, $\ln(\nu_{mut}/\nu_{wt}) = 3.85 \pm 0.73$) carbonylation, calculated using eq 5. The error bars show standard deviation around zero calculated for the change in α-helix and β-sheet propensities for all proteins used by Chiti *et al.*[43] (b) Site-specific intrinsic aggregation propensity, Zagg, for the native (solid black line) and three carbonylated villin headpiece molecules (only single carbonylation hits: Arg15 to Gsa, the lightest gray dotted line; Pro22 to Gsa, dark gray solid line; and Lys31 to Asa, the darkest gray dashed line).

DISCUSSION

This study presents the first-ever computer simulation effort to analyze the effects of carbonylation on protein structure. We have shown that the structure of the completely carbonylated villin headpiece domain unfolds in the course of simulated trajectories. Further analysis showed that extent of destabilization is site-specific, and that the disruption of both specific stabilizing elements (salt bridges and proline kinks) together with surface hydrophobicity change upon carbonylation of a large number of residues is required for protein unfolding. Note, however, that the effect of surface hydrophobicity change cannot be disentangled from other specific effects: for example, the carbonylation event, which disrupts salt-bridge formation, invariably also changes the hydrophobicity of the involved

residue. Carbonylation of the proline residue is the only case where such a binary effect is avoided, as the native and the carbonylated forms of the residue exhibit similar levels of hydrophobicity.

The level of carbonylation at which the villin headpiece, a marginally stable protein,[46] unfolds in our simulations (ca. 14% of all residues) is ~3 orders of magnitude greater than the average level of carbonylation which correlates with cellular senescence and death in vivo (ca. 0.025% of all residues).[3,10,13] This suggests that typical cellular levels of carbonylation are not likely to disrupt protein structure, which is not in contradiction with experimental studies.[16,18] If carbonylation does not induce major structural destabilization of a typical protein and thus lead to aggregation, as widely assumed,[17,47] how does it then promote the formation of cytotoxic aggregates? First, several studies suggest that some proteins are more susceptible to carbonylation than others,[9,48-50] which may result in a situation where these proteins are completely carbonylated, while the majority of other proteins are still intact, in agreement with our results. Second, newly synthesized proteins in an old organism might be misfolded or unfolded due to, for example, accumulated DNA mutations and would therefore be more susceptible to carbonylation, as has been shown by Dukan *et al.*[51] Finally, modifications of specific, functionally important residues may abolish the function of a protein without destabilizing its structure.

We propose two novel possibilities to explain this. First, our results suggest that carbonylation just by itself may increase protein aggregability of unstructured polypeptides (Figure 7a). In this scenario, protein unfolding arises because of either carbonylation or some other environmental factors, but importantly, carbonylation makes the unfolded molecule additionally aggregation-prone. Second, our results suggest that carbonylation may increase protein aggregability even in the absence of major unfolding (Figure 7b). From lysozyme to superoxide dismutase to prolactin, there are numerous examples of proteins which undergo amyloid formation under native conditions and without any major unfolding, simply as a consequence of local thermal fluctuations.[52] Increasing the local propensity to aggregate through carbonylation may speed this process up.

Using TI and MHP approaches, we have shown that the products of carbonylation of Lys and Arg residues are significantly more hydrophobic than these amino acids. Such an increase in

hydrophobicity, together with a concomitant charge removal, significantly increases the intrinsic aggregability of proteins.[43-45] On the other hand, the aggregation propensity of proline residues is decreased by precluding β-sheet formation, and their mutations are known to lead to protein aggregation.[53] Finally, recent findings that the three quantitatively most important carbonylable residues (Arg, Lys, and Pro) in many proteins actually serve as key gatekeepers that flank aggregation-prone regions of proteins and prevent aggregation further support our suggestions.[54]

In general, the list of amyloidogenic, deposition diseases shows a strong overlap with the list of diseases in which high levels of protein carbonylation are detected, including Alzheimer's disease, Parkinson's disease, diabetes, rheumatoid arthritis, dementia with Lewy bodies, familial amyotrophic lateral sclerosis, and others.[7,9,52,55] Because of the complex pathophysiology of these diseases, the causal link between carbonylation and aggregation is not simple and is still to be fully explored. In fact, not all aggregation-prone proteins associated with these diseases are at the same time highly carbonylable and vice versa. However, there are a number of important examples where this is precisely so. For example, ubiquitin C-terminal hydrolase L1 (UCH-L1), which features critically in sporadic variants of Alzheimer's and Parkinson's diseases, forms aggregates but is also highly carbonylable.[7,56] Superoxide dismutase 1 in familial amyotrophic lateral sclerosis,[7,52] human β2-microglobulin in end-stage renal failure,[52,57] and β-actin and α/β-tubulin in multiple sclerosis[58] are further examples of such proteins. Importantly, our results provide a novel explanation that links aggregation and carbonylation in these systems, even at relatively low levels of carbonylation typically seen *in vivo*. One way to test this connection experimentally would be to use Lys to Leu, Arg to Val, and Pro to Val point mutations for high-resolution, site-specific studies of the structural and functional consequences of carbonylation, as these mutations, according to our results, quantitatively match the effects of hydrophobicity increase and charge removal in the course of carbonylation. Such studies, hand in hand with further computational analyses, may provide an atomistic picture behind protein carbonylation and its cellular consequences, protein aggregation, and cell senescence.

ASSOCIATED CONTENT

Supporting information. Figures showing TI, MHP vs rmsd free energy surfaces, and Zagg results; table of physicochemical properties of Asa, Gsa, Leu, and Val; parametrization of Asa and Gsa residues; and details of TI calculation.

ACKNOWLEDGEMENTS

This work was supported in part by the National Foundation of Science, Higher Education and Technological Development of Croatia (EMBO Installation grant to B.Z.), the Unity through Knowledge Fund (UKF 1A grant to B.Z.), and the Austrian Science Fund FWF (START grant Y 514-B11 to B.Z.). We thank A. Polyansky for help with MHP calculations, R. Zubac for help with entropy calculations, R. Santos for help with graphical representation of villin headpiece, and the members of the Laboratory of Computational Biophysics at MFPL for useful advice and assistance.

REFERENCES

1. Stadtman, E. R. Protein oxidation and aging. *Science* **257**, 1220-1224 (1992).
2. Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**, 20313-20316 (1997).
3. Levine, R. L. & Stadtman, E. R. Oxidative modification of proteins during aging. *Exp. Gerontol.* **36**, 1495-1502 (2001).
4. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* **24**, 1311-1317 (2005).
5. Requena, J. R., Levine, R. L. & Stadtman, E. R. Recent advances in the analysis of oxidized proteins. *Amino Acids* **25**, 221-226 (2003).
6. Maisonneuve, E. *et al.* Rules governing selective protein carbonylation. *PLoS ONE* **4**, e7269 (2009).
7. Dalle-Donne, I., Giustarini, D., Colombo, R., Rossi, R. & Milzani, A. Protein carbonylation in human diseases. *Trends Mol. Med.* **9**, 169-176 (2003).
8. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. & Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **52**, 601-623 (2006).
9. Dalle-Donne, I. *et al.* Protein carbonylation, cellular dysfunction, and disease progression. *J. Cell. Mol. Med.* **10**, 389-406 (2006).
10. Levine, R. L. Carbonyl modified proteins in cellular regulation, aging, and disease. *Free Radic. Biol. Med.* **32**, 790-796 (2002).
11. Grune, T., Jung, T., Merker, K. & Davies, K. J. Decreased proteolysis caused by protein aggregates, inclusion bodies, plaques, lipofuscin, ceroid, and 'aggresomes' during oxidative stress, aging, and disease. *Int. J. Biochem. Cell Biol.* **36**, 2519-2530 (2004).
12. Davies, K. J. Degradation of oxidized proteins by the 20S proteasome. *Biochimie* **83**, 301-310 (2001).
13. Krisko, A. & Radman, M. Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14373-14377 (2010).
14. Davies, K. J. A. & Delsignore, M. E. Protein damage and degradation by oxygen radicals. III. Modification of secondary and tertiary structure. *J. Biol. Chem.* **262**, 9908-9913 (1987).
15. Gardner, P. R., Nguyen, D. D. H. & White, C. W. Aconitase is a sensitive and critical target of oxygen poisoning in cultured-mammalian-cells and in rat lungs. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12248-12252 (1994).
16. Lasch, P. *et al.* Hydrogen peroxide-induced structural alterations of RNAse A. *J. Biol. Chem.* **276**, 9492-9502 (2001).
17. Bota, D. A. & Davies, K. J. A. Lon protease preferentially degrades oxidized mitochondrial aconitase by an ATP-stimulated mechanism. *Nat. Cell Biol.* **4**, 674-680 (2002).
18. Li, D. W. & Bruschweiler, R. All-atom contact model for understanding protein dynamics from crystallographic B-factors. *Biophys. J.* **96**, 3074-3081 (2009).
19. Perez, V. I. *et al.* Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3059-3064 (2009).
20. Wu, W., Wu, X. J. & Hu, Y. F. Structural modification of soy protein by the lipid peroxidation product acrolein. *Lebensm. Wiss. Technol.* **43**, 133-140 (2010).
21. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
22. Khurana, S. & George, S. P. Regulation of cell structure and function by actin-binding proteins: villin's perspective. *FEBS Lett.* **582**, 2128-2139 (2008).

23. Friederich, E., Vancompernolle, K., Louvard, D. & Vandekerckhove, J. Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro. *J. Biol. Chem.* **274**, 26751-26760 (1999).
24. Dalle-Donne, I. *et al.* Actin carbonylation: from a simple marker of protein oxidation to relevant signs of severe functional impairment. *Free Radic. Biol. Med.* **31**, 1075-1083 (2001).
25. Dalle-Donne, I., Rossi, R., Milzani, A., Di Simplicio, P. & Colombo, R. The actin cytoskeleton response to oxidants: from small heat shock protein phosphorylation to changes in the redox state of actin itself. *Free Radic. Biol. Med.* **31**, 1624-1632 (2001).
26. McKnight, C. J., Matsudaira, P. T. & Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **4**, 180-184 (1997).
27. Duan, Y. & Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**, 740-744 (1998).
28. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927-937 (2002).
29. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R. & Pande, V. S. Native-like mean structure in the unfolded ensemble of small proteins. *J. Biol. Chem.* **323**, 153-164 (2002).
30. Meng, J. *et al.* High-resolution crystal structures of villin headpiece and mutants with reduced F-actin binding activity. *Biochemistry* **44**, 11963-11973 (2005).
31. Lei, H., Wu, C., Liu, H. & Duan, Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4925-4930 (2007).
32. Cornilescu, G. *et al.* Solution structure of a small protein containing a fluorinated side chain in the core. *Protein Sci.* **16**, 14-19 (2007).
33. Gronwald, W., Hohm, T. & Hoffmann, D. Evolutionary Pareto-optimization of stably folding peptides. *BMC Bioinformatics* **9**, 109 (2008).
34. Schuler, L. D., Walde, P., Luisi, P. L. & van Gunsteren, W. F. Molecular dynamics simulation of n-dodecyl phosphate aggregate structures. *Eur. Biophys. J.* **30**, 330-343 (2001).
35. Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
36. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed B. Pullman) 331-342 (Reidel, Dordrecht, 1981).
37. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
38. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **7**, 306-317 (2001).
39. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
40. Andricioaei, I. & Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys* **115**, 6289-6292 (2001).
41. Levy, R. M., Srinivasan, A. R., Olson, W. K. & McCammon, J. A. Quasi-harmonic method for studying very low-frequency modes in proteins. *Biopolymers* **23**, 1099-1112 (1984).
42. Efremov, R. G. *et al.* Molecular lipophilicity in protein modeling and drug design. *Curr. Med. Chem.* **14**, 393-415 (2007).
43. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808 (2003).

44. Tartaglia, G. G. *et al.* Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).
45. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
46. Godoy-Ruiz, R. *et al.* Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J Phys Chem B* **112**, 5938-5949 (2008).
47. Pan, J. C., Yu, Z. H., Hui, E. F. & Zhou, H. M. Conformational change and inactivation of arginine kinase from shrimp Feneropenaeus chinensis in oxidized dithiothreitol solutions. *Biochem. Cell Biol.* **82**, 361-367 (2004).
48. Das, N., Levine, R. L., Orr, W. C. & Sohal, R. S. Selectivity of protein oxidative damage during aging in Drosophila melanogaster. *Biochem. J* **360**, 209-216 (2001).
49. Jana, C. K., Das, N. & Sohal, R. S. Specificity of age-related carbonylation of plasma proteins in the mouse and rat. *Arch. Biochem. Biophys.* **397**, 433-439 (2002).
50. Choi, J. *et al.* Oxidative modifications and aggregation of Cu,Zn-superoxide dismutase associated with Alzheimer and Parkinson diseases. *J. Biol. Chem.* **280**, 11648-11655 (2005).
51. Dukan, S. *et al.* Protein oxidation in response to increased transcriptional or translational errors. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5746-5749 (2000).
52. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15-22 (2009).
53. Williams, A. D. *et al.* Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J. Biol. Chem.* **335**, 833-842 (2004).
54. Rousseau, F., Serrano, L. & Schymkowitz, J. W. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Biol. Chem.* **355**, 1037-1047 (2006).
55. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
56. Choi, J. *et al.* Oxidative modifications and down-regulation of ubiquitin carboxyl-terminal hydrolase L1 associated with idiopathic Parkinson's and Alzheimer's diseases. *J. Biol. Chem.* **279**, 13256-13264 (2004).
57. Michelis, R., Sela, S. & Kristal, B. Intravenous iron-gluconate during haemodialysis modifies plasma beta(2)-microglobulin properties and levels. *Nephrol. Dial. Transplant.* **20**, 1963-1969 (2005).
58. Bizzozero, O. A. in *Handbook of neurochemistry and molecular neurobiology. Brain and spinal cord trauma*, 3rd ed., Banik, N. L., Ray, S. K., Lajtha, A., eds.; Chapter 23, Springer: New York ; London, 2009

Appendices to Chapter III

Parameterization of Asa and Gsa residues

Lysine residues were changed into Asa by removing the hydrogen atoms (HZ1, HZ2 and HZ3) bonded to the nitrogen atom NZ in the side chain, and by replacing the nitrogen atom (NZ) with oxygen; arginine residue was changed into Gsa by removing all side chain atoms, except the three carbon atoms (CB, CG and CD) and the nitrogen atom NE, which was replaced with an oxygen atom; finally the proline residue was changed into Gsa the same way like arginine by replacing the hydrogen atom HD2 bound to CD carbon atom with oxygen. The double bond between carbon and oxygen atoms, and the potential energy term for an angle between a triplet of one oxygen and two carbon atoms in the GROMOS 45A3 force field[1] are described always using the same parameters regardless of the type of carbon or oxygen atoms (e.g., the carboxyl group in the backbone). In order to be internally consistent, we used the same bond parameters in the description of the bond between the atoms CD and oxygen in Asa and the atoms CG and oxygen in Gsa, and the same angle parameters in the description of the angle between the last two carbon atoms and oxygen atom both in Asa and Gsa. All other bonded parameters in Asa and Gsa residues were the same as in lysine and arginine residues, respectively. GROMOS 45A3 building block files for Asa and Gsa are given at the end of the Supporting Information.

Thermodynamic integration:

We used thermodynamic integration (TI)[2] for the calculation of hydration free energies for all amino-acid side-chain analogues and two carbonylated amino-acid side-chain analogues in order to estimate the difference in hydration free-energy between Lys and Asa; and Arg and Gsa. Hydration free energies of neutral and charged forms of native and carbonylated amino-acid side-chain analogues with the CB atom CH_n replaced by CH_{n+1}, were calculated using TI. Since only hydration free energies of neutral forms of amino-acid side-chain analogues are experimentally measurable, we used them for direct comparison between calculated and experimental values.

The equilibration and free energy calculations in water were carried out using the same conditions as described above for protein simulations, while calculations *in vacuo* were

carried out at 300 K (Berendsen thermostat and τ_T = 0.05 ps were used) without periodical boundary condition, using the simple cut-off method for calculating electrostatics with a cutoff of r_c=1.4 nm. Non-bonded interactions of side chains were scaled down to zero in a stepwise manner using a coupling parameter λ. Free energy changes upon removal of non-bonded interactions were calculated as integrals of the averages of the derivatives of the total system Hamiltonian with respect to λ, between the boundaries λ=0 and λ=1,

$$\Delta G = \int_0^1 \langle \frac{\partial H}{\partial \lambda} \rangle_\lambda \, d\lambda. \quad (1)$$

After initially 26 evenly spaced λ-points were sampled, changes in slope at each point of the $\langle \frac{\partial H}{\partial \lambda} \rangle_\lambda$ versus λ graph were calculated. The number of additional λ-points placed between two given neighbor λ-points in the second step was proportional to the sum of slope changes at these points, for a total of 26 additional λ-points. The slope changes for the first and the last λ-point were considered to be the same as the slope changes in the second and the penultimate λ-point respectively. Trapezoidal integration was used to evaluate the integral in equation (1) using 52 λ-points. Sampling of 50 ps of equilibration and 200 ps of data collection at each point were used. In order to avoid singularities in the non-bonded interaction a soft-core interaction was used

$$V_{sc}(r) = (1-\lambda)V^A(r_A) + \lambda V^B(r_B), \quad (2)$$

$$r_A = (\alpha\sigma_A^6\lambda^p + r^6)^{\frac{1}{6}}, \quad (3)$$

$$r_B = (\alpha\sigma_B^6(1-\lambda^p) + r^6)^{\frac{1}{6}}, \quad (4)$$

where σ_A and σ_B are van der Walls parameters and α = 1.51 and p = 1.[3,4] The hydration free energy was calculated by subtracting the free energy change when side chain was simulated *in vacuo* from the free energy change when side chain was simulated in water.

Fractional contribution of different specific factors to villin headpiece destabilization upon carbonylation

To find how much different specific factors contribute to destabilization of villin headpiece, using our simulated results, we have derived a set of 20 inequalities.

Inequalities based on RMSD:

$$5\Delta G_K < \Delta\Delta G_{f-u}, \quad (5)$$

$$\Delta G_K + \Delta G_R < \Delta\Delta G_{f-u}, \quad (6)$$

$$2\Delta G_K + \Delta G_R > \Delta\Delta G_{f-u}, \quad (7)$$

$$2\Delta G_K + \Delta G_P < \Delta\Delta G_{f-u}, \quad (8)$$

$$3\Delta G_K + \Delta G_P > \Delta\Delta G_{f-u}, \quad (9)$$

$$\Delta G_R + \Delta G_P < \Delta\Delta G_{f-u}, \quad (10)$$

$$\Delta G_K + \Delta G_R + \Delta G_P > \Delta\Delta G_{f-u}. \quad (11)$$

Inequalities based on core compactness:

$$5\Delta G_K < \Delta\Delta G_{f-u}, \quad (12)$$

$$3\Delta G_K + \Delta G_R < \Delta\Delta G_{f-u}, \quad (13)$$

$$4\Delta G_K + \Delta G_R > \Delta\Delta G_{f-u}, \quad (14)$$

$$5\Delta G_K + \Delta G_P < \Delta\Delta G_{f-u}, \quad (15)$$

$$2\Delta G_K + \Delta G_R + \Delta G_P < \Delta\Delta G_{f-u}, \quad (16)$$

$$3\Delta G_K + \Delta G_R + \Delta G_P > \Delta\Delta G_{f-u}. \quad (17)$$

Inequalities based on core α-helicity:

$$5\Delta G_K < \Delta\Delta G_{f-u}, \quad (18)$$

$$3\Delta G_K + \Delta G_R < \Delta\Delta G_{f-u}, \quad (19)$$

$$4\Delta G_K + \Delta G_R > \Delta\Delta G_{f-u}, \quad (20)$$

$$3\Delta G_K + \Delta G_P < \Delta\Delta G_{f-u}, \quad (21)$$

$$4\Delta G_K + \Delta G_P > \Delta\Delta G_{f-u}, \quad (22)$$

$$2\Delta G_K + \Delta G_R + \Delta G_P < \Delta\Delta G_{f-u}, \quad (23)$$

$$3\Delta G_K + \Delta G_R + \Delta G_P > \Delta\Delta G_{f-u}. \quad (24)$$

To find a solution to this set of inequalities numerically, we discretized the space of values for $\Delta\Delta G_K$, $\Delta\Delta G_R$ and $\Delta\Delta G_P$ to integer percentage of $\Delta G_{f\text{-}u}$ and counted the number of fulfilled inequalities for a given set of discrete values of $\Delta\Delta G_K$, $\Delta\Delta G_R$ and $\Delta\Delta G_P$, among all 1030301 variations for $0 \le \Delta\Delta G_K \le 100$, $0 \le \Delta\Delta G_R \le 100$ and $0 \le \Delta\Delta G_P \le 100$ in steps of $0.01\Delta G_{f\text{-}u}$. The number of fulfilled inequalities ranged from 6 to 16.

