MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„A computational exploration of Cereblon binders and its dynamics"

verfasst von / submitted by

Anna Aplenc, MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Thierry Langer

Acknowledgement

Firstly, I would like to thank Dr. Peter Ettmayer for getting me in touch with my direct supervisor at Boehringer & Ingelheim, Vienna. Through his kindness, I got the opportunity to meet my incredible mentor for this master thesis.

My greatest thanks go to my mentor Dr Andreas Bergner, who not only introduced me to the world of Computational Chemistry, but also who I want to express my outmost admiration to for his guidance, patience, and knowledgeability. He guided me through a new area of research that allowed me to expand my knowledge invaluably and challenged me with vivid discussions on-topic. He creates an inspiring and productive working environment and takes time to listen to a master student's endless number of questions. In addition, I enjoyed at lunch breaks the interesting views of an enthusiastic "Wagnerian", which I will keep in mind for the next choice of a classical concert.

Moreover, I want to acknowledge the colleagues from the MedChem group, such as Yvonne Westmaier, Philipp Schmalhorst as well as colleagues from the structural biology and other groups, such as Silvia Arce-Solano, Gerhard Fischer, Leonhard Geist, Lion Herndlhofer, Andrea Kahlbacher, Dirk Kessler, Moriz Mayer, Lena Muenzker, Michael Passegger, Klaus Rumpel, and Alexander Weiss-Puxbaum, to mention a few.
Especially I want to thank Dr Zuzana Jandova, who has guided and taught me all I know about MD Simulations. Without her, the work of this thesis would not have been as near complete or such a successful and huge learning experience.

Further, I want to thank the colleagues and friends I got to know during my internship at Boehringer & Ingelheim. From the orchestra to the choir, from one department to the other, from onboarding to the last round of coffee in the kitchen – thank you for making this experience so unique. It is a special case when one meets only colleagues that are helpful, inspiring, kind and patient.

From the University of Vienna, I would like to express my gratitude to Univ.-Prof. Thierry Langer, whose enthusiasm for my master thesis was unparallel and his advice was invaluable. He and his colleagues Thomas Wieder and Christian Fellinger have contributed and guided my learning experience as well as the development of my work by being part of monthly discussions. Their contributions not only were helpful but inspired further advances in the hotspot analysis tool GRAIL.
Eventually, Prof. Langer played already before this master thesis an important role in my university career. It was due to him that I was accepted into the master programme Drug Discovery and Development. It has undoubtedly transformed my study career in unimaginable ways, and I am very happy to have found myself in this new and extraordinary study programme in Vienna, which I was able to graduate from last November.
I would like to acknowledge everybody that has supported, taught, and challenged me throughout the years of my studies. They have been very challenging, hard-working, unpredictable, and educational years so far. I hope to have gained the knowledge and learned as many lessons as possible to be well prepared for my future career steps.

There is no shortage of words that could express my gratitude and admiration for my family and friends that have been with me throughout my life. I would like to express special thanks to my boyfriend, who continuously surprises me and encourages me with endless support. Finally, I must thank the three most important people in my life who I call family. They have been the most constant, relentless, and inspirational support, they have been my guiding light throughout my life and without whom I would not dare to reach for more.

1. Table of Content

1. Table of Content 6
2. Zusammenfassung 8
3. Summary 9
4. Introduction 10
   4.1. Ubiquitin proteasome system (UPS) 11
      4.1.1. 26S Proteasome 13
      4.1.2. Ubiquitin activating enzyme E1, conjugating enzyme E2 and ligase enzyme E3 14
   4.2. E3 Ligase CRBN 19
      4.2.1. CRBN Introduction 19
      4.2.2. Dynamics of CRBN 21
   4.3. PROTACs 23
      4.3.1. History of IMiDs 28
      4.3.2. CRBN PROTACs 31
         4.3.2.1. Covalent ligands 33
         4.3.2.2. Covalent Docking in the IMiD binding site 37
      4.3.3. Studying ligand interactions to CRBN 39
         4.3.3.1. MD simulations with CRBN 40
         4.3.3.2. Application of a hotspot analysis based on MD trajectories 42
5. Methods 45
   5.1. Overview of Workflow 45
   5.2. Structural Overview of CRBN 46
   5.3. Covalent Docking 47
   5.4. Molecular Dynamics (MD) Simulation 51
      5.4.1. Preparing MD Simulations 51
      5.4.2. Running MD Simulations 52
      5.4.3. Analysis of Trajectories 53
         5.4.3.1. MD Interaction and Pharmacophores Analysis 54
         5.4.3.2. GRids of phArmacophore Interaction fieLds (GRAIL) 55
6. Results 57
   6.1. Structural Findings of CRBN 57
   6.2. Covalent Docking 58
   6.3. MD Simulations 68

6.3.1. Interaction analysis between lenalidomide and both CRBN conformers ... 70
6.3.2. Hotspot analysis: GRAIL ... 80
7. Discussion ... 89
7.1. Covalently binding PROTACs ... 89
7.2. Hotspot analysis of dynamic CRBN ... 94
7.3. Outlook ... 98
8. List of Abbreviations ... 99
9. List of Figures ... 103
10. References ... 112

2. Zusammenfassung

Das Ubiquitin-Proteasom-System (UPS) ist einer von zwei Hauptwegen, über die die Zelle ungefaltete oder beschädigte Proteine abbaut. Der Hauptkomplex von UPS besteht aus drei Enzymen (E1, E2 und E3), die das Ubiquitin als Signalmolekül aktivieren, übertragen und an die Proteinsubstrate binden. Proteine mit solchen Ubiquitinketten werden durch das Proteasom für den Abbau durch das Proteasom markiert. Aufgrund einer der größten medizinischen Tragödien der Geschichte wurde die E3-Ligase Cereblon (CRBN) als Hauptziel sogenannter immunmodulatorischer Medikamente (IMiDs) wie Lenalidomid entdeckt. Trotz ihrer teratogenen Wirkung sind ihre Strukturen oft Teil einer neuen Medikamentenklasse namens PROTACs (Proteolysis Targeting Chimera), die den Abbau von Proteinen für eine therapeutische Modulation beeinflussen sollen. Die Studie dieser Masterarbeit zielte darauf ab, das Verhalten neuer CRBN-Liganden zu verstehen, von denen vorgeschlagen wurde, dass sie kovalent an einen speziellen Histidinrest in der IMiD-Bindungstasche von CRBN binden und nicht nur den Wechsel von der offenen zur geschlossenen Konformation des Proteins bewirken, sondern auch den Abbau von Substrateni nduzieren. Durch kovalentes Docking wurden Wechselwirkungsmuster mit neuen Resten analysiert, die zur Verbesserung des PROTAC-Designs beitragen könnten. Darüber hinaus wurden Moleküldynamiksimulationen (MD) durchgeführt, um die unterschiedlichen Wechselwirkungen von Resten mit Lenalidomid zu verstehen und Hotspots in CRBN vorherzusagen, um das Design neuer Liganden basierend auf den simulierten Konformationen des Proteins voranzutreiben.

3. Summary

The ubiquitin proteasome system (UPS) is one of two major pathways, by which the cell degrades unfolded or damaged proteins. The main complex of UPS is comprised of three enzymes called E1, E2, and E3 that activate, transfer, and ligate the Ubiquitin as a signal molecule onto the protein substrates, respectively. Proteins with such Ubiquitin chains are marked for proteasome degradation by the proteasome. Due to one of the greatest medical tragedies in history, the E3 ligase Cereblon (CRBN) was discovered as the primary target of so-called immunomodulatory drugs (IMiDs) such as lenalidomide. Despite their teratogenic effects, their structures are often a part of a new drug class called PROTACs (proteolysis targeting chimera) that are designed to affect the degradation of proteins for a therapeutic modulation. The work of this master thesis aimed to understand new ligands that were proposed to covalently bind to a special Histidine residue in the IMiD binding pocket of CRBN and cause not only the switch from the open to the closed conformation of the protein, but also induce degradation of substrates. By covalent docking interaction patterns to new residues have been analysed that could help to improve PROTAC design. In addition, molecular dynamic (MD) simulations were run to understand the different interactions of residues to lenalidomide and to predict hotspots on CRBN to advance new ligand design based on the simulated conformations of the protein.

4. Introduction

The human cell has various mechanisms and pathways controlling the quality of synthesized nucleic acids and proteins. Proteins that are damaged, unfolded or not useable anymore are subjected to protein degradation mechanisms. The most prominent pathway for protein degradation is the ubiquitin proteasome system (UPS) that ensures protein homeostasis. This includes maintaining the protein levels at the cellular concentrations needed, folding in the correct state and their presence at the right time and right place. The second degradation system is incorporated in vesicles, called lysosomes. In contrast, these special organelles degrade long-lived or insoluble proteins, organelles, and even intracellular parasites [1]–[3].
In the context of this master thesis, the UPS will be the focus and discussed further in more detail as well as applying its special mechanism of action for therapeutic interventions. The UPS was soon discovered as a mode of action to degrade specific proteins in the cell and showed promise to prompt the development of anticancer therapies. Many cancer strategies are dependent on the UPS such as an abnormal degradation of the cell cycle control proteins and aberrant regulation of proapoptotic factors. Consequently, efforts went into targeted drug design for the individual steps of the UPS pathway. This master thesis will focus on targeted protein degradation (TPD) strategies via the UPS, which embrace the design of particular drug molecules such as proteolysis targeting chimera (PROTACs) and molecular glues (MGs) (Figure 1). Initially, MGs have been described as molecules inducing the protein degradation by interacting with one of the enzymes forming the degradation complex. The E3 ligase is the third enzyme of the multiprotein complex transferring the Ubiquitin signal onto the substrate that are then ready for degradation. These promising drug groups interact with the E3 ligase called cereblon (CRBN) and interact with the same protein binding sites where immunomodulatory drugs (IMiDs) bind. From their first description until now a plethora of these molecules have been developed based on the original IMiD ligand motives. Even though thalidomide as the first IMiD drug on the market caused one of the biggest scandals in the history of the pharmaceutical industry, many PROTACs use the known binding interactions to the E3 ligase and have developed as highly specific targeted protein degraders [4]–[7].

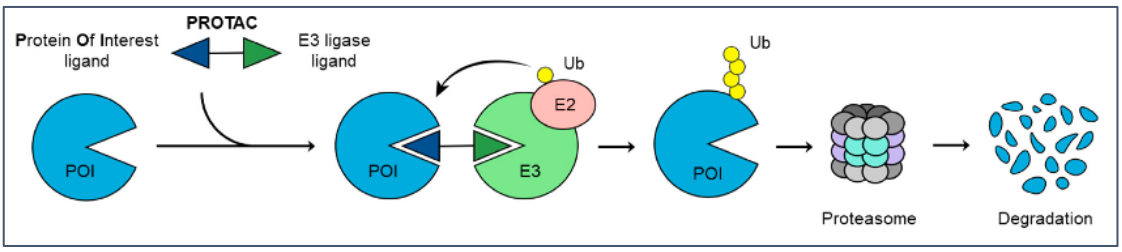


Figure 1 Mechanism of action of heterobifunctional molecules called Proteolysis targeting chimeras (PROTACs): Hijacking the ubiquitin proteasome system (UPS) by bringing E3 ligase and protein of interest (POI) in close proximity and inducing ubiquitination on Lysine residues and subsequent degradation of the POI by the proteasome (schematic figure adopted from [8]

4.1. Ubiquitin proteasome system (UPS)

The UPS removes short-lived and soluble, misfolded proteins from the cell by attaching the ubiquitin (Ub) as a tag on the target proteins. Ub is a 76-residue protein and gets attached to specific Lysine (Lys) residues of the target protein. This post translational modification (PTM) is carried out via a multiple step mechanism involving a multimeric complex [4]–[6]. The UPS carries out the ubiquitinoylation, the de-ubiquitinoylation and finally the proteasomal degradation [2].
Three main enzymes orchestrate the Ub transfer to the substrate. In an ATP dependent step, the Ub activating enzyme (UAE) E1 is the first protein activating the Ub and transferring it to the next catalytic enzyme of the cascade. The Ub conjugating enzyme E2 takes over the Ub to transfer it to Ub ligase enzyme E3. This enzyme connects the PTM via an iso-peptide bond to specific Lys residues of the substrate protein. The multi-step process can be repeated since Ub has several Lys residues that can be ubiquitinated and consequently, polyubiquitin chains can be formed [3] (Figure 2). For the prolongation of the Ub chain a fourth enzyme, Ub-chain elongation factor E4 is needed [2]. On which Lys residue the initial Ub as well as the elongation to a polyubiquitin chain is made, can influence the degradation mechanism. A well-described linkage to Ub is Lys11 and Lys48, in contrast to the Lys6 or Lys29 Ub signals that are also known to induce the UPS and proteasomal degradation but are less described in literature. In addition, there are Lys residues such as Lys63, providing signals outside the UPS machinery, like for DNA repair and activation of transcription factors [1], [6] (Figure 3). Further, different Ub patterns have been observed to influence the substrates E3 ligases recognize. The diverse

patterns of the PTM lead to protein selectivity and specificity for degradation [1]. The specific Ub decorations on the substrates that signals the transport to the 26S proteasome. This multi-subunit protein complex hydrolyses the Ub proteins into small peptides, which also get deubiquitinated in this part of the degradation process by the corresponding deubiquitinating enzymes (DUBs) [4]–[6].

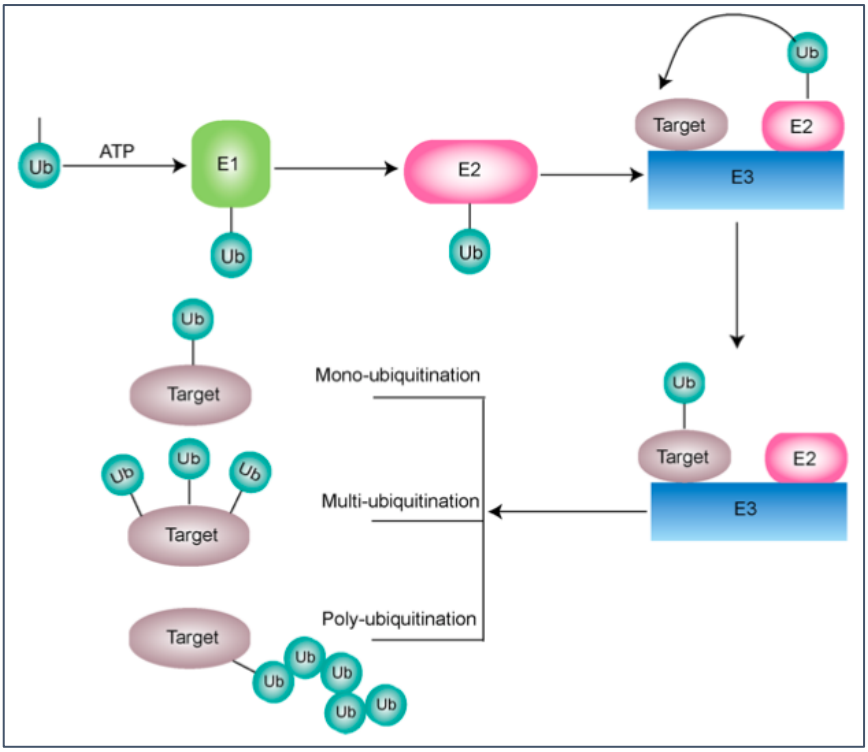


Figure 2 Three main enzymes orchestrate the ubiquitin (Ub) transfer to the substrate. In an ATP dependent step, the Ub activating enzyme (UAE) E1 is the first protein activating the Ub and transferring it to the next catalytic enzyme of the cascade. The Ub conjugating enzyme E2 takes over the Ub to transfer it to Ub ligase enzyme E3. This enzyme connects the post-translational modification (PTM) via an iso-peptide bond to specific Lysine (Lys) of the substrate protein. The multi-step process repeats since also Ub has several Lys residues and polyubiquitin chains can form. (schematic figure adopted from [9])

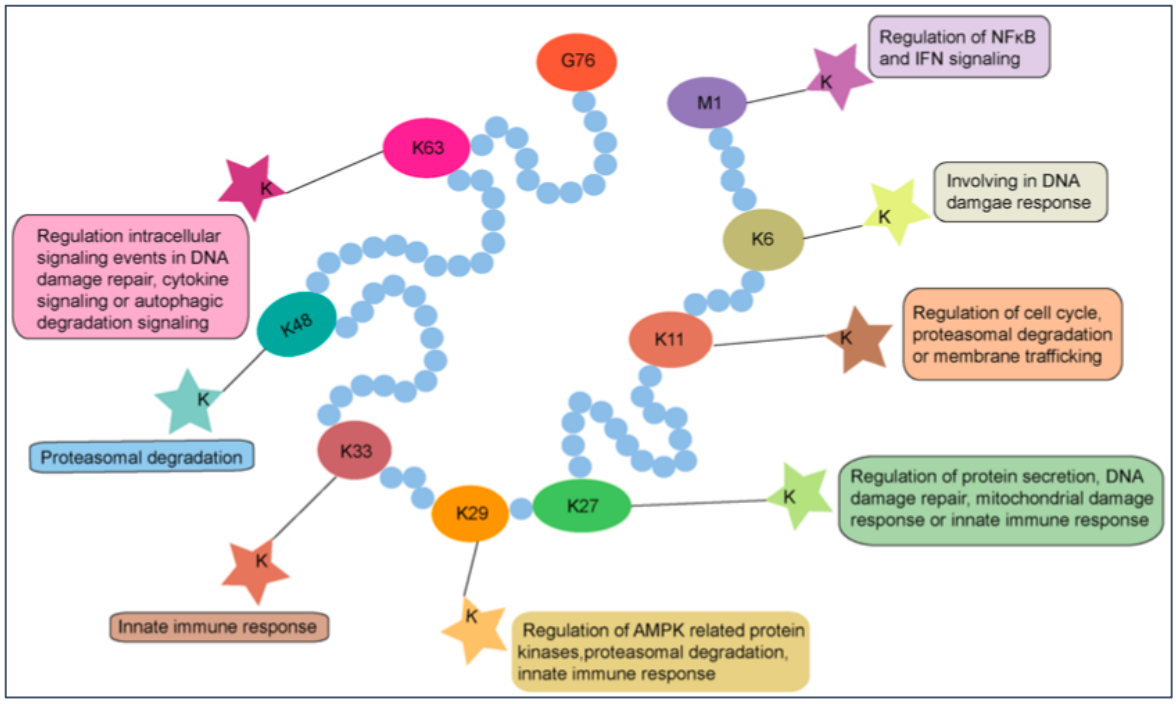


Figure 3 The initial Ubiquitin (Ub) is transferred and attached to a specific Lysine (Lys) residue. From the initial Ub the signal is elongated to a polyubiquitin chain that can influence the degradation mechanism. Depending on which Lys residue the Ub signals are attached different mechanism are activated in the cell. A well-described linkage to Ub chains is Lys11 or Lys48, in contrast to the Lys6 or Lys29 Ub signals are also known to induce the UPS and proteasomal degradation but are less described in literature. In contrast, there are Lys residues such as Lys63, providing signals outside the UPS machinery, like for DNA repair and activation of transcription factors. In addition, different Ub patterns have been observed to influence the substrates E3 ligases recognize. The diverse patterns of the PTM lead to protein selectivity and specificity for degradation. Ubiquitin is attached to 7 lysine sites in figure: K6, K11, K27, K29, K33, K48, K63; and N-terminal Methionine (M1). (schematic figure adopted from [9])

4.1.1. 26S Proteasome

The cytosolic complex of proteins called 26S proteasome degrades the ubiquitinylated substrates and consists of several subunits that are shaped in a tubular form (Figure 4). The major part is comprised of the 20S proteasome, the core part that contains the proteolytic function. It splits the substrates into peptides in an ATP-independent manner at the inner side of the tube-shaped complex, where the active sites reside. The inner side can degrade a range of substrates since its comprised of different subunits harbouring caspase-, tryptic- and chymotryptic-like sites [2]. At the top and bottom of the barrel shape of the 20S proteasome, a 19S cap is bound and ensures specific degradation of the substrate, when ATP is present (Figure 4). This protein subunit is the regulatory part and assures that only

ubiquitinated proteins enter the channel of the proteasome to the active site of degradation [1].

In general, the 26S proteasome plays an important role as it contributes to the regulation of the cell cycle, inflammatory and immune responses, protein misfolding, ER associated degradation and disease progression [1]. Therefore, the entire multi-subunit complex has been also part of investigations trying to inhibit or activate the cascade of degradation, among other steps in the UPS pathway. For the treatment of cancers, the proteasome has been a popular target to slow down or prevent tumour growth. Nowadays, there are several generations of small molecules that have been clinically approved, such as Bortezomib. Its inhibitory effect blocks the chymotryptic active sites in the proteasome and has been used to treat various lymphomas [6]. The downside of all proteasome inhibitors so far, is that they cause the patients to become refractory meaning they are no longer responding to the treatment. In addition, proteasome inhibitors are not as efficient, especially in solid tumours, and require an additional inhibitor or otherwise a combinatorial therapy [2].

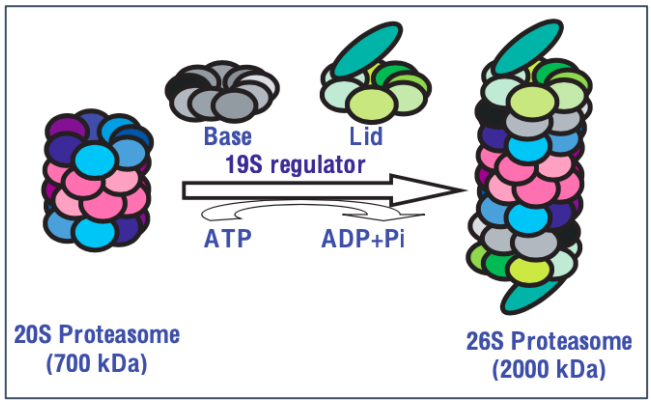


Figure 4 Formation and structural subunits of 26S proteasome (schematic figure adopted from [1])

4.1.2. Ubiquitin activating enzyme E1, conjugating enzyme E2 and ligase enzyme E3

Before a protein substrate gets degraded by the 26S proteasome three main enzymes called E1, E2 and E3 ensure the ubiquitination of the protein for signalling and transfer. In the human genome for each of these proteins there is more than just one subtype of each enzyme. For E1 two enzymes and for E2 about 40 enzymes are encoded, while for the E3 ligase over 600 members have been found [2]. This allows the UPS mechanism to use different E3 ligases with a very high substrate specificity,

since this enzyme is responsible for substrate recognition and interaction [4]. Depending on how the Ub transfer is carried out by the Ub conjugating E2, the ternary complexes consisting of E1, E2 and E3 enzymes belong to different subgroups: (1) RING (Really Interesting New Gene), (2) HECT (Homologous to E6AP C-terminus) and (3) RBR (RING between RING). The RING class being the largest group of E3 enzymes transfers the Ub to the substrate without formation of a thioester bond to E2, while the HECT E3 ligases form a catalytic Cystein (Cys) dependent intermediate with Ub attached to E2 before transfer to the target protein. As the structure and name behind the mechanism already tells, RBR E3 ligases carry out a combinatorial approach. Their RING1 domain recognizes the attached E2 Ub (RING-like), and the RING2 domain offers a Cys-residue to the active site of E2 forming a thioester bond as an intermediate step (HECT-like)[6].

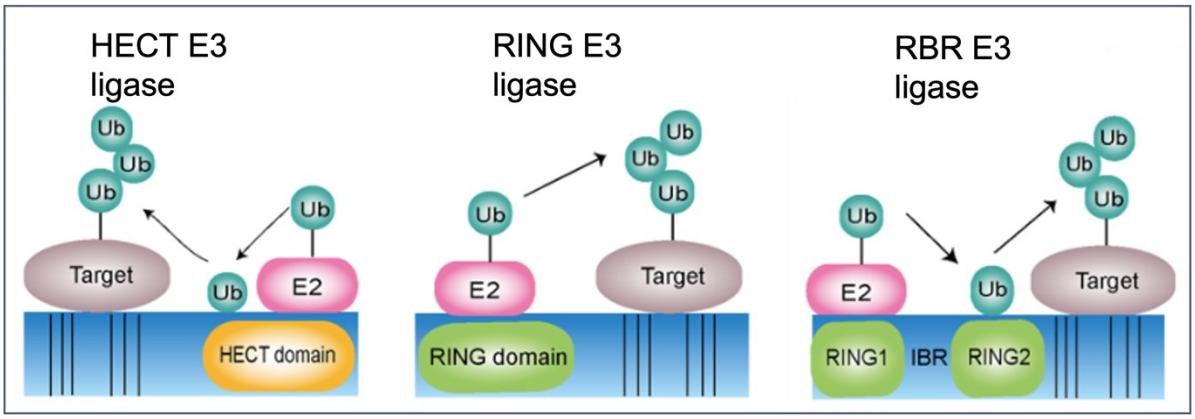


Figure 5 Depending on how the Ub transfer is carried out by the E2, the ternary complexes with E1, E2 and E3 belong to different subgroups. From left to right: (1) RING (Really Interesting New Gene), (2) HECT (Homologous to E6AP C-terminus) and (3) RBR (RING between RING). (schematic figure adopted from [9])

For UAEs only a few small molecules have been reported as inhibitors for either E1 enzyme UBA1 or UBA6. PYR-41 is a pyrazolidine-(di)on derivative and was shown as the first targeting inhibitor by forming a covalent bond to a Cys residue of E1 that is also cell-membrane permeable [2] (Figure 6A). Its mechanism of action found only promising results in targeting cancers with p53 mutations driving these cells into apoptosis. Similar mechanism of action was observed for the small molecule MLN4924, an inhibitor of the closely related E1 enzyme, NEDD8-activating enzyme (NAE) (Figure 6B). Neddylation is a very specific PTM with NEDD8 being a ubiquitin-like protein. MLN4924 is a first-in-class inhibitor and in clinical development also known as TAK-243. The treatment with this adenosine sulfamate analogue mediates

cell death of cancerous cells from haematological and solid tumours. However, inhibition of later steps in this UPS cascade have shown more promising results [2], [6].

E2 conjugates the Ub and in cooperation with a specific E3 ligase, they specify the type of polyubiquitin chain linkage. Due to this essential association with an E3 enzyme, most of the efforts toward an inhibitory ligand have been focusing on E3 blockage, except for a few reported small molecules such as CC0651 and NSC697923 (Figure 6C and 6D). These E2 targeting compounds either cause great structural rearrangements discharging Lys residues as Ub acceptor site or prevent directly the E2-Ub conjugation. Interestingly, CC0651 has been reported to have a very high specificity for the E2 enzyme Ube2R1 and similarly to the first E1 modulator PYR-41, the inhibitor NSC697923 was proposed to bind covalently to a corresponding catalytic Cys residue, too [2], [6].

The last step before delivery of the ubiquitinated substrate to the proteasome, is carried out by E3 ligases. While a huge number of the E3 enzymes exists for different substrate specificities, trying to decipher all their implications in a network of cellular pathways has been a challenging task. Already an overview of the main mechanism would extend vastly the frame of this master thesis.
A similarly challenging task has been to understand how to categorize human encoded E3 ligases in a manageable summary as most of the literature focuses on various aspects of ligand design, their modulation, mechanism of actions or individual complex structures rather than exploring the details of the E3 ligase relations [4], [6], [10], [11] (Figure 7). Due to these reasons, only four out of hundreds of E3 ligases are most often highlighted as they have been the focus of published research, from which one was focused on in this master thesis. Mouse double minute homolog 2 (MDM2), inhibitors of apoptosis proteins (IAP), Von-Hippel-Lindau (VHL) and CRBN are the most researched E3 ligases and have been also used for the development of various ligands, especially those called PROTACs as described below [4], [6], [10], [11]. To note, researchers have developed modulating binders for all three enzymes involved in the UPS, but so far FDA-approved drugs were only those targeting E3 enzymes [2].

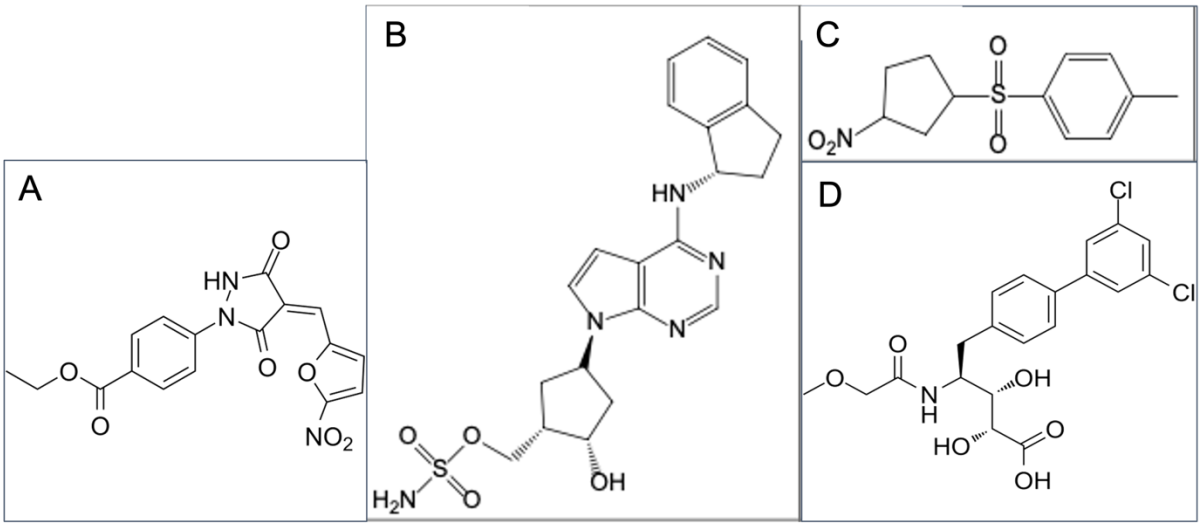


Figure 6 Structures of various small molecules modulating the three enzymes E1, E2 and E3. From left to right: (A) PYR-41 is a known covalently binding E1 ligand, (B) MLN4924 inhibits the closely related E1 protein called NEDD8 activating enzyme (NAE) that signals via NEDD8 molecules, (C) NSC697923 and (D) CC0651 are one of the very few E2 targeting inhibitors that cause great structural rearrangements and subsequently, the Ubiquitin is not ligated to the Lys residues. (molecular structures adapted from [12])

The E3 enzyme MDM2 contains a RING finger domain and is known as a critical player in the negative regulation of the tumour suppressor p53. This interplay of MDM2-p53 was extensively researched showing the loss of MDM2 inducing cell death in an p53-dependent manner and understanding that in over 50% of cancers mutations in p53 are found. Various compounds have been reported to inhibit the interaction between these two players [4], [6], [11]. Among those, PROTACs have been designed to cause ubiquitinoylation of p53 by an MDM2-inducing pathway. These ligands applied a derivative of the first small molecule shown to inhibit the MDM2-p53 interaction as part of their structure. In contrast, CRBN ligands have been found to efficiently and selectively degrade MDM2. However, ligand design has been limited by superior results of other E3 ligase based TPD [4], [6], [9], [13].

IAPs can be categorized into two subgroups, cellular IAPs (cIAP) and X-linked IAPs (XIAP). Both contain a RING domain as E3 ligases and are regulators of apoptosis due to their caspase inhibiting abilities. Several ligands have been developed based on their natural antagonist, second mitochondria-derived activator of caspases (SMAC). For XIAP initiating cell death SMAC ligands have been designed by mimicking endogenous peptides. Interestingly, some ligands have been reported to

induce auto-ubiquitinoylation of a special domain present in all IAPs and thus, degrading cIAPs themselves by proteasomes [6], [11], [13].

The first target of protein degradation by VHL described the protein Hif-1α as part of a co-crystal structure with this E3 ligase. By identification of a structural recognition site with a Proline (Pro) residue, first PROTAC ligands were designed that included the same epitope peptide motif. It was conjugated to a ligand structure that aimed to target the protein substrate. The early development of VHL ligands was based on peptidic structures that showed limitations and urged for non-peptidomimetic molecules able to induce substrate ubiquitinoylation. Nevertheless, Hif-1α-like ligands were able to penetrate through the cell and degrade there several POIs, including BRD4.
VHL is a E3 ligase forming a multiprotein complex consisting of Elongin B and C, cullin 2 and Rbx-1 and has been able to target substrates by VHL binders, which are usually around 5-7 peptides long (Figure 8). This was one disadvantage over PROTACs interacting with CRBN, which design had much smaller molecules in size and molecular weight [13]. Various research groups from academia and industry developed a vast number of non-peptidic ligands for the E3 ligase VHL. PROTACs inducing degradation of estrogen-related receptor-α and receptor-interacting serine/threonine-protein kinase-2 are considered as the first high quality, small molecule ligands for VHL and were reported independently and simultaneously by two different studies [6], [11], [13].
Before focusing on CRBN that has been the E3 ligase used in this work, it can be noted that E3 enzymes are implicated in a complex network of pathways, where they were found to act as both tumour suppressors and promoters. To design ligands that target selectively and efficiently one E3 ligase requires a thorough understanding of the structural, pharmacokinetic and pharmacodynamic properties. Only then a therapeutic intervention of E3 ligases that vary in their expressions in tissue- and tumour dependent way, can be applied to regulate possible substrates or POI degradation. In addition, the specific way of ubiquitinoylation of Lys that lead to efficient proteasomal degradation of marked proteins still requires more research [4].

