UNIVERSITÀ DEGLI STUDI DI PALERMO
Dottorato in Scienze Molecolari e Biomolecolari
Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche
Settore Scientifico Disciplinare CHIM/08

COMPUTATIONAL METHODOLOGIES APPLIED TO PROTEIN-PROTEIN INTERACTIONS FOR MOLECULAR INSIGHTS IN MEDICINAL CHEMISTRY

IL DOTTORE
DOTT.SSA MARIA RITA GULOTTA

IL COORDINATORE E TUTOR
PROF.SSA PATRIZIA DIANA

CO-TUTOR
DOTT. UGO PERRICONE

CICLO XXXIII
ANNO 2021

To my family, current and to be,

and to my tenacity

that has brought me till here

ABSTRACT

In living systems, proteins usually team up into "molecular machinery" implementing several protein-to-protein physical contacts – or protein-protein interactions (PPIs) – to exert biological effects at both cellular and systems levels. Deregulations of protein-protein contacts have been associated with a huge number of diseases in a wide range of medical areas, such as oncology, cancer immunotherapy, infectious diseases, neurological disorders, heart failure, inflammation and oxidative stress.

PPIs are very complex and usually characterised by specific shape, size and complementarity. The protein interfaces are generally large, broad and shallow, and frequently protein-protein contacts are established between non-continuous epitopes, that conversely are dislocated across the protein interfaces. For this reason, in the past two decades, PPIs were thought to be "undruggable" targets by the scientific research community with scarce or no chance of success. However, in recent years the Medicinal Chemistry frontiers have been changing and PPIs have gained popularity amongst the research groups due to their key roles in such a huge number of diseases.

Until recently, PPIs were determined by experimental evidence through techniques specifically developed to target a small group of interactions. However, these methods present several limitations in terms of high costs and labour- and time-wasting. Nowadays, a large number of computational methods have been successfully applied to evaluate, validate, and deeply analyse the experimentally determined protein interactomes. In this context, a high number of computational tools and techniques have been developed, such as methods designed to construct interaction databases, quantum mechanics and molecular mechanics (QM/MM) to study the electronic properties, simulate chemical reactions, and calculate spectra, and all-atom molecular dynamics simulations to simulate temporal and spatial scales of inter- and intramolecular interactions. These techniques have allowed to explore PPI networks and predict the related functional features.

In this PhD work, an extensive use of computational techniques has been reported as valuable tool to explore protein-protein interfaces, identify their hot spot residues, select small molecules and design peptides with the aim of inhibiting six different studied PPIs. Indeed, in this thesis, a success story of *in silico* approaches to PPI study has been described, where MD simulations, docking and pharmacophore screenings led to the identification of a set of PPI modulators. Among these, two molecules, RIM430 and RIM442, registered good inhibitory activity with IC_{50} values even within the nanomolar range against the interaction between MUC1 and CIN85 proteins in cancer disease.

Furthermore, computational alanine scanning, all-atom molecular dynamics simulations, docking and pharmacophore screening were exploited to (1) rationally predict three potential interaction models of $NLRP3_{PYD}$-ASC_{PYD} complex involved in inflammatory and autoimmune diseases; (2) identify a potentially druggable region on the surface of SARS-CoV-2 Spike protein interface and select putative inhibitors of the interaction between Spike protein and the host ACE2 receptor against COVID-19 (CoronaVIrus Disease 2019); (3) investigate intramolecular modifications as a consequence of a point mutation on C3b protein (R102G) associated with the age-related macular degeneration (AMD) disease; (4) design non-standard peptides to inhibit

transcriptional events associated with HOX-PBX complex involved in cancer diseases; and (5) to optimise a patented peptide sequence by designing helix-shaped peptides embedded with the hydrogen bond surrogate approach to tackle cocaine abuse relapses associated with Ras-RasGRF1 interaction.

Although all the herein exploited techniques are based on predictive calculations and need experimental evidence to confirm the findings, the results and molecular insights retrieved and collected show the potential of the computer-aided drug design applied to the Medicinal Chemistry, guaranteeing labour- and time-saving to the research groups. On the other hand, computing ability, improved algorithms and fast-growing data sets are rapidly fostering advances in multiscale molecular modelling, providing a powerful emerging paradigm for drug discovery. It means that more and more research efforts will be done to invest in novel and more precise computational techniques and fine-tune the currently employed methodologies.

PREFACE

The research activities conducted during the PhD three-year period were focused on the analysis and study of six different protein-protein interactions (PPIs) involved in cancer, inflammatory and immune system diseases and addiction to substances of abuse. The research work was funded by Ri.MED Foundation, therefore most of the research activities were conducted at the Molecular Informatics Unit led by Dr Ugo Perricone, exploiting the computing power present at Ri.MED Foundation. Then, between the end of the second year and the beginning of the third year of the PhD course, the research activities were followed up at Cardiff University in the Molecular Modelling laboratory under the supervision of Professor Andrea Brancale.

The first part of the work was focused on the study and deepening of protein-protein interactions, their physicochemical characteristics and the computational techniques applied to date for the exploration of protein-protein interfaces and the identification of potential modulators. Thus, this study made it possible to prepare in 2018 a review article entitled "An overview of recent Molecular Dynamics applications as medicinal chemistry tool for undruggable sites challenge" published in the peer-reviewed journal *Medicinal Chemistry Communications*.

Subsequently, in May and June 2018 I took part in the Erasmus Plus for Traineeship programme through an internship at the laboratories of the company BioAscent Discovery Ltd. located in New House (Glasgow), in the United Kingdom. The activities carried out at this institution have allowed me to acquire skills related to the automated management of large libraries of molecules, in detail about their storage, their dissolution and the preparation of aliquots and plates for biological assays. In fact, the computational techniques for identifying potentially active molecules, such as docking and pharmacophore approach, usually allow to "filter" large libraries of molecules to identify a few hundred compounds responding to precise parameters. These selected compounds should then be acquired or synthesized and stored in specific storage systems, and can undergo transformations, such as dissolutions or dilutions for carrying out biological/biophysical tests. The skills acquired during the traineeship provided me with a comprehensive view of the procedure for acquiring, managing and processing physical libraries of compounds.

The first period of the PhD work was focused on research activities conducted at Ri.MED Foundation under the supervision of Dr Ugo Perricone and focused on the study of two protein-protein interactions, described below:

- Interaction between Mucin 1 (MUC1) and Cbl-interacting protein of 85 kDa (CIN85), which has been associated with invasiveness of tumour cells and with the development of metastasis in the colorectal tract. In pathological conditions, MUC1 protein is in a hypoglycosylated state, which allows it to interact with CIN85 forming a complex capable of migrating and invading new tissues. The computational study made it possible to identify two small molecules that have been shown to inhibit the interaction between these two proteins registering biological activity in the nanomolar range.
- Interaction between the Pyrin domain (PYD) of the protein NACHT, LRR and PYD domains-containing 3 (NLRP3) and the Pyrin domain of the protein Adapter apoptosis-associated speck-like protein containing a CARD (ASC), which has been

associated with numerous chronic inflammatory and autoimmune diseases, such as ulcerative colitis, Crohn's disease, psoriasis, multiple sclerosis, or rheumatic diseases and arthropathies, such as systemic lupus erythematosus, etc. Computational techniques made it possible to create three potential $NLRP3_{PYD}$-ASC_{PYD} interaction models, which have been used for the identification of potentially active small molecules, that will be tested in the laboratories of Ri.MED Foundation.

In June 2019, I continued my research activities at the Molecular Modelling laboratory at the University of Cardiff, in the United Kingdom, under the guidance of Professor Andrea Brancale. During the months I spent in Cardiff, I worked on the study of three projects on the following protein-protein interactions:

- Interaction between the C3b protein and factor H (FH), which has been associated with age-related macular degeneration (AMD). In particular, recent studies have shown that the patients affected by this disease present a mutated form of C3b protein ($C3b_{R102G}$). This mutation causes a reduction in the affinity of FH protein for C3b, thus prolonging the activity of C3b and therefore the activation of the inflammatory system associated with the complement. The computational techniques applied to this study made it possible to identify a peptide that should be able to discriminate the wild-type form of C3b protein from the mutant, in order to develop a diagnostic kit able to identify the mutation $C3b_{R102G}$ using plasma samples from patients.
- Interaction between Homeobox proteins (HOX) and Pre-B-cell Leukaemia Homeobox (PBX) proteins, which has been associated with transcriptional events related to the development of numerous forms of cancer in various human tissues. Computational techniques made it possible to design a peptide motif binding the PBX protein and, based on this, to identify peptides composed of non-natural amino acids, *i.e.* residues that have a different side chain than the amino acids present in nature, with the aim of inhibiting the cooperative binding between the two proteins. The peptides designed will be synthesized and tested in the laboratories of Cardiff University.
- Interaction between the Rat sarcoma protein (Ras) and the protein Ras guanine nucleotide-releasing factor 1 (RasGRF1), which has been associated with relapsing behaviours in patients dependent on substances of abuse, such as cocaine. The computational methods employed have allowed to design and identify some peptides through the employment of the 3_{10}-hydrogen bond surrogate approach. This latter consists in the introduction of a C-C bond between the amino acid i and the amino acid $i+3$ to simulate the hydrogen bond between the backbone of the two involved amino acids. In this way, the peptides are forced to assume an α-helix conformation. The designed peptides will be synthesized and tested at Cardiff University laboratories and should block the interaction between Ras and RasGRF1 proteins to improve behavioural alterations caused by drug addiction.

Finally, after the research period carried out in Cardiff, I worked on another project concerning the Coronavirus, that has caused the pandemic currently underway. Specifically, the research activities relating to this project were focused on the protein-protein interaction described below:

- Interaction between the Spike glycoprotein of the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the receptor Angiotensin-converting enzyme 2 (ACE2) of the host organism, which has been identified as the event triggering the

virus entry into the host cell. In particular, computational methods have made it possible to identify the N-terminal portion of the interaction interface of Spike as the region potentially most suitable for the design of putative inhibitors. Therefore, a group of small molecules was selected and will be tested to verify their potential ability to compete with ACE2 for binding the Spike protein N-terminal region.

This study allowed to prepare a research article entitled "Targeting SARS-CoV-2 RBD interface: a supervised computational data-driven approach to identify potential modulators" published in the peer-reviewed journal *ChemMedChem*.

ACKNOWLEDGEMENTS

I would like to thank all the people that have given their contribution to my professional and personal growth during this PhD three-year path.

First of all, I would like to thank my co-tutor Dr Ugo Perricone that has been more than a guide for me. He introduced me to the computer-aided drug design field teaching me the knowledge basis for carrying out my research activities and pushing me to explore in depth "twists and turns" of this sector. He has been able to transfer me his passion for this research field and make me passionate as well. He has led me to discover my professional inclinations and interests. Finally, he has been also a friend when tiredness and difficulties began to be felt, bringing me back on the right track.

I would like to thank my tutor Professor Patrizia Diana that with her kindness, her wisdom and her smile has always had the right words for addressing my professional growth path and helping me to find the right way.

I would like to thank my colleagues and friends Jessica, who shared with me this PhD path, Giada and Nedra, that made my days happier gladdening the work with our laughs but even always supporting me with discussions and suggestions. A special thanks to Giada for being more than a colleague to me.

I would like to thank the CEO of Ri.MED Foundation, Dr Alessandro Padova for giving me the possibility to spend most of the time of my PhD three-year period at Ri.MED laboratories and for making me feel part of this institution. I would like to thank also all the colleagues at Ri.MED Foundation, Dr Mari De Rosa for her wise advices and tips and her prompt openness towards helping out me when necessary, Dr Claudia Coronnello for her discreet presence and availability and all the Ri.MED administrative and research staff for making me feel welcomed.

I would like to thank Dr Sylviane Boucharens for her generous and maternal hospitality in Scotland, hosting me at BioAscent Discovery Ltd and providing me with the possibility to enrich my professional background. I would like to thank also the BioAscent personnel, my Scottish friends, Stuart for his exceptional kindness and congeniality all the mornings together from the train to the lab, together with Steven and Gina, for teaching me all the "tricks of the trade", Karen and Guthrie for their niceness and Paul for his brilliant humour. I will never forget the human warmth of these Scottish friends.

I would like particularly to thank Professor Andrea Brancale for his infinite hospitality in Cardiff University, for his kindness and niceness and for being always prompt and available to provide me suggestions, driving my research activities and findings, and I would like to thank also all the other members of Professor Brancale group, Salvo, Marcella, Giulio and Carmine for providing me food for thought during my research efforts.

I would like to thank my companions of adventure in Wales, Fabiana, Luce, Beatrice and Silvia, because probably without them my stay in Cardiff would not have been so amazing, and all the Italian colleagues and friends, Elisa, Alessandra, Silvia, Fabio, Alberto and Marika for making me feel home in Wales.

I would like to thank my parents and my brother Antonio for their inexhaustible support, because without their precious presence maybe I would not be here submitting this PhD thesis.

I would like to thank my husband-to-be Francesco, who has shared with me this PhD path and supported me every time I needed advice or soothing words, believing in me and reawakening the determination in going forward when some hurdles made me tottering.

Last but not least, I would like to thank my best friends Giorgia and Giulia, that have always been my rocks and during this three-year period have often listened to my stories advising me with wise suggestions and, above all, I thank them because they have always believed in me.

Finally, I think I have been very lucky to meet these people, new or always been, during this three-year professional and life path, because all of them have made me stronger and definitely a better person.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Aβ – Amyloid β

ACE2 – Angiotensin-Converting Enzyme 2

AIM2 – Absent In Melanoma 2

AIR – Ambiguous Interaction Restraint

AMD – Age-related Macular Degeneration

AP – Alternative Pathway

ASC – Adapter apoptosis-associated Speck-like protein containing a CARD

CAPRI – Critical Assessment of PRedicted Interactions

CARD - Caspase Activation and Recruitment Domain

CCP – Complement-Control-Protein

CIN85 – Cbl-interacting protein of 85 kDa

CLD – Collectrin-Like Domain

COVID-19 - COronaVIrus Disease 2019

CPP – Conditioned Place Preference

Cryo-EM – Cryo-Electron Microscopy

CUB – C1r/C1s, UEGF, BMP1

CV – Collective Variable

DA – Dopamine

DAMPs – Damage-Associated Molecular Patterns

DD – Death Domain

DED – Death Effector Domain

ERK – Extracellular signal-Regulated Kinase

Exd – Extradenticle

FH – Factor H

FI – Factor I

GEF - Guanine nucleotide Exchange Factor

HADDOCK – High Ambiguity Driven DOCKing

HBS – Hydrogen Bond Surrogate

HD – Homeodomain

HOX – Homeobox protein

HSC – Hematopoietic Stem Cell

HX – Hexapeptide

IEG – Immediate Early Gene

IL – Interleukin

LNK – Linker

LPS – LipoPolySaccharide

LRR - Leucine-Rich Repeat

MAC – Membrane Attack Complex

MD – Molecular Dynamics

MEK – Mitogen-activated protein kinase

MG – MacroGlobulin

MM/GBSA – Molecular Mechanics/Generalized Born Surface Area

MM/PBSA – Molecular Mechanics/Poisson-Boltzmann Surface Area

MUC1 – Mucin 1

MW – Molecular Weight

NACHT – Nucleotide-binding oligomerisation domain containing a CARD

NES – Nuclear Export Sequence

NLRC4 – NLR family, CARD domain-containing 4

NLRP3 – NACHT, LRR and PYD domains-containing 3

NLS - Nuclear Localization Signal

NMR – Nuclear Magnetic Resonance

NOD – Nucleotide-binding Oligomerization Domain

PAINS – Pan-Assay Interference compoundS

PAMPs – Pathogen-Associated Molecular Patterns

PBX – Pre-B cell leukaemia transcription factor

PD – Peptidase Domain

PDB – Protein Data Bank

PPI – Protein-Protein Interaction

PPII - PolyProline II

PPIM – Protein-Protein Interaction Modulator

PRR - Pattern Recognition Receptor

PYD – Pyrin Domain

QM/MM – Quantum Mechanics and Molecular Mechanics

Raf – Rat fibrosarcoma protein

Ras – Rat sarcoma protein

RasGRF1 – Ras Guanine nucleotide-Releasing Factor 1

RBD - Receptor-Binding Domain

RBM - Receptor-Binding Motif

RCA - Regulator of Complement Activity

Rem – Ras exchanger motif

REMD – Replica-Exchange Molecular Dynamics

REOS – Rapid Elimination Of Swill

RMSD – Root-Mean-Square Deviation

RMSF – Root-Mean-Square Fluctuation

ROS – Reactive oxygen species

SARS-CoV – Severe Acute Respiratory Syndrome CoronaVirus

SARS-CoV-2 – Severe Acute Respiratory Syndrome CoronaVirus 2

SCRP - Short Consensus RePeat

SH3 – Src Homology 3

Sos – Son of sevenless protein

TA-MUC1 – Tumour-Associated MUC1

TALE – Three-Amino acid Loop Extension

TED - ThioEster-containing Domain

VNTR – Variable Number Tandem Repeat

WHO – World Health Organisation

CONTENTS

CHAPTER ONE

PROTEIN-PROTEIN INTERACTIONS – Computational approaches to study protein-protein interfaces in Medicinal Chemistry

1.1 Introduction

In the last decades, research efforts in cell biology, molecular biology, biochemistry, structural biology, and biophysics have built a significant knowledge about functions and molecular properties of individual proteins. This knowledge has been made available by consulting the major protein databases like UniProt [1, 2]. However, proteins rarely carry out their biological functions alone. On the contrary, they usually team up into "molecular machinery" implementing several physicochemical interactions to exert biological effects at both cellular and systems levels. In this context, it should be crucial to investigate and unravel the complex molecular relationships in living systems to get a complete map of protein-to-protein physical contacts – or protein-protein interactions (PPIs) – occurred in a living organism. This map has been also referred to as interactome [3, 4]. It has been reported that the human interactome consists of about 650,000 PPIs [5, 6], compared to only about 20,000 protein-coding genes [7], and any deregulation of these interactions leads to a disease state. This fact has increased the interest in the PPIs representing a wide source of novel targets for the development of new therapeutics, leading to important breakthroughs in understanding biological pathways, host-pathogen interactions and cancer growth and spreading [8–18]. Indeed, targeting protein-protein interactions is getting particularly attractive due to more and more available data especially about protein-protein complexes and entire signalling pathways.