Asa and Gsa GROMOS 45A3 parameters

ffG45a3.rtp file:

```
[ ASA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH2    0.00000    1
  CE  CH1    0.38000    2
 OE1    O   -0.38000    2
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD   CE   gb_26
  CE  OE1   gb_4
 [ angles ]
;  ai   aj   ak  gromos type
  -C    N    H    ga_31
   H    N   CA    ga_17
  -C    N   CA    ga_30
   N   CA    C    ga_12
  CA    C   +N    ga_18
  CA    C    O    ga_29
   O    C   +N    ga_32
   N   CA   CB    ga_12
   C   CA   CB    ga_12
  CA   CB   CG    ga_14
  CB   CG   CD    ga_14
  CG   CD   CE    ga_14
  CD   CE  OE1    ga_29
 [ impropers ]
;  ai   aj   ak   al  gromos type
   N   -C   CA    H    gi_1
   C   CA   +N    O    gi_1
  CA    N    C   CB    gi_2
 [ dihedrals ]
;  ai   aj   ak   al  gromos type
 -CA   -C    N   CA    gd_4
  -C    N   CA    C    gd_19
   N   CA    C   +N    gd_20
   N   CA   CB   CG    gd_17
  CA   CB   CG   CD    gd_17
  CB   CG   CD   CE    gd_17
  CG   CD   CE  OE1    gd_20


[ GSA ]
 [ atoms ]
   N    N   -0.28000    0
   H    H    0.28000    0
  CA  CH1    0.00000    1
  CB  CH2    0.00000    1
  CG  CH2    0.00000    1
  CD  CH1    0.38000    2
 OE1    O   -0.38000    2
   C    C      0.380    3
   O    O     -0.380    3
 [ bonds ]
   N    H   gb_2
   N   CA   gb_20
  CA    C   gb_26
   C    O   gb_4
   C   +N   gb_9
  CA   CB   gb_26
  CB   CG   gb_26
  CG   CD   gb_26
  CD  OE1   gb_4
```

```
[ angles ]
; ai   aj   ak  gromos type
  -C   N    H   ga_31
   H   N   CA   ga_17
  -C   N   CA   ga_30
   N  CA    C   ga_12
  CA   C   +N   ga_18
  CA   C   O    ga_29
   O   C   +N   ga_32
   N  CA   CB   ga_12
   C  CA   CB   ga_12
  CA  CB   CG   ga_14
  CB  CG   CD   ga_14
  CG  CD  OE1   ga_29
[ impropers ]
; ai   aj   ak   al  gromos type
   N   -C   CA   H   gi_1
   C   CA   +N   O   gi_1
  CA   N    C   CB   gi_2
[ dihedrals ]
; ai   aj   ak   al  gromos type
 -CA  -C    N   CA   gd_4
  -C   N   CA    C   gd_19
   N  CA    C   +N   gd_20
   N  CA   CB   CG   gd_17
  CA  CB   CG   CD   gd_17
  CB  CG   CD  OE1   gd_20
```

ffG45a3.hdb file:

```
ASA     1
1       1       H       N       -C      CA
GSA     1
1       1       H       N       -C      CA
```

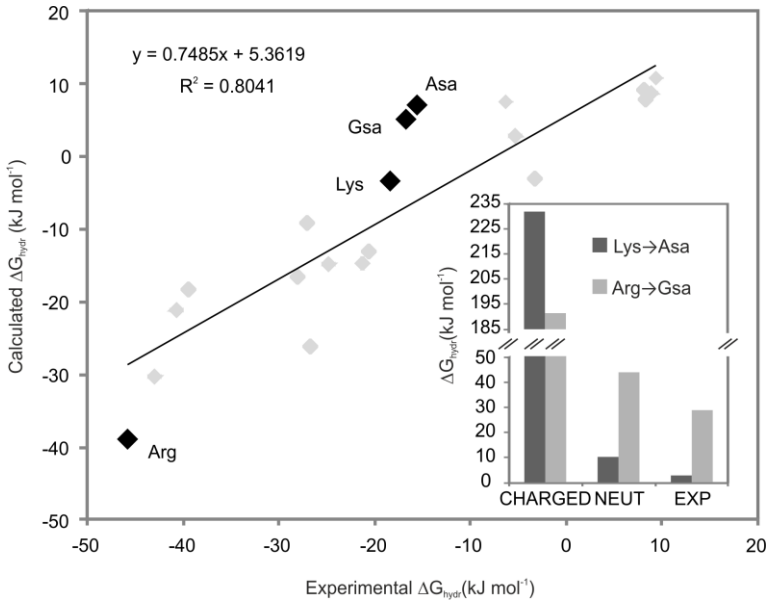


Figure S1. Hydration free energies of native and carbonylated side-chains analogues. Comparison of experimental and calculated hydration free energies using thermodynamic integration. Large black diamonds represent carbonylable side chains Lys, Arg in neutral forms and their carbonylated side chains, aminoadipic semialdehyde (Asa) and glutamic semialdehyde (Gsa), while small gray diamonds represent all other side chains. The regression line with R^2=0.804 shows that experimental and calculated values are well correlated. Inset - differences ($\Delta\Delta G_{hydr}$) between hydration free energies of Asa and Lys; and Gsa and Arg from experiment and simulation. Note that the difference in hydrophobicity is significantly greater than estimated by thermodynamic integration simulation because lysine and arginine residues at biologically relevant pH are usually in charged forms, which are much more hydrophilic than their neutral forms used here.

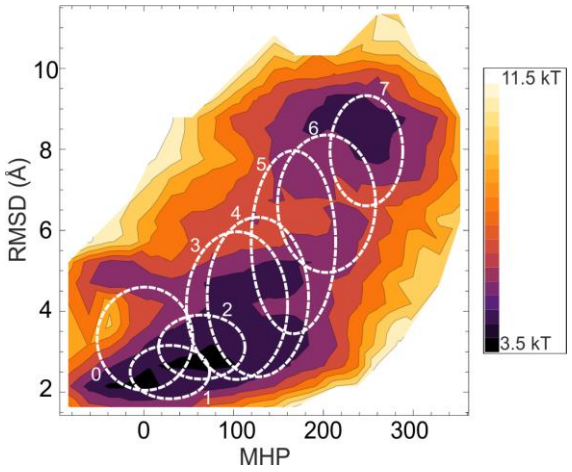


Figure S2. Two dimensional projection of the free-energy landscape as a function of MHP and RMSD from the experimental NMR structure for the last 25 ns of all the simulations. Since ensembles with different number of carbonylated residues contain structures form different number of simulations (e.g., native ensemble contains structures from 5 simulations while 5-carbonylations ensemble contains structures from 21 simulations), the fractions of total population are rescaled in such a way so that the total sums of the fractions for each

ensemble are equal. The relative free energy values were calculated as the negative logarithm of the rescaled fractions. The ellipses are centered at the average values of MHP and RMSD with the major and minor semi-axes equal to the standard deviations of the distributions.

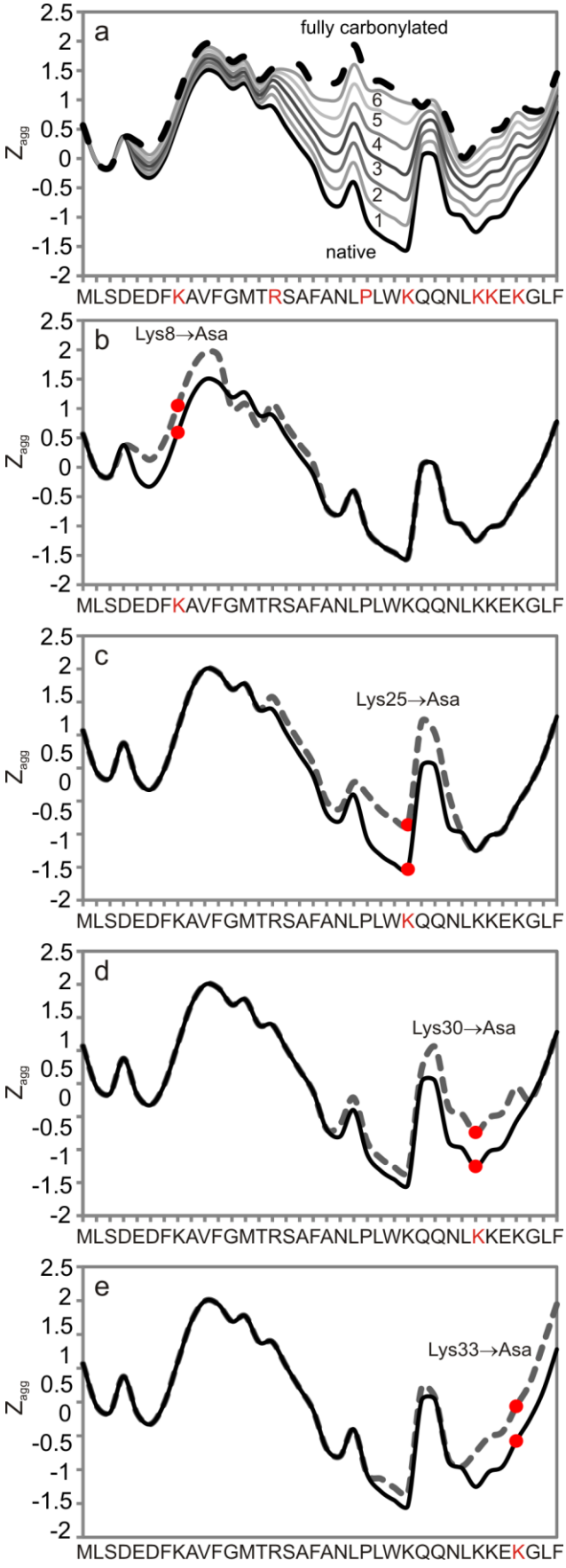

a
fully carbonylated
6
5
4
3
2
1
native
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
b
Lys8→Asa
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
c
Lys25→Asa
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
d
Lys30→Asa
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
e
Lys33→Asa
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
Zagg
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2

Figure S3. Site-specific intrinsic aggregation propensity, Zagg, for: (a) The native (solid black line), fully carbonylated (dashed black line) villin headpiece and averages over all combinations of villin headpiece with 1, 2, 3, 4, 5 and 6 carbonylated residues (thin solid gray lines); comparison between villin headpiece in the native state and with one carbonylated residue: (b) native - solid black line, Lys8 to Asa - dashed gray line; (c) native - solid black line, Lys25 to Asa - dashed gray line; (d) native - solid black line, Lys30 to Asa - dashed gray line; and (e) native - solid black line, Lys33 to Asa - dashed gray line.

Table S1. Comparison of basic physico-chemical properties of Asa, Gsa, Leu and Val. Solvent accessible surface area (SASA) and molecular volume were calculated using Gromacs tools.[5]

Properties	Asa	Leu	Gsa	Val
Chemical formula	C_4H_8O	C_4H_{10}	C_3H_6O	C_3H_8
MHP	1.56	1.67	1.27	1.28
Mw (g mol^{-1})	72.11	58.12	58.08	44.1
SASA (nm^2)	1.9	1.8	1.7	1.6
Volume (nm^3)	0.23	0.21	0.19	0.18
Charge	0	0	0	0

REFERENCES

1. Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
2. Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431-492 (1989).
3. Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529-539 (1994).
4. van der Spoel, D. *et al.* Gromacs User Manual version 3.3. www.gromacs.org 2005.
5. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306-317 (2001).

Chapter IV

Are current atomistic force fields accurate enough to study proteins in crowded environments?

Petrov, D. & Zagrovic, B. Manuscript currently being processed for *Proc. Natl. Acad. Sci. U.S.A.*

DP and BZ conceived and designed the study. DP collected and DP and BZ analyzed the data. DP and BZ wrote the manuscript.

ABSTRACT

High concentration of macromolecules in the crowded cellular interior influences different thermodynamic and kinetic properties of proteins, including structural stabilities, intermolecular binding affinities and enzymatic rates. Moreover, various structural biology methods, such as NMR or different spectroscopies, typically involve samples with relatively high protein concentration. Due to large sampling requirements, however, the accuracy of classical molecular dynamics (MD) simulations in capturing protein behavior at physiologically or experimentally relevant concentrations still remains largely untested. Here, we use MD simulations to study native (aggregation-resistant) and oxidatively damaged (aggregation-prone) forms of villin headpiece at 6 mM and 9.2 mM protein concentration. We perform an exhaustive set of simulations using different force fields, electrostatics treatments and solution ionic strengths. Surprisingly, the two villin headpiece variants exhibit similar aggregation behavior, although their estimated aggregation propensities differ markedly. Importantly, regardless of the simulation protocol applied, native villin headpiece consistently aggregates even under conditions at which it is experimentally known to be fully soluble. We demonstrate that aggregation of native molecules, the same as that of carbonylated molecules, is accompanied by a dramatic decrease in total potential energy, with not only hydrophobic, but also polar residues and backbone contributing substantially. Strikingly, the same effect is observed for six major atomistic force fields examined, suggesting that they artificially promote protein aggregation likely due to the overestimation of protein-protein interactions originally intended to stabilize protein 3D structures. Overall, our results suggest that current MD force fields may strongly affect the picture of protein behavior in biologically relevant crowded environments.

INTRODUCTION

Cell interior is comprised of various macromolecules occupying even up to 40% of the total cytoplasmic volume, with proteins being the most abundant class of molecules.[1,2] Importantly, such densely packed environment can drastically affect thermodynamic and kinetic properties of proteins.[3,4] Moreover, as a consequence of high intracellular concentration, proteins can form cytotoxic aggregates that have been linked with numerous pathologies.[5,6] Additionally, most solution-based biophysical experimental methods, such as NMR or different spectroscopies, involve samples with relatively high protein concentration. Volume-exclusion and confinement effects in the context of crowding have been qualitatively well-understood by statistical mechanical theories and computer simulations.[4,7] What is more, Brownian dynamics and coarse-grained simulations have provided a detailed description of multipart mixtures of biomolecules and have sometimes even matched real systems in their complexity.[8-10] Nevertheless, although highly successful, such approaches still fall short of capturing the fully atomistic, dynamic picture of high-concentration macromolecular systems. To this end, molecular dynamics (MD) simulations, a high-resolution computational biology tool,[11] have recently been employed to model in detail different aspects of crowding when it comes to protein structure, dynamics and interactions as well as solvent behavior.[12-14] Moreover, MD simulations have been applied to explore early events in the formation of protein aggregates, focusing predominantly on short peptides with high aggregation propensity and their association.[15-20] However, arguably due to high computational expenses, control tests to show that non-aggregating polypeptides do not aggregate are rarely performed, with only a few attempts in this direction. For example, Gsponer *et al.* and Tsai et al. have shown that control mutant peptides aggregate into structures with reduced amyloid character as compared to aggregation-prone peptides.[15,18] However, to the best of our knowledge, not a single study so far has provided clear evidence of a known non-aggregating peptide as a negative control.

Here, we use classical MD simulations to study the behavior of the 36-residue villin headpiece mini-protein[21-23] at atomistic resolution in the presence of multiple copies of the same molecule in the simulation box. In contrast to other MD simulation studies of protein-protein interactions or aggregation that mainly focus on highly interacting or aggregation-

prone polypeptides, our choice of the model system was primarily motivated by the fact that villin headpiece does not self-associate or aggregate at low and moderate protein concentrations. For example, this well-studied actin-binding polypeptide remains soluble at protein concentration of 1-2 mM, as shown by infrared spectroscopy.[24,25] Additionally, Fourier transform infrared spectra of the peptide indicate no aggregation at 6 mM when suspended in a 50 mM sodium acetate buffer with a possibility of aggregate or dimer formation only at a higher protein concentration of ~18 mM.[23] Similarly, circular dichroism experiments suggest that there is no significant aggregation of villin headpiece at the concentration of 9.2 mM, although in this case the data was collected at -40 °C in glycerol/water solution.[22] Finally, NMR spectra of villin headpiece can successfully be recorded even at the concentration of 32 mM, although with changes in chemical shifts of surface residues as a consequence of increased protein-protein interactions upon crowding.[26] Here, it should also be mentioned that villin headpiece adopts a completely α-helical fold,[21] a secondary structure element known to prevent aggregation, as opposed to β-sheets that promote it.

Increased levels of protein aggregation have been repeatedly related to protein oxidative damage with highly oxidized proteins being a frequent component of potentially cytotoxic aggregates.[27-29] In a recent study, we have shown that metal-catalyzed carbonylation, arguably one of the most important types of irreversible oxidative modifications, drastically increases intrinsic aggregability of villin headpiece by directly affecting its hydrophobicity, net charge and secondary structure,[30] protein properties shown to strongly influence aggregation propensity.[31] In particular, using the formalism proposed by Chiti *et al.*[31], we have shown that a single carbonylation event can increase the aggregation rate of villin headpiece by more than an order of magnitude even in the absence of unfolding, and have additionally corroborated this finding using Zyggregator[32] predictor. Additionally, simultaneous damage to lysine, arginine and proline residues, i.e., the main targets of protein carbonylation, causes unfolding of the protein with concomitant solvent-exposure of its hydrophobic core and backbone,[30] which is believed to be one of the main mechanisms of protein aggregation.[5,29] Therefore, the carbonylated villin with 7 affected residues is expected to have a significantly elevated propensity to self-associate as compared to the native variant. Accordingly, here we study two different villin headpiece

systems: 1) native, properly folded protein resistant to aggregation, and 2) carbonylated, completely unfolded and aggregation-prone protein. The principal question that we ask is what drives the formation of villin headpiece self-aggregates in MD simulations and, in the process, we explore the limitations of current atomistic force fields in describing biologically and experimentally realistic protein solutions.

RESULTS

We have examined the behavior of native and carbonylated villin headpiece at biologically relevant protein concentrations of 6 mM and 9.2 mM, using MD simulations with 4 or 8 chemically identical villin headpiece copies in the simulation box, respectively, at different sodium chloride concentrations (0 to 0.8 M). Furthermore, to account for potential inaccuracies in the physical description of the system, we have employed two different force fields (GROMOS45a3,[33] GROMOS54a7[34]) to generate MD trajectories and four additional force fields to analyze them (AMBER94,[35] AMBER99SB-ILDN,[36] CHARMM27-CMAP[37] and OPLS-AAL[38]) together with two types of electrostatics treatment (particle mesh Ewald[39] and reaction-field[40]). In total, we have applied 10 different simulation protocols to generate 5.2 μs of explicit solvent MD trajectories. Such extensive sampling is essential given the complexity of the systems comprised of multiple polypeptides at finite protein concentration. In order to characterize thermodynamic properties of native (depicted in black throughout) and carbonylated (depicted in red throughout) systems, in several instances we have calculated averages over the entire simulated time (50 ns, 100 ns or 200 ns depending on the simulation protocol applied) and over all simulation setups (unless stated otherwise) in order to enhance the coverage of the phase space of studied systems or simply for illustration purposes. Note, however, that we exclude data obtained from MD simulations at 9.2 mM protein concentration (only 1 out of 10 simulation setups) from the analysis if the property in question is either extensive or concentration dependent (see Methods for more details).

Structural analysis of native and carbonylated villin

Villin headpiece is a 3-helix bundle protein with a tightly packed hydrophobic core comprised of three phenylalanine residues (Fig. 1A).[21] To structurally characterize the

simulated proteins, we have used: 1) atom-positional root-mean-square deviation (RMSD) from the native villin headpiece structure,[21] 2) the number of residues in α-helical conformation (#α), and 3) the sum of distances between the centers of mass of core phenylalanines (Σd_{PHEs}). In Figure 1B, we present aggregate averages of these three measures over all simulated conditions, with the average values for individual simulation protocols given in Table S1. Overall, native polypeptides remain folded at infinite dilution as previously shown by us[30] and demonstrated here by averages of all three independent measures of foldedness (Figure 1B, single simulations): <RMSD> = 2.7 ± 1.2 Å, <#α> = 22.3 ± 2.3 and $<\Sigma d_{PHEs}>$ = 20.3 ± 2.4 Å. On the other hand, carbonylated molecules at infinite dilution populate the unfolded state,[30] with an increase in the average RMSD (<RMSD> = 8.1 ± 0.8 Å) and the sum of the distances between the solvent-exposed core phenylalanines ($<\Sigma d_{PHEs}>$ = 33.3 ± 5.8 Å) and a significant decrease in the α-helical content (<#α>= 7.0 ± 3.9) (Figure 1B, single simulations). Importantly, both native and carbonylated villin headpiece, simulated with multiple copies in the simulation box, closely resemble the structural properties of the equivalent polypeptides at infinite dilution (Figure 1B, aggregation simulations). Note that initial configurations for aggregation simulations were chosen randomly from the equilibrium ensemble of the last 25 ns (also used for the analysis) of previously simulated 110-ns-long, infinite dilution trajectories,[30] thus ensuring that the two sets of simulations are comparable.