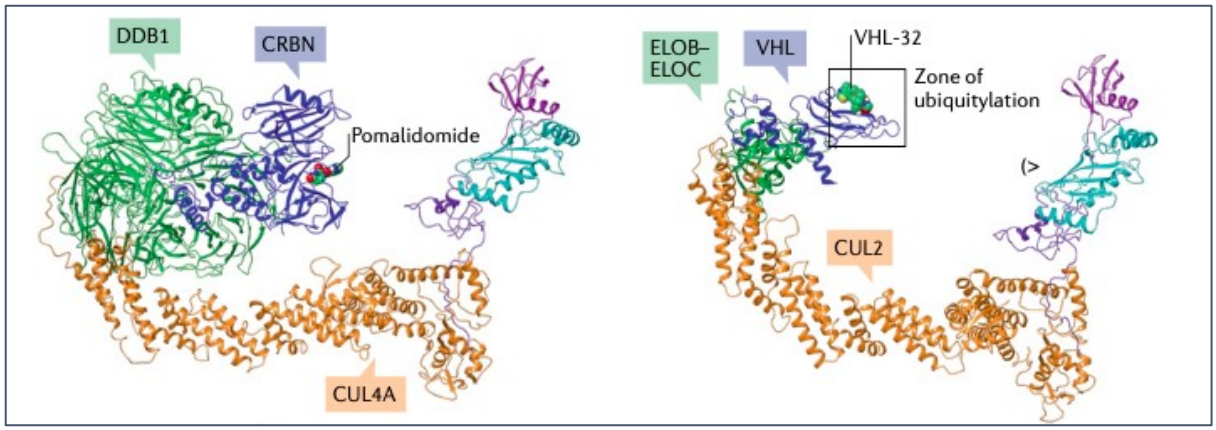


Figure 7 Structural examples of RING E3 ligases with adaptor proteins, protein of interest (POI) chains, small molecules in respective binding pockets. From right to left: CRBN ligase in a complex of damaged DNA binding protein (DDB1)-Cullin-4 (CUL4); VHL in complex with Elongin B/C (ELOB-ELOBC)-cullin 2 (CUL2)-Rbx-1 (adopted from [10])

4.2. E3 Ligase CRBN

4.2.1. CRBN Introduction

The protein CRBN in an E3 ligase 442 amino acids long and forms a similar degradation complex like the briefly discussed RING complex with VHL. The multimeric degradation complex is built of Cullin-4-RING E3 Ub ligase (CRL4), an adaptor protein called damaged DNA binding protein 1 (DDB1) and a RING finger protein ROC1 (Figure 7). The 50 kDa CRBN protein contains three domains: an N-terminal Lon protease- like domain (hereafter, Lon domain), an intermedial helical bundle, and a C-terminal thalidomide-binding domain (TBD). The latter part of CRBN harbours the conserved IMiD binding domain and is thus also called CULT (CRBN domain of unknown activity, binding cellular ligands and thalidomide) domain. It contains not only the binding pocket where the ligands such as thalidomide were first described to bind, but also where the so-called sensor loop is found. The sensor loop has been observed to interact with the ligands in the IMiD binding pocket and is structurally a beta insert hairpin [8], [14], [15].

As for any E3 enzyme, CRBN makes up the substrate specific part of the degradation complex and interacts with the substrates. After the Cullin4-ROC1 ligase module is formed CRBN gets recruited to the complex by DDB1. The adaptor protein consists of three beta propeller chains, where in between the helical bundle of CRBN docks. One domain of DDB1 interacts with Cullin-4 and positions the CRBN specific

substrates in a way so that ubiquitination can occur [14]. CRBN was only found after many years of focused research that aimed to understand how thalidomide caused one of the biggest medical tragedies in history and which factors make them induce teratogenicity. The IMiD compound was found to target very selectively the E3 ligase. Compounds were designed based on the three generations of IMiD ligands thalidomide, lenalidomide and pomalidomide. These small molecules and their derivatives were later described as “molecular superglues”. In general, compounds modulating with CRBN activity were applied initially for multiple myelomas (MM) and further used for other hematologic and solid tumours. Most often these derivatives of IMiDs keep including the glutarimide moiety while the rest of the structure varies [2]. For the three IMiDs lenalidomide, thalidomide and pomalidomide, it is known that they all bind to the IMiD binding pocket of CRBN, and all can induce the modifications on the transcription factors Ikaros and Aiolos by Ub signals. These molecules were the first MGs, even though this distinction was only described many years later. Furthermore, all three ligands are structurally very similar and share the interaction motives of their glutarimide moieties. It is the essential part of their IMiD structures that ensure not just interaction to the residues of the binding pocket but truly anchors them there. This is in contrast to the novel generations of IMiD like ligands that then vary in the rest of the structure to direct POI specificity. In addition, and of importance for the content of this work, lenalidomide can also induce specifically the ubiquitination of casein kinase 1α (CK1α) [2]. For a good overview of the vast diversity of ligands developed based on the three IMiDs the published review of Bircelj *et al* (2021) is recommendable [8].

Eventually, a new group of ligands was developed including heterobifunctional moieties but differ from MGs and IMiD compounds by their rational pairing of a glutarimide structure and a POI-targeting warhead moiety by a linker. These ligands are called PROTACs and they were designed with a specific target substrate in mind [10]. While among the endogenous substrates, CRBN induces adenosine monophosphate-activated protein kinase subunit a1, amyloid precursor protein, argonaute 2, c- Jun, calcium-activated potassium channel subunit a1, chloride channel protein 1 and glutamine synthetase [16].

4.2.2. Dynamics of CRBN

The TBD of CRBN contains a hydrophobic pocket where thalidomide was found to bind after first crystal structures were determined for the E3 ligase. So-called CELMoD (Cereblon E3 ligase modulatory drugs) ligands interact in this conserved side as well as to induce ubiquitinoylation of (neo)substrates and subsequent degradation. “Neosubstrates” are considered any proteins that are distinctively targeted by the ligand by which therapeutic interventions are rendered [14].
More than a dozen crystal structures of CRBN were published so far and the majority shows overall the same CRBN conformation. The experimentally determined closed conformation of CRBN has co-crystallized DDB1 in a complex that showed Lon and TBD domains of CRBN to interact with DDB1. Most often an IMiD or IMiD like ligand is bound. In contrast, only two publicly available structures of the open CRBN conformation have been experimentally solved. In that conformation Lon and TBD domain are separated by a 45° rotation creating an open palm like structure of CRBN [14]. This led the researchers of Watson *et al* (2022) to conclude that stable quaternary structures of CRBN and its complexed proteins are most often only possible to define and to observe upon ligand binding in the IMiD site as well as in presence with co-crystallized chains of DDB1 and substrates of the targeted degradation. Interestingly, not all CRBN structures have a sensor loop as part of the CULT domain that is defined by electron density. The sensor loop (residues from 341 to 361) has been proposed to directly interact with IMiD ligands in the binding pocket and with neosubstrates [14]. In addition, crystallographic data observed regions of the protein structure (residues from 210 to 220) that showed a high disorder and were difficult to capture as a structure. These two areas with high flexibility have been reported as characteristics of maintaining the open CRBN conformation. The question remains at which state ligands make the crucial interactions with CRBN to cause the switch between the open and closed conformation and what other conformations does CRBN assumes during this process [14] (Figure 9).
Watson *et al* (2022) could already observe that binding of ligands in the IMiD pocket such as that pomalidomide at various concentrations induces structural re-arrangements of CRBN to different extents. Due to the stabilizing effect of the

binding ligand in the pocket site, the closed conformation could form and was observable (Figure 10). Consequently, the sensor loop area was not found in a disordered state but supported the stable structure. Experiments showed that truncations at the N-terminus of the chain caused unstable and flexible loop states and rendering CRBN in an open conformation. The hypothesis predicts that efficacy of substrate ubiquitinoylation depends on the extent of successful conversion from the open to the closed CRBN conformers that is associated with ligand interaction. Increased ligand binding was reported to improve stabilization of the closed conformation and thereby enhance ubiquitination of substrates and degradation efficacy. Independent of the ligand and DDB1 interactions, the sensor loop as part of the IMiD binding pocket showed to be have great influence for the transitions between the two extremes of CRBN forms so far captured by crystal structures [14]. Interestingly, independent researchers have proposed this “unfolding” of the CRBN structure upon the loss of the IMiD ligand thalidomide, where among other primary findings also the tryptophan (Trp) residues were challenged as key interacting residues [15].

Obtaining different conformations of CRBN and the challenge to define certain areas by electron density data or other experimental means shows the flexibility of this protein. Often it is co-crystallized in a modulatory complex with proteins like DDB1, where the interfaces are determined in conformational states between the open and closed form. These findings also included the prediction of the “ubiquitination zone” that considers the extent of ubiquitinoylation of Lys residues as a geometrical issue [17], [18].

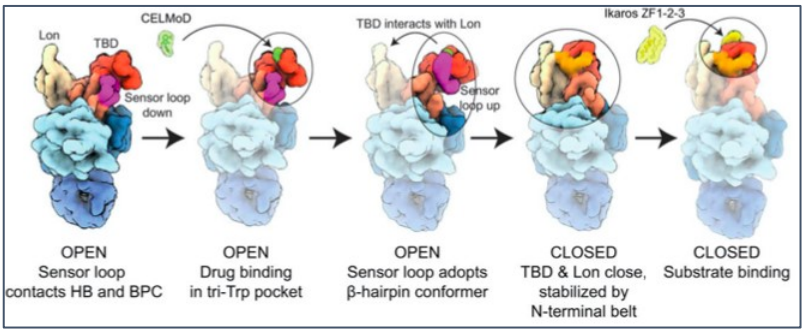


Figure 8 The question remains at which state ligands make the interactions to CRBN to cause the switch between the open and closed conformation and what other conformations does CRBN explore during this process. The figure represents the hypothesis based on the two conformational states captured and described by Watson *et al* (2022) [14].

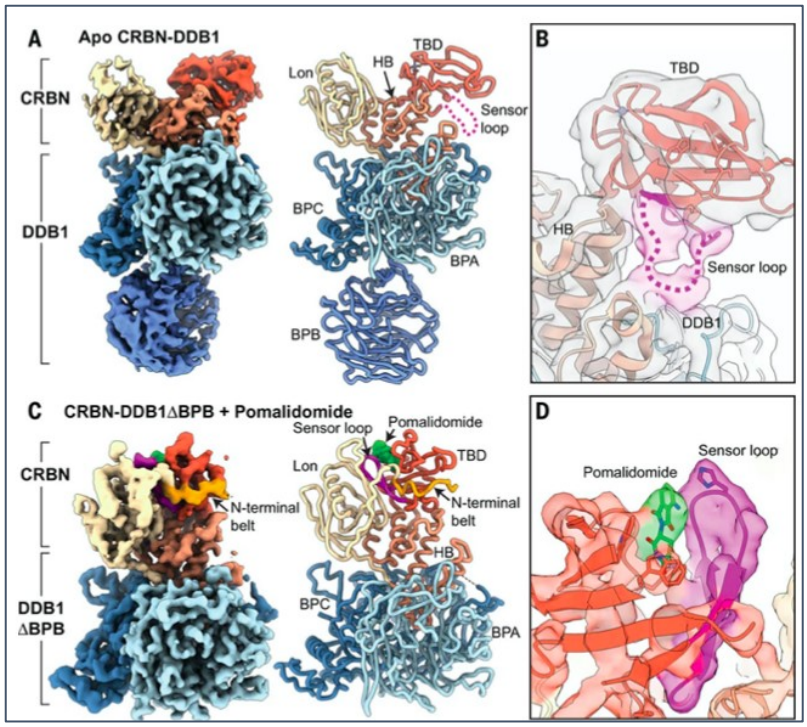


Figure 9 Watson *et al* (2022) could already observe that ligands binding to the IMiD pocket such as pomalidomide at various concentrations would induce conformation re-arrangement of CRBN to different extents. This fraction was reported to retain the closed conformation. The sensor loop area was not found in a disordered state but supported stable conformation [14].

4.3. PROTACs

In the past decade, two main types of these ligands inducing TPD have been developed and are called either molecular glues (MGs) or PROTACs. Their names are sometimes used synonymously, despite their structural distinctions. Namely, PROTACs consist of three building blocks: a warhead, a linker, and an anchor. In contrast MGs only consist of two structural elements, the warhead and anchor. The warheads are the part of the ligands that are interacting with the POI, while the anchors interact with the E3 ligases, like CRBN. The linker connects those two structural elements and can vary in length, as shortly described below in the case study of dBet1 and ARV-825. Such ligands are not based on any peptide like structures. In contrary, MGs do not include a linker structure but bring the POI and E3 ligase much closer together, expectedly increasing protein-protein interactions (PPIs), with the IMiD compounds being the structurally simplest and first ligands of this group. Another well-known MG is Mezigdomide (CC-92480). Overall, MGs and PROTACs have their hetero-bifunctional small molecule design in common [10], [11].

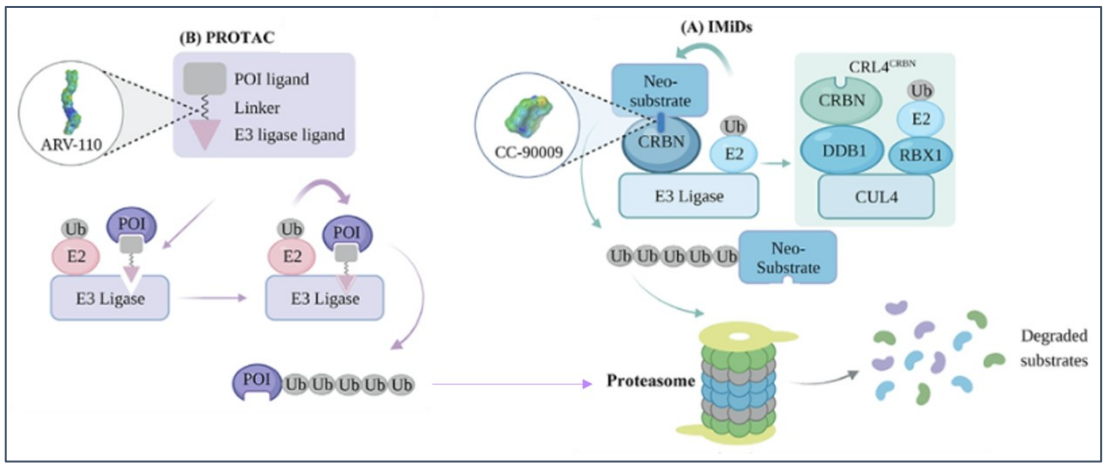


Figure 10 Schematic overview of the targeted protein degradation induced by (A) IMiDs and (B) PROTACs. (schematic figure adapted from [19])

In general, PROTACs have been of great interest as their mode of action allows them to bind to “undruggable” proteins and does not need an active or functional site in proteins [3], [10]. Small molecules acting as inhibitors are typically binding the active sites of receptors and enzymes and by that block the mechanism of action of these proteins. These ligands bind with high affinity to their targets, but also are not active unless the interaction is established. This type of mode of action is called “occupancy-driven pharmacology”. Consequently, the traditional drug design has left out many favourable targets for much needed therapeutic interventions. The research on PROTACs showed its mechanism of action is “event driven”. This means that these ligands bring the targeted POI and E3 ligase in closer proximity and are not used up after the reaction. By this process they act as catalysts. PROTAC interaction sites can be pockets of small binding molecules, from where ligand design continues to develop [3], [10], [19]. In addition, PROTACs bring the POIs and E3 ligase in close proximity, which favours PPIs that may also induce beneficial cooperativity to the complex interactions [3], [10]. It has been proposed that if the ligand was selective enough for a recruitment of a specific E3 ligase, the compound would not even require high affinity, but moderate binding affinity would suffice to associate the POI close to the binding pocket for degradation. Simultaneously, the same mode of action is a concern for drug toxicity and concentration [19]. This redefinition of druggable targets introduced a plethora of

possible substrates for therapeutic interventions and is the key advantage of PROTACs over other traditional small molecule designs [20].

While PROTAC-like acting ligands for VHL were comprised of several peptides and thus were quite large for E3 binders, CRBN ligands are more favourable as they are smaller in size and initially based on IMiD structures. VHL recruiting peptidic ligands showed limited cellular permeability and consequently, low degradation induction [11]. To ensure higher chances during ligand design to generate a molecular structure that has favourable properties, many researchers apply the infamous Rule of 5. Lipinski established the guidelines that describe properties for favourable pharmacokinetics considering the absorption, distribution, metabolism, and excretion ("ADME"). To gain good oral availability, these molecules should not have a molecular weight over 500 g/mol, include more than 5 H-bond donors or 10 H-bond acceptors and have a partition coefficient logP of less than 5 [20], [21], [19], [22]. However, PROTACs are still larger than what the Rule of 5 by Lipinski dictates, by which small molecules were proposed to have favourable properties in terms of oral availability and subsequently, predicted to be more successful in reaching the target substrate. Therefore, the chemical space of linker moieties has been proposed to be the most likely route for increase solubility. Linker lengths were varied as one possible solution to this issue and were observed to influence their properties. While lower molecular weights can be achieved by shorter linker sizes, linker moieties can also contribute to PROTAC conformations in solutions and degradation efficiency. Finally, these variations in ligand design can influence the zone of ubiquitination of POIs and the PPIs between POIs and E3 ligase. Nevertheless, PROTACs still go beyond the Rule of 5 principles. The modification and optimization processes of these compounds have been applied to different methodologies such as fragment-based drug design (FBDD) aiming to generate molecular structures with improved pharmacological properties. Further, different researcher groups have reported various methods on how to determine, compare and even predict efficiencies of PROTACs based on their structures and with that their properties. While certain aspects make them favourable over other strategies of small molecules, but the ligand design still requires improvement on certain challenging physiochemical properties and ideally, influence the targeting efficiency based on structure-activity relationship (SAR) [10], [23] [11], [13], [16], [20], [20].

Therapeutic modality	PROTACs/ molecular glues	Small-molecule inhibitors
Target scaffolding functions	✓	
Potential to treat undruggable proteins	✓	
Iterative mechanism of action	✓	
Broad tissue penetration	✓	
Orally bioavailable	✓	✓
Ease of manufacturing	✓	✓
Preclinical validation, proof-of-concept established	✓	✓
Clinical validation	✓ (Phase II for PROTACs)	✓

Features	PROTAC	IMiD
Representative structure		
MoA	Heterobifunctional molecules bind to E3 ligase and target	Molecules reshape the surface of CRBN to interact with neosubstrate
Molecular weight (g/mol)	700–1500	200–600
Route of administration	Oral, IV	Oral
Type of malignancy	Solid tumor and hematologic malignancy	Hematologic malignancy
Limitations	High molecular weight Low oral bioavailability Low brain penetration	Unpredictable target may induce unknown toxicity Limited rational design approaches
Advantages	Target scaffolding function Potential to treat undruggable targets	Low molecular weight Good oral bioavailability Binding pocket is not essential Potential to treat undruggable targets

Table 1 Left: Table showing the therapeutic modality of PROTACs/molecular glues vs. Small-molecule inhibitors showcasing the advantages of targeted protein degradation by application of PROTACs/MGs (adapted from [10]). Right: Table showing the different features and characteristics between PROTACs and IMiDs (adapted from [19]).

For rationally designed PROTACs and MGs, the first crystal structures of CRBN in 2014 helped to improve the understanding for the structural ligand development. Two X-ray structures showed the phthalimide ring of the co-crystalized thalidomide exposed to the environment towards the opening of the binding site and was proposed as a starting point for further linker designs or substituent structures [10], [11]. Over the last years, various research groups have provided CRBN structures with adaptor, helper proteins and even with co-crystallized POIs and ligands. These studies have allowed to analyse the interactions of IMiDs to the residues in the binding pocket. Their interaction pattern includes the so-called “tryptophan cage” or “aromatic cage” in the binding site, where observable H-bonds to Trp residues contribute to the stability of the closed CRBN conformation and are able to increase the degradation effect for neo-substrates [15]. These three specific aromatic residues have been found to be strictly conserved from humans to bacterial CRBN structures and are as such part of a universal ligand or recognition motif, as suggested by Heim *et al* (2022). In terms of the degree of CULT domain conservation, the root-mean-square deviation (RMSD) was determined between all human to bacterial CRBN structures with DDB1 chains and were found to vary less than 1 Å over 100 Cα positions [15].

While the degrader complex with CRBN has endogenous substrates that are targeted for degradation, IMiDs and other PROTAC like ligands are designed to induce degradation of new POIs. Compounds with highest affinities of several studies showed a common structure, all small molecules included a cyclic imide moiety (Table 2, Figure 15) [11], [24], [25]. In contrast, the glutarimide moiety of IMiDs was experimentally determined to bind to the hydrophobic cavity of the binding pocket [3]. In addition, the crystal structures of CRBN with POIs such as CK1α support the hypothesis that PPIs stabilize the complex further, a feature especially predicted for MGs. Miñarro-Lleonar *et al* (2023) established this finding with the H-bond disassociation energy profiles of various residues pairs at the CRBN-CK1α interface based on the same PDB structure used in this master thesis, 5FQD [11], [24], [25]. However, there is still an ongoing discussion whether PPI-mediated cooperativity or rather to which extent this cooperativity contributes to the effects of complex formation of CRBN with POIs and consequently, to the degradation of the (neo)substrates. As reported by Bai *et al* (2021), examples for cooperativity as well as non-cooperativity and evene anti-cooperativity have been published [18].

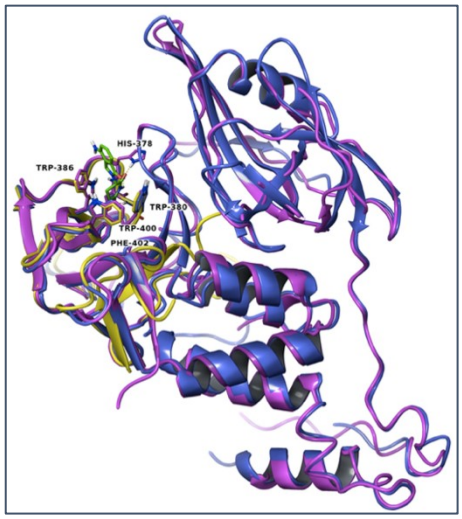


Figure 11 IMiD binding site residues His-378, Trp-380, Trp-386, and Phe-402 and ligands thalidomide (yellow), lenalidomide (green) and pomalidomide (blue) are shown as not the only conserved regions of CRBN as overlay almost perfect between E3 ligase of various species, including human (purple; PDB code 4TZ4), chicken (blue; PDB code 4CI2), mouse (yellow; PDB code 4TZ4) cereblon (adapted from [26]).

4.3.1. History of IMiDs

The history of IMiD drugs has not been filled only with promises of new drugs and successful treatments of cancers or various other diseases as mentioned before. Its evolution has its origins in one of the biggest pharmaceutical scandals of the last century, which consequences are still impacting today's regulatory and developments of drugs. In the 1950ies a compound was synthesized and was observed to be highly tolerable in mice at massive dosages. The compound thalidomide was a nonbarbiturate drug and was observed to be non-addictive. The producing German company Chemie-Grunenthal had a worldwide success with this new drug [7], [27]. Within a decade, it was a very popular drug prescribed as an antiemetic, especially in pregnant women to treat their morning sickness. Up until then, the world's best-selling drug was also intensively marketed as being completely safe and as such, was provided to an unprecedented number of patients without prescription (Figure 13). However, the thalidomide hype was soon over after an increasing and unpredictable series of cases appeared reporting infants with severe birth defects and malformations. Finally, in 1962 two independent researchers confirmed the common cause of these medical disaster, all mothers have been treated with thalidomide. This men-made tragedy of over 10,000 children born with the consequences of the thalidomide treated mothers has impacted the regulatory and procedure on drug testing like no other pharmaceutical scandal that far [7], [27]. For the first time, a difference between species in drug response and reaction was recognized, since mice were only later shown to be far less sensitive to thalidomide than any other non-human test animals such as primates and rabbits. Since the 1950s and 1960s the regulations on drug testing have changed drastically and become increasingly strict with numerous standards to be met before entering the clinical phase with humans [7].

Nevertheless, IMiD drugs have found new application areas to treat and become promising drugs again, such as in the treatment of leprosy shortly after the medical scandal of thalidomide. However, the truly bigger success was made with thalidomide when it was introduced as a treatment for patients with myeloma. It soon became a standard care for these cancer patients and was applied in combination with dexamethasone. Thalidomide was then released by the second generation of an IMiD drug called lenalidomide that was also widely used in off label therapies, next to

patients with refractory or relapse myeloma. A third-generation drug called pomalidomide took over by the 2010s for patients where treatments with either lenalidomide or the proteasome inhibitor bortezomib have failed before [27]. In addition, lenalidomide marketed as Revlimid was not only approved for multiple myeloma, but also found applications in treating other diseases such myelodysplastic syndrome, mantle cell lymphoma, diffuse large B cell lymphoma, chronic lymphoblastic leukaemia, and other lymphatic tumours as well as several solid tumours.

All IMiD drugs share a similar structure consisting of a phthalimide and glutarimide moiety but slightly differ in the functional groups. In context of the IMiD history, it needs to be noted that all three drugs thalidomide, lenalidomide and pomalidomide have teratogenic activities, because of which patients and providers of treatments with any these small molecules need to be registered in special programs taking care and supervision of risk evaluation and its mitigation strategy. In hindsight to the historical scandal of thalidomide, until today these means ensure no possible embryonic or foetal exposure to these drugs [27].

Figure 12 Thalidomide was marketed under the name Distaval in UK to pregnant women as a very safe sedative (adapted from [7]).

After the historical disaster of Grunenthal's drug, a great deal of research was invested to understand the mechanisms behind thalidomide's teratogenic effects. These efforts defined CRBN as the primary target of the teratogenicity. Its ubiquitous expression and ability to form the complex for protein degradation was part of the hypothesis aiming to explain the long-lasting effect and consequences of thalidomide exposure to unborn children. However, this simplistic view begins to fail when researchers tried to understand the mechanism of action as various studies have shown. As an example, numerous changes in expression of genes and proteins were detected, which are all intertwined and complicate the elucidation of the effects

of IMiD drugs diffuse the picture [7], [27]. Another hypothesis for thalidomide's teratogenicity led to the demonstration that the chirality of the racemic mixture during the production was the origin for the malformations in the unborn children (Figure 14). However, findings eventually showed teratogenic consequences of the (R)-enantiomer in rabbit models, and conversion between the two isomers was found to undergo at physiological conditions, too. Mori *et al* (2018) observed in their research a 10-fold higher interaction to CRBN with the *(S)*-enantiomer and a more relaxed glutarimide ring conformation over the *(R)*-isoform, while it induced greater teratogenic effects in zebra fish [28]. Similarities between the two structures in the IMiD binding site and their modes of interactions were reported as possible implications for the differences in binding affinity and therefore hypothesized to be at least one of the reasons for the induced teratogenicity.

An overview of the structures, side effects and indications for the three generations of IMiD drugs is provided in the table adapted from Holstein (2014) (Table 2) [27]. To sum up, IMiD and IMiD like drugs are still in application and under development for different disease but especially for cancer treatments and under continuously improving regulatory guidance [3].

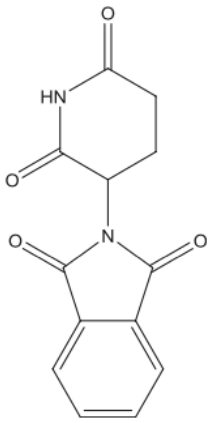
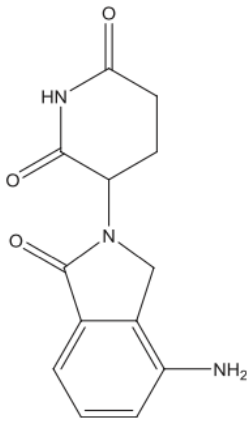
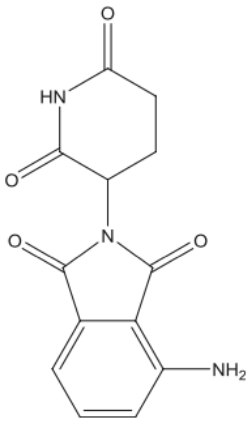
	Thalidomide	Lenalidomide	Pomalidomide
Structure			
Daily dosage (mg)	50–200	2.5–25	1–4
Half-life (h)	5.5–7.3	3–5 (Threefold increase with moderate/severe renal impairment)	7.5–9.5
Metabolism	Minimal	Minimal	CYP1A2 (major), CYP2C19 (minor), CYP2D6 (minor), and CYP3A4 (major)
Side effects	Sedation, constipation, neuropathy, and muscle weakness	Myelosuppression, fatigue, diarrhea/constipation, and muscle cramps	Myelosuppression, fatigue, and diarrhea/constipation
Nonmyeloma indications labeled by the US Food and Drug Administration	Erythema nodosum leprosum	Del-5q myelodysplastic syndrome and mantle cell lymphoma	—
Indications not labeled by the US Food and Drug Administration	Behcet's syndrome, lupus erythematosus, and systemic mastocytosis	Chronic lymphoid leukemia, light-chain amyloidosis, and myelofibrosis	Light-chain amyloidosis and myelofibrosis

Table 2 Overview of the three IMiD molecules showing common and differing pharmacological properties (adapted from [27])

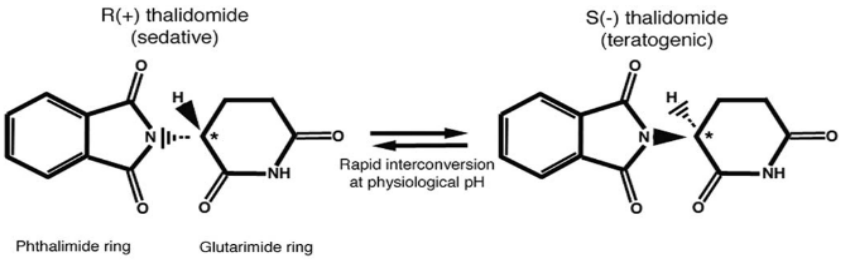


Figure 13 Thalidomide as a stereoisomer comes in two forms. Its production contained a racemic mixture with both enantiomers (left: (R+) vs right: (S-) isomer) and was reasoned to be one of the causes for the tragic consequences for the unborn children or patients taking the IMiD drug that was prescribed as a sedative to pregnant women. However, the conversion from isomer tot h other can happen during physiological conditions. (adapted from [7]).

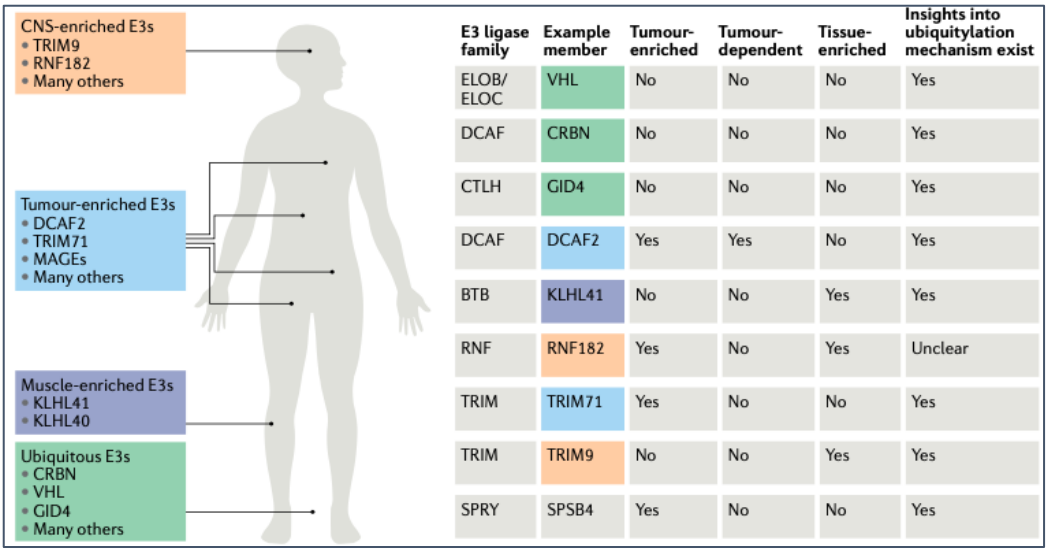
4.3.2. CRBN PROTACs

CRBN was described to have a plethora of substrates for ubiquitinoylation depending on the tissue and cell type (Figure 14). CRBN targets are involved in ion channel regulation, cancer development, and regulation of immunity, energy metabolism as well as cardiovascular and brain diseases. This great variety in

substrates is one of the reasons for the extensive research into PROTACs, MGs and similar ligands inducing CRBN specific protein degradation [29]. The first known ligands of CRBN were molecules such as thalidomide and other IMiD derivatives like lenalidomide and pomalidomide. This knowledge was used as a basis for the development of PROTACs that would interact in the same binding pocket such as those very first CRBN ligands. More than 30 proteins were successfully targeted as substrates by PROTACs not only increasing the range of therapeutic modalities of anticancer drugs but they are also used as drugs in immune disorders, neurodegenerative diseases and even against viruses [8].

Once IMiD drugs were shown to target CRBN as their main target in cells, various pathways were described and elucidated more and more parts of the complex mechanism of action that caused in the 1960s teratogenicity in the patients' unborn children treated with thalidomide. These research findings advanced the rational design and development on IMiD ligands aiming to degrade specific POIs in a CRBN dependent manner [3], [10]. In time, IMiDs were delivering first promising results on treating various cancers based on CRBN that was observed to degrade Ikaros proteins, specifically Ikaros zinc finger 1 and 3 (IKZF1/3). The small molecules based on IMiD like structures were developed due to their high selectivity and high affinity known for CRBN [3], [10].

As the first PROTAC targeting CRBN and inducing TPD of BRD4, the ligand was structurally comprised of pomalidomide and BRD4 inhibitor OTX015. It was shown to effectively degrade BRD4 in Burkitt's lymphoma cells. Next to the compound ARV-825, the pomalidomide based ligand dBet1 were also designed as BRD4 selective degraders (Figure 15). In addition, the comparison of these two small molecules showed significantly different intracellular potencies as PROTACs due to their structural differences in the linker moiety. dBet1 was observed to degrade 90% of BRD4 at 0.5 μM in contrast to the 1 nm concentration of ARV-825 [3]. These and similar findings increased the knowledge in the PROTAC field and provided in a decade various "PROTAC toolboxes" by various research groups. Approaches used included virtual screening (VS), high throughput screening (HTS), DNA encoded libraries (DELs) and fragment based ligand design (FBLD), just to mention a few [29], [13].



E3 ligase family	Example member	Tumour-enriched	Tumour-dependent	Tissue-enriched	Insights into ubiquitylation mechanism exist
ELOB/ ELOC	VHL	No	No	No	Yes
DCAF	CRBN	No	No	No	Yes
CTLH	GID4	No	No	No	Yes
DCAF	DCAF2	Yes	Yes	No	Yes
BTB	KLHL41	No	No	Yes	Yes
RNF	RNF182	Yes	No	Yes	Unclear
TRIM	TRIM71	Yes	No	No	Yes
TRIM	TRIM9	No	No	Yes	Yes
SPRY	SPSB4	Yes	No	No	Yes

Figure 14 Schematic representation of E3 ligases highlighting examples of ligases in the human body (adapted from [10]).