PPIs have been shown to be significant therapeutic targets for a wide range of medical areas, such as oncology [19–23], cancer immunotherapy [24], tropical infectious diseases [25], neurological disorders [26], heart failure [27], inflammation and oxidative stress [7]. In the past two decades, PPIs were thought to be "undruggable" targets by the scientific research community with scarce or no chance of success [28]. However, recent efficient large-scale technologies on genomics and proteomics programmes, such as the high-throughput experimental technologies, allowed to identify and measure broad networks of protein interactions between protein pairs by reaching a comprehensive knowledge of the protein-protein interactome. Thus, in recent years a collection of large- and small-scale efficient technologies have notably identified and increased the number of reported protein-protein interactions [29], by building public repositories of PPIs. A compendium of these PPI databases can be found in "Pathguide" [30].

Recent successes in the inhibition of PPIs with small molecules have emerged from both academic and private research by applying several new strategies to modulate the activity of proteins and identify new drugs against this tremendous reservoir of potential targets [31–46]. Targeting PPIs usually means specifically interfere with dimer and oligomer formation or disrupting antibody-antigen interaction [47] or specifically targeting protein-protein interactions to alter a signalling pathway within a cell process.

As the binding region of PPIs is significantly different as compared with protein-traditional drug interactions, different approaches and techniques have been developed. For example, gene-editing methods allowed to perform point mutations within the genomes of mammalian cells [48], making possible the validation process of individual PPIs as putative drug targets with unprecedented precision. Finally, the protein-protein interfaces have been shown to be less conserved among species than traditional active sites, therefore PPI inhibitors are also commonly thought to have a greater opportunity for being selective [49].

1.1.1 Protein-protein interfaces characteristics and composition

Protein-protein interactions are very complex and usually characterised by specific shape, size and complementarity [50]. The protein interfaces are generally large, broad and shallow, and frequently their contacts are established between non-continuous epitopes, that conversely are dislocated across the protein interfaces.

Their recognition sites can exhibit standard sizes 1200-2000 $Å^2$ [51], while a few smaller interfaces normally present sizes ranging 750-1200 $Å^2$, and they usually make short-term and low-stability complexes [52, 53]. On the contrary, large interfaces range sizes 2000-4660 $Å^2$ and they essentially occur between G-proteins and other components of the signal transduction system and between proteases and a particular class of inhibitor protein partners [51, 54].

Furthermore, it is often assumed that the energy of protein-protein binding is directly related to the buried hydrophobic surface area, when the protein heterodimer interfaces sizes exceed 600 $Å^2$ [55–59]. This size cut-off has been considered the minimum area required to make an almost water-free environment around a critical set of energetically favourable interactions [56]. In this context, a particular aspect of interest is that, during the protein folding process, hydrophobic residues shield themselves from the solvent, yielding a hydrophobic protein interior (also called *core*) and a hydrophilic surface (termed *rim*) [52, 60, 61]. Amino acids at the protein-protein interface *core* are more hydrophobic than those at the *rim* [13, 52, 61–63] and are more frequently identified as hot spots [56]. This physicochemical diversity between interior amino acids and surface amino acids is highly connected to protein stability and evolution. Indeed, it was demonstrated that buried amino acids at the *core* are more conserved than those at the *rim* [64–67], while residues at the surface of proteins show a major inclination in mutating [68–72]. This fact is probably connected to a higher protein destabilisation when interior amino acids mutate resulting in hard damages to the structure and function of the protein [73] or the cell [74].

As above mentioned, protein-protein interface amino acids are frequently hydrophobic and bury a large extent of non-polar surface area, hence it has been assumed that the hydrophobicity is a leading force in PPIs [75–77]. The non-polar regions originated by the presence of hydrophobic residues establish van der Waals contacts, resulting in the expulsion of water molecules in the interface, and causing an increase in entropy that favours complex formation [78] and results in a gain in free energy [79]. Indeed, the free energy gain generated by all the individual van der Waals contacts together with the energy gain produced by the desolvation process represents a

considerable free energy increase, that may provide a higher stabilization of the protein-protein complex [50].

In fact, these contacts provide tight packing between protein residues organized as patches including protrusions from the surface. The number of these patches may vary from 1 to 15 together with their sizes that are between 200 and 400 Å [80].

Another driving force of protein-protein interactions is represented by the electrostatic contacts [55, 81–85], whereas their importance is to the electrostatic complementarity of interacting protein surfaces [54, 55, 86–88], that promotes complex formation [89, 90] and defines the lifetime of the complexes [91]. Furthermore, the average number of hydrogen bonds is proportional to the width of the protein surfaces [92, 93], that is one hydrogen-bond is usually found each 100–200 Å of surface. In general, it was found that 76% of protein-protein hydrogen bonds are established by the side chains of amino acids, while the other hydrogen bonds are generated between the surrounding water molecules and the protein contact surfaces [94–96].

A systematic analysis of the key amino acids at the interface, also termed hot spots, unveiled a non-random composition on protein interfaces [97]. The most conserved and frequent amino acids at protein-protein interfaces are usually tryptophan (21%), arginine (13.3%), and tyrosine (12.3%) [56], whereas the average percentage of aromatic residues as hot spots clearly demonstrate their importance to protein interactions.

Leucine, serine, threonine, and valine residues are less present or overall absent as hot spots even if in some cases they can be important for some protein-protein complexes [56]. In particular, tryptophan has been shown to play a unique function due to its aromatic nature [98], that can contribute with aromatic π-interactions and hydrophobic contacts. Moreover, tryptophan can also donate hydrogen bond and can shield fragile hydrogen bonds from water with its hydrophobic nature [99]. Finally, tryptophan mutation to an alanine generates a large cavity, due to the significant difference in sizes [56] by triggering a highly complex destabilisation.

Arginine can establish multiple types of favourable interactions arranging up to five hydrogen bonds and a salt bridge thanks to its positive charge on the guanidinium motif [56].

As above mentioned, tyrosine was the third more conserved amino acids among protein-protein complexes database. It shows a hydrophobic surface, and both aromatic π-interactions and the hydrogen bonding ability of its 4-hydroxyl group [56]. On the contrary, phenylalanine, the most similar amino acid to tyrosine, has three times lower probability of being a hot spot, probably because it cannot participate in hydrogen bonds as for tyrosine due to the lack of the hydroxyl group [56].

The analysis of several complexes has highlighted that aspartate and asparagine are more frequent at protein interfaces compared to glutamate and glutamine. This fact can be explained because aspartate and asparagine present less conformational freedom due to the shorter side chains. This can result even presumably in differences in side-chain conformational entropy [100]. Moreover, although leucine and isoleucine are isomers with essentially identical chemistry, the first amino acid is present only for 0.83%, while the isoleucine reported a frequency of 9.62% as a hot spot [56].

Complementarity is a significant characteristic of protein-protein interfaces essentially defined based on the size of the buried surface, the alignment of polar and non-polar residues, the number of buried waters, and the packing densities of atoms

involved in the PPI [101]. As already mentioned, most protein-protein interactions are characterised by optimal tight-fitting regions [102] with complementary pockets distributed across the binding interface and including structurally conserved residues [103]. Indeed, residues across the protein-protein interface often coevolve [104, 105], and they create complemented pockets rich in conserved residues [102]. These pockets are defined as complementary [106] because it exists a strong complementarity both in shape and in the closeness of hydrophobic and hydrophilic hot spots. Moreover, charged residues can establish salt bridges while hydrophobic residues belonging to a protein surface fit into small recesses on the opposite face [107]. Usually, this complementarity may provide druggable sites for the identification of modulators [102] based on the hot spots of one face that pack against the hot spot of the other face establishing a region determinant for complex binding [108, 109].

The number of these hot spots is tightly correlated with the interface size [102], and their local distribution and packing are crucial factors determining the PPI stabilisation [57].

It has been demonstrated that on average 79% of the hot spot amino acids are located on complemented pockets [57], and 93% of residues that upon alanine scanning mutagenesis report a free energy difference of binding ($\Delta\Delta G_{binding}$) higher than 4 kcal/mol are usually found as protruding or complemented pocket residues [57]. Complemented pockets usually show a few polar and ionizable amino acids (*e.g.* arginine, lysine, glutamate, and aspartate) compared to other surface pockets [102]. This fact contributes to decreasing the desolvation barrier necessary for protruding residues to contact into the complemented pocket. However, at the bottom of these pockets, several polar and ionisable residues have been found to increase binding stability, with the effect of enhancing polar-polar interactions in a hydrophobic environment [102]. Among the most frequent residues located in complemented pockets, tryptophan is often the most abundant and may act by shielding the complemented pocket from the solvent [110]. On the other hand, glycine is a conserved amino acid when located within a complemented pocket rather than if found in another region of the protein interface. As glycine is a residue that lacks a side-chain, hence it usually generates tight packing with aromatic, polar, and small hydrophobic amino acids in the interacting chain and establishing some backbone H-bonds across the interface [111].

1.1.2 Drug design of protein-protein inhibitors

Unlike classical targets, such as membrane receptors or enzymes, that include a well-defined binding site, the complexity of PPIs impacts also on designing modulators and their chemical and physicochemical features [112].

Often for protein-protein interactions, there are no natural ligands or known active compounds to be exploited as a reference to guide the drug design process, requiring an unconventional drug discovery strategy to be applied for the identification of hit compounds [55, 86, 87, 113, 114].

Successful PPI modulators reported in literature often show molecular weights two or three times higher than traditional small molecules [53, 115]. Indeed, the protein interacting surfaces are usually shallow and are widely exposed to the solvent molecules, and usually hits show low affinities for protein-protein interactions, reporting K_D values of 0.1-5 mM [115–117].

Generally, PPI modulators (PPIMs) are classified as stabilisers or disruptors according to their mechanism of action [112, 118]. PPI stabilisers provide an increase in protein-protein complex binding affinity and stability by directly binding the interaction interface (orthosteric stabilisation) or binding to a remote site of the protein and causing an increase of protein-protein affinity (allosteric stabilisation) [112, 118]. On the other hand, the PPI disruptors may compete in binding one of the two protein partners at the binding region (orthosteric disruption) or destabilise a PPI through an interaction with a distal or proximal site on the protein surface (allosteric disruption), generating a decrease in protein-protein affinity (figure 1.1).

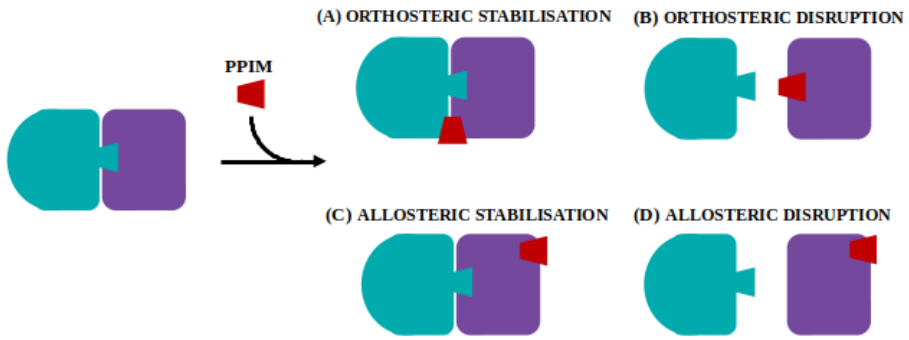


Figure 1.1. PPI modulators mechanism of action, where green and purple structures represent two different proteins interacting with each other. On the left, two interacting proteins are illustrated. When PPIM interfere with these proteins four scenarios can occur: **(A)** an orthosteric stabilization, where the PPIM binds both proteins in a different region than the protein-protein interface stabilising the complex; **(B)** an orthosteric disruption, where the PPIM is positioned within the interfacial binding region; **(C)** an allosteric stabilisation, where the PPIM binds one of the two protein partners at a distal site of the protein surface generating a stabilisation of the protein-protein interaction; and **(D)** an allosteric disruption, where the PPIM binds a region of the two proteins at a distal site of the protein surface triggering conformational changes within this protein and causing protein-protein disruption [112].

However, several troubles have been found in developing small molecule antagonists that target protein-protein interactions. First, a database of starting structures is not available to draw from and build novel potential modulators. Indeed, small ligand databases designed for traditional targets are usually inappropriate to target PPIs through for example virtual high-throughput screening. A principal component analysis (including topological surface area, logP, and MW) reported by Pagliaro *et al*. [119] on three commercial databases for PPIs (Maybridge, Asinex and Chemical Diversity Database International Diversity Collection) showed that only 50% by the diversity space was covered. It means that the current chemical libraries exhibit scarce diversity to cover mostly the PPI drugs chemical space. However, nowadays a significant number of validated cheminformatic and machine learning methods has been applied to fill in this gap. Indeed, general observations concerning protein-protein interaction chemical features of PPI modulators reporting nanomolar K_D values highlighted that the molecular weights are usually at least 650 KDa and therefore the PPIMs will not, in general, observe Lipinksi's rule of 5 [53].

As already mentioned, protein-protein interfaces are flat making it more difficult to design small molecules able to accommodate a binding pocket. However, usually some specific residues involved in contacts, termed hot spots, can be identified as crucial for stabilising the complex. Indeed, the hot spots contribute to most of the free energy of

binding and can be targeted for drug design [120]. In this context, fragment-based ligand design is often very successful for designing drugs targeting PPIs, because it allows enlarging the chemical diversity of the designed compounds to be screened [121].

Although the above-described issues in PPIMs design, in the past decades, an increasing number of success stories have appeared. Several small molecules have been reported to target more than 40 different PPIs, corresponding to various topological spaces – primarily helix-based domains, β-strand domains, mixed folding (helix + - β-strand) and loop-binding groove domains [122]. Some of these drugs have been demonstrated to be very effective as PPI disruptors, such as p53-MDM2 inhibitors [123], IL2-IL2R inhibitors [124], AMA1-RON2 inhibitors [125], and a few of these (*e.g.* ABT-263) have reached pre-clinical or clinical trials [126–131].

The PPI inhibitors are usually classified into the following three categories: antibodies, peptides and peptidomimetics. These compounds can be designed based on certain protein recognition motifs and they include specific molecular scaffolds that have been shown to exhibit biological or pharmacological activity when incorporated into drug design [132]. PPI inhibitors exploiting secondary structures as scaffolds, such as α-helix [133–135], β-sheet [136], or β-turns [137, 138] are also referred as "proteomimetics" [139, 140]. Other examples of PPI modulators are the extended structures and proline-rich segments [141] that also reproduce molecular motifs, but also organic molecules such as benzodiazepines that can introduce recognition sites for proteins [142].

Although peptides present some disadvantages such as metabolic instability, poor oral bioavailability and scarce ability to cross membranes, most of them showed high selectivity and potency [143]. However, in physiological conditions, the peptides consisting of less than 15 amino acids generally are expected to be essentially unstable due to a low nucleation probability according to the helix-coil transition theory [144, 145]. They show certain flexibility and can take a huge amount of different conformations [146–148]. Therefore, several synthetic strategies have been developed to create peptides with stable folded structure [149], *e.g.* the hydrogen bond surrogate (HBS) approach [150] (figure 1.2). This strategy is expected to overwhelm the intrinsic nucleation propensities of the amino acids by providing upstream a preorganization of the residues, that causes the helix formation initialization [150–153].

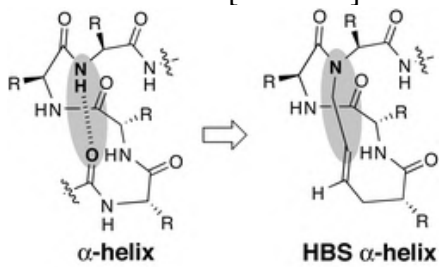


Figure 1.2. Hydrogen bond surrogate approach is based on the generation of an artificial α-helix where the C=O · · · H-N hydrogen bond between the *i*th and the (*i* + *4*)th residues is mimicked by a covalent bond of the type C=X–Y–N, where X and Y are usually carbon atoms [150, 153].

