



universität  
wien

# DIPLOMARBEIT

Titel der Diplomarbeit

A Mutational Study of the Allosteric Pathway in KIX

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat)

Verfasserin / Verfasser: Reiner Ribarics  
Matrikel-Nummer: 0400169  
Studienrichtung / Studienzweig A490 Molekulare Biologie  
(lt. Studienblatt):  
Betreuer: Dr. Martin Tollinger

Wien, im Jänner 2010



# Acknowledgements

The following lines are dedicated to everyone who supported me during my thesis. Especially, I want to thank my parents for offering me the great possibility to study. After weeks of working in the lab, I enjoyed the distracting and jauntily moments back home. I am very grateful for my girlfriend's mental support, for standing the rather stressful time of writing the thesis with me and for cheering me up in times of bad mood.

My gratitude goes to Robert Konrat for giving me the opportunity to do my diploma thesis in his research lab, to Martin Tollinger for guiding me through theoretical matters and Sven Brüsweiler for helping me with practical concerns. Furthermore, I want to appreciate my colleagues at the Department of Biomolecular Structural Chemistry at the University of Vienna for guiding me and being there for me whenever I needed you.

# Abstract

## English

Allostery provides a way of regulating protein function and activity. Side chain dynamics and structural conformation shifts are the basis of allosteric regulatory mechanisms. In an allosteric protein an arriving signal changes its properties. In the case of the KIX-domain of the CREB-binding protein (CBP), a  $\approx 2$ -fold increase in affinity to the phosphorylated kinase-inducible domain (pKID) upon binding of the mixed lineage leukaemia (MLL) transcription factor is observed. It is thought that communication between the two binding sites of KIX is mediated by a dense network of hydrophobic amino acids, including isoleucines 611, 657 and 660, phenylalanine 612 and tyrosine 658. Binding of MLL induces a global conformational switch to an excited state exhibiting a higher affinity to pKID. The dynamics of conformational rearrangements are amenable by NMR relaxation measurements. To verify their role in allostery, the dynamics of a series of mutant KIX forms were analysed. Highly conserved amino acids that form an essential part of the allosteric network were substituted for similar ones via site-directed mutagenesis in order to probe the importance of their proposed evolutionary function. NMR  $R_2$  relaxation dispersion experiments were used for quantitative determination of exchange contribution to transverse relaxation. Our data showed that the allosteric network is very sensitive to mutations. Dynamics of all mutants were strongly impaired when compared to wildtype KIX with the exception of KIX Y658V. Isothermal titration calorimetry was used for determining KIX/pKID ( $K_d = 4.11 \mu\text{M}$ ) and KIX·MLL/pKID ( $K_d = 1.84 \mu\text{M}$ ) binding affinities. We came to the same conclusion as Goto et al., 2002 [1]: in-vitro, the KIX·MLL complex displays a 2-fold higher affinity to pKID than free KIX. The free I660V mutant form shows a slightly increased affinity to pKID ( $K_d = 2.42 \mu\text{M}$ ). Binding of the MLL transcription factor does not increase the pKID binding affinity. Allosteric signal transmission and conformational shifts to the excited state seem to be abolished. In order to verify a correlation between dynamics and binding affinity, residual KIX mutants have to be analysed with respect to pKID affinity.

## Deutsch

Allosterie stellt eine Möglichkeit dar Proteinfunktion und Aktivität zu regulieren. Seitenket- tendynamik und strukturelle Konformationsänderungen sind zwei Möglichkeiten allosterische Regulationsmechanismen zu realisieren. In einem Protein, das allosterischer Regulation unter- liegt, kann ein herankommendes Signal die Eigenschaften eben jenes Proteins verändern. Im Fall der KIX-Domäne des CREB-bindenden Proteins wird, nach Bindung des mixed lineage leukaemia (MLL) Transkriptionsfaktors, eine  $\approx 2$ -fach gesteigerte Affinität zur phosphorylated kinase-inducible domain (pKID) gemessen. Nach heutigen Theorien wird die Kommunikation zwischen den beiden Bindungsstellen von KIX über ein dichtes Netz hydrophober Aminosäuren vermittelt. Dazu gehören die Isoleucine 611, 657 und 660, Phenylalanin 612 und Tyrosin 658. Die Interaktion zwischen KIX und MLL induziert eine globale Konformationsänderung hin zu einer Struktur die stärker an pKID bindet, ein sogenannter angeregter Zustand (excited state). Die Dynamik von Konformationsänderungen kann mittels NMR Relaxation gemessen werden. Es wurde die Dynamik einer Reihe an mutierten KIX Varianten analysiert um die Rolle der hydrophoben Aminosäuren in dem allosterischen Netzwerk zu überprüfen. Stark konservierte Aminosäuren die das allosterische Netzwerk bilden, wurden durch Ähnliche ausgetauscht um die Hypothese ihrer evolutionären Funktion zu verifizieren. NMR  $R_2$  Relaxationsdispersion Experimente liefern quantitative Daten über den Beitrag chemischen Austauschs zur transver- salen Relaxation. Laut unseren Daten ist das allosterische Netzwerk sehr sensitiv gegenüber den Austausch einzelner, hochkonservierter Aminosäuren. Mit Ausnahme der Y658V Mu- tante, ist die Dynamik aller mutierter KIX Varianten stark vermindert. Bindungsaffinitäten von KIX/pKID ( $K_d=4.11 \mu\text{M}$ ) und KIX·MLL/pKID ( $K_d = 1.84 \mu\text{M}$ ) wurden mittels isothermaler Titrationskalorimetrie (ITC) gemessen. Die bereits publizierten Ergebnisse von Goto et al., 2002 stimmen mit unseren überein: in-vitro zeigt der binäre KIX·MLL Komplex eine 2-fach höhere Affinität zu pKID als KIX alleine. Die nicht-komplexierte I660V Variante der KIX-Domäne zeigt eine leicht erhöhte Affinität zu pKID ( $K_d = 2.42 \mu\text{M}$ ). Bindung des MLL Transkriptions- faktor zeigt hier keine Erhöhung der pKID Bindungsaffinität. Unsere Ergebnisse deuten darauf hin, dass die allosterische Regulation von KIX und die Konformationsänderungen hin zu einem angeregten Zustand in der I660V Mutante nicht mehr funktionieren. Um jedoch einen Zusam- menhang zwischen Dynamik und Bindungsaffinität, sprich Allosterie, zu verifizieren, müssen auch die restlichen KIX Mutanten auf ihre pKID Bindungseigenschaften analysiert werden.

# Contents

<b>Introduction</b>	<b>10</b>
1.1 Proteins . . . . .	10
1.1.1 From DNA to RNA to protein . . . . .	10
1.1.2 Proteins and their structural properties . . . . .	12
1.1.3 Allosteric regulation of proteins . . . . .	13
1.1.4 Manipulating proteins . . . . .	14
1.2 Eukaryotic cellular transcription: a detailed overview . . . . .	15
1.2.1 Initiation . . . . .	16
1.2.2 Elongation . . . . .	16
1.2.3 Termination . . . . .	17
1.2.4 CREB-binding protein and its KIX domain . . . . .	17
1.3 Nuclear magnetic resonance spectroscopy . . . . .	23
1.3.1 Introduction to quantum mechanics . . . . .	23
1.3.2 Chemical shift . . . . .	26
1.3.3 The vector model . . . . .	27
1.3.4 Multi-dimensional NMR spectroscopy . . . . .	29
1.3.5 Assignment strategies . . . . .	30
1.3.6 Relaxation dispersion experiments . . . . .	34
1.4 Isothermal titration calorimetry . . . . .	37
<b>Methods and Materials</b>	<b>39</b>
2.5 Cell strains . . . . .	39
2.5.1 DH5 $\alpha$ . . . . .	39
2.5.2 BL21(DE3) . . . . .	39
2.6 Site-directed mutagenesis . . . . .	39
2.6.1 Plasmid pHisKIX . . . . .	39
2.6.2 Polymerase chain reaction . . . . .	40
2.6.3 Agarose gel electrophoresis . . . . .	41
2.6.4 Cell transformation . . . . .	41
2.6.5 Plasmid preparation . . . . .	41
2.7 Expression of KIX . . . . .	42

2.7.1	Cell transformation . . . . .	42
2.7.2	Starter culture . . . . .	42
2.7.3	Expression culture . . . . .	42
2.7.4	Purification . . . . .	43
2.7.5	Polyacrylamid gel electrophoresis (PAGE) . . . . .	44
2.8	NMR measurements and data processing . . . . .	45
2.8.1	NMR preparation of the KIX domain of CBP . . . . .	45
2.8.2	Data processing . . . . .	45
2.8.3	$^1\text{H}, ^{15}\text{N}$ -HSQC . . . . .	45
2.8.4	$^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY . . . . .	45
2.8.5	CPMG . . . . .	46
2.9	Isothermal titration calorimetry . . . . .	46
2.9.1	Sample preparation . . . . .	46
2.9.2	Experimental setup . . . . .	47
2.9.3	Data evaluation . . . . .	47
2.10	Media . . . . .	48
2.10.1	M9 minimal medium . . . . .	48
2.10.2	Trace elements . . . . .	48
	<b>Results</b>	<b>49</b>
3.11	NMR dispersion relaxation . . . . .	49
3.12	Isothermal titration calorimetry . . . . .	53
	<b>Discussion</b>	<b>55</b>
	<b>A Appendix</b>	<b>56</b>
A.1	Plasmid map of the KIX domain of CBP . . . . .	56
A.2	Sequence of the KIX domain of CBP . . . . .	57

# List of Figures

1.1	The DNA double helix . . . . .	11
1.2	Schematic of a protein backbone . . . . .	13
1.3	Schematic of polymerase chain reaction . . . . .	14
1.4	Signalling pathways of the CREB-binding protein . . . . .	18
1.5	KIX domain of the CREB-binding protein. . . . .	19
1.6	Comparison of KIX amino acid sequence with other organisms . . . . .	22
1.7	Energy levels and wave functions in quantum mechanics . . . . .	25
1.8	General pulse sequence in two-dimensional NMR spectra . . . . .	29
1.9	$^1\text{H}, ^{15}\text{N}$ -HSQC . . . . .	31
1.10	$^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY . . . . .	32
1.11	Geometry of an excerpt of KIX $\alpha$ 1-helix . . . . .	33
1.12	Sequential walking in $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY spectra . . . . .	34
1.13	Spin echo pulse sequence . . . . .	35
1.14	Schematic of a spin echo . . . . .	36
2.15	Titration curve of free KIX titrated with pKID . . . . .	47
3.16	Comparison of amino acids . . . . .	49
3.17	Relaxation dispersion profile of KIX wildtype and KIX mutant L620A . . . . .	50
3.18	Relaxation dispersion profile of KIX mutants I611L and I611V . . . . .	51
3.19	Relaxation dispersion profile of KIX mutants I611V pH5.9 and F612Y . . . . .	51
3.20	Relaxation dispersion profile of KIX mutants I657L and I657V . . . . .	52
3.21	Relaxation dispersion profile of KIX mutants Y658V and I660V . . . . .	52
3.22	Excerpt of superimposing $^1\text{H}, ^{15}\text{N}$ -HSQC spectra of wildtype KIX·MLL and mutant KIX I660V·MLL . . . . .	53
A.1	Plasmid map of pHisKIX . . . . .	56



# List of Tables

1.1	Experiments providing information for NMR resonance assignment . . . . .	30
2.2	PCR temperature-cycle scheme for site-directed mutagenesis . . . . .	40
A.1	Amino acid sequence of KIX . . . . .	57
A.2	Nucleotide sequence of KIX . . . . .	57

# Introduction

Various minds have faced the question of inheritance over the last millennia and even today people might ask themselves why related species share many characteristics, yet each individual looks different and makes a species what it is. Lots of theories have emerged, however it was not until the 1950s when the mystery of how life transmits hereditary information to its offspring was solved. The discovery of deoxyribonucleic acid, now widely known as DNA, as the hereditary substance of life is a fundamental principle of the science of molecular biology and genetics we know today. [2] The knowledge of inheritance has given researchers of the 20th century the possibility to create new approaches to resolve more of life's mysteries.

## 1.1 Proteins

### 1.1.1 From DNA to RNA to protein

The next question in resolving inheritance was how DNA stores hereditary information and how it is processed in order to yield the characteristics of an organism. Genetic information of DNA is stored similar to a computer's hard drive. A computer stores data using the numbers 0 and 1, but it is the sequence where the information lies in. DNA however is a bit more sophisticated and uses four different building blocks arranged in a linear manner along a backbone. Two anti-parallel DNA strands align to build a double helical shape. The bonding forces between these two linear strands are relayed by the four building blocks, chemically seen as bases. The bases' structure and chemical nature allow only distinct pairs to interact; this is why the two strands have to be complementary in respect to their sequence in order to form the observed double helical shape.

As stated previously, it is the unique order of bases, the DNA sequence, where the information of inheritance lies in. The entirety of DNA inside a cell is referred to as genome and is subdivided into several different groups of sequence types, each having different functions. The most commonly known type is the gene that codes either for a protein or RNA and serves as a blueprint for the construction of these two biopolymers. [3] Proteins are made up of joining together amino acids, whilst RNA, very similar to DNA, consists of four bases and a sugar-phosphate

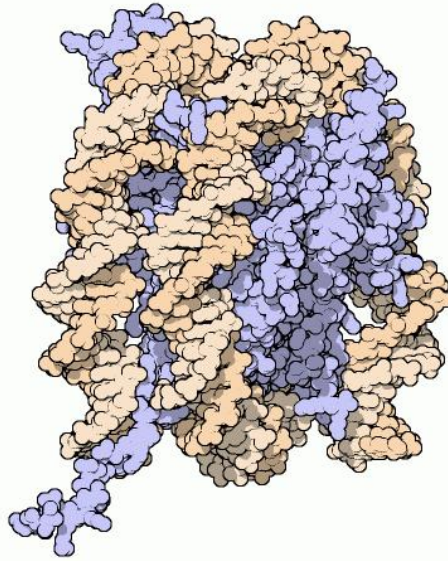


Figure 1.1: *The DNA double helix (wheat) wound around a histone octamer (blue). Illustration by David S. Goodsell, PDB's Molecule of the Month, July 2000.*

backbone. Merely a small part of the genome comprises genes, while most of it is non-coding and represents repetitive motives or introns. The non-coding DNA was for some time referred to as junk DNA; however science is starting to understand now the importance of these sequences as they have been shown to be involved in some illnesses and have therefore proven to possess important biological activities.

The eukaryotic genome is split into chromosomes, their number specific for a species, each consisting of exactly one DNA molecule. The nucleus of a human cell for example contains 23 chromosome pairs with a total length of  $3 \times 10^9$  base pairs, a number so large that without efficient packaging it would not be possible to fit the whole genome into the nucleus. The smallest structures of packaging are achieved by wrapping the DNA double helix twice onto an octameric protein complex referred to as the histone core. [3] The core comprises of four different histone proteins, H2A, H2B, H3 and H4, each occurring two times. Histone H1 is not part of the core particle, but belongs to the nucleosome structure and is needed for the assembly of further DNA superstructures. The wound up DNA strand renders a major part of the genome quite inaccessible for other proteins. Mechanisms exist to regulate packaging intensity; the most commonly known are chemical modification of parts of the histone core. Acetylation, phosphorylation and methylation can influence the affinity of DNA binding by the core particle, therefore providing means of transcriptional regulation. The CREB-binding protein (CBP), which is part of the study of this thesis, and its paralog p300 both possess histone acetyl transferase (HAT), activity. Detailed functions of CBP will be elaborated in chapter 1.2.

Everything we see while looking at any organism is either a protein or a product of such. [3] Since proteins emerge from genes it is clear why DNA and its products are so important for life sciences.

Here it should be noted that DNA itself does not serve as a template for protein synthesis. In fact, special proteins transcribe the antisense DNA strand of a gene to a complementary single stranded RNA that serves as an intermediate for protein synthesis and is generally referred to as messenger RNA (mRNA). [2]

These RNA transcripts serve as a template for ribosomes, themselves made up of protein and RNA, to synthesise proteins. An mRNA sequence is made up of a coding sequence carrying the blueprint of a protein and is flanked by untranslated regions, so-called UTRs. UTRs comprise various binding sites including a ribosomal binding site residing in the 5' UTR. Further processing includes the addition of a 5' CAP and a 3' polyadenine tail further rendering the RNA recognisable by the translation machinery. The coding sequence of a mRNA is arranged in base triplets, whereas each triplet encodes for one amino acid. Amino acids are in turn the building blocks of proteins and during the process of translation ribosomes link amino acids by forming covalent peptide or amide bonds. In doing so a linear amino acid sequence emerges that will, in most cases, adopt a more complex structure to fulfil its function.

The rate at which genes are copied to RNA and further translated into an amino acid sequence, in other words, the gene expression, is precisely regulated by the cell machinery. Virtual any different environmental condition a cell is exposed to needs a special pattern of gene expression. Simple examples are the heat shock proteins, whose expression is increased with elevating temperature levels [4] in order to cope with the new situation. When the temperature is decreased the expression levels will go down again. This can be seen as some sort of energy management, where unnecessary expenses are reduced.

### 1.1.2 Proteins and their structural properties

Proteins are molecules of various shapes and functions. They can catalyse chemical reactions, convey structural properties, control synthesis of other molecules and much more. Proteins consist of a linear string of amino acids held together by covalent amide bonds. Life uses 20 amino acids, each with different chemical properties, to build proteins. Ribosomes function in synthesising a linear sequence of amino acids, also referred to as a protein's primary structure. Various chemical interactions between amino acids lead to a three-dimensional structure peculiar for a protein. When comparing the structure of many different proteins, two folding patterns are often recognised: The  $\alpha$ -helix and  $\beta$ -sheet. Both patterns are very common, because they result from interactions of backbone amide (-N-H) and carbonyl groups (C=O) and are not dependent on a specific amino acid sequence. [2] These two structural elements form a protein's secondary structure, whereas the overall three-dimensional fold or conformation is referred to as the tertiary structure.

But why do proteins need a specific structure? Enzymes for example need a binding site for their substrates, the chemical compounds they will modify, in order to catalyse a chemical reaction. A suitable analogy would be the lock and key mechanism, where the protein holds a

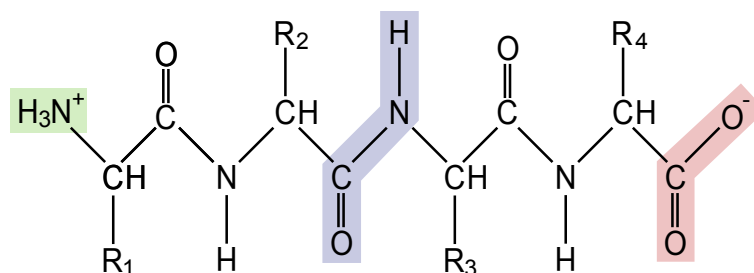


Figure 1.2: *Schematic of a protein backbone. Alpha carbon atoms of amino acids are connected through planar amide bonds (blue). One end carries an amino- (green) and the other a carboxyl-group (red).*

pocket for the substrate to bind. Only if the substrate fits into the pocket, binding and chemical modifications can occur. It is the chemical interactions between an enzyme and its ligand that trigger modifications, so it becomes clear that a change in structure can change the binding pocket and hence the substrate will not fit in anymore. Even slight sterical changes can abolish these interactions and render the function of an enzyme useless. Now it becomes clear why so much effort has been put into developing efficient and reliable methods for protein structure determination.