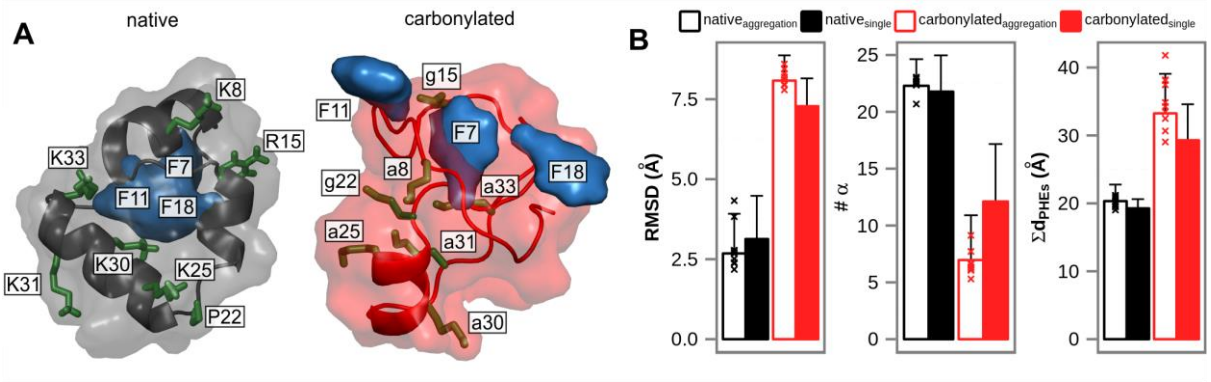


Figure 1. Structural characterization of villin headpiece in the folded native and the unstructured carbonylated state. A) Typical structures of native and carbonylated villin showing the core phenylalanines in blue and carbonylated residues in green; B) three measures of villin foldedness averaged over all simulated trajectories and given with standard deviations (data for infinite dilution taken from ref[30]). The averages were calculated

over the entire simulated time for finite protein-concentration systems, whereas only the last 25 ns of simulated time were used for infinite dilution systems.

Analysis of the aggregation process and villin headpiece aggregates

As predicted, carbonylated villin headpiece molecules, initially placed to maximize their intermolecular distances, start to associate after only a few nanoseconds of free diffusion, leading to an exponential decrease in the number of free monomers in solution, together with a rapid increase in the number of intermolecular atomic contacts at all conditions examined (Figure 2, see Methods for more details). Namely, configurations in which the number of free monomers reaches 0 are observed in 23 out of 26 simulated trajectories with aggregates comprised of at least 75% of all monomers observed in every single trajectory (Figure 2). Surprisingly, in contradiction with experimental findings,[13,14] analysis of the native villin headpiece system shows that the fraction of free monomers decreases and the number of inter-protein atomic contacts increases with simulated time in the same manner as for carbonylated villin headpiece (Figure 2). Altogether, 10 different simulation setups have been examined (Table S1) using different force fields, types of electrostatics treatment and ionic strength, and strikingly, all fail to keep native villin headpiece mono-disperse and reproduce experimental observations.

To further characterize the aggregation process of the studied systems, we have analyzed the kinetics of protein-protein association and dissociation. The average dissociation time of native molecules from aggregates is approximately 45 ns, which is significantly longer than the average time required for their association of approximately 17 ns. Similarly, the average timescales of protein dissociation and association for the carbonylated system are approximately 50 ns and 19 ns, respectively. Importantly, approximately 65% of native and 70% of carbonylated complexes that have formed never dissociate during the simulated time, including all those with life-times longer than 55 ns (SI Figure S1). This fact clearly indicates that the actual life-times of villin headpiece in the aggregated state may be much longer than the average values obtained from simulations, which are limited by sampling. On the other hand, longer simulations would not significantly influence the estimated association times, since only less than 5% of proteins in both systems remain free in solution in the course of simulations. This, in turn, suggests that even much lower protein

concentration would likely not prevent native villin headpiece from aggregating, but only increase the search time needed for proteins to find each other by free diffusion.

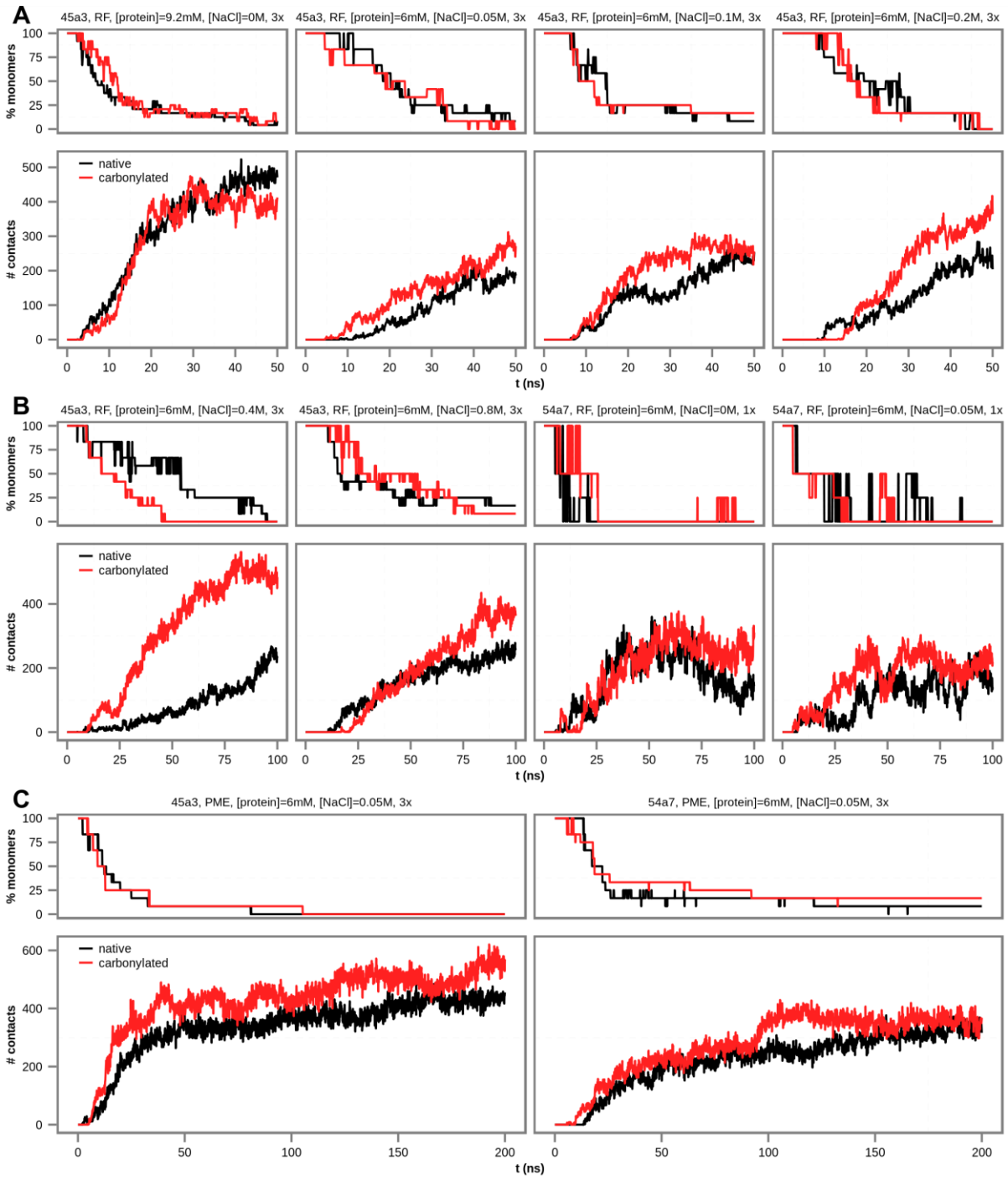


Figure 2. Aggregation of native and carbonylated villin headpiece under various simulation conditions monitored by the percentage of free monomers in solution and the number of intermolecular atomic contacts as a function of time, averaged over independent simulations performed at the same conditions. Altogether, 10 simulations setups were examined and are grouped here according to length: A) 50-ns-long, B) 100-ns-long,

and C) 200-ns-long simulations. Exact simulation setup is given above each plot including number of replicate trajectories used to generate the curves (1x or 3x).

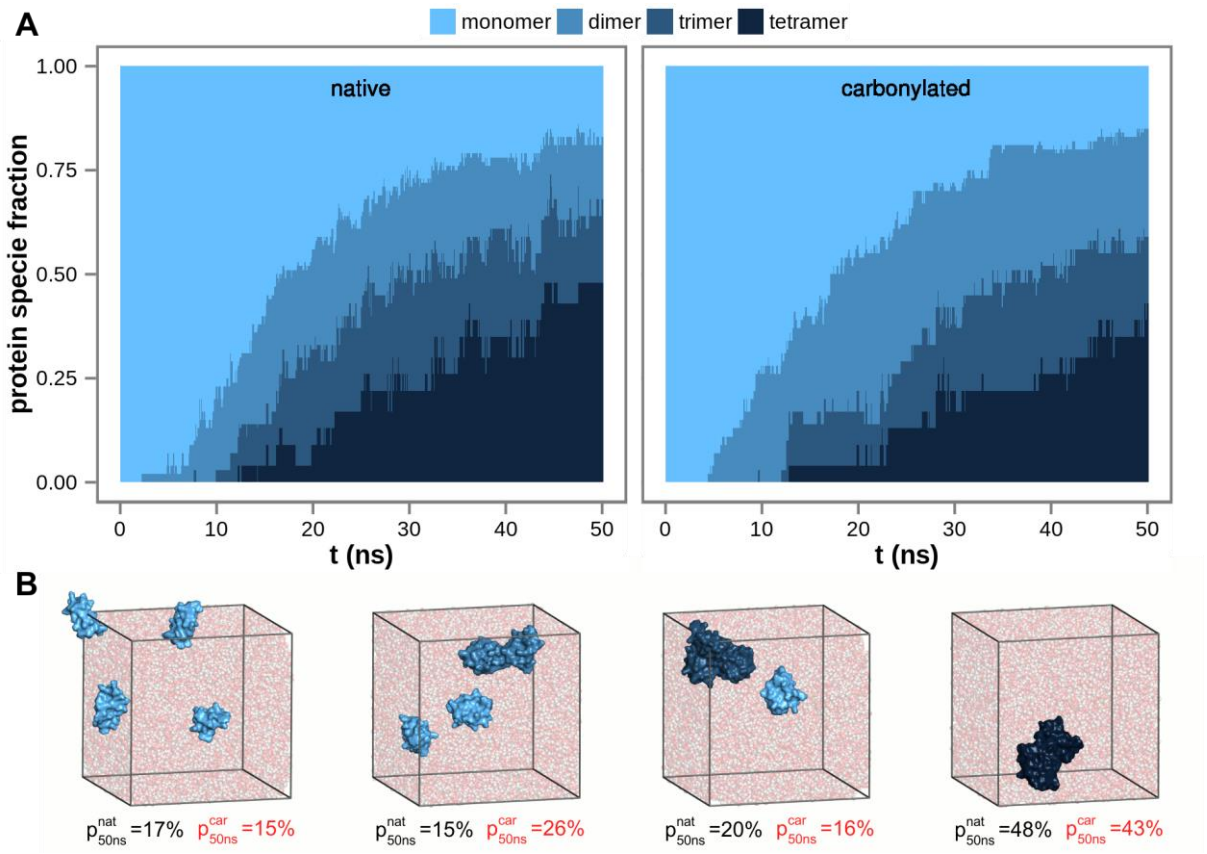


Figure 3. A) Average, aggregate progression of villin headpiece self-association as a function of time calculated over all simulated trajectories. B) Snapshots from simulated trajectories depicting a typical sequence of formation of villin headpiece self-aggregates from free monomers. Fraction of monomers, dimers, trimmers and tetramers in the ensembles at 50 ns are given explicitly.

In both studied systems, starting from free monomers, peptide dimers begin to form first, followed by the formation of trimers and tetramers, leading to the maximum number of tetrameric aggregates after approximately 50 ns of simulated time. Overall, tetrameric aggregates are the most abundant and free monomers the least abundant species in our simulations, representing on average approximately 45% and 15% of the total protein content at 50 ns in native simulations, respectively (Figure 3). Furthermore, in order to identify specific residues involved in villin headpiece self-association, we have calculated the fraction of intermolecular atomic contacts formed by each residue (side chains only) over all native and carbonylated simulations, normalized by solvent-accessibility to account for surface exposure, i.e., the probability to interact. Surprisingly, while association of

hydrophobic residues is believed to be a key element in protein aggregation, our analysis reveals that interactions in both native and carbonylated aggregates are dominated by either backbone or glutamine and asparagine residues (Figure 4). In addition to these hydrophilic moieties, carbonylated aggregates are also characterized by contacts involving hydrophobic ring-containing phenylalanines and tryptophans (Figure 4). Finally, we have also analyzed the average contact maps for both native and carbonylated systems (Figure S2), which demonstrate that the N-terminal residues form fewer intermolecular contacts than the rest of the molecule in both studied systems. Additionally, the region between residues 15 and 25 of the native contact map is depleted in contacts, whereas the carbonylated map displays a more even distribution of contacts. However, no system-specific pattern among the residues exhibiting many intermolecular contacts has been observed.

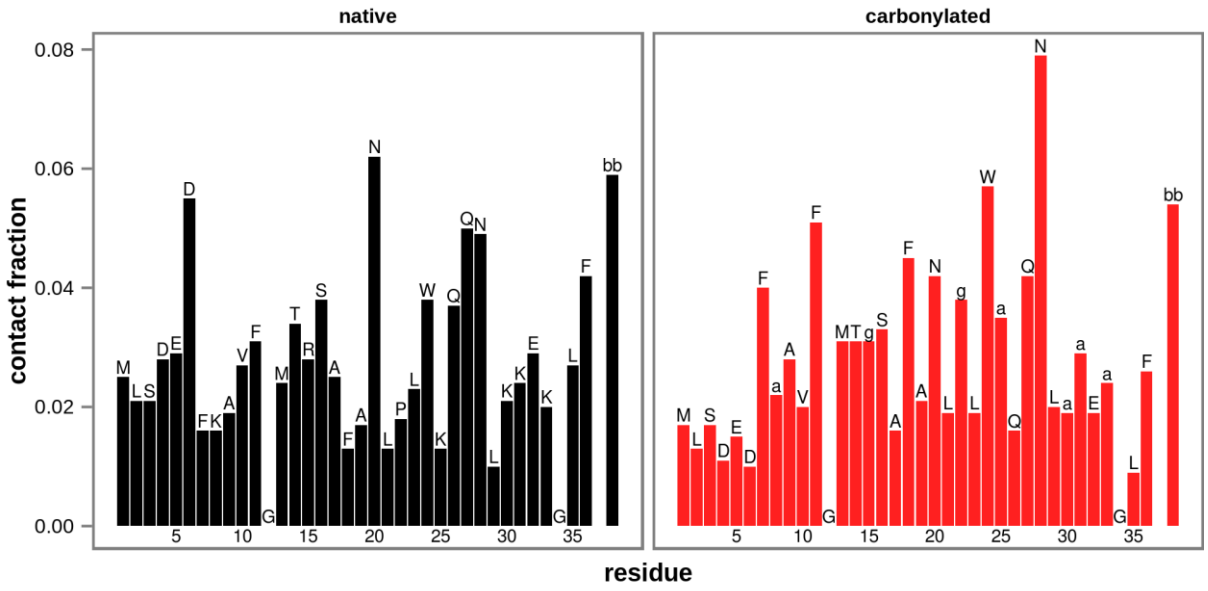


Figure 4. Sequence-wise peptide-peptide interaction propensity estimated by the number of inter-peptide atomic contacts normalized by solvent-accessibility, i.e., surface-exposure per amino acid (side chains only), with the peptide backbone treated as a separate residue, with the values obtained by averaging over all simulated trajectories (*bb* – backbone, *a* – aminoadipic semialdehyde, *g* – glutamic semialdehyde, while natural amino acids are indicated using standard 1-letter code).

What drives villin headpiece aggregation in MD simulations?

To address this, we have explored the role of enthalpic contributions in the aggregation of the studied systems, including solvent-solvent, protein-solvent and protein-protein

interactions, by calculating the average of the difference in potential energy between fully aggregated (tetrameric) and non-aggregated (monomeric) states. Expectedly, solvent-solvent and protein-protein interactions provide favorable contribution, while protein-solvent interactions provide unfavorable contribution to the total enthalpy of aggregation, as seen in the case of simulations with the most extensive sampling (200 ns), the PME electrostatics treatment and 0.05 M salt concentration (Figure 5A). However, the total potential energy of the same systems is significantly lower in the aggregated state (by approximately 150 kJ/mol and 330 kJ/mol for the native and the carbonylated system simulated by GROMOS45a3 force field, respectively, and approximately 100 kJ/mol and 270 for the native and the carbonylated system simulated by GROMOS54a7 force field, respectively, Figure 5A), suggesting that self-association may be an enthalpically driven process. Further analysis of the simulations obtained using RF electrostatics treatment (the GROMOS 45a3 parameter set only) shows that the aggregated states are also favored for every salt concentration examined (Figure 5B).

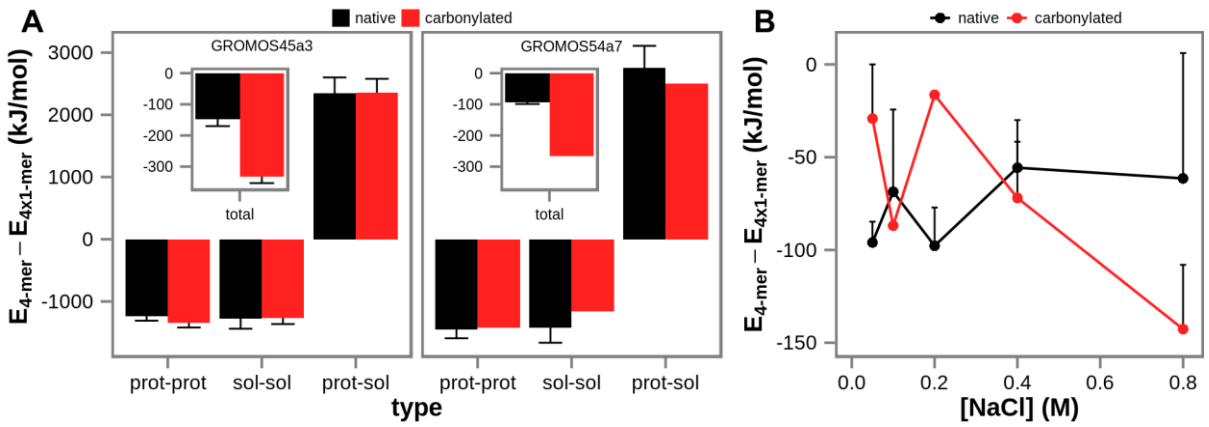


Figure 5. The difference in potential energy between aggregated (tetramers) and non-aggregated (free monomers) states of native and carbonylated villin headpiece calculated directly from simulated data. A) Average difference in total potential energy (inset, the y-axis the same as in the main figure), and contributions from solvent-solvent, protein-protein and protein-solvent interactions over: 1) simulations obtained using the GROMOS 45a3 parameter set, PME electrostatics treatment and 0.05 M salt concentration (left panel), and 2) simulations obtained using the GROMOS 54a7 parameter set, PME electrostatics treatment and 0.05 M salt concentration (right panel). B) Average difference in total potential energy over simulations obtained using GROMOS 45a3 and RF electrostatics treatment at different salt concentrations.