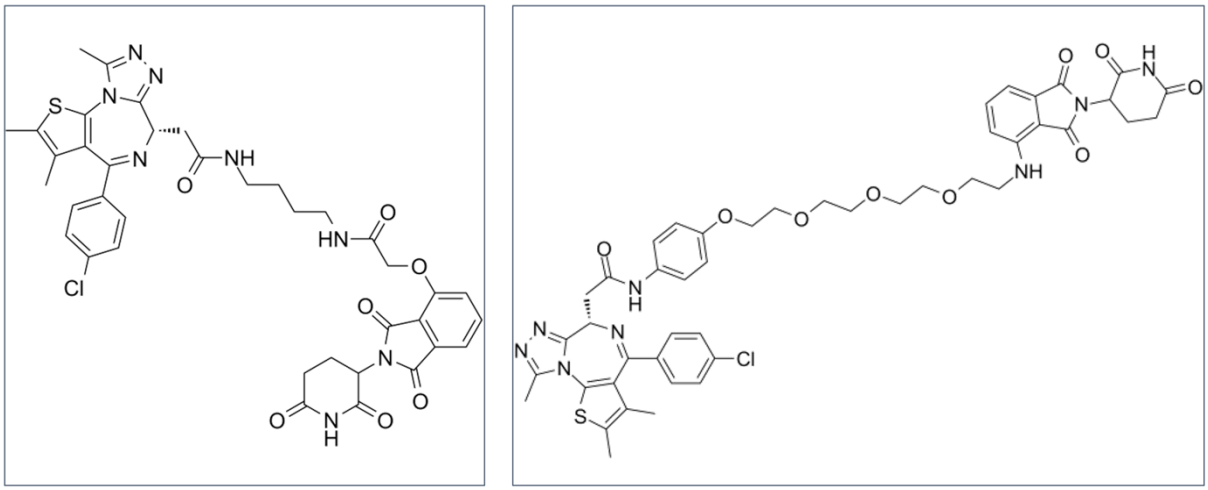


Figure 15 Molecular structures of two PROTACs that were observed to have varying pharmacological properties while only the linker length is different. Left: dBet1 vs Right: ARV-825 (adapted from [12])

4.3.2.1. Covalent ligands

The rational drug design of small ligands that are used for therapeutic interventions aims to improve in potency and selectivity. As PROTACs act by event driven pharmacology, as described above, reversible binders interacting with residues of the binding pocket were changed to irreversible ligands. Covalent ligands have been successfully designed for other indications, but little has been reported in terms of covalent PROTAC design [10], [30].

Developing ligands for the binding pocket in a defined space based on weak reversible structures poses as challenging. Fragment based drug design (FBDD) provides a promising starting point to design and circumvent certain E3 ligase PROTAC difficulties. In addition, using mass spectrometric techniques allows screening of electrophilic fragment libraries, which compounds can covalently bind to a residue and address challenging shallow pockets between PPIs sites. Most often Cys is the binding residue, such as generated fragments covalently indicating druggable residues of E3 ligases [13].

One of the fewer examples was described for a very specific E3 ligase that has not been studied as much as the E3 ligases VHL or CRBN. Kelch-like ECH associated protein 1 (KEAP1) is part of a degradation complex that induces mainly Nrf2 ubiquitination and degradation. PROTACs for the KEAP1 degrading complex have been developed by using bioactive compounds or by applying various methods such as HTS, VS or FBDD. This led to a ligand design, where the small molecule forms a covalent bond to a Cys residue from KEAP1 via a Michael acceptor moiety. The ligand subsequently recruits BRD4 to this E3 ligase. There are several potent inhibitors that bind covalently KEAP1 as PROTACs (Figure 16). This approach has been gaining increasing popularity and promised to bring new strategies for the "PROTAC toolbox". Initially, covalent warheads were screened against reactive Cys residues, but are now in development for all kinds of amino acid residues. This provides a different perspective on bringing POIs and E3 ligases together than traditional PROTACs have done so far. Overall, there has been only a handful of published literature on covalent ligands binding to E3 ligases [4], [11], [13].

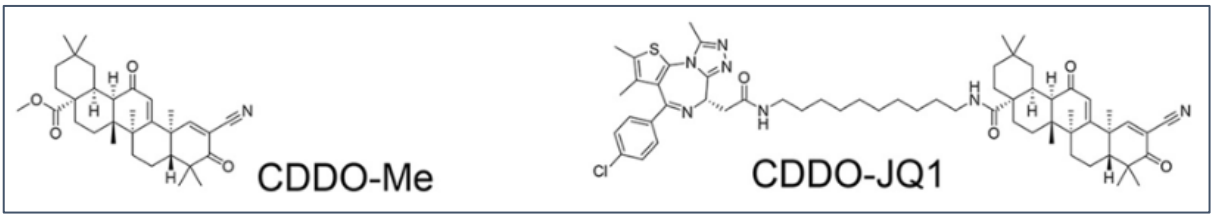


Figure 16 Structures from two PROTACs developed for the E3 ligase KEAP1 that are non-peptidic and were proposed to bind covalently to Cysteine residues of the E3 enzyme (adapted from [11])

However, Cys residues may not be the only residue to target for covalent ligands. Che and Jones (2022) highlighted the advances in sulfonyl exchange chemistry and its potential for site-selectively capturing residues for covalent bonds such as

tyrosine (Tyr), Lys, and serine (Ser) in protein pockets. Interestingly, though, data analysis of measured pKa values for ionizable residues in proteins have resulted in a newly predicted, powerful amino acid residue. Histidine (His) shows barely a pKa over 7.4 and is found to be located proximal to acidic residues. Unprotonated His residues can be targeted by electrophilic fragments made by nucleophilic chemical reactions of the side chain. In addition to their physiochemical potential at a physiological pH, His residues occur most commonly at catalytic sides of enzymes and appear 8.2 times higher than expected based on natural frequency [31]. More importantly for ligand design, their presence on protein surfaces is very low and shows one of the highest preferences to be within 4.5 Å of drug-like compounds in binding sites, which is twice higher than the expected natural abundance [31]. The reactions of probe labelling methods that were used to map amino acids in proteins, was used for the development of special ligands selectively binding residues. Recently, such a nucleophilic derivative comprised of a sulfonyl fluoride and IMiD like structure was reported to label a His residue in a protein binding site, next to other catalytic site residues. Based on this knowledge, the researches around Che and Jones (2022) developed a first covalent PROTAC binding to a His residue for the IMiD binding pocket of CRBN [31]. Based on their publications in the last year only two compounds were characterized as covalently binding PROTACs. One additional compound called EM12 is a derivative from lenalidomide and is the phthalimide that has been reported before on its physiochemical as well as biochemical implications for protein degradation with CRBN [31], [32]. In a previous study, EM12 was compared in its properties by an *in vitro* profiling of CRBN modulating drugs for lipophilicity, solubility, metabolism, permeability, intracellular bioavailability and cell-based potency that included thalidomide, lenalidomide and pomalidomide [32] (Table 3).

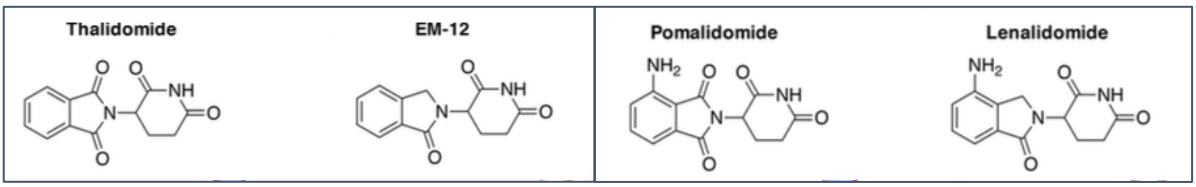


Figure 17 Electronic structure calculations for the three IMiDs and their derivative compound EM12 (adapted from [32])

	thalidomide	EM-12	pomalidomide	lenalidomide
molecular weight	258	244	273	259
cLogP	0.3	0.2	0.2	−0.5
Chrom LogD (pH 7.4)	0.900 ± 0.001	0.498 ± 0.001	0.787 ± 0.001	−0.058 ± 0.001
Caco-2 (10^{-6} cm/s)[b]	35.6 ± 1.6 (0.9)	22.2 ± 0.9 (0.99)	24.9 ± 2.0 (1.1)	1.45 ± 0.05 (1.7)
kinetic solubility (μM)	345 ± 47	451 ± 48	78 ± 17	461 ± 47
HLM $T_{1/2}$ (min)	>186	>186	>186	124
human plasma $T_{1/2}$ (min)	22	119	159	284
$f_{u,cell}$	0.34 ± 0.2	1.0[c]	0.90 ± 0.1	1.0[c]
K_p	8.9 ± 3.4	2.2 ± 0.1	4.5 ± 0.4	5.0 ± 0.4
F_{ic}	3.0 ± 0.68	2.2 ± 0.1	4.1 ± 0.04	5.0 ± 0.4
IKZF1 DC_{50} (nM)	>10 000	192.7 ± 52.5	26.7 ± 3.8	90 ± 15.2

Table 3 In vitro Profiles of the three IMiD molecules and their derivative EM12, including results for lipophilicity, solubility, metabolism, permeability, intracellular bioavailability and cell-based potency (adapted from [32])

There has been already reports for potent ligands and CRBN modulators against tumour activities that contained sulfonyl moieties. Among those, the small molecule AO6355 showed hydrogen bonding and π-π stacking with CRBN residues His353, His378 and the tryptophan triad Trp380, Trp386 and Trp400 [22]. In general, sulfonyl fluoride and fluorosulfate moieties have been shown as reactive warheads of small molecules that make the ligand site-selectively modify residues in proteins. Based on all rational designs of PROTACs and chemistry, respectively, two ligands have been reported as covalent ligands for the His residue His353 in the IMiD binding pocket and part of the sensor loop that was discussed previously as a key residue area stabilizing CRBN conformations. The lenalidomide derivative EM12 served as a starting point for combination with the sulfonyl fluoride chemistry that is known to covalently label His residues. Eventually, the two ligands called EM12FS and EM12SO2F were characterized and their labelling of His353 was confirmed by peptide mass spectrometry (MS) [33].

EM12SO2F is the more potent ligand binder of the two covalent ligands of CRBN in cells and showed adduct formation as predicted. It was also found to inhibit CRBN function in native cells by the absence of IKZF1 degradation. In comparison, EM12FS forms a weaker covalent bond to CRBN as it only partially labelled His353. However, it showed greater stability in human plasma, microsomes, and hepatocytes over EM12 and IMiD ligands lenalidomide and pomalidomide (Figure 18). These results were not challenged by the properties of EM12SO2F. First experiments also pointed out that the subtle structural distinction between the two covalent ligands already allowed detection of differences in complex formation with its POI, affected ubiquitination and thus, substrate degradation [31], [33]. However important

metabolic stability as a necessary requirement is, covalent ligands can have advantages in their pharmacokinetics and pharmacodynamics over reversibly interacting molecules. Although EM12FS has a high stability, optimizing its equilibrium binding and modifying the intrinsic reactivity of the warhead could provide overall more favourable profile for this ligand that was observed to degrade proteins. Even though the difference lies in only one oxygen atom, EM12SO2F induced barely any degradation [31], [33].

While these two ligands provide a great promise in advancing the “PROTAC toolbox” and improving the design of efficient and POI-selective CRBN compounds, there is no crystal structure captured yet that would confirm experimentally the covalent link. So far only MS data propose the covalent bond of EM12FS and EM12SO2F.

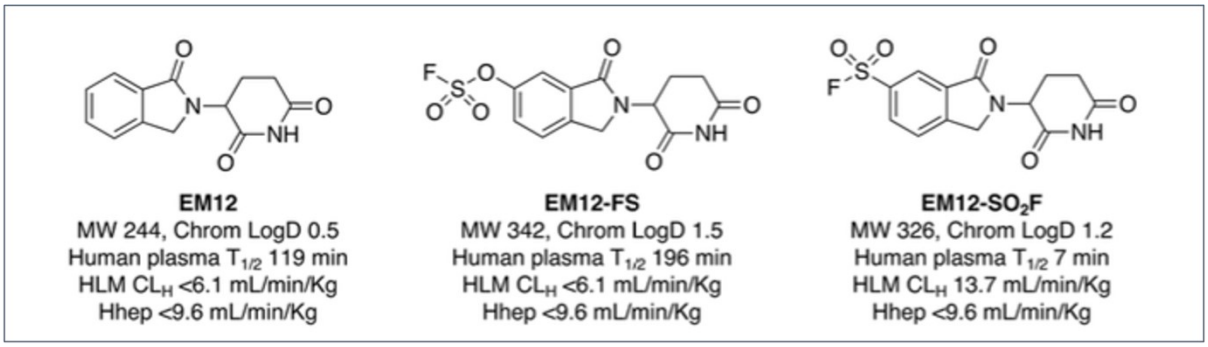


Figure 18 Molecular structures of the IMiD derivative EM12 and the two covalent ligands EM12FS and EM12SO2F (from left to right), including individually depicted molecular weights (MW), logD values and results for stability in human plasma, microsomes (HLM), and hepatocytes (Hhep) (from top to bottom) (adapted from [31]).

4.3.2.2. Covalent Docking in the IMiD binding site

In this master thesis, an *in-silico* method was applied to predict plausible poses in the binding site of CRBN. Docking is a computational approach that assesses potential binding conformations of small molecules in protein binding pockets. Docking results aim to reduce the conformational search space and in so provide helpful guidance to reduce time, labour, and resource consuming *in vitro* and eventually, *in vivo* experiments. Different computational programs apply individual calculations for retrieving best poses acting as surrogate experiments. Often considerations such as poses with lowest energy or conformations avoiding sterical clashes, generate several possible outcomes [34].

Molecular docking is used to predict best conformations of ligands in the binding site but also estimate the binding affinity using different scoring functions. Reducing the steric aspect of a scoring function to simulate protein flexibility, relaxation methods and varying the stable protein backbones and flexible side chains are only a few consideration that can be implemented in docking workflows [35] (Figure 19).

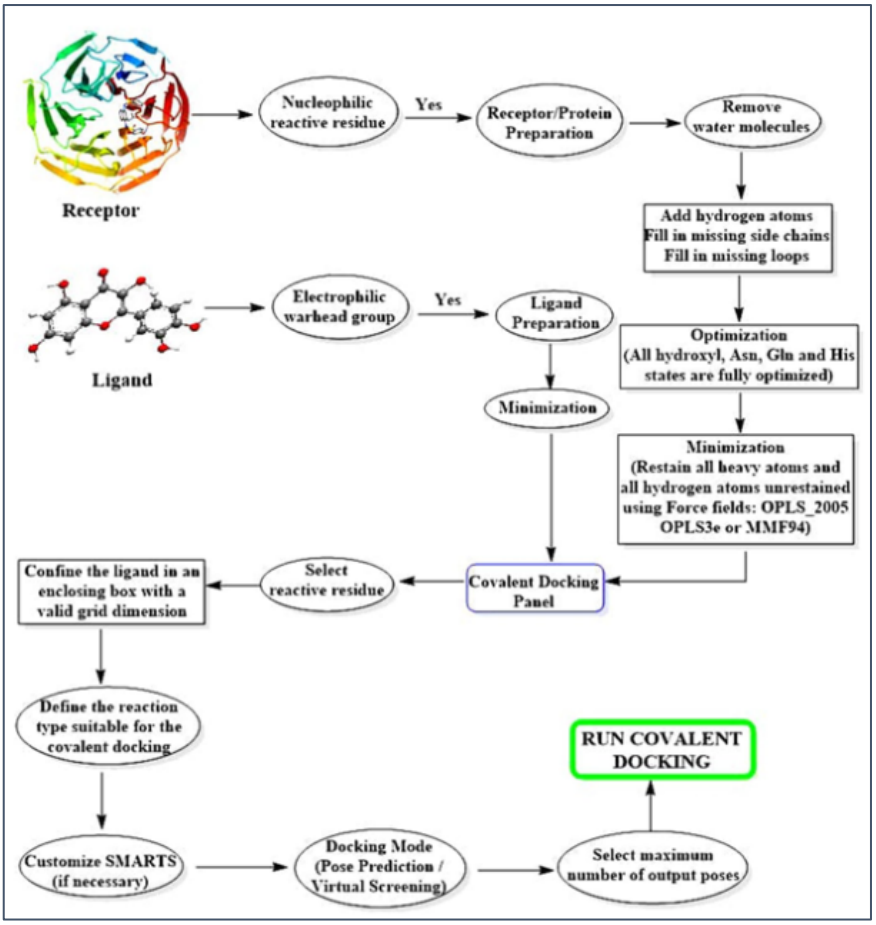


Figure 19 Overview of typical covalent docking workflow that was to a greater extent followed (adapted from [36]

In this work, His353 was used to covalently dock the two ligands EM12FS and EM12SO2F using the tool “GOLD” and to find plausible conformations for both covalent ligands in the IMiD binding pocket. It is based on a genetic algorithms that generates a covalent link by super-positioning a “link atom” of both the ligand and protein structure to yield post-docking ligand conformations in the binding pocket area. The link atom of the ligand forces the covalent link to fit onto the link atom of the protein and the bending energy term uses generated by the applied parameters. A proper geometry of the covalent bond is eventually verified by the addition of the fitness docking score to the energy term. Further, non-covalent interactions such as

H-bond formations are considered in the process, too, for the generation of plausible poses [37]. GOLD has been applied in other case studies with covalent ligands, including targets involved in the UPS pathway. The briefly discussed ligand MLN4924 was used for docking experiments that mapped interactions and finally led to new NAE covalent inhibitors [36]. Here, covalent docking with GOLD was applied to understand the binding mode and possible interactions that have not been characterized yet and provide guidance for new ligand design, while rationally evaluating His353 as the true amino acid residue for the covalent linkage of the two IMiD derivatives with fluorosulfate moieties.

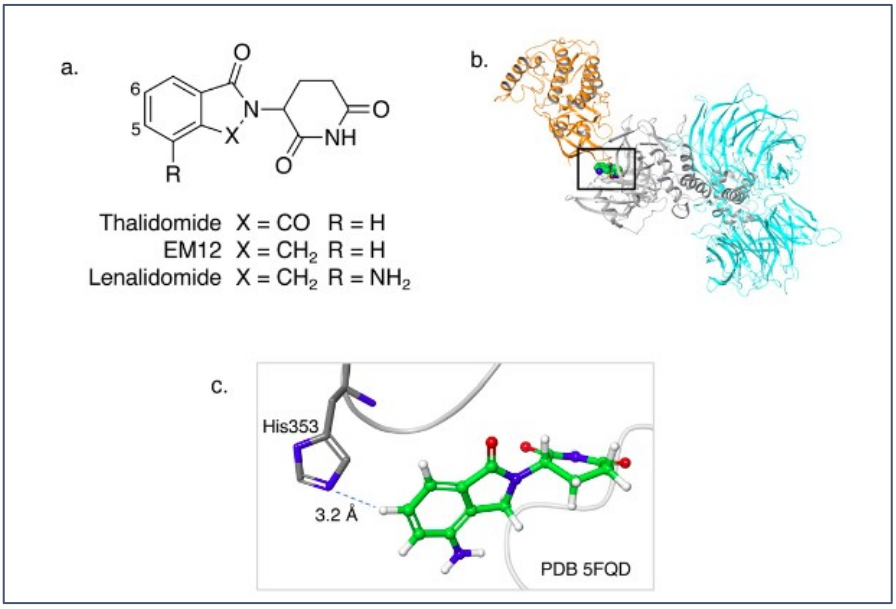


Figure 20 (a) Structure for thalidomide, EM12 and lenalidomide shown that vary only in the functional groups of X and R. (b) crystal structure of CRBN (grey) with CK1α (gold), DDB1 (cyan) and lenalidomide (green) in the binding pocket. (c) Zoom into binding pocket with specifically depicting residue His353 and lenalidomide showing the distance of 3.2 Å (adapted from [33]).

4.3.3. Studying ligand interactions to CRBN

Most MGs inducing CRBN driven protein degradation have been based on thalidomide, such as CC-885 and CC-90009. New developments based on IMiD structures included the replacement of phthalimide with a new ring moiety and the addition of new functional groups, such as aromatic moieties. These developments in ligand design introduce new interactions with the IMiD binding pocket residues [38]. The development of the good ligand design was based also on dynamics

studies in recent literature that would include more information than the experimentally observable interactions between co-crystallizable ligands and captured rigid CRBN conformational states. Yan and Zheng (2023) report on the interactions between binding pocket residues of their simulated CRBN and IMiD ligands. Comparable to findings based on crystal structures of CRBN, the glutarimide part of the IMiDs was found essential for anchoring ligands in the hydrophobic pocket. Residue glutamic acid (Glu) 377 was included in their considerations for the development of a IMiD like ligand and addition of a basic group providing a salt bridge interaction [38].

One advantage of such ligand interaction studies includes that initially *in silico* tools can provide useful information and guide towards a more successful ligand design with guided optimizations of ligand properties. Molecular dynamics (MD) simulations have been a powerful tool and are increasingly applied for prediction and analysis of biological macromolecules, such as proteins, and small molecules. Like the research reported in Yan and Zheng (2023), protein-ligand complexes were set up for MD simulations with CRBN and IMiD compounds to gain more insights into possible interaction sites. There are a few basic analysis that can be calculated from the generated MD trajectories, but different case studies include different analysis workflows. Here are two plot examples for basic analysis of an MD trajectory: The root-mean-square deviation (RMSD) are the calculated deviations of the various conformations from an MD trajectory compared against the first set of coordinates of the $C\alpha$ atoms. RMSF provides valuable insights into the flexibility of the individual amino acids of the protein. In addition for the identification of key interaction residues, often an analysis of H-bonds between the protein and ligand are conducted [38].

4.3.3.1. MD simulations with CRBN

Comprehensive *in silico* methods like MD simulations have supported at many different instances the discovery and development of drug design and started to guide and aide the understanding of PROTACs and CRBN dynamics, too [11], [22]. Approaches like MD simulations help to understand and ideally capture all or most of the intrinsic switches of proteins. However, to simulate distinctively different

conformations of a protein under native conditions and their transitions, it requires deep knowledge to generate the motion and mobility of the molecules and atoms in a system. Initially, the captured rigid structures by X-ray crystallography or Cryo-EM are used and ideally, the protein has been found in different conformational states. Most often, when certain regions of proteins are too flexible, these residues cannot be experimentally defined and are modelled in, if previous homologous structures are available. In parallel, NMR spectroscopy can provide information on protein dynamics within a structural ensemble, but the method is limited to a certain protein size [39]. Regardless of the advances in computational power, MD simulations can highlight the mobility of biological and chemical systems on an atomic level from femto- to milliseconds [40]. In this master thesis, a collection of scripts and tools of the “AMBER” suite has been applied to set up MD simulations and study the lenalidomide interactions to CRBN. MD simulations generate trajectories of atomic positions as a function of time and produce energetically accessible conformations of a system. Thus, trajectories are large amounts of data, for which analysis tools have been developed. Overall, MD simulations have been used for gaining an understanding of protein folding, drug-receptor interaction, and fast time scale motions of molecules [41]. For a detailed overview of the content that is included in “AMBER”, Salomon-Ferrer *et al* (2012) provides a good insight into the process and analysis possibilities of MD simulations [40].

Recently, few studies applied MD simulations to capture possible interactions of ligands with CRBN binding pockets [26], [36], [38]. In one of these published reports a handful of ligands have been used for docking experiments and its most probable conformations of the ligands were applied further for MD simulations. This was carried out in parallel to the IMiD ligand lenalidomide [38]. In addition, lenalidomide interactions to CRBN residues and CK1α were explored by MD simulations to verify by *in silico* findings the stabilization effect of the H-bonds between the two proteins in the complex. In that study, PPIs were assessed to provide evidence for the hydrophobic shielding of lenalidomide that subsequently, increases stability of the two-protein complex. The study applied the same PDB based CRBN structure like in this master thesis, but simulated the interactions for only a timeframe of 2 fs by a steered MD (SMD) protocol [25]. This MD simulation differs in the manner how atoms are allowed to move by the calculations for the system. In an SMD simulation,

a force is applied to certain atoms while other atoms are fixed. During the time of the simulation, the atoms unwind and unfold under the defined conditions and thus, their behaviour can be studied [42].
Here, the work focused on the IMiD lenalidomide, but the systems used as an initial protein structure both the open and closed CRBN conformational states for the rationalization of lenalidomide interactions over a simulated time of 1,000 ns and ideally, capturing new residue interactions by exploring and including all distinctive conformations per frame.

4.3.3.2. Application of a hotspot analysis based on MD trajectories

MD simulations have been applied to sample various protein conformations in the modelled aqueous environment and to try and address new residue interactions for guiding ligand design and to ensure more stable interactions. However, MD trajectories contain a large amount of data, for which only little analysis can be conducted in a simple manner and often require a deep understanding of the script collections available in the software package of “AMBER” and similar tools. Alternatives for such analysis incorporate the pharmacophore concept. Pharmacophore models can come from ligand protein complexes, which are based on a set of ligands with known activities that can be used for ligand-based VS [35]. Pharmacophores are defined as an ensemble of steric and electronic features that make up the optimal supramolecular interactions with a specific biological substrate structure and eventually to induce its biological response. These features include among others H-bond acceptors (HBA), H-bond donors (HBD), positive and negative ionizable groups (PI/NI), hydrophobic regions (H-H) and aromatic rings (AR) [35], [43]. It is important to consider that pharmacophore models are dependent on atomic coordinates of the system molecules and can over- or underestimate the features as all coordinates are retrieved from the structures provided in protein data bases, such as the PDB. In addition, the structures represent one rigid conformation of a protein and thus do not include features caused by protein dynamics. The pharmacophore concept can be extended when applied to sampled conformations of an MD trajectory. Some studies used individually determined conformational states from MD simulations to calculate pharmacophoric features. However, hotspot analysis can include the merged structural information in trajectories to generate pharmacophore

features based on the dynamic protein data. This was proposed to advance ligand design by prioritization of features, which incorporation could improve ligand interactions. In this way, features that appear often during an MD simulation but are not present in rigid PDB structures would not be omitted but could rather be considered in ligand design [44].

Such considerations were incorporated in the method GRAIL (GRids of phArmacophore interaction fieLds) [45] (Figure 21). GRAIL applies a traditional grid approach for the identification of interaction sites, while using the pharmacophore concept to generate grids for numerical analysis and eventually, an abstract visualization of the features for a target system. This hotspot analysis method uses grids, which are volumetric areas covering the region of interest. This space is often the binding pocket of the target protein [45]. Initially, the GRID program was developed by Goodford and each map describes the interaction energy between a given probe molecule and the protein at each point of a regular grid, based on a classical empirical energy function. The probe to target interaction with energies below a certain threshold and within the area of a grid can be visually displayed by corresponding iso-surfaces. In addition, the generated 3D grids can be also used for various other numerical operations. Regions in the grid with negative energies are favourable for the individual pharmacophore probe, while those with positive energies indicate areas with repelling forces to the probe. In contrast to other grid-based methods, GRAIL includes the calculation of interaction scores for an array of grid points in a defined region and produces maps of pharmacophoric features of defined types. The calculated scores are representative of how well geometric conditions such as angle ranges and allowed distances are fulfilled upon the probe set at a certain grid point. It rationalizes the extent how well a probe feature interacts with complementary features in a protein environment [45]. Eventually, graphical information generated by GRAIL for MD simulated data can be shown by importing the necessary structures and calculated grids into programs such as LigandScout [45]. The generated pharmacophore interactions fields ideally guide the development of ligand design and to identify new beneficial moieties as well as to improve the molecular structure that leads to an optimized compounds binding the target protein. This master thesis uses the GRAIL tool to highlight potential interaction sites with distinct pharmacophore features. As a special feature, GRAIL can use all relevant grid maps from the large coordinate input of an MD trajectory to assess conformation

dependent characteristics of key interactions over time and not just rigid protein structures [43]. Therefore, the generated MD trajectories of CRBN in a closed and open state interacting with lenalidomide were the basis for this hotspot analysis with the grid covering the residues of the CULT domain that were proposed to be essential for the switch of the two conformational states.

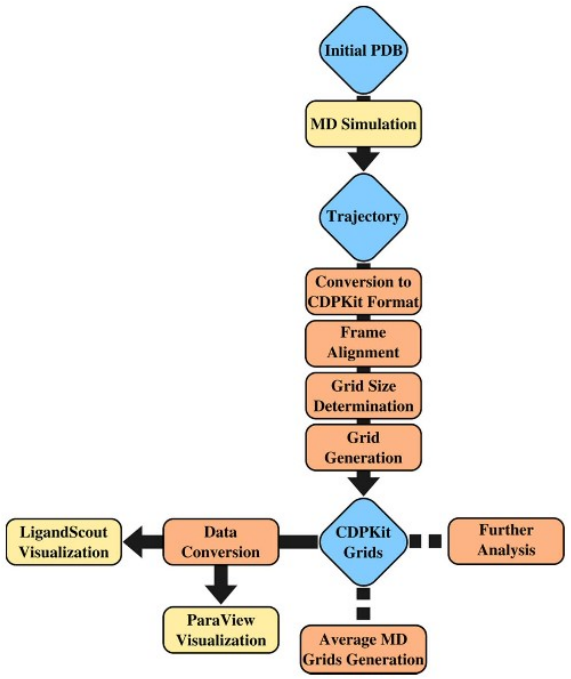


Figure 21 Schematic overview of a typical workflow starting with a PDB structure as the basis for an MD simulation and using the generated trajectory with all conformations of the simulated molecules for the calculations of the hotspot analysis called GRAIL that includes the pharmacophore concept to display pharmacophore fields in a defined grid (adapted from [43]).

5. Methods

5.1. Overview of Workflow

This Master thesis has two pillars, each with their own method aiming to answer the discussed questions on ligand interaction influencing CRBN protein conformations (Figure 22). One topic used covalent docking to assess potential interaction of residues in the binding pocket of two previously described ligands that were shown experimentally by mass spectrum data to be covalently bound to sensor loop of the protein. This was carried out under varying conditions. Covalent links for the different nitrogen atoms in the Histidine residue imidazole ring and various tautomeric forms of the side chain were tested as well as a water molecule was included in the dockings. Eventually, all docked poses were filtered according to several aspects as discussed below in more detail to generate the most plausible poses for the ligands. The second part of the thesis consisted of setting up and running MD simulations with CRBN. Again, different conformations of the CRBN protein were applied and their trajectories were analysed. In addition to the interactions of the ligand in the binding pocket with the residues, fluctuations per residue were assessed. To gain more insights on potential interactions in the binding pocket a grid-based approach called GRAIL was applied, which includes the pharmacophore concept [43]. Overall, the thesis aimed to understand the dynamics of the CRBN protein and the ligand interactions in the IMiD binding pocket. Considering that experimental data showing only the open and closed state of the protein, little is reported about the open CRBN conformation, which can be also confirmed by only the handful of open structures in the public protein databases. Further, the role of the sensor loop in the vicinity of the binding pocket was inspected. Subsequently, docking and GRAIL data shed some light on favourable and untapped interactions that would advance ligand design in the future.

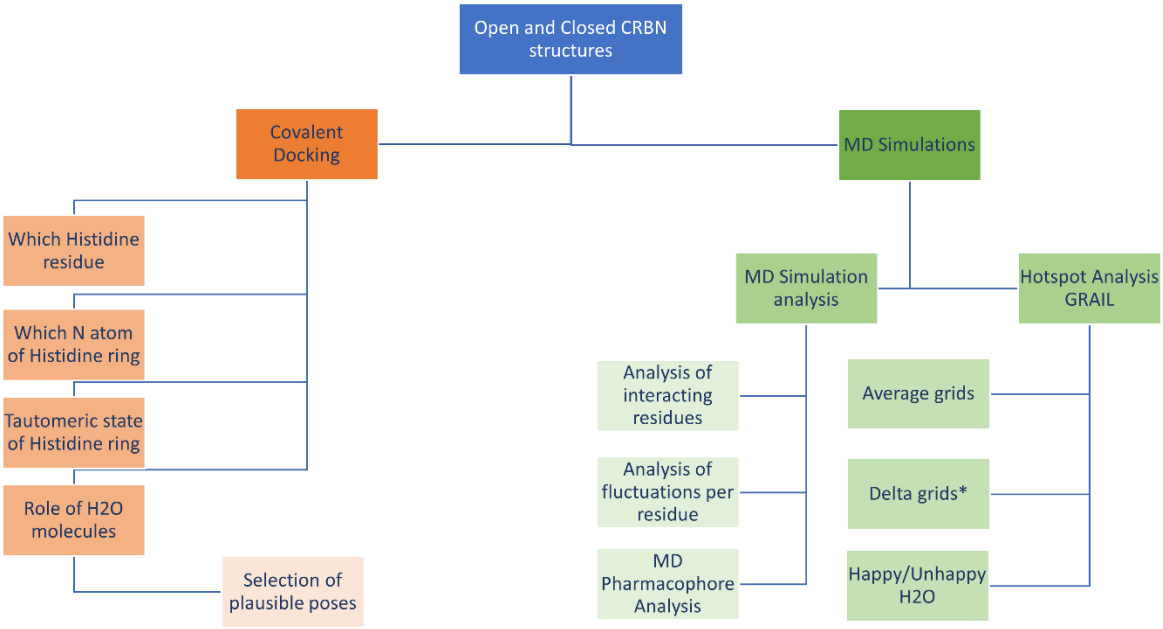


Figure 22 Overview of the workflow included in this master thesis.
Two main pillars with one introducing Covalent Docking and the other MD simulations aiming to assess possible interaction of covalent ligands to CRBN and the dynamics of CRBN conformers, respectively.