### 1.1.3 Allosteric regulation of proteins

In many cases the structure of a protein is not totally rigid, but shows an overall dynamic behaviour and can even switch between different conformations. At the level of structure there are mechanisms of regulation of protein function. If a protein undergoes a change in structure, this change will in most cases have an impact on its function. Such conformational shifts can be triggered by ligand binding, chemical modifications or other changes in the environment, for example temperature, pH or ion strength. Dynamics and conformational shifts represent the basis for allosteric regulatory mechanisms of proteins. A simple example of allostery is found in the protein haemoglobin, which is relevant for oxygen distribution of mammals and is responsible for the red colour of blood. Haemoglobin consists of four globular subunits arranged in a circular manner. Each of the subunits can bind one oxygen molecule ( $O_2$ ) via its haem co-factor. Now, if oxygen binds to a subunit, it induces a conformation shift altering the three-dimensional structure of the whole protein aggregate and increasing the affinity of the other, unoccupied subunits to oxygen. So the affinity for oxygen increases with each molecule of oxygen bound and the same is true for oxygen release. While four molecules are bound to the haem co-factor, the first one is hard to be released, but the affinity will decrease for each oxygen molecule released. To get an idea of how these regulatory systems really work, modified versions of proteins can be analysed with respect to their behaviour. Scientific research of allosteric regulatory mechanisms is of particularly interest as the manipulations of such are possible targets for drug design to tackle diseases.

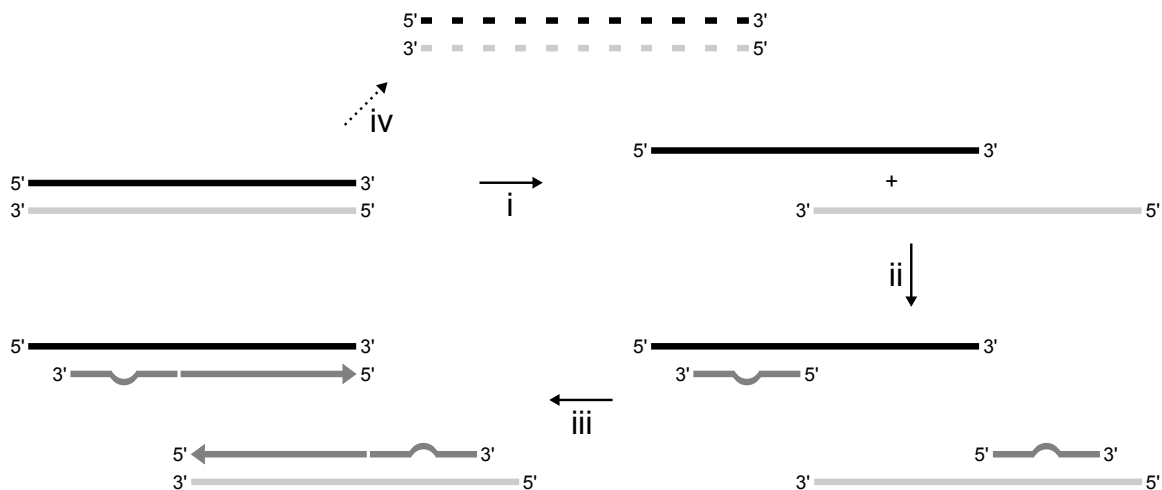


Figure 1.3: *Scheme of the polymerase chain reaction process. Amazingly, replication and amplification of DNA needs very few components: A DNA polymerase that can work at high temperatures, appropriate primers and four deoxyribonucleotide triphosphates: dATP, dCTP, dGTP and dTTP. (i) Melting of the DNA double helix, (ii) Annealing, (iii) Elongation. These three steps are repeated several times to exponentially amplify the DNA template. The primers' bulges indicate non-complementary regions of the DNA strand, where mutations are introduced. (iv) As the last step, methylated template DNA is degraded by DpnI restriction endonuclease.*

#### 1.1.4 Manipulating proteins

Molecular biology in its beginnings and even today is based on manipulating genetic information and analysing the resulting phenotypes. A phenotype can be a morphological or physiological change in the organism of interest. Comparing the unchanged, or wild-type organism, with a mutant one where a gene product is abolished, the change in phenotype can give the researcher a hint about the gene's function. In the former days of molecular biology a mutant organism was produced by exposing the offspring to mutagenic conditions, like radiation or DNA damaging chemicals. Typically known as forward genetics, this approach generates random mutants that are selected by a phenotype of interest by the researcher. However, fast screening methods have to exist in order to analyse thousands or more individual organisms for a specific phenotype. Nowadays, reverse genetics are commonly used, where the sequence of an organisms' genome is typically known and precise methods exist to alter it.

Site-directed mutagenesis is a technique to target for the need of introducing single and even multiple point mutations into the coding sequence of a gene. One or more base pairs can either be substituted with other bases, deleted or new base pairs can be introduced. Polymerase chain reaction is essentially a highly simplified in-vitro reproduction of a cell's DNA replication system and provides the basis for site-directed mutagenesis. Figure 1.3 presents the process of amplifying DNA. Artificially synthesised primers carrying subtle changes in the nucleotide sequence are needed to introduce alterations in the reaction product. [5]

With the possession of the sequence of an organism's genome nearly any gene or region of interest can be changed. Since the DNA sequence of a protein coding gene directly resembles the amino acids sequence, although there are exceptions where the messenger RNA is modified after transcription (RNA editing), it is possible to analyse protein function by producing various mutants, each having one amino acid substituted through another chemically distinct one. This approach can reveal sites highly necessary for protein function.

Another common approach to find important amino acid residues is to compare coding sequences of the same protein of different organisms. Residues not important for gene function can be substituted through other amino acids or deleted by spontaneous mutations at the level of DNA during the progress of evolution. Changes to important residues can have negative effects on the organism or even result in its death, so the spreading of these alleles will most probably be inhibited. What we are likely to find in the genomes of different organisms are alleles that have shown to be still functional. Bioinformatical alignment algorithms can reveal conserved stretches of amino acids. If they are present over a broad range of different organisms they are very likely to be associated with the function of a protein.

## 1.2 Eukaryotic cellular transcription: a detailed overview

Chapter 1.1.1, From DNA to RNA to protein, gives a short overview about the biosynthesis of polypeptides. This section will deal with more complex details of transcription and its control mechanisms<sup>1</sup>.

Cells of a human body contain a genome made up of  $3 \cdot 10^9$  base pairs that stretches across two meters in length. It is subdivided into 46 chromosomes each consisting of one continuous DNA molecule. The enormous length of DNA molecules has led to the need for efficient packaging in order to fit the huge hereditary molecule into the nucleus of a cell. The first level of packaging is achieved by wrapping the DNA twice onto an octameric protein complex referred to as the histone core [3]. Together with DNA and an additional histone H1, this complex is called the nucleosome and comprises the basic repeating element of chromatin. The overall structure of this assembly resembles that of beads on a string and has been shown to prevent initiation of transcription in eukaryotic cells [6], so that genes are typically inactive and not expressed. In order to overcome repression, the DNA has to be made accessible for the transcriptional complex. Eukaryotes have three different RNA polymerases each transcribing a different class of RNA: RNA Polymerase I transcribes ribosomal RNA (rRNA), RNA Polymerase II (Pol II) messenger RNA and RNA Polymerase III is used to transcribe transfer RNA (tRNA).

Products of RNA Polymerase I and III are functional RNA, whereas mRNA serves as a template for protein synthesis. A gene is a small subunit of a usually very long DNA sequence. Finding

---

<sup>1</sup>If not stated otherwise all the information provided in chapter 1.2.1 to 1.2.3 is based on Anthony J. F. Griffiths - Introduction To Genetic Analysis (2004).

the beginning of a gene and placing Polymerase II at the starting site is a demanding process that has to be done very accurately by the transcription machinery. Transcription is an asymmetrical process where only one of the two DNA strands is used as a template [3]. Closely upstream the 5' site of the transcription start lays a regulatory region referred to as promoter to which transcription factors bind and in turn recruit other proteins of the transcriptional complex. Additionally, there exist enhancer and silencer sequences that regulate gene expression and are usually found further upstream or downstream the coding region. However, through the spacial coiling of DNA they come near the transcriptional complex, where they can stimulate gene activation. Transcription comprises three distinct stages termed initiation, elongation and termination.

### 1.2.1 Initiation

In eukaryotes, the Pol II enzyme has the ability to unwind DNA, synthesise RNA and rewind DNA, but is incapable of recognising promoters and initiating transcription [7]. These two tasks are assigned to transcription factors, whereas regulation is performed by proteins known as mediators or co-activators. Mediators are essential to eukaryotic transcription of almost all Pol II promoters [8, 9] and interact directly with Pol II, the general transcription factors and activators that are bound to enhancers [10, 11]. Mediators regulate not only transcriptional activation, but also repression and generally functions as a signal integrator.

Nearly 60 proteins constitute the transcriptional complex. The assembly begins with binding of sequences in the promoter region of a gene by general transcription factors (GTFs) and with attaching of GTFs to one another. The consensus sequence TATA is seen in alignments of eukaryotic promoters and is therefore commonly referred to as TATA box. GTFs attract Pol II and position it at the transcription start. A total of six GTFs together with Pol II comprise the preinitiation complex (PIC). The TATA binding protein (TBP) is part of the TFIID complex and recruits Pol II as well as other GTFs to the promoter hence assembling the PIC.

### 1.2.2 Elongation

After the necessary proteins have been attracted to the promoter the RNA Polymerase II has to escape from this complex to elongate the RNA transcript, a process known as promoter clearance. During promoter clearance there is a chance of abortive initiation, where a truncated RNA transcript is released prior to completion and initiation is interrupted. If elongation continues, most of the GTFs dissociate from RNA Polymerase II, while some of them remain at the promoter site to assemble the next PIC so that multiple transcripts can be synthesised simultaneously. The beginning of the elongation phase is marked by phosphorylation of Pol II carboxyl tail domain (CTD) by one of the GTFs. This modification is thought to weaken the interactions between Pol II and GTFs, therefore releasing it to start transcription. Like DNA, the nascent



RNA transcript is synthesised from its 5' to the 3' end. During the process of transcription, the newly synthesised RNA strand, the pre-mRNA, becomes modified in several ways: First, a 7-methylguanosine residue linked to the transcript by three phosphate groups is appended to the 5' end. This structure is also referred to as cap. Second, a tail of adenine nucleotides is added to the 3' end. Third, the coding sequence is interrupted by non-coding regions, so-called introns, and has to be spliced prior to protein synthesis.

### 1.2.3 Termination

Termination of transcription is not well understood in eukaryotes; however it involves cleavage of the RNA transcript and polyadenylation of such. After the pre-mRNA is properly modified and matured to messenger RNA it is exported from the nucleus to the cytosol where it serves as a template for protein synthesis.

### 1.2.4 CREB-binding protein and its KIX domain

An equally important and complex part of transcription, besides activation, is its regulation. A cell needs control over gene expression in order to adjust a protein's concentration according to its demands and environmental influences. Transcriptional initiation is mediated by promoter or enhancer DNA-binding proteins that recruit the basal transcription complex and RNA polymerase II. [12, 13] Besides, non-DNA-binding proteins called mediators play an important role in efficient recruitment of the transcription complex. Mediators fulfil their function by interlinking components of the basal transcription complex with DNA-binding proteins and also seem to participate in chromatin remodelling.

The CREB-binding protein (CBP) is one of the transcriptional mediators mentioned before that exhibit control over various biological processes such as cellular differentiation, development and growth control [14]. It has been shown that CBP participates in controlling cell cycle and proliferation by being bound to transcription factors, including NF- $\kappa$ B or the tumour suppressor gene p53. It is therefore not surprising that alterations in critical parts of CBP are linked to various tumours and other diseases like Rubinstein-Taybi syndrome.

CBP was shown to be regulated through phosphorylation by the mitogen-activated protein (MAP) kinase pathway [15], therefore giving extracellular signals access to gene expression control. Activity of the mediator can further be regulated by cyclinE/Cdk2, p21 and viral oncogenes, thus intimately tethering CBP to the cell cycle control. Cyclins and cyclin-dependent kinases (Cdks) along with cyclin-dependent kinase inhibitors (CKIs) give rise to the different phases of the cell cycle. Cdks form complexes with specific cyclines, enabling Cdk kinase activity, while CKIs provide a way to negatively regulate the activity of these complexes. Together, the balance of positive (CDK and cyclines) and negative (CKIs) regulators decide when a cell will proliferate, differentiate or arrest in cell cycle [16]. Furthermore, CBP has the ability to

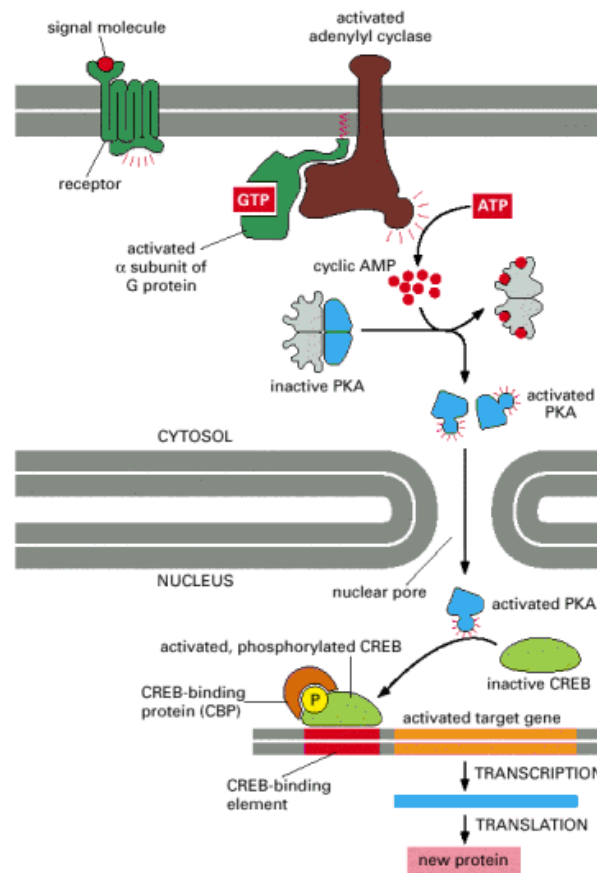


Figure 1.4: *Involvement of CBP in membrane signalling. Membrane receptors activate a signal cascade including adenylyl cyclase, protein kinase (PKA) and cyclic-AMP response element binding protein, CREB. Illustration taken from Alberts, Molecular Biology Of The Cell.*

covalently add acetyl groups to histones of the nucleosome and even non-histone proteins [17] or recruit other histone acetyltransferases (HATs) to loosen up DNA super-structure and thus facilitating transcription by providing access to the DNA double helix. Transcription can also be stimulated through acetylation of transcription factors themselves.

CBP consists of several independently folded protein binding domain modules, including the nuclear receptor interaction domain (RID), the CREB and MYB interaction domain (KIX), the cysteine/histidine regions (CH1 and CH3) and the interferon response binding domain (IBiD). Each module mediates and cross-links various signalling pathways and is separated from the others by unfolded linker sequences. CBP acts as a scaffold for the assembly of the transcriptional machinery by binding general transcription factors and activators to turn on a gene. It integrates signals from activators and relays them to general transcription factors that in turn bind to DNA promoter sequences [18]. Figure 1.4 shows how CPB is included in a signal transduction pathway emerging from the cell plasma membrane. The work of this thesis deals with the KIX domain of the CREB-binding protein, which will be further discussed in detail.

The KIX domain is offering two distinct binding sites for a great variety of transcription factors,

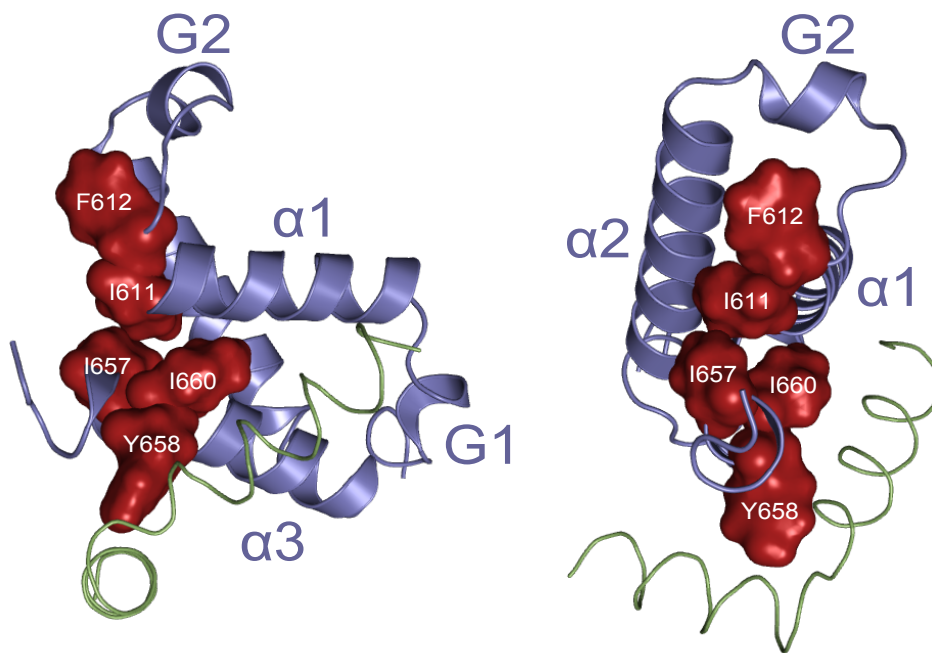


Figure 1.5: *KIX domain of CREB-binding protein (blue) together with pKID (green). The domain comprises three  $\alpha$ -helices ( $\alpha1$ ,  $\alpha2$ ,  $\alpha3$ ) and two  $3_{10}$ -helices (G1, G2). Red coloured electron clouds represent the allosteric network proposed by Brüscheweiler et al. Phe-612 was found to be necessary for MLL binding. Solution structure by Radhakrishnan et al. (1997) [19] and the picture rendered by PyMol [21].*

including phosphorylated kinase-inducible domain (pKID) and mixed lineage leukaemia (MLL) transcription factor. Both proteins can simultaneously bind to KIX with pKID occupying one and MLL the other binding site. Studies have shown that the affinity of KIX to pKID is about a factor of two higher when MLL is already bound to KIX, suggesting an allosteric mechanism behind this effect. Precise knowledge of the structure is needed to understand the subtle details of the proposed allosteric model.