To study other force fields in addition to GROMOS 45a3 and 54a7, we have re-evaluated the potential energy of the simulated trajectories by using 6 widely used MD force fields

(GROMOS45a3,[33] GROMOS54a7,[34] AMBER94,[35] AMBER99SB-ILDN,[36] CHARMM27-CMAP[37] and OPLS-AAL[38]), and two electrostatic treatment methods treatments (RF and PME), on energy-minimized snapshot configurations, where the force field used for energy minimization was also used to re-evaluate the potential energy. Note that we have only re-evaluated potential energies for native systems since parameters for carbonylated residues are only available for the GROMOS force field. Remarkably, regardless of the type of electrostatics treatment, all of the evaluated force fields show the same trend in favoring tetrameric aggregated over non-aggregated conformers of native villin headpiece (Figure 6 and S3), with the extent of such bias ranging from approximately -50 kJ/mol to -240 kJ/mol (PME calculations) for GROMOS54a7 and OPLS-AAL force fields, respectively. Importantly, in order to decrease the uncertainty in estimation, we have averaged the re-evaluated differences in potential energy between the tetrameric and monomeric configurations for each force field over all trajectories no matter which simulation condition they came from (Figure 6A and S3A).

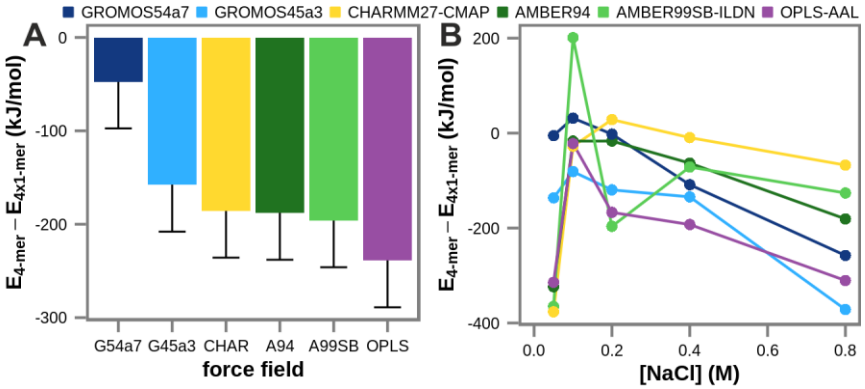


Figure 6. Difference in total potential energy between aggregated (tetramers) and non-aggregated (free monomers) conformations of native villin headpiece calculated from re-evaluated energies using PME electrostatics treatment and 6 widely used MD force fields on energy-minimized configurations taken from simulation. A) averages over all simulations with the estimated standard errors (50 kJ/mol) shown using one-sided error bars. B) averages over simulations at different salt concentrations.

Under the assumption that the entropic contribution to the free energy of association is comparable for all different force fields, this result suggests that utilizing any of these force fields would most likely lead to aggregation of native villin headpiece in MD simulations, and therefore be at odds with experimental observations. The fact that the two GROMOS force fields, which were used to generate configurations, exhibit the lowest preference towards

the aggregated state when compared to other evaluated force fields further supports this claim (Figure 6). Parenthetically, the primary strategy in parameterizing the GROMOS 54a7 force field was to reproduce experimental hydration free energies of amino-acid side-chain analogs, a property closely related to solubility in water, while the other force fields examined here significantly underestimate this property.[41,42]

DISCUSSION

This study explores the behavior of the villin headpiece domain in its native and carbonylated forms at experimentally and biologically relevant protein concentrations in MD simulations. We have shown that native and carbonylated villin headpiece molecules remain folded and unfolded throughout the simulated trajectories, respectively, as previously seen in simulations at infinite dilution (Figure 1). Even though 4 individual native villin molecules unfolded in the course of simulated trajectories, this is to be expected in the ensemble of more than 100 polypeptides simulated herein, since the villin headpiece is only a marginally stable protein.[43] By probing various simulation conditions, we have observed that native and carbonylated molecules aggregate in a very similar fashion, despite the fact that they display markedly different aggregation propensities. This finding is even more remarkable given the fact that we have focused on conditions experimentally known to render villin headpiece soluble. Even though villin headpiece can begin to aggregate at protein concentration of 6 mM after a period of a few days,[23] it is highly unlikely that this is related to the nanosecond-time-scale aggregation observed here. Overall, our choice of the "borderline" concentrations at which villin headpiece remains soluble was based on three principal rationales: 1) we wanted to minimize the computational costs by minimizing the size of the simulation box (even at this protein concentrations each simulated system contains more than 100,000 atoms) and consequently the search time needed for peptides to find each other by free diffusion, 2) we wanted to maximize the likelihood of observing the difference in aggregation between the native, supposedly aggregation-resistant, and the carbonylated, aggregation-prone systems, and 3) 6 mM (90% of simulation protocols) and 9.2 mM (10% of protocols) concentrations of villin headpiece correspond to the protein mass fraction of approximately 3-4%, which is experimentally relevant, but also significantly lower than what one observes in typical crowded cellular environments. Likely, any

aggregation that was observed here would only be potentiated at higher concentrations. However, in order to address the possibility that aggregation of the native system is induced by the "borderline" villin headpiece concentrations used in this study, we have analyzed the kinetics of protein-protein association and dissociation. This analysis has revealed that a large number of protein-protein complexes never dissociate in the course of simulated trajectories, and that the average protein dissociation time is markedly longer that the average association time (Figure S1), suggesting that aggregation in MD simulations would occur even at lower concentrations than applied here.

Furthermore, in order to characterize the nature of villin headpiece intermolecular interactions, we have identified the residues involved in the formation of protein-protein atomic contacts. Surprisingly, this analysis shows that interactions in villin headpiece aggregates are mainly dominated by polar glutamines and asparagines as well as hydrophobic ring-containing phenylalanines and tryptophans (Figure 4). Although the two former hydrophilic amide-containing amino acids are generally considered to be aggregation reducing,[31,44] they have been repeatedly linked to protein aggregation and deposition disorders, most notably in poly-Q diseases.[5,18,45] In addition to this, the high occurrence of peptide backbone atoms among villin headpiece intermolecular contacts (Figure 4) supports the hypothesis that poly-peptide chains have a general tendency to aggregate due to an intrinsic aggregability of the protein backbone.[46] This is further corroborated by recent findings that poly-glycine and poly-alanine chains aggregate readily.[47]

Finally, to investigate what drives villin headpiece aggregation in MD simulations, we have analyzed how the potential energy of the system, and different contribution thereof, change upon aggregation. Expectedly, villin headpiece self-association is accompanied by a decrease in protein-protein and solvent-solvent, and an increase in protein-solvent short-range potential energy. However, the total potential energy drastically decreases upon aggregation, suggesting that the utilized force fields (the GROMOS 45a3 and 54a7 parameter sets) favor aggregated conformers (Figure 5), in contradiction with experimental findings. Strikingly, by re-evaluating the potential energy of the aggregated and non-aggregated conformers using 6 widely-used atomistic force fields (GROMOS45a3,[33]

GROMOS54a7,[34] AMBER94,[35] AMBER99SB-ILDN,[36] CHARMM27-CMAP[37] and OPLS-AAL[38]), we have shown that all of them favor the aggregated states of native villin headpiece (Figure 6). However, one should emphasize that this result has been obtained by averaging over conformers originally obtained by differently defined Hamiltonians and using a limited number of energy minimized configurations, significantly increasing the error in the estimated potential energy differences. These shortcomings notwithstanding, the calculated differences in the total potential energy come with an error of approximately 200 kJ/mol for each simulation and 50 kJ/mol for the aggregate average difference (Figure 6), as estimated using the root-mean-square errors from the potential energies derived directly from simulations, calculated for the same force field as used in the simulation.

Taken together, our results suggest that typical classical MD force fields bias protein aggregation by overestimating protein-protein and solvent-solvent as opposed to protein-solvent interactions. Additionally, the demonstrated imbalance between these three components of total potential energy displayed by the force fields may be partly a consequence of the widely-used force field validation approaches, aimed mostly at reproducing tertiary and secondary structures of well-characterized proteins. Simply put, strengthening of protein-protein and weakening of protein-solvent interactions leads to stabilization of protein structure, one of the principal targets of most parameterization strategies. This, if true, suggests that realistic polypeptides may display more dynamics and unstructuredness than generally observed in MD-simulation studies. Interestingly, a recent study exploring the limitations of MD simulations by employing a state-of-the-art designated supercomputer to perform a 200-μs-long simulation of an intrinsically unfolded protein revealed that the modeled protein appears to be more compact and collapsed than observed by NMR,[48] further supporting this speculation. Importantly, these potential flaws of current force fields may have strong implications when it comes to the accuracy of MD models in describing protein dynamics and interactions in biologically relevant crowded environment. Finally, recent validation studies of force fields show that they improve over time, but are still not able to reproduce all relevant experimental data.[49,50] Such validations and synergy between experimental methods and MD simulations should in the years to come bring improvements in computational models and physical descriptions of biomolecular systems of interest as well as provide a better understanding of biologically

important processes such as protein folding, protein-ligand binding and protein aggregation in the context of experimentally or biologically realistic conditions. We hope that the results presented herein will provide an additional source of motivation in this direction.

METHODS

Molecular dynamics simulations setup

We have used classical MD simulations to study the behavior of villin headpiece with multiple copies of the molecule in the simulation box. Two systems were examined: 1) native villin headpiece (sequence: MLSDEDF**K**AVFGMT**R**SAFANL**P**LW**K**QQNL**KK**E**K**GLF), and 2) its carbonylated form. The seven letters in bold mark the most important carbonylable amino acids (K, R and P) in villin headpiece, which were all modified into their carbonylated versions in the carbonylated form of the molecule. Upon carbonylation, lysine is converted into aminoadipic-semialdehyde, while arginine and proline are converted into glutamic-semialdehyde, for which force field parameters were taken from refs[30,51]. In total, 10 simulation protocols were applied (Table S1), varying in protein (6 mM and 9.2 mM) and salt (0 M, 0.05 M, 0.1 M, 0.2 M, 0.4 M and 0.8 M) concentration, force field (GROMOS 45a3[33] and 54a7[34] parameter sets), electrostatics treatment (RF - reaction-field[40] and PME - particle mesh Ewald[39]), the number of protein molecules in a simulated box (4 and 8), simulation time (50 ns, 100 ns and 200 ns), and the number of independent simulations (1 and 3). All MD simulations were carried out using GROMACS biomolecular simulation package,[52] keeping the system at 300 K and 1 bar using a Berendsen thermostat (τ_T = 0.05 ps) and barostat (τ_p = 1 ps and compressibility = 4.5×10^{-5} bar^{-1}).[53] A cutoff of 1.4 nm and the dielectric constant of 65 were used for RF, and the Fourier spacing of 0.1 nm for PME calculations. Starting from either native or carbonylated free monomers (4 and 8 for 6 mM and 9.2 mM protein concentration, respectively) maximally separated in a simulation box, polypeptides were allowed to diffuse freely and interact with each other and the solvent. The initial configurations were prepared by placing villin headpiece molecules in a cubic simulation box, with the size defined by the protein concentration, i.e., sides of approximately 10.4 nm (6 mM) and 11.3 nm (9.2 mM), for a total number of atoms exceeding 100,000 in all cases. Villin headpiece monomers were randomly selected from the

ensemble of the last 25 ns of five 110-ns-long independent simulated trajectories of both the peptide in the native and carbonylated form from ref[30], only taking into account structures with atom-positional root-mean-square deviation (RMSD) from the native villin headpiece structure[21] in the range of the average ensemble plus or minus one standard deviation. After filling the simulation box with SPC water molecules[54] and sodium chloride at a given concentration, and steepest descent energy minimization (1500 steps), the system was equilibrated by gradually increasing the temperature (from 100 to 300 K) over 100 ps with gradually decreasing position restraints (from 25 000 to 5000 kJ mol^{-1} nm^{-2}) at constant volume and temperature, and finally additionally equilibrated for 20 ps at constant pressure and temperature of 1 bar and 300 K. Atom coordinates and velocities were saved every 50,000 integration steps, i.e., 100 ps.

Potential energy re-evaluation

To examine the performance of 6 widely used MD force fields (GROMOS45a3,[33] GROMOS54a7,[34] AMBER94,[35] AMBER99SB-ILDN,[36] CHARMM27-CMAP[37] and OPLS-AAL[38]) in the context of villin headpiece aggregation, the trajectories of the native villin headpiece simulated using GROMOS 45a3 and 54a7 force fields were re-evaluated as follows: first, utilizing a given force field, each saved configuration was energy minimized by steepest descent in 1500 steps, and second, the total potential energy was calculated for the minimized configurations using the same force field as in the energy minimization step, together with contributions from short-range non-bonded solvent-solvent, protein-protein and protein-solvent interactions evaluated for all atom pairs within a cutoff distance of 1.4 nm. Electrostatic contribution to the potential energy was calculated using RF and PME methods with the same parameters as used in the simulation setups. The water TIP3P[55] model was used to re-evaluate potential energies of AMBER94,[35] AMBER99SB-ILDN,[36] CHARMM27-CMAP[37] and OPLS-AAL[38] force fields.

Analysis of the simulated data and figure preparation

GROMACS,[52] DSSP,[56] PYMOL[57] and ggplot2[58] (from R) tools were used to analyze the collected trajectories and to generate the figures. Backbone atoms were used for rotational-translational fitting and calculation of atom-positional RMSD with respect to the NMR villin

headpiece structure (first model).[21] The number of inter-peptide atomic contacts and villin headpiece association and dissociation kinetics were evaluated by defining atomic contacts to be present any time two atoms from different villin headpiece molecules come within 0.4 nm from each other, and a peptide-peptide complex to be present any time a pair of villin headpiece molecules remain continuously in atomic contact in the course of at least 1 ns. The site-specific intermolecular interaction propensity along the villin headpiece sequence was estimated by calculating the fraction of the number of intermolecular atomic contacts (heavy atoms only) per residue (normalized at each time-point by the solvent-accessibility of a residue in question) in each simulation, and subsequently averaged over all simulated trajectories (either native or carbonylated). The solvent-accessibility, i.e., the surface-exposure of a given residue was calculated as the fraction of the solvent-accessible surface area (SASA) of the residue in the total SASA of the peptide. Note that only amino-acid side chains were used, while the backbone atoms were considered as a separate collective residue. Finally, the differences in potential energy between aggregated and non-aggregated villin headpiece conformers were calculated as the difference between the average potential energy of aggregated (all-tetramer) snapshots and the average potential energy of non-aggregated (free monomers) snapshots from each simulated trajectory, averaged over different simulation sets. In particular, energy differences evaluated directly from simulations were averaged over different simulation conditions, whereas those calculated from re-evaluated energies were averaged over all simulated systems (excluding simulations at 9.2 mM protein concentration for these cases). Only simulated trajectories with more than 20 snapshots of both aggregated and non-aggregated states were taken into account. Note that villin headpiece systems with protein concentration of 9.2 mM (1 out of 10 simulated setups) were excluded from calculation of average properties if the property in question was either extensive or concentration dependent, including evaluations of villin headpiece association/dissociation kinetics, time-dependent formation of higher order villin headpiece complexes (e.g., dimers) and potential energy.

ACKNOWLEDGEMENTS

This work was supported in part by the Austrian Science Fund FWF (START grant Y 514-B11 to BZ, http://www.fwf.ac.at/) and European Research Council (ERC Starting Independent

grant 279408 to BZ, http://erc.europa.eu/). We thank members of the Laboratory of Computational Biophysics at MFPL for useful advice and assistance.

REFERENCES

1. Ellis, R. J. Macromolecular crowding: Obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597-604 (2001).
2. Ellis, R. J. & Minton, A. P. Cell biology - Join the crowd. *Nature* **425**, 27-28 (2003).
3. Zhou, H. X., Rivas, G. N. & Minton, A. P. in *Annual Review of Biophysics* Vol. 37 *Annual Review of Biophysics* 375-397 (2008).
4. Elcock, A. H. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **20**, 196-206 (2010).
5. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10-S17 (2004).
6. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15-22 (2009).
7. Minton, A. P. Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited. *Biophys. J.* **88**, 971-985 (2005).
8. Periole, X., Huber, T., Marrink, S. J. & Sakmar, T. P. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J. Am. Chem. Soc.* **129**, 10126-10132 (2007).
9. Scott, K. A. *et al.* Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16**, 621-630 (2008).
10. McGuffee, S. R. & Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comp. Biol.* **6** (2010).
11. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
12. England, J. L., Lucent, D. & Pande, V. S. A role for confined water in chaperonin function. *J. Am. Chem. Soc.* **130**, 11838-11839 (2008).
13. Feig, M. & Sugita, Y. Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *J Phys Chem B* **116**, 599-605 (2012).
14. Harada, R., Sugita, Y. & Feig, M. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* **134**, 4842-4849 (2012).
15. Gsponer, J., Haberthur, U. & Caflisch, A. The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5154-5159 (2003).
16. Hwang, W., Zhang, S. G., Kamm, R. D. & Karplus, M. Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12916-12921 (2004).
17. Nguyen, H. D. & Hall, C. K. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16180-16185 (2004).
18. Tsai, H. H. *et al.* Energy landscape of amyloidogenic peptide oligomerization by parallel-tempering molecular dynamics simulation: Significant role of Asn ladder. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8174-8179 (2005).
19. Ma, B. & Nussinov, R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Curr. Opin. Chem. Biol.* **10**, 445-452 (2006).
20. Agrawal, N. J. *et al.* Aggregation in protein-based biotherapeutics: Computational studies and tools to identify aggregation-prone regions. *J. Pharm. Sci.* **100**, 5081-5095 (2011).

21. McKnight, C. J., Matsudaira, P. T. & Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **4**, 180-184 (1997).
22. Havlin, R. H. & Tycko, R. Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3284-3289 (2005).
23. Chung, J. K., Thielges, M. C. & Fayer, M. D. Dynamics of the folded and unfolded villin headpiece (HP35) measured with ultrafast 2D IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3578-3583 (2011).
24. Brewer, S. H. *et al.* Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16662-16667 (2005).
25. Bunagan, M. R., Gao, J., Kelly, J. W. & Gai, F. Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis. *J. Am. Chem. Soc.* **131**, 7470-7476 (2009).
26. Harada, R., Tochio, N., Kigawa, T., Sugita, Y. & Feig, M. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.* **135**, 3696-3701 (2013).
27. Levine, R. L. & Stadtman, E. R. Oxidative modification of proteins during aging. *Exp. Gerontol.* **36**, 1495-1502 (2001).
28. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. & Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **52**, 601-623 (2006).
29. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
30. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).
31. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808 (2003).
32. Tartaglia, G. G. *et al.* Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).
33. Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
34. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
35. Cornell, W. D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
36. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
37. Chen, J. H., Im, W. P. & Brooks, C. L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **128**, 3728-3736 (2006).
38. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* **105**, 6474-6487 (2001).
39. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577-8593 (1995).
40. Tironi, I. G., Sperb, R., Smith, P. E. & van Gunsteren, W. F. A generalized reaction field method for molecular-dynamics simulations. *J Chem Phys* **102**, 5451-5459 (1995).
41. Shirts, M. R., Pitera, J. W., Swope, W. C. & Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J Chem Phys* **119**, 5740-5761 (2003).

42. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676 (2004).
43. Godoy-Ruiz, R. *et al.* Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J Phys Chem B* **112**, 5938-5949 (2008).
44. Pawar, A. P. *et al.* Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392 (2005).
45. Nelson, R. *et al.* Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**, 773-778 (2005).
46. Gsponer, J. & Vendruscolo, M. Theoretical approaches to protein aggregation. *Protein Pept. Lett.* **13**, 287-293 (2006).
47. Lorusso, M., Pepe, A., Ibris, N. & Bochicchio, B. Molecular and supramolecular studies on polyglycine and poly-L-proline. *Soft Matter* **7**, 6327-6336 (2011).
48. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134**, 3787-3791 (2012).
49. Best, R. B., Buchete, N.-V. & Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **95**, 4494-4494 (2008).
50. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS ONE* **7** (2012).
51. Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C. & Zagrovic, B. A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comp. Biol.* **9**, e1003154 (2013).
52. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.* **4**, 435-447 (2008).
53. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
54. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration. Reidel, Dordrecht, 1981.
55. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935 (1983).
56. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
57. Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3r1* (2010).
58. Wickham, H. ggplot2: elegant graphics for data analysis. Springer, New York, 2009.