5.2. Structural Overview of CRBN

An analysis of published structures revealed that 26 human experimental structures with CRBN are reported to the PDB database, to the best of our knowledge at the beginning of this work. Furthermore, 26 other structures were found from different species, which divide into three from *Gallus gallus*, 16 from *Magnetospirillum gryphiswaldense* MSR-1 and seven from *Mus musculus*. All these structures were inspected regarding their resolution, their chain lengths and experimentally defined presence of the His353 in the sensor loop (a table with an overview is provided in the result section, Table 4). In more detail, the collected structures were filtered on the selection whether the sensor loop had been experimentally defined by electron density. Finally, the human structures were

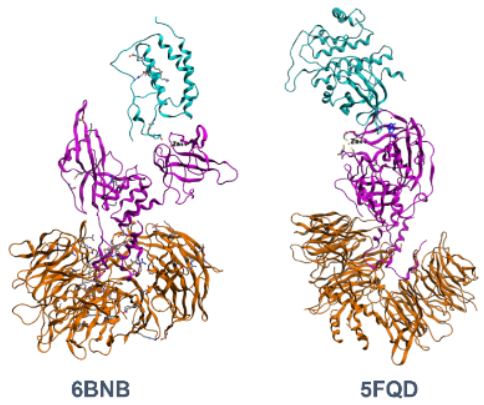


Figure 23 Two PDB structures shown used in this master thesis and coloured by chain ID with DDB1 (orange), CRBN (violet) and protein of interest (light blue). Left: open CRBN state (6BNB); Right: closed CRBN state (5FQD);

categorized into two main groups depending on their conformation of the CRBN protein. A visual inspection of all structures and their characterizations were made in MOE after alignment of the sensor loops.
In respect to the aims of this work that were started with the findings and covalent ligands reported by Jones and Che (2022), one open and one closed conformation of the CRBN protein were chosen as the suitable structures. 6BNB and 5FQD will be from here on be the mainly discussed and used structures as representatives for open and closed conformations, respectively (Figure 23). The structure 6BNB is defined by the X-Ray method at a resolution of 6.34 Å. The PDB entry consists in addition to CRBN, of DDB1 and BRD4. The other x-ray defined structure 5FQD has the ligand lenalidomide in the IMiD binding pocket co-crystallized. The other chains in the PDB entry include DDB1 and CK1α with an overall at a resolution of 2.45 Å.

5.3. Covalent Docking

As before any docking procedure, 5FQD and 6BNB structures were imported from the PDB and prepared in MOE by the QuickPrep workflow. The panel is comprised of a list of considerations to prepare structures that include deletion of distant solvents, addition of hydrogens and to perform calculations on charges and initial refinements. Endpoints were chosen to be neutralized and empty residues that are defined by sequence only, were preserved. The panel uses the Protonate3D methodology to define ionization states and the addition of hydrogens in addition to the default settings that allow terminal amides, sulfonamides, and imidazole groups to flip and to optimize the hydrogen bond network [44], [45]. For structures with a ligand, atoms further than 8 Å away from any ligand atom are fixed, while hydrogens close to the binding pocket should not be excluded and able to move unrestrained within 4.5 Å. For further details on the QuickPrep Panel and its applied default settings, the MOE Handbook provides many more details [44], [45].
The PDB derived 6BNB structure has no ligand co-crystallized and required modelling of the ligand from 5FQD. Both PDB structures were imported into MOE and initially aligned by sequence, along their common entry chain of DDB1 and subsequently superposed. In a second step, the sensor loop region was aligned and superposed. Finally, tagging of the ligand to 6BNB and removal of the remaining

5FQD structure left 6BNB with the modelled lenalidomide molecule to be energy minimized. Here, it can be noted that for all work carried out in MOE the default Amber10:EHT force field was applied [44], [45].

Files were saved in “mol2” format, which enabled the further workflow with the docking program GOLD. Among many available programs, covalent docking was conducted with GOLD, a genetic algorithm-based approach for docking of flexible ligands into binding sites [37]. GOLD based covalent docking was guided by the wizard panel of the Hermes visualizer, which allows the preparation of all necessary files and their structures in a stepwise manner. After import of the prepared structures, either 5FQD or 6BNB, with the ligand, missing hydrogens were added and protonation states of His residues as well as Asparagine (Asn)/Glutamine (Gln) flips were defined. All docking runs included the addition of hydrogens but varied in their settings of His tautomerization, as detailed below in table 5. Similarly, water molecules from the structure were all deleted, unless otherwise specified, as with the water molecule HOH2037 closed to the binding pocket (Table 5, Figure 33). Furthermore, the ligand extraction was necessary to define the binding pocket of the to be docked ligands as they are automatically re-loaded for defining the binding site. In both cases, 5FQD and 6BNB files combined the protein structure and a co-crystallized or modelled-in lenalidomide in the respective pocket. The extracted ligand was saved in a mol2 file. The protein settings for the docking procedure defined the binding site by extraction of the ligand and all protein atoms within an 8 Å radius were used for this definition of the pocket. The ligands were prepared as mol2 files as well and were changed in a way that a part of the substructure modified the original compounds [37]. This modification was carried out in MOE and energy minimized before the mol2-files were imported into the GOLD docking procedure [44]. In this thesis, two ligands called EM12FS and EM12SO2F were of interest that were shown to be covalent binders to CRBN in the IMiD binding pocket [46].

The wizard panel offers many variations but to define the covalent link by the protein and ligand atoms was the key feature for the discussed dockings. The algorithm connected the ligand with the specified atom in the protein by providing a substructure of the linkage. These atoms of ligands and protein structures were defined and are available in the GOLD conf file that can be used to load the wizard panel and adjust further for the next docking run. All other settings, including scoring functions, constraints and ligand flexibility were applied with the default.

Next to the conf file, all output structures of the dockings and pose solutions were saved as mol2 files. These allowed further processing of individual poses in MOE. All ligand poses were energy minimized with the covalent linkage to His353 established. Interactions with residues from the binding pocket were assessed by MOE's ligand interaction tool and were extracted for further differential analysis.

All poses were subjected to one energy minimization step prior to ligand-residue interaction analysis. This was sufficient to gain a visually suitable torsion. All covalently docked poses were filtered accordingly to two main criteria. Initially, a visual inspection of the ligand poses, and the bond geometry was carried out and compared to the original co-crystallized ligand. Further, the atoms of the ligand and the substructure of the individual poses were exported and subjected to the Rarey Torsion Scan tool [47]. The substructure included part of the His side chain and was applied during the GOLD docking procedure (Figure 24). The curated SMARTS library is a collection of torsion motifs that are associated with torsion angle distributions based on published crystallographic data. The automated output of preferred angles is based on the torsion rules and is visualized according to the scheme of a traffic light: green torsions are considered relaxed, yellow ones tolerable and red ones strained (Figure 25). In addition to the aspects of the visual inspection, all poses with a minimum of one red torsion angle were excluded. Based on the left poses, further analysis regarding residue interactions was conducted.

Covalent docking runs were carried out with both the open and closed conformations of the CRBN. In more detail, the dockings differed in their settings. The conditions varied in their tautomeric state of the imidazole ring, the nitrogen atom from which the covalent link was built and the His residue it was made with. To verify the experimentally determined His353 as the main residue for the covalent link, all other His side chains close to the binding pocket were also investigated. This included His357, His378 and His397. Along these lines, two docking runs also included a specified water molecule that was found very close to the opening of the binding site. An overview of all docking runs, and their variations is provided in Table 5.

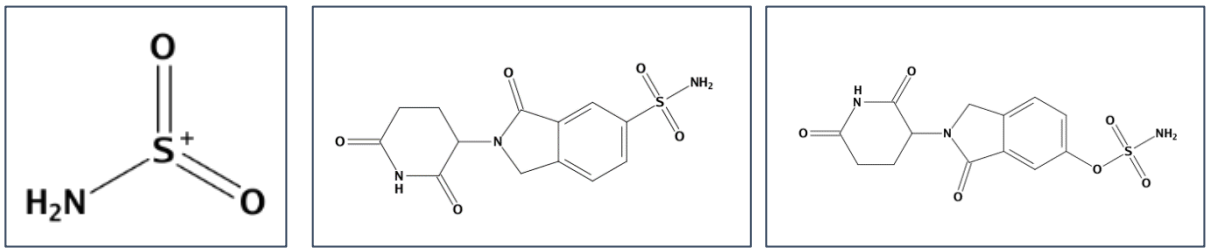


Figure 24 Left: Sulfonyl amide substructure proposed to label Histidine residues and perform a covalent bond. Middle: One of two small molecules hypothesized to form a covalent bond with CRBN residue His353, called EM12SO2F, shown with an N-link atom towards a Histidine. Right: the other small molecule hypothesized to form a covalent bond with CRBN residue His353, called EM12FS, shown with an N-link atom towards a Histidine.

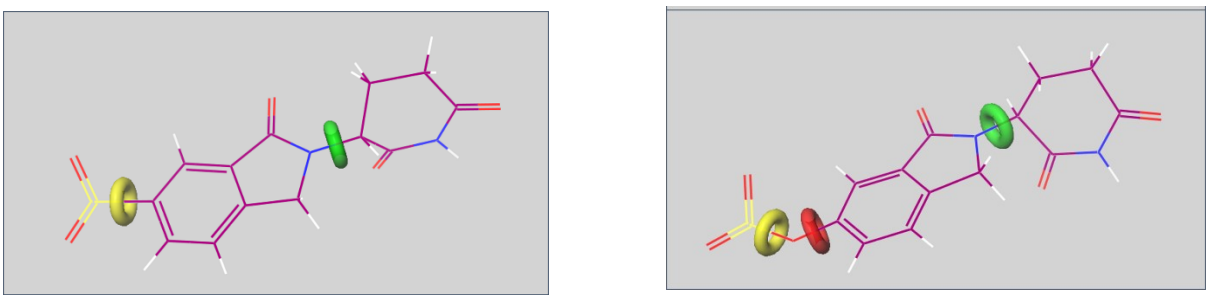


Figure 25 Both ligands hypothesized to bond covalently to the CRBN residues His353 shown without the leaving group F-atom. Left: EM12SO2F; Right: EM12FS. Exemplary small molecules for the visual output of the traffic-light system-based display whether the torsions of the rotatable bonds strained. Red to green standing for strained to relaxed torsion angles. Torsion Scan carried out based on Rarey Tool in Maestro [41].

	Docking_GOLD	ND1 proton. State	NE2 proton. State	Cov. Bond to Nx	HOH2037	Number of poses
5FQD	Run1	+	+	NE2	-	7
	Run2	-	+	NE2	-	6
	Run3	+	-	NE2	-	6
	Run4	-	-	NE2	-	13
	Run5	+	+	ND1	-	6
	Run6	-	+	ND1	-	10
	Run7	+	-	ND1	-	6
	Run8	-	-	ND1	-	6
	Run9	+	+	NE2	✓	6
	Run10	+	+	ND1	✓	12
6BNB	Run11	+	+	NE2	-	7
	Run12	-	+	NE2	-	10
	Run13	+	-	NE2	-	10
	Run14	+	+	ND1	-	12
	Run15	-	+	ND1	-	14
	Run16	+	-	ND1	-	13
HIS3xy	HIS357_RunA	+	-	NE2	-	6
	HIS357_RunB	-	+	ND1	-	13
	HIS378_RunA	+	-	NE2	-	20
	HIS378_RunB	-	+	ND1	-	6
	HIS397_RunA	+	-	NE2	-	7
	HIS397_RunB	-	+	ND1	-	12

Table 4 Overview of all the covalent docking runs. From left to right: Name of docking run carried out by GOLD; ND1 protonation state; NE2 protonation state; covalent bond defined whether ND1 or NE2; Runs including the water molecule HOH2037 that is closed to the binding pocket; Number of poses generated by GOLD;

5.4. Molecular Dynamics (MD) Simulation

All MD simulations in this work have been set up with “AMBER”, the Assisted Model Building with Energy Refinement, a collection of programs allowing the preparation and running of MD simulations for a variety of molecule types. The following MD simulations allowed studying the dynamics of the CRBN protein and its interactions to the ligand lenalidomide in the IMiD binding pocket.

5.4.1. Preparing MD Simulations

MD simulations were prepared with *antechamber*, which provides the force fields for the protein as well as for organic molecules. Eventually, the ligand was written in a mol2 file as well as the protein in a PDB format.

1. Ligand Preparation:

The lenalidomide ligand that is co-crystallized in the binding pocket of CRBN PDB structure 5FQD was modelled for 6BNB at the corresponding position as described above. The *antechamber* program was used to prepare the mol2 ligand file that could be used further for the *LEaP* step and for its package functions to assign atomic charges and atom types. The force field GAFF2 was applied for the preparation of the ligand. By running parmchk2 subsequently, the missing force field parameters were applied to generate a frcmod file for the ligand.

2. Protein Preparation:

The starting protein structure needed modifications to apply it for the *LEaP* step. Firstly, both protein conformers were prepared in MOE by QuickPrep default settings as described above, except endings of chains were capped. By this step, missing atoms or gaps were filled, too. For the simulation runs with “AMBER” the PDB file of the prepared CRBN structures containing the atom names, residue names, chain identifiers and the coordinates of heavy atoms were used. The deletion of all hydrogen atoms and all ligand atoms as well as all connectivity records were excluded from the files. Finally, the terminals of chains needed capping. Small, neutral groups are added to the end of the chains during the preparation of the MD simulations to prevent charged endings. During the structural preparation in MOE, acetyl (ACE) caps on N-terminals of all residues were added. However, to have the right caps on C-terminals of chains, N-methyl amide (NMA) endings from the MOE

preparation as well as CA atoms were changed to amine group (NME). File outputs were saved in a PDB format for both CRBN conformations individually and as such were prepared for *LEaP* after dealing with the Zinc ion as described below.

3. Metal Ion Coordination:

Residues around the Zinc ion of CRBN needed adaptations as the coordination of the CYS residues are not defined in the PDB file. By using the program Pymol for visualization, the metal coordinating CYS residues 394, 391, 326 and 323 were identified and were changed to CYM via a shell command. In line with the ZAFF modeling tutorial from “AMBER”, it deprotonated charges of the CYS residues around the Zn^{2+} ion. With this final modification the prepared protein PDB files were subjected to the *LEaP* script.

4. *LEaP* step:

LEaP was run by a shell script with the prepared protein files, and the output gave coordinate and topology files for further “AMBER” steps, respectively. Several commands in the *LEaP* script will be shortly described here. The tleap scripts loaded leaprc files for both energy fields ff19SB and gaff2 for the protein and ligand, respectively, as described above. The parameter files were loaded, and a standard water model called TIP3P was used as water parameter set. Finally, the protein and ligand file were taken to solvate the molecules in the systems. The water box for the solvation indicates that at least 14 Å of buffer were between the molecules and the periodic box wall (Figure 26). After solvation, the *LEaP* script also specified the addition of ions in a way that Na^+ and Cl^- ions were added to the system keeping the unit neutral. Topology and coordination files were saved as parm7 and rst7 files, respectively.

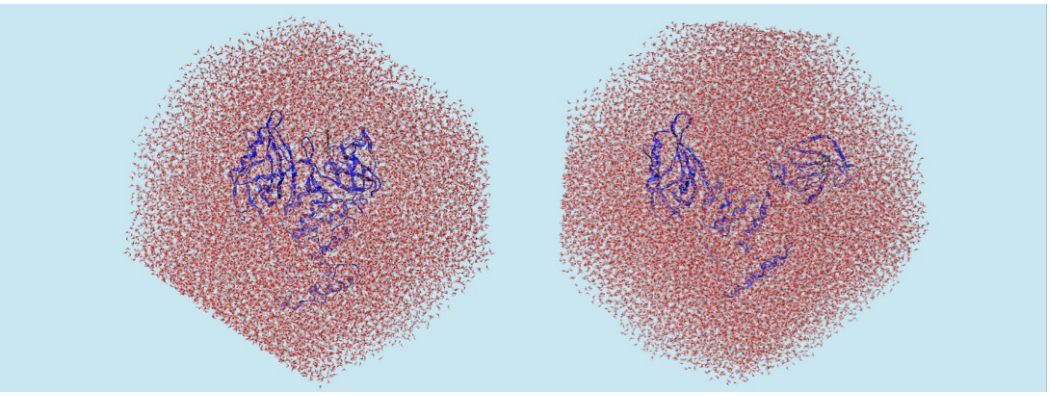


Figure 26 First frames from both CRBN conformers with ligand extracted and shown without stripping the environment, containing water molecules and ions, Na^+ and Cl^-. Left: open CRBN; Right: closed CRBN;

The equilibration of the system consists of three steps, starting with the energy minimization to create a stable system. The systems were heated up in several steps before the relaxed systems were run as MD simulations generating trajectory files.

1. Minimization:

An “AMBER” module of *sander*, the simulated annealing with NMR-derived energy restraints, called *pmemd* was applied to carry out energy minimization before the heating step of the system. During minimization atoms are moved based on the forces in the system in such a manner to find the structure closes to the energy minimum. This step ensures that the MD simulation does not fail due to bad contacts that induce unnecessary high energies, and could cause the simulations to crash, melt parts of the structures or in general, produce unrealistic trajectories.

2. Heating with constant variables:

The heating steps during the relaxation procedure reached 300 K gradually and gives randomized velocities from the beginning. Equilibration was carried out under NPT conditions, with N standing for the number of particles, P for pressure and T for temperature, where pressure dynamics are controlled.

3. Production run of the MD simulation:

Finally, the MD simulations were run by subjecting the “AMBER” job scripts to the HPC workload manager slurm. The MD simulations were run in replicates for all systems, which included the open and closed conformations of CRBN with the ligand in the binding site. All MD simulations were run for a time of 1,000 ns.

5.4.3. Analysis of Trajectories

The trajectories from all MD simulations were subjected to the same analysis. An overview of the analysis workflow can be found in Figure 22. Each trajectory was loaded into LigandScout in a .nc file format to create MD pharmacophores over the whole trajectories. This was done loading a topology file in PDB format first, either by taking the first frame of the trajectories or by using the topology file output from the MD simulation run. Only then, the corresponding trajectories were loaded. The creation of MD pharmacophores allowed for analysis of pharmacophore-based interactions between lenalidomide and the residues in the CRBN structures and generated a visual output of three different plots. Unique pharmacophore features

could be observed by their appearances in count and in which frames as well as in over the whole timeline (Figure 43, 44, and 45).
In addition, a CCPTRAJ based analysis was carried out that is provided by the "AMBER" tool package to perform analysis on trajectory information by a variety scripts. Among the many possibilities of the collection of scripts, CCPTRAJ was used by using text input files with all commands for handling trajectory sizes, file formats and their analysis. Prior to using any trajectory file generated, the all residues of the structures were fitted against the first frame and to include the possibility to remove water molecules and ions from the solvation step in order to decrease the file size. Only after this step, the trajectory files were loaded into programs, like LigandScout, or were further used for analysis. This was executed by the commands of "autoimage", RMS fit for all residues and "strip" of solvation atoms. Trajectories were also possible to transform into a multi-PDB file as well as merging of the two replicates per setting of the MD simulations.
Furthermore, the recipe collection of scripts in the CCPTRAJ hub, a hydrogen bond analysis between ligand and protein as well as a detection of regions for high mobility from the protein simulations was carried out.

5.4.3.1. MD Interaction and Pharmacophores Analysis

MD Pharmacophores were created in LigandScout for all trajectories that included a CRBN protein with lenalidomide. This gave three plots for the analysis of unique pharmacophores, which are defined as a pharmacophore feature between exactly one ligand atom and one protein atom. The interaction map gives all interactions plotting relative appearance and type of the individual pharmacophores features to the respective residues of CRBN. All 50,000 frames were used for the analysis. In addition, the MD pharmacophores were created for trajectories still containing the water molecules and ions and those that were stripped from them. The interaction maps made from trajectories still including the waters and ions of the solvation step could also show possible interactions to this environment.
Next to the interaction map, sets of unique pharmacophore features can be presented in the pharmacophore plot by their number of appearances throughout all frames. Here, the set of unique pharmacophore features stands for several unique pharmacophores that are found in one frame. These sets of specific

pharmacophores based on unique ligand atom to protein atom interaction is counted throughout the trajectories and presented in a bar chart starting with the highest number of appearances for the sets and the number of unique pharmacophores. Here, in the interactive visual window or in the hierarchy view of LigandScout, individual pharmacophores were chosen to highlight their presence in the bars giving insights in how many sets of pharmacophore features these exist. This was carried out with pharmacophores taken from first appearances from protein atoms of so-called core residues and included pharmacophores from His378, Trp380 and Asn351.
Finally, the pharmacophore feature timeline plot is the third visual output after creating MD pharmacophores over any trajectory loaded into LigandScout. The relative presence of unique pharmacophore features is plotted against binned frames. Binning size was set at 5,000 frames providing a rough and timely overview of the presence of unique pharmacophore features. Once more, the same features from the so-called core residues were picked and visually highlighted.

5.4.3.2. GRids of phArmacophore Interaction fieLds (GRAIL)

GRAIL is a grid-based method that uses the plethora of conformations of the simulated protein and the pharmacophore concept to generate a visual output of surfaces in the individually defined grid box [43]. For each type of pharmacophore feature a GRAIL map is generated based on the trajectory in PDB format that has aligned residues of all frames. In addition, residues which should be included in the grid box can be individually selected and were defined in this part of the master thesis by residue numbers 318 to 392 for the closed and 318 to 364 for the open CRBN conformer. These residues represent the CULT domain of CRBN and as such comprise all residues from the binding site. From the six types of interactions that can be visualized by GRAIL maps only those yielding information will be shown below. These interactions include the pharmacophore features like the following: hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic (H), positive ionisable (PI) area, negative ionisable (NI) area and aromatic ring (AR), halogen bond donor (XBD) and halogen bond acceptor (XBA). These pharmacophore grids only take the protein conformations of the trajectories into account and exclude ligands from the calculations. To generate the right file formats

to run calculations for GRAIL maps, the trajectory files were transformed into multi-PDB files with and without stripping of the water molecules and ions, next to aligning the residues. Topology files were subjected to stripping water molecules and ions, too. In the first step, a grid file was generated using the respective trajectory and topology file in PDB format that produces one output file. Secondly, residues that should be included in the grid box were defined and saved in a text file that was used for the grid generation from the previously generated output file. The same step can be done for calculations of an average grid. Finally, the calculated grids were unpacked into the individual pharmacophore GRAIL maps to a KONT file format that is compatible with the program LigandScout. Next to average grid calculations, outputs were also generated from scripts producing GRAIL maps for happy/unhappy waters and converted to KONT files. A visual analysis of GRAIL maps showing individual surfaces of the pharmacophore interactions was carried out in LigandScout. The generated trajectories in LigandScout were inserted after the respective topology file. Visualization of the pharmacophore interactions fields can be done over the entire protein view, or by zooming into the binding pocket and only then loading the individual KONT files. The visualized streaks and blobs of colour show potential pharmacophore feature interactions that were assessed by adjusting the score value of the individual GRAIL maps above a contouring level of a minimal 0.5.

6. Results

6.1. Structural Findings of CRBN

Two conformations of the CRBN were used for further experiments. From the PDB database, 26 human structures were identified, from which only two structures had the sensor loop of the CRBN protein defined with electron density data (Table 4). The structure 6BNB was applied for further experiments as the open CRBN conformation. In contrast, there are much more entries with a closed CRBN structure, from which the majority has a ligand co-crystallized in the IMiD binding pocket. Only three of the human CRBN proteins reported have no ligand, while one of these is one of the open conformations.

Structural inspection of all 26 human CRBN structures showed that only the two structures 6BNB and 6H0F have the two-palm state. Going into more detail, the sensor loop and especially the His353 conformations were inspected. This showed two main groups of His conformations. One group included all closed CRBN protein entries of the PDB database. In contrast, the second group showed diverging His353 conformation for three individual PDB entries that comprised 6BN7, 6BN8, 6BN9. However, these three protein structures showed an overall closed CRBN conformer and interestingly, do not have an electron density defined sensor loop (Figure 27).

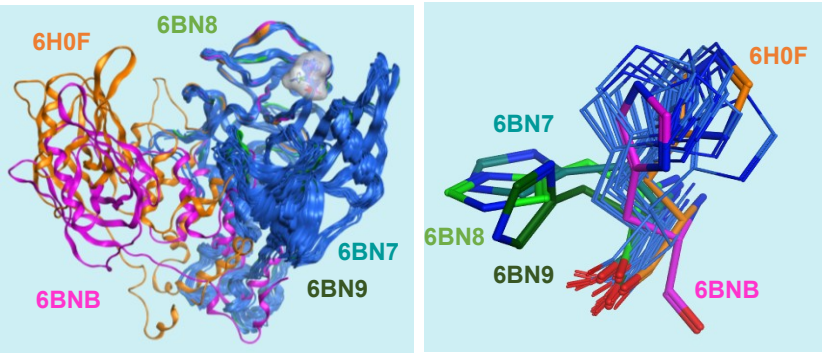


Figure 27 Structural display of the closed and open CRBN showing all 26 human PDB entries with CRBN available. Left: whole CRBN protein structures (blue = all closed conformers; pink and orange = only two open conformers). Right: conformation of residue His353 from all 26 human CRBN structures shown, dividing into two major conformational groups (same colour coding as left figure). Open CRBN structures 6H0F and 6BNB are part of the bigger group of His353 conformations. Second group comprises closed CRBN structures 6BN7, 6BN8 and 6BN9. None of the latter have the His353 and surrounding residues from the sensor loop experimentally defined by electron density.

PDB Entry	Method	Resolution	B value	R factor	R-free	Clash score	Ramachandran outliers	Mutations in CRBN	CRBN Length	His353	His357	His378	His397	Ligands	opened/closed
4M91	X-ray	1.10 Å	10	0,137	0,16	5	0%	0	12	yes	-	-	-	yes	-
4TZ4	X-ray	3.01 Å	30,29	0,202	0,272	12	0.7%	0	381	yes	yes	yes	yes	yes	closed
5FQD	X-ray	2.45 Å	96,76	0,181	0,21	3	0.5%	0	426	yes	yes	yes	yes	yes	closed
5HXB	X-ray	3.60 Å	112,3	0,224	0,273	26	1.3%	0	406	yes	yes	yes	yes	yes	closed
5V3O	X-ray	3.20 Å	95,2	0,204	0,267	2	1.2%	0	406	yes	yes	yes	yes	yes	closed
6BN7	X-ray	3.50 Å	NULL	0,212	0,256	5	0.5%	1 (BRD4)	463	yes	yes	yes	yes	yes	closed
6BN8	X-ray	3.99 Å	189,7	0,289	0,333	3	1.0%	2 (BRD4)	463	yes	yes	yes	yes	NO	closed
6BN9	X-ray	4.38 Å	NULL	0,275	0,301	3	0.8%	1 (BRD4)	463	yes	yes	yes	yes	NO	closed
6BNB	X-ray	6.34 Å	NULL	0,337	0,381	5	0.7%	1 (BRD4)	463	yes	yes	yes	yes	NO	opened
6BOY	X-ray	3.33 Å	NULL	0,199	0,234	5	0.5%	1 (BRD4)	463	yes	yes	yes	yes	yes	closed
6H0F	X-ray	3.25 Å	97,73	0,212	0,234	0	0.9%	0	426	yes	yes	yes	yes	yes	opened
6H0G	X-ray	4.25 Å	152,8	0,221	0,256	1	1.1%	0	426	yes	yes	yes	yes	yes	closed
6UML	X-ray	3.58 Å	85,16	0,204	0,267	9	0.4%	0	406	yes	yes	yes	yes	yes	closed
6XK9	X-ray	3.64 Å	93,1	0,208	0,245	8	0.3%	0	406	yes	yes	yes	yes	yes	closed
7BQU	X-ray	1.90 Å	NULL	0,191	0,236	6	0%	1 (CRBN)	114	yes	yes	yes	yes	yes	closed
7BQV	X-ray	1.80 Å	NULL	0,197	0,223	3	0%	1 (CRBN)	106	yes	yes	yes	yes	yes	closed
7LPS	X-ray	3.78 Å	124,3	0,289	0,303	3	0.6%	0	390	yes	yes	yes	yes	yes	closed
8CVP	EM	3.40 Å	NULL			4	0.1%	0	442	NO	no	yes	yes	yes	openedHis
8D7U	EM	3.10 Å	NULL			10	0.8%	0	442	yes	yes	yes	yes	yes	closed
8D7V	EM	3.20 Å	NULL			10	0.4%	0	442	yes	yes	yes	yes	yes	closed
8D7W	EM	3.10 Å	NULL			8	0.3%	0	442	yes	yes	yes	yes	yes	closed
8D7X	EM	3.40 Å	NULL			9	0.2%	0	442	NO	no	yes	yes	yes	openedHis
8D7Y	EM	3.40 Å	NULL			11	0.6%	0	442	NO	no	yes	yes	yes	openedHis
8D7Z	EM	3.10 Å	NULL			10	0.5%	0	442	yes	yes	yes	yes	yes	closed
8D80	EM	3.60 Å	NULL			8	0.6%	0	442	yes	yes	yes	yes	yes	closed
8D81	EM	3.90 Å	NULL			9	0.6%	0	442	yes	yes	yes	yes	yes	closed

Table 5 shows an overview of all 26 human CRBN including PDB structures that were potential candidates to be used in this master thesis. From left to right column: PDB entry code; method used to define experimentally structures; Resolution achieved for the respective structure in Ångstrom; B-value; R-factor value (light orange for values above 0.25, dark orange for values below 0.25); Clash score value; Ramachandran outliers in percentages; Mutations recorded for CRBN or other protein chains in the PDB entry (BRD4); CRBN length giving the residue length of CRBN; His353 showing whether His353 was experimentally defined in the sensor loop of CRBN (yellow); columns for His357, His378, and His397 giving the same information for these Histidine residues (light yellow); Ligands showing whether a ligand co-crystallized in the IMiD binding pocket (light grey); open/closed defining which CRBN conformer structure (red) as well as His353 conformation (pink);

6.2. Covalent Docking

Two ligands called EM12FS and EM12SO2F have been reported to be covalent binders for the CRBN protein (Figure 18) [46]. However, there is no experimentally determined structure with either lenalidomide derivative. It has been shown that the binding of an IMiD or IMiD like ligand is an essential step before degradation can happen. While the two ligands only differ by one oxygen atom, one has been shown to be a degrader while the other one has been not. Nevertheless, the ligand designs could be used for further development for PROTACs. The advantageous covalent bond to either the CRBN or the POI in the degradation complex would change the therapeutic PROTAC modality to an improved 2-body kinetics. Therefore, covalent docking was carried out to assess possible poses with new interactions to residues in the binding pocket that could add ideas to the development of a covalently linked degrader.

Covalent docking was performed with two structures available from the PDB. 5FQD is a reasonably well defined experimental closed structure that includes next to CRBN also DDB1 and CK1α. The structure has the IMiD lenalidomide co-

crystallized. Here, the small molecule is found in the IMiD binding pocket site. In addition, an open form of the CRBN protein was used to observe possible interactions of the ligands with the protein. It has been hypothesized that upon binding the CRBN complex closes and changes its two-palm conformation to one fist-like state.

The docking runs varied in the tautomeric states of the His353 side chain, the nitrogen atom that was used for the covalent link to the ligands (NE2 or ND1), the His residues in the binding pocket area, and the protein structure (Overview in Table 5 and Figure 29). From the docked poses only, those were used that remained after the application of the filtering conditions. The ligands in the individual poses were energy minimized and checked for a proper geometry of the generated bond (Figure 28). The filtering of the poses that showed plausible positions of the ligand in the binding pocket area in comparison to the pocket are of the co-crystallized ligand that among other geometric considerations also included interactions to specific residues. These were defined as “core residues”. Core residues are all those amino acids that interact with the co-crystallized IMiD ligand in the pocket, too, and include of Trp380, Trp386, His378 and Asn351 (Figure 29). Finally, from the remaining poses the covalently bound ligands were extracted including the atom linkage to the His residue and were tested with the Rarey Torsion Scan tool [47]. Poses that twisted the covalently bound ligands in a manner that caused red marked torsions were excluded from the subsequent analysis (Figure 30). An overview of the remaining poses per run is shown in Table 6.

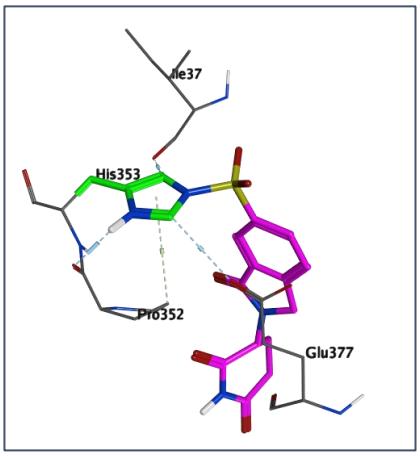


Figure 28 Visual extraction of an exemplary pose with the ligand EM12SO2F (pink) linked to His353 (green) showing a good geometry of the covalent bond after energy minimization.

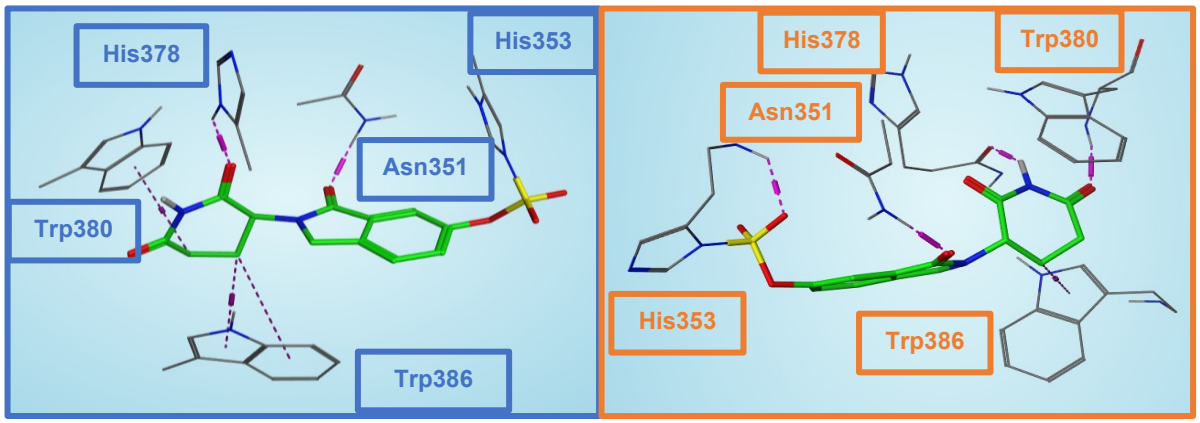


Figure 29 Overview of a selection aspect for which poses remain for interaction analysis of the covalently docked ligands, here two examples shown with small molecule EM12FS docked covalently to His353 for the closed (left; blue) and open (right; orange) CRBN structure. Both CRBN conformers displaying the so-called “core” residues in the IMiD binding pocket: Asn351, His378, Trp380, and Trp386.