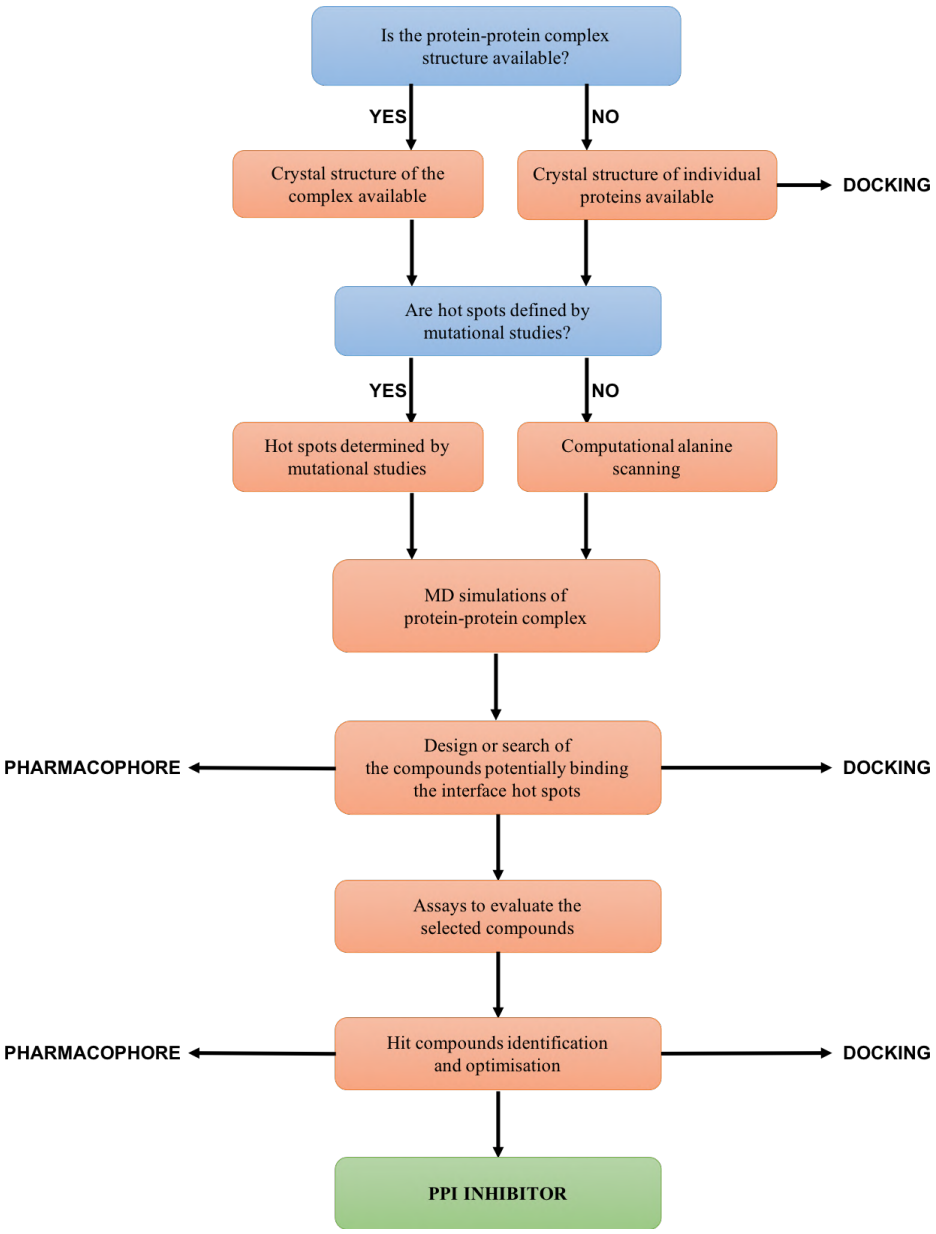
1.2 Molecular modelling in protein-protein interactions

Nowadays, the PDB structures are used as a benchmark for unveiling specific key host-guest interactions, structural protein insights, protein conformational changes and so

on. However, in some cases, structural complexes are unavailable often because of the difficulty in experimentally solving them through NMR or X-ray crystallography methodologies. Therefore many research efforts have been focused on developing methods to predict protein-protein structural insights and molecular properties [121].

In this context, CAPRI [154], the critical assessment of predicted interactions, was created in 2001 as a forum for evaluating methods for computational protein-protein docking and protein complex interaction prediction. This forum is biannually updated with computational data from international research groups that are invited to test their developed computational methods and predict the structures of protein complexes, that will be experimentally solved and made public later than year.

The main steps of a conventional drug discovery process for protein-protein interactions are summarised in scheme 1.1 [155].



Scheme 1.1. A schematic PPI inhibitor design procedure involving computational techniques (computational alanine scanning, docking and pharmacophore approaches) across the different stages of the PPI drug design [132].

The first step to undertake a drug discovery programme on PPIs is the experimental evidence of interaction between the two proteins of interest [155]. Indeed, before performing extensive docking and pharmacophore calculations, hot-spot identification studies and design of PPI, first experimental methods should highlight and confirm an interaction between the two proteins of interest or identify the partner of a known protein of interest. When the structure of the protein-protein complex is available and the hot spots are known from mutational studies, docking or pharmacophore screening can be the first step of the work to design drug-like molecules. On the contrary, if the PPI structure is not available but the structures of the protein monomers of interest have been experimentally solved alone and mutational data are known, the protein complex can be simulated by performing protein-protein docking. The next sections describe in detail the steps of a drug discovery programme on PPIs, starting from the protein-protein complex generation through protein-protein docking and contacts investigation by performing computational alanine scanning and molecular dynamics simulations, and continuing with docking and pharmacophore screening to identify putative PPI inhibitors.

1.2.1 Protein-protein complex prediction through docking

There are two types of protein-protein docking: (1) the template-based docking, where the structures of individual proteins are docked using as a reference a template structure of a known dimer of proteins belonging to the same homologous family. This method is usually fast because no automatic docking and scoring algorithms are employed. The generated protein-protein complex is then minimised to get the docked conformation; (2) The template-free docking, where the protein complex is generated using docking algorithms with or without the support of experimental data [132].

In general, when template-free docking is performed without the support of experimental data, the complex structures is usually generated by fixing in the space a protein and rotating and translating the second one around the first. Each new protein-protein configuration is scored based on the energy of interaction calculated according to terms such as surface complementarities, electrostatic interactions, van der Waals interactions, and other terms depending on the method applied. These calculations are very time-consuming and the disadvantage is that it is very unlikely to find every possible rotation and translation for two interacting monomers. On the other hand, when experimental mutational data are available a docking method that incorporates this information can be employed to dock the two protein structures. An example of application performing this docking method is HADDOCK (High Ambiguity Driven DOCKing) [156–158], where the user should provide information obtained from biochemical and chemical shift perturbation data from NMR titration, as well as mutagenesis experiments [157]. On the basis of the input data about active and passive residues, ambiguous interaction restraints (AIRs) are introduced during the docking to provide the most likely orientation of the two proteins. The resulting structures are ranked according to their intermolecular energy, that is calculated based on the three different docking stages. The first stage is characterised by randomization of orientations and rigid-body energy minimization, that generates several thousand of models including rotational

and translational optimisation. For this stage, the "Rigid score" is calculated as in equation (1),

$$\textit{Rigid Score} = 0.01\ E_{AIR} + 0.01\ E_{vdW} + 1.0\ E_{elect} + 1.0\ E_{desolv} - 0.01\ BSA \qquad (1)$$

where E_{AIR} is the ambiguous interaction restraint energy, E_{vdW} is the van der Waals energy, E_{elect} is the electrostatic energy, E_{desolv} is the desolvation energy and BSA is the buried surface area.

The second stage of HADDOCK docking protocol is a semi-flexible simulated annealing, that introduces flexibility to the protein partners through three-step molecular dynamics-based refinement. For this stage, the "Flexible score" is defined as in equation (2).

$$\textit{Flexible Score} = 0.1\ E_{AIR} + 1.0\ E_{vdW} + 1.0\ E_{elect} + 1.0\ E_{desolv} - 0.01\ BSA \qquad (2)$$

Finally, the third stage is the refinement in explicit solvent (TIP3P model [159]) performing short molecular dynamics simulation to refine contacts. For this stage, the "Water score" is calculated as in equation (3).

$$\textit{Water Score} = 0.1\ E_{AIR} + 1.0\ E_{vdW} + 0.2\ E_{elect} + 1.0\ E_{desolv} \qquad (3)$$

A specific type of protein-protein interactions can occur when a small fragment of a protein of interest establishes contacts with another protein. In this case, it is referred to as protein-peptide interaction. It is very frequent that the peptide of interest exhibits a specific secondary structure. Examples of protein-peptide interactions in literature are the SH3 domains, the WW domains and the PDZ domains [160–164].

Most of the protein-protein docking software is often not suitable for performing protein-peptide docking because they do not incorporate the relevant flexibility of side chains of both partner molecules. At the same time, software used for docking of small molecules show limitations in the number of rotatable bonds for flexibility [132]. In this context, London *et al*. [161] developed a protein-peptide docking method that applies a coarse model of interaction and Monte-Carlo simulations to refine the complex using energy minimization. The resulting protein-peptide complex includes refinement of both protein and peptides backbone and side chains in their bound state.

Once the protein-protein or protein-peptide complexes (experimentally solved or docked) are available, the next step is the analysis of the hot spot residues of both protein partners.

1.2.2 Computational alanine scanning for predicting hot spots

One of the most valuable procedure for detecting crucial amino acids within protein-protein interfaces is the alanine scanning mutagenesis. This methodology allows to identify key residues and analyse a wide range of protein-protein interfaces [52, 56, 120, 165–167]. Although this technique is very costly and time-consuming, alanine scanning mutagenesis is definitely suitable for mapping functional epitopes, by introducing alanine substitutions in place of other amino acids in order to remove side-

chain atoms from the β-carbon without introducing additional conformational freedom [168–171].

Alanine is the amino acid of choice to perform computational residue scanning, because it shows a propensity to form α-helices, but can also occur in β-sheets. It is also generally equivalent to simply truncating a side chain back to the β-carbon, which is the first side chain atom. The β-carbon position depends upon the backbone dihedral angles (φ and ψ) of the polypeptide, so it is really part of the main chain structure of the protein. Thus, alanine is generally an accepted single residue of first choice for mutational scanning, because it retains the β-carbon but no other side chain chemistry.

On the other hand, glycine is not suitable for the substitutions because it lacks side chain, thus it is unusually flexible and can take on polypeptide backbone conformations generally not allowed by other amino acids [172]. Therefore, mutations to glycine may cause flexibility and possible conformational changes convoluted with the effects of removing the side chain atoms making experimental data interpretation more complex than for alanine. Moreover, replacing side chains with larger, more constrained (such as branched β-carbon side chains of valine and isoleucine), more polar, differently charged, or more hydrophobic atoms may all cause changes in structures and conformation along with the side chain chemistry, thereby further complicating the analysis of results.

In light of the above, alanine substitutions are chosen to get reliable measurements of the energetic contributions of individual side chains to protein binding at specific positions on protein interfaces. The development of this technique allowed to unveil a highly uneven distribution of energetic contributions of individual residues across each interface. Only a little number of residues (the hot spots) can be considered crucial by significantly contributing to the binding free energy of protein-protein complexes [52, 55, 173–178]. Hot spots have been defined as those amino acids that upon alanine mutation generate a binding free energy difference ($\Delta\Delta G_{binding}$) higher than 2.0 kcal/mol [120], while residues with $\Delta\Delta G_{binding} < 2$ kcal/mol are defined neutral [179]. Binding free energy values higher than 4 kcal/mol have been associated with a strong impact on protein binding affinity. However, these values are quite unusual and the most accepted and reliable threshold for mutation results is over 2 kcal/mol. Thorn and Bogan [120] analysed interfacial residues of a protein database and interestingly found that an average of 9.5% of these residues were hot spots.

Nowadays, when alanine scanning mutagenesis results are not available or they required too much time, the energy contributions of individual amino acids at a binding interface can also be theoretically predicted *in silico*, by performing a computational alanine scanning with molecular mechanics energy calculations combined with Poisson–Boltzmann [121] or generalized Born [180] and surface area continuum solvation (MM/PBSA and MM/GBSA) methods. This technique has been shown to be reasonably accurate, by reporting several success rates in literature [181].

This fully atomistic computational methodological approach consists of a computational Molecular Dynamics simulation protocol performed in a continuum medium using the Poisson-Boltzmann or generalized Born model. This method allows to perform a systematic scanning mutagenesis of protein-protein interfaces and it is able of affordably predicting the experimental results of mutagenesis.

The mutant complexes are usually generated by performing a single truncation of the mutated side chain, replacing Cα with a hydrogen atom and setting the Cα-H direction to that of the former Cα-Cβ. The $\Delta\Delta G_{binding}$ is calculated as the difference between the mutant ($\Delta G_{binding\_mut}$) and wild type ($\Delta G_{binding\_wt}$) complexes free energy defined as follows [182] in equation (4).

$$\Delta\Delta G_{binding} = \Delta G_{binding\_mut} - \Delta G_{binding\_wt} \quad (4)$$

Typical contributions to the free energy ($G_{complex}$) are based on the internal energy (bond, dihedral, and angle) (E_{int}), the electrostatic (E_{elect}) and the van der Waals (E_{vdW}) interactions, the free energy of polar solvation (G_{polar_solv}), the free energy of non-polar solvation ($G_{non\text{-}polar\_solv}$), and the entropic contribution (S) according to the following equation (5).

$$G_{complex} = E_{int} + E_{elect} + E_{vdW} + G_{polar\_solv} + G_{non\text{-}polar\_solv} - TS \quad (5)$$

1.2.3 Molecular dynamics applications in protein-protein interactions

Protein-protein interactions are crucial actors in most biological processes, therefore detecting specific amino acid residues can contribute to investigating the specificity and the strength of protein interactions. For this reason, an atomic exploration of the protein-protein systems may provide a better understanding of the driving forces for PPIs, by elucidating also the molecular recognition processes between the protein partners [9, 47, 183–186, 52, 55, 56, 75, 86, 87, 102, 107]. Furthermore, protein structures may change when performing their biological functions or undertake transitions from unfolded to folded state and vice versa [187].

Classical all-atom molecular modelling is very useful for exploring local motions by playing a crucial role for generating realistic molecular representations of biological protein functions, exploration of experimental data and other tasks involved into a drug discovery process [188–194]. On the other hand, atomistic molecular dynamics (MD) simulations can generally simulate folding processes of small and relatively quick folding proteins [195, 196] or the potential dimerization process [197]. Therefore, in the last ten years, the application of coarse-grained modelling of proteins [198, 199] has been increased [200] as an important technique to analyse large biomolecular systems [201]. The coarse-grained modelling has been successfully used in investigating protein folding mechanisms and protein structure prediction [187]. Coarse-grained models are computationally time-saving compared to all-atom simulations and may reproduce much longer time-scales and/or larger sizes of the systems under study (figure 1.3).

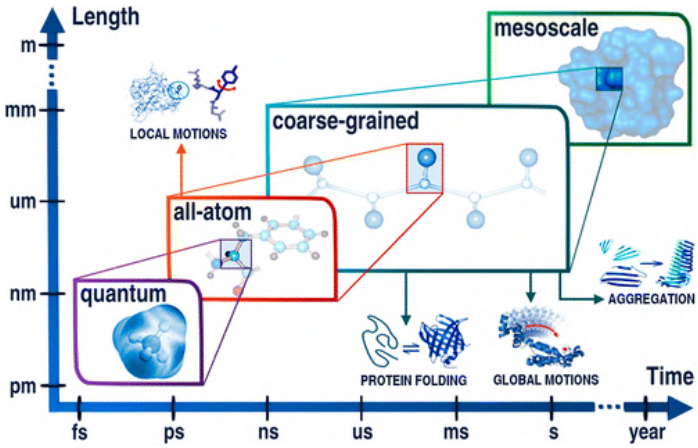


Figure 1.3. The molecular modelling techniques applied at different resolutions, quantum, all-atom, coarse-grained, and mesoscale, can range several time scales and system sizes [187].

The coarse-grained protein modelling can generate several levels of reduced models of polypeptide chain [202–204]. However, to achieve an atomistic precision level to characterise protein-protein interactions, in synergy with other computational and experimental techniques, all-atom MD simulations represent a valuable method [205], especially when some interacting hot spots are not visible from the crystal structure and conversely, motions can unveil them. Hence, interactions unreported before can also be detected [206]. MD simulations can also be useful to identify transient pockets that play a role in protein-protein binding but could not be captured by crystallography [207, 208]. In fact, PPIs are also characterised by a dynamic nature, therefore transient pockets and buried binding hot spots can emerge on the protein surfaces and guide the design of small molecules to these transient areas, as demonstrated by several studies reported in the literature. The conformational changes on protein-protein interactions usually involve motions of side chains and small loop perturbations [206, 209]. Thus, MD can allow exploring binding interfaces in terms of flexibility, and previously undetected important interactions can be identified. Indeed, protein-protein interfaces are flexible, show adaptability, and can change from the unbound to bound state. It can be noticed in some proteins that upon protein-protein binding a partner can undertake conformational changes exhibiting novel cavities usually not visible in the unbound state [206].