Appendices to Chapter IV

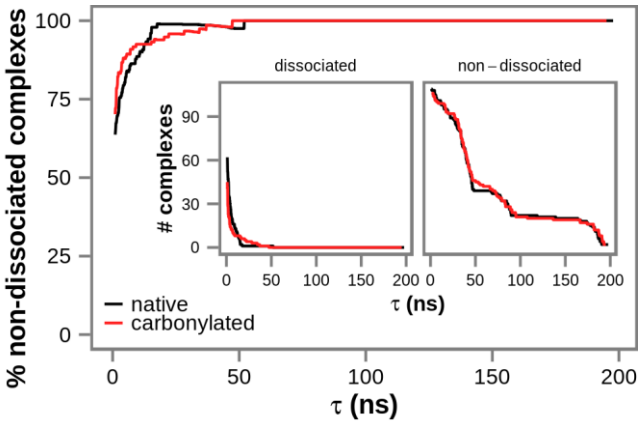


Figure S1. Dissociation kinetics of villin complexes in native and carbonylated forms. Main panel - the percentage of peptide-peptide complexes that never dissociate in the course of simulated trajectories shown as a function of complex life-time. Inset – inverse of the cumulative distribution of the number of villin headpiece intermolecular complexes as a function of complex life-time, i.e., the function that at each value of τ (complex life-time) gives the number of complexes with a longer life-times than the given τ-value, shown for complexes that dissociate (left) and never dissociate (right) over simulated time.

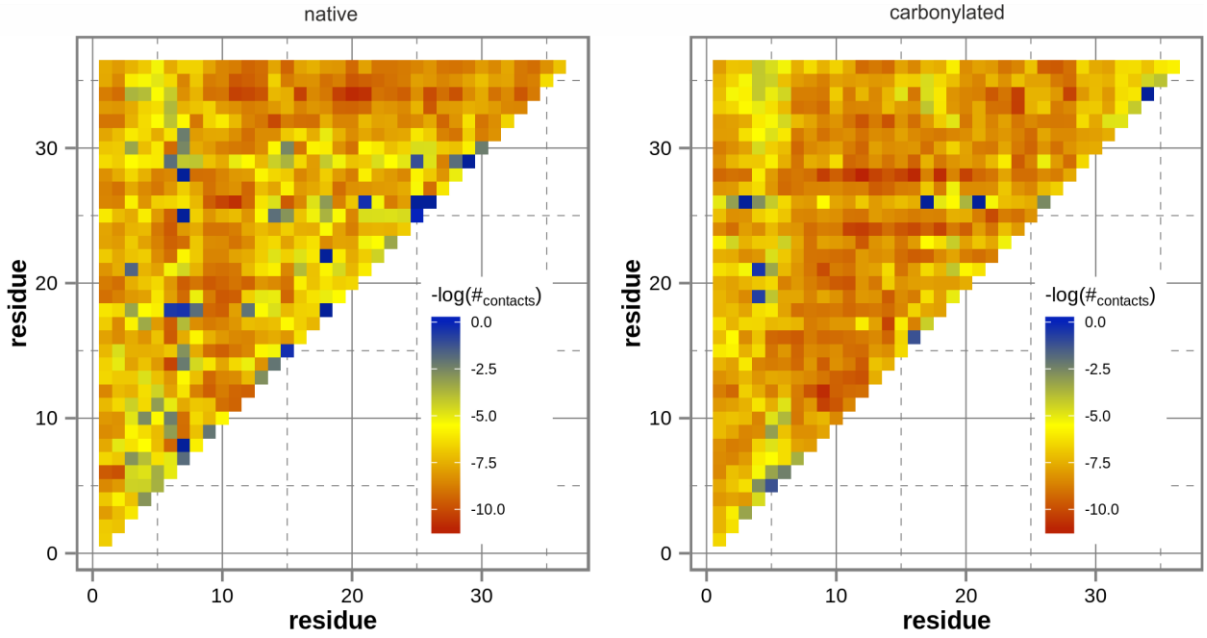


Figure S2. Sequence-wise contact map of intermolecular atomic contacts for native and carbonylated villin headpiece estimated by the number of intermolecular atomic contacts normalized by solvent-accessibility, i.e., surface-exposure per amino acid. The color code of the heat map corresponds to the negative logarithm of the number of contacts for each pair of residues, additionally rescaled in such a way that 0 represents the pair of residues with the smallest number of contacts.

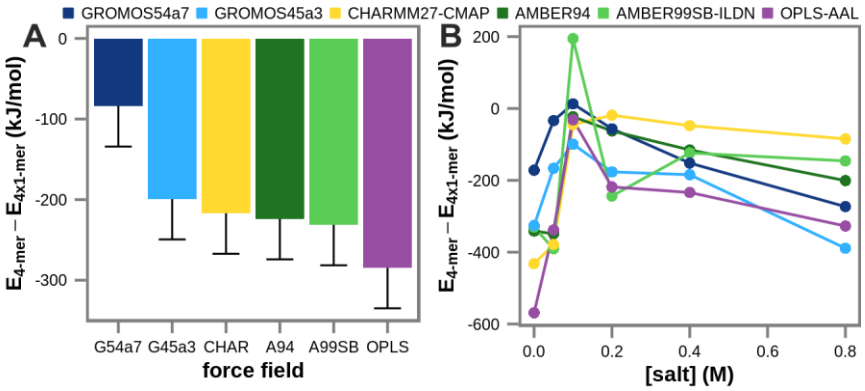


Figure S3. The difference in total potential energy between aggregated (tetramers) and non-aggregated (free monomers) states of native villin headpiece calculated from the re-evaluated energies using RF electrostatics treatment and 6 widely used MD force fields on energy-minimized configurations, where the force field used for re-evaluation of the potential energy was also used for energy-minimization. The averages over all simulations are shown with the estimated standard errors (50 kJ/mol) depicted by one-sided error bars.

TABLE S1. Simulation conditions and setups together with the averages and standard deviations of three measures of foldedness: atom-positional root-mean-square deviation (RMSD) from the native villin headpiece structure,[1] number of residues in α-helical conformation (# α) and the sum of distances between the centers of mass of core phenylalanines (ΣdPHEs). The last two rows highlighted in gray are the aggregate averages over all 10 simulated systems at finite protein concentration and the averages over the infinitely diluted systems (data taken from ref[2]), together with their respective standard deviations. The averages for the finite protein-concentration systems were calculated over the entire simulated time, whereas for the infinitively diluted systems the last 25 ns of the simulated trajectories were used. Each simulation setup was used to model both native and carbonylated villin and accordingly, the averages and standard deviations are shown for native (left) and carbonylated (right) systems.

REFERENCES

1. McKnight, C. J., Matsudaira, P. T. & Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **4**, 180-184 (1997).
2. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).

#	[protein] (mM)	N peptides	[salt] (M)	force field	electrostatics	t (ns)	N simulations	RMSD (A)	# alpha	dPHE
1	9.2	8	0	GROMOS45a3	RF	50	3	2.7 ± 1.1 | 8.2 ± 0.8	22.6 ± 1.9 | 6.3 ± 3.8	19.9 ± 1.7 | 34.9 ± 2.3
2	6	4	0.05	GROMOS45a3	RF	50	3	2.4 ± 0.8 | 8.5 ± 0.7	22.9 ± 1.5 | 5.3 ± 2.0	20.3 ± 1.7 | 38.2 ± 3.4
3	6	4	0.1	GROMOS45a3	RF	50	3	2.8 ± 1.8 | 8.4 ± 0.6	22.6 ± 1.9 | 6.2 ± 3.2	19.5 ± 1.0 | 37.5 ± 3.6
4	6	4	0.2	GROMOS45a3	RF	50	3	2.2 ± 0.5 | 8.1 ± 0.6	23.0 ± 1.4 | 6.6 ± 3.0	20.7 ± 1.6 | 33.5 ± 6.0
5	6	4	0.4	GROMOS45a3	RF	100	3	2.5 ± 0.8 | 8.3 ± 0.6	22.7 ± 1.8 | 6.7 ± 2.7	19.5 ± 1.9 | 38.1 ± 6.0
6	6	4	0.8	GROMOS45a3	RF	100	3	2.7 ± 1.3 | 8.1 ± 0.6	22.4 ± 2.0 | 6.1 ± 3.0	19.5 ± 1.3 | 32.5 ± 4.5
7	6	4	0	GROMOS54a7	RF	100	1	3.8 ± 3.0 | 8.6 ± 1.0	22.6 ± 2.1 | 7.7 ± 2.3	19.9 ± 1.2 | 41.8 ± 10.2
8	6	4	0.05	GROMOS54a7	RF	100	1	4.3 ± 2.5 | 8.0 ± 0.6	22.4 ± 2.4 | 6.6 ± 4.6	19.0 ± 1.0 | 34.3 ± 3.4
9	6	4	0.05	GROMOS45a3	PME	200	3	2.8 ± 0.8 | 7.9 ± 0.8	20.7 ± 3.2 | 6.4 ± 3.7	20.9 ± 3.2 | 29.0 ± 3.7
10	6	4	0.05	GROMOS54a7	PME	200	3	2.4 ± 0.5 | 7.8 ± 0.9	23.1 ± 1.5 | 9.1 ± 5.0	21.2 ± 2.9 | 30.7 ± 4.0
aggregate average								2.7 ± 1.2 | 8.1 ± 0.8	22.3 ± 2.3 | 7.0 ± 3.9	20.3 ± 2.4 | 33.3 ± 5.8
Ref[2]	0	0	0	GROMOS45a3	RF	110	5	3.1 ± 1.4 | 7.3 ± 0.9	21.8 ± 3.2 | 12.1 ± 5.1	19.3 ± 1.4 | 29.3 ± 5.3

Chapter V

Effect of oxidative damage on the stability and dimerization of superoxide dismutase 1

Petrov, D., Daura, X. & Zagrovic, B. Further results (in preparation to be submitted to *PLoS Comput. Biol.*)

DP and BZ conceived and designed the study. DP collected and DP, XD and BZ analyzed the data. DP, XD and BZ wrote the manuscript.

ABSTRACT

Superoxide dismutase 1 (SOD1), a homo-dimeric metalloenzyme, is one of the key antioxidative agents in human cells. It has been linked to amyotrophic lateral sclerosis (ALS), a devastating, late-onset neurodegenerative disorder, with more than 150 ALS-related mutations identified in the SOD1 gene. Additionally, oxidatively damaged and aggregated species of the enzyme have been reported in ALS patients and animal models. However, the molecular mechanisms and possible implications of oxidative stress in the process of SOD1-mediated ALS development still remain elusive. Here, we use classical molecular dynamics (MD) simulations together with thermodynamic integration to study the effects of different oxidative modifications on SOD1 folding and dimerization free energies. In particular, we focus on lysine, arginine and proline carbonylation and cysteine oxidation, two of the most prominent modification types induced by oxidative stress, and examine residues at the homo-dimer interface. Our results show that oxidative damage to SOD1 can drastically destabilize both its homo-dimer and monomer structures, supporting the hypothesis that age-related increase in oxidative stress may be the main trigger of ALS, with the mutations in SOD1 gene being only an additional factor in the development of the disease. Finally, we propose and discuss further objectives of this ongoing project, including examination of ALS-related mutations in SOD1, exploration of binding free energies of metal ions bound to SOD1 and optimization of free energy calculation methods based on MD simulations.

INTRODUCTION

Reactive oxygen species (ROS) participate in a large number of different chemical reactions with proteins leading to modified amino-acid side chains and backbone or even cross-linked and fragmented proteins.[1] Importantly, such oxidative modifications have been associated with aging and age-related disorders such as neurodegenerative diseases, cancer or diabetes.[2,3] Additionally, highly oxidized proteins have been found in potentially cytotoxic protein aggregates and amyloid fibrils.[4,5] Furthermore, recent evidence has indicated that carbonylation, arguably the most studied irreversible oxidative modification, increases protein aggregation propensity and can therefore trigger the formation of insoluble protein inclusions.[6] However, a direct causal relationship between oxidative stress on the one hand and protein aggregation, aging and development of late onset diseases on the other has not been established.[7] An important and widely studied system for exploring this relationship has been Cu/Zn superoxide dismutase (SOD1).[8,9] In particular, this enzyme has been found to associate with 20% of cases of familial amyotrophic lateral sclerosis (fALS), an age-related neurodegenerative disease, with more than 150 different mutations in the SOD1 gene having been linked with fALS.[8,10,11] Moreover, the presence of SOD1 in protein inclusions in motor neurons and astrocytes of fALS patients[12-14] and animal models[15,16] has been observed and documented in detail. Interestingly, SOD1, which breaks down free superoxide radicals as one of the key antioxidant enzymes in human cells, is by the nature of its function as well as its cellular localization exposed to higher levels of oxidative stress than most other proteins. However, the molecular cause and potential implications of oxidative stress in the development of fALS still remain elusive.

Under native conditions SOD1 forms a stable homo-dimer, which is thought to prevent its aggregation and subsequent cytotoxicity,[17] and binds zinc and copper ions, which are important for its dismutase activity as well as its tertiary and quaternary structure formation and stability.[18-20] The two main mechanisms that have been proposed to explain the toxicity of SOD1 mutants are: 1) reduced dismutase activity or increased peroxidase activity leading to an overall increase in oxidative damage of cellular proteins, and 2) formation of insoluble aggregates through a decrease in the stability of SOD1 monomers and/or SOD1 dimer destabilization. Concerning the latter, it has been suggested that it is possibly due to a

decrease in binding affinities of copper and zinc ions with a concomitant increase in their cellular levels, which by itself is thought to be neurotoxic.[8,18]

Here, we use molecular dynamics (MD) simulations[21,22] to investigate whether and under what circumstances direct oxidative damage of the SOD1 enzyme could trigger cytotoxicity. In particular, we explore how different oxidative modifications at the homo-dimer interface affect dimer stability, and further ask how these modifications modulate the stability of free monomers in solution. Importantly, while significant efforts have been directed at experimentally characterizing the thermodynamic properties of SOD1 wild-type and fALS mutants,[23,24] a systematic investigation of oxidatively modified SOD1, i.e., of oxidation "mutants", has so far not been performed. With the continued advance of computer power, this problem has become tractable by different theoretical and computational approaches. On the one hand, various efficient semi-empirical methods, utilizing force-field- and knowledge-based scoring functions to predict protein stability upon mutation, have been developed.[25-27] Although validation against experimental data has shown that such methods in general reproduce correct trends, they, however, often fail in providing a precise quantitative measure of stability.[28] On the other hand, two recent studies have employed perturbation techniques in combination with classical MD, a rigorous, physically based and arguably more accurate approach, to estimate changes in folding free energies upon mutation. In particular, Seeliger and de Groot[29] have successfully calculated thermodynamic stability differences for 109 mutations in the microbial ribonuclease barnase, achieving a remarkable accuracy, with a Pearson correlation coefficient of R = 0.86 against experimental data and an average absolute error of only 3.31 kJ/mol, using non-equilibrium fast-growth thermodynamic integration simulations. Furthermore, Lin *et al.*[30] have performed one-step perturbation calculations to explore the effects of four different side-chain substitutions in a hepta-β-peptide on the folding equilibrium, obtaining results in agreement with experimentally available NMR and circular dichroism data.

Here, in light of these computer-simulation studies, we use the thermodynamic integration (TI) method[31], one of the most widely used and most thoroughly tested thermodynamic techniques available,[32-34] in combination with equilibrium MD simulations to estimate changes in free energy of SOD1 homo-dimer formation and its monomer folding. While

computationally extremely expensive, TI is in principle capable of yielding free energy differences at the limit of accuracy of the force field used.[35-37] We test convergence of the free-energy calculations using block averaging. Finally, we discuss further plans and directions of this study.

METHODS

Molecular dynamics simulations and free energy calculations

We have used the thermodynamic integration approach[31] to calculate the impact of different types of oxidative modifications on the stability of SOD1 dimer and monomer. Alchemical modifications from native residues of interest to their oxidized forms, in the context of the folded SOD1 homo-dimer and monomer or its unfolded state (modeled by a GGXGG pentapeptide, where X stands for the affected residue), were performed by smoothly modifying the force field parameters from those defining the initial to those defining the end state. These processes were coupled to a parameter λ, ranging from λ=0 to λ=1, with the end points representing the native and the modified residue, respectively. Starting from a fully stretched pentapeptide or a 3D structure of SOD1 (PDB code 3KH4[38], using the chains A and B for dimer and the chain A for monomer simulations), each system was solvated in a cubic box filled with explicit SPC[39] water molecules, energy minimized and subsequently equilibrated in three independent copies by gradually increasing the temperature (from 100 to 300 K) over 100 ps and decreasing position restraints (from 25,000 to 5,000 kJ mol^{-1} nm^{-2}) at constant volume and temperature. Additional equilibration for 20 ps at constant pressure (1 bar) and temperature (300 K) was performed subsequently. Starting from each equilibrated system, three independent, 500-ps-long MD simulations were run at each of 21 equally spaced λ-points, with 2 additional λ-points near both ends of the λ range, for a total of 112.5 ns simulation time per system. The change in free energy of an alchemical modification was then calculated as the integral of the ensemble average of the derivative of the system Hamiltonian with respect to λ, between the boundaries λ=0 and λ=1. The integrals were evaluated by the generalized Simpson's rule for non-equidistant nodes using averages over the nine independent simulations at each λ-point, including only the last 150 ps of each 500-ns-long simulation. The change upon

oxidative modification in the stability of SOD1 homo-dimer ($\Delta\Delta G^{nat\rightarrow oxi}_{mono\rightarrow dim}$) and monomer ($\Delta\Delta G^{nat\rightarrow oxi}_{unf\rightarrow mono}$) were calculated according to the thermodynamic cycle in Figure 1 as follows:

$$\Delta\Delta G^{nat\rightarrow oxi}_{mono\rightarrow dim} = \Delta G^{oxi}_{mono\rightarrow dim} - \Delta G^{nat}_{mono\rightarrow dim} = \Delta G^{nat\rightarrow oxi}_{dim} - 2\Delta G^{nat\rightarrow oxi}_{mono} \quad (1)$$

and

$$\Delta\Delta G^{nat\rightarrow oxi}_{unf\rightarrow mono} = \Delta G^{oxi}_{unf\rightarrow mono} - \Delta G^{nat}_{unf\rightarrow mono} = \Delta G^{nat\rightarrow oxi}_{mono} - \Delta G^{nat\rightarrow oxi}_{unf}. \quad (2)$$

Here, $\Delta G^{oxi}_{mono\rightarrow dim}$ and $\Delta G^{nat}_{mono\rightarrow dim}$ are free energies of dimer formation of the oxidatively modified and native SOD1 homo-dimer, respectively, while $\Delta G^{oxi}_{unf\rightarrow mono}$ and $\Delta G^{nat}_{unf\rightarrow mono}$ are the folding free energies of the oxidatively modified and native SOD1 monomer, respectively. Finally, $\Delta G^{nat\rightarrow oxi}_{dim}$, $\Delta G^{nat\rightarrow oxi}_{mono}$ and $\Delta G^{nat\rightarrow oxi}_{unf}$ are free energy changes upon alchemical modification in the context of the folded SOD1 homo-dimer and monomer, and its unfolded state, respectively. Statistical errors were estimated using block averaging and standard propagation of error.[40] MD simulations were run using the GROMACS 4.0.7 biomolecular simulation package,[41] with the GROMOS force field 54A7 parameter set,[42,43] a reaction field electrostatic scheme with a cutoff r_c = 1.4 nm and dielectric constant ε_{rf} = 65 and Berendsen thermostat and barostat keeping the temperature at 300 K (τ_T = 0.05 ps) and pressure at 1 bar (τ_p = 1 ps and compressibility = 4.5 × 10^{-5} bar^{-1}).[44] A soft-core formalism[45] was used to avoid singularities in the potential energy function when removing the non-bonded interactions of atoms, with parameters *sc-α* = 0.7 and *sc-power* = 1, except for threonine carbonylation for which *sc-α* = 1.51 and *sc-power* = 2 were used.

Using the above free energy calculation protocol, the effect of oxidative modifications of nine residues at the homo-dimer interface (THR2, LYS3, CYS6, LYS9, THR54, PRO62, CYS111, ARG115 and THR116) was explored. These particular residues were chosen because they were expected to affect the homo-dimer stability more strongly than others. Namely, out of 46 interfacial residues (defined to be all residues in an SOD1 monomer that are within 8 Å from any atom in the second monomer in an SOD1 dimer), we focused on carbonylation of threonines, lysines, arginines and prolines, and oxidation of cysteines to cysteic acids, two of the most frequent types of oxidative modification found in nature. Furthermore, these

modifications lead to a major change in local hydrophobicity as explained in more detail in Chapter I and ref[46]. Parameters of the modified residues were taken from Chapter I.