The KIX domain of the CREB-binding protein is a small bundle of three mutually interacting  $\alpha$ - and two short  $3_{10}$ -helices (see figure 1.5) that enclose an extensive hydrophobic core. [19] The  $\alpha$ -helices are denominated  $\alpha1$ ,  $\alpha2$  and  $\alpha3$  and the  $3_{10}$ -helices G1 and G2. Helix  $\alpha1$  comprises residues Gln597-Ile611, while helices  $\alpha2$  and  $\alpha3$  are formed by residues Arg623-Tyr640 and Arg646-Lys662, respectively [20]. The linker sequence of eleven residues that connect helices  $\alpha1$  and  $\alpha2$  harbours the short G2 helix that spans residues Pro617-Lys621. G1 is located near the N-terminus and mapped to residues Trp591-His594. The MLL binding is mediated by hydrophobic interactions with Phe-612 being essential for, while pKID essentially wraps around helix  $\alpha3$  and interacts with a hydrophobic patch formed by helices  $\alpha2$  and  $\alpha3$  [20].

Two hydrophobic cores, whose residues are highly conserved among organisms of the eukaryotic domain (see figure 1.6) are accommodated by the surrounding secondary structural elements,

with Leu628 interlinking both cores. At the carboxy-terminus of helix  $\alpha 1$ , near the MLL binding site, resides Ile611, which contacts Phe612 and is part of the primary hydrophobic core. Its side chain forms part of a dense hydrophobic core along with two aliphatic amino acids, Ile657 and Ile660. The close contacts of Ile611 to Ile657 interlink the carboxy terminus of helix  $\alpha 1$ , which is close to where MLL binds, to the remote pKID binding surface. Further interactions of Ile660 interlink residues in the N-terminal region of MLL with Ile611. Summarising the experimental data so far, the incident of MLL binding to KIX is transmitted through Phe612 and Ile660 to Ile611, which, in turn communicates with Ile657. As part of the hydrophobic surface that consists of a tightly coupled network of interactions, the side chain dynamics of Ile657 can influence pKID binding properties. NMR studies suggested a global switch between two conformations. Further analysis revealed that conformational rearrangements within the highly conserved isoleucine core and the protein backbone occur in a collective manner. [18]

As stated previously, in-vitro studies comparing the binding affinities of free, unbound KIX and the binary KIX·MLL complex to pKID showed that KIX is much more likely to bind pKID when MLL is already present. Such regulation at a distance is commonly referred to as allostery, an effect described in chapter 1.1.3 when discussing the increasing affinity of haemoglobin to oxygen by oxygen binding. A signal originating at one binding site is thereby communicated through changes in the protein structure to trigger a response at the second, remote site [18]. Here, the signal comprises MLL binding to KIX and the response is an increased affinity to the other transcription factor, pKID. By changing the affinity between two proteins, their interactions have to be changed accordingly. This can, for example, involve masking or introduction of strong electrostatic groups or uncovering of hydrophobic clews. All of these changes go along with at least small conformational rearrangements.

However, conformational shifts are not the only mechanism of communication through protein structure. Besides, changes in protein dynamics can trigger allosteric effects. All atoms residing in environments with temperatures over  $0^\circ$  K are constantly in motion; the same being true for large biomolecules like proteins that exhibit thermal motions over timescales ranging from picoseconds to seconds. Different types of motions include, for example, aromatic ring flipping ( $\mu$ s-ms), methyl group rotations (ps-ns), correlated backbone rotations ( $\mu$ s-ms) and protein folding (ms-s and slower). Even proteins of stable structure undergo backbone- and side chain fluctuations that are susceptible to alterations in dynamics. Ligand binding can lead to decreasing entropy and spreading rigidification of side chain residues even to a remote binding site. A less dynamic structure can lead to enhanced ligand binding specificity, while loosened up structures will cause the opposite effect.

NMR relaxation dispersion techniques were applied to KIX by Brüschweiler et al. to quantitatively resolve processes occurring at the micro- to millisecond timescale ( $\approx 0.5$  to 5 ms). Contributions to relaxation dispersion arise from any kind of motion occurring at a given timescale that go along with chemical shift changes. Chemical shifts will be discussed in detail in chapter 1.3, but for now it is sufficient to know that they are sensitive reporters of local structure and

that signals with the same chemical shifts cannot be resolved by NMR. Motions from backbone rearrangements, protein folding or unfolding and aromatic ring flipping occur in the range of micro- to milliseconds. Many amino acids of KIX show dynamic behaviour and after ruling out global unfolding as an underlying process, slow aromatic ring flipping could still contribute, Brüschweiler et al. suggested a continuous and global conformational switch between two differently dispersed states.

The two conformers are referred to as ground- and excited state, with the excited state exhibiting a higher affinity to pKID than the ground state. The excited state is not populated to an appreciable degree [18] in the absence of MLL and is therefore invisible to NMR spectroscopy. However, the population of the excited state increases until saturating concentration of MLL is reached. The excited state structurally resembles the ternary KIX·MLL·pKID complex, therefore it is thought that there exists communication between the binding sites of MLL and pKID. Transmission of allosteric regulation is thought to be mediated by amino acid residues tyrosine 650, histidine 651, alanine 654, isoleucine 657 and tyrosine 658 that form a hydrophobic groove on the KIX surface and serve as a binding site for pKID. [18] Phenylalanine 612, residing at the carboxy-terminal end of helix  $\alpha 1$  and the short loop between  $\alpha 1$  and  $G_1$ , has in turn been found to be essential for MLL binding. Alignments reveal that KIX is well preserved among the eukaryotic domain, as figure 1.6 points out. Blue highlighted sites show residues that are invariant throughout the alignment, including residues of the allosteric network, MLL binding site and the hydrophobic groove of helix  $\alpha 3$ .

This thesis deals with the allosteric properties of the KIX domain of the CREB-binding protein and tries to further characterise them by NMR relaxation dispersion techniques. Site-directed mutagenesis was used as a means of introducing point mutations to give rise to single amino acids substitutions in the KIX primary sequence. The study of mutant KIX variants should provide further insights into the allosteric pathway and the evolutionary conserved function of conserved residues.

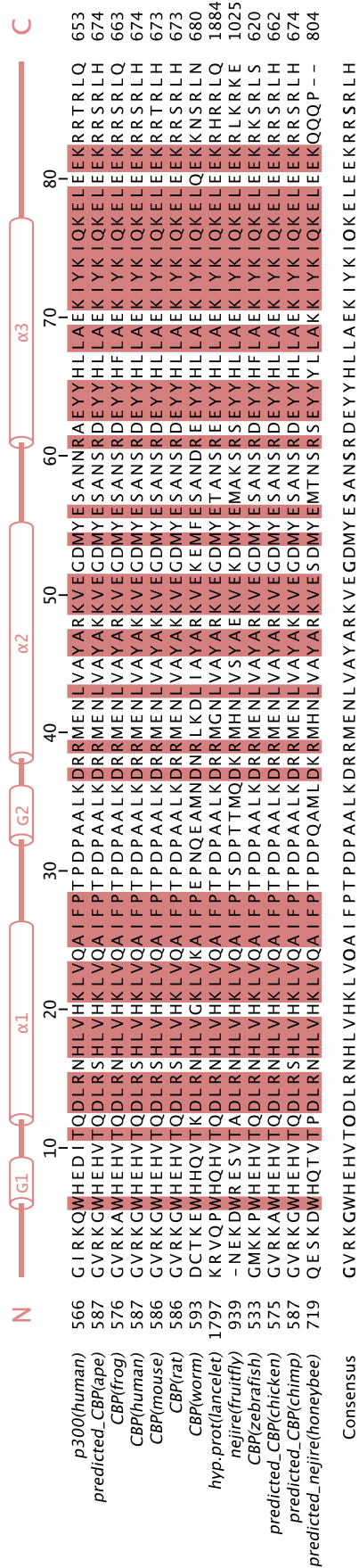


Figure 1.6: Alignment of the KIX domains of various organisms: *Homo sapiens sapiens* (human), *Pan troglodytes* (chimp), *Macaca mulatta* (ape), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Xenopus laevis* (frog), *Caenorhabditis elegans* (roundworm), *Danio rerio* (zebrafish), *Apis mellifera* (honeybee) *Gallus gallus* (chicken), *Drosophila melanogaster* (fruitfly), *Brachyostoma floridae*(lancelet). Predicted and hypothetical proteins (hyp.prot) are marked accordingly. Sites of the primary sequences that are equal in all organisms are highlighted in blue. Sequences were identified using the BLAST search tool [22] for protein sequences in the NCBI database. Multiple alignments were done by T-Coffee [23]. The graphical presentation was compiled by Jalview [24].

## 1.3 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance, NMR for short, was developed in 1946 by F. Bloch and E. M. Purcell. The technique is based upon the interaction between electromagnetic radio waves and atomic nuclei residing in a strong magnetic field. Nuclear magnetic resonance spectroscopy has become a convenient tool of protein analysis in the recent years and is thereby not limited to structure determination, but is capable of measuring protein dynamics as well. From structure follows function: an idea that is only half the truth as became apparent in the recent years of research. Dynamics accompanies function, therefore rendering NMR as an indispensable tool for characterising protein function and regulation. Before going into the complex details of nuclear magnetic resonance spectroscopy, the attributes of matter at molecular and atomic scale have to be elaborated in the next chapter.

### 1.3.1 Introduction to quantum mechanics

A basic knowledge of the assembly of matter as well as modern quantum mechanics is needed in order to understand the principles of NMR. During time, scientists found out more and more about the constitution of matter, thereby discovering molecules, atoms and even subatomic particles. Physicists first tried to describe the behaviour of these micro-objects using Newton's law of motion, but failed to explain the puzzling results of some experiments.

Absorption of electromagnetic waves is a fundamental occurrence in nature and describes a way in which matter can take up energy. The reverse process is called emission, whereupon energy is released to the environment, again as electromagnetic waves. These waves can be seen as consisting of both electric and magnetic fields that propagate in an oscillating manner. The energy of an electromagnetic wave is dependent on the frequency of oscillation. But how do electromagnetic waves actually arise? Physics says they originate whenever electrically charged particles are in motion. In classical physics, electrons were seen as a one-dimensional harmonic oscillator radiating at the oscillation frequency. A harmonic oscillator is a system swinging or vibrating in a periodical manner following a sine or cosine function. Pursuing this view, electrons were thought to resemble negatively charged particles in motion and hence emitting a propagating electromagnetic wave.

A quantum harmonic oscillator, however, behaves a bit different and was adapted in order to fit experimental results that could not be explained by classical physics. Max Planck was the first to introduce discrete energy quanta in his equation to describe black body radiation. He stated that electrons can only absorb energy a multiple of  $h \cdot \nu$ , where  $h$  describes a constant and  $\nu$  the frequency of oscillation. This resulted in the awareness that a quantum harmonic oscillator, in contrast to its classical counterpart, cannot take up infinitesimal amounts of energy. The same is true for many other atomic properties that acquire merely distinct states.

Equation 1.1 summarizes the relation between energy and the frequency given in  $\text{rad s}^{-1}$ , where  $\hbar$  equals  $\frac{h}{2\pi}$ . When an atom absorbs a discrete energy quantum, one of its states is said to be excited to a higher energy level. Absorption can thereby result from electrons or the atomic nucleus. When an electron spontaneously falls back to its ground state it results in emission of a photon. Figure 1.7 points out two energy levels, where  $E_\alpha$  denotes the lower and  $E_\beta$  the upper energy level. Excitation occurs only when the energy of the photon matches the energy difference of the upper and lower level, given by  $\Delta E = E_\beta - E_\alpha$ . Depending on the nature of the observed system, the number of energy levels is not limited to two different ones. To describe the effects of nuclear properties important in biochemical NMR it is, in most cases, nevertheless sufficient.

$$E = \hbar \cdot \nu \tag{1.1}$$

In classical spectroscopic techniques, absorption will weaken an irradiated light beam and hence result in a line in the absorption spectrum. Spectral lines of UV-VIS or infrared spectroscopy are adequately described by the idea of molecules existing in different energy levels. However, this model is not sufficient for NMR. Modern quantum mechanics instructs to reconsider the perception of the world aloof the daily life and think in discrete states. One of the most important and maybe even more puzzling postulate is that '[...] a measurement of energy yields a value which corresponds to one of the energy levels available to the system' [25]. In other words, a measurement of the energy of a molecule will always result in a value of one of the energy levels. This, however, does not necessarily mean that the system needs to be in one of the available energy levels.

In quantum mechanics, the wave function of a particle fully describes all of its properties. The energy levels are each associated with a corresponding wave function, but again this does not mean that the wave function of the particle has to be the one associated with its measured energy level. During a measurement of a quantum mechanical object that is described by a wave function it is transformed in a state corresponding to one of its energy levels. Before the measurement it is rather a mixture of the wave functions corresponding to each energy level. This condition is commonly referred to as mixed- or superposition state. Superposition states arise from molecular vibrations and collisions and have very short lifetimes that cannot be grasped by infrared or UV-VIS spectroscopy. NMR however is capable of manipulating and exploiting these mixed states.

So far, energy quantisation of matter that will contribute to further understanding of principles of NMR spectroscopy has been elaborated. Back in the year 1922, Stern and Gerlach observed quantisation of the nuclear spin of atoms. Besides mass and charge, the nuclear spin is an intrinsic property of atomic nuclei. A spin is an observed effect in the nature of matter that has no equivalent counterpart in the dimension of humans. Even if there are no spinning movements of atomic nuclei, it can sometimes be useful to think of the spin property as a



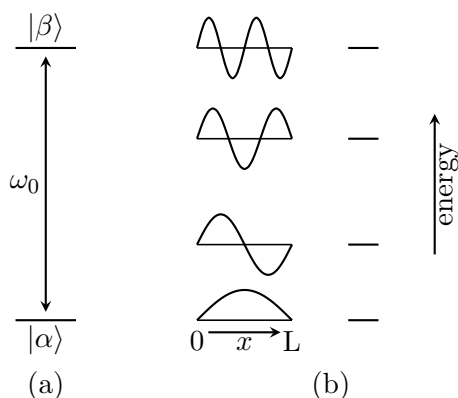


Figure 1.7: (a) Energy diagram showing the transition of the lower to the upper level and the corresponding energy. (b) Energy levels (right) and their corresponding wave functions (left). A particle is always described as a mixture of these wave functions. Illustration based upon Keeler, *Understanding NMR Spectroscopy* (2005)

rotational movement. However, some other times this point of view is totally wrong and fails to explain the spin behaviour. In an atomic nucleus, protons and accordingly neutrons align to pairs of antiparallel spins. Observations show that a coupled spin pair annihilates its magnetic properties. The nuclear spin quantum number is based on the observation of physical effects. Depending on the composition of an atomic nucleus, it can adopt positive integral or half-integral numbers. Some nuclei, including  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  exhibit nuclear spin properties, they are called spin-active nuclei, because their nuclei comprise unequal numbers of protons. Atoms without nuclear spins are invisible to NMR as they show no interactions with magnetic fields. Spin-active carbon and nitrogen isotopes have mean occurrence in nature and therefore have to be artificially enriched in the sample of interest.

The nuclear spin gives rise to a magnetic moment that interacts with magnetic fields, a physical effect whereupon NMR spectroscopy is built. Atoms with an even number of protons and neutrons are therefore invisible to NMR spectroscopy as their nuclear spin is zero and no magnetic moment is present. Having a nonzero nuclear spin gives rise to a magnetic moment, thus the atom is then comparable to a tiny bar magnet that interacts with magnetic fields. As a vectorial property, the magnetic moment describes the strength of the atomic dipole and has both a magnitude and direction. The magnetic moment  $\mu_m$  is given by

$$\mu_m = \gamma p \tag{1.2}$$

where  $p$  denotes the angular momentum and  $\gamma$  the magnetogyric ratio. On the other hand, the angular momentum is associated with the nuclear spin quantum number  $I$  and the Planck constant  $h$  as equation 1.3 points out.

$$p = h \cdot \sqrt{I(I + 1)} \tag{1.3}$$

While not influenced by a magnetic field, the magnetic moments emerging from the nuclear spin point in all kinds of directions. Any direction of the magnetic moments is equally favoured, thus they will point in random directions and therefore cancel out one another. Even a very small sample contains such a vast number of atoms that the magnetic moments take the mean and therefore no macroscopic magnetisation is observed. Interaction of nuclei with a strong, magnetic field aligns the magnetic moments of these atoms along the axis of the applied field. In some way it is allowed to think of these nuclei as little bar magnets. Introductions to quantum theories usually tell that the magnetic moments of spin one-half nuclei become polarised in a magnetic field, adopting either a parallel ( $\alpha$ ) or antiparallel ( $\beta$ ) state, with the eigenvalues  $+\frac{1}{2}$  and  $-\frac{1}{2}$  representing their energies. In fact, quantum mechanics say that the atoms exist in a mixed or superposition state, where measurement of the z-component will always yield the same value. The state of the nucleus for the x and y-components is, however, not defined. However, the impact of a static magnetic field to spin-active nuclei causes the magnetic moments to align along the axis of the magnetic field and the energy of spin orientation to split up into  $2I+1$  states according to the Zeeman effect, where  $I$  represents the nuclear spin quantum number. The number of possible spin orientations is an intrinsic property of a nucleus and depends on its composition. In a static magnetic field there are just a few possible nuclear spin orientations that are unequally favourable, so to speak they appear in different energy levels. Only a very small part of the magnetic moments align with the external field, the others are randomized by thermal motion and collisions. Adding up the magnetic moments of all properly aligned orientations results in a macroscopic or bulk magnetisation of the sample that is aligned parallel to the external magnetic field.

### 1.3.2 Chemical shift

The preceding subsection provided the basics of quantum mechanics, now the consecutive one will deal with important NMR parameters and models to further understand the methods and experiments that have been part of this thesis. The chemical shift is the unit of measurement in NMR spectra having the dimension of parts per million (ppm). Resonance energies are usually expressed in terms of wavelengths or frequencies, however chemical shifts have been introduced due to the fact that the differences in absolute frequencies are very small in NMR spectroscopy. The chemical shift of a nucleus is the difference between the absolute resonance frequency of  $\omega_{0,i}$  and a reference compound  $\omega_{ref}$  divided by  $\omega_{ref}$ . Multiplication by a factor of  $10^6$  is applied to scale the chemical shifts to manageable dimensions, as equation 1.4 illustrates.

$$\delta = \frac{\omega_0 - \omega_{ref}}{\omega_{ref}} \times 10^6 \quad (1.4)$$

Absorption spectra show that chemically different nuclei have different resonance frequencies. Shielding effects provide the basis for this observation. The external magnetic field  $B_0$  induces

a current in the electron cloud of an atom, thereby generating a local magnetic field. The effective magnetic field  $B_{eff}$  affecting the nucleus is the sum of the external and local magnetic field. Due to the difference in the local magnetic field of chemically non-equivalent nuclei, their resonances appear as separated signals in the absorption spectrum. Looking at the facts, it comes to no surprise that chemical shifts are sensitive reporters of molecule structure. Every change in structural shape of a protein creates a new chemical environment and is accompanied by changes in the electron density.

$$\omega_0 = -\gamma B_{eff} \quad (1.5)$$

The elements of the protein backbone show nearly identical resonances in unfolded proteins as the random coil structure is flexible. Formation of secondary structures show an upfield or downfield shift of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei relative to random coil resonances. Site specific information can be gained by assigning the resonances to the residues of a protein.