However, in some cases, all-atom molecular dynamics simulations employing classical mechanics have been demonstrated to present limitations in suitably exploring all the energy landscapes of biological molecules with many local minima frequently separated by high-energy barriers [210–212]. Indeed, in long simulations, several free energy minima can trap proteins in non-relevant conformations, impeding going forward to relevant conformations, leading to poor dynamic characterization of protein structures [210, 213]. For this purpose, enhanced sampling techniques can be employed, such as replica-exchange molecular dynamics (REMD), metadynamics and simulated annealing [212].

Moreover, in some cases, the predictive potential of MD simulation could be limited by the accuracy of the tuned force fields applied to define inter- and intramolecular interactions. To date, the atomistic physics-based force fields, such as OPLS3e [225], DES-Amber [226], CHARMM36m [227] and others, have reached high levels of accuracy especially for the simulation of biological macromolecules comparing the predicted data with the experimental ones (often from single-chain systems). In the case of proteins, the force field parameters have been widely refined with improved water

models and torsion-angle potentials ensuring highly accurate simulations of both folded and disordered protein systems [226, 228, 229].

Furthermore, MD simulations have reported several success cases in supporting the design of novel PPI inhibitors by complementing experimental screening techniques [113]. In this context, MD simulations can predict the binding mode of potential PPI modulators, whereas static structures conversely can provide only a few structural insights [207]. Thus, MD can be exploited in the validation process of high-affinity PPI binders, as multiple host-guest conformations are sampled during the trajectory.

Moreover, MD simulations can be applied to refine structures of low-quality resolution protein-protein complexes solved via experimental techniques, such as X-ray crystallography and cryo-electron microscopy (cryo-EM) [214]. Indeed, the generation of high-quality resolution structures of multiprotein complexes via these classical experimental methods still reports several challenges. Therefore, computational all-atom explorations such as MD simulations can help to elucidate structural and conformational dynamics of complex macromolecular structures relevant for biological processes [215].

During the MD simulations, the Newton equations of motions for all atoms of the system under study are integrated numerically, whereas the forces applied on every atom i (f_i) are defined by calculating the gradient of the potential energy function, $V(r^N)$, as reported in equation (6) [113].

$$f_i = m_i a_i = -\frac{\partial V(r^N)}{\partial r_i} \tag{6}$$

where m_i and a_i are respectively the mass and the acceleration of each atom i. The potential energy function $V(r^N)$ is parametrised through the selected force field and it depends on the positions of all N particles (r^N) in the system. The force field is applied to model bonded interactions, such as bonds, angles and torsion angles, and nonbonded interactions, such as van der Waals and electrostatic contacts. Therefore, during the trajectories, each atom can move depending on the force f_i exerted on it and equation 6 is integrated repeatedly for a predefined number of timesteps.

In this equation, each generated conformation of the proteins is associated with a potential energy and the forces f_i applied on each atom of the system are directly proportional to the gradient of the potential energy function. The forces are computed by using MD applications, such as the most common Desmond [216], NAMD [217] and Gromacs [218], including a very small integration time step (Δt) in the order of femtoseconds (10^{-15} second) in the attempt of reproducing biologically relevant timescales [113].

Moreover, when setting MD parameters, the solvent, *i.e.* the surrounding water molecules for physiological environments, can be modelled explicitly (*e.g.* TIP3P [159]) or included via continuum models [219]. The resulting MD trajectory reproduces the dynamics of protein conformations in a short-time simulation. However, nowadays the powerful and specialised hardware [220, 221] can massively compute longer MD simulations ranging from microseconds to milliseconds to deeply investigate conformational motions and identify novel potential molecular recognition sites between proteins [222–224].

1.2.4 Docking and pharmacophore screening

Once the protein-protein or protein-peptide complexes (experimentally solved or docked) are available and the hot spots have been identified, virtual screening campaigns can be run. Among several computational methods applied in a drug discovery process for PPIs, docking is a virtual screening method most used to identify potential modulators and for the lead optimisation. It has been extensively used and plays a key role in several stages during the design of PPI inhibitors [132]. Molecular docking works predicting ligand orientation and potential conformations within the binding pocket or on the protein surface. The overall docking process is based on two steps: a conformational search for the ligands within the defined grid box on a protein surface to generate several possible conformations; and a scoring algorithm that allows ranking the different generated conformations bound to the receptor, or "poses", on the basis of an energy function (from lowest to highest energy values) [230–233].

Docking screening presents three main issues related to the protein-protein surfaces characteristics that are not found in a traditional binding pocket. The first issue is the flexibility of proteins in conformational search and scoring, especially for transient PPIs or in presence of conformational changes in the binding site from bound to unbound states. The second issue is represented by water bridges that small molecules or peptides could establish in the bound state. Indeed, in most of the PPI hot spots, water molecules arrange around the hot spots leaving them dry. Hence ligands interacting with these hot spots can exhibit water bridges that usually are not included in docking procedure [18]. Even if some docking applications can incorporate water molecules, most of them are used for the active sites of enzymes, where the interactions with water molecules are usually well known or there are a few water molecules within the pocket. Finally, the third issue is related to the absence of protein-protein complex PDB structures in some cases and neither experimental data that can guide the complex generation [132].

The designed molecules can be docked to the protein surface defining the grid box on the hot spots, in order to test the ability of these compounds to interact with the key amino acids. For this purpose, there are several docking applications that can be employed, such as the most used AutoDock [234], Glide [235], MOE [236], GOLD [237] and others [232, 238–240].

In parallel with the docking technique, one of the employed computational methods for virtual screening campaign is the pharmacophore approach. For PPI projects the pharmacophore model can be built based on the chemical functionalities of one of the two protein partners [241]. Pharmacophore modelling derived from protein-protein interfaces have proven to perform efficient virtual screening for PPI drug discovery [242]. Employing a classical approach of pharmacophore modelling and virtual screening, several success stories have been reported on identifying inhibitors for PPIs [243–246]. For this purpose, many pharmacophore applications can be used, such as LigandScout [245], Phase [247, 248] and others [249].

A pharmacophore model can be defined as the 3D representation of the chemical features that are necessary for molecular recognition of a ligand by a protein and to obtain a desired biological effect, such as inhibition. Therefore, different molecules that contain functional groups in the appropriate 3D conformation and report high pharmacophore

scores can be considered promising modulators. The pharmacophore model generally consists of sphere-like features corresponding to the molecular functionalities such as hydrophobic, aromatic, hydrogen bond acceptor or donor, positive or negative ionisable groups [250]. This computational technique can be exploited into different molecular modelling strategies. For example, the generated models can be used as queries to identify molecules from a virtual compound database, but also they can address the docking of the molecules within the receptor by retaining the conformations extracted from the pharmacophore screening [251]. Finally, the model can also be used to compare pharmacophore and docking results and guide in selecting the consensus molecules as the most promising [241].

Thus, selecting and testing the compounds retrieved from virtual screening campaigns, hit molecules can be found by reporting experimental evidence that they may act as PPI modulators. Hence, once again docking and pharmacophore screening can be exploited to perform hit optimisation protocols. Indeed, the results from both techniques (docking and pharmacophore approach) are predictions that even need experimental validations. For this reason, several methods, such as surface plasmon resonance (SPR) [164], and NMR techniques, are available to provide information about respectively kinetics of binding and structure-activity relationship (SAR) of small molecules to proteins [252, 253].

Other techniques used for PPI modulators design are the proximity ligation assays [254] and enzyme fragment complementation assays [255], that respectively analyse protein dimerization and the related inhibition implemented by small molecules; FRET [256], and mass spectrometry [257–262]. However, these experimental techniques are no subjects of this work, hence they will not be discussed in this thesis.

1.3 Structure of the thesis

This PhD thesis is overall focused on computational approaches to the exploration of protein-protein interactions. Therefore, in this chapter, an overview of the protein-protein interactions has been provided by describing protein-binding interfaces characteristics and composition and the most frequently used computational techniques within a PPI drug discovery programme. Next chapters present molecular modelling strategies to address the exploration and drug design approaches on six different PPI projects, as below listed:

- *CHAPTER 2* – deals with the design strategy of small molecule inhibitors of the interaction between MUC1 (Mucin 1) and CIN85 (Cbl-interacting protein of 85 kDa) to prevent the invasiveness of cancer cells together with metastatic effects [263–265];
- *CHAPTER 3* – reports the study of the interaction between the pyrin domain (PYD) of NLRP3 protein (NACHT, LRR and PYD domains-containing 3) and the PYD of ASC protein (adapter apoptosis-associated speck-like protein containing a CARD) involved into inflammatory diseases, in order to reproduce putative models of the $NLRP3_{PYD}$-ASC_{PYD} complex [266, 267];
- *CHAPTER 4* – describes computational studies to identify a potential interacting region on the Spike protein surface of the novel coronavirus SARS-CoV-2 to

select putative inhibitors of the binding to human ACE2 protein (Angiotensin-converting enzyme 2), in order to prevent the viral entry [268, 269];

- *CHAPTER 5* – is focused on the structural analysis of the interaction between the mutated C3b protein ($C3b_{R102G}$) and its protein partner Factor H in patients affected by Age-Macular Degeneration disease [270];
- *CHAPTER 6* – reports a computational approach to modify the structure of a patented peptide including non-standard amino acids with the aim of inhibiting the trimeric complex formation between HOX (homeobox protein), PBX (Pre-B cell leukaemia transcription factor) and DNA to prevent cancer diseases [271, 272];
- *CHAPTER 7* – describes a structural optimisation strategy of a patented peptide by employing a 3_{10}-helix hydrogen bond surrogate approach to tackle the interaction between Ras (*rat sarcoma* protein) and RasGRF1 (Ras guanine nucleotide-releasing factor 1) proteins that have been associated with cocaine addiction and abuse [273, 274].

The above-listed PPI projects aim to provide comprehensive methodologies to address computational strategies for structural insights in PPIs and the design of protein-protein inhibitors such as small molecules or peptides. The work performed on Spike protein of SARS-CoV-2 has already been published in *ChemMedChem* Journal (DOI: 10.1002/cmdc.202000259), while the other works have been considering for research articles that in a short time will be submitted to scientific peer-reviewed journals.

References – Chapter One

1. Jain E, Bairoch A, Duvaud S, et al (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics 10:136. https://doi.org/10.1186/1471-2105-10-136
2. (2010) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38:D142–D148. https://doi.org/10.1093/nar/gkp846
3. Cusick ME, Klitgord N, Vidal M, Hill DE (2005) Interactome: gateway into systems biology. Hum Mol Genet 14:R171–R181. https://doi.org/10.1093/hmg/ddi335
4. De Las Rivas J, Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol 6:e1000807. https://doi.org/10.1371/journal.pcbi.1000807
5. Venkatesan K, Rual J-F, Vazquez A, et al (2009) An empirical framework for binary interactome mapping. Nat Methods 6:83–90. https://doi.org/10.1038/nmeth.1280
6. Stumpf MPH, Thorne T, de Silva E, et al (2008) Estimating the size of the human interactome. Proc Natl Acad Sci 105:6959–6964. https://doi.org/10.1073/pnas.0708078105
7. Lu M-C, Tan S-J, Ji J-A, et al (2016) Polar Recognition Group Study of Keap1-Nrf2 Protein–Protein Interaction Inhibitors. ACS Med Chem Lett 7:835–840. https://doi.org/10.1021/acsmedchemlett.5b00407
8. Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol 7:188–197. https://doi.org/10.1038/nrm1859
9. Chothia C, Janin J (1975) Principles of protein–protein recognition. Nature 256:705–708. https://doi.org/10.1038/256705a0
10. Lievens S, Eyckerman S, Lemmens I, Tavernier J (2010) Large-scale protein interactome mapping: strategies and opportunities. Expert Rev Proteomics 7:679–690. https://doi.org/10.1586/epr.10.30
11. Deremble C, Lavery R (2005) Macromolecular recognition. Curr Opin Struct Biol 15:171–175. https://doi.org/10.1016/j.sbi.2005.01.018
12. Han J-H, Batey S, Nickson AA, et al (2007) The folding and evolution of multidomain proteins. Nat Rev Mol Cell Biol 8:319–330. https://doi.org/10.1038/nrm2144
13. Janin J, Bahadur RP, Chakrabarti P (2008) Protein–protein interaction and quaternary structure. Q Rev Biophys 41:133–180. https://doi.org/10.1017/S0033583508004708
14. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of Protein−Protein Interactions: What are the Preferred Ways For Proteins To Interact? Chem Rev 108:1225–1244. https://doi.org/10.1021/cr040409x
15. Levy ED, Pereira-Leal JB (2008) Evolution and dynamics of protein interactions and networks. Curr Opin Struct Biol 18:349–357. https://doi.org/10.1016/j.sbi.2008.03.003
16. Nooren IMA (2003) Diversity of protein-protein interactions. EMBO J 22:3486–3492. https://doi.org/10.1093/emboj/cdg359
17. Ponstingl H, Thomas Kabir, Gorse D, Thornton JM (2005) Morphological aspects of oligomeric protein structures. Prog Biophys Mol Biol 89:9–35. https://doi.org/10.1016/j.pbiomolbio.2004.07.010
18. Reichmann D, Rahat O, Cohen M, et al (2007) The molecular architecture of protein–protein binding sites. Curr Opin Struct Biol 17:67–76. https://doi.org/10.1016/j.sbi.2007.01.004
19. Arrowsmith CH, Bountra C, Fish P V., et al (2012) Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov 11:384–400. https://doi.org/10.1038/nrd3674
20. Nero TL, Morton CJ, Holien JK, et al (2014) Oncogenic protein interfaces: small molecules, big challenges. Nat Rev Cancer 14:248–262. https://doi.org/10.1038/nrc3690
21. Arkin MR, Whitty A (2009) The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. Curr Opin Chem Biol 13:284–290. https://doi.org/10.1016/j.cbpa.2009.05.125
22. Arkin MR, Tang Y, Wells JA (2014) Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. Chem Biol 21:1102–1114. https://doi.org/10.1016/j.chembiol.2014.09.001
23. Li Z, Ivanov AA, Su R, et al (2017) The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. Nat Commun 8:14356. https://doi.org/10.1038/ncomms14356
24. Li Q, Quan L, Lyu J, et al (2016) Discovery of peptide inhibitors targeting human programmed

death 1 (PD-1) receptor. Oncotarget 7:64967–64976. https://doi.org/10.18632/oncotarget.11274
25. Dawidowski M, Emmanouilidis L, Kalel VC, et al (2017) Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. Science (80-) 355:1416–1420. https://doi.org/10.1126/science.aal1807
26. Hayes MP, Soto-Velasquez M, Fowler CA, et al (2018) Identification of FDA-Approved Small Molecules Capable of Disrupting the Calmodulin–Adenylyl Cyclase 8 Interaction through Direct Binding to Calmodulin. ACS Chem Neurosci 9:346–357. https://doi.org/10.1021/acschemneuro.7b00349
27. Anand P, Brown JD, Lin CY, et al (2013) BET Bromodomains Mediate Transcriptional Pause Release in Heart Failure. Cell 154:569–582. https://doi.org/10.1016/j.cell.2013.07.013
28. Ran X, Gestwicki JE (2018) Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. Curr Opin Chem Biol 44:75–86. https://doi.org/10.1016/j.cbpa.2018.06.004
29. Blow N (2009) Untangling the protein web. Nature 460:415–417. https://doi.org/10.1038/460415a
30. Pathguide. http://www.pathguide.org/. Accessed 21 Sep 2020
31. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. Nature 450:1001–1009. https://doi.org/10.1038/nature06526
32. Fry D (2008) Drug-Like Inhibitors of Protein-Protein Interactions: A Structural Examination of Effective Protein Mimicry. Curr Protein Pept Sci 9:240–247. https://doi.org/10.2174/138920308784533989
33. Popowicz GM, Dömling A, Holak TA (2011) The Structure-Based Design of Mdm2/Mdmx–p53 Inhibitors Gets Serious. Angew Chemie Int Ed 50:2680–2688. https://doi.org/10.1002/anie.201003863
34. Rinaldi M, Tintori C, Franchi L, et al (2011) A Versatile and Practical Synthesis toward the Development of Novel HIV-1 Integrase Inhibitors. ChemMedChem 6:343–352. https://doi.org/10.1002/cmdc.201000510
35. Voet A, Callewaert L, Ulens T, et al (2011) Structure based discovery of small molecule suppressors targeting bacterial lysozyme inhibitors. Biochem Biophys Res Commun 405:527–532. https://doi.org/10.1016/j.bbrc.2011.01.053
36. Fotouhi N, Graves B (2005) Small Molecule Inhibitors of p53/MDM2 Interaction. Curr Top Med Chem 5:159–165. https://doi.org/10.2174/1568026053507705
37. Wang S (2010) Design of Small-Molecule Smac Mimetics as IAP Antagonists. pp 89–113
38. White PW, Faucher A-M, Goudreau N (2010) Small Molecule Inhibitors of the Human Papillomavirus E1-E2 Interaction. pp 61–88
39. Lessene G, Czabotar PE, Colman PM (2008) BCL-2 family antagonists for cancer therapy. Nat Rev Drug Discov 7:989–1000. https://doi.org/10.1038/nrd2658
40. Azmi AS, Mohammad RM (2009) Non-peptidic small molecule inhibitors against Bcl-2 for cancer therapy. J Cell Physiol 218:13–21. https://doi.org/10.1002/jcp.21567
41. Cossu F, Mastrangelo E, Milani M, et al (2009) Designing Smac-mimetics as antagonists of XIAP, cIAP1, and cIAP2. Biochem Biophys Res Commun 378:162–167. https://doi.org/10.1016/j.bbrc.2008.10.139
42. Pinacho Crisóstomo FR, Feng Y, Zhu X, et al (2009) Design and synthesis of a simplified inhibitor for XIAP-BIR3 domain. Bioorg Med Chem Lett 19:6413–6418. https://doi.org/10.1016/j.bmcl.2009.09.058
43. Christ F, Voet A, Marchand A, et al (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 6:442–448. https://doi.org/10.1038/nchembio.370
44. De Luca L, Ferro S, Gitto R, et al (2010) Small molecules targeting the interaction between HIV-1 integrase and LEDGF/p75 cofactor. Bioorg Med Chem 18:7515–7521. https://doi.org/10.1016/j.bmc.2010.08.051
45. Azmi AS, Wang Z, Philip PA, et al (2011) Emerging Bcl-2 inhibitors for the treatment of cancer. Expert Opin Emerg Drugs 16:59–70. https://doi.org/10.1517/14728214.2010.515210
46. De Luca L, De Grazia S, Ferro S, et al (2011) HIV-1 integrase strand-transfer inhibitors: Design, synthesis and molecular modeling investigation. Eur J Med Chem 46:756–764. https://doi.org/10.1016/j.ejmech.2010.12.012
47. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. Nat Rev Drug Discov 3:301–317.