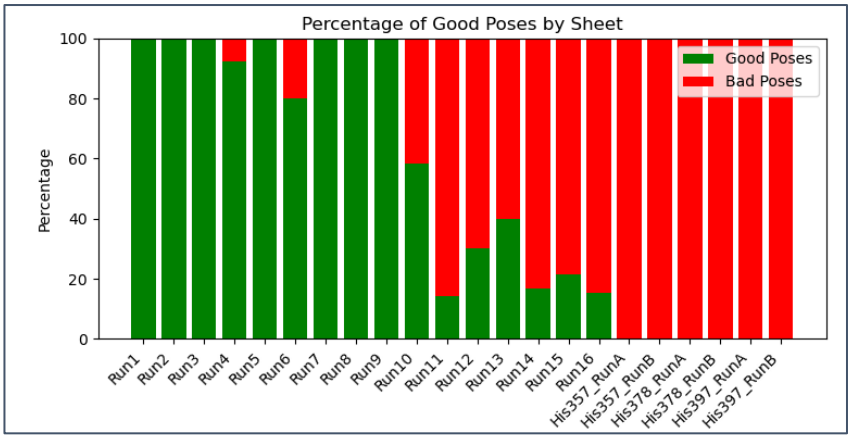


Figure 30 Overview of „Good Poses“ per docking run in percentages. Good poses were defined by scanning all docked poses of covalently docked ligands with the Rarey Torsion Scan Tool. Torsion angles of all rotatable bonds were categorized into very strained to relaxed. Only those poses without any strained torsion angles remained after the selection.

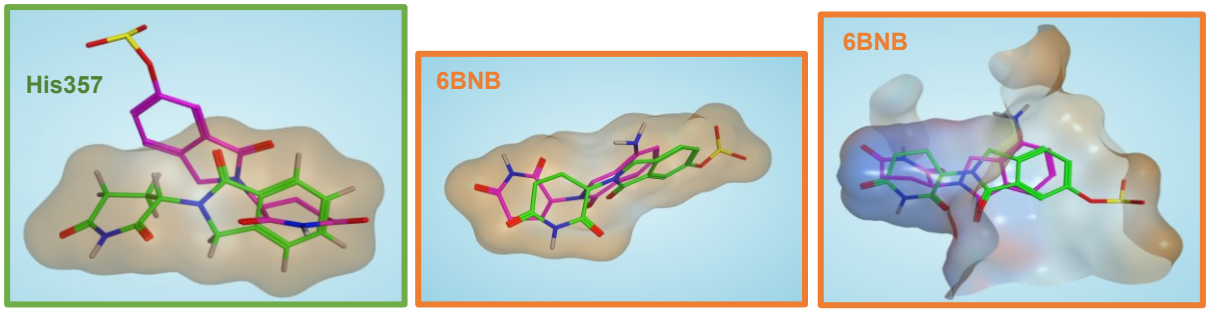


Figure 31 Exemplary poses shown that were excluded from the docking runs due to geometric considerations of the poses. Left: Both ligands were also docked to other Histidine residues, such as the pose shown here for His357 docked ligand, in comparison to the co-crystallized ligand lenalidomide in the closed CRBN structure (5FQD). Middle and Right: Both ligands were also docked to the binding site in the open CRBN structure (6BNB). Strained small molecule geometry and torsion angles of the poses excluded the poses from further interaction analysis. Binding pockets (brown) based on lenalidomide (green) in the IMiD binding site visualized.

Dockings also tested potential covalent links with other His residues in the vicinity of the binding pocket, which included His357, His378 and His397. Each a run was performed with the covalent link to either N-atom in the imidazole ring, so NE2 as well as ND1 individually. None of these generated poses showed plausible placements in the pocket in terms of orientation of the docked ligands (Figure 31). In addition, interactions with the core residues were seldomly or only partially observed. Similarly, individual runs with the open CRBN structure 6BNB did not yield any plausible poses with the same exclusion aspects. A tendency for NE2-linked poses with reasonable orientation, core residue interactions and agreeable torsion angles was observed for 6BNB docked runs to remain after the selection. This was observable to some extent also for 5FQD dockings. In general, more NE2 poses with the closed CRBN conformation were left after filtering with the above-described aspects in comparison to ND1 docked poses.

	Docking_GOLD	ND1 proton. State	NE2 proton. State	Cov. Bond to Nx	HOH2037	Number of poses	% used poses
5FQD	Run1	+	+	NE2	-	4	57,14
	Run2	-	+	NE2	-	4	66,67
	Run3	+	-	NE2	-	3	50
	Run4	-	-	NE2	-	11	84,61
	Run5	+	+	ND1	-	3	50
	Run6	-	+	ND1	-	10	50
	~~Run7~~	~~+~~	~~-~~	~~ND1~~	~~-~~	~~6~~	~~0~~
	Run8	-	-	ND1	-	1	16,67
	Run9	+	+	NE2	✓	5	83,33
	~~Run10~~	~~+~~	~~+~~	~~ND1~~	~~✓~~	~~12~~	~~0~~
6BNB	Run11	+	+	NE2	-	1	14,28
	Run12	-	+	NE2	-	3	30
	Run13	+	-	NE2	-	1	10
	~~Run14~~	~~+~~	~~+~~	~~ND1~~	-	~~12~~	0
	~~Run15~~	-	~~+~~	~~ND1~~	-	~~14~~	0
	Run16	+	-	ND1	-	1	7,69
HIS3xy	~~HIS357_RunA~~			~~NE2~~	~~-~~	~~6~~	~~0~~
	~~HIS357_RunB~~			~~ND1~~	~~-~~	~~13~~	~~0~~
	~~HIS378_RunA~~			~~NE2~~	~~-~~	~~20~~	~~0~~
	~~HIS378_RunB~~			~~ND1~~	~~-~~	~~6~~	~~0~~
	~~HIS397_RunA~~			~~NE2~~	~~-~~	~~7~~	~~0~~
	~~HIS397_RunB~~			~~ND1~~	~~-~~	~~12~~	~~0~~

Table 6 Overview of all the covalent docking runs with marked runs excluded by selection of visual inspections, presence of core residues and Rarey Torsion Scan red marked poses. From left to right: Name of docking run; ND1 protonation state; NE2 protonation state; covalent bond linkage point (ND1 or NE2); Runs including the water molecule HOH2037 that is closed to the binding pocket; Number of poses generated by GOLD; remaining poses per run; all red marked runs did not leave any poses to work with after selection.

Overall, after processing of the poses according to the selected criteria, the runs with the highest cut of remaining poses that were used for further analysis included a few dockings with the closed CRBN conformation. In contrast to runs with the open conformation, lower percentages remained for analysis and the highest number of remaining poses within this subgroup were from run 12 with 30%, a NE2 ligand docked run (Figure 32 and Table 6). In addition, it can be noted that the absolute numbers of poses did not vary between the 5FQD and 6BNB based dockings.

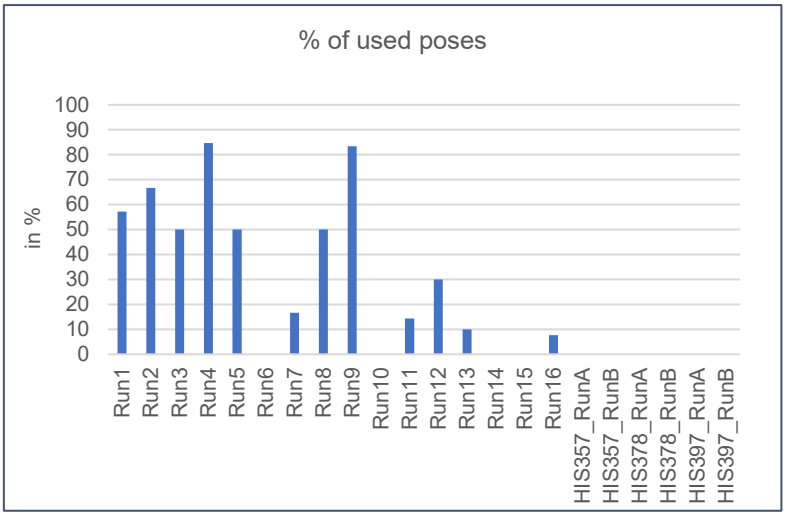


Figure 32 shows remaining number of docked poses in percentages per run. Bar graph output from Table 3.

Moreover, two runs included the water molecule HOH2037, a molecule found close to the binding pocket (Figure 33). To try and assess whether this molecule was necessary in some way for the ligand interactions and their stability in the docked poses, it was extracted and considered part of the protein structure for these dockings. Only the run with the closed CRBN structure constructed with the NE2 linked ligands and the water molecule showed reasonable poses and remained after the selection.

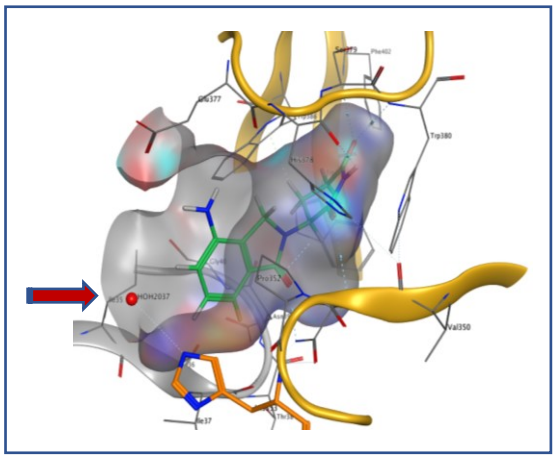


Figure 33 Visual display of the binding pocket with lenalidomide (green) shown, water molecule HOH2037 (red) pointed out by arrow.

His378, Trp386, Trp380 and Asn351 were considered as core residues and were observed to have the highest frequency and stability of interactions with the ligand. Moreover, the co-crystallized IMiD ligand was observed to interact with the same residues. Together these criteria defined the core residues. Among the set of other residues, a variety of amino acids close to the binding pocket was observed to interact with the two covalent ligands EM12FS and EM12SO2F. Residues like His353, Isoleucin (Ile) 35, Ile37, Glu377, Lys18, Pro352 and Trp400 were found. As the docking runs differed in certain settings as described above, differences in interaction patterns to the ligands were observable. All plausible poses after filtering showed that the individual core residues as well as the sum of other residues had similar relative interaction patterns (Figure 34).

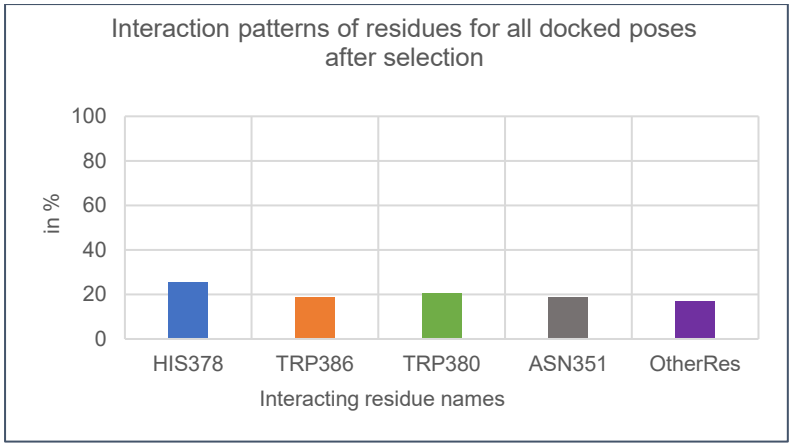


Figure 34 Interaction pattern of the two ligands covalently docked to both CRBN conformers. “Core” residues are shown individually, His378, Trp386, Trp380 and Ans351 (from left to right). Other residues are shown collectively, before further differentiating interaction analysis. Bars show relative interaction amounts in percentages from all docked poses remaining after the filtering.

Overall, most of the docked poses showed additional interactions with Ile37 (34%) and Lys18 (29%) next to core residues, both residues of another chain identity and only from the PDB entry 5FQD. In addition to the lower interaction levels of Ile35 with only 8% as another residue that is not part of the CRBN protein. Both ligands made interactions with either CRBN conformers by residues like Pro352 (17%), Glu377 (8%), Trp400 (2%) and His353 (2%) (Figure 35). More differential analysis was carried out to gather a better insight into the respective interaction patterns between NE2 versus ND1 as well as EM12FS and EM12SO2F docked poses. The same analysis was conducted for differences in patterns of the other residues between the open and closed conformer of the CRBN protein after filtering of the docked poses.

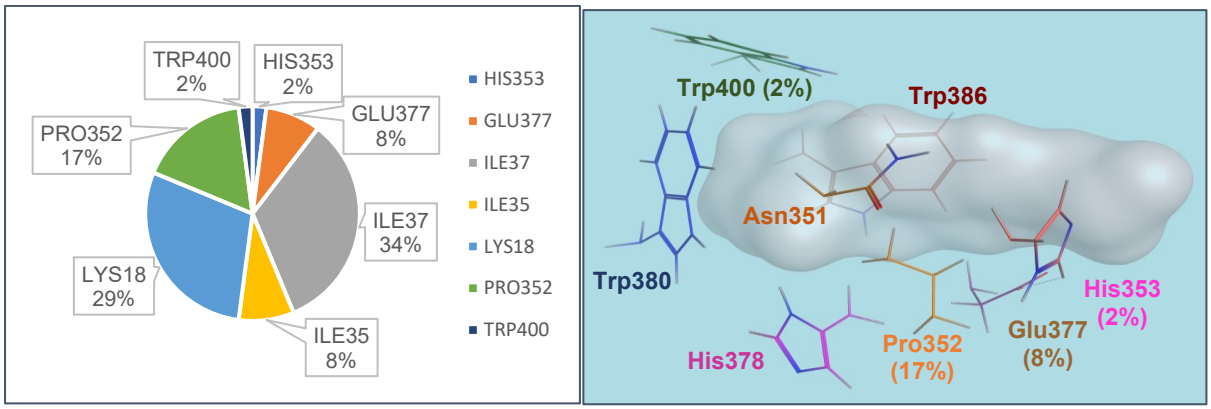


Figure 35 Pie chart showing all „other“ residues interacting with covalently docked ligands in all remaining poses after filtering, next to core residues. Other residues were defined as shown: His353, Glu377, Ile37, Ile35, Lys18, Pro352 and Trp400.

For all NE2 covalently docked ligands, no Trp400 interaction was observable, but interactions with Lys18 (38%), Ile37 (21%), Pro352 (19%), Ile35 (11%), and Glu377 (11%) (Figure 36). In contrast to ND1 covalently linked ligands, the highest interacting residue was Ile37 (73%) and in some other settings, His353 (9%), Trp400 (9%) and Pro352 (9%) (Figure 36). To sum up, for NE2 docked poses only the residues Pro352 and Glu377, and for ND1 docked poses next to Pro352, Trp400 and His353 were ligand interactions made with the CRBN protein and not with the respective POI of the PDB structures (Figure 36). While the respective POI for the open and closed CRBN entries differ, either ligand interact with the Ile residues of the closed structure. There is no interaction to the POI when the ligands were docked in the open conformation.

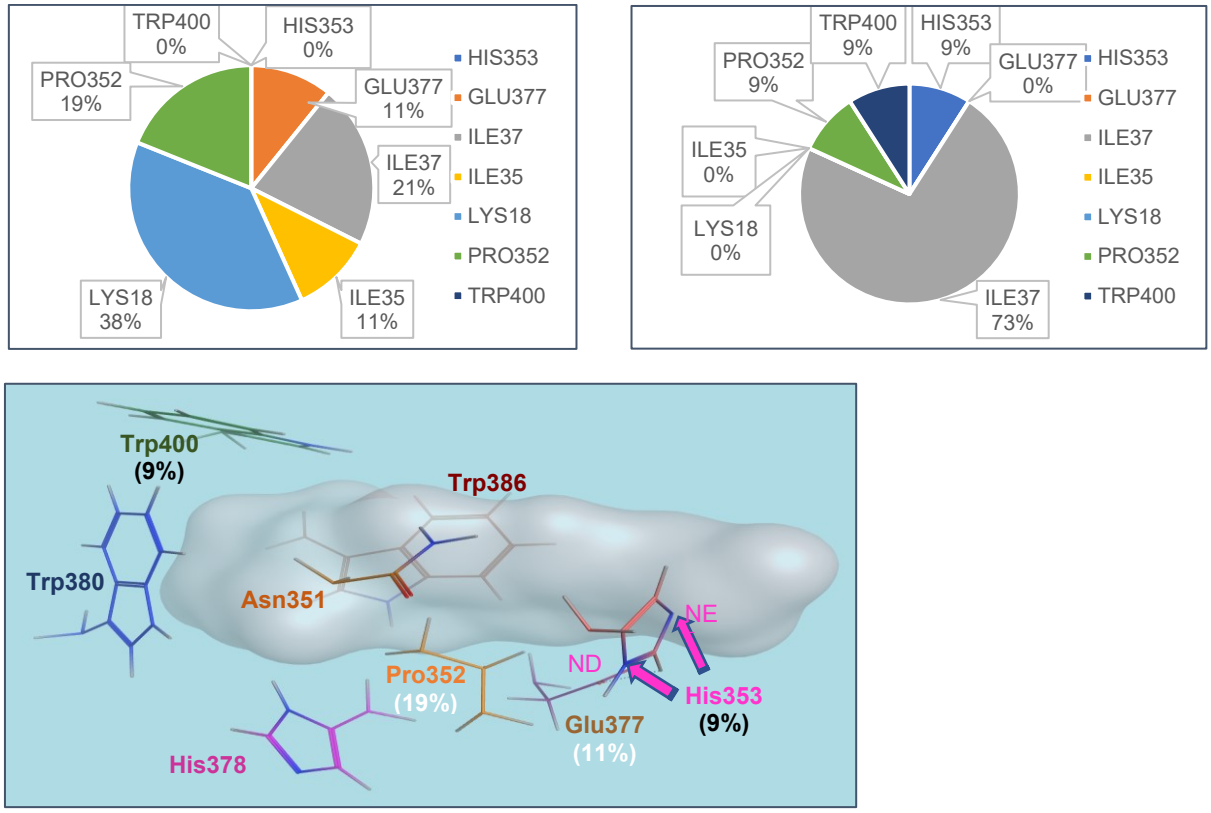


Figure 36 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: NE2 covalently docked poses. Right: ND1 covalently docked poses. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also NE2 (white) and ND1 (black) docked results in percentages. Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37).

Almost all binding site residues were found to be part of the various interaction patterns of the closed CRBN structure 5FQD. The observed interactions were like the results for all docked poses. Interacting residues were observed for both ligands as the following: Ile37 (34%), Lys18 (30%), Pro352 (17%), Ile35 (9%), Glu377 (8%) and Trp400 (2%) (Figure 37). Once more, Pro352 and Glu377 were shown as the most frequently interacting residues from the IMiD pocket. In contrast, His353 interaction was observed to be the only additional interaction with the open conformer. While the closed conformation shows a greater variety of possible interaction partners in the binding pocket next to the core residues, the ligand poses of the open structure remained only with His353 backbone interaction next to the covalent bond of the ligand to either N-atom in the imidazole ring of the same residue (Figure 37).

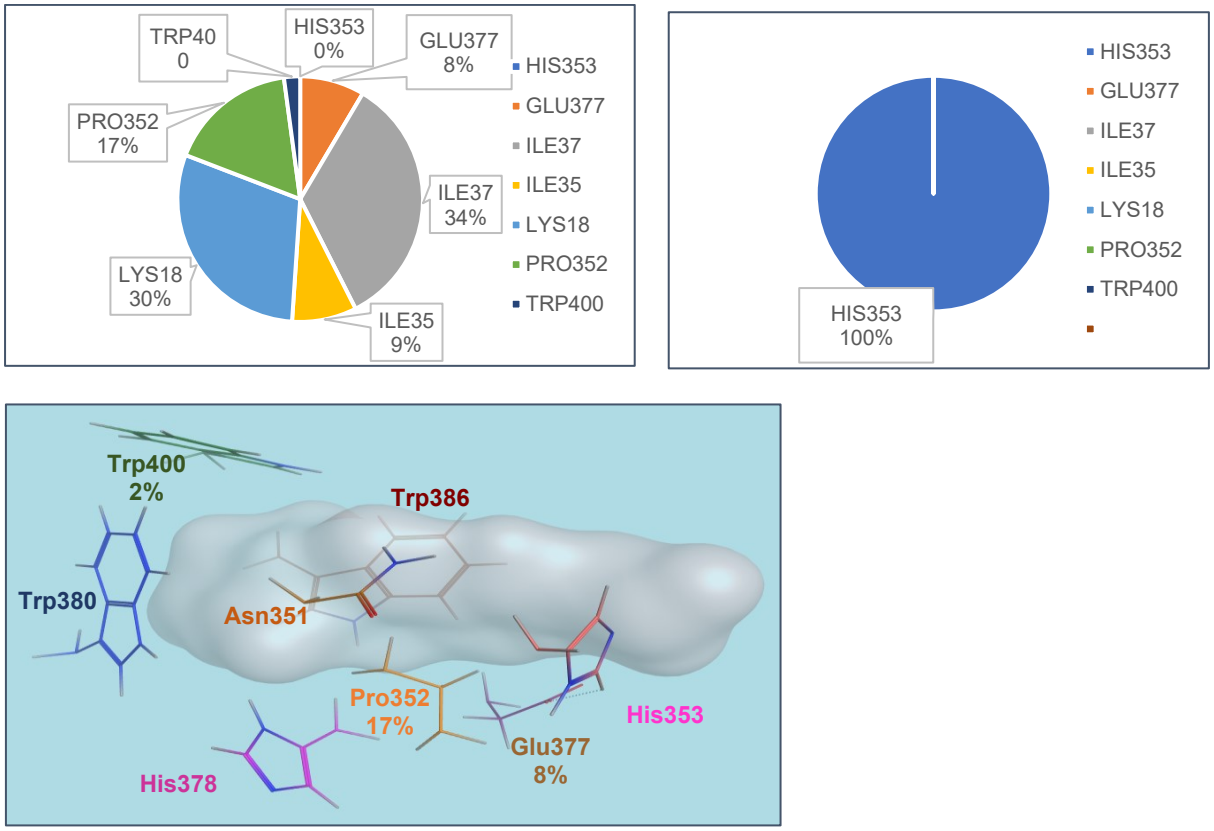


Figure 37 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: all covalently docked poses with the closed CRBN structure. Right: all covalently docked poses with the open CRBN structure. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also docked results in percentages with the closed CRBN. Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37).

Finally, the differential analysis showed differences in the interaction patterns between the two covalently docked ligands. The experimentally characterized degrader molecule EM12FS showed interactions from highest to lowest like the following: Lys18 (42%), Ile37 (25%), His353 (9%), Ile35 (8%), Pro352 (8%) and Trp400 (8%) (Figure 38). Next to the most frequent interactions being with the POI of the PDB entries, the major interacting residues were Pro352, His353 and Trp400. In contrast, no Glu377 interaction was observable. Comparing the interaction patterns to those of the other ligand EM12FS, there was no His353 and Trp400 interactions found. In addition, the non-degrading molecule EM12FS showed interactions with Pro352 (20%) and Glu377 (11%), next to POI residues like Lys18 (25%), Ile37 (36%) and Ile35 (8%) (Figure 38). These results were similar to those of the analysis differentiating interacting residues between NE2 and ND1 docked poses. EM12FS and ND1 based poses showed no Glu377 interactions, while both conditions had the interacting residues Pro352 and Trp400. In contrast, EM12SO2F and NE2 docked poses had no Trp400 interaction, while residues Pro352 and Glu377 were found at comparable levels. Overall, closed CRBN poses with NE2 linked ligands showed interactions with Pro352 (Figure 38).

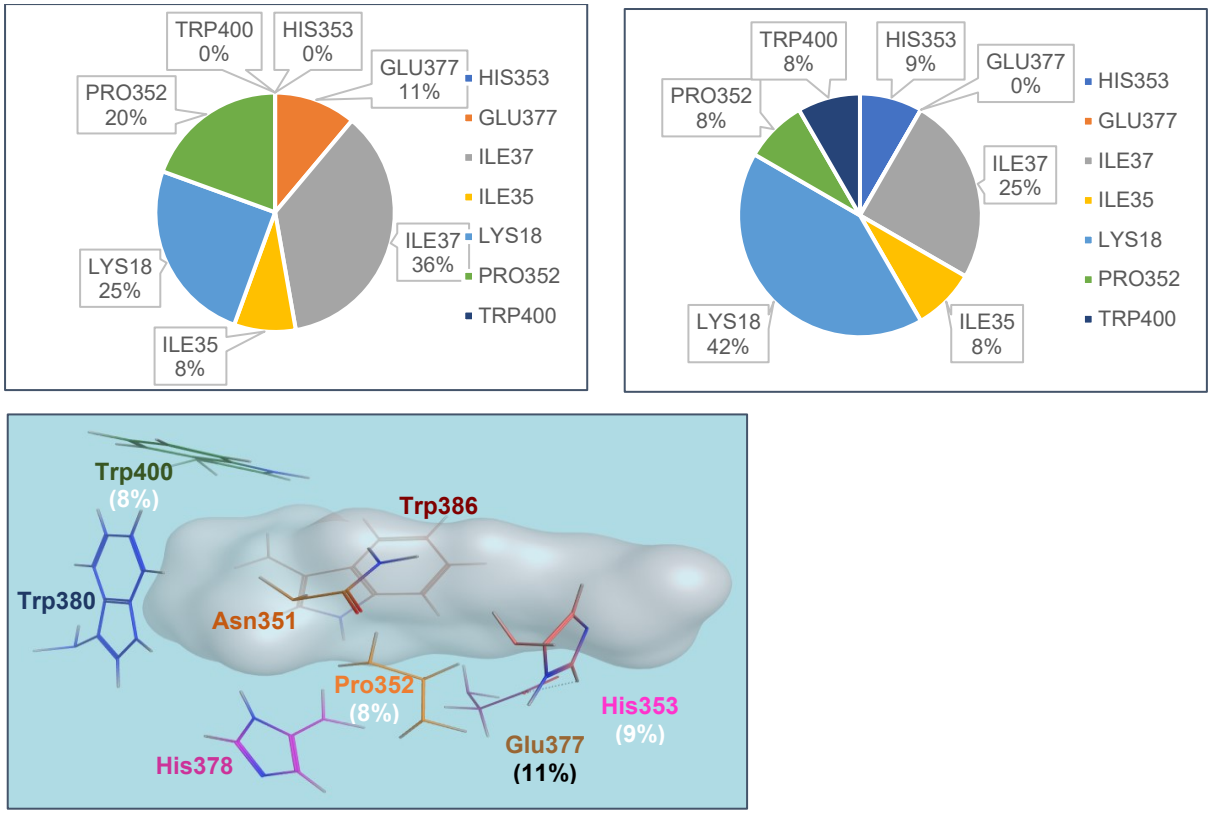


Figure 38 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: all covalently docked poses with the ligand EM12SO2F. Right: all covalently docked poses with the ligand EM12FS. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also docked results in percentages of ligand EM12FS (white) and EM12SO2F (black). Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37).

6.3. MD Simulations

MD simulations have several advantages in regard to analysis of ligand interactions in comparison to dockings, which allows only little motility and structures are more rigid. Therefore, MD simulations were run with “AMBER” with a closed and open form of the CRBN protein. The same PDB structures for open and closed CRBN state were applied as basis for this part of the thesis. As such, 5FQD was prepared as described above and 6BNB included the modelled in ligand from the closed PDB structure. The small molecule lenalidomide is the co-crystallized ligand of 5FQD in the IMiD binding pocket.

The generated trajectories from the MD simulations allow for an analysis of the ligand interaction and dynamics of the CRBN protein based on an open and closed

form. With CCPTRAJ, an “AMBER” related collection of scripts, the flexibility per residue was calculated and is shown as an RMSF plot (Figure 39). In addition, an analysis of H-bond interactions of the ligand to the residues in the binding pocket were assessed (Figure 40). The program LigandScout was applied to analyse MD-derived pharmacophores from entire trajectories. This powerful tool offers a user-friendly interface, especially for less informatics affine scientists. It allowed the generation of an interaction map, unique feature set and feature timeline plot, as described above in more detail. Finally, the trajectories were also applied for using a grid-based method called GRAIL that was recently published by Schuetz *et al* (2018). It considers the various conformations of the protein in the trajectory and by using the pharmacophore concept generates a visual output that can be used for the rational design of ligands. Based on these grids, the individual GRAIL maps can be visualized in various programs, such as LigandScout and can be used to display pharmacophore interaction fields for the individual pharmacophoric features [43]. The generated RMSF plot for both trajectories with the open and closed form including lenalidomide showed similarities in certain residue areas such as the regions between 30 and 150 as well as 220 to 300, respectively (Figure 39). Overall, the open CRBN conformation with the ligand shows a greater flexibility than the closed one. Other regions were assessed to have more flexibility. The CULT domain (residues 318 to 426) that includes the sensor loop, shows several peaks with differences of up to 6 Å. In addition, there is one particular maximum for the closed CRBN starting approximately from residue 370 until the end of the chain. Interestingly, the highest peak showing the greatest flexibility for both conformers of the CRBN structure were calculated for the residues 150 to 200. Especially, for the trajectory with the open CRBN conformation and the ligand was observed to have a flexibility of a distance greater than 18 Å in comparison to the first frame of the trajectory. This area is known to be in contact with the DDB1 protein providing further evidence for the extreme flexibility of CRBN and therefore the challenge of experimentally obtaining a crystal structure solely with CRBN. This movement of the residues in the region 150 to 200 is also observable in the trajectories to the naked eye (Figure 39).

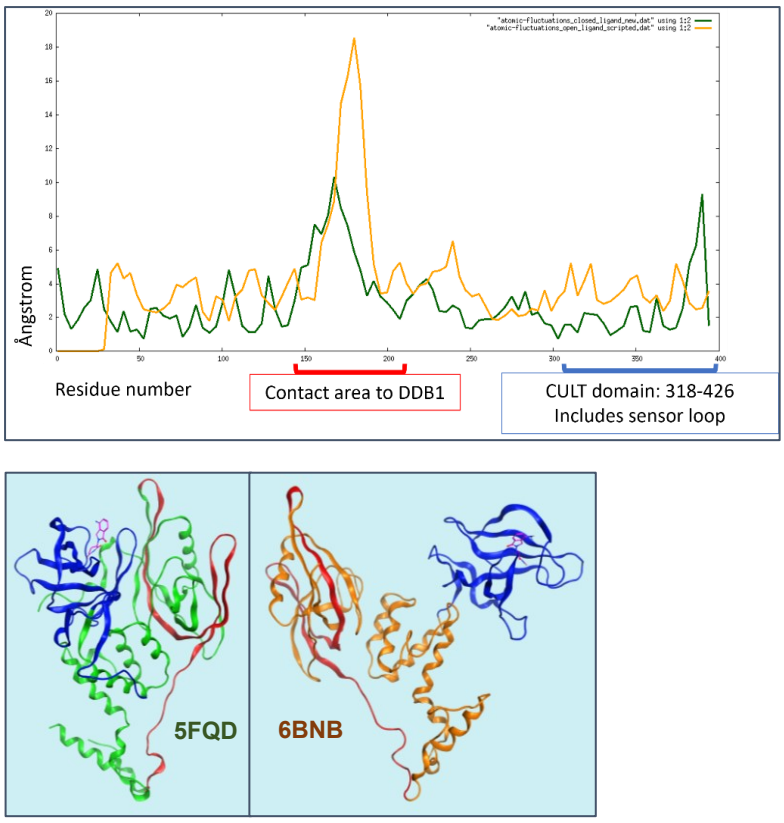


Figure 39 Above: RMSF plot showing the residue number (x-axis) against the Ångstrom difference (y-axis). Both trajectories of the open (orange) and closed (green) CRBN structure with the ligand lenalidomide showed similarities over most of the calculated flexibility of the residues. Areas of increased movement are displayed by the peak height, such as marked residues from 150 to 200 or 318 to the end of the protein chains, respectively. Below: Both closed (5FQD, green) and open (6BNB, orange) CRBN structure shown with respectively coloured residue lengths corresponding to the demarcations in the plot above before, such as the DDB1 contact area (red) and CULT domain (blue). Lenalidomide is shown in the IMiD binding pocket (pink).

6.3.1. Interaction analysis between lenalidomide and both CRBN conformers

In addition, the scripts from CCPTRAJ were applied to analyse the H-bond interactions of the ligand to the open and closed CRBN protein. The H-bond interaction plots show the occurrence per frame against a set of interaction between a specific lenalidomide atom to a specific residue atom.

The residues in the pocket of the open CRBN interact with the following residues in order of the highest number with one set consisting of each specifically a residue atom and ligand atom: Glu377 (6 sets), His378 (3 sets), Asn351 (3 sets), Trp380 (2 sets), His353 (2 sets), Trp400 (1 set) and Ser379 (1 set). The residues with the highest number of interactions via H-bonds is observable for residues like His378,

Asn351 and Trp380, with the latter being the most abundant H-bond interaction throughout the entire trajectory. Most diverse set of different linker atoms was found for Glu377 showing six unique ligand atoms to residue atoms H-bond interactions (Figure 40).

The same was assessed for the closed CRBN where similar residues were found to interact via H-bonds with the ligand. The following sets of unique ligand residue atom pairs were calculated and plotted: Glu377 (6 sets), His378 (3 sets), Trp380 (2 sets), Asn351 (2 sets), His353 (1 set), His357 (1 set), Trp400 (1 set) and Ser379 (1 set). The most abundant H-bond interaction throughout the entire trajectory was observable for the same residue of the binding pocket as the analysis for the open CRBN interacting with lenalidomide showed, Trp380. In addition, all other core residues identified from the docking experiments were found in this analysis, too, Asn351 and His378. Differences between the closed and open CRBN interactions to lenalidomide were only detected for the residue His357 (Figure 40).