### 1.3.3 The vector model

The signals recorded by a nuclear magnetic resonance spectrometer are often represented using the vector model that vividly describes what happens to the ensemble of nuclear spins. The vector model is used to exactly describe a rather limited set of pulse sequences concerning uncoupled spins. It fails however to explain more complex experiments, including even basic two-dimensional spectra, where information about coupled spins is obtained.

The vector model represents the arrangement of the NMR device in a three dimensional coordinate system. The external magnetic field generated by a superconducting magnet goes along the z-axis. The sample giving rise to the spin magnetisations is centred at the origin, whereas a coil that is placed around the sample is used for detecting signals and producing a radiofrequency field in the x, y- or transverse plane. The equilibrium- or bulk-magnetisation is represented by a vector pointing along the direction of the applied field (figure 1.14(a)). For the equilibrium magnetisation caused by the static, external magnetic field  $B_0$  the vector starts at the origin and points to +z. The direction of the equilibrium magnetisation appears to be stationary, however the magnetic moments of individual spins execute precessional motions. The magnetisation vector is tilted away from the z-axis and executes a precessional movement, similar to that of a gyroscope. It is because of the ensemble of spins precessing out-of-phase that the net magnetisation does not move. Random distribution of the tilted magnetisation vectors cancel out their contribution to transverse magnetisation. The speed of this rotation is referred to as Larmor frequency and its dependency resolved in equation 1.5, where  $B_{eff}$  is the effective magnetic field present at a nucleus.

The external magnetic field  $B_0$  decreases through shielding effects of the electron cloud of a nucleus. The chemical environment differs from one nucleus to the other, therefore even slight

changes in protein structure can result in detectable changes in Larmor precession. To further simplify the model, a little trick is used to make the irradiated RF field appear static. A rotary co-ordinate system is introduced, where the co-ordinate system is not static, but rotates with a carrier frequency  $\omega_{rot.frame}$ . If the carrier frequency matches exactly the Larmor frequency, then a transverse magnetisation will appear to be static. Given an offset  $\Omega = \omega_0 - \omega_{rot.frame}$ , the transverse magnetisation will rotate either clock- or counter-clockwise in the x,y-plane.

Due to magnetic field inhomogeneity or individual Larmor frequencies, the magnetisation vectors are randomly distributed in the transverse plane and cancel out each other. If a magnetic field oscillating at the Larmor frequency is applied in the transverse plane, the resonance condition is met and the net magnetisation starts to rotate around the axis of irradiation. Using this principle, the magnetisation vector can be conducted to any position.  $90^\circ$  and  $180^\circ$  pulses are most commonly used in NMR spectroscopy. A  $90^\circ$  pulse deflects the equilibrium magnetisation into the transverse layer, whereas a  $180^\circ$  pulse mirrors the magnetisation with respect to the plane given by the z-axis and the axis of irradiation (see figure 1.14(e)). Assuming the simple example of one uncoupled spin, whose magnetisation has been deflected to the transverse plane using a  $90^\circ$  pulse. As stated previously, the magnetic moments precess around the z-axis, therefore giving rise to a rotational movement of the net magnetisation in the transverse plane. Such an oscillating magnetisation then again induces an electric current in the coil giving rise to the NMR signal, commonly referred to as free induction decay (FID).

For the above example, the FID will oscillate at the Larmor frequency of the spin. In a system where more than one spin exists, the FID is a result of superimposed oscillating signals. Subjecting the FID to Fourier transform will convert the time domain, the detection, into a frequency domain. It thereby does not matter of how many different signals the FID is composed. The resulting frequency spectrum will show peaks of the Larmor precession. Detecting the signal in the transverse layer results in an oscillating signal, where the x- and y-components are given by:

$$S_x = S_0 \cos \Omega t \quad S_y = \pm S_0 \sin \Omega t \quad (1.6)$$

By assuming that the signal decays exponentially the model extends to:

$$S_x = S_0 \cos \Omega t e^{-\frac{t}{T_2}} \quad S_y = \pm S_0 \sin \Omega t e^{-\frac{t}{T_2}} \quad (1.7)$$

Both components differ from each other only by a phase shift of  $90^\circ$ . Rotational motions are usually expressed in terms of complex numbers, with the x- and y-component being the real and imaginary part, respectively. [25]. However, due to relaxation processes the FID decays with increasing time and the system approaches the equilibrium state. The following equation

$$S(t) = S_x + iS_y = S_0 \times e^{\pm i\Omega t} \underbrace{e^{-\frac{t}{T_2}}}_{\text{damping}} \times \underbrace{e^{i\phi}}_{\text{phase}} \times \underbrace{e^{i\phi_{corr}}}_{\text{correction}} \quad (1.8)$$

describes the observed signal over time  $S(t)$ . The model also includes a damping term describing relaxation processes and a term for phase correction. NMR spectrometers produce time domain data with an arbitrary phase [25]. By trial and error,  $\phi_{corr}$  can be chosen to be  $-\phi$ , annihilating the correction term and yielding a spectrum with peaks of absorption mode Lorentzian lineshape.

### 1.3.4 Multi-dimensional NMR spectroscopy

Scalar coupling or J-coupling describes the interactions between covalently bound atoms in NMR spectroscopy. J-couplings lead to a splitting of the resonance signal of chemically non-equivalent nuclei in multiplets. The effect operates across a maximal distance of three covalent bonds. A spin one-half nucleus has two states of spin energy:  $+\frac{1}{2}$  and  $-\frac{1}{2}$ . A nucleus I that is attached to another spin one-half nucleus S, is confronting either a  $+1/2$  or  $-1/2$  spin state of S. The energy discrepancy of the states of S leads to an upfield or downfield shift of the nucleus I resonance depending on the spin state of S. Observations in an ensemble of nuclei I and S show that the signal of nucleus I splits into two lines symmetrically about the chemical shift, as both states are equally probable. The distance between the two peak maxima describes the coupling constant  $J_{12}$ . The J-couplings are usually expressed in terms of  $s^{-1}$  or Hertz (Hz), as they happen to be independent of temperature and the strength of the external magnetic field  $B_0$ .

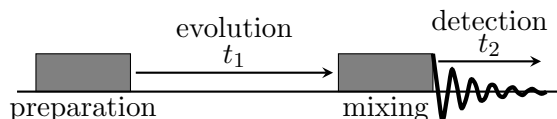


Figure 1.8: *General pulse sequence in two-dimensional NMR spectra.*

In a one-dimensional spectrum the intensity is plotted against the frequency. The plot of a two-dimensional spectrum has two frequency axes where the size of the spot representing the peak is equivalent to its intensity. The most applied experiments designed to tackle challenges in protein structure or dynamics are the ones, where each peak shows a correlation between small numbers of quantities. Figure 1.8 shows the basic pulse sequence of a 2D spectrum. Starting with the preparation time, the equilibrium magnetisation is usually transferred from hydrogen to another heteronucleus in question, e.g.  $^{13}\text{C}$  or  $^{15}\text{N}$ . Hydrogen atoms are the most sensitive elements in NMR spectroscopy due to the magnitude of their magnetogyric ratio and hence equilibrium magnetisation and signal intensity. It is therefore common to excite hydrogen atoms and transfer their magnetisation to another nucleus using the INEPT style pulse sequence. The equilibrium magnetisation is furthermore transformed into some kind of coherence that evolves during time  $t_1$ . It is important to note that during  $t_1$  there is no observation of the evolving signal. In fact, the signal is recorded after the mixing time, where the coherence is manipulated

into an observable signal. This feature of indirectly following unobservable signals like multiple-quantum coherence is an essential feature of modern NMR spectroscopy. Moreover, it is the mixing period that determines the content of the spectrum. Data acquisition is followed after the end of the mixing period for the time  $t_2$  at a regular sampling interval. Fourier transform converts the time-domain function  $S(t)$  into a frequency-domain function  $S(\omega)$ , thus yielding a one-dimensional spectrum.

The important question that should arise now is how a second frequency-domain function is introduced. It turns out to be rather simple. The same pulse sequence is repeated each time with a different evolution time, therefore generating a second time-domain function of the signal. In the first pulse sequence, the evolution period is skipped and the FID recorded during detection time  $t_2$ . In subsequent iterations, the evolution time is set to  $\Delta_1$ ,  $2\Delta_1$ ,  $3\Delta_1$  and so on. The data recorded in such a way can be represented as a matrix  $S(t_1, t_2)$  with  $t_2$  forming the rows and  $t_1$  increments the columns. To manipulate the data into a spectrum, first, the rows are subjected to Fourier transform. Afterwards, each data point in  $t_2$  is subjected to Fourier transform along the column or  $t_1$  increments resulting in a matrix with two frequency-domain functions, or the spectrum  $S(\omega_1, \omega_2)$ . [25] More than one evolution period in a pulse sequence leads to multiple resonance spectra.

### 1.3.5 Assignment strategies

Chemical shift resonances of proteins have to be assigned to amino acids after the spectrum is recorded. There exist a wide variety of experiments to establish a complete assignment of uniform  $^{13}\text{C}$  and  $^{15}\text{N}$  labelled proteins from scratch. A summary of experiments is shown in table 1.1. Assignment of the KIX backbone resonances was already done by Tollinger and Brüschweiler.

Experiment	Observed nuclei
HNCO	H(i), N(i), C'(i - 1)
HNCA	H(i), N(i), C $_{\alpha}$ (i), C $_{\alpha}$ (i - 1)
HN(CO)CA	H(i), N(i), C $_{\alpha}$ (i - 1)
HN(CA)CO	H(i), N(i), C'(i)
CBCA(CO)NH	H(i), N(i), C $_{\alpha}$ (i - 1)
HBHA(CO)NH	H(i), N(i), H $_{\alpha}$ (i - 1)
CBCANH, HNCACB	H(i), N(i), C $_{\alpha}$ (i), C $_{\beta}$ (i), C $_{\alpha}$ (i - 1), C $_{\beta}$ (i - 1)
(H)CC(CO)NH-TOCSY	H(i), N(i), C $^{\text{aliph.}}$ (i - 1)
H(CC)(CO)NH-TOCSY	H(i), N(i), H $^{\text{aliph.}}$ (i - 1)
HCCH-TOCSY	H $^{\text{aliph.}}$ , C $^{\text{aliph.}}$

Table 1.1: Experiments providing information for NMR resonance assignment. Table taken from Sattler et al. (1999) [26]

Mutant variants of KIX that are structurally similar to the wildtype form show merely small resonance shifts in double correlated  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectra. Most resonances can be correlated by comparison without the need of more complex experiments that need  $^{13}\text{C}$  labelling. Remaining peaks can be assigned by using experiments that provide spacial information of adjacent residues, e.g. the  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{15}\text{N}$ -NOESY experiment. In the next subsection it will become clear that the HSQC provides the basis for Carr, Purcell, Meiboom and Gill (CPMG) relaxation dispersion and  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{15}\text{N}$ -NOESY experiments.

### Heteronuclear single quantum coherence (HSQC)

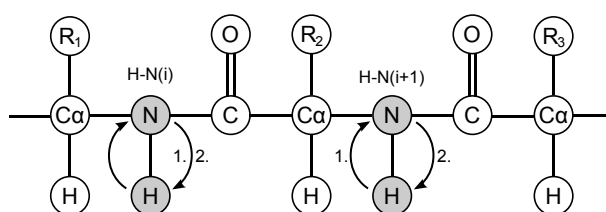


Figure 1.9: Schematic of magnetisation transfer in a  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC experiment. (i) Initial transverse magnetisation is transferred to nitrogen nuclei. (ii) Following an evolution time  $t_1$ , the magnetisation is transferred back to hydrogen, where it is manipulated into an observable signal. A spectrum showing double resonance correlation between hydrogen and nitrogen atoms is obtained in this way. Figure based upon *Protein NMR - A Practical Guide*, <http://www.protein-nmr.org.uk>

Organic life-forms mainly comprise three elements: hydrogen, carbon and nitrogen. Natural abundance of NMR-active  $^{13}\text{C}$  carbon and  $^{15}\text{N}$  nitrogen isotopes is very low, thus enrichment of these elements is a prerequisite for NMR analysis of proteins. A 2D experiment that is commonly used to study biomolecules is the double resonance  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single quantum coherence experiment. Each peak of the spectrum shows a one-bond correlation between the observed nuclei. Nitrogen atoms are repeatedly present in protein backbone amide groups and some amino acid side chains. Secondary structures like  $\alpha$ -helices or  $\beta$ -sheets result in an upfield or downfield shift, as was elaborated in section 1.3.2. Chemical shifts are sensitive reporters of local structure, therefore rendering an HSQC spectrum well suited for detecting structural alterations. It should be noted at this point that the amino acid proline does not show any signal in a  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC as its nitrogen atom is attached to the side chain and not bonded to a hydrogen atom.

### Nuclear Overhauser enhancement spectroscopy (NOESY)

Nuclear Overhauser enhancements (NOEs) arise through coupling of the magnetic moments of nuclei in close spacial proximity. The important difference in contrast to experiments based upon J-couplings is that the interactions of nuclei are mediated through space and not covalent bonds.

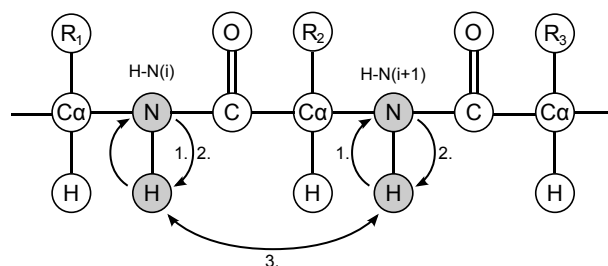


Figure 1.10: Schematic of magnetisation transfer in a  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY experiment. (i+ii) Initially generated transverse magnetisation is transferred from hydrogen to nitrogen nuclei, where, after an evolution time  $t_{1,1}$ , the magnetisation is transferred back to hydrogen. (iii) Magnetisation is exchanged through space via dipolar couplings between hydrogen nuclei of amide groups. (iv-v) Like in the initial phase, magnetisation is transferred to nitrogen and after an evolution time  $t_{1,2}$  transferred back to hydrogen. The signal is manipulated into an observable signal and data recorded. Figure based upon *Protein NMR - A Practical Guide*, <http://www.protein-nmr.org.uk>

NOEs are effective over a maximum distance of  $5 \text{ \AA}$  with the signal intensity being proportional to the space between coupling nuclei. Therefore, they can provide useful information about the conformation of biomolecules by generating structural restraints. The  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY experiment is basically a  $^1\text{H},^{15}\text{N}$ -HSQC extended in a third dimension that provides information about closely related hydrogen atoms. The pulse sequence of the  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY experiment consists of an HSQC-like sequence, followed by a mixing time that allows dipolar couplings to evolve and is finished by executing an HSQC-like sequence again. Each HSQC-like sequence implies an evolution time that is incremented to finally get a three-dimensional matrix  $S(t_{1,1}, t_{1,2}, t_2)$ . Fourier transform yields a spectrum with three frequency axes  $S(\omega_1, \omega_2, \omega_3)$ .

Assignment strategies and the underlying experiments were discussed at the beginning of section 1.3.5. Determination of protein backbone amide NOEs can be resolved in the  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY experiment. Again, cross- and diagonal peaks have to be discriminated. In a three dimensional  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY spectra the requirement for a cross peak to appear is the spacial proximity of amide group protons to each other. The distance from a backbone amide group to next in random coil conformation is sufficiently small to give NOE signal. In helical secondary structures, amide groups of adjacent residues come in close proximity to yield short range NOEs with stronger signals. Provided with a partly assigned HSQC spectra as well as the primary amino acid sequence of the protein in question, a  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY experiment is sufficient to determine residues adjacent of already assigned ones. Typical geometry of an  $\alpha$ -helix is shown in figure 1.11, in which an excerpt of KIX  $\alpha$ -helix is pictured. From these distances one can predict that amide hydrogen atoms in helically folded proteins will show intermediate to strong NOE signals.

Three parts of a  $^1\text{H},^{15}\text{N},^{15}\text{N}$ -NOESY spectrum are shown in figure 1.12. Starting with the excerpt in the middle of the figure (H-N(i)), three distinct resonances can be discriminated. The strongest signal represents the diagonal peak that is accompanied by two weaker signals, the



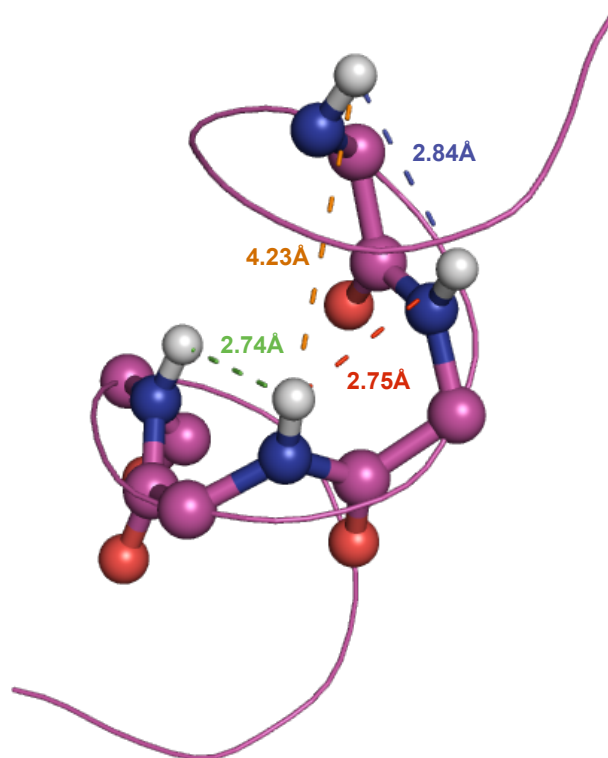


Figure 1.11: *Geometry of an excerpt of KIX  $\alpha$ -helix. Distances between amide hydrogens of His602, Leu603, Val604 and His605 are shown. Strong NOE signals of amide hydrogen atoms are observed from adjacent residues. The rather long distance between residue  $i$  and  $i + 2$  merely yields weak NOEs that were mostly overwhelmed by noise in observed  $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY spectra. Carbon atoms are coloured in pink, nitrogen in blue, hydrogen in white and oxygen atoms are stained in red. Solution structure by Radhakrishnan et al. (1997) [19] and the picture rendered by PyMol [21].*

cross-peaks. Cross-peaks appear above or below the diagonal peak, all at the same proton resonance frequency. Arrows indicate peaks with the same chemical shifts in the nitrogen dimension. Adjacent residues  $i - 1$  and  $i + 1$  have diagonal peaks at the nitrogen frequency of the cross-peaks of residue  $i$ . This symmetrical spectrum comes to no surprise as the magnetisation is transferred from one residue to the other and vice versa. The last question to resolve is how the diagonal peaks are assigned. To understand this it is obvious to understand how the  $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY 3D spectrum is structured. A series of 2D spectra are beaded along an axis that is called the z-dimension, whereas the spectra themselves span the axes x and y. Axes y, z correspond to nitrogen and axis x to the hydrogen frequency dimension. Collapsing the y-dimension that provides the NOE information will result in a  $^1\text{H}, ^{15}\text{N}$  spectrum occupying axes x and z. From this it follows that the nitrogen co-ordinate of a peak in an HSQC spectrum can be transformed into the z-dimension to locate the corresponding diagonal peak. Note that the z-dimension is made up of a series of 2D spectra, thus the HSQC peak co-ordinates are used to identify the correct spectrum. In that way, the sequence can be migrated and assigned step by step.