https://doi.org/10.1038/nrd1343
48. Komor AC, Kim YB, Packer MS, et al (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533:420–424. https://doi.org/10.1038/nature17946
49. Cesa LC, Mapp AK, Gestwicki JE (2015) Direct and Propagated Effects of Small Molecules on Protein–Protein Interaction Networks. Front Bioeng Biotechnol 3:. https://doi.org/10.3389/fbioe.2015.00119
50. Fernandez A, Scheraga HA (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc Natl Acad Sci 100:113–118. https://doi.org/10.1073/pnas.0136888100
51. Horton N, Lewis M (1992) Calculation of the free energy of association for protein complexes. Protein Sci 1:169–181. https://doi.org/10.1002/pro.5560010117
52. Conte L Lo, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285:2177–2198. https://doi.org/10.1006/jmbi.1998.2439
53. Sperandio O, Reynès CH, Camproux A-C, Villoutreix BO (2010) Rationalizing the chemical space of protein–protein interaction inhibitors. Drug Discov Today 15:220–229. https://doi.org/10.1016/j.drudis.2009.11.007
54. Zhao S, Morris GM, Olson AJ, Goodsell DS (2001) Recognition templates for predicting adenylate-binding sites in proteins. J Mol Biol 314:1245–1255. https://doi.org/10.1006/jmbi.2000.5201
55. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci 93:13–20. https://doi.org/10.1073/pnas.93.1.13
56. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9. https://doi.org/10.1006/jmbi.1998.1843
57. Keskin O, Ma B, Nussinov R (2005) Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. J Mol Biol 345:1281–1294. https://doi.org/10.1016/j.jmb.2004.10.077
58. Novotny J, Bruccoleri RE, Saul FA (1989) On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. Biochemistry 28:4735–4749. https://doi.org/10.1021/bi00437a034
59. Privalov PL (1979) Stability of Proteins Small Globular Proteins. pp 167–241
60. Gething M-J, Sambrook J (1992) Protein folding in the cell. Nature 355:33–45. https://doi.org/10.1038/355033a0
61. Bahadur RP, Chakrabarti P, Rodier F, Janin J (2003) Dissecting subunit interfaces in homodimeric proteins. Proteins Struct Funct Genet 53:708–719. https://doi.org/10.1002/prot.10461
62. Janin J, Rodier F, Chakrabarti P, Bahadur RP (2007) Macromolecular recognition in the Protein Data Bank. Acta Crystallogr Sect D Biol Crystallogr 63:1–8. https://doi.org/10.1107/S090744490603575X
63. Chakrabarti P, Janin J (2002) Dissecting protein-protein recognition sites. Proteins Struct Funct Genet 47:334–343. https://doi.org/10.1002/prot.10085
64. Eames M, Kortemme T (2007) Structural Mapping of Protein Interactions Reveals Differences in Evolutionary Pressures Correlated to mRNA Level and Protein Abundance. Structure 15:1442–1451. https://doi.org/10.1016/j.str.2007.09.010
65. Mintseris J, Weng Z (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci 102:10930–10935. https://doi.org/10.1073/pnas.0502667102
66. Guharoy M, Chakrabarti P (2005) Conservation and relative importance of residues across protein-protein interfaces. Proc Natl Acad Sci 102:15447–15452. https://doi.org/10.1073/pnas.0505425102
67. Caffrey DR (2004) Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci 13:190–202. https://doi.org/10.1110/ps.03323604
68. Franzosa EA, Xia Y (2009) Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. Mol Biol Evol 26:2387–2395. https://doi.org/10.1093/molbev/msp146
69. Conant GC, Stadler PF (2009) Solvent Exposure Imparts Similar Selective Pressures across a Range of Yeast Proteins. Mol Biol Evol 26:1155–1161. https://doi.org/10.1093/molbev/msp031
70. Zhou T, Drummond DA, Wilke CO (2008) Contact Density Affects Protein Evolutionary Rate from Bacteria to Animals. J Mol Evol 66:395–404. https://doi.org/10.1007/s00239-008-9094-4
71. Sasidharan R, Chothia C (2007) The selection of acceptable protein mutations. Proc Natl Acad

Sci 104:10080–10085. https://doi.org/10.1073/pnas.0703737104
72. Chothia C, Lesk AM (1987) The Evolution of Protein Structures. Cold Spring Harb Symp Quant Biol 52:399–405. https://doi.org/10.1101/SQB.1987.052.01.046
73. Tokuriki N, Stricher F, Schymkowitz J, et al (2007) The Stability Effects of Protein Mutations Appear to be Universally Distributed. J Mol Biol 369:1318–1332. https://doi.org/10.1016/j.jmb.2007.03.069
74. Drummond DA, Wilke CO (2008) Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell 134:341–352. https://doi.org/10.1016/j.cell.2008.05.042
75. Tsai C-J, Nussinov R (1997) Hydrophobic folding units at protein-protein interfaces: Implications to protein folding and to protein-protein association. Protein Sci 6:1426–1437. https://doi.org/10.1002/pro.5560060707
76. Young L, Jernigan RL, Covell DG (1994) A role for surface hydrophobicity in protein-protein recognition. Protein Sci 3:717–729. https://doi.org/10.1002/pro.5560030501
77. Korn AP, Burnett RM (1991) Distribution and complementarity of hydropathy in mutisunit proteins. Proteins Struct Funct Genet 9:37–55. https://doi.org/10.1002/prot.340090106
78. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155. https://doi.org/10.1021/bi00483a001
79. Kauzmann W (1959) Some Factors in the Interpretation of Protein Denaturation. pp 1–63
80. Lijnzaad P, Argos P (1997) Hydrophobic patches on protein subunit interfaces: Characteristics and prediction. Proteins Struct Funct Genet 28:333–343. https://doi.org/10.1002/(SICI)1097-0134(199707)28:3<333::AID-PROT4>3.0.CO;2-D
81. Stevens JM, Armstrong RN, Dirr HW (2000) Electrostatic interactions affecting the active site of class sigma glutathione S-transferase. Biochem J 347 Pt 1:193–7
82. Sheinerman F (2000) Electrostatic aspects of protein–protein interactions. Curr Opin Struct Biol 10:153–159. https://doi.org/10.1016/S0959-440X(00)00065-8
83. Xu D, Lin SL, Nussinov R (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations 1 1Edited by B. Honig. J Mol Biol 265:68–84. https://doi.org/10.1006/jmbi.1996.0712
84. Xu D, Tsai CJ, Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng Des Sel 10:999–1012. https://doi.org/10.1093/protein/10.9.999
85. Veselovsky A V., Ivanov YD, Ivanov AS, et al (2002) Protein-protein interactions: mechanisms and modification by drugs. J Mol Recognit 15:405–422. https://doi.org/10.1002/jmr.597
86. Janin J (1995) Elusive affinities. Proteins Struct Funct Genet 21:30–39. https://doi.org/10.1002/prot.340210105
87. Janin J (1995) Protein-protein recognition. Prog Biophys Mol Biol 64:145–166. https://doi.org/10.1016/S0079-6107(96)00001-6
88. Schmitt S, Kuhn D, Klebe G (2002) A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. J Mol Biol 323:387–406. https://doi.org/10.1016/S0022-2836(02)00811-2
89. Vijayakumar M, Wong K-Y, Schreiber G, et al (1998) Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar 1 1Edited by B. Honig. J Mol Biol 278:1015–1024. https://doi.org/10.1006/jmbi.1998.1747
90. Camacho CJ, Weng Z, Vajda S, DeLisi C (1999) Free Energy Landscapes of Encounter Complexes in Protein-Protein Association. Biophys J 76:1166–1178. https://doi.org/10.1016/S0006-3495(99)77281-4
91. Ivanov YD, Ivanov A V., Kaysheva AL, et al (2009) Productive and non-productive complexes in cytochrome P450-containing systems. Biochem Suppl Ser B Biomed Chem 3:183–197. https://doi.org/10.1134/S1990750809020103
92. Jones S, Thornton JM (1997) Analysis of protein-protein interaction sites using surface patches. J Mol Biol 272:121–132. https://doi.org/10.1006/jmbi.1997.1234
93. Jones S, Thornton JM (1997) Prediction of protein-protein interaction sites using patch analysis. J Mol Biol 272:133–143. https://doi.org/10.1006/jmbi.1997.1233
94. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM (1996) Protein clefts in molecular recognition and function. Protein Sci 5:2438–52. https://doi.org/10.1002/pro.5560051206
95. Vaughan CK, Buckle AM, Fersht AR (1999) Structural response to mutation at a protein-protein interface. J Mol Biol 286:1487–1506. https://doi.org/10.1006/jmbi.1998.2559

96. Davies DR, Cohen GH (1996) Interactions of protein antigens with antibodies. Proc Natl Acad Sci U S A 93:7–12. https://doi.org/10.1073/pnas.93.1.7
97. Lichtarge O, Bourne HR, Cohen FE (1996) An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. J Mol Biol 257:342–358. https://doi.org/10.1006/jmbi.1996.0167
98. Samanta U, Pal D, Chakrabarti P (2000) Environment of tryptophan side chains in proteins. Proteins Struct Funct Genet 38:288–300. https://doi.org/10.1002/(SICI)1097-0134(20000215)38:3<288::AID-PROT5>3.0.CO;2-7
99. Fernández A (2002) Desolvation shell of hydrogen bonds in folded proteins, protein complexes and folding pathways. FEBS Lett 527:166–170. https://doi.org/10.1016/S0014-5793(02)03204-0
100. Lee KH, Xie D, Freire E, Amzel LM (1994) Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. Proteins Struct Funct Genet 20:68–84. https://doi.org/10.1002/prot.340200108
101. Lawrence MC, Colman PM (1993) Shape Complementarity at Protein/Protein Interfaces. J Mol Biol 234:946–950. https://doi.org/10.1006/jmbi.1993.1648
102. Li X, Keskin O, Ma B, et al (2004) Protein–Protein Interactions: Hot Spots and Structurally Conserved Residues often Locate in Complemented Pockets that Pre-organized in the Unbound States: Implications for Docking. J Mol Biol 344:781–795. https://doi.org/10.1016/j.jmb.2004.09.051
103. Hubbard SJ, Argos P (1994) Cavities and packing at protein interfaces. Protein Sci 3:2194–2206. https://doi.org/10.1002/pro.5560031205
104. del Sol Mesa A, Pazos F, Valencia A (2003) Automatic Methods for Predicting Functionally Important Residues. J Mol Biol 326:1289–1302. https://doi.org/10.1016/S0022-2836(02)01451-1
105. Goh C-S, Cohen FE (2002) Co-evolutionary Analysis Reveals Insights into Protein–Protein Interactions. J Mol Biol 324:177–192. https://doi.org/10.1016/S0022-2836(02)01038-0
106. Binkowski TA, Adamian L, Liang J (2003) Inferring Functional Relationships of Proteins from Local Sequence and Spatial Surface Patterns. J Mol Biol 332:505–526. https://doi.org/10.1016/S0022-2836(03)00882-9
107. Russell RB, Alber F, Aloy P, et al (2004) A structural perspective on protein–protein interactions. Curr Opin Struct Biol 14:313–324. https://doi.org/10.1016/j.sbi.2004.04.006
108. Moreira IS, Fernandes PA, Ramos MJ (2006) Detailed microscopic study of the full zipA:FtsZ interface. Proteins Struct Funct Bioinforma 63:811–821. https://doi.org/10.1002/prot.20944
109. Moreira IS, Fernandes PA, Ramos MJ (2006) Unraveling the Importance of Protein−Protein Interaction: Application of a Computational Alanine-Scanning Mutagenesis to the Study of the IgG1 Streptococcal Protein G (C2 Fragment) Complex. J Phys Chem B 110:10962–10969. https://doi.org/10.1021/jp054760d
110. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N (2001) Residue frequencies and pairing preferences at protein-protein interfaces. Proteins 43:89–102
111. Halperin I, Wolfson H, Nussinov R (2004) Protein-Protein Interactions. Structure 12:1027–1038. https://doi.org/10.1016/j.str.2004.04.009
112. Perricone U, Gulotta MR, Lombino J, et al (2018) An overview of recent molecular dynamics applications as medicinal chemistry tools for the undruggable site challenge. Medchemcomm 9:920–936. https://doi.org/10.1039/C8MD00166A
113. Rakers C, Bermudez M, Keller BG, et al (2015) Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations? Wiley Interdiscip Rev Comput Mol Sci 5:345–359. https://doi.org/10.1002/wcms.1222
114. Metz A, Ciglia E, Gohlke H (2012) Modulating Protein-Protein Interactions: From Structural Determinants of Binding to Druggability Prediction to Application. Curr Pharm Des 18:4630–4647. https://doi.org/10.2174/138161212802651553
115. Modell AE, Blosser SL, Arora PS (2016) Systematic Targeting of Protein–Protein Interactions. Trends Pharmacol Sci 37:702–713. https://doi.org/10.1016/j.tips.2016.05.008
116. Breckenridge A, Mello M, Psaty BM (2012) New horizons in pharmaceutical regulation. Nat Rev Drug Discov 11:501–502. https://doi.org/10.1038/nrd3787
117. Magee T V. (2015) Progress in discovery of small-molecule modulators of protein–protein interactions via fragment screening. Bioorg Med Chem Lett 25:2461–2468. https://doi.org/10.1016/j.bmcl.2015.04.089
118. Fischer G, Rossmann M, Hyvönen M (2015) Alternative modulation of protein–protein interactions by small molecules. Curr Opin Biotechnol 35:78–85.