Interestingly, interaction with the residue Ser379 that is close to the binding site, was defined, too. This particular Serin residue was found in the docking experiments, too, but only with poses that were excluded with the selection process, such as an NE2 docked pose with 6BNB.

To sum up, the H-bond interaction analysis did not show any striking differences between the two CRBN forms. Nevertheless, it can be noted that the same core residues and some of the other residues such as Glu377, His353, and Trp400, were found to establish H-bond interactions like those in the docking experiments. In contrast, the trajectories-based analysis of interactions did not show any H-bond interaction with residue such as Pro352 that was observed as one of the most frequently occurring other residues in the docking experiments for the closed CRBN conformation.

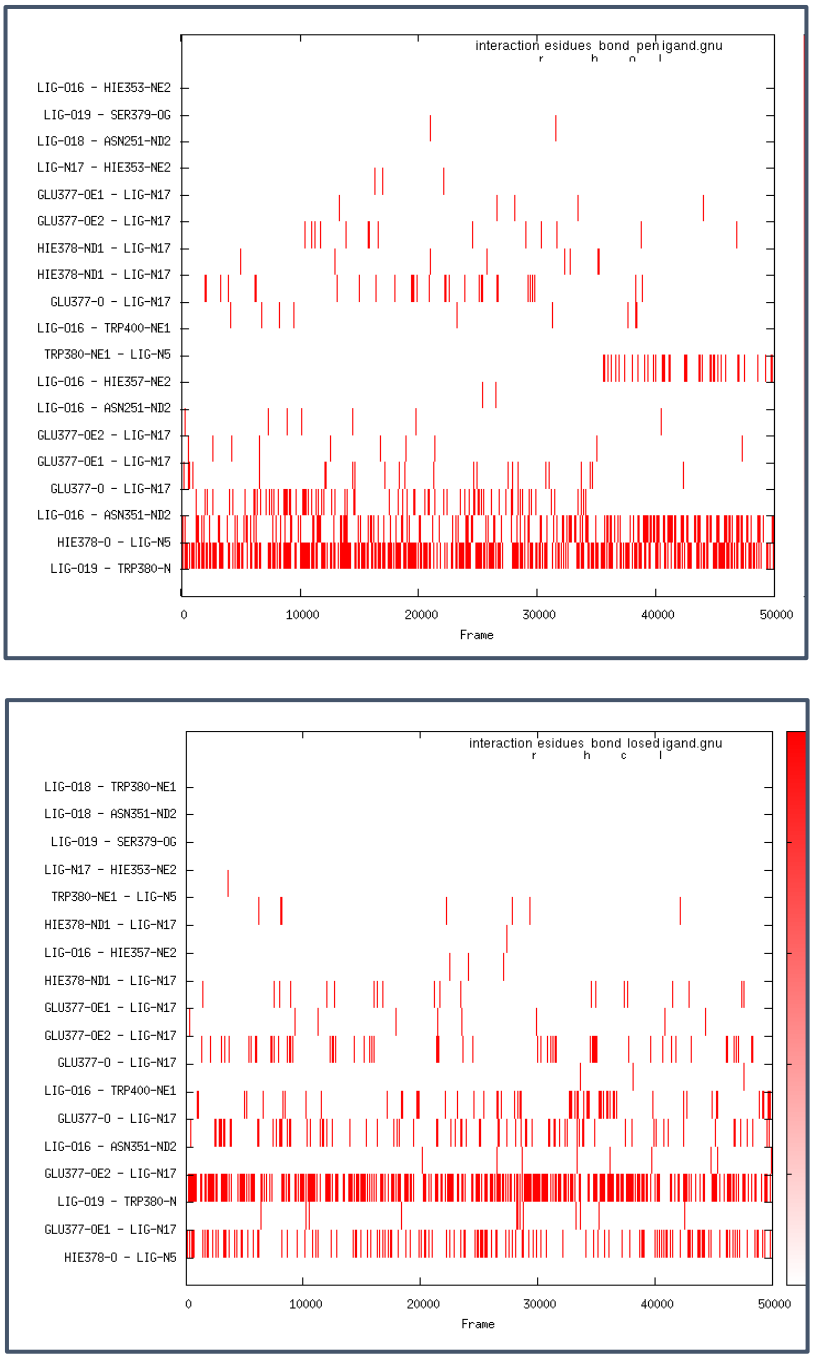


Figure 40 Plots showing the H-bond interactions (red) between protein and ligand residue atoms (y-axis) against frame (x-axis). Open (above) and closed (below) CRBN trajectories with ligands subjected to CCPTRAJ based analysis.

Additionally, the ligand interactions to both CRBN forms were assessed by the program LigandScout. The interaction map shows pharmacophore features of the ligand plotted against their relative appearance in the trajectory as well as the

residue involved in the interaction. MD-based pharmacophores were created and plotted for all trajectories with lenalidomide and included the water molecules, too. The interaction maps showed the highest interaction pattern with Trp380 by an HBA (Figure 41). It occurred almost through the entire trajectory with over 98%, independent of which CRBN form with the ligand was run. The highest appearance of the HBD feature was observed for the His378 residue at 62% for the trajectories with the open CRBN and ligand. In contrast, the trajectories with the closed CRBN showed the highest appearance of the same feature for the Glu377 residue with 78% interacting with the ligand. In contrast to the open CRBN, the His378 residue interacted with the ligand via an HBD during 39% of the trajectories. The same relative appearance was plotted for Glu377 in the open CRBN trajectories. These two residues showed the largest pharmacophore feature differences in their relative appearance for the trajectories with the two CRBN forms. More comparable levels of appearances showed the pharmacophore feature of HBA with the interactions to Ser379 at 16% and 21% in the open and closed CRBN trajectories, respectively. Similarly, the other HBA feature from the residue Asn351 to the ligand interacted during the trajectories with the open and closed CRBN for 34% and 37%, respectively. One of the lowest interaction appearances of a residue with the ligand to CRBN conformers was calculated for Trp400 between 1% and 3%. All other interactions with residues were plotted with 0% that stands for all pharmacophore features occurring less than 1% in the trajectories. All interaction maps showed two residues differing in their interaction pattern with lenalidomide between the closed and open CRBN. These residues included Phe381 and His357 that were calculated to interact via an HBA and via an HBD feature during less than 1% of the open CRBN trajectory, respectively. Interestingly, His357 was shown as the only differing residue interacting via an H-bond with lenalidomide in the analysis carried out via CCPTRAJ.

One particular pharmacophore appearance should be noted from the MD-based pharmacophore feature calculation of the trajectories with all water molecules. An HBA feature was observed for a relative appearance of 4% in the trajectory of the closed CRBN with the ligand (Figure 42). It can be noted that it was the only pharmacophore feature above 1% with a water molecule in either trajectory. In addition, all other pharmacophore features interact at the same level to the residues

in the binding pocket with the ligand as in the resulted plots calculated based on the trajectories without the water molecules.

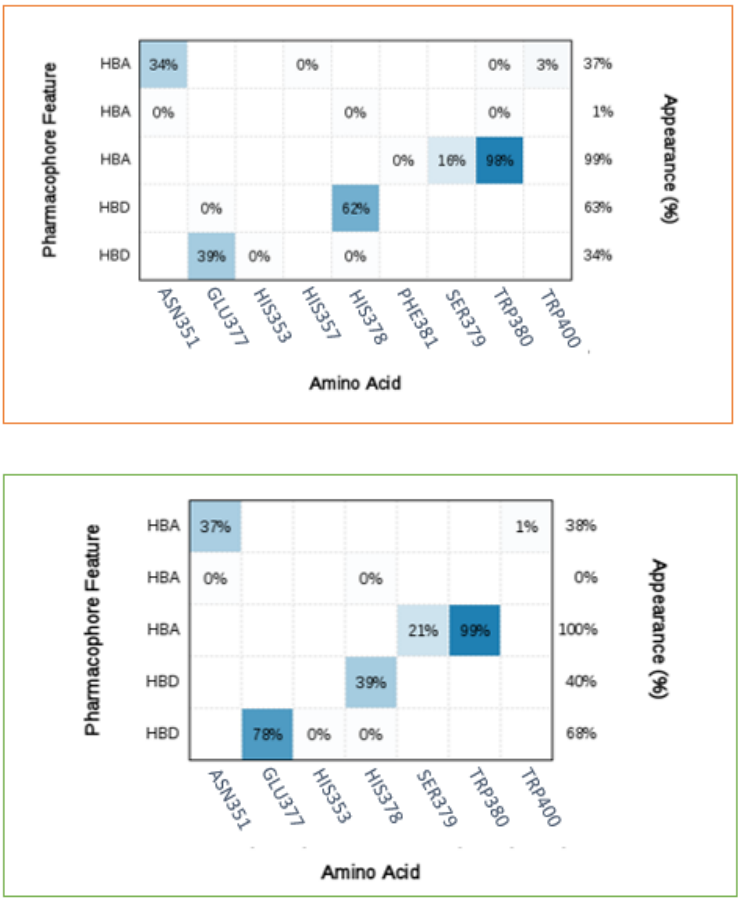


Figure 41 Pharmacophore interaction map generated by LigandScout MD Pharmacophore creation for the trajectories of the open (orange) and closed (green) CRBN structure with lenalidomide and stripped environment, without water molecules and ions. The map plotted residues interacting with the lenalidomide in the binding site and the corresponding pharmacophore feature. The relative interaction of the residues with the pharmacophore feature are shown in percentages by appearances during the entire trajectories. Appearances below 1% are shown as 0%.

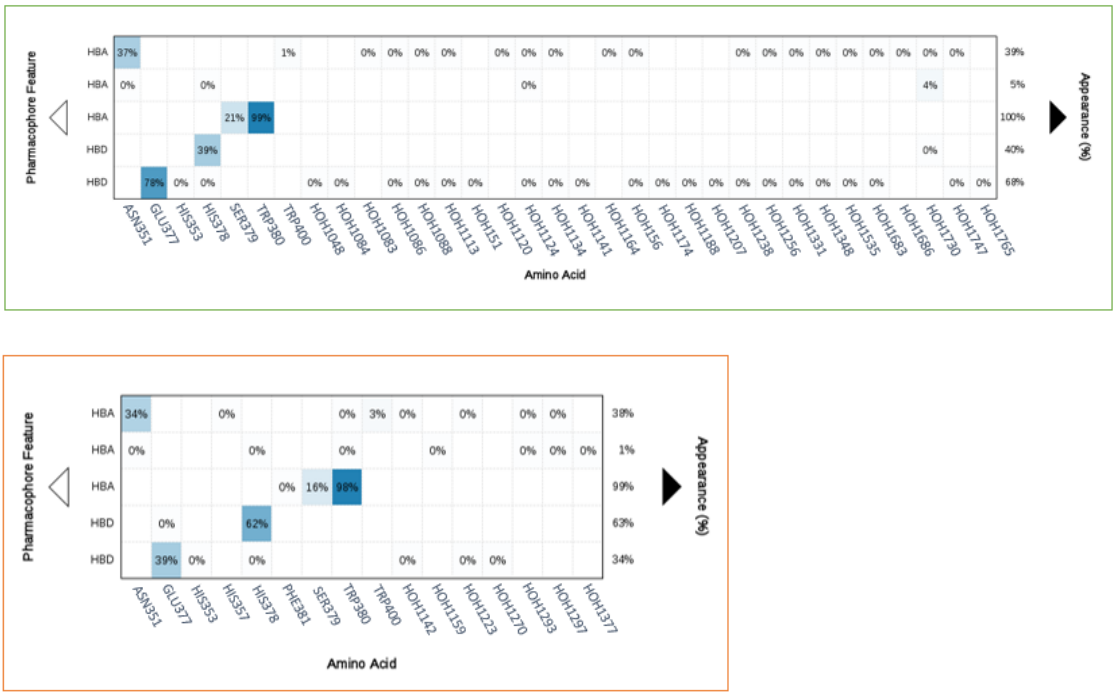


Figure 42 Pharmacophore interaction map generated by LigandScout MD Pharmacophore creation for the trajectories of the open (orange) and closed (green) CRBN structure with lenalidomide and the environment, including the water molecules to be part of the calculation. The map plotted residues interacting with the lenalidomide in the binding site and the corresponding pharmacophore feature. The relative interaction of the residues with the pharmacophore feature are shown in percentages by appearances during the entire trajectories. Appearances below 1% are shown as 0%.

Secondly, LigandScout provides a pharmacophore plot, where a set of unique pharmacophore features are plotted against their number of appearances in the frames of a trajectory from highest to lowest presence. By selection of a specific pharmacophore feature the bars from the plot appear green showing in which set and therefore how often it is part of a set. To verify the interaction map, the three pharmacophores from the core residues Trp380, His378 and Asn351 were selected in all trajectories (Figure 43). The two HBA and one HBD feature were selected as a minimum set and were observed to turn most of the unique pharmacophore feature sets green showing presence in most of the sets with different combinations of unique pharmacophores. This showed the core residues truly can be found as key interaction partners in most of the frames of the individual trajectories.
For those frames denoted with a zero (“0”) no pharmacophores were detected. For the trajectory with the closed CRBN and ligand, frames were found with only one pharmacophore feature set and a maximum of 6 pharmacophores in one unique set. Similarly, for the trajectory with the open CRBN and ligand unique sets were found

from none to a maximum of 6 pharmacophore features per frame. In addition, in figure 43 the pharmacophore plot allowed the pharmacophore feature of the water molecule from the closed CRBN trajectory with the water molecules included. This HBA interaction was found to be part of several unique sets. The feature set with the highest number of appearances in frames of the trajectories are furthermore shown.

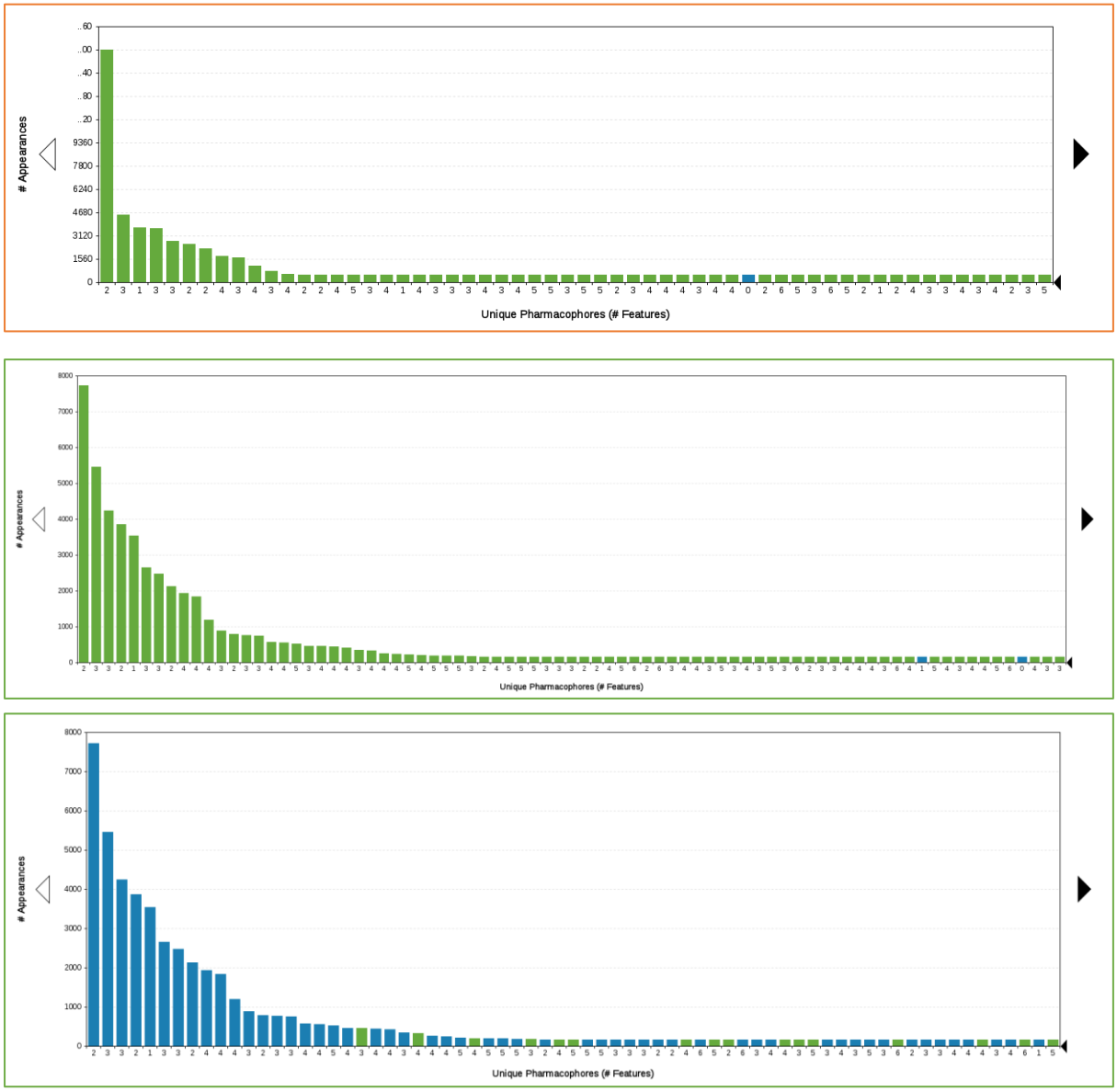


Figure 43 Pharmacophore Plots were generated by LigandScout after MD pharmacophores were created. Pharmacophore plots show sets of unique pharmacophore features with individual bars against the number of appearances in the entire trajectories. A unique pharmacophore feature is defined by one specific atom of the ligand interaction with a specific protein atom via a pharmacophore feature. By choosing a specific pharmacophore feature, bars in the interactive pharmacophore plot are shown in green. For the two plots at the top three pharmacophore features were chosen from frame 1 of both trajectories, interacting from protein atoms of the residues His378, Trp380 and Asn351. Above: open CRBN (orange); Middle: closed CRBN (green); Below: closed CRBN (green) including the water molecules: green bars are showing the sets of unique pharmacophore

features where water molecule HOH2037 was observed to interact with lenalidomide by a pharmacophore feature;

Finally, a third plot provides insights into the timely behaviour of the pharmacophore features during a simulation. The pharmacophore feature timeline provides all pharmacophore types occurring during the trajectory and are plotted against the number of appearances. The frames were binned with the size of 5,000, which subsequently calculated each 5,000 frames as one data point shown as a rough timeline in the plot. The trajectory of the closed CRBN and ligand showed in the sum 10 HBA and 7 HBD features during the entire time (Figure 44, 45). For the trajectory with the closed CRBN and the ligand, pharmacophores included 6 HBD and 6 HBA. The three core pharmacophores towards the residues Trp380, His378 and Asn351 were selected and shown over the timeline plot. Once more, the HBA feature from Trp380 can be observed to have the highest number of appearances over the whole timeline. In contrast, the other two pharmacophore features of His378 and Asn351 show a slight variation in number of appearances throughout the binned frames (Figure 44). While the trajectory with the open CRBN and ligand, showed a similar result for the HBA from Trp380, the other two core pharmacophore features diverge heavily in the last third of the trajectory. The HBA feature from Asn351 decreases to less than 300 appearances, whereas the HBD from His378 increases to a similar level as the constantly present Trp380 pharmacophore feature with over 4,500 appearances in the last third of the timeline (Figure 45).

Figure 44 The pharmacophore feature timeline shows pharmacophore feature presence of individual pharmacophore features after MD pharmacophores were created in LigandScout. For the trajectory of the open CRBN with lenalidomide several pharmacophores were detected and plotted as binned frames with a size of 5,000 frames per data point. H-bond donor (HBD) pharmacophores are shown in green, while H-bond acceptor (HBA) pharmacophores are shown in red. Above: Pharmacophore feature timeline shown for all pharmacophores from the open CRBN with several appearing in low numbers during the entire trajectory, while one prominent HBA that was shown as feature interacting with Trp380 as the highest and constantly present pharmacophore. Below: Three pharmacophore features were selected from the first frame of the open CRBN trajectory. Two HBA features were observed to interact with Trp380 and Asn351, while one HBD feature interacting by His378 residue was found to be present in similar presence like the Asn351 pharmacophore feature.

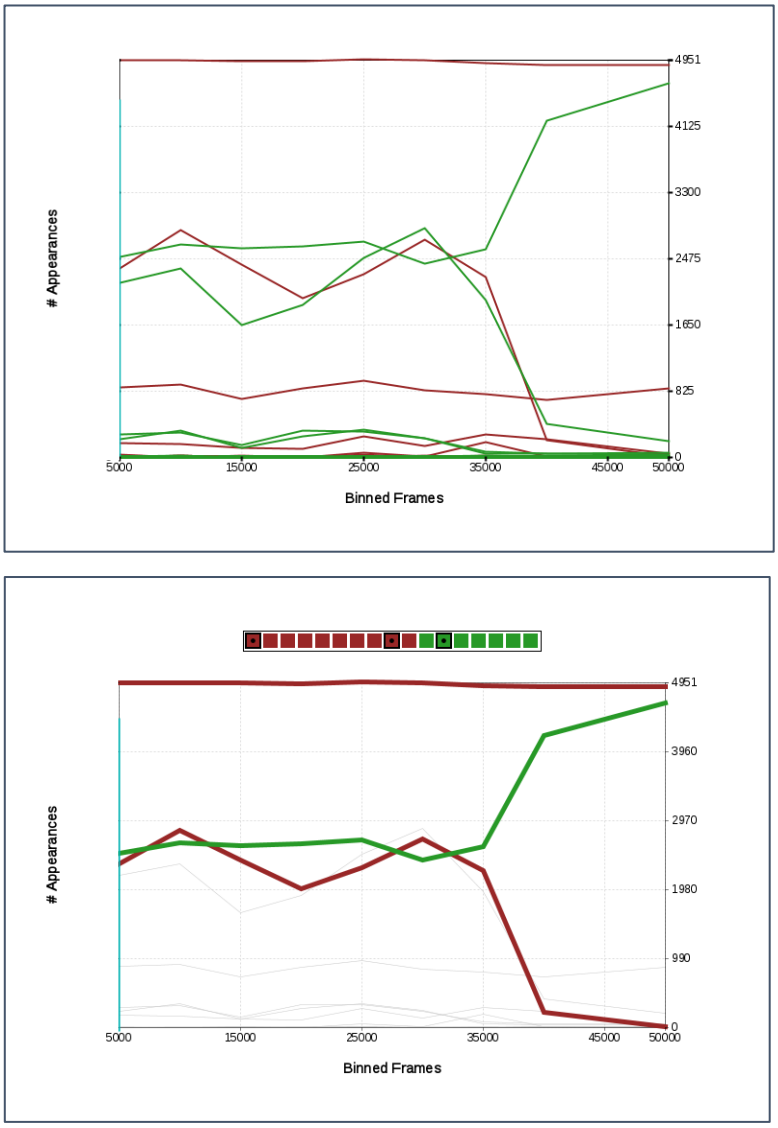


Figure 45 The pharmacophore feature timeline shows pharmacophore feature presence of individual pharmacophore features after MD pharmacophores were created in LigandScout. For the trajectory of the closed CRBN with lenalidomide several pharmacophores were detected and plotted as binned frames with a size of 5,000 frames per data point. H-bond donor (HBD) pharmacophores are shown in green, while H-bond acceptor (HBA) pharmacophores are shown in red. Above: Pharmacophore feature timeline shown for all pharmacophores from the closed CRBN with most features appearing in low numbers during the entire trajectory, while one prominent HBA that was shown as feature interacting with Trp380 as the highest and constantly present pharmacophore. Below: Three pharmacophore features were selected from the first frame of the closed CRBN trajectory. Two HBA features were observed to interact with Trp380 and Asn351, while one HBD feature interacting by His378 residue was plotted increasingly and taking over by the end of the trajectory over the Asn351 pharmacophore feature.

To sum up, the core pharmacophore features of the MD-based pharmacophore analysis by both analysis tools CCPTRAJ and LigandScout resulted in the core interaction features being captured for the residues Trp380, His378 and Asn351.

These residue interactions are the same three out of four core residues that were defined in the docking experiment as described above. The MD trajectories allowed to observe the dynamics of these unique features and find more difference in residue interactions to CRBN than the docking results made possible. HBA feature from Trp380 was found not only to be the most abundant but also throughout the entire trajectories the most stable feature of the binding pocket. Moreover, the analysis showed distinct differences between residue interactions of the ligand with the open or closed CRBN protein. CCPTRAJ showed overlaps in the differential analysis of the interacting residues to LigandScout. His357 was not generated as an interacting residue during the dockings but was only captured by CCPTRAJ H-bond and LigandScout interaction analysis. There it was plotted as one of two residues differentiating between the closed and open CRBN form, where it was observed, next to Phe381. In comparison to the docking results, for both trajectories interactions with the residue Ser379 from the binding pocket were observable.
The greatest differences were found with the residues Glu377 and His378, the latter being one of the core residues, for the MD-based pharmacophore analysis by LigandScout. While with the open CRBN the interaction of His378 was predicted to be more important than with Glu377, the results are almost perfectly reverse with the closed CRBN. Finally, this pharmacophore feature of His378 was found to be more stable and increasingly prevail towards the end of the trajectory with the open CRBN and ligand. In contrast to the closed CRBN trajectory, where the HBA from the His378 showed a slightly varying but steady number of appearances over the entire trajectory and as such was very similar to the HBD feature of Asn351. Finally, it can be noted that there was no significant interaction or pharmacophore feature found with any water molecule, but one in the trajectory of the closed CRBN and ligand at 4%.

6.3.2. Hotspot analysis: GRAIL

A recent publication by Schuetz *et al* (2018) provided a new tool for hotspot analysis of individual rigid protein structures or of whole structure ensembles from MD simulations, such as those of PDB entries or protein trajectories, respectively. A grid-based approach was applied to calculate and visually display so called GRAIL maps to find possibly new pharmacophore moieties that could guide rational ligand design.

All conformations of the CRBN trajectories were included into the calculations for the GRAIL maps [43]. The output allows for a relatively comprehensive overview since only 6 pharmacophores are defined and are addressing potential hotspots in the dynamic structures of the open and closed CRBN. Grid points of this approach including the pharmacophore concept were based on the CULT domain only instead of including the entire CRBN protein to decrease computation time by defining a reduced size of the grid box. The CULT domain was chosen as it comprises the IMiD binding pocket and therefore, the sensor loop. The visual output that was generated by GRAIL will be restricted in this part of the master thesis to the analysis of the average pharmacophore interaction fields from which useful information was retrievable. One GRAIL map is provided as an example for a visualization that does not contribute any information and as such adds no value to the ligand design development (Figure 51).

Some hotspot iso-surfaces support already known IMiD interaction patterns such as interactions to aromatic residues in the binding pocket. As such GRAIL maps of pharmacophores like aromatic to aromatic (AR-AR) and positive ionizable to aromatic (PI-AR) interactions were observable at the end of the binding pocket where not only aromatic residues reside but also lenalidomide shows H-bond interactions. These were found at a contouring level of 0.9 giving a high confidence to this visual display (Figure 46).

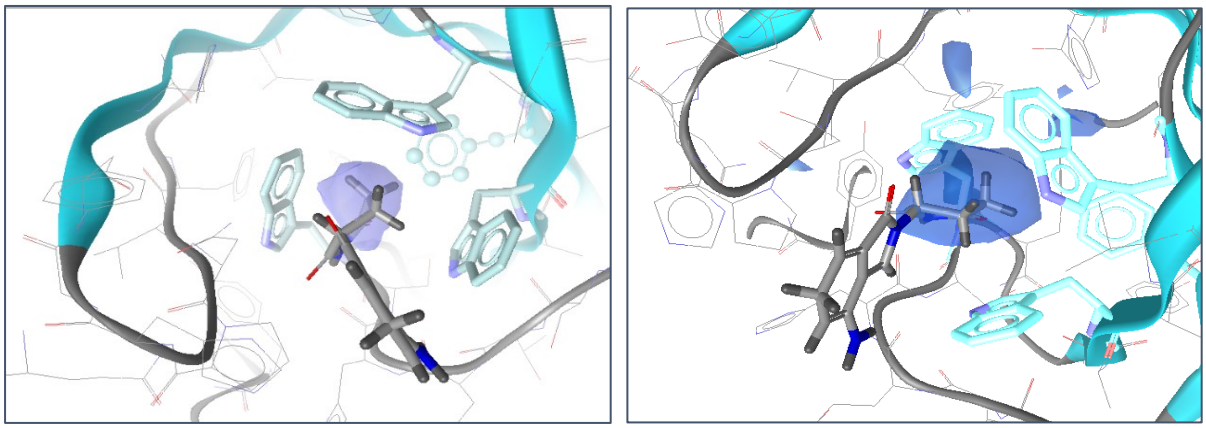


Figure 46 GRAIL map was generated for the closed CRBN with lenalidomide within a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields. Left: positive ionizable (PI) to aromatic (AR) (violet); Right: AR to AR (dark blue); both were displayed at a contouring level of 0.9 with a score going to 1.0. Bulky feature presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields.

Another common hotspot feature was captured by loading the interaction field of PI to negative ionizable (NI) pharmacophores based on the closed CRBN trajectory. At the opening of the binding pocket the residue Glu377 is located, where the average of the GRAIL map showed potential hotspot for an NI moiety in order to catch a residue interaction. This finding was not observable for the trajectory calculations based on the open CRBN conformation. Interestingly, lenalidomide was observed to interact with Glu377 in the closed CRBN state by two-fold more than with the open conformer in the pharmacophore-based interaction analysis generated by LigandScout. The contrary interaction field of NI to PI pharmacophore feature showed over the whole CULT domain only an iso-surfaces away from the binding pocket, around the residues Lys401. Both GRAIL maps were visualized at a score of 0.8 (Figure 47).

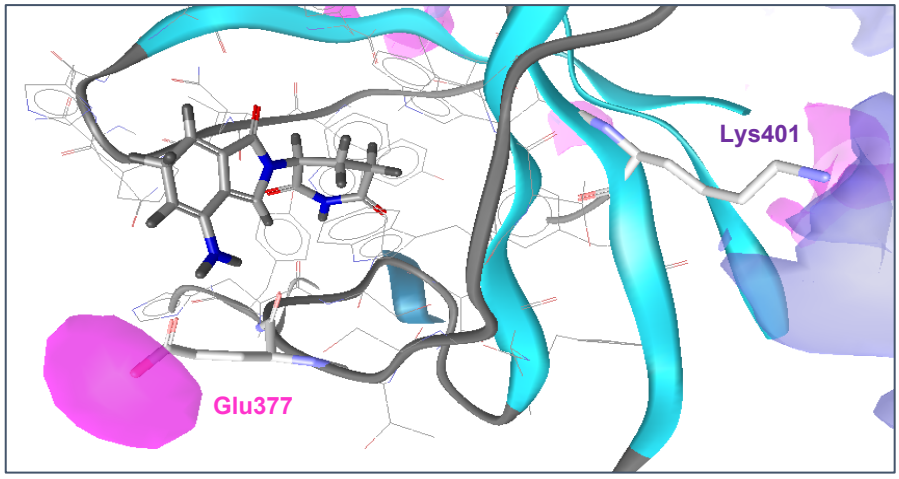


Figure 47 GRAIL map was generated for the closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: positive ionizable (PI) to negative ionizable (NI) (pink) and NI to PI (violet); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky feature presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields; GRAIL map of PI-NI around Glu377, while NI-PI field displayed around Lys401.

In contrast to the same GRAIL maps generated based on the open CRBN trajectory and displayed at the same score of 0.8 PI-NI interaction fields were captured in the vicinity of the sensor loop, where His353 moves throughout the trajectory. In addition, at the very end of the binding pocket the iso-surface covers Phe402. While NI-PI hotspots were only displayed on the surface of CRBN, stretching between

Lys401 and Thr387, which are not located close to the binding pocket area (Figure 48).

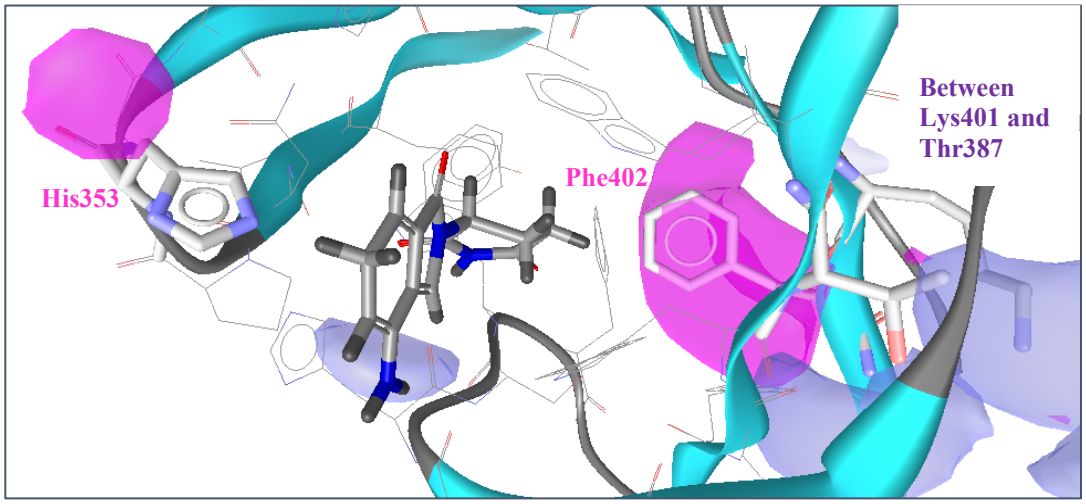
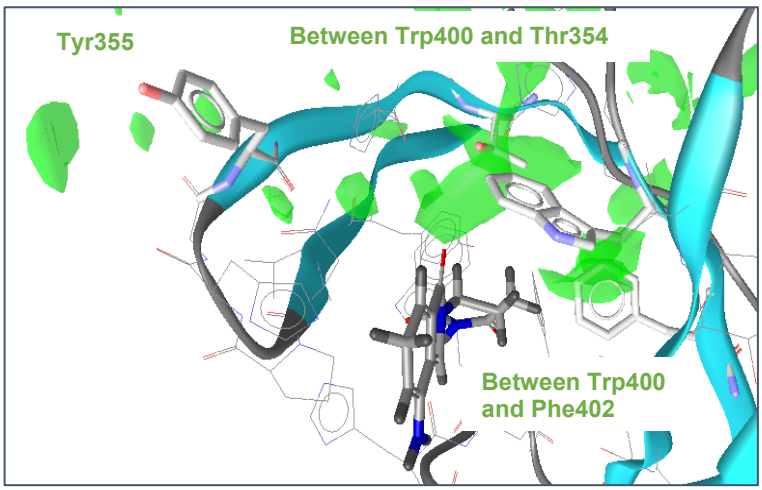


Figure 48 GRAIL map was generated for the open CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: positive ionizable (PI) to negative ionizable (NI) (pink) and NI to PI (violet); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields; GRAIL map of PI-NI around His353 and Phe402, while NI-PI field displayed between Lys402 and Thr387.