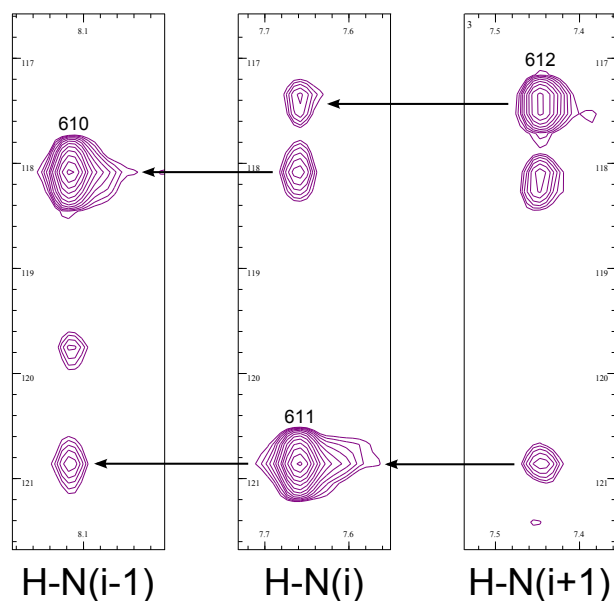


Figure 1.12: *Sequential walking on a  $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY spectrum: Starting at an arbitrary residue  $i$  there appear a diagonal- and a number of cross-peaks. Amide groups lying in close spacial proximity of maximal  $5 \text{ \AA}$  result in symmetric cross- and diagonal peaks between adjacent residues, as the arrows point out.*

### 1.3.6 Relaxation dispersion experiments

The functional properties of protein regulation or catalysed biochemical reactions are very often intimately paired to rearrangements in structure. Conformational changes can regulate protein activity by extruding side chains of catalytically important residues from the active centre, open up or close substrate entry gates, fine-tune molecular recognition processes by order-to-disorder transitions or transmit information between remote sites. [27] All these functions emphasize the importance of protein dynamics in biological processes. Dynamics of biomolecules in solution span a large region of times lying between  $10^{-12}$  and  $10^4$  s. [28] NMR spectroscopy provides a unique tool for investigating protein structure and dynamics that span many time scales at atomic resolution. Characterisation of fast molecular motions that range from  $10^{-12}$  to  $10^{-9}$  is tackled by  $T_1$ ,  $T_2$  and NOE measurements. Slow motions can be studied by  $T_{1\rho}$  experiments. [28] Chemical exchange rates of proteins usually lie in the milli- to microsecond time domain, leading to broadened lines with weaker intensities of exchange-correlated sites. However, conformers are not always populated to an appreciable degree that allows detection with traditional structural approaches [27]. One way to overcome these limitations is the Carr, Purcell, Meiboom and Gill (CPMG) relaxation dispersion experiment that allows for quantitatively analysing protein dynamics in the millisecond time scale even of highly skewed populations [18]. Measurement of exchange contribution to the transverse relaxation rate  $R_2$  provides the bases for the experiment. The CPMG method will be further elaborated in the following subsection.

In the aforementioned case of a protein that can exist in two distinct conformational states,

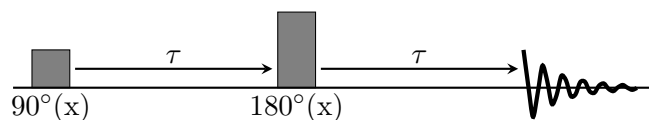


Figure 1.13: *Spin echo pulse sequence.*

$A \xrightleftharpoons[k_{-1}]{k_1} B$ , each state is likely to have a distinct chemical shift, because they reside in different chemical and magnetic environments. Each conformer gives rise to a peak at position  $\Omega_A$  and  $\Omega_B$ , separated by the chemical shift difference  $\Delta\omega_{A,B}$ , with their intensities proportional to the relative populations  $p_A$  and  $p_B$ , respectively. The exchange rate  $k_{\text{ex}}$  is given by  $k_{\text{ex}} = k_1 + k_{-1}$ . Considering the case that the exchange rate is much less than  $2\pi\Delta\omega$ , then separate peaks for the two distinct states are observed as long as the lower populated state is in order of a few percent. [27] For the case of intermediate exchange, where  $\Delta\omega_{A,B}$  is in the range of  $2\pi\Delta\omega$ , line broadening is observed. Contribution to linebroadening can be allocated to three parameters  $k_{\text{ex}}$ ,  $\Delta\omega$  and  $p_b$ . Relaxation dispersion experiments allow for measuring contribution of chemical exchange to the transverse magnetisation.

This experiment is built upon an effect first detected by Erwin Hahn in the 1950, called a spin echo. The pulse sequence producing a spin echo starts with a  $90^\circ(x)$  pulse followed by a  $180^\circ(x)$  pulse separated by time  $\tau$  (see figure 1.13). Figure 1.14 shows the behaviour of the spin magnetisation subjected to a spin echo pulse sequence. The first pulse generates transverse magnetisation. It is convenient for nuclear magnetic resonance studies to describe the spin precession in a rotating co-ordinate system, spinning with  $\omega_{\text{rf}}$ , that will eliminate the impact of the oscillating RF-pulse. Due to magnetic field inhomogeneity, the effective field strength acting on a spin and its related Larmor precession are different throughout the sample cell (see equation 1.5). After some time elapses, the magnetisation vectors fan out. Some spins will rotate faster and some slower than the carrier frequency of the rotary co-ordinate frame. Subsequently, the  $180^\circ(x)$  or inversion pulse rotates the magnetisations around the x-axis, thus mirroring them with respect to the x, z-plane. It causes the signal to rephase and producing a spin echo showing its maximum after a time of  $2\tau$ .

The CPMG experiment uses a variant of the spin echo pulse sequence. After the initial  $90^\circ(x)$  pulse and the delay  $\tau$ , the  $180^\circ(x)$  pulse followed by another delay  $\tau$  are repeated n-times. The rotation frequency of a spin around the external magnetic field represents a resonance transition between energy levels and its chemical shift. The spin echo sequence compensates for magnetic field inhomogeneity and prevents that precessing spin vectors fan out. After a time interval of  $2\tau$  the spins are perfectly refocused. Residual fanning out will result solely from the exchange process and of course the machine's imperfection.

Regarding the case where the chemical shift difference between two states A and B is zero, so to say they precess with the same resonance frequency, the  $180^\circ$  pulse compensates for the magnetic field inhomogeneity. The spins will be perfectly refocused after a time interval  $\tau$ . If

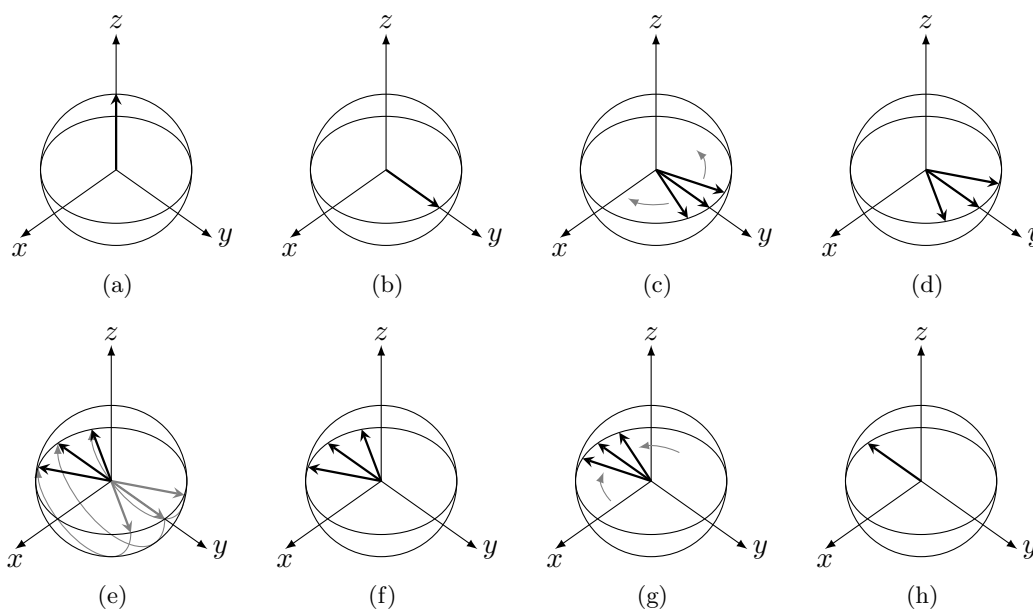


Figure 1.14: *Schematic representation of the spin echo effect. (a) The equilibrium magnetisation aligns along the external magnetic field,  $B_0$ . (b) A  $90^\circ(x)$  pulse rotates the magnetisation around the  $x$ -axis. (c, d) After some time, the chemical shifts come out of phase due to magnetic field inhomogeneity. (e) A  $180^\circ(x)$  pulse inverts the magnetisations, so to say they get mirrored with respect to  $x, z$ -plane. Dashed lines represent the situation before the refocusing pulse. (f, g) The magnetisation further precess in the same direction. Slow and fast precessing spins catch up after a given amount of time. (h) The spin vectors reach  $-y$  simultaneously after a time of  $2\tau$ , that is double the time between the initial  $90^\circ(x)$  and  $180^\circ(x)$  pulse. The signal will decay due to transverse relaxation processes.*

no exchange is present, all spins will be perfectly refocused, so to say they will reach the  $-y$  axis simultaneously in the given co-ordinate system (see figure 1.14) after a time period of  $2\tau$ . In a plot where the number of spins crossing the line is plotted against time, a sharp line will be observed. However, in the presence of a conversion between fast and slow precessing spins, this line broadens with the magnitude of the conversion rate. From the peak breath the following information can be extracted: the relative number of fast and slow spins ( $p_A$  and  $p_B$ ), the rate of interconversion ( $k_{\text{ex}} = k_{-1} + k_1$ ) and the chemical shift difference ( $\Delta\omega_{\text{disp}} = \omega_B - \omega_A$ ) between exchanging conformers [27].

A  $180^\circ(x)$  pulse rotates the spin vectors around the  $x$ -axis, whereupon a slow precessing spin will be ahead of a fast precessing one. During the delay between  $180^\circ(x)$  pulses the spins will refocus. Stochastic interconversion between states of different chemical shifts, so to say different precession frequencies, will however disturb the process of refocusing. In the example mentioned above, the fast spin will catch up with slow one after a time of  $2\tau$ . If however the slow precessing spin converts to a fast, the distance between the spins will stay the same until either one of fast spins changes back to the state giving rise to the slow precessing vector. Playing through this example, not all of the spins will be coherent after a time of  $2\tau$ . The larger the spins precess with

different chemical shifts (Lamor frequency can be interconverted to chemical shifts), the broader the peak will be. The less time the spin magnetisations precess against each other unaffected, the less is the probability of an exchange process. Hence, the effect of an exchange process will accumulate to a lesser extent and is minimised with an increasing number of refocusing pulses. In other words, shorter intervals between inversion pulses decrease the evolution time of the exchange process and thus sharpens the peak. The number of inversion pulses in a fixed time interval is referred to as CPMG field strength and given by  $s^{-1}$  or Hz.

An actual relaxation dispersion experiment comprises a series of  $^1\text{H}, ^{15}\text{N}$ -HSQC experiments recorded at variable CPMG field strengths. A factor proportional to the peak breath  $R_2$  is plotted against the CPMG field strength. As mentioned before, linebroadening is observed for residues showing intermediate exchange rates ( $\mu\text{s}$  -  $\text{ms}$ ). A typical dispersion curve of intermediate exchanging residues has an exponential shape. Fit of the data to Carver-Richards equation extracts the aforementioned values for the relative populations  $p_A$  and  $p_B$ ,  $k_{\text{ex}}$  and  $\Delta\omega_{\text{disp}}$ .

## 1.4 Isothermal titration calorimetry

Calorimetry deals with measuring heat changes of a system. All chemical reactions, dilutions or physical changes are accompanied by a change in heat or enthalpy. Environmental alterations consuming energy are referred to as endothermic processes while a reaction that releases heat is called an exothermic process. A calorimeter can therefore be seen as an universal detector. Calorimetric experiments have proven to be useful in measuring interactions of biomolecules, for example affinities of protein-ligand or protein-protein complexes. A typical biochemical reaction is shown in the following equation:



The mode of operation of an isothermal titration calorimeter is actually simple in contrast to its technical realisation. A power compensation calorimeter consists of a measurement and a reference cell residing in an adiabatic shield and is equipped with sensitive circuits to detect temperature differences between the cells. The measurement cell is controlled at a constant temperature (isothermal), by applying constant cooling and using a temperature controller and heater to keep the temperature. A feedback mechanism controls the power supply of the measurement cell, whereas constant power is applied to the reference cell. An exothermic reaction taking place inside the measurement cell causes the temperature to rise. As a consequence, the power applied to the heater is reduced. The opposite is true for endothermic reactions. It is the power supplied to the heater of the measurement cell as a function of time that represents the raw signal of the detector. [29]

Thermodynamic parameters are calculated from the integrated raw data, plotted against the

molar ratio of receptor and complex, along with a nonlinear regression fit to a pre-estimated model that has to be chosen depending on the number of possible binding sites of the receptor. Under optimal conditions it is possible to determine values for the equilibrium constant  $K$ , the change in enthalpy  $\Delta H$  and entropy  $\Delta S$  and stoichiometry of interaction.

The below-mentioned equations summarise the correlation between these variables:

$$\Delta G^0 = -RT \ln K_{eq} \quad (1.10)$$

$$K_{eq} = \left\{ \frac{[Complex]}{[Receptor] \cdot [Ligand]} \right\}_{equilibrium} \quad (1.11)$$

$$\Delta G = \Delta G^0 + RT \ln \left\{ \frac{[Complex]}{[Receptor] \cdot [Ligand]} \right\}_{actual} \quad (1.12)$$

$$\Delta G = \Delta H - T\Delta S \quad (1.13)$$

Chemical reactions at equilibrium show no net conversion between the reaction compounds, so to say their concentrations stay the same. The equilibrium constant  $K_{eq}$  shows the relative ratio of concentrations between receptor and ligand to the receptor/ligand complex at equilibrium concentrations. The higher  $K$  is, the higher the complex will be concentrated as equation 1.11 points out. A high value of  $K$  implies for a huge binding affinity of the receptor and ligand. In biochemistry it is however far more common to express binding affinities in terms of  $\frac{1}{K}$  with the dimension of  $\text{mol L}^{-1}$ , which would refer to the reaction:



# Methods and Materials

## 2.5 Cell strains

### 2.5.1 DH5 $\alpha$

Like many cloning strains, *E. coli* DH5 $\alpha$  cells feature a number of properties that render them well suitable for preparation of plasmid DNA. The strain carries mutations in various genes, including *endA1*. The product of *endA1* is an endonuclease that will degrade plasmid DNA in many miniprep methods. The gene *hsdR17* codes for a restriction enzyme that will degrade DNA that has not been methylated by the EcoKI restriction-modification system. The lack of these enzymes greatly boosts efficiency and simplicity of cell transformation.

### 2.5.2 BL21(DE3)

The BL21(DE3) *E. coli* strain contains a  $\lambda$ -prophage carrying the T7 RNA polymerase gene rendering it suitable for expression of any gene under the control of the T7 promoter. It furthermore overexpresses the lac repressor (*lacI<sup>q</sup>*) that turns off lac promoter inducible genes more completely. The lac-repressor interacts with a short DNA region called the lac-operator and functions in gene silencing. Expression of a lac-repressor controlled gene is initiated by addition of IPTG, a sugar-analogue that binds to the lac-repressor and inactivates it.

## 2.6 Site-directed mutagenesis

### 2.6.1 Plasmid pHisKIX

The vector that contains the KIX domain was initially cloned by the group of Stacey E. Rutledge in 2003. Their resulting plasmid, pHisKIX, features a functional copy of the KIX domain of the CREB-binding protein, residues 586 to 672, in-frame with a N-terminal hexahistidine tag under the control of a T7 promoter and a lac operator (for additional details see plasmid map in figure A.1).

In order to alter the KIX coding sequence and introduce single amino acids substitution, site-directed mutagenesis was performed. Small artificial primer of about 20 to 50 nucleotides long, that are, with one or two exceptions, complementary to some part of the KIX coding sequence, were synthesised. Even not perfectly matched complementary sequences will form a somewhat distorted double helix. During polymerase chain reaction, the modified construct will be exponentially amplified. Due to the lack of appropriate enzymes in the in-vitro reaction, the resulting DNA is not methylated. The PCR reaction is afterwards treated with the enzyme DpnI, which will degrade residual methylated template vector, thereby strongly decreasing the chance of false-positive transformed cells.

## 2.6.2 Polymerase chain reaction

Site-directed mutagenesis was performed using Stratagene's QuikChange II Site-Directed Mutagenesis Kit or equivalent standard reagents<sup>2</sup>.