https://doi.org/10.1016/j.copbio.2015.04.006
119. Pagliaro L, Felding J, Audouze K, et al (2004) Emerging classes of protein–protein interaction inhibitors and new tools for their development. Curr Opin Chem Biol 8:442–449. https://doi.org/10.1016/j.cbpa.2004.06.006
120. Thorn KS, Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 17:284–285. https://doi.org/10.1093/bioinformatics/17.3.284
121. Bienstock RJ (2012) Computational Drug Design Targeting Protein-Protein Interactions. Curr Pharm Des 18:1240–1254. https://doi.org/10.2174/138161212799436449
122. Reynès C, Host H, Camproux A-C, et al (2010) Designing Focused Chemical Libraries Enriched in Protein-Protein Interaction Inhibitors using Machine-Learning Methods. PLoS Comput Biol 6:e1000695. https://doi.org/10.1371/journal.pcbi.1000695
123. Khoury K, Dömling A (2012) P53 mdm2 inhibitors. Curr Pharm Des 18:4668–78. https://doi.org/10.2174/138161212802651580
124. Wilson CGM, Arkin MR (2010) Small-Molecule Inhibitors of IL-2/IL-2R: Lessons Learned and Applied. pp 25–59
125. Pihan E, Delgadillo RF, Tonkin ML, et al (2015) Computational and biophysical approaches to protein–protein interaction inhibition of Plasmodium falciparum AMA1/RON2 complex. J Comput Aided Mol Des 29:525–539. https://doi.org/10.1007/s10822-015-9842-7
126. Tse C, Shoemaker AR, Adickes J, et al (2008) ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. Cancer Res 68:3421–3428. https://doi.org/10.1158/0008-5472.CAN-07-5836
127. Ackler S, Mitten MJ, Foster K, et al (2010) The Bcl-2 inhibitor ABT-263 enhances the response of multiple chemotherapeutic regimens in hematologic tumors in vivo. Cancer Chemother Pharmacol 66:869–880. https://doi.org/10.1007/s00280-009-1232-1
128. Tahir SK, Wass J, Joseph MK, et al (2010) Identification of Expression Signatures Predictive of Sensitivity to the Bcl-2 Family Member Inhibitor ABT-263 in Small Cell Lung Carcinoma and Leukemia/Lymphoma Cell Lines. Mol Cancer Ther 9:545–557. https://doi.org/10.1158/1535-7163.MCT-09-0651
129. Vogler M, Furdas SD, Jung M, et al (2010) Diminished Sensitivity of Chronic Lymphocytic Leukemia Cells to ABT-737 and ABT-263 Due to Albumin Binding in Blood. Clin Cancer Res 16:4217–4225. https://doi.org/10.1158/1078-0432.CCR-10-0777
130. Gandhi L, Camidge DR, Ribeiro de Oliveira M, et al (2011) Phase I Study of Navitoclax (ABT-263), a Novel Bcl-2 Family Inhibitor, in Patients With Small-Cell Lung Cancer and Other Solid Tumors. J Clin Oncol 29:909–916. https://doi.org/10.1200/JCO.2010.31.6208
131. Sakuma (2011) ABT-263, a Bcl-2 inhibitor, enhances the susceptibility of lung adenocarcinoma cells treated with Src inhibitors to anoikis. Oncol Rep 25:. https://doi.org/10.3892/or.2010.1123
132. Sable R, Jois S (2015) Surfing the Protein-Protein Interaction Surface Using Docking Methods: Application to the Design of PPI Inhibitors. Molecules 20:11569–11603. https://doi.org/10.3390/molecules200611569
133. Saraogi I, Hamilton AD (2008) α-Helix mimetics as inhibitors of protein–protein interactions. Biochem Soc Trans 36:1414–1417. https://doi.org/10.1042/BST0361414
134. Orner BP, Ernst JT, Hamilton AD (2001) Toward Proteomimetics: Terphenyl Derivatives as Structural and Functional Mimics of Extended Regions of an α-Helix. J Am Chem Soc 123:5382–5383. https://doi.org/10.1021/ja0025548
135. Rodriguez JM, Nevola L, Ross NT, et al (2009) Synthetic Inhibitors of Extended Helix-Protein Interactions Based on a Biphenyl 4,4′-Dicarboxamide Scaffold. ChemBioChem 10:829–833. https://doi.org/10.1002/cbic.200800715
136. Tsai JH, Waldman AS, Nowick JS (1999) Two New β-strand Mimics. Bioorg Med Chem 7:29–38. https://doi.org/10.1016/S0968-0896(98)00225-9
137. Fasan R, Dias RLA, Moehle K, et al (2004) Using aβ-Hairpin To Mimic anα-Helix: Cyclic Peptidomimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction. Angew Chemie Int Ed 43:2109–2112. https://doi.org/10.1002/anie.200353242
138. Hirschmann RF, Nicolaou KC, Angeles AR, et al (2009) The β-D-Glucose Scaffold as a β-Turn Mimetic. Acc Chem Res 42:1511–1520. https://doi.org/10.1021/ar900020x
139. Ko E, Liu J, Burgess K (2011) Minimalist and universal peptidomimetics. Chem Soc Rev 40:4411. https://doi.org/10.1039/c0cs00218f
140. Jubb H, Higueruelo AP, Winter A, Blundell TL (2012) Structural biology and drug discovery for protein–protein interactions. Trends Pharmacol Sci 33:241–248.

https://doi.org/10.1016/j.tips.2012.03.006
141. Ball LJ, Kühne R, Schneider-Mergener J, Oschkinat H (2005) Recognition of Proline-Rich Motifs by Protein-Protein-Interaction Domains. Angew Chemie Int Ed 44:2852–2869. https://doi.org/10.1002/anie.200400618
142. Huang Y, Khoury K, Chanas T, Dömling A (2012) Multicomponent Synthesis of Diverse 1,4-Benzodiazepine Scaffolds. Org Lett 14:5916–5919. https://doi.org/10.1021/ol302837h
143. Craik DJ, Fairlie DP, Liras S, Price D (2013) The Future of Peptide-based Drugs. Chem Biol Drug Des 81:136–147. https://doi.org/10.1111/cbdd.12055
144. Siedlecka M, Goch G, Ejchart A, et al (1999) α-Helix nucleation by a calcium-binding peptide loop. Proc Natl Acad Sci 96:903–908. https://doi.org/10.1073/pnas.96.3.903
145. Yang J, Zhao K, Gong Y, et al (1998) α-Helix Nucleation Constant in Copolypeptides of Alanine and Ornithine or Lysine. J Am Chem Soc 120:10646–10652. https://doi.org/10.1021/ja982319d
146. Jochim AL, Arora PS (2009) Assessment of helical interfaces in protein–protein interactions. Mol Biosyst 5:924. https://doi.org/10.1039/b903202a
147. Garner J, Harding MM (2007) Design and synthesis of α-helical peptides and mimetics. Org Biomol Chem 5:3577. https://doi.org/10.1039/b710425a
148. Henchey LK, Jochim AL, Arora PS (2008) Contemporary strategies for the stabilization of peptides in the α-helical conformation. Curr Opin Chem Biol 12:692–697. https://doi.org/10.1016/j.cbpa.2008.08.019
149. Gonzalez-Ruiz D, Gohlke H (2006) Targeting Protein-Protein Interactions with Small Molecules: Challenges and Perspectives for omputational Binding Epitope Detection and Ligand Finding. Curr Med Chem 13:2607–2625. https://doi.org/10.2174/092986706778201530
150. Patgiri A, Jochim AL, Arora PS (2008) A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in α-Helical Conformation. Acc Chem Res 41:1289–1300. https://doi.org/10.1021/ar700264k
151. Wang D, Chen K, Kulp JL, Arora PS (2006) Evaluation of Biologically Relevant Short α-Helices Stabilized by a Main-Chain Hydrogen-Bond Surrogate. J Am Chem Soc 128:9248–9256. https://doi.org/10.1021/ja062710w
152. Kemp DS, Curran TP, Boyd JG, Allen TJ (1991) Studies of N-terminal templates for .alpha.-helix formation. Synthesis and conformational analysis of peptide conjugates of (2S,5S,8S,11S)-1-acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricyclo[2.8.1.04,8]tridecane (Ac-Hel1-OH). J Org Chem 56:6683–6697. https://doi.org/10.1021/jo00023a038
153. Cabezas E, Satterthwait AC (1999) The Hydrogen Bond Mimic Approach: Solid-Phase Synthesis of a Peptide Stabilized as an α-Helix with a Hydrazone Link. J Am Chem Soc 121:3862–3875. https://doi.org/10.1021/ja983212t
154. (2020) CAPRI - Critical Assessment of PRedicted Interactions
155. Pierce B, Weng Z (2007) Structure Prediction of Protein Complexes. pp 109–134
156. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc 125:1731–1737. https://doi.org/10.1021/ja026939x
157. de Vries SJ, van Dijk ADJ, Krzeminski M, et al (2007) HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins Struct Funct Bioinforma 69:726–733. https://doi.org/10.1002/prot.21723
158. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014
159. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
160. Cesareni G, Panni S, Nardelli G, Castagnoli L (2002) Can we infer peptide recognition specificity mediated by SH3 domains? FEBS Lett 513:38–44. https://doi.org/10.1016/S0014-5793(01)03307-5
161. London N, Raveh B, Cohen E, et al (2011) Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. Nucleic Acids Res 39:W249–W253. https://doi.org/10.1093/nar/gkr431
162. Madden DR, Gorga JC, Strominger JL, Wiley DC (1992) The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. Cell 70:1035–1048. https://doi.org/10.1016/0092-8674(92)90252-8
163. Songyang Z (1997) Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains.

Science (80-) 275:73–77. https://doi.org/10.1126/science.275.5296.73
164. Wilson WD (2002) Analyzing Biomolecular Interactions. Science (80-) 295:2103–2105. https://doi.org/10.1126/science.295.5562.2103
165. Hu Z, Ma B, Wolfson H, Nussinov R (2000) Conservation of polar residues as hot spots at protein interfaces. Proteins Struct Funct Genet 39:331–342. https://doi.org/10.1002/(SICI)1097-0134(20000601)39:4<331::AID-PROT60>3.0.CO;2-A
166. DeLano WL (2000) Convergent Solutions to Binding at a Protein-Protein Interface. Science (80-) 287:1279–1283. https://doi.org/10.1126/science.287.5456.1279
167. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 12:14–20. https://doi.org/10.1016/S0959-440X(02)00283-X
168. Wells JA (1990) Additivity of mutational effects in proteins. Biochemistry 29:8509–8517. https://doi.org/10.1021/bi00489a001
169. Cunningham B, Wells J (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science (80-) 244:1081–1085. https://doi.org/10.1126/science.2471267
170. Cunningham BC, Wells JA (1993) Comparison of a Structural and a Functional Epitope. J Mol Biol 234:554–563. https://doi.org/10.1006/jmbi.1993.1611
171. Skolnick J, Fetrow JS, Kolinski A (2000) Structural genomics and its importance for gene function analysis. Nat Biotechnol 18:283–287. https://doi.org/10.1038/73723
172. Morrison KL, Weiss GA (2001) Combinatorial alanine-scanning. Curr Opin Chem Biol 5:302–307. https://doi.org/10.1016/S1367-5931(00)00206-4
173. Wells JA (1991) Systematic mutational analyses of protein-protein interfaces. pp 390–411
174. Wells JA (1994) Structural and functional basis for hormone binding and receptor oligomerization. Curr Opin Cell Biol 6:163–173. https://doi.org/10.1016/0955-0674(94)90132-5
175. Wells JA (1996) Binding in the growth hormone receptor complex. Proc Natl Acad Sci 93:1–6. https://doi.org/10.1073/pnas.93.1.1
176. Thornton JM (2008) The Hans Neurath Award lecture of The Protein Society: Proteins-A testament to physics, chemistry, and evolution. Protein Sci 10:3–11. https://doi.org/10.1110/ps.90001
177. Ma B, Wolfson HJ, Nussinov R (2001) Protein functional epitopes: hot spots, dynamics and combinatorial libraries. Curr Opin Struct Biol 11:364–369. https://doi.org/10.1016/S0959-440X(00)00216-5
178. Buckle AM, Schreiber G, Fersht AR (1994) Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-.ANG. resolution. Biochemistry 33:8878–8889. https://doi.org/10.1021/bi00196a004
179. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A 99:14116–21. https://doi.org/10.1073/pnas.202485799
180. Beard H, Cholleti A, Pearlman D, et al (2013) Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. PLoS One 8:e82849. https://doi.org/10.1371/journal.pone.0082849
181. Moreira IS, Fernandes PA, Ramos MJ (2007) Computational alanine scanning mutagenesis—An improved methodological approach. J Comput Chem 28:644–654. https://doi.org/10.1002/jcc.20566
182. Martins SA, Perez MAS, Moreira IS, et al (2013) Computational Alanine Scanning Mutagenesis: MM-PBSA vs TI. J Chem Theory Comput 9:1311–1319. https://doi.org/10.1021/ct4000372
183. Ma B, Kumar S, Tsai C-J, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng Des Sel 12:713–720. https://doi.org/10.1093/protein/12.9.713
184. Tsai C-J, Lin SL, Wolfson HJ, Nussinov R (1997) Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. Protein Sci 6:53–64. https://doi.org/10.1002/pro.5560060106
185. Tsai C-J, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. Protein Sci 8:1181–1190. https://doi.org/10.1110/ps.8.6.1181
186. Kortemme T, Baker D (2004) Computational design of protein–protein interactions. Curr Opin Chem Biol 8:91–97. https://doi.org/10.1016/j.cbpa.2003.12.008
187. Kmiecik S, Gront D, Kolinski M, et al (2016) Coarse-Grained Protein Models and Their Applications. Chem Rev 116:7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163
188. Schmidt T, Bergner A, Schwede T (2014) Modelling three-dimensional protein structures for

applications in drug design. Drug Discov Today 19:890–897. https://doi.org/10.1016/j.drudis.2013.10.027
189. Schwede T, Sali A, Honig B, et al (2009) Outcome of a Workshop on Applications of Protein Models in Biomedical Research. Structure 17:151–159. https://doi.org/10.1016/j.str.2008.12.014
190. Khoury GA, Smadbeck J, Kieslich CA, Floudas CA (2014) Protein folding and de novo protein design for biotechnological applications. Trends Biotechnol 32:99–109. https://doi.org/10.1016/j.tibtech.2013.10.008
191. Lounnas V, Ritschel T, Kelder J, et al (2013) Current progress in structure-based rational drug design marks a new mindset in drug discovery. Comput Struct Biotechnol J 5:e201302011. https://doi.org/10.5936/csbj.201302011
192. Ferreira L, dos Santos R, Oliva G, Andricopulo A (2015) Molecular Docking and Structure-Based Drug Design Strategies. Molecules 20:13384–13421. https://doi.org/10.3390/molecules200713384
193. Wang T, Wu M-B, Zhang R-H, et al (2015) Advances in Computational Structure-Based Drug Design and Application in Drug Discovery. Curr Top Med Chem 16:901–916. https://doi.org/10.2174/1568026615666150825142002
194. Ozboyaci M, Kokh DB, Corni S, Wade RC (2016) Modeling and simulation of protein–surface interactions: achievements and challenges. Q Rev Biophys 49:e4. https://doi.org/10.1017/S0033583515000256
195. Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr Opin Struct Biol 24:98–105. https://doi.org/10.1016/j.sbi.2013.12.006
196. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How Fast-Folding Proteins Fold. Science (80-) 334:517–520. https://doi.org/10.1126/science.1208351
197. Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomistic Description of the Folding of a Dimeric Protein. J Phys Chem B 117:12935–12942. https://doi.org/10.1021/jp4020993
198. Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253:694–698. https://doi.org/10.1038/253694a0
199. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103:227–249. https://doi.org/10.1016/0022-2836(76)90311-9
200. Takada S (2012) Coarse-grained molecular simulations of large biomolecules. Curr Opin Struct Biol 22:130–137. https://doi.org/10.1016/j.sbi.2012.01.010
201. Levitt M (2014) Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture). Angew Chemie Int Ed 53:10006–10018. https://doi.org/10.1002/anie.201403691
202. Clementi C (2008) Coarse-grained models of protein folding: toy models or predictive tools? Curr Opin Struct Biol 18:10–15. https://doi.org/10.1016/j.sbi.2007.10.005
203. Kolinski A, Skolnick J (2004) Reduced models of proteins and their applications. Polymer (Guildf) 45:511–524. https://doi.org/10.1016/j.polymer.2003.10.064
204. Tozzini V (2005) Coarse-grained models for proteins. Curr Opin Struct Biol 15:144–150. https://doi.org/10.1016/j.sbi.2005.02.005
205. Corradi V, Mancini M, Manetti F, et al (2010) Identification of the first non-peptidic small molecule inhibitor of the c-Abl/14-3-3 protein–protein interactions able to drive sensitive and Imatinib-resistant leukemia cells to apoptosis. Bioorg Med Chem Lett 20:6133–6137. https://doi.org/10.1016/j.bmcl.2010.08.019
206. Arkin MR, Randal M, DeLano WL, et al (2003) Binding of small molecules to an adaptive protein-protein interface. Proc Natl Acad Sci 100:1603–1608. https://doi.org/10.1073/pnas.252756299
207. Eyrisch S, Helms V (2007) Transient Pockets on Protein Surfaces Involved in Protein−Protein Interaction. J Med Chem 50:3457–3464. https://doi.org/10.1021/jm070095g
208. Eyrisch S, Medina-Franco JL, Helms V (2012) Transient pockets on XIAP-BIR2: toward the characterization of putative binding sites of small-molecule XIAP inhibitors. J Mol Model 18:2031–2042. https://doi.org/10.1007/s00894-011-1217-y
209. Eyrisch S, Helms V (2009) What induces pocket openings on protein surface patches involved in protein–protein interactions? J Comput Aided Mol Des 23:73–86. https://doi.org/10.1007/s10822-008-9239-y
210. Chen C, Huang Y, Xiao Y (2013) Enhanced sampling of molecular dynamics simulation of peptides and proteins by double coupling to thermal bath. J Biomol Struct Dyn 31:206–214.