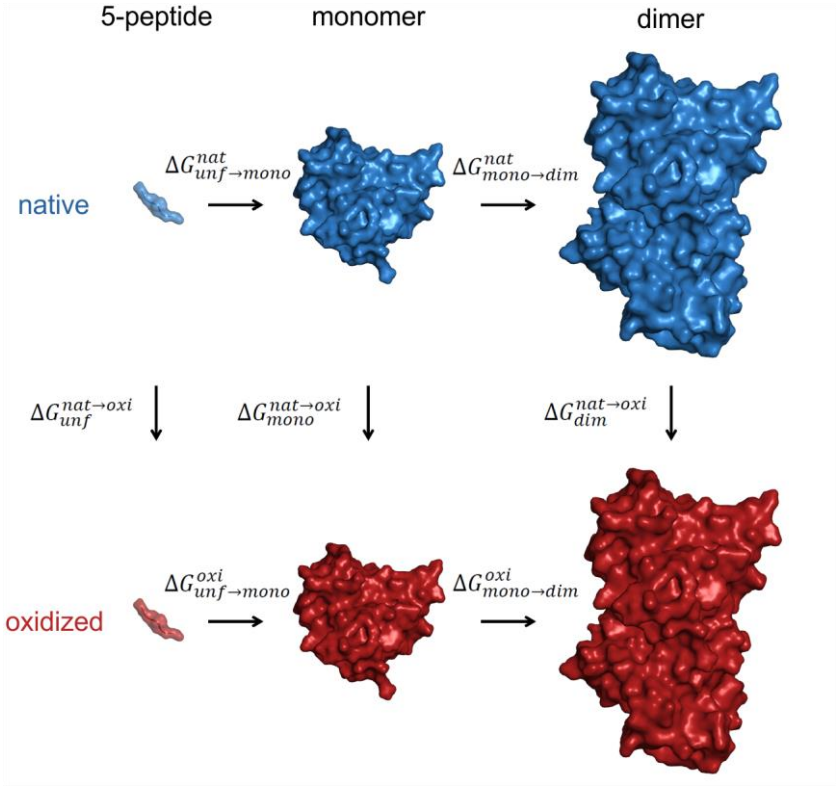


Figure 1. Thermodynamic cycle. Since free energy is a state function, sum of free energy changes between states that form a thermodynamic cycle add up to zero. This property allows one to assess the relative free energy difference between the native and oxidized SOD1 dimer/monomer stability, a quantity rather difficult to calculate directly, by evaluating changes in the free energy of an alchemical modification from a native residue of interest to its oxidized form ($\Delta G^{nat\rightarrow oxi}_{dim}$ for an alchemical modification in the dimer, $\Delta G^{nat\rightarrow oxi}_{mono}$ for an alchemical modification in the monomer and $\Delta G^{nat\rightarrow oxi}_{unf}$ for an alchemical modification in the pentapeptide, i.e., unfolded state), according to the above thermodynamic cycle and Equations 1 and 2.

RESULTS AND DISCUSSION

How do oxidative modifications affect SOD1 homo-dimer and monomer stability?

To address this question, we have applied advanced free energy calculations based on MD simulations, focusing on the following residues at the SOD1 homo-dimer interface: THR2, LYS3, CYS6, LYS9, THR54, PRO62, CYS111, ARG115 and THR116. In particular, we have used the TI method[31] to calculate free energy changes upon alchemical switching from native to oxidatively modified residues, an approach recently applied to successfully estimate the

effect of point mutations on protein stability.[29] However, in contrast to the aforementioned study in which fast, non-equilibrium free energy calculations were performed, we have employed a widely used equilibrium simulation approach, using thermodynamically reversible paths for the calculations. One of the reasons for opting for this more traditional approach has been the recently observed and studied convergence problems of non-equilibrium free energy calculations.[47] Before discussing the specific effects of individual oxidative modifications studied here, we would first like to discuss the important issues of sampling and convergence in our simulations.

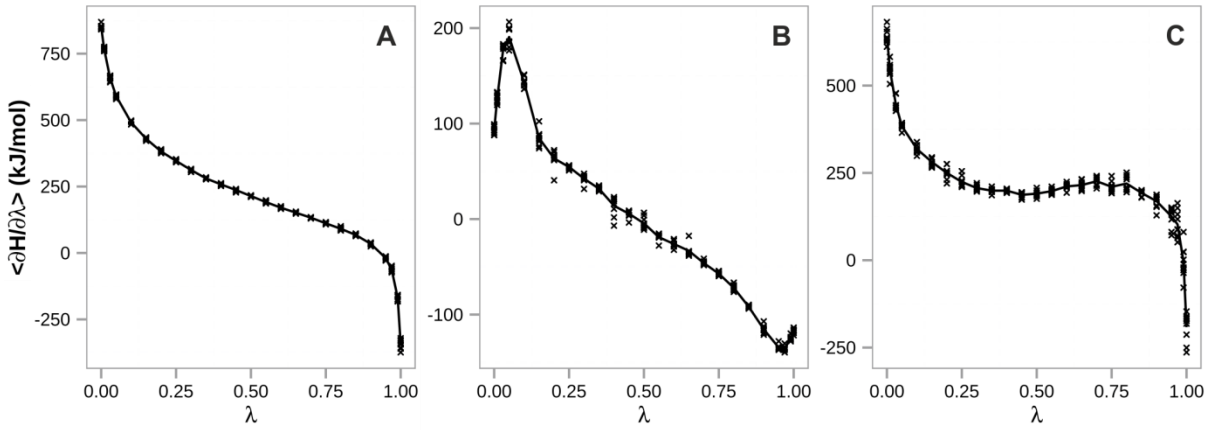


Figure 2. Typical <∂H/∂λ> curves. Ensemble average derivative of the system Hamiltonian with respect to λ shown as a function of λ: A – LYS9, B – THR54 and C – ARG115. Multiple points at given λ come from independent simulations, with the average curves shown in solid lines.

In general, insufficient sampling is one of the major limitations of the MD method, especially in free energy calculations applied to complex systems with rough free energy landscapes. Thus, we have used block averaging (see Methods for details) to estimate the errors of the ensemble average derivative of the system Hamiltonian with respect to λ (<∂H/∂λ>). The analysis of the distribution of errors at individual λ–points shows that approximately 90% of errors are smaller than 10 kJ/mol, with the average over all SOD1 dimer and monomer simulations of 5.2 kJ/mol, which is comparable to 2RT ≈ 5 kJ/mol at room temperature. Additionally, smooth <∂H/∂λ> versus λ curves (Figure 2) allow for an adequate numerical estimation of the calculated integrals, i.e., free energy differences, with all errors of the calculated integrals smaller than 4 kJ/mol. Importantly, this suggests that our free energy

calculations are reasonably converged, and this is likely due to the extensive sampling performed (a total of 112.5 ns for each alchemical perturbation).

Small error bars of the calculated free energy changes, i.e., indicators of convergence, must be distinguished from accuracy of the obtained results. In addition to convergence, the latter depends primarily on the degree of systematic error of the employed method. In particular, alchemical modifications involving net charge perturbation are a sizeable source of systematic error in free energy calculations, introducing a significant reduction in accuracy. However, since the change in stability is defined as the difference between free energy changes in two separate legs of the thermodynamic cycle (Figure 1), the inclusion of the same alchemical perturbation and presumably the same systematic error would cancel out by subtraction. As reported by Seeliger and de Groot[29], this would then provide acceptable accuracy with more than 50% of their calculated values within 1 kcal/mol (< 4.18 kJ/mol) of the experimental data. Importantly, corrections to free energies of charging can be applied *ex post* to obtain more accurate and reliable results.[48,49]

Notably, more than 20% of the errors from simulations probing oxidative modification of CYS6 residue are greater than 20 kJ/mol, with the average of 14 kJ/mol, clearly showing that the convergence has not been reached for this system. Additional inspection of the simulated trajectories revealed that the affected residue (CYS6) flips upon modification, concomitantly causing partial unfolding of the β-strand formed by the native residue (Figure 3). For this reason, we have excluded CYS6 from further analysis.

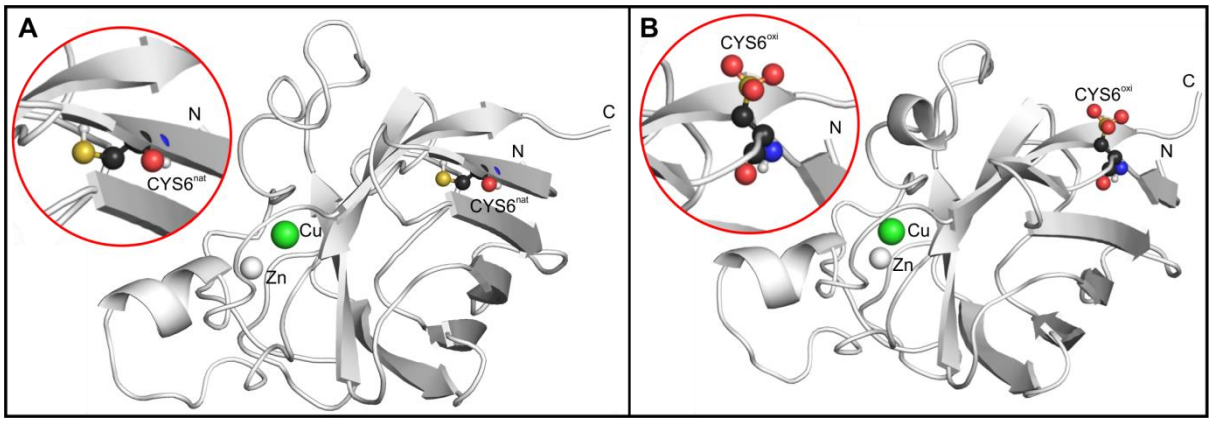


Figure 3. Local unfolding as a consequence of oxidative damage of CYS6. While the affected residue is buried in its native form (A), it flips to protein surface and becomes solvent exposed upon oxidative modification, additionally destabilizing local β-sheet structure (B). Inset: a close-up picture of the affected residue.

How do oxidative modifications affect homo-dimer and monomer stability of SOD1? The stability of the SOD1 homo-dimer is markedly decreased by carbonylation of LYS9 residue ($\Delta\Delta G^{nat\rightarrow oxi}_{mono\rightarrow dim}$ = 23.8 ± 1.1 kJ/mol, Figure 4).

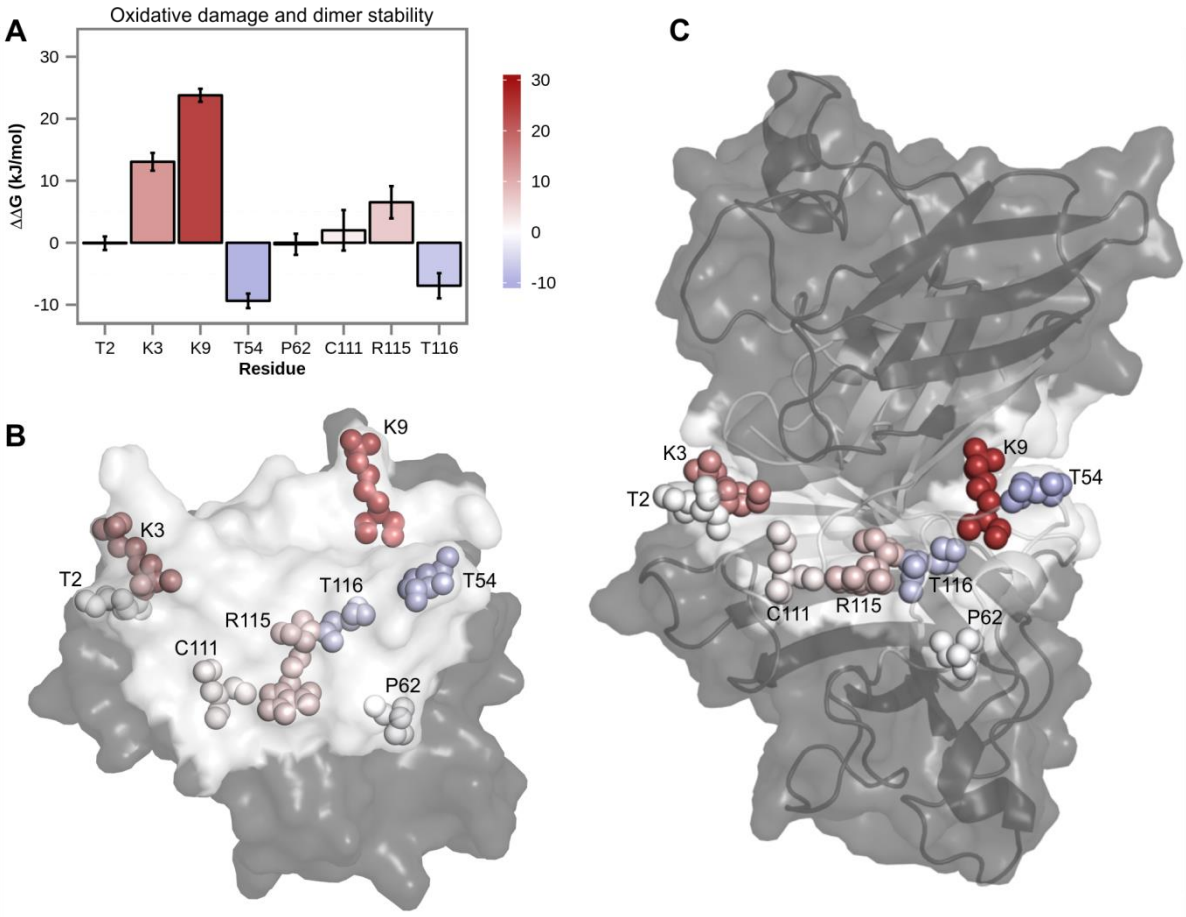


Figure 4. Impact on the SOD1 dimer stability on oxidative damage of residues at the homo-dimer interface. A) changes in free energy of the homo-dimer formation with the error bars calculated by block averaging and standard propagation of error. Location of the studied interface residues and effects of their oxidative modifications on the SOD1 dimer in the context of: B) one of the monomers (view at the interface) and C) SOD1 dimer. *Color code for free energy changes:* blue – stabilization, white - no effect and red – destabilization; color code for *protein structure:* white – interface and grey – rest of the protein.

This is striking when compared to the stability of the dimer itself, which was estimated to be approximately -50 kJ/mol,[23] and to the experimentally measured destabilization effects of some ALS-causing mutations with $\Delta\Delta G^{nat\rightarrow mut}_{mono\rightarrow dim}$ smaller than 5 kJ/mol.[23] Furthermore, oxidatively damaged LYS3 and ARG115 residues exhibit notable destabilization as well, although smaller in extent when compared to LYS9 (13.1 ± 1.4 kJ/mol and 6.5 ± 2.6 kJ/mol, respectively, Figure 4). On the other hand, threonine carbonylation events stabilize the

homo-dimer by -9.4 ± 1.2 kJ/mol (THR54) and -6.9 ± 2.0 kJ/mol (THR116). It must be noted, however, that threonines are significantly less prone to carbonylation compared to other carbonylable amino acids. Lastly, oxidative modifications of THR2, PRO62 and CYS111 show no effect, with the changes in the free energy of the dimer formation smaller than the calculated errors (Figure 4).

When it comes to the folding free energy of the SOD1 monomer, oxidative damage of ARG115, CYS111, LYS9 and LYS3 residues significantly affect monomer stability, decreasing it by 31.0 ± 1.3 kJ/mol, 28.7 ± 1.5 kJ/mol, 21.9 ± 0.3 kJ/mol and 17.3 ± 0.5 kJ/mol, respectively (Figure 5). Although such drastic destabilization effects correspond to an increase in the ratio between the unfolded and folded states of SOD1 by about four orders of magnitude, it is highly unlikely that these oxidative modifications are able to cause complete unfolding of the monomer, given that SOD1 is a hyperstable protein.[50] However, they might induce partial local unfolding, as observed for CYS6 (Figure 3), potentially leading to the formation of insoluble aggregates and consequently cytotoxicity, a mechanism already identified for SOD1 and other well-structured polypeptides involved in protein deposition disorders.[51] In contrast, carbonylation modifications of the remainder of the studied residues only marginally alter monomer stability, with all of them having destabilizing effects (Figure 5).

Taken all together, the results show that a majority of oxidative modifications, regardless of the type of modification and nature of the targeted residue, destabilize both the SOD1 homo-dimer and monomer, with a greater absolute impact on monomer stability. This, in turn, supports the idea that age-related increase in oxidative stress might lead to cytotoxicity through impaired dimerization and increased aggregation propensity of SOD1, consequently leading to ALS. Moreover, this suggests, as speculated before,[17] that increased level of oxidative damage of SOD1, triggered by age, may be a key element in ALS development, with fALS-linked SOD1 mutations being only an additional factor making the protein more susceptible to gain cytotoxic properties and increasing the probability of an early onset of the disease. Importantly, different lines of evidence support these speculations: 1) fALS is a late-onset disease, suggesting that other cause(s) in addition to the reported SOD1 mutations may be required for disease development, 2) as repeatedly shown, SOD1 is involved in sporadic ALS as well,[52] 3) increased oxidative stress induces

SOD1 aggregation both *in vitro* and in *vivo*,[17,53] 4) oxidative modifications increase protein aggregation propensity and this is particularly true for carbonylation, which can drastically promote aggregability even at low concentrations,[6] and 5) one of the predominant determinants of longevity appears to be resistance of proteome integrity and protein stability to oxidative stress, as recently shown.[54,55]

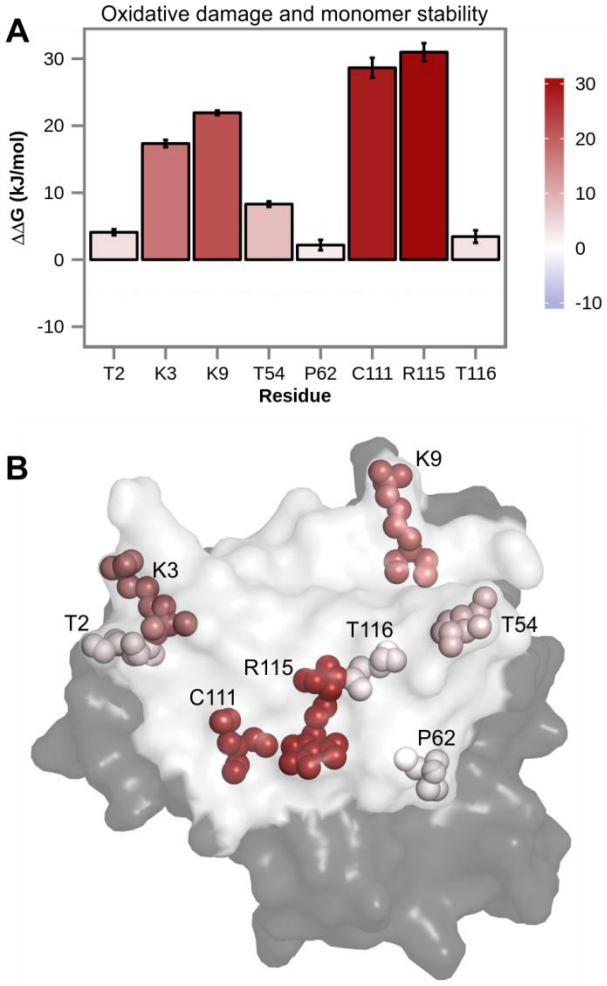


Figure 5. Impact on the SOD1 monomer stability of oxidative damage of residues at the homo-dimer interface. A) changes in free energy of the folding free energy of with the error bars calculated by block averaging and propagation of error. B) Location of the studied interface residues and effects of their oxidative modifications on the SOD1 monomer (view at the interface). *Color code for free energy changes:* blue – stabilization, white - no effect and red – destabilization; color code for *protein structure:* white – interface and grey – rest of the protein.