Amplification reaction:

5 $\mu$ l	10x Pfu-Ultra HF DNA polymerase reaction buffer
2 $\mu$ l	double stranded DNA template (50 ng $\mu$ l <sup>-1</sup> )
10 $\mu$ l	forward oligonucleotide primer (20 ng $\mu$ l <sup>-1</sup> )
10 $\mu$ l	reverse oligonucleotide primer (20 ng $\mu$ l <sup>-1</sup> )
1 $\mu$ l / 5 $\mu$ l	Stratagene's proprietary dNTP mix / standard 2 mM dNTP mix

The reaction mixture was filled up with distilled water (referring to water always means distilled and deionised water) to a final volume of 50  $\mu$ l. Before starting the polymerase chain reaction (PCR), 1  $\mu$ l of Pfu-Ultra HF DNA polymerase (2.5 U  $\mu$ l<sup>-1</sup>) was added. PCR temperature cycling was carried out as indicated in table 2.2 in order to amplify the whole pHisKIX plasmid. After completion of thermo-cycling, the reactions were cooled down to 4°C to prevent further enzyme activity. Afterwards, the template plasmid was degraded using 10 units DpnI for at least 5 hours or overnight digestion. DpnI will only digest methylated DNA synthesised from dam<sup>+</sup> cell lineages and leave non-methylated DNA produced by the PCR reaction intact. It is therefore essential to use methylated plasmid DNA that was extracted from dam<sup>+</sup> cell lineages like DH5 $\alpha$ .

Segment	Cycles	Temperature	Time
1	1	95°C	2 minutes
2	20	95°C	30 seconds
		55-65°C	1 minute
		68°C	14 minutes

Table 2.2: PCR temperature-cycle scheme for site-directed mutagenesis

---

<sup>2</sup>The protocol is based upon the QuikChange II Site-Directed Mutagenesis Kit instruction manual that has been modified in subtle ways



### 2.6.3 Agarose gel electrophoresis

Amplification may be checked by agarose gel electrophoresis of the DpnI treated PCR product. For the separation of the six kilo bases (kb) pHisKIX plasmid, a 1 % agarose gel (1g agarose in 100 ml TBE-buffer (0.89 M Trizma Base, pH 8.3 in 0.89 M boric acid, 0.02 M ethylenediaminetetraacetic acid (EDTA) and 0.1 % bromphenol blue)) is sufficiently dense. Appropriate amounts of agarose and TBE buffer (1x) were filled into a clean falcon tube and heated in a microwave oven to properly dissolve the agarose. The mixture was cooled down for a few minutes at room temperature. Afterwards 1  $\mu$ l per 10 ml TBE buffer (1x) of the DNA dye SYBR®Safe (10,000x in DMSO) was added. The falcon tube was inverted several times to properly dissolve the dye and poured into a gel caster. When using SYBR®Safe the gel should be stored protected from light, because the compound is very sensitive. After some time the agarose was polymerised and samples can be loaded.

About 5-6  $\mu$ l of the PCR reaction were mixed with the an appropriate amount of DNA sample buffer (20 mM EDTA, 30 % glycerine, 0.5 % SDS and 0.1 % bromphenol blue) and loaded into the slots of the agarose gel along with 5  $\mu$ l of DNA marker. Separation was performed at constant electric field strength of 5 V cm<sup>-1</sup> until the dye of the marker reached the end of gel. The bands were visualised with the aid of ultra-violet (UV) light. In any case, if bands were visible or not, cell transformation was continued.

### 2.6.4 Cell transformation

100  $\mu$ l of competent DH5 $\alpha$  cells are thawed on ice for 15 minutes. Subsequently, 1-3  $\mu$ l of the DpnI treated PCR product are added. The cells are kept on ice for another five to ten minutes, followed by a 90 seconds heat-shock at 42°C. After two more minutes on ice to cool the cells down, 300  $\mu$ l of LB full medium are added. Given a 45 minutes recovery phase in a shaking incubator at 37°C, 80-100  $\mu$ l of the transformed cells are plated onto ampicillin (0.28 mM) containing plates.

### 2.6.5 Plasmid preparation

Preparation of plasmid DNA is performed using QIAprep miniprep kit. For that purpose, a single colony of transformed DH5 $\alpha$  cells is picked, inoculated in 10 ml LB Broth media and grown overnight to proliferate cells for plasmid preparation. Cells are harvested by centrifuging them at 4000 rotations per minute (rpm) and 4°C for 15 minutes. Residual media is carefully removed, the pellet resuspended in 500  $\mu$ l of buffer P1 and then transferred to a micro-centrifuge tube.

After adding 500  $\mu$ l of buffer P2, the micro-centrifuge tubes are carefully inverted several times, because fiercely handling could shear genomic DNA, which would contaminate the isolate. After

adding 700  $\mu\text{l}$  of buffer N3, the solution is mixed immediately until it became cloudy. Centrifugation for ten minutes at 13000 rpm will result in a compact pellet containing cell debris. The supernatant, where the plasmid DNA lies in, is then transferred to a QIAprep spin column by pipetting or decanting. The columns are centrifuged for one minute to load plasmid DNA onto the column and the flow-through discarded.

The column was then loaded with 750  $\mu\text{l}$  of buffer PE and centrifuged for one minute to wash. Flow-through was discarded and the column centrifuged again for one minute to remove residual washing buffer. The columns were put into clean 1.5 ml micro-centrifuge tubes and 25-30  $\mu\text{l}$  autoclaved, distilled water was dropped onto the centre of each column. After incubation of up to 15 minutes, the columns were centrifuged for one minute and the flow-through collected. Elution was redone using 50  $\mu\text{l}$  of autoclaved, distilled water to collect residual plasmid DNA. Successful mutations in the KIX coding region were verified by DNA sequencing.

## **2.7 Expression of KIX**

### **2.7.1 Cell transformation**

To uniformly label proteins for NMR analysis, cells have to be grown in media containing a source of isotopes. Cells were transformed according to the procedure described in chapter 2.6.4, but using the BL21(DE3) E.coli strain and 10 ng of plasmid DNA instead of PCR product in order to obtain a fresh culture of cells containing the pHisKIX plasmid.

### **2.7.2 Starter culture**

One to three colonies of BL21(DE3) are picked from the plate and inoculated into 1 ml LB medium containing ampicillin. The cell suspension is shaken at 37°C and 180 rpm until the optical density  $\text{OD}_{600}$  exceeds 1.0, then 200-250  $\mu\text{l}$  of the suspension are added to 1 litre M9 medium containing a  $^{15}\text{N}$  nitrogen source to uniform label protein amide groups. The suspension is shaken at 28-30°C and 180 rpm over night. The  $\text{OD}_{600}$  is around 0.3-0.5 in the next morning.

### **2.7.3 Expression culture**

Cells were further grown to an  $\text{OD}_{600}$  of 0.7-0.8, then adding IPTG to achieve a final concentration of 1 mM. Expression of KIX is activated by IPTG inhibiting the lac repressor. The cell suspension is incubated for four hours and cells harvested afterwards by centrifuging them at 4000 rpm and 4°C for 15 minutes. The pellet is resuspended in 8 ml 10 mM imidazole in 20 mM Trizma Base, pH 7.5 and 250 mM NaCl (loading buffer) and the resulting cell suspension stored at -20°C. At that point the cells can be stored for a long period, before going on with the purification step.

## 2.7.4 Purification

Purification of his-tagged proteins is straightforward using immobilised metal-ion affinity chromatography (IMAC). A stretch of histidines will form a strong chelate complex with metal ions bound to the matrix of the column, therefore trapping the desired tagged protein. Due to its similarity to a histidine side chain, imidazole will compete for binding sites with matrix-bound proteins. However, there happen to be other E.coli proteins co-purifying along with the artificially tagged protein and also various partly digested, truncated versions of the protein of interest will bind to the matrix. To remove these contaminants, size exclusion chromatography (SEC) is attached as a further purification step. In order to convey the KIX domain of CBP to a more native state, the artificial histidine-tag is cleaved off by thrombin protease. Another size exclusion chromatography step following thrombin cleavage separates uncleaved KIX and thrombin protease from the desired cleavage product.

All columns are operated with ÄKTA Explorer fast protein liquid chromatography (FPLC) device. Before applying any buffer to the column the pumps are washed with appropriate buffers and the inlets cleaned with water to avoid contamination. Frozen samples are thawed at room temperature and the same amount of 6M guanidine hydrochloride is added to the cell suspension. To disrupt plasma membranes and lyse cells, the suspension is treated with ultrasonic for 20 minutes at 50 % pulsation and 60 % power using a Bandelin Sonopuls HD 2070 homogenizer. During sonication the suspension has to be cooled with ice to avoid sample heating and minimize protease activity. The suspension of broken up cells is then centrifuged at 4°C for 20 minutes to remove cells debris. The overexpressed KIX domain is present in the supernatant and subjected to affinity chromatography for further purification.

A HiTrap<sup>TM</sup>Chelating HP column of 5 ml volume was used for affinity chromatography. The matrix of the column has to be treated with Ni<sup>+</sup> ions prior to use to allow histidine stretches to form chelate complexes. The column is first washed with four column volumes (CV) of water to remove the storage solution. Half a column volume of 1M NiCl<sub>2</sub> provides Ni<sup>+</sup> ions that bind to the column matrix. Another four CV of water are used to remove residual, unbound NiCl<sub>2</sub> from the column. After nickel loading, the HiTrap<sup>TM</sup>Chelating HP column is equilibrated with five CV loading buffer, five CV of 500 mM imidazole in 20 mM Trizma Base, pH 7.5 and 250 mM NaCl (elution buffer) and again five CV of loading buffer. The elution buffer contains very high imidazole concentration that displaces his-tagged proteins from the column. After equilibration, the supernatant of the aforementioned homogenised sample is directly loaded onto the HiTrap<sup>TM</sup>Chelating HP column that is afterwards washed with six to eight CV loading buffer to remove unbound proteins. Elution is started by applying elution buffer at a flowrate of 1 ml min<sup>-1</sup>. The process of elution is followed with a UV spectrometer at wavelengths 280 and 260 nm. Fractions of 0.75 ml are collected and pooled according to the peak dimensions.

The pooled sample is loaded onto a HiLoad<sup>TM</sup>26/60 Superdex<sup>TM</sup>75pg column that was equilibrated over night with 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 5.5 in 2 mM K<sub>2</sub>HPO<sub>4</sub> and 25 mM NaCl (potassium

phosphate buffer). Size exclusion chromatography separates KIX from other proteins that co-purified along with it. Amicon centrifugal filters with a molecular weight cut-off (MWCO) of 3 kilo Dalton (kDa) are used to reduce the sample volume if it exceeded 15 ml. Otherwise the whole sample is directly loaded onto the column. Elution is performed with potassium phosphate elution buffer at 3 ml min<sup>-1</sup> and 1.5 ml fractions were collected. Again the purification is monitored using a UV spectrometer at wavelengths 260 and 280 nm, whereupon the KIX protein peak should appear at around 80 minutes. The protein-containing fractions are pooled and concentrated to 2 ml using a 3 kDa MWCO Amicon centrifugal filter. Ten units of thrombin protease are added to the concentrated sample and incubated over night with the solution stirred. The thrombin cleaved sample is directly loaded onto HiLoad<sup>TM</sup>16/60 Superdex<sup>TM</sup>75pg and another SEC step is performed to separate the KIX domain from residual uncleaved products and thrombin protease. Fractions containing high concentration of protein are pooled and absorption measured using a UV-VIS spectrometer. Protein concentration is calculated according to the relation between absorption and quantity as shown in the following equation derived from Beer-Lambert law,

$$c = \frac{A_{280nm}}{\epsilon L} \quad (2.15)$$

where  $c$  is the concentration,  $A_{280\text{ nm}}$  the absorption at wavelength 280 nm,  $\epsilon$  the molar extinction co-efficient (12090 M<sup>-1</sup>cm<sup>-1</sup> for wildtype KIX domain and 13370 M<sup>-1</sup>cm<sup>-1</sup> for KIX F612Y mutant) and  $L$  the length of the cuvette.

### 2.7.5 Polyacrylamid gel electrophoresis (PAGE)

Following aliquots are taken during the purification procedure to analyse protein purity: homogenised cell suspension, flow-through of HiTrap, elution of HiTrap and elution of HiLoad 26/60. The aliquots are mixed with the same amount of protein sample buffer (120 mM Trizma Base, pH 6.8, 6 % (w/v) SDS, 20 % (v/v) glycerine, 0.01 % (w/v) bromphenol blue and 10 % (v,v)  $\beta$ -mercaptoethanol) to unfold and evenly charge proteins and prepare them for electrophoresis.

Resolving gel (33 % (w/v) acrylamide and 0.9 % (w/v) bisacrylamide (N,N-methylene bisacrylamide)) and stacking gel (20 % (w/v) acrylamide and 1 % (w/v) bisacrylamide (N,N-methylene bisacrylamide)) are prepared and assembled according to BioRad Mini-PROTEAN Cell instruction manual. About 10-12  $\mu$ l of each aliquot mixed with sample buffer is filled into a slot. SigmaAldrich SDS7 or Fermentas PageRuler<sup>TM</sup>Plus is used as molecular weight markers to estimate the size of proteins in the aliquots.

Protein bands are separated at 20 volts cm<sup>-1</sup> in Lämmli buffer (25 mM Trizma Base, 190 mM glycine and 0.1 % SDS) until the dye of the marker reached the end of gel. Subsequently, the resolving gel is cut off from the stacking gel and the stained with Coomassie protein staining

solution (0.5 % (w/v) Coomassie Brilliant Blue R-250, 50 % (v/v) ethanol and 10 % (v/v) acetic acid (99 %)) for 30 minutes. To destain the gel and leave the Coomassie dye in the protein bands, it is treated with destaining solution (50 % ethanol and 10 % acetic acid (99 %)) for 30 minutes and then placed in tap water until residual colour is washed out.

## 2.8 NMR measurements and data processing

### 2.8.1 NMR preparation of the KIX domain of CBP

All samples prepared for NMR measurements are concentrated to contain about 1 mM KIX in potassium phosphate buffer. MLL activation domain peptide is added to a concentration of 2 mM and the sample dialysed against 50 mM  $\text{KH}_2\text{PO}_4$  pH 5.5, 25 mM NaCl and 1 mM  $\text{NaN}_3$  (dialysis buffer) for 24 hours using a 0.5 kDa MWCO. Dissolving synthetic MLL yields to contaminating side reactions that have to be removed prior to NMR measurement. After dialysis,  $\text{D}_2\text{O}$  is added to achieve 90 % (v/v) dialysis buffer / 10 % (v/v)  $\text{D}_2\text{O}$ .

### 2.8.2 Data processing

All NMR measurements are performed at 27°C (300 K) and in 50 mM  $\text{KH}_2\text{PO}_4$  pH 5.5, 25 mM NaCl and 1 mM  $\text{NaN}_3$  (dialysis buffer). For transformation of raw NMR data to interpretable two- and three-dimensional spectra following programs are used: NMRpipe, var2pipe and seriesTab. For peak assignment CcpNmr is used.

### 2.8.3 $^1\text{H}, ^{15}\text{N}$ -HSQC

2D  $^1\text{H}, ^{15}\text{N}$ -HSQC spectra are acquired using a 800 MHz Varian INOVA Spectrometer, at 300 K. Spectra dimension span 256 ( $t_1$ )  $\times$  1344 ( $t_2$ ), with a spectral width of 2199.978 Hz  $\times$  8445.946 Hz. The spectra are processed with the NMRPipe software package [30] and analysed using CcpNmr [31].

### 2.8.4 $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY

3D  $^1\text{H}, ^{15}\text{N}, ^{15}\text{N}$ -NOESY spectra are acquired using a 800 MHz Varian INOVA Spectrometer, at 300 K. Spectra dimension span 128 ( $t_{1,1}$ )  $\times$  128 ( $t_{1,2}$ )  $\times$  1638 ( $t_2$ ), with a spectral width of 1920.031 Hz  $\times$  1920.031 Hz  $\times$  12775.471 Hz. The spectra are processed with the NMRPipe software package [30] and analysed using CcpNmr [31].

## 2.8.5 CPMG

A series of 16 HSQC spectra are acquired at different CPMG field strengths using a 500 or 600 MHz and 800 MHz Varian INOVA Spectrometer, at 300 K. Spectra dimension span 224 ( $t_1$ )  $\times$  1638 ( $t_2$ ), with a spectral width of 1920.031 Hz  $\times$  12775.471 Hz. The spectra are processed with the NMRPipe software package [30] and analysed using CcpNmr [31]. Averaged peak heights  $R_{2,\text{eff}}$  are calculated from a  $5 \times 5$  grid of data points centred at the maximum peak height using seriesTab. Carver-Richards equation is used to fit  $R_{2,\text{eff}}$  dispersion curves of residues showing intermediate exchange,

$$R_2 = R_2^0 + \frac{1}{2} \left\{ k_{ex} - \frac{1}{\tau_{CP}} \cosh^{-1} \{ D_+ \cosh(\eta_+) - D_- \cosh(\eta_-) \} \right\} \quad (2.16)$$

$$D_{\pm} = \frac{1}{2} \left\{ \pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \xi^2)^{\frac{1}{2}}} \right\}^{\frac{1}{2}} \quad (2.17)$$

$$\eta_{\pm} = \frac{\tau_{CP}}{\sqrt{2}} \left\{ \pm\psi + (\psi^2 + \xi^2)^{\frac{1}{2}} \right\}^{\frac{1}{2}} \quad (2.18)$$

where,  $\tau_{CP}$  is the duration of the spin-echo,  $\psi = k_{ex}^2 - \Delta\omega^2$ ,  $\eta = -2\Delta\omega k_{ex}(1 - 2p_B)$  [32],  $R_2$  the transverse relaxation rate constant and  $R_2^0$  the natural transverse relaxation rate governed by dipole-dipole and/or chemical shift anisotropy effects [28]. From the above equations, values for the exchange rate constant  $k_{ex}$ , chemical shift difference  $\Delta\omega$  and relative populations of states  $p_a$  and  $p_b$  can be extracted.

## 2.9 Isothermal titration calorimetry

### 2.9.1 Sample preparation

Concentration of stock solutions containing KIX are calculated from UV light absorption (see equation 2.15), whereas MLL and pKID stock solutions are prepared by weighting. Dilution and dissolving are all done in 50 mM  $\text{KH}_2\text{PO}_4$ , pH 5.5, 2 mM  $\text{K}_2\text{HPO}_4$ , 25 mM NaCl and 1 mM  $\text{NaN}_3$  (dialysis buffer). All stock solutions are simultaneously dialysed in two litres of dialysis buffer for one to three days during gentle mixing. Following solutions are prepared from stock solutions by dilution: 40  $\mu\text{M}$  KIX, 40  $\mu\text{M}$  KIX·MLL and 350  $\mu\text{M}$  pKID. Sample volumes should at least allow for three measurements to produce statistically reliable data. Measurements are performed with wildtype KIX and the KIX I660V mutant.