https://doi.org/10.1080/07391102.2012.698244
211. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The Energy Landscape Perspective. Annu Rev Phys Chem 48:545–600. https://doi.org/10.1146/annurev.physchem.48.1.545
212. Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta - Gen Subj 1850:872–877. https://doi.org/10.1016/j.bbagen.2014.10.019
213. Bergonzo C, Henriksen NM, Roe DR, et al (2014) Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. J Chem Theory Comput 10:492–499. https://doi.org/10.1021/ct400862k
214. Lawson CL, Baker ML, Best C, et al (2011) EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res 39:D456–D464. https://doi.org/10.1093/nar/gkq880
215. Dror RO, Dirks RM, Grossman JP, et al (2012) Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annu Rev Biophys 41:429–452. https://doi.org/10.1146/annurev-biophys-042910-155245
216. Bowers KJ, Chow DE, Xu H, et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: ACM/IEEE SC 2006 Conference (SC'06). IEEE, pp 43–43
217. Phillips JC, Hardy DJ, Maia JDC, et al (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys 153:044130. https://doi.org/10.1063/5.0014475
218. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun 91:43–56. https://doi.org/10.1016/0010-4655(95)00042-E
219. Koehl P (2006) Electrostatics calculations: latest methodological advances. Curr Opin Struct Biol 16:142–151. https://doi.org/10.1016/j.sbi.2006.03.001
220. Martínez-Mayorga K, Pitman MC, Grossfield A, et al (2006) Retinal Counterion Switch Mechanism in Vision Evaluated by Molecular Simulations. J Am Chem Soc 128:16502–16503. https://doi.org/10.1021/ja0671971
221. Shaw DE, Maragakis P, Lindorff-Larsen K, et al (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. Science (80-) 330:341–346. https://doi.org/10.1126/science.1187409
222. Shirts M (2000) COMPUTING: Screen Savers of the World Unite! Science (80-) 290:1903–1904. https://doi.org/10.1126/science.290.5498.1903
223. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1−39). J Am Chem Soc 132:1526–1528. https://doi.org/10.1021/ja9090353
224. Lane TJ, Shukla D, Beauchamp KA, Pande VS (2013) To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin Struct Biol 23:58–65. https://doi.org/10.1016/j.sbi.2012.11.002
225. Roos K, Wu C, Damm W, et al (2019) OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. J Chem Theory Comput 15:1863–1874. https://doi.org/10.1021/acs.jctc.8b01026
226. Piana S, Robustelli P, Tan D, et al (2020) Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. J Chem Theory Comput 16:2494–2507. https://doi.org/10.1021/acs.jctc.9b00251
227. Huang J, Rauscher S, Nawrocki G, et al (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14:71–73. https://doi.org/10.1038/nmeth.4067
228. Best RB, Zheng W, Mittal J (2014) Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput 10:5113–5124. https://doi.org/10.1021/ct500569b
229. Robustelli P, Piana S, Shaw DE (2018) Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci 115:E4758–E4766. https://doi.org/10.1073/pnas.1800690115
230. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps. J Med Chem 49:5851–5855. https://doi.org/10.1021/jm060999m
231. Moitessier N, Englebienne P, Lee D, et al (2008) Towards the development of universal, fast and

highly accurate docking/scoring methods: a long way to go. Br J Pharmacol 153:S7–S26. https://doi.org/10.1038/sj.bjp.0707515
232. Grinter S, Zou X (2014) Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. Molecules 19:10150–10176. https://doi.org/10.3390/molecules190710150
233. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins Struct Funct Bioinforma 56:235–249. https://doi.org/10.1002/prot.20088
234. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256
235. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430
236. Vilar S, Cozza G, Moro S (2008) Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. Curr Top Med Chem 8:1555–1572. https://doi.org/10.2174/156802608786786624
237. Jones G, Willett P, Glen RC, et al (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748. https://doi.org/10.1006/jmbi.1996.0897
238. Cross JB, Thompson DC, Rai BK, et al (2009) Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. J Chem Inf Model 49:1455–1474. https://doi.org/10.1021/ci900056c
239. Schneidman-Duhovny D, Nussinov R, Wolfson HJ (2007) Automatic prediction of protein interactions with large scale motion. Proteins Struct Funct Bioinforma 69:764–773. https://doi.org/10.1002/prot.21759
240. Ewing TJA, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15:411–428. https://doi.org/10.1023/A:1011115820450
241. Koch O, Cappel D, Nocker M, et al (2011) Virtual screening using structure-based consensus pharmacophore models and ensemble docking based on MD-generated conformations. J Cheminform 3:O23. https://doi.org/10.1186/1758-2946-3-S1-O23
242. Voet A, Banwell EF, Sahu KK, et al (2013) Protein Interface Pharmacophore Mapping Tools for Small Molecule Protein: Protein Interaction Inhibitor Discovery. Curr Top Med Chem 13:989–1001. https://doi.org/10.2174/1568026611313090003
243. De Luca L, Barreca ML, Ferro S, et al (2009) Pharmacophore-Based Discovery of Small-Molecule Inhibitors of Protein-Protein Interactions between HIV-1 Integrase and Cellular Cofactor LEDGF/p75. ChemMedChem 4:1311–1316. https://doi.org/10.1002/cmdc.200900070
244. Murgueitio MS, Henneke P, Glossmann H, et al (2014) Prospective Virtual Screening in a Sparse Data Scenario: Design of Small-Molecule TLR2 Antagonists. ChemMedChem 9:813–822. https://doi.org/10.1002/cmdc.201300445
245. Wolber G, Langer T (2005) LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 45:160–169. https://doi.org/10.1021/ci049885e
246. Wolber G, Dornhofer AA, Langer T (2007) Efficient overlay of small organic molecules using 3D pharmacophores. J Comput Aided Mol Des 20:773–788. https://doi.org/10.1007/s10822-006-9078-7
247. Dixon SL, Smondyrev AM, Knoll EH, et al (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20:647–671. https://doi.org/10.1007/s10822-006-9087-6
248. Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. Chem Biol <html_ent glyph="@amp;" ascii="&"/> Drug Des 67:370–372. https://doi.org/10.1111/j.1747-0285.2006.00384.x
249. Koes DR, Camacho CJ (2011) Pharmer: Efficient and Exact Pharmacophore Search. J Chem Inf Model 51:1307–1314. https://doi.org/10.1021/ci200097m
250. Schaller D, Šribar D, Noonan T, et al (2020) Next generation 3D pharmacophore modeling. WIREs Comput Mol Sci 10:. https://doi.org/10.1002/wcms.1468
251. Voet A, Y.J. Zhang K (2012) Pharmacophore Modelling as a Virtual Screening Tool for the

Discovery of Small Molecule Protein-protein Interaction Inhibitors. Curr Pharm Des 18:4586–4598. https://doi.org/10.2174/138161212802651616
252. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering High-Affinity Ligands for Proteins: SAR by NMR. Science (80-) 274:1531–1534. https://doi.org/10.1126/science.274.5292.1531
253. Erlanson DA (2006) Fragment-based lead discovery: a chemical update. Curr Opin Biotechnol 17:643–652. https://doi.org/10.1016/j.copbio.2006.10.007
254. Fredriksson S, Gullberg M, Jarvius J, et al (2002) Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol 20:473–477. https://doi.org/10.1038/nbt0502-473
255. Yin H, Chu A, Li W, et al (2009) Lipid G Protein-coupled Receptor Ligand Identification Using β-Arrestin PathHunter™ Assay. J Biol Chem 284:12328–12338. https://doi.org/10.1074/jbc.M806516200
256. Karpova T, McNally JG (2006) Detecting Protein–Protein Interactions with CFP-YFP FRET by Acceptor Photobleaching. Curr Protoc Cytom 35:. https://doi.org/10.1002/0471142956.cy1207s35
257. Dias DM, Van Molle I, Baud MGJ, et al (2014) Is NMR Fragment Screening Fine-Tuned to Assess Druggability of Protein–Protein Interactions? ACS Med Chem Lett 5:23–28. https://doi.org/10.1021/ml400296c
258. Galdeano C, Gadd MS, Soares P, et al (2014) Structure-Guided Design and Optimization of Small Molecules Targeting the Protein–Protein Interaction between the von Hippel–Lindau (VHL) E3 Ubiquitin Ligase and the Hypoxia Inducible Factor (HIF) Alpha Subunit with in Vitro Nanomolar Affinities. J Med Chem 57:8657–8663. https://doi.org/10.1021/jm5011258
259. Mani T, Wang F, Knabe WE, et al (2013) Small-molecule inhibition of the uPAR·uPA interaction: Synthesis, biochemical, cellular, in vivo pharmacokinetics and efficacy studies in breast cancer metastasis. Bioorg Med Chem 21:2145–2155. https://doi.org/10.1016/j.bmc.2012.12.047
260. Hain AUP, Bartee D, Sanders NG, et al (2014) Identification of an Atg8-Atg3 Protein–Protein Interaction Inhibitor from the Medicines for Malaria Venture Malaria Box Active in Blood and Liver Stage Plasmodium falciparum Parasites. J Med Chem 57:4521–4531. https://doi.org/10.1021/jm401675a
261. Hain AUP, Weltzer RR, Hammond H, et al (2012) Structural characterization and inhibition of the Plasmodium Atg8–Atg3 interaction. J Struct Biol 180:551–562. https://doi.org/10.1016/j.jsb.2012.09.001
262. Duong-Thi M-D, Bergström G, Mandenius C-F, et al (2014) Comparison of weak affinity chromatography and surface plasmon resonance in determining affinity of small molecules. Anal Biochem 461:57–59. https://doi.org/10.1016/j.ab.2014.05.023
263. Cascio S, Finn O (2016) Intra- and Extra-Cellular Events Related to Altered Glycosylation of MUC1 Promote Chronic Inflammation, Tumor Progression, Invasion, and Metastasis. Biomolecules 6:39. https://doi.org/10.3390/biom6040039
264. Cascio S, Finn OJ (2015) Complex of MUC1, CIN85 and Cbl in Colon Cancer Progression and Metastasis. Cancers (Basel) 7:342–52. https://doi.org/10.3390/cancers7010342
265. Cascio S, Farkas AM, Hughey RP, Finn OJ (2013) Altered glycosylation of MUC1 influences its association with CIN85: the role of this novel complex in cancer cell invasion and migration. Oncotarget 4:1686–1697. https://doi.org/10.18632/oncotarget.1265
266. Bae JY, Park HH (2011) Crystal Structure of NALP3 Protein Pyrin Domain (PYD) and Its Implications in Inflammasome Assembly. J Biol Chem 286:39528–39536. https://doi.org/10.1074/jbc.M111.278812
267. Stutz A, Kolbe C-C, Stahl R, et al (2017) NLRP3 inflammasome assembly is regulated by phosphorylation of the pyrin domain. J Exp Med 214:1725–1736. https://doi.org/10.1084/jem.20160933
268. Gulotta MR, Lombino J, Perricone U, et al (2020) Targeting SARS-CoV-2 RBD Interface: a Supervised Computational Data-Driven Approach to Identify Potential Modulators. ChemMedChem cmdc.202000259. https://doi.org/10.1002/cmdc.202000259
269. Hoffmann M, Kleine-Weber H, Schroeder S, et al (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 1–10. https://doi.org/10.1016/j.cell.2020.02.052
270. Heurich M, Martinez-Barricarte R, Francis NJ, et al (2011) Common polymorphisms in C3, factor B, and factor H collaborate to determine systemic complement activity and disease risk.

Proc Natl Acad Sci 108:8761–8766. https://doi.org/10.1073/pnas.1019338108
271. Ando H, Natsume A, Senga T, et al (2014) Peptide-based inhibition of the HOXA9/PBX interaction retards the growth of human meningioma. Cancer Chemother Pharmacol 73:53–60. https://doi.org/10.1007/s00280-013-2316-5
272. LaRonde-LeBlanc NA (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev 17:2060–2072. https://doi.org/10.1101/gad.1103303
273. Fasano S, D'Antoni A, Orban PC, et al (2009) Ras-Guanine Nucleotide-Releasing Factor 1 (Ras-GRF1) Controls Activation of Extracellular Signal-Regulated Kinase (ERK) Signaling in the Striatum and Long-Term Behavioral Responses to Cocaine. Biol Psychiatry 66:758–768. https://doi.org/10.1016/j.biopsych.2009.03.014
274. Papale A, Morella IM, Indrigo MT, et al (2016) Impairment of cocaine-mediated behaviours in mice by clinically relevant Ras-ERK inhibitors. Elife 5:. https://doi.org/10.7554/eLife.17111

CHAPTER TWO

MUC1-CIN85 INTERACTION – Supervised multi-computational approach to identify hit compounds

2.1 Introduction

Cancer is one of the most severe diseases and the second leading cause of death globally. The World Health Organisation (WHO) states that in 2018 the cancer burden rose to 18.1 million new cases and 9.6 million cancer deaths. Many efforts have been done to tackle this global threat, and novel treatments have shown to be reliable and mainly improve the patient's life quality. However, it is expected that even more research efforts will be necessary to fight cancer morbidity and mortality [1]. Furthermore, to complicate already hard clinical cases, late stages of the disease are often associated to invasiveness and metastasis, a process composed by multiple steps when cancer cells detach from the basement membrane, degrade the surrounding matrix, and finally invade the neighbouring tissues and enter the bloodstream [2]. In this context, experimental evidence, such as lung experimental assays performed at the University of Pittsburgh, unveiled the crucial role of a protein-protein interaction between the Mucin 1 (MUC1) and the Cbl-interacting protein of 85 kDa (CIN85) throughout the invasiveness and metastasis process [3].

MUC1 is a transmembrane glycoprotein that is normally expressed in the glandular or luminal epithelial cells of mammary gland, oesophagus, stomach, duodenum, pancreas, uterus, prostate, and lungs, and to a lesser extent, in hematopoietic cells [4, 5]. MUC1 exerts a protecting role of the epithelia. The extracellular domain, that extends up to 200-500 nm from the cell surface, is heavily glycosylated with extended negatively charged sugar branches [6]. This glycosylation generates a physical barrier that prevents pathogenic accessibility and adhesion. Furthermore, the sugar chains oligomerise to form a mucinous gel, that lubricates and protect the underlying epithelia from pollutants, pH changes, desiccation and microbes [7, 8].

MUC1 consists of two peptide fragments extracellularly associated through stable hydrogen bonds: the longer N-terminal subunit (MUC1-N) and the shorter C-terminal subunit (MUC1-C) [9]. The first fragment is composed of two domains: the proline, threonine and serine-rich (PTS) domain and the Sea urchin sperm protein enterokinase and agrin (SEA) domain. The PTS domain is also designated as the variable number tandem repeat (VNTR) region usually consisting of 20-21 amino acids, that are 20 to 120-fold repeated in normal cells [10]. In the VNTR region, serine and threonine residues compose about 40% of the amino acids, and they are extensively *O*-glycosylated, while *N*-glycosylation occurs on the five asparagine residues of VNTR. Indeed, it is estimated that 50-90% of MUC1 weight is to be ascribed to the *O*-glycosylation and a lesser extent to the *N*-glycosylation. The sugar moieties mask the peptide core working as a shield from undergoing proteolytic cleavage by environmental enzymes. On the contrary, the shorter subunit of MUC1 (MUC1-C) consists of three domains: the 58 amino acid extracellular domain (ECD), a 28 amino acid transmembrane domain (TMD), and a 72 amino acid cytoplasmic tail (CT) (figure 2.1) [4, 11].

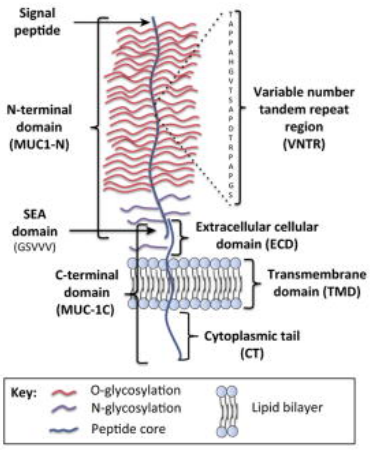


Figure 2.1. MUC1 composition [11]

Under normal conditions, MUC1 is located on the plasma membrane as a heterodimeric complex. On the other hand, it has been associated with pathological conditions, when presenting aberrant glycosylation in cancer cells (figure 2.2).

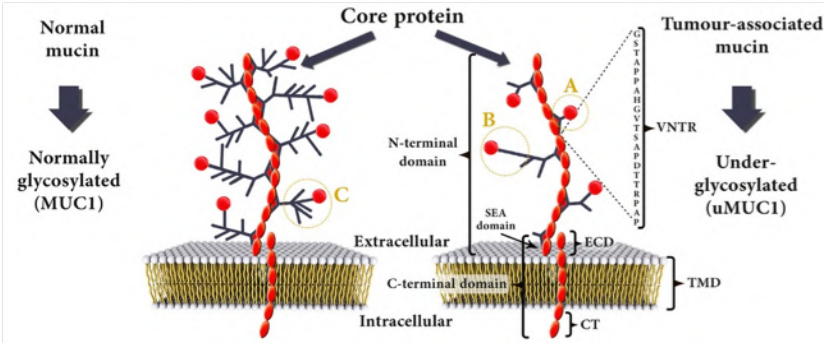


Figure 2.2. Glycosylated MUC1 VNTR vs hypoglycosylated MUC1 VNTR at the endothelial cell membrane [12]

Normal MUC1 and the tumour-associated MUC1 (TA-MUC1) differ for biochemical features, but also their cellular distribution. Normally, the glycosylation process starts when the α-GalNAc transferase catalyses the addition of N-acetyl galactosamine (GalNAc) to serine and threonine residues of the MUC1-N backbone, generating the Tn antigen. Then, the galactose is added to Tn antigen by a transferase enzyme (Core 1 β-1,3-galactose transferase) to generate Core 1 *O*-glycan or T antigen. This latter is again added with N-acetyl glucosamine (GlcNAc) by another transferase enzyme (Core 2 β-1,6-N-acetyl glucosamine), catalysing the formation of Core 2 *O*-glycan. The sugar branches of the Core 2 *O*-glycans undergo further chain elongation followed by termination upon addition of fucose or sialic acid to the terminal sugar [13] (figure 2.3A).