GRAIL maps were also calculated for the HBD-HBA pharmacophores and observed as interaction fields at a contouring level of 0.8 for both the open and closed CRBN protein. The open conformer showed more iso-surfaces around the binding pocket and close-by residues, such as the visualized area around Tyr355. While the other displayed potential interaction fields captured above discussed residues from the pocket such as around Trp380 and Trp400 and stretched even to parts of the pocket such as around residue Phe402. In comparison to the closed CRBN, where GRAIL maps pointed mainly an area between Pro352 and Trp380, again aforementioned residues of the binding pocket (Figure 49).



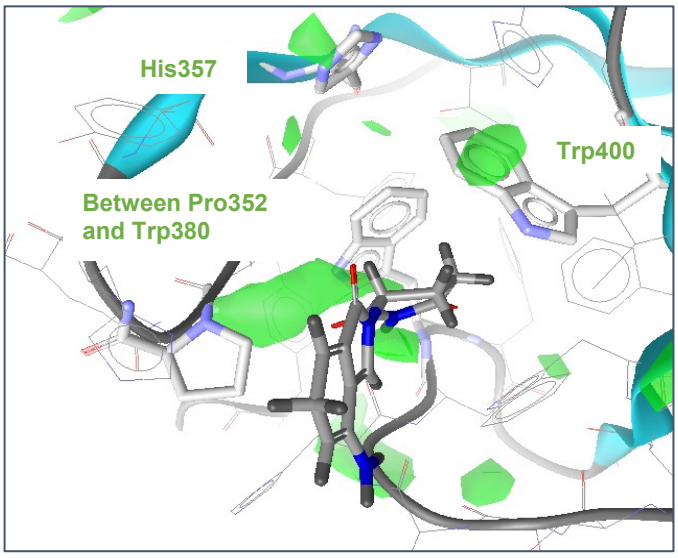


Figure 49 GRAIL map was generated for the open and closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: H-bond donor (HBD) to H-bond acceptor (HBA) (green); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN GRAIL map of HBD-HBA around Tyr354, between Trp400 and Thr387 as well as between Trp400 and Phe402; while open CRBN HBD-HBA interaction fields displayed around His357, Trp400 as well as between Pro352 and Trp380.

Like the HBD-HBA pharmacophore fields, GRAIL can also calculate the Halogen-bond donor (XBD) and acceptor (XBA) maps. The trajectories of the open CRBN showed GRAIL maps in the area between Phe402 and Trp400. In contrast to the closed CRBN trajectory where iso-surfaces were displayed between Phe352 and

Trp380. Both displays were set at a contouring level of 0.95 showing the distinct areas for potential XBD to XBA interactions with high confidence (Figure 50).

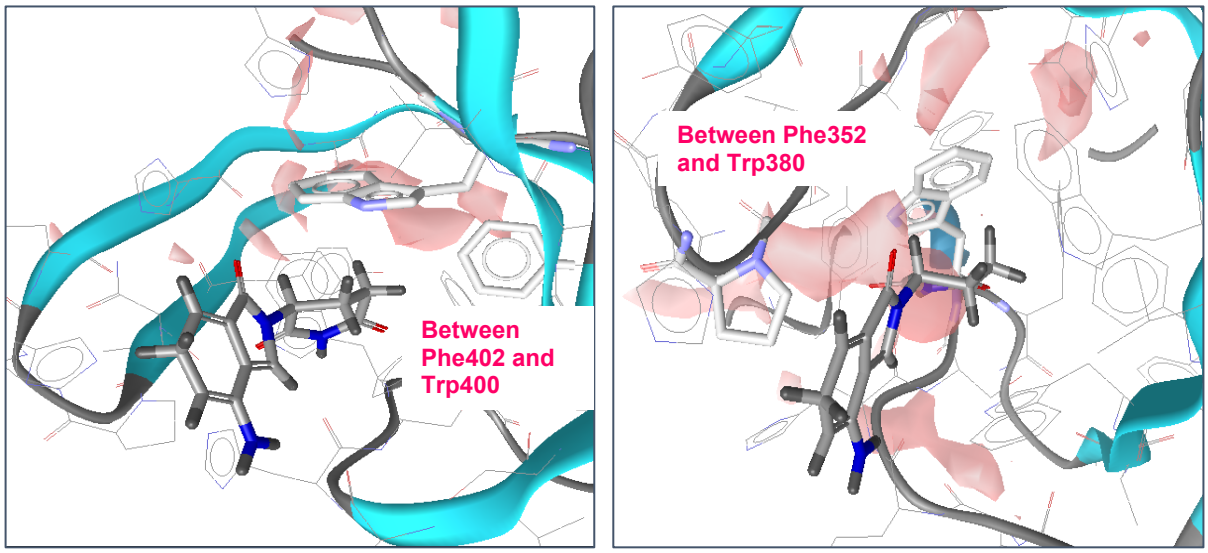


Figure 50 GRAIL map was generated for the open and closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: Halogen-bond donor (XBD) to Halogen-bond acceptor (XBA) (light pink); both were displayed at a contouring level of 0.95 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN GRAIL map of XBD-XBA between Trp400 and Phe402; while open CRBN XBD-XBA interaction fields displayed between Phe352 and Trp380.

One GRAIL map showed a distinctive difference between the calculations based on the closed and open CRBN trajectory. In Figure 51, the closed and open CRBN GRAIL maps for hydrophobic to hydrophobic (H-H) pharmacophore interactions are shown at the set score levels of 0.9 and 0.95. The GRAIL map of the open CRBN were barely observable and did not increase when the score was reduced to 0.5. In contrast, the GRAIL iso-surfaces of the closed CRBN where the majority of the binding pocket was observed to be covered by the H-H pharmacophore interaction field. The increase of the score to 0.95 showed an improved visual information output. The GRAIL map was only visualized between Trp380 and Trp400 (Figure 51). This example should show the challenge to retrieve valuable information from GRAIL maps and is dependent from the structural input as well as pharmacophore interaction field.

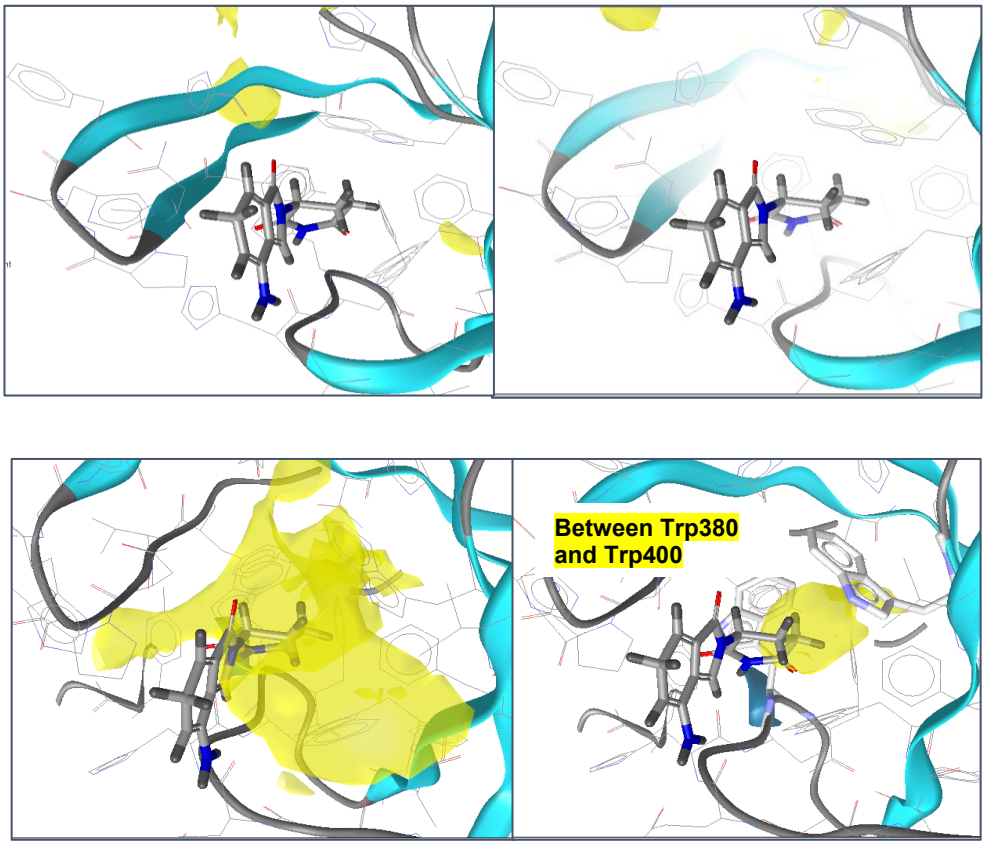


Figure 51 GRAIL map was generated for the open (above) and closed (below) CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: Hydrophobic (H) to H interaction (yellow); both were displayed at a contouring level of 0.90 (above) and 0.95 (below) with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN the GRAIL map for the H-H pharmacophore were displayed at a score of 0.95 in an improved way to retrieve better information for the ligand design development showing between residues Trp390 and Trp400.

A special GRAIL map was developed just recently and introduces calculations about the water molecules in the environment. As already once before in the docking experiments, water molecules also played a role here in the hotspot analysis to understand whether such water interactions to a protein or ligand could be exploited. The imported trajectories in LigandScout were not stripped of water molecules or ions, but otherwise were subjected to the same workflow. These grids calculate whether a water molecule from the environment should be considered “happy” or

“unhappy”. Happy water areas can be worked into the ligand design by replacing them, interacting with them, or even be avoided. However, unhappy water iso-surfaces should be considered to be displaced with an appropriately designed extension of the ligand. These special hotspots were carried out for both states of the CRBN protein. The interaction fields based on the open conformation showed only happy water areas such as around water in the near vicinity of the binding pocket and therefore to the lenalidomide ligand. Similar areas were found around residues Asn351, Trp380 and Tyr349. In contrast the GRAIL maps based on the closed CRBN showed both happy and unhappy water fields. One iso-surface embedded two water molecules WAT467 and WAT886 that were located at the site of the binding pocket. In contrast, happy water areas were found at the lower end of the binding pocket displayed, such as around residues Trp386 and Glu377 as well as including a water molecule WAT7720. The visual outputs for both CRBN conformers were generated at a score of 0.95 giving a high confidence to these results (Figure 52).

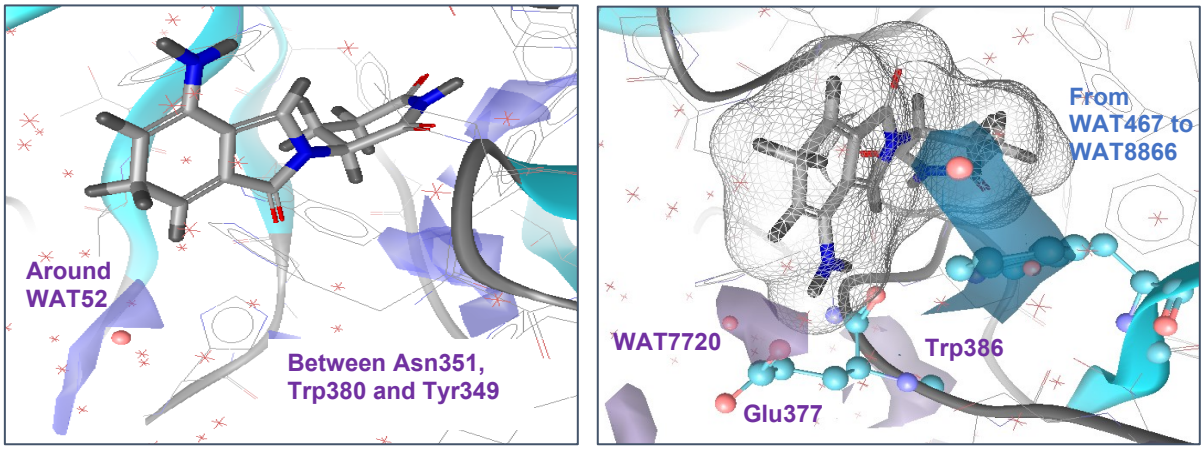


Figure 52 GRAIL map was generated for the open (left) and closed (right) CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields for a special GRAIL map: happy waters (happyH2O) (violet) and unhappy waters (unhappyH2O) (dark blue); both were displayed at a contouring level of 0.95 with a score going to 1.0. Bulky presenting amino acid side chains and relevant water molecules (red) are highlighted to improve the visualization of the GRAIL interaction fields: for the open CRBN GRAIL map of the happy H2O was around between Trp380, Tyr349 and Asn351 as well as the water molecule WAT522; for closed CRBN both water iso-surfaces were displayed: unhappy H2O interaction fields were visualized around water molecules WAT467 to WAT8866, while happy H2O were visualized between Trp386, Glu377 and the water molecule WAT7720.

To sum up, hotspot analysis is a powerful tool moving forward toward ligand design. In particular, GRAIL analysis includes all conformations from a trajectory and provides a user-friendly visual output based on a set of pharmacophores. For the trajectories in this master thesis containing the open and closed CRBN with the IMiD ligand lenalidomide, GRAIL maps showed distinctive differences and commonalities among the displayed iso-surfaces of the individual pharmacophore interaction fields. While some pharmacophore fields confirmed known H-bond interactions to residues of the binding pocket, such as AR-AR or PI-AR to the Trp residues at the end of the binding site, other GRAIL maps reproduced similar results to the docking experiments and MD-based pharmacophore analysis, such as the PI-NI iso-surface around Glu377. However, the GRAIL calculations were also able to display hypothesized interaction sites such as the open CRBN trajectory displaying an PI-NI area around the residue His353. Most often aforementioned residues such as Trp380 and Phe352 and residues more towards the opening of the binding pocket such as Trp400 and Phe402 were captured. Next to the pharmacophore interaction fields, a special GRAIL map also showed water molecules from the environment that were proposed to be displaced or should be replaced by the ligand developing the design and including similar interactions to these areas.

7. Discussion

7.1. Covalently binding PROTACs

The effects of IMiDs have been discovered after one of the greatest medical tragedies in history that led to the discovery of the E3 ligase CRBN. It is one of over 600 E3 ligase enzymes encoded in the human genome and so far, only a handful of them have been a topic of research [10]. Due to the simultaneous discovery of thalidomide inducing teratogenicity and natural peptidic compounds inducing protein degradation, two E3 ligases have been the focus of research to modulate the UPS mechanism for therapeutic purposes. However, ligands that were observed to induce protein degradation for E3 ligase CRBN and VHL differ not just in their physiochemical properties but also structurally [3], [10], [20]. While VHL-based PROTACs have comprised structures of several peptides linked together to interact with POIs, CRBN ligands are often derivatives of the IMiD structures. These drugs have been applied for various indications and thus were excessively researched, next to continuous efforts to try and understand the reasons of their teratogenicity. Ideally, PROTACs and MGs binding CRBN or other E3 ligases should recruit specific proteins for ubiquitination and subsequent degradation. These ligands induce PPIs not just with endogenous substrates but also selectively with neosubstrates that are degraded for therapeutic interventions [10], [14].

In this master thesis, IMiD like compounds were applied to understand possible binding modes to the protein CRBN. All IMiD drugs have been observed to share similar interaction patterns due to their common structure. The glutarimide moiety is shared by all of them and has been described to interact by H-bonds with the so-called aromatic cage that is comprised of three Trp residues in the end of the IMiD binding pocket of CRBN. In addition, residues such as His378 and Ans351 are found to interact with IMiD like compounds that share the glutarimide structure (Figure 29). This chemical moiety has been also proposed to stabilize the CRBN complex with the POIs. So far, CRBN was captured in two main conformational states. Most often the closed conformation of CRBN has been experimentally determined and structures are available in public PDB files. These structures have experimentally

defined coordinates for the atoms and include for the closed conformations most often a defined ligand in the IMiD binding pocket. IMiDs or IMiD like ligands are observed to contribute to the stability of CRBN structures that are co-crystallized with the adaptor protein DDB1 and POIs such as BRD4 and CK1α. Due to the flexibility of certain regions, it is more challenging to capture CRBN conformations without ligand and those in an open state. There are only two X-ray structures with open CRBN conformations, which have low resolution. The applied open structure in the master thesis 6BNB has a resolution of 6.34 Å (Figure 27). Nevertheless, obtaining experimentally a structure with an open CRBN conformation or any other conformation between the two states that have been so far described has been a challenge, especially due to the highly flexible regions of CRBN. Therefore, researchers have come up with different experimental ways to define the hypothesized switch from the open to the closed state upon binding of a ligand in the IMiD pocket of CRBN.

One leading hypothesis is based on the findings of a special flexible region in the CULT domain of CRBN that is called the sensor loop. It is comprised of several residues that are part of the IMiD binding pocket. In most of the PDB structures the loop is experimentally determined, but some CRBN structures define this loop in a different position. This handful of structures have the sensor loops modelled in. It is one of the reasons behind the challenge of generating a complete open CRBN structure. The sensor loop has been found to be involved in the switch between the two extremes of the described CRBN states. Especially, the His353 residue has become of great importance for a new approach of PROTAC design.

If ligands would covalently bind as PROTACs or MGs to E3 ligases, it would turn the pharmacokinetics into a 2-body reaction and thus, imply favourable properties. However, covalently binding compounds have been rarely explored for E3 ligases. With only a few above-mentioned examples from a handful of publications, where such compounds were only designed to form a link to Cys residues, the published data has rarely considered His residues for ligand design [2], [6], [13]. Developments in the synthetic chemistry have shown that compounds with sulfonyl fluoride and fluorosulfate moieties are reactive and lead to selective labelling of His residues in proteins. In addition to the physiochemical implications of such small molecules and the advances in chemistry for ligand design, His residues have been reported to be

promising covalent anchor points. They have a low surface presence and appear to be frequently part of catalytic sides. Further, data has shown that His residues are 8.2 times more often present than the statistically expected natural frequency as well as preferably are available within 4.5 Å to drug like molecules in binding sites, which is at least twice higher than by natural abundance [33], [46]. These promising findings resulted in rationally designed ligand that would induce CRBN-driven TPD. Such compounds would include the glutarimide moiety to capture known stabilizing interactions in the deeper part of the IMiD binding pocket and include reactive, site selective structures that form a covalent link to a His residue.

Two ligands called EM12FS and EM12SO2F have been synthesized and tested for their ability to form a covalent bond to a His residue in the IMiD binding pocket. Their structure includes a glutarimide moiety and sulfonyl fluoride acting as electrophilic warhead towards His residues at the surface of the binding pocket. Data from experiments with mass spectrometry suggest a covalent link to His353, a residue on the surface of the pocket [33], [46]. Since there is no crystallized structure of CRBN, where either of these ligands would show the covalent bond defined by electron density, an *in-silico* approach was applied to define most likely poses of these two compounds.

By the covalent docking approach, His353 was evaluated as the targeted residue for the covalent bond. Further, poses with an open conformation of CRBN were also tested for plausibility, and possible interaction patterns there as well as differences to the binding pocket of the closed state were investigated. Finally, a covalent bond with either N atom in the side chain of His353 was formed but the various docking runs had different settings. The generated poses by covalent docking were filtered and selected by presence of core residues including H-bonds to Trp380 and Trp386 as well as often previously described residues such as His378 and Asn351 (Figure 29). An additional selection criterion included the scan by the Rarey Torsion Tool, where poses with energetically unfavourable torsions were excluded [47] (Figure 25, 30). Eventually, plausible poses were retrieved for both forms of CRBN with extremely different residue interaction patterns next to the core residues. However, more diverse interaction patterns were found with the closed CRBN conformation. These results included residues such as Pro352, Glu377 and Trp400. Interestingly, the patterns diversified when the analysis focused on which N atom of the His imidazole were covalently docked onto. For NE2 anchored poses additional H-bond

interactions with Pro352 and Glu377 were shown, while ND1 docked ligands made interactions with Pro352, Trp400 and His353 (Figure 36, 37, 38). Pro352 has not been found as often described as a residue that is part of other PROTACs binding network in the pocket. However, the aromatic triad is a well-known interaction pattern of IMiD like ligands and included next to Trp380 and Trp386 also Trp400 that was also predicted as an interaction site for the covalently binding compounds. It was interesting to observe that the H-bond was formed to the backbone of His353 in addition to the covalent link. These poses were exclusively found for the open conformation, which supports the hypothesis that this residue as a part of the sensor loop could have special implications in the switch from the open to the closed conformation and even serve as a first interaction point between ligand and E3 ligase CRBN [14].

Eventually, the differential analysis was also conducted for the interaction patterns of the two ligands, which were found to have distinctive results for those poses differing in the anchoring protein link. The EM12SO2F compound has been characterized as very unstable in human plasma and to block the degrading efficiency of CRBN rather than to induce it. H-bonds for poses with EM12SO2F have been found to residues Pro352 and Glu377. In contrast, the more stable and degrader ligand EM12FS was observed to generate poses where H-bonds were made with residues such as Pro352, Trp400 and His353. These are similar findings to ND1 covalently docked poses that did not discriminate in the input between both ligands. Poses that were anchored by the ND1 atom link showed preferred tendencies to dock the ligand EM12FS. In contrast, NE2 anchored poses remained after the selection mostly with the ligand EM12SO2F (Figure 36, 37, 38).

Overall, it needs to be noted that the greatest overlap of interactions was not detected for residues in the binding pocket but with residues from the co-crystallized POIs. The POIs of both PDB entries 6BNB and 5FQD were chosen as models, not just for the open and closed CRBN conformation but also for showing exemplary ligand interactions with the POIs BRD4 and CK1α, respectively. In the context for a rational ligand design, docking without the POIs would have provided little information if only CRBN was used. Here the two POIs can be used as an example for possible PPIs and as such be applied in the ligand development in these specific case studies. These residues included specifically Lys18, Ile35 and Ile37 (Figure 35).

Some distinctive interaction patterns between the different settings that were discussed above, were comparable to the analysis for these residues, too. While H-bond interactions were only established with the closed CRBN conformation to the specifically mentioned three residues of POI, none were found when the ligands were docked in the open structure. This supports the notion that PROTACs bring the two proteins E3 ligase and POI together before ubiquitination that induces signal transfer and subsequent degradation. Further, these results can refer to the ongoing discussion on the contribution of PPIs established or enhanced by mechanism of action of PROTACs and MGs. Different relative presence of H-bond interactions to the POI residues were also observed between the differently N-atom docked poses. ND1 covalently linked compounds were found to only induce interactions with Ile37, while NE2 docked poses showed a more diverse interaction pattern including all three residues.

Recently published literature has reported on certain residues to be common for IMiD like compounds in the binding site of CRBN. While many of these results were established by similar *in silico* approaches, specific interactions such as those of the glutarimide moiety have been described repeatedly. Firstly, the covalent docking could reproduce the interactions to the same residues that included Trp380, Trp386, His378 and Ans351. In addition, the covalent ligands were proposed to selectively bind to His353. This result can be supported since no plausible poses was generated with any other His residues close to the binding pocket. Nevertheless, including the open conformation of CRBN to the workflow allowed to observe a smaller number but still achieved plausible poses for the two covalent ligands. Consequently, the most interesting findings were the differentiating interaction patters only between EM12SO2F or EM12FS docked poses that were found to be analogous to interaction results only between NE2 or ND1 docked poses, respectively. All these dockings included Pro352, which thus far, is a rarely described interacting residue when reviewing results of previously publications. However, Glu377 and Trp400 formed H-bonds in the covalent dockings and confirm their involvement in binding IMiD like ligands to the pocket.

Next to the special role of His353, the data suggests a correlation between NE2 docked poses and the non-degrader molecule EM12SO2F. Even more worth mentioning are the findings of ND1 docked poses and the degradation inducing EM12FS compound that showed H-bond interactions to Trp400, Pro352 and to the

backbone of His353. This supports the key role in the switch between open and closed conformation of CRBN states as well as the stabilizing ligand interactions from the binding pocket for a degradation inducing complex.

7.2. Hotspot analysis of dynamic CRBN

Advances in computational power have made simulations and their analysis of CRBN dynamics possible and more reliable. MD simulations have become a valuable tool for the investigation of molecular interactions and their dynamic behaviour [40], [41]. Their impact has grown especially since force fields and sampling methods have improved. In this work it allowed to investigate simulated movements and flexibility of the open and closed CRBN with lenalidomide in the IMiD pocket. The advantage of this *in silico* method over other approaches is the ability to generate several conformations that have not been captured in rigid crystal structures. Analysis based on experimentally defined crystal structures of CRBN are an important first step in discovering the binding pocket and its residue interactions to ligands that were determined only for the closed conformation of CRBN. Several publications have already reported on the challenging fact that CRBN has very flexible regions, which makes the experimental determination of structures more difficult. Only two structures have been experimentally obtained and defined in the open conformation providing two extremes of the CRBN dynamic. The two conformational states were used as a starting point for setting up MD simulations. The MD simulations provide the frame for the motility of the CRBN protein in either state. Interactions to the ligand lenalidomide in the binding pocket were able to be observed during the time of the simulation and provided new insights into residue interactions and a possible outlook onto new ligand design.
MD simulations generate huge amount of data collected as so-called trajectories that can be subjected to a plethora data analysis tools and approaches. In this work a better understanding of the dynamics of CRBN and its simulated conformations that were impossible to be captured by traditional crystallography methods was aimed for. Therefore, the flexibility per residue presented in an RMSF plot supported the limitations of conventional structure determination (Figure 39). Certain areas of CRBN independent of the conformation were found to be very flexible. Next to the residues of the CULT domain that showed partially great distances in terms of

motility, the contact area to DDB1 was plotted as the most flexible region in the MD simulations. This is because only CRBN protein was included in the simulation and lacked the stabilizing factor of the adaptor protein DDB1 or any other chains such as the co-crystallized POIs. These residues were found to be more flexible with the open conformation than with the closed state, even though both had the proposed to stabilize IMiD ligand in their pocket as part of the simulated system (Figure 39).
The ligand lenalidomide was present for both MD simulations and allowed interaction studies with the residues of the binding pocket that were found to partially differentiate depending on the conformational state of CRBN protein. The analysis of the H-bond network of CRBN to lenalidomide was carried out with two tools, the "AMBER" based CCPTRAJ and LigandScout. Both programs showed the important and stable H-bonds with the core residues that were previously described in literature and were captured in the docking experiments alike (Figure 40, 41, 42). Trp380, Trp386, His378 and Asn351 residues are interacting for the largest part of the time scale during the trajectory with the glutarimide moiety of the IMiD ligand. The "AMBER" based analysis tools of CCPTRAJ showed the same interacting residues with the open and closed CRBN conformations including Glu377, His353, Trp400 and Ser379 (Figure 40). Only the interaction with residue Ser379 was not observed in the docking experiments that included lenalidomide derivatives. This shows one of the advantages of MD simulations over docking experiments with more rigid protein structures. Other residues such as Glu377 and Trp400 have been hypothesized previously to form H-bonds with various IMiD like compounds and were also observed. Of special interest is the residue His353 that is part of the sensor loop, which was proposed to play a key role in the conformational switch as well as in the stabilization of the CRBN upon ligand binding. While this residue was covalently docked in the above-mentioned experiments and was suggested to form a covalent bond with the lenalidomide derivative ligands, it also was only found for the open conformation to interact with an additional H-bond to the backbone of His353. Here, a contributing finding can be extracted from the MD simulation experiments. His353 interaction with lenalidomide was found in two sets during the trajectory based on the open CRBN conformation over the one set that was observed for the closed CRBN structure to interact with the same residue.
The second tool for an efficient analysis of MD simulations was provided by the program LigandScout. It allowed to generate MD-based pharmacophores from the

entire trajectory and thus, create several plots where the individual pharmacophores and corresponding interactions were made observable in a relative and timely manner. Especially the MD pharmacophore plot showed a comparable output of the H-bond interactions that was analysed by CCPTRAJ, too. While the above-mentioned core residues were calculated to provide the highest interaction results with residues Trp380, His378 and Asn351, interactions with other residues such as Glu377, Ser379 and Trp400 were also reproduced (Figure 41, 42). However, some of these interactions relatively differed between the open and closed CRBN trajectory. MD-based pharmacophoric feature was the second highest for Glu377 to form H-bonds with lenalidomide binding in the closed CRBN pocket. In contrast, such high interactions percentages were caught for the open CRBN trajectory only with His378. These two residues almost perfectly invert their H-bond contribution to lenalidomide between the two conformational states. In addition, two more residues were detected in the interaction patterns of the MD simulation with the open CRBN protein showing His357 and Phe381 to form H-bonds with lenalidomide, despite being present below 1% during the entire trajectory. Interestingly, to note is also the interaction to His353 which interaction is captured below 1% for both trajectories simulating the open and closed CRBN conformations.

H-bond analysis from trajectories and docking results based on more rigid conformational structures already provide a good understanding of which residues could be beneficial to improve interactions with and consequently, how to modify ligand design for more stable interactions with CRBN. Their results show partially the interactions to the same residues. MD-based pharmacophore models have increasingly been applied in this area to advance ligand design. A recently published hotspot analysis tool was reported to include the pharmacophore concept to provide information on the most likely interaction type that can be formed with individual residues. Its advantage is based on using the trajectory that includes the various conformations of the protein. After defining the area of the binding pocket in a grid, the potential pharmacophoric interactions are calculated over all conformations in the trajectory and can be displayed as interaction fields. These calculations are based on the grid-based method GRAIL [43]. An advantage of this approach is that the generated output is manageable in terms of size. For each unique pharmacophoric feature an interaction field is produced. GRAIL maps were calculated for the open and closed CRBN trajectory. Some pharmacophore interaction fields verified the

correct application and previously discussed interaction patterns. Such as the AR-AR or PI-AR iso-surfaces that were displayed around the IMiD binding pocket, where the aromatic cage of three Trp is present and known to form H-bonds to the glutarimide moiety (Figure 46). Likewise, GRAIL maps showed HBD-HBA interaction fields around above-mentioned residues, such as Trp400, His357, Pro352 and core residues known to establish H-bond interactions such as Trp380 (Figure 49). Interestingly, at a high iso-contour value, the GRAIL maps for XBD-XBA visualized a signal around the same residues (Figure 50).

The IMiD binding pocket is known for its hydrophobic surfaces and conserved interaction patterns to its pocket residues. Nevertheless, at high iso-contour values the visualization of GRAIL maps of hydrophobic interactions displayed a different view between the open and closed CRBN (Figure 51). While almost none of the hydrophobic pharmacophoric fields were visually observable for the open CRBN trajectory, at the same score level the hydrophobic GRAIL map covered almost the entire binding pocket area of the closed CRBN pocket. Such findings support that the GRAIL calculations are a valuable tool for CRBN as in this study case, but also strengthen the results from recently published results. These *in silico* studies tried to assess the PPIs of CRBN to its POI where lenalidomide was suggested to induce hydrophobic shielding that would increase and add to the stability of a degradation inducing protein complex.

GRAIL maps for PI-NI pharmacophores showed the most interesting iso-surfaces for CRBN ligand design and development. The residue Glu377 in the closed CRBN conformation was found by GRAIL-based grid calculations to have a pharmacophore potential, where an IMiD based ligand from the binding pocket with a positive ionizable moiety is predicted to interact with the negatively ionizable structural part of the residue and thus, form a potentially potent interaction. In contrast to the open CRBN conformation, where the GRAIL maps based on the respective trajectory generated iso-surfaces of the same score level around two different residues, Phe402 and His353. Especially, the residue His353 often showed pharmacophoric feature potentials based on *in silico* findings with the IMiD ligand lenalidomide in this master thesis.

7.3. Outlook

Overall, IMiD like ligands to bind the CRBN pocket are proposed to include structural moieties that would interact with residues such as Glu377, Pro352 and His353. These residues have been observed in both types of approaches as the analysis of docking and of the MD trajectories showed and should be explored further. In contrast, other residues have been mentioned more rarely by *in silico* methods before. Among those, primarily Ser379 can be mentioned. Interactions to this residue have been found for the docking experiment, too, but those poses have been excluded due to the filtering criteria. In the MD simulations, this specific residue was observed to establish H-bond interactions to lenalidomide that is in binding pockets of both closed and open CRBN conformations. Thus, experiments targeting this residue could be engaging and perhaps highlight some other design possibilities since it has been hardly explored. However, the unique position of His353 at the outer end of the sensor loop and the difficulty to experimentally determine this flexible region should not be underestimated. The advances in synthetic chemistry and the potential of selectively targeting His residues combines the potential for a new mechanism of action for CRBN inducing degradation.

The mechanism of PROTACs bears the potential to selectively induce TPD for specific substrates in the cell. Many efforts have been subjected to design such ligands for the E3 ligase CRBN and yielded in the great diversity of structures available today. The special glutarimide moiety was found in previous publication as well as in the frame of this master thesis to be a core feature for anchoring IMiD like ligands in the binding pocket and to achieve selective interaction with CRBN. Exploration for new moieties in the surfacing part of the potential compound should be based on the rational findings that could be provided by *in silico* methods. Nevertheless, conducting further experiments will confirm or disprove the predicted residues of this work.

In conclusion, this master thesis provides several starting strategies for ligand design that propose the mode of covalently binding molecules inducing degradation. In addition, these results were for the larger part reproduced with a recently reported hotspot analysis based on the simulated dynamics of CRBN protein and the modelling as such can guide further rational ligand design.