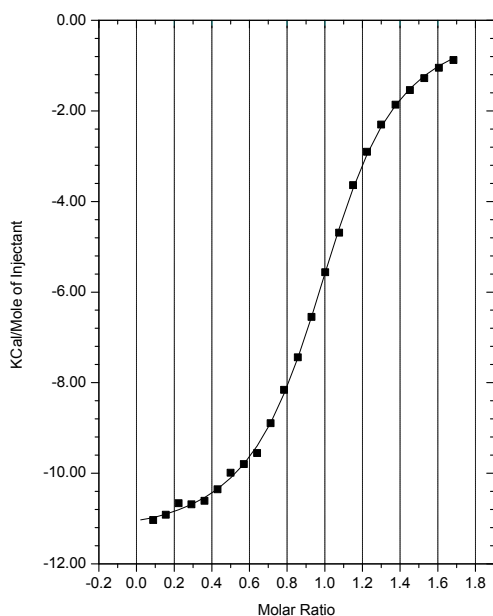


Figure 2.15: *Sigmoid titration curve of free KIX titrated with pKID.*

## 2.9.2 Experimental setup

Setting up the experiment starts with extensively cleaning the calorimeter in order to remove residual contaminations. The sample cell is washed twice with water and one time with dialysis buffer and the syringe is rinsed with approximately three syringe volumes of water. Subsequently, the clean equipment was loaded with titrand (sample cell) and titration solution (syringe). The titrand comprises of either free KIX or the KIX·MLL complex (40  $\mu$ M) and the titration solution contains 350  $\mu$ M pKID. After setting up the device titration was performed according to the following scheme: first injection of 0.5  $\mu$ l titration solution, followed by 23 injections of 1.5  $\mu$ l titration solution. Titration was performed in intervals of 180 seconds at 27°C using the MicroCal iTC<sub>200</sub> isothermal titration calorimeter.

## 2.9.3 Data evaluation

Data analysis carried out by Origin Data Analysis and Graphing Software package. Fitting the titration curve (see figure 2.15 extracts binding affinity, enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) change. The binding affinity  $K_d$  is expressed as the reciprocal dissociation constant defined by the following equations:



$$K_d = \frac{[KIX \cdot MLL \cdot pKID]}{[KIX \cdot MLL] \cdot [pKID]} \quad (2.20)$$

## 2.10 Media

### 2.10.1 M9 minimal medium

Concentration	Amount
34 mM	$\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$
22 mM	$\text{KH}_2\text{PO}_4$
8.5 mM	$\text{NaCl}$
18.7 mM	$^{15}\text{NH}_4\text{Cl}$

The medium was autoclaved at 120°C for 20 minutes and stored at 4°C until use. Before inoculating cells, following filter (0.2  $\mu\text{M}$ ) sterilized compounds were added to complete the medium: 2 ml 1 M  $\text{MgSO}_4$ , 10 ml trace elements (see subsection 2.10.2), 20 ml 20 % glucose, 0.3 ml 1 M  $\text{CaCl}_2$  and ampicillin to a final concentration of 0.28 mM.

### 2.10.2 Trace elements

Concentration	Amount
340 mM	EDTA
62 mM	$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$
2 mM	$\text{ZnCl}$
1.5 mM	$\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$
0.84 mM	$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$
3.2 mM	$\text{H}_3\text{BO}_3$
0.14 mM	$\text{MnCl}_2 \cdot 6\text{H}_2\text{O}$



# Results

## 3.11 NMR dispersion relaxation

Having been introduced to the rather complex nuclear magnetic resonance spectroscopy and its proper methods, the results and their interpretation will follow. As specified in detail in chapter 1.3,  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{15}\text{N}$ -NOESY and  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra were used to properly link amide group resonances to their position in the protein backbone. Relaxation dispersion experiments were used to quantitatively analyse exchange processes. A two-state model of the process is used to locally fit  $R_2$  relaxation dispersion data of each residue using the Carver-Richards equation. The relaxation dispersion profiles of figures 3.17 to 3.21 show dispersion curves of residues throughout the sequence of KIX. Acting on the assumption that a two-state exchange process is present in the KIX protein, three parameters, including the reaction rate constants  $k_1$  and  $k_{-1}$  and the chemical shift differences between the two states  $\Delta\omega_{\text{disp}}$ , are fit to the data. Thereby, non-flat curves provide evidence for a dynamic process occurring. It is yet important to note that dynamics outside the detection range or timescale will lead to flat curves even if an exchange process is present.

Lots of residues in the relaxation dispersion profile of the wildtype KIX·MLL complex show

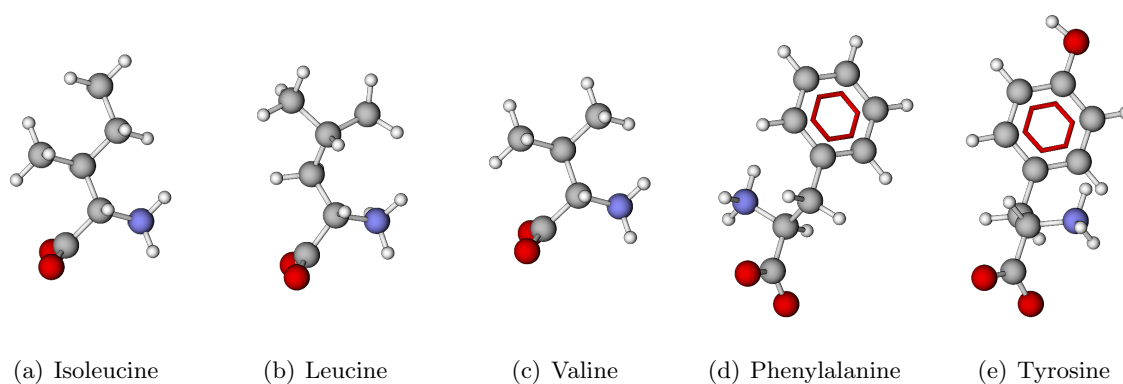


Figure 3.16: *Comparison of amino acids that were used in KIX mutation study. The amino acids are shown in their zwitterionic form, where hydrogen atoms are represented by white spheres, carbon atoms by gray spheres, nitrogen atoms by blue spheres and oxygen atoms by red spheres.*

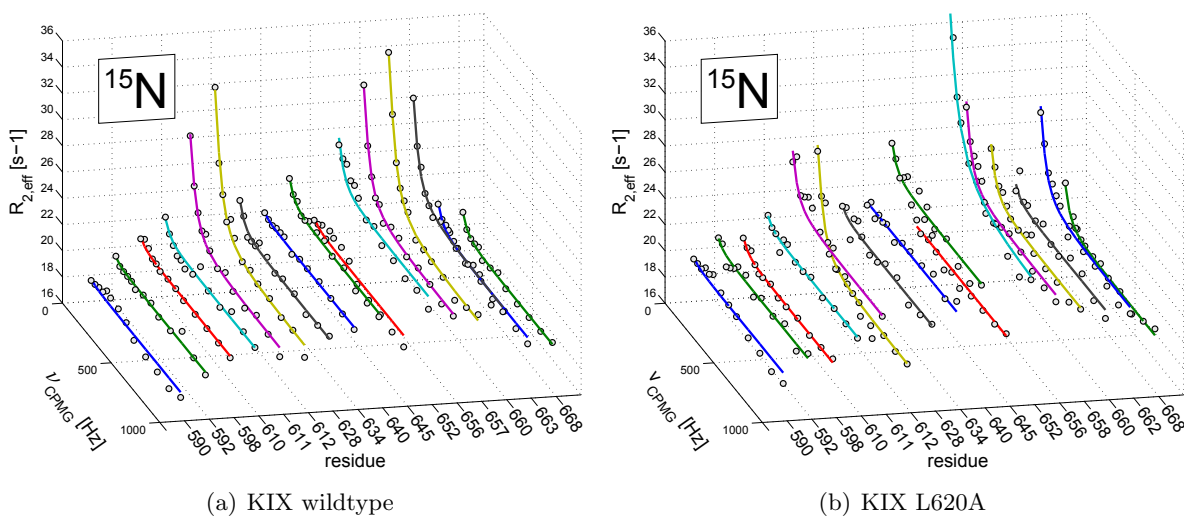


Figure 3.17: *Relaxation dispersion profile of (a) KIX wildtype (data from Brüschweiler et al., December 2008) (b) KIX mutant L620A. Comparison of the residual relaxation dispersion rates,  $R_{2,eff}$ . Non-flat fits indicate an observable exchange process in the  $\mu$ s-ms time domain. Relative relaxation rates are plotted against the CPMG field. Note that the L620A dispersion profile does not show data for residues Ile657 and Glu663, but Tyr658 and Lys662 instead.*

non-flat relaxation. Fitting of the data to Carver-Richards equation suggests a conformational transition on the micro- to millisecond timescale. Global fits of the data of Brüschweiler et al. (2008), further suggest that the excited state of the binary complex is structurally similar to the ternary complex. They reported that chemical shift differences between the ground and excited state transition of the KIX·MLL complex and the shifts of the binary to the ternary complex upon pKID binding are similar. This observation is also in agreement with KIX·MLL complex showing increased affinity to pKID when compared to free KIX. MLL binding is therefore suggested to alter the KIX structure so as to increase pKID binding affinity. NMR relaxation dispersion experiments of mutant KIX were performed to get a clue about which residues are important for the observed dynamics of the KIX·MLL complex. It should be memorised at this point that free KIX as well as the ternary KIX·MLL·pKID do not show any contribution to  $R_2$  relaxation. Such a behaviour can be explained by assuming that pKID binding will locally change side chain entropy and trigger their solidification to adopt the a more stable conformation. A series of specific residues were substituted in order to determine their evolutionary preserved role in the proposed allosteric pathway. The mutations did not impair the overall structure of KIX as chemical shifts of amide bond peaks in HSQC spectra are comparable to wildtype KIX throughout all mutants. Strong exchange throughout the sequence is observed in wildtype KIX·MLL complex, especially for residues Ile611, Phe612, Ile657 and Ile660 that comprise the hydrophobic core connecting both binding sites of KIX. A comparison of the amino acids that were used to substitute isoleucines 611, 657, 600, Phe612 and Tyr658 are shown in figure 3.16

Dynamics of nearly all KIX mutants are strongly decreased. Mutant KIX Y658V and L620A

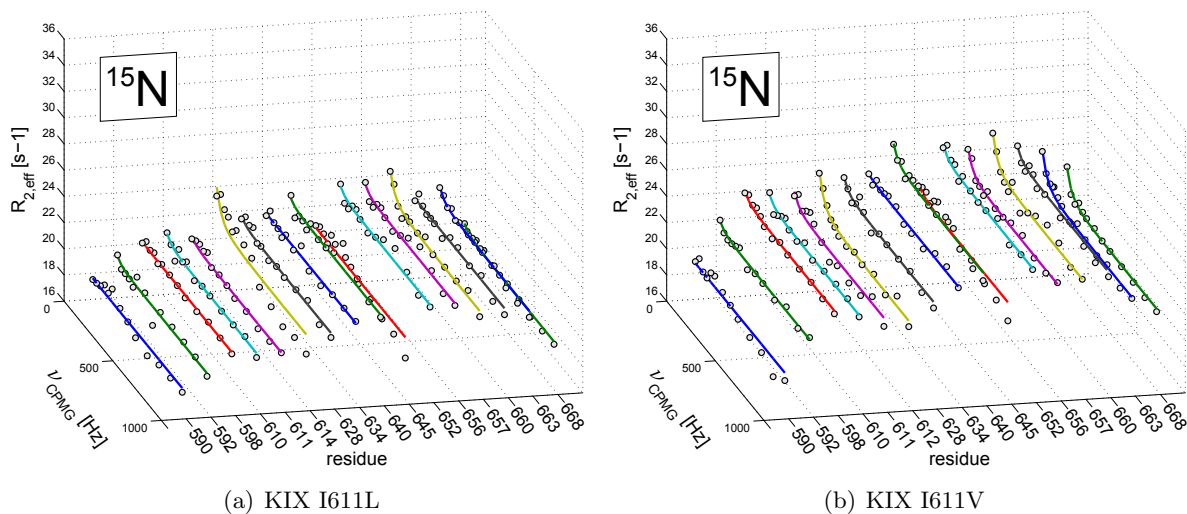


Figure 3.18: *Relaxation dispersion profile of (a) KIX mutant I611L (b) KIX mutant I611V. Note that the I611L dispersion profile does not show data for residue Ile612, but Tyr614 instead.*

seem to be the exceptions. L620A does not participate in formation of the hydrophobic core; therefore it serves as a control. Y658V defines the end of the proposed allosteric pathway (see figure 1.5) and mutations are not thought to strongly interfere. Mutations inside the transmission chain, in the hydrophobic core built by residues isoleucine 611, 657 and 660. It therefore came as a surprise that F612Y has a rather flat relaxation dispersion profile too, but all the more this suggests a very tightly coupled network of those side chain residues. NMR relaxation dispersions deliver information about the dynamic of the binary KIX complex, it is, however, not capable of measuring the consequence of absence and presence of exchange. For that purpose, binding affinity studies were done using isothermal titration calorimetry.

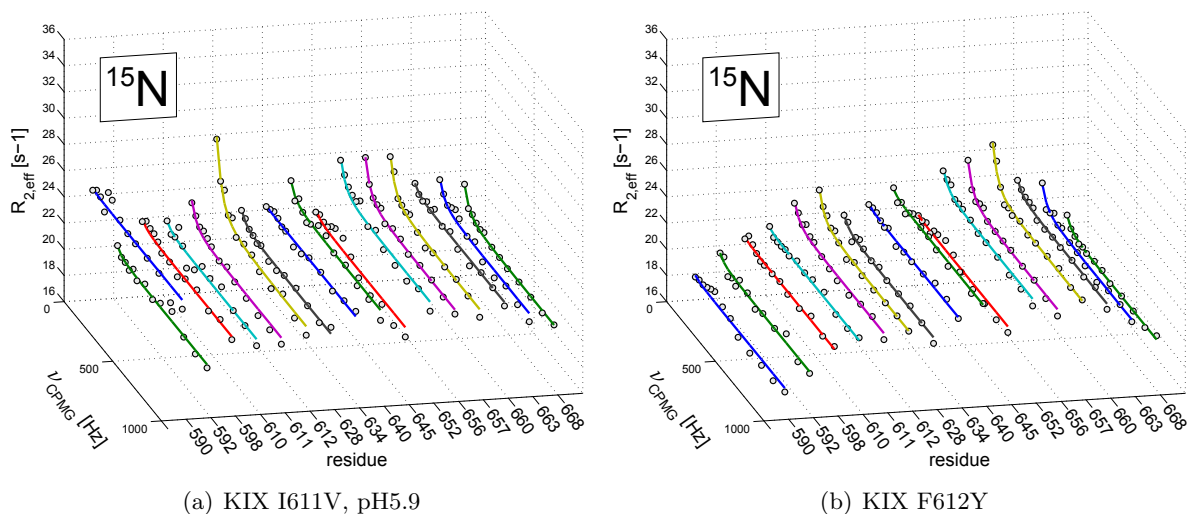
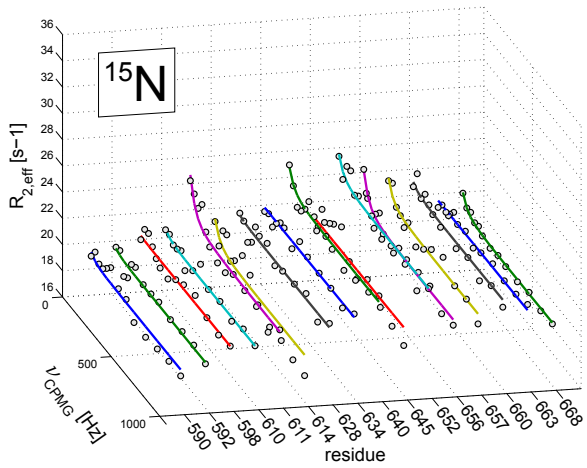
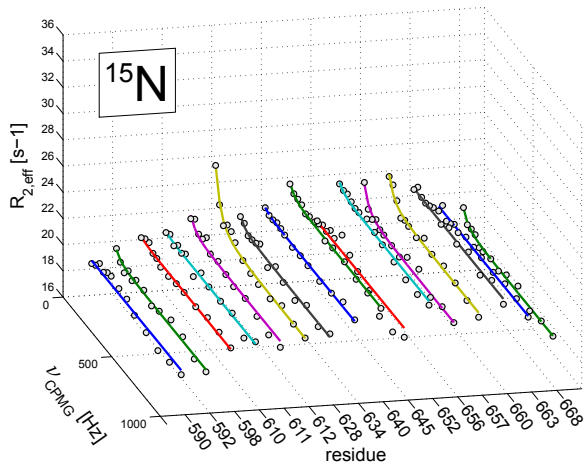


Figure 3.19: *Relaxation dispersion profile of (a) KIX mutant I611V, pH5.9 (b) KIX mutant F612Y.*

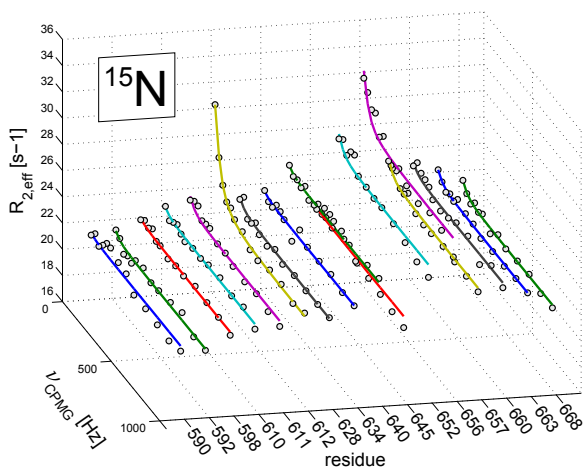


(a) KIX I657L

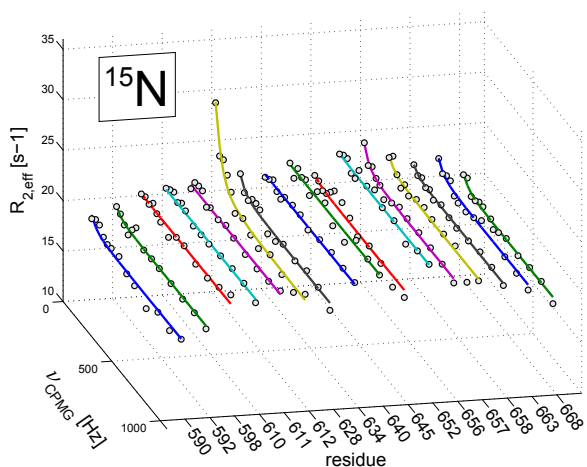


(b) KIX I657V

Figure 3.20: Relaxation dispersion profile of (a) KIX mutant I657L (b) KIX mutant I657V.



(a) KIX Y658V



(b) KIX I660V

Figure 3.21: Relaxation dispersion profile of (a) KIX mutant Y658V (b) KIX mutant I660V.