Finally, even though 5 of the 9 studied modifications involve net charge perturbation, thus potentially weakening the quantitative aspect of the estimated changes in SOD1 monomer and dimer stability, we strongly believe that the calculated data support the qualitative

interpretations given, considering that 1) we use multiple-fold longer simulation time per alchemical modification than in related studies,[29,30,56] arguably achieving better convergence and accuracy, 2) the majority of the estimated changes in free energy point in the direction of destabilization, with a sizeable fraction showing drastic effects and the most extreme examples in excess of 10RT at room temperature, and 3) since lysine and arginine carbonylation is a charge-removing and cysteine oxidation is a charge-introducing modification, one would in principle expect opposite impact on the calculated SOD1 stability (if affected by a systematic error), yet they both exhibit significant destabilization effects.

FURTHER PLANS

In order to test the presented hypotheses and speculations in this chapter, we plan to analyze effects of other oxidizable residues of SOD1 on its homo-dimer and monomer stability. Moreover, we plan to explore if oxidative stress can lead to cytotoxicity through a decrease in the binding affinities to Zn and Cu ions. Additionally, we aim to compare the effects of fALS-related SOD1 point mutations with those of oxidative modifications, particularly those of the mutated residues. To address such a broad range (both in number and type) of mutations and modifications, an efficient and accurate automated protocol for free energy calculations upon alchemical perturbations is essential. In particular, improvements in the protocol that was applied herein could be achieved through: 1) separation of perturbations in electrostatics and van der Waals interactions, 2) calculations of charging corrections in free energy, 3) reducing the number of independent simulations, the number of λ points and/or the length of independent simulations by a more judicious choice of λ points, 4) quantitative estimation of the phase space overlap of MD simulations at neighboring λ points, and 5) resorting to alternative MD-based approaches for free energy calculation.[34,48,49,57,58] Importantly, as a control, we plan to calculate and compare free energy differences with available experimental data,[29] prior to applying calculations to SOD1 systems of interest. Furthermore, we have already developed an automated tool for introduction of alchemical modifications (GROMACS topology format), based on the VF algorithm for graph isomorphism matching.[59] However, extensive testing and debugging is still required to ensure desirable reliability and efficiency of the algorithm. Notably, these efforts might potentially result in a significant advance when it comes to the methodology

for free energy estimation, which may be applicable in different contexts ranging from estimation of protein stability to binding affinity calculations to rational drug development.[29,60,61]

ACKNOWLEDGEMENTS

This work was supported in part by the Austrian Science Fund FWF (START grant Y 514-B11 to BZ, http://www.fwf.ac.at/), European Research Council (ERC Starting Independent grant 279408 to BZ, http://erc.europa.eu/) and the HPC-EUROPA2 project with the support of the European Commission - Capacities Area - Research Infrastructures (project number: 1036 to DP, http://www.hpc-europa.eu/). We thank members of the Laboratory of Computational Biophysics at MFPL for useful advice and assistance.

REFERENCES

1. Levine, R. L. & Stadtman, E. R. Oxidative modification of proteins during aging. *Exp. Gerontol.* **36**, 1495-1502 (2001).
2. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* **24**, 1311-1317 (2005).
3. Dalle-Donne, I., Rossi, R., Colombo, R., Giustarini, D. & Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **52**, 601-623 (2006).
4. Davies, K. J. Degradation of oxidized proteins by the 20S proteasome. *Biochimie* **83**, 301-310 (2001).
5. Grune, T., Jung, T., Merker, K. & Davies, K. J. Decreased proteolysis caused by protein aggregates, inclusion bodies, plaques, lipofuscin, ceroid, and 'aggresomes' during oxidative stress, aging, and disease. *Int. J. Biochem. Cell Biol.* **36**, 2519-2530 (2004).
6. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).
7. Andersen, J. K. Oxidative stress in neurodegeneration: cause or consequence? *Nat. Med.* **10**, S18-S25 (2004).
8. Hand, C. K. & Rouleau, G. A. Familial amyotrophic lateral sclerosis. *Muscle Nerve* **25**, 135-159 (2002).
9. Barber, S. C. & Shaw, P. J. Oxidative stress in ALS: Key role in motor neuron injury and therapeutic target. *Free Radic. Biol. Med.* **48**, 629-641 (2010).
10. Ince, P. G. *et al.* Molecular pathology and genetic advances in amyotrophic lateral sclerosis: an emerging molecular pathway and the significance of glial pathology. *Acta Neuropathol.* **122**, 657-671 (2011).
11. Abel, O., Powell, J. F., Andersen, P. M. & Al-Chalabi, A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum. Mutat.* **33**, 1345-1351 (2012).
12. Kato, S. *et al.* New consensus research on neuropathological aspects of familial amyotrophic lateral sclerosis with superoxide dismutase 1 (SOD1) gene mutations: Inclusions containing SOD1 in neurons and astrocytes. *Amyotroph. Lateral Scler.* **1**, 163-184 (2000).
13. Liu, H.-N. *et al.* Lack of evidence of monomer/misfolded superoxide dismutase-1 in sporadic amyotrophic lateral sclerosis. *Ann. Neurol.* **66**, 75-80 (2009).
14. Forsberg, K., Andersen, P. M., Marklund, S. L. & Brannstrom, T. Glial nuclear aggregates of superoxide dismutase-1 are regularly present in patients with amyotrophic lateral sclerosis. *Acta Neuropathol.* **121**, 623-634 (2011).
15. Bruijn, L. I. *et al.* Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. *Science* **281**, 1851-1854 (1998).
16. Johnston, J. A., Dalton, M. J., Gurney, M. E. & Kopito, R. R. Formation of high molecular weight complexes of mutant Cu,Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12571-12576 (2000).
17. Rakhit, R. *et al.* Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **279**, 15499-15504 (2004).
18. Valentine, J. S., Doucette, P. A. & Potter, S. Z. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annu. Rev. Biochem.* **74**, 563-593 (2005).
19. Banci, L. *et al.* Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: A possible general mechanism for familial ALS. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11263-11267 (2007).

20. Kurahashi, T., Miyazaki, A., Suwan, S. & Isobe, M. Extensive investigations on oxidized amino acid residues in H2O2-treated Cu,Zn-SOD protein with LC-ESI-Q-TOF-MS, MS/MS for the determination of the copper-binding site. *J. Am. Chem. Soc.* **123**, 9268-9278 (2001).
21. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
22. Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**, 52-61 (2012).
23. Vassall, K. A., Stathopulos, P. B., Rumfeldt, J. A. O., Lepock, J. R. & Meiering, E. M. Equilibrium thermodynamic analysis of amyotrophic lateral sclerosis-associated mutant apo Cu,Zn superoxide dismutases. *Biochemistry* **45**, 7366-7379 (2006).
24. Rumfeldt, J. A. O., Stathopulos, P. B., Chakrabarrty, A., Lepock, J. R. & Meiering, E. M. Mechanism and thermodynamics of guanidinium chloride-induced denaturation of ALS-associated mutant Cu,Zn superoxide dismutases. *J. Mol. Biol.* **355**, 106-123 (2006).
25. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using rosetta. *Methods Enzymol.* **383**, 66-93 (2004).
26. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, W306-W310 (2005).
27. Benedix, A., Becker, C. M., de Groot, B. L., Caflisch, A. & Boeckmann, R. A. Predicting free energy changes using structural ensembles. *Nat. Methods* **6**, 3-4 (2009).
28. Potapov, V., Cohen, M. & Schreiber, G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng. Des. Sel.* **22**, 553-560 (2009).
29. Seeliger, D. & de Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **98**, 2309-2316 (2010).
30. Lin, Z., Kornfeld, J., Maechler, M. & van Gunsteren, W. F. Prediction of folding equilibria of differently substituted peptides using one-step perturbation. *J. Am. Chem. Soc.* **132**, 7276-7278 (2010).
31. Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431-492 (1989).
32. Straatsma, T. P. & McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **43**, 407-435 (1992).
33. Peter, C., Oostenbrink, C., van Dorp, A. & van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. *J Chem Phys* **120**, 2652-2661 (2004).
34. de Ruiter, A. & Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **15**, 547-552 (2011).
35. Shirts, M. R. & Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J Chem Phys* **122** (2005).
36. Bruckner, S. & Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *J. Comput. Chem.* **32**, 1303-1319 (2011).
37. Bruckner, S. & Boresch, S. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **32**, 1320-1333 (2011).
38. Gazdag, E. M. *et al.* Purification and crystallization of human Cu/Zn superoxide dismutase recombinantly produced in the protozoan Leishmania tarentolae. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 871-877 (2010).

39. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration. Reidel, Dordrecht, 1981.
40. Boned, R., van Gunsteren, W. E. & Daura, X. Estimating the temperature dependence of peptide folding entropies and free enthalpies from total energies in molecular dynamics simulations. *Chemistry* **14**, 5039-5046 (2008).
41. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.* **4**, 435-447 (2008).
42. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676 (2004).
43. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
44. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
45. Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R. & van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529-539 (1994).
46. Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**, 20313-20316 (1997).
47. Daura, X., Affentranger, R. & Mark, A. E. On the relative merits of equilibrium and non-equilibrium simulations for the estimation of free-energy differences. *Chemphyschem* **11**, 3734-3743 (2010).
48. Kastenholz, M. A. & Huenenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J Chem Phys* **124** (2006).
49. Hunenberger, P. & Reif, M. in *Single-ion solvation: Experimental and theoretical approaches to elusive thermodynamic quantities RSC Theoretical and Computational Chemistry Series* (ed Jonathan Hirst) 1-664 (Royal Society of Chemistry, London, 2011).
50. Forman, H. J. & Fridovic.I. Stability of bovine superoxide dismutase - effects of metals. *J. Biol. Chem.* **248**, 2645-2649 (1973).
51. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15-22 (2009).
52. Gagliardi, S. *et al.* SOD1 mRNA expression in sporadic amyotrophic lateral sclerosis. *Neurobiol. Dis.* **39**, 198-203 (2010).
53. Oeda, T. *et al.* Oxidative stress causes abnormal accumulation of familial amyotrophic lateral sclerosis-related mutant SOD1 in transgenic Caenorhabditis elegans. *Hum. Mol. Genet.* **10**, 2013-2023 (2001).
54. Perez, V. I. *et al.* Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3059-3064 (2009).
55. Krisko, A. & Radman, M. Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14373-14377 (2010).
56. Lin, Z. & van Gunsteren, W. F. Combination of Enveloping Distribution Sampling (EDS) of a Soft-Core Reference-State Hamiltonian with One-Step Perturbation to Predict the Effect of Side Chain Substitution on the Relative Stability of Right- and Left-Helical Folds of beta-Peptides. *J. Chem. Theory. Comput.* **9**, 126-134 (2013).

57. Wu, D. & Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J Chem Phys* **123** (2005).
58. Wu, D. & Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J Chem Phys* **123** (2005).
59. Cordella, L. P., Foggia, P., Sansone, C. & Vento, M. Subgraph transformations for the inexact matching of attributed relational graphs. *Graph Based Representations in Pattern Recognition* **12**, 43-52 (1998).
60. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* **21**, 150-160 (2011).
61. de Ruiter, A. & Oostenbrink, C. Efficient and accurate free energy calculations on trypsin inhibitors. *J. Chem. Theory. Comput.* **8**, 3686-3695 (2012).

Concluding discussion

From enzymatic activation to transcription and translation regulation to disease development and aging, post-translational modifications (PTMs) of proteins play an essential role in various biological processes.[1-3] However, despite the great importance of understanding atomistic-level effects of PTMs, molecular dynamics (MD) simulations,[4] a premier high-resolution computational biology tool, has been limited to unmodified, native proteins due to a surprising deficiency of suitable tools and systematically developed parameters for treating modified proteins. To fill this gap, we have obtained novel force field parameters for more than 250 different types of enzymatic and non-enzymatic PTMs occurring at amino-acid side chains and protein termini as presented in **Chapter I**. These parameters have been generated in the context of the GROMOS force field (45a3[5] and 54a7[6,7] parameter sets), chosen primarily because of its wide-spread usage, accuracy in reproducing experimental data and general transferability of parameters between chemically similar groups in different compounds.[6] Importantly, such parameterization strategy has yielded a remarkable, near chemical-accuracy matching of calculated and experimentally measured hydration free energies (RMSE = 4.2 kJ/mol over the validation set comprised of small molecules), a thermodynamic property closely related to hydrophobicity and originally used to develop the GROMOS 53a6 force field in 2004,[6] lending support for the general validity of the reported parameters and soundness of the parameterization philosophy underlying the GROMOS force field. However, further verification is expected to be carried out at a community-wide level through simulations of post-translationally modified proteins and direct comparison with relevant experimental data. Additionally, this study has aimed at systematic parameterization of a large majority of known PTMs, providing human curated, fully-consistent set of parameters, which has to be contrasted with typical MD studies focusing on a single modification using different procedures and force fields,[8-10] and available tools for automated generation of force field parameters (e.g., the AMBER[11] feature *antechamber* and online tools SwissParam[12], PRODRG[13], ATB[14] and *q4md-forcefieldtools*[15]), directed at small chemicals rather than protein PTMs. Moreover, the reported parameters have been used to quantitatively show that most PTMs significantly alter physico-chemical properties (e.g., hydrophobicity) of target amino acids, a mechanism shown to have broad biological implications.[16,17]

In an accompanying study presented in **Chapter II**, we have developed the Vienna-PTM web server designated to automated incorporation of PTMs into protein 3D structures and supplying the associated force field parameters described in Chapter I. Notably, in addition to reliability and compatibility with different operating systems and web browsers, the emphasis of servers' design has been primarily put on flexible structure allowing for easy and straightforward introduction of new features, tools and modification types or even completely new force fields. Therefore, even though the server currently provides exclusively GROMOS force field parameters of protein PTMs in GROMOS[18] and GROMACS[19] formats, our hope is that the implemented service will expand to other widely-used force fields (e.g., AMBER,[20] CHARMM[21] and OPLS[22]) and simulation packages (e.g., AMBER[11], CHARMM[23] and NAMD[24]) as well as modifications of other biologically important molecules such as nucleic acids and phospholipids.[25,26] Collectively, the developed PTM parameters together with the Vienna-PTM server provide a comprehensive and user-friendly platform for preparing, running and analyzing MD simulations of modified proteins.

While enzymatic PTMs regulate a large number of cellular processes, non-enzymatic, oxidative damage to proteins is associated with aging and age-related disorders, including neurodegenerative diseases, cancer and diabetes.[3,27-29] Additionally, potentially cytotoxic aggregates comprised of highly oxidized proteins have consistently been reported in these pathologies.[30,31] However, a direct causal relationship between oxidative stress on the one hand and protein aggregation, aging and development of late onset diseases on the other has not been established. In order to examine this relationship, we study the effects of non-enzymatic oxidative modifications on protein structure, dynamics and aggregation propensity. In particular, in **Chapter III** we focus on metal-catalyzed carbonylation, one of the most important and the most studied non-enzymatic PTMs,[32] and examine how it affects the villin headpiece domain, a structurally well characterized polypeptide.[33,34] Using MD simulations, we have modeled all possible variants with Lys, Arg and Pro residues carbonylated. This microscopic, site-specific analysis reveals that disruption of specific structural elements (e.g., a salt bridge) together with high overall levels of carbonylation, i.e., approximately three orders of magnitude above those related to cellular death,[3] are required to unfold villin headpiece, only a marginally stable protein.[35] Surprisingly, even though it is widely assumed that carbonylation leads to cytotoxicity through destabilization

of protein structure,[32] this result, additionally corroborated with experimental evidence,[36,37] suggests that protein structure most likely remains intact at typical cellular levels of carbonylation. On the other hand, we show that local changes in physico-chemical properties upon single carbonylation event, i.e., *point mutation* to a charge-neutral and more hydrophobic residue when compared to the cognate amino acid, drastically increase protein aggregation propensity. Importantly, these findings potentially provide a direct link between biologically relevant conditions and cytotoxicity induced by formation of insoluble protein inclusions.

To investigate this further, we have simulated aggregation of the villin headpiece domain in its native and carbonylated forms as described in **Chapter IV**. As expected, carbonylated, aggregation-prone version of the peptide self-associate to dimers and higher order oligomers, however, the carbonylated residues play only a small role in this process. Surprisingly, in addition to hydrophobic ring-containing residues, i.e., phenylalanine and tryptophan, polar asparagines and glutamines as well as hydrophilic backbone are predominantly involved in the formation of inter-peptide contacts, supporting the hypothesis that aggregability is an intrinsic property of polypeptide chains.[38] Importantly, the native villin displays the same aggregation mechanism and behavior as the carbonylated form, accompanied by a drastic decrease in the potential energy of the system, despite the fact that the applied simulation conditions correspond to those, which are experimentally known to render villin soluble.[13,14] Strikingly, we demonstrate that such aggregation-biasing imbalance between protein-protein, protein-solvent and solvent-solvent interactions is also shared by six widely used force fields, namely: GROMOS45a3,[5] GROMOS54a7,[7] AMBER94,[39] AMBER99SB-ILDN,[20] CHARMM27-CMAP,[21] and OPLS-AAL.[22] One may speculate that the favoring of protein-protein over protein-solvent interactions by current force fields is an artifact of a commonly accepted verification procedure, in which the ability of a given force field to reproduce experimentally obtained secondary and tertiary protein structure is tested. This could also be related to the finding that most force fields exhibit a tendency to describe amino acids as being more hydrophobic than they actually are.[40] This, in turn, implies that MD simulations potentially underestimate protein dynamics and unstructuredness, as indeed shown in a recent study of an intrinsically disordered protein.[41] Importantly, this suggests that current force fields in general might have limited accuracy

when it comes to describing protein dynamics and behavior in biologically relevant crowded environments.

In conditions of increased oxidative stress, carbonylated proteins are most likely affected by other oxidative modifications as well, since they are more readily introduced when compared to carbonylation modifications.[42] However, one should emphasize that as an irreversible reaction, carbonylation permanently damages target residues, and is for this reason arguably one of the most widely-used biomarkers of protein damage, oxidative stress and aging.[32,43] In Chapters III and IV, the villin headpiece domain has been used as a model system primarily due to its small size, fast folding kinetics and well-characterized folding mechanism.[13,14] Even though this actin-binding peptide is not directly linked to aging and oxidative stress, actin cytoskeleton morphology and structure are known to get compromised upon oxidative damage.[44,45] In order to extend the scope of our study, in **Chapter V** we have explored different types of non-enzymatic oxidative PTMs and their potential cytotoxicity as mediated through damage to superoxide dismutase 1 (SOD1), a key antioxidant protein in humans which, by the nature of its function, is exposed to higher levels of oxidative stress than most other proteins. Importantly, this homo-dimeric metalloenzyme has been repeatedly linked to familial amyotrophic lateral sclerosis (fALS), a devastating, late-onset neurodegenerative disease, with more than 150 fALS-related mutations in the SOD1 gene.[46-48] Additionally, aggregated SOD1 has been reported in motor neurons and astrocytes of fALS patients[49-51] and animal models.[52,53] However, the exact molecular mechanism of cytotoxicity triggered by SOD1 remains elusive. As a first step towards better understanding of the microscopic effects of oxidative damage to SOD1, we have employed classical MD simulations in combination with the thermodynamic integration method[54] to estimate changes in the free energy of SOD1 folding and dimerization. In particular, we have focused on nine residues at the homo-dimer interface, which are affected by presumably the most prominent non-enzymatic oxidative PTMs. Our results show that most of the studied oxidative modifications destabilize both homo-dimer and monomer structures, the mechanisms that have been shown to increase aggregation propensity of SOD1 protein.[55,56] Markedly, the extent of destabilization observed herein is significantly greater than that experimentally determined for fALS-related mutations,[57] even up to 10-fold in the most extreme cases. This suggests that, as speculated before,[55] age-

related increase in oxidative damage to SOD1 may play a key role in ALS development, with the reported SOD1 mutations being only an additional factor in promoting early onset of the disease. Although in this study we provide only calculated estimates of the changes in stability without experimental verification, two recent studies applying similar methods involving alchemical perturbations have successfully accessed effects of side-chain substitutions on the thermodynamic stability of microbial ribonuclease barnase and a hepta-β-peptide,[58,59] obtaining results in a good agreement with experimental measurements. Notably, we have performed significantly longer simulations than in these related studies,[58,59] arguably achieving better convergence and accuracy. However, such extensive sampling comes at a price of high computational requirements, which further engenders the need for an improved and more efficient protocol in order to further investigate the relationship between oxidative stress and SOD1-mediated cytotoxicity. In particular, such improvements would be needed in order to analyze the effects of additional oxidizable residues as well as fALS-related mutations (including oxidatively-modified versions of mutated residues) on the stability of SOD1 homo-dimer and monomer, and binding affinities to Zn and Cu ions which are important for the dismutase activity as well as the tertiary and quaternary structure formation and stability of SOD1.[60-62] In particular, by using more advanced approaches for sampling and free energy evaluation (e.g., local elevation method[63] and Bennett acceptance ratio[64]) and quantitative estimation of the phase space overlap between states along the perturbation path, or methods for calculating free energy corrections of charging a part of the system,[65-70] one should in principle decrease the computational costs without affecting, if not even enhancing, the accuracy. Additionally, an efficient and reliable tool for automated introduction of alchemical modifications is essential for treating a wide range of modifications and mutations, as required to systematically address biologically relevant variants of SOD1. We have already made a significant progress in developing an automated dual topology builder (GROMACS format) based on the VF algorithm for graph isomorphism matching,[71] designed to find the greatest overlap between two given compounds, i.e., the largest common substructure, and create an arguably optimal alchemical perturbation path between the two compounds. Importantly, in addition to the analysis of the discussed SOD1 systems, such a powerful tool can be utilized in various contexts involving alchemical free energy calculations.[72] We plan

to: firstly, introduce this tool as an additional feature to the Vienna-PTM server to provide dual topologies of modified and cognate amino acids, and secondly, to calculate relative changes in physico-chemical properties (e.g., hydration free energy and solvation free energy in non-polar solvents) of canonical amino acids upon different PTMs. Furthermore, the range of applicability of this tool extends to research areas involving calculation of binding free energies, including protein ligand binding affinities, screening procedures for a large number of candidate compounds and rational drug design.[69,72,73]

Lastly, driven by the extreme biological importance of more than 400 existing PTMs, and how they have surprisingly been largely neglected by the MD community, this thesis is primarily aimed at providing a systematic, comprehensive and user-friendly toolkit designed for computational modeling of post-translationally modified proteins. From direct MD simulations to biomolecular structure refinement to computational free energy estimation and drug design, the developed force field parameters in combination with the Vienna-PTM server greatly expand the range of MD methodology to a large class of biomolecular systems of paramount importance. It is our hope that this advance will play a catalytic role in bringing together realistic cell biology, dominated by PTMs, and the quantitative, reductionist power of structural biology and chemistry, as embodied in the MD method, and help shed light on a broad spectrum of important biological questions. By using these tools, we have explored potential molecular mechanisms of aging and development of age-related diseases in the context of oxidatively damaged proteins. By probing atomic-level dynamics and gaining high-resolution insight into effects of oxidative modifications on protein structure, dynamics and interactions, we have: linked microscopic determinants of protein structure stability and aggregation to aging through oxidative stress, made and provided experimentally testable predictions, and tested limitations of current models to describe biologically relevant crowded environments. We believe that such computational studies, in synergy with experimental efforts, will improve our microscopic-level understanding of oxidative stress, disease development and aging and possibly lead to novel and more successful therapeutic strategies.