8. List of Abbreviations

Å	Ångstrom
ACE	Acetyl group
ADME	Absorption, Distribution, Metabolism, Excretion
AMBER	Assisted Model Building with Energy Refinement
AR	Aromatic
Asn	Asparagine
ATP	Adenosine Triphosphate
BRD4	Bromodomain containing protein 4
CELMoD	Cereblon E3 ligase modulatory drug
cIAP	Cellular Inhibitor of Apoptosis Protein
CK1α	Casein Kinase 1 alpha
Cl-	Chloride ion
CRBN	Cereblon
CRL4	Cullin-4-RING E3 ligase
CUL2	Cullin 2
CUL4	Cullin 4
CULT	cereblon domain of unknown activity, binding cellular ligands and thalidomide
Cys	Cysteine
DDB1	Damaged DNA Binding protein
DEL	DNA encoded library
DNA	Deoxiribbonucleic acid
DUB	De-Ubiquitinating enzyme
E1	Enzyme 1
E2	Enzyme 2
E3	Enzyme 3
ELOB/ELOBC	Elongin B / C
FBDD	Fragment Based Drug Design
FBLD	Fragment Based Ligand Design
fs	Femtoseconds
Gln	Glutamine

Glu	Glutamic acid
GRAIL	GRids of phArmacophore interaction fieLds
H	Hydrogen
H-bonds	Hydrogen Bonds
H-H	Hydrophobic
HBA	H-Bond Acceptor
HBD	H-Bond Donor
HECT	Homologous to E6AP C-terminus
Hhep	Human Hepatocytes
Hif-1α	Hypoxia-inducible factor 1 alpha
His	Histidine
HLM	Human Microsome Liver
HOH	Water
HTS	High throughput Screening
IAP	Inhibitor of Apoptosis Protein
IKZF1/3	Ikaros Zinc Finger 1/3
Ile	Isoleucin
IMiD	Immunomodulatory Drugs
kDa	kilo Dalton
KEAP1	Kelch-like ECH associated protein 1
Lys	Lysine
M1	Methionine 1
MD	Molecular Dynamics
MDM2	Mouse Double Minute homolog 2
MG	Molecular Glues
MM	Multiple myeloma
MOE	Molecular Operating Environment
MS	Mass Spectrometry
MW	Molecular Weight
N	Nitrogen
Na+	Sodium ion
NAE	NEDD8-activating enzyme
ND1	delta Nitrogen

NE2	epsilon Nitrogen
NEDD8	Neural precursor cell expressed developmentally down regulated 8
NI	Negative Ionizable
NMA	N-methyl amide group
NME	Amine group
ns	Nanoseconds
PDB	Protein Data Bank
Phe	Phenylalanine
PI	Positive Ionizable
POI	Protein Of Interest
PPI	Protein-Protein Interaction
Pro	Proline
PROTAC	Proteolysis Targeting Chimaeras
PTM	Post-Translational Modification
RBR	RING between RING
RING	Really Interesting New Gene
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
SAR	Structure-Activity Relationship
Ser	Serine
SMAC	Second Mitochondria derived Activator Caspases
SMARTS	SMILES arbitrary target specification
SMD	Steered MD
TBD	Thalidomide-binding Domain
TPD	Targeted Protein Degradation
Trp	Tryptophane
Tyr	Tyrosine
UAE	Ubiquitin Activating Enzyme
Ub	Ubiquitin
UPS	Ubiquitin Proteasome System
VHL	Von-Hippel-Lindau
VS	Virtual Screening

WAT	Water
XBA	Halogen-Bond Acceptor
XBD	Halogen-Bond Donor
XIAP	X-linked Inhibitor of Apoptosis Protein
μM	Micromolar

9. List of Figures

Figure 1 Mechanism of action of heterobifunctional molecules called Proteolysis targeting chimeras (PROTACs): Hijacking the ubiquitin proteasome system (UPS) by bringing E3 ligase and protein of interest (POI) in close proximity and inducing ubiquitination on Lysine residues and subsequent degradation of the POI by the proteasome (schematic figure adopted from [8]..... 11
Figure 2 Three main enzymes orchestrate the ubiquitin (Ub) transfer to the substrate. In an ATP dependent step, the Ub activating enzyme (UAE) E1 is the first protein activating the Ub and transferring it to the next catalytic enzyme of the cascade. The Ub conjugating enzyme E2 takes over the Ub to transfer it to Ub ligase enzyme E3. This enzyme connects the post-translational modification (PTM) via an iso-peptide bond to specific Lysine (Lys) of the substrate protein. The multi-step process repeats since also Ub has several Lys residues and polyubiquitin chains can form. (schematic figure adopted from [9]) 12
Figure 3 The initial Ubiquitin (Ub) is transferred and attached to a specific Lysine (Lys) residue. From the initial Ub the signal is elongated to a polyubiquitin chain that can influence the degradation mechanism. Depending on which Lys residue the Ub signals are attached different mechanism are activated in the cell. A well-described linkage to Ub chains is Lys11 or Lys48, in contrast to the Lys6 or Lys29 Ub signals are also known to induce the UPS and proteasomal degradation but are less described in literature. In contrast, there are Lys residues such as Lys63, providing signals outside the UPS machinery, like for DNA repair and activation of transcription factors. In addition, different Ub patterns have been observed to influence the substrates E3 ligases recognize. The diverse patterns of the PTM lead to protein selectivity and specificity for degradation. Ubiquitin is attached to 7 lysine sites in figure: K6, K11, K27, K29, K33, K48, K63; and N-terminal Methionine (M1). (schematic figure adopted from [9])..... 13
Figure 4 Formation and structural subunits of 26S proteasome (schematic figure adopted from [1]) 14
Figure 5 Depending on how the Ub transfer is carried out by the E2, the ternary complexes with E1, E2 and E3 belong to different subgroups. From left to right: (1) RING (Really Interesting New Gene), (2) HECT (Homologous to E6AP C-terminus) and (3) RBR (RING between RING). (schematic figure adopted from [9])..... 15
Figure 6 Structures of various small molecules modulating the three enzymes E1, E2 and E3. From left to right: (A) PYR-41 is a known covalently binding E1 ligand, (B) MLN4924 inhibits the closely related E1 protein called NEDD8 activating enzyme (NAE) that signals via NEDD8 molecules, (C) NSC697923 and (D) CC0651 are one of the very few E2 targeting

inhibitors that cause great structural rearrangements and subsequently, the Ubiquitin is not ligated to the Lys residues. (molecular structures adapted from [12]) 17
Figure 7 Structural examples of RING E3 ligases with adaptor proteins, protein of interest (POI) chains, small molecules in respective binding pockets. From right to left: CRBN ligase in a complex of damaged DNA binding protein (DDB1)-Cullin-4 (CUL4); VHL in complex with Elongin B/C (ELOB-ELOBC)-cullin 2 (CUL2)-Rbx-1 (adopted from [10]) 19
Figure 8 The question remains at which state ligands make the interactions to CRBN to cause the switch between the open and closed conformation and what other conformations does CRBN explore during this process. The figure represents the hypothesis based on the two conformational states captured and described by Watson et al (2022) [14]. 22
Figure 9 Watson et al (2022) could already observe that ligands binding to the IMiD pocket such as pomalidomide at various concentrations would induce conformation re-arrangement of CRBN to different extents. This fraction was reported to retain the closed conformation. The sensor loop area was not found in a disordered state but supported stable conformation [14]. 23
Figure 10 Schematic overview of the targeted protein degradation induced by (A) IMiDs and (B) PROTACs. (schematic figure adapted from [19]) 24
Figure 11 IMiD binding site residues His-378, Trp-380, Trp-386, and Phe-402 and ligands thalidomide (yellow), lenalidomide (green) and pomalidomide (blue) are shown as not the only conserved regions of CRBN as overlay almost perfect between E3 ligase of various species, including human (purple; PDB code 4TZ4), chicken (blue; PDB code 4CI2), mouse (yellow; PDB code 4TZ4) cereblon (adapted from [26]). 27
Figure 12 Thalidomide was marketed under the name Distaval in UK to pregnant women as a very safe sedative (adapted from [7]). 29
Figure 13 Thalidomide as a stereoisomer comes in two forms. Its production contained a racemic mixture with both enantiomers (left: (R+) vs right: (S-) isomer) and was reasoned to be one of the causes for the tragic consequences for the unborn children or patients taking the IMiD drug that was prescribed as a sedative to pregnant women. However, the conversion from isomer tot h other can happen during physiological conditions. (adapted from [7]). 31
Figure 14 Schematic representation of E3 ligases highlighting examples of ligases in the human body (adapted from [10]). 33
Figure 15 Molecular structures of two PROTACs that were observed to have varying pharmacological properties while only the linker length is different. Left: dBet1 vs Right: ARV-825 (adapted from [12]) 33

Figure 16 Structures from two PROTACs developed for the E3 ligase KEAP1 that are non-peptidic and were proposed to bind covalently to Cysteine residues of the E3 enzyme (adapted from [11]) 34
Figure 17 Electronic structure calculations for the three IMiDs and their derivative compound EM12 (adapted from [32]) 35
Figure 18 Molecular structures of the IMiD derivative EM12 and the two covalent ligands EM12FS and EM12SO2F (from left to right), including individually depicted molecular weights (MW), logD values and results for stability in human plasma, microsomes (HLM), and hepatocytes (Hhep) (from top to bottom) (adapted from [31]). 37
Figure 19 Overview of typical covalent docking workflow that was to a greater extent followed (adapted from [36] 38
Figure 20 (a) Structure for thalidomide, EM12 and lenalidomide shown that vary only in the functional groups of X and R. (b) crystal structure of CRBN (grey) with CK1α (gold), DDB1 (cyan) and lenalidomide (green) in the binding pocket. (c) Zoom into binding pocket with specifically depicting residue His353 and lenalidomide showing the distance of 3.2 Å (adapted from [33]). 39
Figure 21 Schematic overview of a typical workflow starting with a PDB structure as the basis for an MD simulation and using the generated trajectory with all conformations of the simulated molecules for the calculations of the hotspot analysis called GRAIL that includes the pharmacophore concept to display pharmacophore fields in a defined grid (adapted from [43]). 44
Figure 22 Overview of the workflow included in this master thesis. Two main pillars with one introducing Covalent Docking and the other MD simulations aiming to assess possible interaction of covalent ligands to CRBN and the dynamics of CRBN conformers, respectively. 46
Figure 23 Two PDB structures shown used in this master thesis and coloured by chain ID with DDB1 (orange), CRBN (violet) and protein of interest (light blue). Left: open CRBN state (6BNB); Right: closed CRBN state (5FQD); 46
Figure 24 Left: Sulfonyl amide substructure proposed to label Histidine residues and perform a covalent bond. Middle: One of two small molecules hypothesized to form a covalent bond with CRBN residue His353, called EM12SO2F, shown with an N-link atom towards a Histidine. Right: the other small molecule hypothesized to form a covalent bond with CRBN residue His353, called EM12FS, shown with an N-link atom towards a Histidine. 50
Figure 25 Both ligands hypothesized to bond covalently to the CRBN residues His353 shown without the leaving group F-atom. Left: EM12SO2F; Right: EM12FS. Exemplary small molecules for the visual output of the traffic-light system-based display whether the torsions

of the rotatable bonds strained. Red to green standing for strained to relaxed torsion angles. Torsion Scan carried out based on Rarey Tool in Maestro [41]. 50
Figure 26 First frames from both CRBN conformers with ligand extracted and shown without stripping the environment, containing water molecules and ions, Na^+ and Cl^-. Left: open CRBN; Right: closed CRBN; 52
Figure 27 Structural display of the closed and open CRBN showing all 26 human PDB entries with CRBN available. Left: whole CRBN protein structures (blue = all closed conformers; pink and orange = only two open conformers). Right: conformation of residue His353 from all 26 human CRBN structures shown, dividing into two major conformational groups (same colour coding as left figure). Open CRBN structures 6H0F and 6BNB are part of the bigger group of His353 conformations. Second group comprises closed CRBN structures 6BN7, 6BN8 and 6BN9. None of the latter have the His353 and surrounding residues from the sensor loop experimentally defined by electron density. 57
Figure 28 Visual extraction of an exemplary pose with the ligand EM12SO2F (pink) linked to His353 (green) showing a good geometry of the covalent bond after energy minimization.. 59
Figure 29 Overview of a selection aspect for which poses remain for interaction analysis of the covalently docked ligands, here two examples shown with small molecule EM12FS docked covalently to His353 for the closed (left; blue) and open (right; orange) CRBN structure. Both CRBN conformers displaying the so-called “core” residues in the IMiD binding pocket: Asn351, His378, Trp380, and Trp386. 60
Figure 30 Overview of „Good Poses“ per docking run in percentages. Good poses were defined by scanning all docked poses of covalently docked ligands with the Rarey Torsion Scan Tool. Torsion angles of all rotatable bonds were categorized into very strained to relaxed. Only those poses without any strained torsion angles remained after the selection. 60
Figure 31 Exemplary poses shown that were excluded from the docking runs due to geometric considerations of the poses. Left: Both ligands were also docked to other Histidine residues, such as the pose shown here for His357 docked ligand, in comparison to the co-crystallized ligand lenalidomide in the closed CRBN structure (5FQD). Middle and Right: Both ligands were also docked to the binding site in the open CRBN structure (6BNB). Strained small molecule geometry and torsion angles of the poses excluded the poses from further interaction analysis. Binding pockets (brown) based on lenalidomide (green) in the IMiD binding site visualized. 61
Figure 32 shows remaining number of docked poses in percentages per run. Bar graph output from Table 3. 62
Figure 33 Visual display of the binding pocket with lenalidomide (green) shown, water molecule HOH2037 (red) pointed out by arrow. 63

Figure 34 Interaction pattern of the two ligands covalently docked to both CRBN conformers. “Core” residues are shown individually, His378, Trp386, Trp380 and Ans351 (from left to right). Other residues are shown collectively, before further differentiating interaction analysis. Bars show relative interaction amounts in percentages from all docked poses remaining after the filtering. 64
Figure 35 Pie chart showing all „other“ residues interacting with covalently docked ligands in all remaining poses after filtering, next to core residues. Other residues were defined as shown: His353, Glu377, Ile37, Ile35, Lys18, Pro352 and Trp400. 64
Figure 36 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: NE2 covalently docked poses. Right: ND1 covalently docked poses. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also NE2 (white) and ND1 (black) docked results in percentages. Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37)..... 65
Figure 37 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: all covalently docked poses with the closed CRBN structure. Right: all covalently docked poses with the open CRBN structure. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also docked results in percentages with the closed CRBN. Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37). 66
Figure 38 Above: two pie charts showing the relative interaction of the two covalently docked ligands in the binding pocket (below). Left: all covalently docked poses with the ligand EM12SO2F. Right: all covalently docked poses with the ligand EM12FS. All covalent poses considered for the displayed results were linked to His353. Below: Binding pocket showing area of lenalidomide (grey) interacting with CRBN showing next to core residues (Asn351, His378, Trp380, Trp386), also docked results in percentages of ligand EM12FS (white) and EM12SO2F (black). Interactions of other residues from protein of interest were excluded from this visualization (included residues like Lys18, Ile35 and Ile37). 68
Figure 39 Above: RMSF plot showing the residue number (x-axis) against the Ångstrom difference (y-axis). Both trajectories of the open (orange) and closed (green) CRBN structure with the ligand lenalidomide showed similarities over most of the calculated flexibility of the residues. Areas of increased movement are displayed by the peak height, such as marked residues from 150 to 200 or 318 to the end of the protein chains, respectively. Below: Both closed (5FQD, green) and open (6BNB, orange) CRBN structure shown with respectively

coloured residue lengths corresponding to the demarcations in the plot above before, such as the DDB1 contact area (red) and CULT domain (blue). Lenalidomide is shown in the IMiD binding pocket (pink). 70
Figure 40 Plots showing the H-bond interactions (red) between protein and ligand residue atoms (y-axis) against frame (x-axis). Open (above) and closed (below) CRBN trajectories with ligands subjected to CCPTRAJ based analysis. 72
Figure 41 Pharmacophore interaction map generated by LigandScout MD Pharmacophore creation for the trajectories of the open (orange) and closed (green) CRBN structure with lenalidomide and stripped environment, without water molecules and ions. The map plotted residues interacting with the lenalidomide in the binding site and the corresponding pharmacophore feature. The relative interaction of the residues with the pharmacophore feature are shown in percentages by appearances during the entire trajectories. Appearances below 1% are shown as 0%. 74
Figure 42 Pharmacophore interaction map generated by LigandScout MD Pharmacophore creation for the trajectories of the open (orange) and closed (green) CRBN structure with lenalidomide and the environment, including the water molecules to be part of the calculation. The map plotted residues interacting with the lenalidomide in the binding site and the corresponding pharmacophore feature. The relative interaction of the residues with the pharmacophore feature are shown in percentages by appearances during the entire trajectories. Appearances below 1% are shown as 0%. 75
Figure 43 Pharmacophore Plots were generated by LigandScout after MD pharmacophores were created. Pharmacophore plots show sets of unique pharmacophore features with individual bars against the number of appearances in the entire trajectories. A unique pharmacophore feature is defined by one specific atom of the ligand interaction with a specific protein atom via a pharmacophore feature. By choosing a specific pharmacophore feature, bars in the interactive pharmacophore plot are shown in green. For the two plots at the top three pharmacophore features were chosen from frame 1 of both trajectories, interacting from protein atoms of the residues His378, Trp380 and Asn351. Above: open CRBN (orange); Middle: closed CRBN (green); Below: closed CRBN (green) including the water molecules: green bars are showing the sets of unique pharmacophore features where water molecule HOH2037 was observed to interact with lenalidomide by a pharmacophore feature; 76
Figure 44 The pharmacophore feature timeline shows pharmacophore feature presence of individual pharmacophore features after MD pharmacophores were created in LigandScout. For the trajectory of the open CRBN with lenalidomide several pharmacophores were detected and plotted as binned frames with a size of 5,000 frames per data point. H-bond donor (HBD) pharmacophores are shown in green, while H-bond acceptor (HBA)

pharmacophores are shown in red. Above: Pharmacophore feature timeline shown for all pharmacophores from the open CRBN with several appearing in low numbers during the entire trajectory, while one prominent HBA that was shown as feature interacting with Trp380 as the highest and constantly present pharmacophore. Below: Three pharmacophore features were selected from the first frame of the open CRBN trajectory. Two HBA features were observed to interact with Trp380 and Asn351, while one HBD feature interacting by His378 residue was found to be present in similar presence like the Asn351 pharmacophore feature. 78
Figure 45 The pharmacophore feature timeline shows pharmacophore feature presence of individual pharmacophore features after MD pharmacophores were created in LigandScout. For the trajectory of the closed CRBN with lenalidomide several pharmacophores were detected and plotted as binned frames with a size of 5,000 frames per data point. H-bond donor (HBD) pharmacophores are shown in green, while H-bond acceptor (HBA) pharmacophores are shown in red. Above: Pharmacophore feature timeline shown for all pharmacophores from the closed CRBN with most features appearing in low numbers during the entire trajectory, while one prominent HBA that was shown as feature interacting with Trp380 as the highest and constantly present pharmacophore. Below: Three pharmacophore features were selected from the first frame of the closed CRBN trajectory. Two HBA features were observed to interact with Trp380 and Asn351, while one HBD feature interacting by His378 residue was plotted increasingly and taking over by the end of the trajectory over the Asn351 pharmacophore feature. 79
Figure 46 GRAIL map was generated for the closed CRBN with lenalidomide within a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields. Left: positive ionizable (PI) to aromatic (AR) (violet); Right: AR to AR (dark blue); both were displayed at a contouring level of 0.9 with a score going to 1.0. Bulky feature presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields. 81
Figure 47 GRAIL map was generated for the closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: positive ionizable (PI) to negative ionizable (NI) (pink) and NI to PI (violet); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky feature presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields; GRAIL map of PI-NI around Glu377, while NI-PI field displayed around Lys401..... 82
Figure 48 GRAIL map was generated for the open CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD

binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: positive ionizable (PI) to negative ionizable (NI) (pink) and NI to PI (violet); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields; GRAIL map of PI-NI around His353 and Phe402, while NI-PI field displayed between Lys402 and Thr387..... 83
Figure 49 GRAIL map was generated for the open and closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: H-bond donor (HBD) to H-bond acceptor (HBA) (green); both were displayed at a contouring level of 0.8 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN GRAIL map of HBD-HBA around Tyr354, between Trp400 and Thr387 as well as between Trp400 and Phe402; while open CRBN HBD-HBA interaction fields displayed around His357, Trp400 as well as between Pro352 and Trp380. 84
Figure 50 GRAIL map was generated for the open and closed CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: Halogen-bond donor (XBD) to Halogen-bond acceptor (XBA) (light pink); both were displayed at a contouring level of 0.95 with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN GRAIL map of XBD-XBA between Trp400 and Phe402; while open CRBN XBD-XBA interaction fields displayed between Phe352 and Trp380. 85
Figure 51 GRAIL map was generated for the open (above) and closed (below) CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields: Hydrophobic (H) to H interaction (yellow); both were displayed at a contouring level of 0.90 (above) and 0.95 (below) with a score going to 1.0. Bulky presenting amino acid side chains are highlighted to improve the visualization of the GRAIL interaction fields: for the closed CRBN the GRAIL map for the H-H pharmacophore were displayed at a score of 0.95 in an improved way to retrieve better information for the ligand design development showing between residues Trp390 and Trp400..... 86
Figure 52 GRAIL map was generated for the open (left) and closed (right) CRBN with lenalidomide with a grid box covering the residues of the CULT domain which includes the residues forming the IMiD binding pocket. Iso-surfaces were visualized by pharmacophore interaction fields for a special GRAIL map: happy waters (happyH2O) (violet) and unhappy waters (unhappyH2O) (dark blue); both were displayed at a contouring level of 0.95 with a score going to 1.0. Bulky presenting amino acid side chains and relevant water molecules

(red) are highlighted to improve the visualization of the GRAIL interaction fields: for the open CRBN GRAIL map of the happy H2O was around between Trp380, Tyr349 and Asn351 as well as the water molecule WAT522; for closed CRBN both water iso-surfaces were displayed: unhappy H2O interaction fields were visualized around water molecules WAT467 to WAT8866, while happy H2O were visualized between Trp386, Glu377 and the water molecule WAT7720. 87

10. References

[1] D. Nandi, P. Tahiliani, A. Kumar, and D. Chandu, 'The ubiquitin-proteasome system', *J. Biosci.*, 2006.

[2] I. E. Wertz and X. Wang, 'From Discovery to Bedside: Targeting the Ubiquitin System', *Cell Chemical Biology*, vol. 26, no. 2, pp. 156–177, Feb. 2019, doi: 10.1016/j.chembiol.2018.10.022.

[3] D. P. Bondeson and C. M. Crews, 'Targeted Protein Degradation by Small Molecules', *Annu. Rev. Pharmacol. Toxicol.*, vol. 57, no. 1, pp. 107–123, Jan. 2017, doi: 10.1146/annurev-pharmtox-010715-103507.

[4] X. Zhang, S. Linder, and M. Bazzaro, 'Drug Development Targeting the Ubiquitin–Proteasome System (UPS) for the Treatment of Human Cancers', *Cancers*, vol. 12, no. 4, p. 902, Apr. 2020, doi: 10.3390/cancers12040902.

[5] L. Zhao, J. Zhao, K. Zhong, A. Tong, and D. Jia, 'Targeted protein degradation: mechanisms, strategies and application', *Sig Transduct Target Ther*, vol. 7, no. 1, p. 113, Apr. 2022, doi: 10.1038/s41392-022-00966-4.

[6] J. Park, J. Cho, and E. J. Song, 'Ubiquitin–proteasome system (UPS) as a target for anticancer treatment', *Arch. Pharm. Res.*, vol. 43, no. 11, pp. 1144–1161, Nov. 2020, doi: 10.1007/s12272-020-01281-8.

[7] N. Vargesson, 'Thalidomide-induced teratogenesis: History and mechanisms', *Birth Defect Res C*, vol. 105, no. 2, pp. 140–156, Jun. 2015, doi: 10.1002/bdrc.21096.

[8] A. Bricelj, C. Steinebach, R. Kuchta, M. Gütschow, and I. Sosič, 'E3 Ligase Ligands in Successful PROTACs: An Overview of Syntheses and Linker Attachment Points', *Front. Chem.*, vol. 9, p. 707317, Jul. 2021, doi: 10.3389/fchem.2021.707317.

[9] Q. Yang, J. Zhao, D. Chen, and Y. Wang, 'E3 ubiquitin ligases: styles, structures and functions', *Mol Biomed*, vol. 2, no. 1, p. 23, Dec. 2021, doi: 10.1186/s43556-021-00043-2.

[10] M. Békés, D. R. Langley, and C. M. Crews, 'PROTAC targeted protein degraders: the past is prologue', *Nat Rev Drug Discov*, vol. 21, no. 3, pp. 181–200, Mar. 2022, doi: 10.1038/s41573-021-00371-6.

[11] T. Ishida and A. Ciulli, 'E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones', *SLAS Discovery*, vol. 26, no. 4, pp. 484–

502, Apr. 2021, doi: 10.1177/2472555220965528.
[12] ‘MedChemExpress’.
[13] I. N. Michaelides and G. W. Collie, ‘E3 Ligases Meet Their Match: Fragment-Based Approaches to Discover New E3 Ligands and to Unravel E3 Biology’, *J. Med. Chem.*, vol. 66, no. 5, pp. 3173–3194, Mar. 2023, doi: 10.1021/acs.jmedchem.2c01882.
[14] E. R. Watson *et al.*, ‘Molecular glue CELMoD compounds are regulators of cereblon conformation’, *Science*, vol. 378, no. 6619, pp. 549–553, Nov. 2022, doi: 10.1126/science.add7574.
[15] M. D. Hartmann, I. Boichenko, M. Coles, A. N. Lupas, and B. Hernandez Alvarez, ‘Structural Dynamics of the Cereblon Ligand Binding Domain’, *PLoS ONE*, vol. 10, no. 5, p. e0128342, May 2015, doi: 10.1371/journal.pone.0128342.
[16] J. Tao, J. Yang, and G. Xu, ‘The interacting domains in cereblon differentially modulate the immunomodulatory drug-mediated ubiquitination and degradation of its binding partners’, *Biochemical and Biophysical Research Communications*, vol. 507, no. 1–4, pp. 443–449, Dec. 2018, doi: 10.1016/j.bbrc.2018.11.058.
[17] E. S. Fischer *et al.*, ‘Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide’, *Nature*, vol. 512, no. 7512, pp. 49–53, Aug. 2014, doi: 10.1038/nature13527.
[18] N. Bai *et al.*, ‘Modeling the CRL4A ligase complex to predict target protein ubiquitination induced by cereblon-recruiting PROTACs’, *Journal of Biological Chemistry*, vol. 298, no. 4, p. 101653, Apr. 2022, doi: 10.1016/j.jbc.2022.101653.
[19] Y. Fang *et al.*, ‘Targeted protein degrader development for cancer: advances, challenges, and opportunities’, *Trends in Pharmacological Sciences*, vol. 44, no. 5, pp. 303–317, May 2023, doi: 10.1016/j.tips.2023.03.003.
[20] R. I. Troup, C. Fallan, and M. G. J. Baud, ‘Current strategies for the design of PROTAC linkers: a critical review’, *Exploration of Targeted Anti-tumor Therapy*, vol. 1, no. 5, Oct. 2020, doi: 10.37349/etat.2020.00018.
[21] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, ‘Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settingsq’, *Advanced Drug Delivery Reviews*, 2001.
[22] U. Nayek, S. I. Basheer Ahamed, U. H. Mansoor Hussain, M. K. Unnikrishnan, and A. A. Abdul Salam, ‘Computational investigations of indanedione and indanone derivatives in drug discovery: Indanone derivatives inhibits cereblon, an E3 ubiquitin

ligase component', *Computational Biology and Chemistry*, vol. 101, p. 107776, Dec. 2022, doi: 10.1016/j.compbiolchem.2022.107776.
[23] V. Poongavanam *et al.*, 'Predictive Modeling of PROTAC Cell Permeability with Machine Learning', *ACS Omega*, vol. 8, no. 6, pp. 5901–5916, Feb. 2023, doi: 10.1021/acsomega.2c07717.
[24] C. Heim, A.-K. Spring, S. Kirchgäßner, D. Schwarzer, and M. D. Hartmann, 'Identification and structural basis of C-terminal cyclic imides as natural degrons for cereblon', *Biochemical and Biophysical Research Communications*, vol. 637, pp. 66–72, Dec. 2022, doi: 10.1016/j.bbrc.2022.11.001.
[25] M. Miñarro-Lleonar, A. Bertran-Mostazo, J. Duro, X. Barril, and J. Juárez-Jiménez, 'Lenalidomide Stabilizes Protein–Protein Complexes by Turning Labile Intermolecular H-Bonds into Robust Interactions', *J. Med. Chem.*, vol. 66, no. 9, pp. 6037–6046, May 2023, doi: 10.1021/acs.jmedchem.2c01692.
[26] A. A. Akuffo *et al.*, 'Ligand-mediated protein degradation reveals functional conservation among sequence variants of the CUL4-type E3 ligase substrate receptor cereblon', *Journal of Biological Chemistry*, vol. 293, no. 16, pp. 6187–6200, Apr. 2018, doi: 10.1074/jbc.M117.816868.
[27] S. A. Holstein, 'The Evolving Tale of Immunomodulatory Drugs and Cereblon', *Clin Pharmacol Ther*, vol. 96, no. 5, pp. 538–541, Nov. 2014, doi: 10.1038/clpt.2014.168.
[28] T. Mori *et al.*, 'Structural basis of thalidomide enantiomer binding to cereblon', *Sci Rep*, vol. 8, no. 1, p. 1294, Jan. 2018, doi: 10.1038/s41598-018-19202-7.
[29] H. K. Kim, J. E. Seol, S. W. Ahn, S. Jeon, C.-S. Park, and J. Han, 'Cereblon: promise and challenges for combating human diseases', *Pflugers Arch - Eur J Physiol*, vol. 473, no. 11, pp. 1695–1711, Nov. 2021, doi: 10.1007/s00424-021-02624-0.
[30] Y. Fang *et al.*, 'Targeted protein degrader development for cancer: advances, challenges, and opportunities', *Trends in Pharmacological Sciences*, vol. 44, no. 5, pp. 303–317, May 2023, doi: 10.1016/j.tips.2023.03.003.
[31] J. Che and L. H. Jones, 'Covalent drugs targeting histidine – an unexploited opportunity?', *RSC Med. Chem.*, vol. 13, no. 10, pp. 1121–1126, 2022, doi: 10.1039/D2MD00258B.
[32] N. R. Kong, H. Liu, J. Che, and L. H. Jones, 'Physicochemistry of Cereblon Modulating Drugs Determines Pharmacokinetics and Disposition', *ACS Med. Chem.*

Lett., vol. 12, no. 11, pp. 1861–1865, Nov. 2021, doi: 10.1021/acsmedchemlett.1c00475.

[33] J. T. Cruite *et al.*, 'Cereblon covalent modulation through structure-based design of histidine targeting chemical probes', *RSC Chem. Biol.*, vol. 3, no. 9, pp. 1105–1110, 2022, doi: 10.1039/D2CB00078D.

[34] R. P. Nowak *et al.*, 'Plasticity in binding confers selectivity in ligand-induced protein degradation', *Nat Chem Biol*, vol. 14, no. 7, pp. 706–714, Jul. 2018, doi: 10.1038/s41589-018-0055-y.

[35] M. Wieder *et al.*, 'Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations', *J. Chem. Inf. Model.*, vol. 57, no. 2, pp. 365–385, Feb. 2017, doi: 10.1021/acs.jcim.6b00674.

[36] A.-Q. K. Oyedele *et al.*, 'Docking covalent targets for drug discovery: stimulating the computer-aided drug design community of possible pitfalls and erroneous practices', *Mol Divers*, Sep. 2022, doi: 10.1007/s11030-022-10523-4.

[37] G. et al Jones, 'Development and Validation of a Genetic Algorithm for Flexible Docking'.

[38] J. Yan and Z. Zheng, 'Discovery of Highly Potent CRBN Ligands and Insight into Their Binding Mode through Molecular Docking and Molecular Dynamics Simulations', *ChemMedChem*, vol. 18, no. 5, Mar. 2023, doi: 10.1002/cmdc.202200573.

[39] A. Ahmed, F. Rippmann, G. Barnickel, and H. Gohlke, 'A Normal Mode-Based Geometric Simulation Approach for Exploring Biologically Relevant Conformational Transitions in Proteins', *J. Chem. Inf. Model.*, vol. 51, no. 7, pp. 1604–1622, Jul. 2011, doi: 10.1021/ci100461k.

[40] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, 'An overview of the Amber biomolecular simulation package: Amber biomolecular simulation package', *WIREs Comput Mol Sci*, vol. 3, no. 2, pp. 198–210, Mar. 2013, doi: 10.1002/wcms.1121.

[41] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham, 'Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms', *J. Chem. Theory Comput.*, vol. 3, no. 6, pp. 2312–2334, Nov. 2007, doi: 10.1021/ct700119m.

[42] namd@ks.uiuc.edu, 'Steered Molecular Dynamics'.

[43] D. A. Schuetz *et al.*, 'GRAIL: GRids of phArmacophore Interaction fieLds', *J. Chem. Theory Comput.*, vol. 14, no. 9, pp. 4958–4970, Sep. 2018, doi:

10.1021/acs.jctc.8b00495.
[44] ‘Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.’
[45] ‘Application_Name, Scientific Vector Language (SVL) source code provided by Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.’
[46] J. Che and L. H. Jones, ‘Covalent drugs targeting histidine – an unexploited opportunity?’, *RSC Med. Chem.*, vol. 13, no. 10, pp. 1121–1126, 2022, doi: 10.1039/D2MD00258B.
[47] P. Penner, W. Guba, R. Schmidt, A. Meyder, M. Stahl, and M. Rarey, ‘The Torsion Library: Semiautomated Improvement of Torsion Rules with SMARTScompare’, *J. Chem. Inf. Model.*, vol. 62, no. 7, pp. 1644–1653, Apr. 2022, doi: 10.1021/acs.jcim.2c00043.