### 3.12 Isothermal titration calorimetry

The information derived from relaxation dispersion data is very reliable, but they alone do not allow for a conclusion of what happens to KIX when amino acid substitutions clearly reduce its overall dynamics. If the dynamics happen to be really necessary for the observed allosteric effect, it could be assumed that reduction of the conformational interconversion should also decrease MLL binding affinity. Isothermal titration calorimetry provides a way of measuring protein-protein interactions with great sensitivity and check if dynamics correlate with pKID binding affinity. Figure 3.22 shows a comparison between wildtype KIX and the I660V mutant. Chemical shifts of I660V differ in the range of approximately 0.5-0.6 ppm. As these small values indicate, no large structural rearrangement that could distort the ITC measurements has taken place upon mutation. However, the difference in pKID binding affinity of wildtype KIX and the binary complex is very low, so the data has to be interpreted with care. Even tiny structural rearrangements could compromise the data.

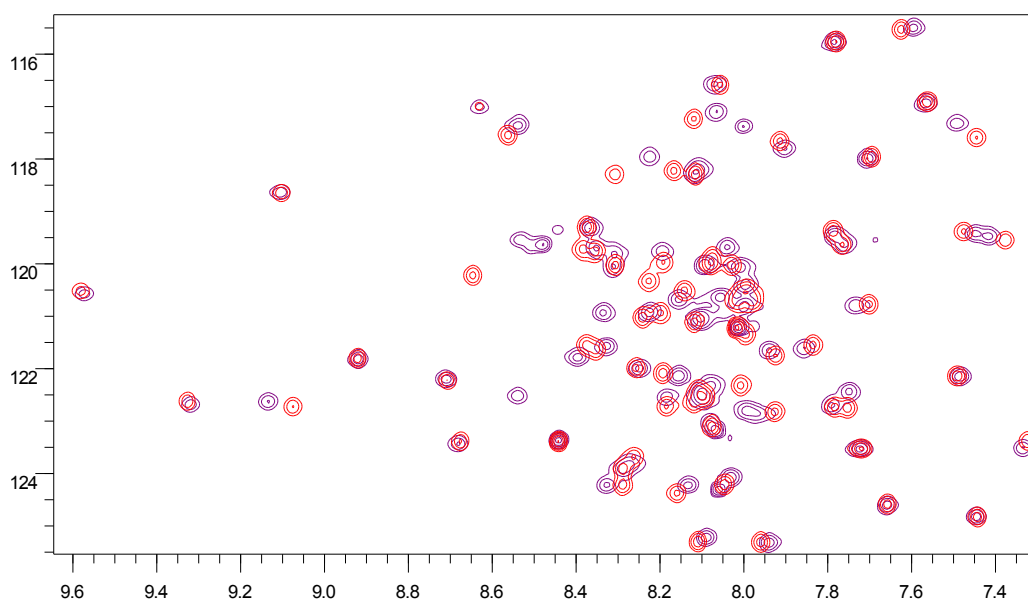


Figure 3.22: Excerpt of superimposing  $^1\text{H},^{15}\text{N}$ -HSQC spectra of wildtype KIX·MLL (blue) and mutant KIX I660V·MLL (red) complex. The peak dispersion is nearly the same for both KIX forms, with chemical shifts differing at a maximum of 0.5 - 0.6 ppm.

Fitting of the titration curve to a two-site model yields the dissociation constant  $K_d$  representing the binding affinity, as well as other thermodynamic parameters ( $\Delta H$  and  $\Delta S$ ). Data for wildtype KIX shows an approximately two-fold increase in pKID affinity upon binding of MLL and is in agreement with previously published data by Goto et. al, 2002 [1]. A two-fold increase in binding affinity is actually not very much. Speaking in terms of energy, this effect can be accomplished by the addition of a single hydrogen bond ( $5\text{-}30 \text{ kJ mol}^{-1}$ ) into a network of interactions. Even a tiny change in structure could lead to the observed effect. KIX I660V shows a dissociation constant of 2.42 and 2.45 for free KIX and the binary KIX·MLL complex, respectively. The

binding affinity of the mutant's ground state is weakly increased compared to the wildtype form. Interestingly, the binary complex does not show any increase in binding affinity of pKID upon MLL binding. Allosteric effects seem to be completely abolished at least in the I660V mutant form of KIX.

<b>Form of KIX</b>	<b>Dissociation constant <math>K_d</math></b>
KIX	$4.11 \pm 0.638 \mu\text{M}$
KIX·MLL	$1.84 \pm 0.055 \mu\text{M}$
KIX I660V	$2.42 \pm 0.491 \mu\text{M}$
KIX I660V·MLL	$2.45 \pm 0.480 \mu\text{M}$

# Discussion

Summarising the data, relaxation dispersion experiments show that the highly conserved residues participating in the allosteric pathway proposed by Brüsweiler et al. are necessary for inter-conversion between the ground state of low- and the minor state of higher pKID-affinity. The minor state was suggested to bind pKID more strongly, because calculated chemical shift differences between the ground and minor state strongly resemble the chemical shift differences observed between the transition of the binary to the ternary complex [18]. Our data clearly show that even very subtle changes of conserved amino acids comprising the binding sites and hydrophobic core impair the dynamic behaviour of the KIX domain.

ITC data of the KIX I660V mutant show that the allosteric effect was eliminated. This correlates with NMR experiments showing strongly reduced dynamic exchange, further reinforcing the model that the conformational transition is correlated to the allosteric effects. The affinity of KIX to pKID is increased by a factor of two upon MLL binding, an effect that is actually very weak when speaking of protein interactions. One has to be cautious when interpreting the data, because mutations themselves can lead to small conformational rearrangements altering the binding affinity.

The double correlated  $^1\text{H},^{15}\text{N}$ -HSQC spectrum of the KIX I660V·MLL complex is nearly identical to that of wildtype KIX·MLL, thus the mutation does not impair the structure. Further HSQC spectra of the KIX I660V·MLL complex titrated with pKID shows the transition from binary to ternary complex. The HSQC spectrum of KIX I660V·MLL with saturating pKID concentration shows decreased chemical shifts between the binary and ternary complex when compared to the wildtype form. Conformational rearrangements are not that distinctive in the KIX I660V mutant. It is therefore questionable if the mutant ternary complex adopts a conformation comparable to that of the wildtype form. The changes in structure of wildtype KIX are actually very small and so is the allosteric effect. Further experiments need to be done in order to draw this secret out of KIX. Full ITC analysis of the mutant forms will show if there really exists a correlation between allostery and exchange, while X-ray crystallography of the binary and ternary complex could originate the points of interactions between the three proteins and provide more insights of the conformational transition.

# Appendix A

## Appendix

### A.1 Plasmid map of the KIX domain of CBP

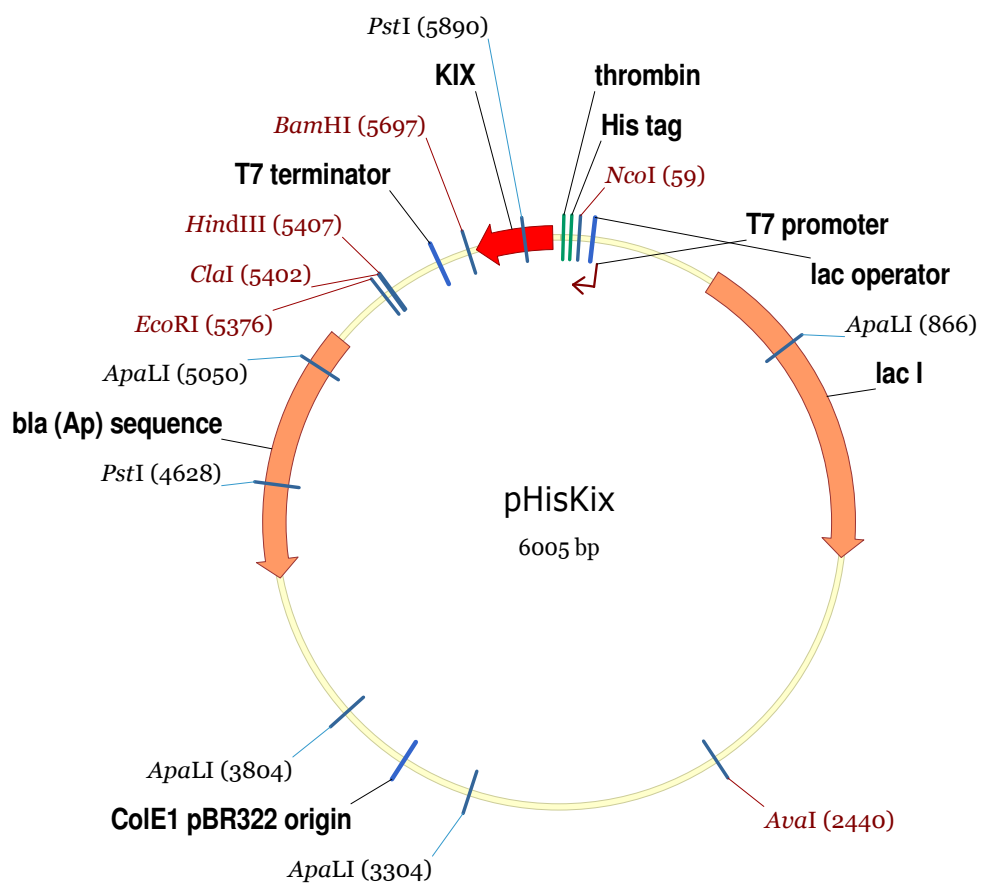


Figure A.1: *pHisKIX* plasmid map. *pHisKIX* is a derivative from *pET15b* plasmid of Novagen. The plasmid map was reconstructed following the cloning procedures of Rutledge et al. (2003) [33].



## A.2 Sequence of the KIX domain of CBP

Region	Amino acid sequence
KIX	<b>M G S S H H H H H S S G L V P R G S</b> H M G V R K G W H E H V T Q D L R S H L V H K L V Q A I F P T P D P A A L K D R R M E N L V A Y A K K V E G D M Y E S A N S R D E Y Y H L L A E K I Y K I Q K E L E E K R R S R L

Table A.1: Amino acid sequence of the KIX domain of CBP ranging from residues 586 to 672 (black), including the hexahistidine stretch (blue), thrombin cleavage site (red) and the KIX sequence. Note that cleavage of the thrombin recognition site L V P R — G S between arginine and glycine leaves artificially engineered residues G S H M behind.

Region	Nucleotide sequence
5' UTR	C G G A T A C A T T C C C T C T A G A A A T A A T T T T G T T T A A C T T T A A G A A G G A G A T A T A A A
KIX	A T G G G C A G C A G C C A T C A T C A T C A T C A T C A C A G C A G C G G C T T G G T G C C G C G C G G C A G C C A T A T G G G T G T T C G A A A A G G G T G G C A T G A A C A T G T G A C T C A G G A C C T A C G G A G T C A T C T A G T C C A T A A A C T C G T T C A A G C C A T C T T C C C A A C T C C G G A C C C T G C A G C T C T G A A A G A T C G C C G C A T G G A G A A C C T G G T T G C C T A T G C T A A G A A A G T G G A G G G A G A C A T G T A T G A G T C T G C T A A T A G C A G G G A T G A A T A C T A T C A T T T A T T A G C A G A G A A A A T C T A T A A A A T A C A A A A A G A A C T A G A A G A A A A G C G G A G G T C A C G T T T A
3' UTR	T G A A T A C T A T C A T T T A T T A G C A G A G A A A A T C T A T A A A A T A C A A A A A G A A C T A G A A G A A A A G C G G A G G T C A C G T T T A T G A T G A G G A T C C G G C T G C T A A C A A A G C C

Table A.2: Promoter and coding region of the KIX written from 5' to 3' end. 5' and 3' untranslated regions (UTRs) correspond to parts of the pHisKIX plasmid, where KIX is controlled by a T7 promoter and terminator. The coding region of KIX is dedicated to residues 586 to 672.

# References

- [1] N. K. Goto, T. Zor, M. Martinez-Yamout, H. J. Dyson, and P. E. Wright, “Cooperativity in Transcription Factor Binding to the Coactivator CREB-binding Protein (CBP),” *Journal of Biological Chemistry*, vol. 277, no. 45, pp. 43168–43174, 2002.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of THE CELL*. Garland Science, 4 ed., 2002.
- [3] A. J. F. Griffiths, W. M. Gelbart, and R. C. Lewontin, *Introduction To Genetic Analysis*. Palgrave Macmillan, 8 ed., 2004.
- [4] D. M. A., “Heat shock proteins: facts, thoughts, and dreams,” *Shock*, vol. 11, pp. 1–12, 1999.
- [5] C. A. Hutchison, S. Phillips, M. H. Edge, S. Gillam, P. Jahnke, and M. Smith, “Mutagenesis at a specific position in a dna sequence,” *Biological Chemistry*, vol. 253,, pp. 6551–6560, 1978.
- [6] M. Han and M. Grunstein, “Nucleosome loss activates yeast downstream promoters in vivo,” *Cell*, vol. 55, pp. 1137–1145, 1988.
- [7] R. D. Kornberg, “Themolecular basis of eukaryotic transcription,” *Proc Natl Acad Sci U S A*, vol. 104, pp. 12955–12961, 2007.
- [8] C. M. Thompson and R. A. Young, “General requirement for rna polymerase ii holoenzymes in vivo,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 10, pp. 4587–4590, 1995.
- [9] Y. Takagi and R. D. Kornberg, “Mediator as a general transcription factor,” *J. Biol. Chem.*, vol. 281, no. 1, pp. 80–89, 2006.
- [10] S. Malik and R. G. Roede, “Dynamic regulation of pol ii transcription by the mammalian mediator complex,” *Trends in Biochemical Sciences*, vol. 30, no. 5, pp. 256 – 263, 2005.
- [11] R. D. Kornberg, “Mediator and the mechanism of transcriptional activation,” *Trends in Biochemical Sciences*, vol. 30, no. 5, pp. 235 – 239, 2005.

- [12] R. Tjian and T. Maniatis, “Transcriptional activation: A complex puzzle with few easy pieces,” *Cell*, vol. 77, no. 1, pp. 5 – 8, 1994.
- [13] G. Orphanides, T. Lagrange, and D. Reinberg, “The general transcription factors of RNA polymerase II.,” *Genes & Development*, vol. 10, no. 21, pp. 2657–2683, 1996.
- [14] R. H. Goodman and S. Smolik, “CBP/p300 in cell growth, transformation, and development,” *Genes & Development*, vol. 14, no. 13, pp. 1553–1577, 2000.
- [15] R. Janknecht and A. Nordheim, “Map kinase-dependent transcriptional coactivation by elk-1 and its cofactor cbp,” *Biochemical and Biophysical Research Communications*, vol. 228, no. 3, pp. 831 – 837, 1996.
- [16] A. W. Snowden and N. D. Perkins, “Cell cycle regulation of the transcriptional coactivators p300 and creb binding protein,” *Biochemical Pharmacology*, vol. 55, no. 12, pp. 1947 – 1954, 1998.
- [17] K. J. McManus and M. J. Hendzel, “Cbp, a transcriptional coactivator and acetyltransferase,” *Biochem. Cell Biol.*, vol. 79, pp. 253 – 266, 2001.
- [18] S. Brüschweiler, P. Schanda, K. Kloiber, B. Brutscher, G. Kontaxis, R. Konrat, and M. Tollinger, “Direct observation of the dynamic process underlying allosteric signal transmission,” *JACS*, vol. 131, pp. 3063–3068, 2009.
- [19] I. Radhakrishnan, G. C. Pérez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright, “Solution structure of the kix domain of cbp bound to the transactivation domain of creb: A model for activator:coactivator interactions,” *Cell*, vol. 91, no. 6, pp. 741 – 752, 1997.
- [20] R. Janknecht and A. Nordheim, “Map kinase-dependent transcriptional coactivation by elk-1 and its cofactor cbp,” *Biochemical and Biophysical Research Communications*, vol. 228, no. 3, pp. 831 – 837, 1996.
- [21] W. DeLano, “The pymol molecular graphics system,” 2002.
- [22] Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, “Gapped blast and psi-blast: a new generation of protein database search programs,” *Nucleic Acids Res.*, vol. 25, pp. 3389–3402, 1997.
- [23] C. Notredame, D. G. Higgins, and J. Heringa, “T-coffee,” *J. Mol. Biol.*, vol. 302, pp. 205–217, 2000.
- [24] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, “Jalview Version 2—a multiple sequence alignment editor and analysis workbench,” *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.
- [25] J. Keeler, *Understanding NMR Spectroscopy*. John Wiley & Sons, 1 ed., 2005.

- [26] M. Sattler, J. Schleucher, and C. Griesinger, "Heteronuclear multidimensional nmr experiments for the structure determination of proteins in solution employing pulsed field gradients," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 34, no. 2, pp. 93 – 158, 1999.
- [27] A. Mittermaier and L. E. Kay, "New Tools Provide New Insights in NMR Studies of Protein Dynamics," *Science*, vol. 312, no. 5771, pp. 224–228, 2006.
- [28] V. I. Bakhmutov, *Practical Nuclear Magnetic Resonance Relaxation for Chemists*. John Wiley & Sons, 1 ed., 2004.
- [29] M. W. Freyer and E. A. Lewis, "Isothermal titration calorimetry: Experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions," *Methods in cell biology*, vol. 48, pp. 79–112, 2008.
- [30] F. Delaglio, S. Grzesiek, V. GW, Z. G, P. J, and B. A., "Nmrpipe: a multidimensional spectral processing system based on unix pipes," *J. Biomol. NMR*, vol. 6, pp. 93–277, 1995.
- [31] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, "The ccpn data model for rnmr spectroscopy: Development of a software pipeline," *Proteins*, vol. 59, pp. 687–696, 2005.
- [32] V. Z. Miloushev, F. Bahna, C. Ciatto, G. Ahlsen, B. Honig, L. Shapiro, and A. G. P. III, "Dynamic properties of a type ii cadherin adhesive domain: Implications for the mechanism of strand-swapping of classical cadherins," *Structure*, vol. 16, no. 8, pp. 1195 – 1205, 2008.
- [33] S. E. Rutledge, H. M. Volkman, and A. Schepartz, "Molecular recognition of protein surfaces: High affinity ligands for the cbp kix domain," *JACS*, vol. 125, no. 47, pp. 14336–14347, 2003.

I was bestirred to find all proprietors of images and figures and ask for permission of utilisation in this thesis. In any case of copyright violation, please notify me.

# Curriculum vitae

**Personal data** Name: Reiner Josef Ribarics  
Date of birth: 1st October 1984  
Nationality: Austrian  
Primary residence: Erlengasse 23, 7312 Horitschon, Austria  
Secondary residence: Ybbsstraße 14/8, 1020 Wien, Austria

## Education

2009 Diploma thesis with the Department of Biomolecular Structural Chemistry at the University of Vienna  
2008 Two months of practical work with the Department of Chromosome Biology in the group of Peter Schlögelhofer  
2007 Two months of practical work with the Department of Molecular Cell Biology in the group of Gerhard Wiche  
2004 - present Diploma study of Molecular Biology with special emphasis in Biochemistry, Plant Biology and Bioinformatics  
1995 - 2003 BORG Oberpullendorf  
1991 - 1995 Volksschule Horitschon