REFERENCES

1. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342-7372 (2005).
2. Lothrop, A. P., Torres, M. P. & Fuchs, S. M. Deciphering post-translational modification codes. *FEBS Lett.* **587**, 1247-1257 (2013).
3. Levine, R. L. & Stadtman, E. R. Oxidative modification of proteins during aging. *Exp. Gerontol.* **36**, 1495-1502 (2001).
4. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064-4092 (2006).
5. Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* **22**, 1205-1218 (2001).
6. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676 (2004).
7. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843-856 (2011).
8. Petrov, D. & Zagrovic, B. Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**, 7016-7024 (2011).
9. Seeliger, D. *et al.* Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem. Biol.* **7**, 150-154 (2012).
10. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17857-17862 (2012).
11. Case, D. A. *et al.* AMBER 12. University of California, San Francisco, 2012.
12. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359-2368 (2011).
13. Schuttelkopf, A. W. & van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60**, 1355-1363 (2004).
14. Malde, A. K. *et al.* An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory. Comput.* **7**, 4026-4037 (2011).
15. Vanquelef, E. *et al.* RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511-W517 (2011).
16. Huq, M. D. M., Tsai, N. P., Khan, S. A. & Wei, L. N. Lysine trimethylation of retinoic acid receptor-alpha - A novel means to regulate receptor function. *Mol. Cell. Proteomics* **6**, 677-688 (2007).
17. Hlevnjak, M., Zitkovic, G. & Zagrovic, B. Hydrophilicity matching - A potential prerequisite for the formation of protein-protein complexes in the cell. *PLoS ONE* **5(6)** (2010).
18. Schmid, N., Christ, C. D., Christen, M., Eichenberger, A. P. & van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **183**, 890-903 (2012).
19. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854 (2013).
20. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
21. Chen, J. H., Im, W. P. & Brooks, C. L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **128**, 3728-3736 (2006).

22. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* **105**, 6474-6487 (2001).
23. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545-1614 (2009).
24. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802 (2005).
25. Fu, Y. & He, C. A. Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* **16**, 516-524 (2012).
26. Nyska, A. & Kohen, R. Oxidation of biological systems: Oxidative stress phenomena, antioxidants, redox reactions, and methods for their quantification. *Toxicol. Pathol.* **30**, 620-650 (2002).
27. Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J. Biol. Chem.* **272**, 20313-20316 (1997).
28. Sohal, R. S. Role of oxidative stress and protein oxidation in the aging process. *Free Radic. Biol. Med.* **33**, 37-44 (2002).
29. Garcia-Garcia, A. *et al.* Biomarkers of Protein Oxidation in Human Disease. *Curr. Mol. Med.* **12**, 681-697 (2012).
30. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10-S17 (2004).
31. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
32. Nystrom, T. Role of oxidative carbonylation in protein quality control and senescence. *EMBO J.* **24**, 1311-1317 (2005).
33. Zagrovic, B., Snow, C. D., Shirts, M. R. & Pande, V. S. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* **323**, 927-937 (2002).
34. Lei, H., Wu, C., Liu, H. & Duan, Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4925-4930 (2007).
35. Godoy-Ruiz, R. *et al.* Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J Phys Chem B* **112**, 5938-5949 (2008).
36. Lasch, P. *et al.* Hydrogen peroxide-induced structural alterations of RNAse A. *J. Biol. Chem.* **276**, 9492-9502 (2001).
37. Le, H. T. *et al.* Impact of hydrogen peroxide on the activity, structure, and conformational stability of the oxidized protein repair enzyme methionine sulfoxide reductase A. *J. Mol. Biol.* **393**, 58-66 (2009).
38. Gsponer, J. & Vendruscolo, M. Theoretical approaches to protein aggregation. *Protein Pept. Lett.* **13**, 287-293 (2006).
39. Cornell, W. D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
40. Shirts, M. R., Pitera, J. W., Swope, W. C. & Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J Chem Phys* **119**, 5740-5761 (2003).
41. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134**, 3787-3791 (2012).
42. Stadtman, E. R. Protein oxidation and aging. *Science* **257**, 1220-1224 (1992).
43. Dalle-Donne, I., Giustarini, D., Colombo, R., Rossi, R. & Milzani, A. Protein carbonylation in human diseases. *Trends Mol. Med.* **9**, 169-176 (2003).

44. Dalle-Donne, I. *et al.* Actin carbonylation: from a simple marker of protein oxidation to relevant signs of severe functional impairment. *Free Radic. Biol. Med.* **31**, 1075-1083 (2001).
45. Dalle-Donne, I., Rossi, R., Milzani, A., Di Simplicio, P. & Colombo, R. The actin cytoskeleton response to oxidants: from small heat shock protein phosphorylation to changes in the redox state of actin itself. *Free Radic. Biol. Med.* **31**, 1624-1632 (2001).
46. Hand, C. K. & Rouleau, G. A. Familial amyotrophic lateral sclerosis. *Muscle Nerve* **25**, 135-159 (2002).
47. Ince, P. G. *et al.* Molecular pathology and genetic advances in amyotrophic lateral sclerosis: an emerging molecular pathway and the significance of glial pathology. *Acta Neuropathol.* **122**, 657-671 (2011).
48. Abel, O., Powell, J. F., Andersen, P. M. & Al-Chalabi, A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum. Mutat.* **33**, 1345-1351 (2012).
49. Kato, S. *et al.* New consensus research on neuropathological aspects of familial amyotrophic lateral sclerosis with superoxide dismutase 1 (SOD1) gene mutations: Inclusions containing SOD1 in neurons and astrocytes. *Amyotroph. Lateral Scler.* **1**, 163-184 (2000).
50. Liu, H.-N. *et al.* Lack of evidence of monomer/misfolded superoxide dismutase-1 in sporadic amyotrophic lateral sclerosis. *Ann. Neurol.* **66**, 75-80 (2009).
51. Forsberg, K., Andersen, P. M., Marklund, S. L. & Brannstrom, T. Glial nuclear aggregates of superoxide dismutase-1 are regularly present in patients with amyotrophic lateral sclerosis. *Acta Neuropathol.* **121**, 623-634 (2011).
52. Bruijn, L. I. *et al.* Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. *Science* **281**, 1851-1854 (1998).
53. Johnston, J. A., Dalton, M. J., Gurney, M. E. & Kopito, R. R. Formation of high molecular weight complexes of mutant Cu,Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12571-12576 (2000).
54. Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431-492 (1989).
55. Rakhit, R. *et al.* Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **279**, 15499-15504 (2004).
56. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15-22 (2009).
57. Vassall, K. A., Stathopulos, P. B., Rumfeldt, J. A. O., Lepock, J. R. & Meiering, E. M. Equilibrium thermodynamic analysis of amyotrophic lateral sclerosis-associated mutant apo Cu,Zn superoxide dismutases. *Biochemistry* **45**, 7366-7379 (2006).
58. Lin, Z., Kornfeld, J., Maechler, M. & van Gunsteren, W. F. Prediction of folding equilibria of differently substituted peptides using one-step perturbation. *J. Am. Chem. Soc.* **132**, 7276-7278 (2010).
59. Seeliger, D. & de Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **98**, 2309-2316 (2010).
60. Valentine, J. S., Doucette, P. A. & Potter, S. Z. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annu. Rev. Biochem.* **74**, 563-593 (2005).
61. Banci, L. *et al.* Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: A possible general mechanism for familial ALS. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11263-11267 (2007).

62. Kurahashi, T., Miyazaki, A., Suwan, S. & Isobe, M. Extensive investigations on oxidized amino acid residues in H2O2-treated Cu,Zn-SOD protein with LC-ESI-Q-TOF-MS, MS/MS for the determination of the copper-binding site. *J. Am. Chem. Soc.* **123**, 9268-9278 (2001).
63. Huber, T., Torda, A. E. & van Gunsteren, W. F. Local elevation - A method for improving the searching properties of molecular-dynamics simulation. *J. Comput. Aided Mol. Des.* **8**, 695-708 (1994).
64. Bennett, C. H. Efficient estimation of free-energy differences from Monte-Carlo data. *J Comput Phys* **22**, 245-268 (1976).
65. Wu, D. & Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J Chem Phys* **123** (2005).
66. Wu, D. & Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J Chem Phys* **123** (2005).
67. Kastenholz, M. A. & Huenenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *J Chem Phys* **124** (2006).
68. Hunenberger, P. & Reif, M. in *Single-ion solvation: Experimental and theoretical approaches to elusive thermodynamic quantities RSC Theoretical and Computational Chemistry Series* (ed Jonathan Hirst) 1-664 (Royal Society of Chemistry, London, 2011).
69. de Ruiter, A. & Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **15**, 547-552 (2011).
70. Garate, J. A. & Oostenbrink, C. Free-energy differences between states with different conformational ensembles. *J. Comput. Chem.* **34**, 1398-1408 (2013).
71. Cordella, L. P., Foggia, P., Sansone, C. & Vento, M. Subgraph transformations for the inexact matching of attributed relational graphs. *Graph Based Representations in Pattern Recognition* **12**, 43-52 (1998).
72. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: Progress and challenges. *Curr. Opin. Struct. Biol.* **21**, 150-160 (2011).
73. Shirts, M. R., Mobley, D. L. & Chodera, J. D. in *Annual Reports in Computational Chemistry, Vol 3* Vol. 3 *Annual Reports in Computational Chemistry* (eds D. C. Spellmeyer & R. A. Wheeler) 41-59 (2007).

Acknowledgments

In the end, I would like to express my gratitude to kind people who have helped and supported me along this exciting journey.

First and foremost, I would like to thank my supervisor **Bojan Zagrovic** for giving me the opportunity to join his group. I am deeply grateful for the vision, guidance, immense knowledge, necessary help and advice that you have provided. Most of all, your thoughtful approach to science and the courage to ask the unexpected, almost eccentric, questions have been an enormous source of ideas and inspiration for my work. Simply put, in my humble attempt to match your impressive message-conveying ability as the master of analogies, you have been a compass navigating me through rough waters of research.

I would like to show my greatest appreciation to **Chris Oostenbrink** by assigning my partial gratitude as follows: 0.7 on the wizard-like chemistry knowledge and intuition as the main contribution to our interaction, 0.2 on the old-schoolness to account for the hand-edited manuscript and 0.1 on the great and fruitful collaboration to add up to 1 net gratitude. I would also like to offer my special thanks to **Christian** for a great deal of help, work and contribution to this thesis, especially for stoically accepting web-design suggestions, and **Melanie** for countless comments on the Vienna-PTM Board (you should earn some additional experience points though).

I have greatly benefited from spending two months in Barcelona, receiving generous support and insightful comments and suggestions from my host **Xavier Daura**. In addition, *muchas gracias* goes to people from the institute for making my visit a joyful experience, especially to **Txino** for letting me steal his coffee and sugar and all his help, and to **Alicia** and **Montse** for helping me integrate with the locals and showing me the best bar in Cerdanyola.

I am indebted to the **members** of the Laboratory of Computational Biology (including the coffee machine). It has been a great pleasure and a privilege to work in such relaxed and creative environment. I would like to thank you all for enlightening discussions, kind assistance and valuable criticism, as well as friendly atmosphere. In particular, I would like to mention **Antonija**, for all the great moments that we have shared (back from high-school

days to date), for revealing the secrets of online shopping to me and for imposing on me good-quality TV-shows and books (I have almost finished reading Calvin and Hobbes), and even a guitar that unleashed a long-forgotten artist in me; and **Mario**, for being such a great colleague, a flat mate (that *kulen* you always bring is just fabulous) and a friend above all.

I appreciate the feedback, constructive discussion and suggestions offered by the members of my thesis committee: **Stefan Boresch, Robert Konrat** and **Egon Ogris**.

I also thank **Nico** for granting me (a person who has never touched a pipette beforehand) permission to use his laboratory equipment and helping me with protein purification. Unfortunately, these experiments have not been included in the thesis.

I owe a very important debt to **Davor Juretic**, who has been an encouraging and supportive supervisor, in whose group I have made my very first scientific steps at the University of Split. I would also like to mention **Dinko** for endless coffee breaks and discussions at Mediterranean Institute for Life Sciences that have sparked a few ideas incorporated in this thesis.

Surely, without the help from the administration **team** at Max F. Perutz Laboratories, this thesis would have never been submitted, and most probably I would have gotten deported from Austria due to the expiration of my residence permit.

For bringing the essentials to the life of us, poor students (Zwickel, lamb and tons of fun), a well-deserved thank you goes to **O'Connor's Old Oak**, a place that has been a good second home to me and my **friends** to whom I owe a massive *cheers*: a magnificent couple **Ana** and **Maja,** always ready to slack off and join for a pint of milk and beer, respectively; **Vanja**, an honorary organizer of unforgettable adventures, from trips to Mexico to cycling down the Danube to Bratislava; and a number of others that have been sharing laughs and good times with me on a daily basis. In addition, I am particularly grateful to my **mates** from Split, who make a tremendous effort to stay in touch with me and throw at least 2 welcome and 3 farewell parties each time I go back home.

Lastly, I owe my heartfelt gratitude to my **family**, who I endlessly love: my mother **Vesna**, ever caring and loving, the only person able to diligently prepare a life-time supply of the very best food, stuffing up my freezer beyond the laws of physics, bending the space to fill

yet another portion of her magic; my father **Rato**, who has taught me that finding the right answer begins always with asking the right question and whose infectious curiosity rolled me down the path of research and science; and my brother **Vedran**, whose first and foremost wish at the age of 5 was to get an older brother (I hope I managed to match the expectations), who has been watching after me ever since as a patient 'younger' brother, tolerant of my mischief.

Curriculum vitae

PERSONAL DATA

Date and place of birth	7th April 1986, Split, Croatia
Current address	Muhrengasse 35/6, 1100 Vienna Austria
Phone	+43 681 20366574
Email	drazen.petrov@univie.ac.at

EDUCATION

Oct 2010 - present	**University of Vienna, Vienna, Austria** Molecular Biology Doctoral Programme (transfer from Split)
Oct 2009 - Oct 2010	**University of Split, Faculty of Sciences, Split, Croatia** Biophysics Doctoral Programme (transfer to Vienna)
Oct 2004 - Jan 2009	**University of Split, Faculty of Sciences, Split, Croatia** MSc in Physics Thesis: "Dynamics of antimicrobial peptide Adepantin in interaction with membranes"

WORK EXPERIENCE

Apr 2010 - present	**Max F. Perutz Laboratories GmbH, Vienna, Austria** Effects of post-translational modifications on protein structure, dynamics and aggregability Supervision: Bojan Zagrovic, PhD
Mar 2009 - Mar 2010	**University of Split, Faculty of Sciences, Split, Croatia** Signal peptide and propeptide sequences analysis in host defense peptides Supervision: Davor Juretic, PhD
Nov 2008 - Mar 2010	**Mediterranean Institute for Life Sciences, Split, Croatia** Protein dynamics and stability upon carbonylation damage Supervision: Bojan Zagrovic, PhD

SKILLS

Molecular dynamics simulations	GROMACS simulation package
Programing	Python, Perl, R, PyMOL, HTML, PHP, JavaScript, css

SCHOLARSHIPS AND AWARDS

2012 **HPC-Europa2 Transnational Access program**
Collaborative research visits using High Performance Computing grant

2009 **University of Split, Faculty of Sciences, Split, Croatia**
Award for the best graduated student of the year

2007 **University of Split, Faculty of Sciences, Split, Croatia**
Rector's award

2004 **Croatian Ministry of Science, Technology and Sports**
Scholarship for talented students

PUBLICATIONS

- Petrov D, Zagrovic B (2013) Are current atomistic force fields accurate enough to study proteins in crowded environments? (in preparation)
- Petrov D, Juretic D (2013) Identification of novel antimicrobial peptides using conserved signal peptide and propeptide sequences. (in preparation)
- Petrov D, Margreitter C, Grandits M, Oostenbrink C, Zagrovic B (2013) A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.* **9**(7):e1003154.
- Petrov D, Margreitter C, Zagrovic B (2013) Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* **41**(Web Server issue):W422-6.
- Petrov D & Zagrovic B (2011) Microscopic analysis of protein oxidative damage: effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* **133**(18):7016-24.
- Juretic D, Vukicevic D, Petrov D, Novkovic M, Bojovic V, Lucic B, Ilic N, Tossi A (2011) Knowledge-based computational methods for identifying or designing novel, non-homologous antimicrobial peptides. *Eur Biophys J.* **40**(4):371-85.

CONFERENCES AND WORKSHOPS

2012 Functional Dynamics of Biomolecules - computational and experimental approaches, Lugano, Switzerland

2012 High Performance Computing in Computational Chemistry and Molecular Biology: Challenges and Solutions provided by ScalaLife project, Lausanne, Switzerland

2012 **Careers in Science – Insights from Structural Biologists, Vienna, Austria** - talk presentation

2011 **6th Christmas Biophysics Workshop, Varazdin, Croatia** - talk presentation

2011 3rd EMBO meeting, Vienna, Austria - poster presentation

2010 **Regional Biophysics Conference, Primosten, Croatia** - organization of the conference and poster presentation

2010 1st Prague Protein Spring, Prague, Czech Republic - poster presentation

2009 3rd Adriatic Meeting on Computational Solutions in Life Sciences, Primosten, Croatia - poster presentation

2009 FEBS Practical Course in Computational Biology, hands-on training in numerical simulation of diffusional processes, Split, Croatia

2009 7th European Biophysics Congress, Genoa, Italy - poster presentation

2008 23rd International Conference of Physics Students, Krakow, Poland - poster presentation

2006 21st International Conference of Physics Students, Bucharest, Romania - poster presentation

LANGUAGES

- Croatian (native)
- English (fluent)
- Italian (basic)
- German (basic)