In pathological conditions, this sugar-coating process on MUC1-N backbone is disrupted because Tn and T antigens are sialylated to sialyl Tn and sialyl T by the enzymes α-2,6-sialyltransferase and α-2,3-sialyltransferase, respectively, due to high levels of expression of these enzymes in cancer cells. This sialylation causes a premature termination of chain elongation resulting in hypoglycosylation of MUC1, that impacts the stability and subcellular localization of MUC1 [14] (figure 2.3B). In particular, in colon cancer cells, a specific mechanism dominates, *i.e.* MUC1 overexpresses the sialyl

Lewis^{x} (sLe^{x}) and sialyl Lewis^{a} (sLe^{a}) epitopes and a decrease in *O*-acetylation appears to contribute to such expression in these cells [14].

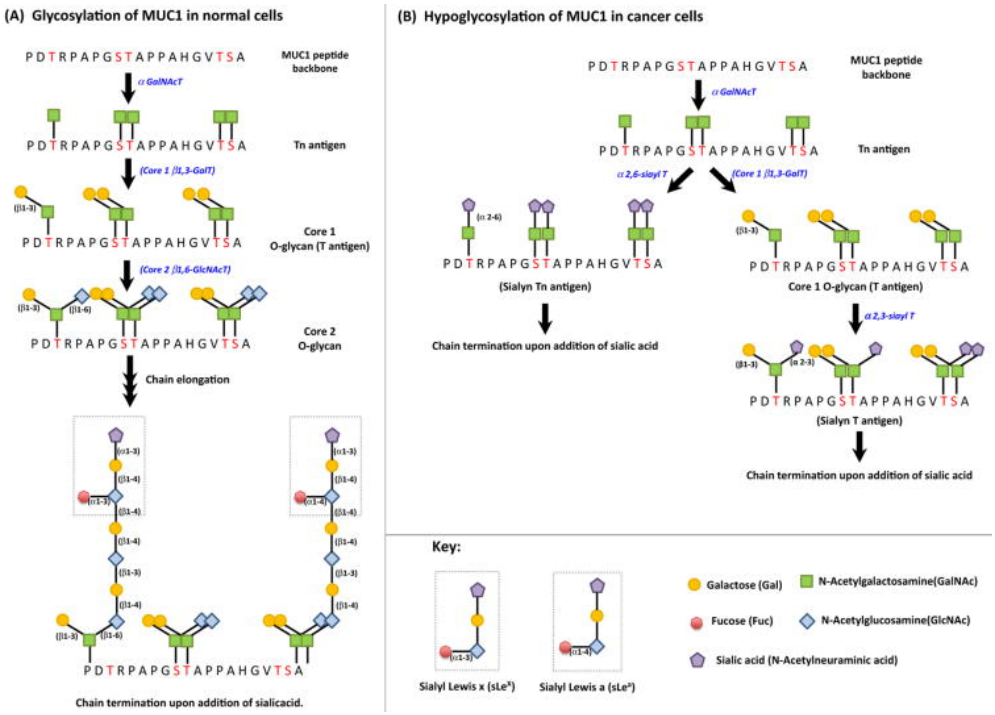


Figure 2.3. A) Mechanism of sugar branches elongation on normal MUC1 VNTR; **B)** Aberrant sugar elongation of TA-MUC1 [11]

Fewer sugars on the VNTRs pave the way to new scenarios of interaction with other proteins, owing to the increase of MUC1 peptide backbone accessibility to more efficient or completely new protein-protein interactions, that can profoundly change intracellular signalling in tumours compared to normal cells. Indeed, experimental *in vitro* and *in vivo* assays highlighted a new protein-protein association between MUC1 and CIN85 [3].

CIN85 is a protein usually associated to the Cbl proteins (Cbl-b and c-Cbl), multi-adaptor-associated ubiquitin ligases, that recruit CIN85 and initiate endocytic internalization and trafficking and sorting of several other proteins [15, 16].
CIN85 contains three Src homology 3 (SH3) domains at its N-terminus followed by a proline-rich region and a C-terminal coiled-coil region [17] (figure 2.4).

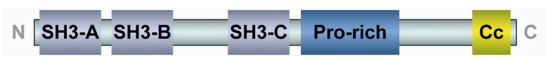


Figure 2.4. CIN85 protein consisting of three Src homology 3 (SH3) domains, a proline-rich region and a C-terminal coiled-coil region [17]

The SH3 domains are small modular interacting domains, that generally bind to proline-containing targets (PXXP motifs). It has been shown that SH3 domains of CIN85 recognise an atypical proline-arginine motif, PXXXPR (where X is any amino acid), in Cbl and many other proteins implicated in the control of clathrin-mediated receptor endocytosis, receptor recycling, and cytoskeletal rearrangements [15–18]. Indeed, it is noteworthy that each repeat in MUC1 VNTR contains a highly conserved sequence,

PDTRPA (figure 2.5), representing a good candidate for binding to CIN85, as it was confirmed by the experimental assays [3].

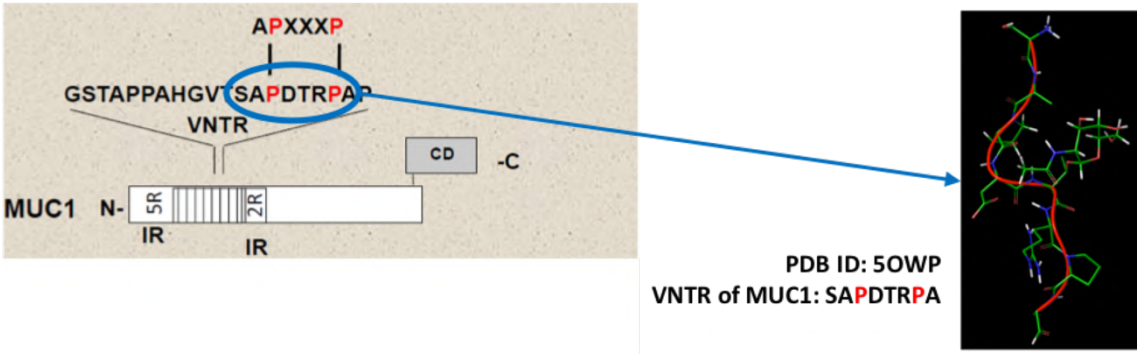


Figure 2.5. On the left, an example of the amino acid composition of MUC1 including VNTR [19]; on the right, 3D structure of a VNTR of MUC1 (PDB ID: 5OWP)

CIN85 has also been implicated in a number of important cellular processes including signal transduction, vesicle-mediated transport, cytoskeleton remodelling, immunological synapse, cell migration and invasion [16, 20–23]. CIN85 was reported to be detected on lamellopodia and invadopodia, which are involved in cell adhesion and migration, suggesting that overexpression of CIN85 could promote invasiveness of cancer cells [24].

Indeed, co-precipitation experiments performed on mouse ovarian cancer and human breast cancer cell lines highlighted an association between MUC1 and CIN85, while the use of confocal immunofluorescence microscopy identified the co-localization of these two proteins both intracellularly and on the cell surface. In fact, the protein-protein complexes were found mainly in the membranes (including plasma membrane, mitochondria, Golgi apparatus, endosomes and endoplasmic reticulum), the cytosol and cytoskeleton fractions, with very little or none detected in the nuclear fraction. Furthermore, an important discovery was that these protein-protein complexes co-localised on large invadopodia-like protrusions (figure 2.6A-C), whereas the invadopodia are structures that are the first step bringing to tumour cell adhesion and invasion.

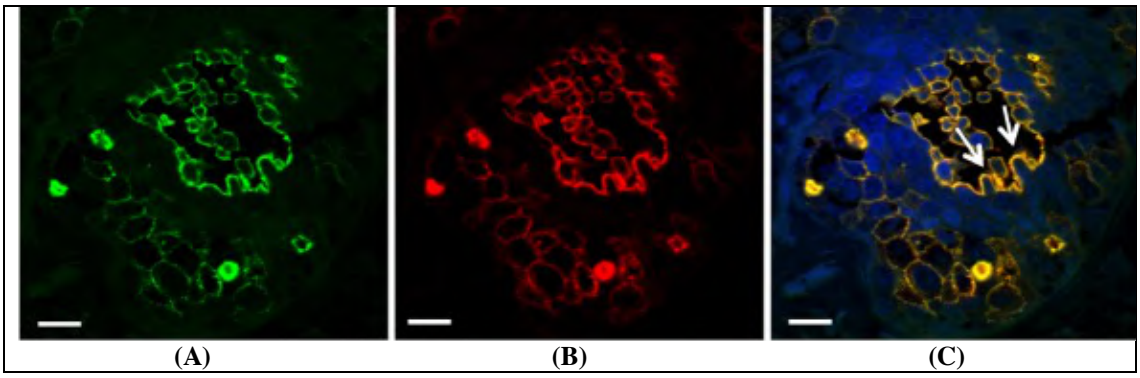


Figure 2.6. Confocal immunofluorescence microscopy assays on mouse ovarian cancer and human breast cancer cell lines revealed highly stained invadopodia-like protrusions of **(A)** anti-MUC1 antibodies (green-stained), **(B)** anti-CIN85 antibodies (red-stained) and **(C)** both together (yellow-stained). White arrows indicate the invadopodia-like protrusions [3].

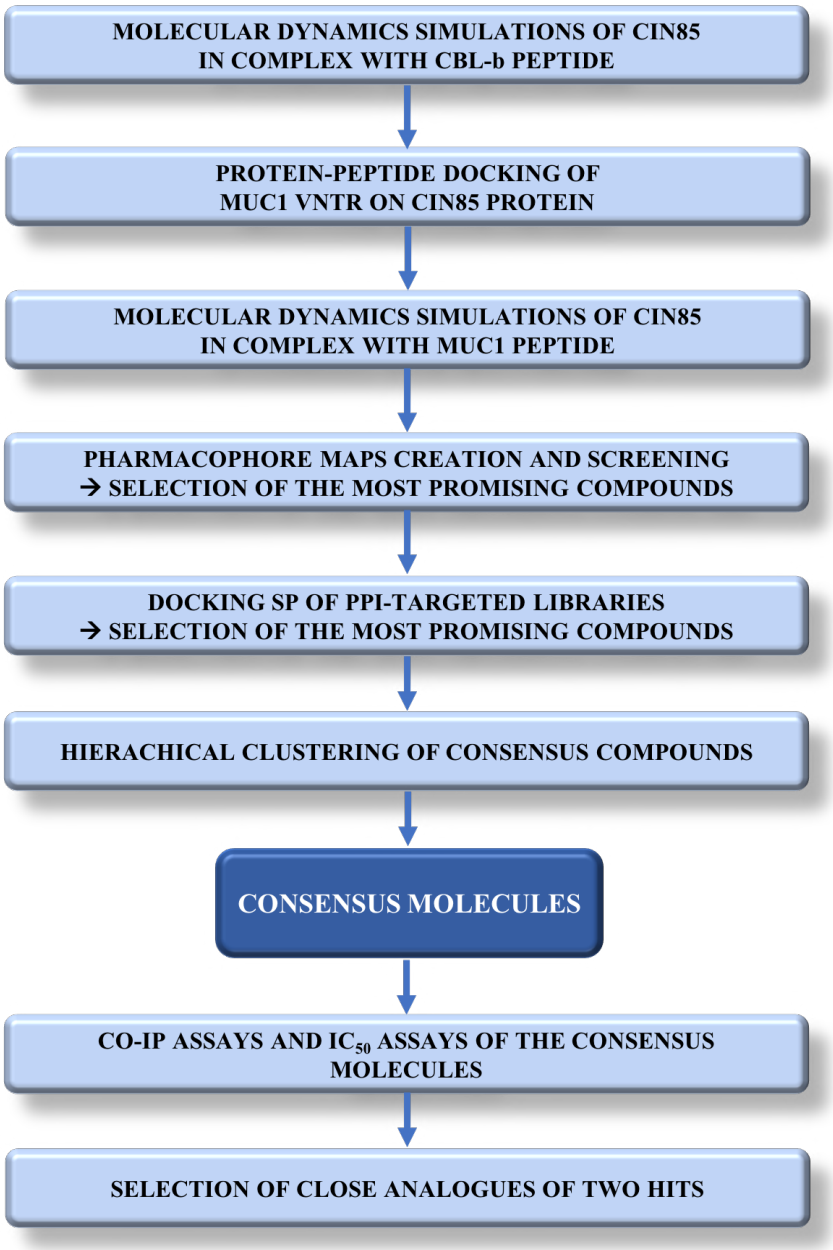
Therefore, the MUC1-CIN85 association suggested that MUC1 regulates also CIN85 role in cell migration and invasion. It was also confirmed by assays *in vivo* using

a well-characterised model of experimental lung metastasis produced by intravenous injection of B16 mouse melanoma cells into mice. The results highlighted the detection of metastatic lung nodules in the mice, thus confirming the previous deduction from *in vitro* assays [3].

In light of the above, CIN85-MUC1 interaction appears to be an interesting pharmacological target. The druggability of this protein-protein interface should be very challenging, considering that an X-ray crystal structure of the complex is not available. Therefore, the strategy of targeting a protein-protein interaction is usually associated with intrinsic issues, where a binding pocket is not detectable and the interaction surface is very shallow and broad.

The SH3 domains of CIN85 have evolved to incorporate different modes of recognition for cognate ligands, providing a level of specificity to achieve its multifunctional capability. Binding of peptides containing proline-rich sequences to SH3 domains can occur in two opposing orientations. In some cases, the stoichiometry is 1:1, where a single peptide of a protein partner binds one SH3 domain. On the other hand, recent studies reported that a single proline-rich motif can simultaneously bind to both N-terminal SH3 domain A (referred to as CIN85A) and SH3 domain B (referred to as CIN85B), due to the long linker between these regions [25]. An example of this heterotrimeric complex was detected by Ceregido *et al.,* that resolved an X-ray crystal structure of a dimer of CIN85 in complex with a proline-rich peptide of Cbl-b with the sequence 902-PARPPKPRPRR-912 [26]. Mutational studies performed on CIN85 protein in complex with this Cbl-b peptide allowed to identify crucial residues, that could guide the design of novel modulators of CIN85 by potentially inhibiting the interaction with MUC1. Therefore, starting from the PDB structure analysis of CIN85 protein (PDB ID: 2BZ8) and from literature data, a hypothesis was formulated: CIN85 could share the same interacting surface for Cbl-b protein and MUC1, considering the high similarity between PXXXP motifs of these two proteins. Moreover, experimental assays performed by Cascio *et al*. [27] reported that CIN85 interact with both proteins (Cbl-b and MUC1) in a dimeric form. In fact, the authors used a MUC1 peptide agonist capable of inducing CIN85 dimerization, that showed to enhance the binding between CIN85 homodimer and MUC1 peptide in two times higher extent than in its absence [27]. This evidence suggested that the binding region of MUC1 on CIN85 homodimer could be the same of Cbl-b as visible in PDB 2BZ8 [26].

Therefore, based on these considerations, the goal of this project was to identify inhibitors of the dimeric form of CIN85 that was expected establishing interaction with MUC1, in order to tackle the metastatic process triggered by this interaction. For this purpose, computational techniques, such as molecular dynamics, docking and pharmacophore screenings were applied. In detail, first a structural analysis of the proteins was conducted by performing MD simulations, and then a consensus approach based on docking and pharmacophore techniques was applied to select putative compounds able to target the dimeric form of CIN85. Scheme 2.1 summarises the steps of this workflow, that are in detail described below leading to the identification of two potential hit compounds. Indeed, the biological activity of these molecules was confirmed by co-immunoprecipitation assays and interaction inhibition assays, reporting promising IC_{50} values in low micromolar (μM) and even nanomolar (nM) ranges. Close analogues of these two compounds have been identified and will be tested to confirm the results for the previous singletons.



Scheme 2.1. Overview of the computational workflow performed to identify consensus molecules potentially modulating CIN85 protein by inhibiting the binding interface

2.1.1 A computational analysis of CIN85 dimer interacting interface

Src Homology 3 (SH3) domains of CIN85 exhibit a characteristic beta-barrel fold, that usually consists of five or six β-strands arranged as two tightly packed anti-parallel β-sheets [17]. As previously mentioned, CIN85 was found binding MUC1 in a dimeric form, generating a heterotrimeric complex. To date, the only currently available PDB structure of CIN85 dimer complexed with a peptide is the PDB 2BZ8 [28]. In this complex, a proline-arginine motif of Cbl-b protein is sandwiched by two SH3 domains of CIN85 arranged in a dimeric form. Figure 2.7 depicts the heterotrimeric complex.

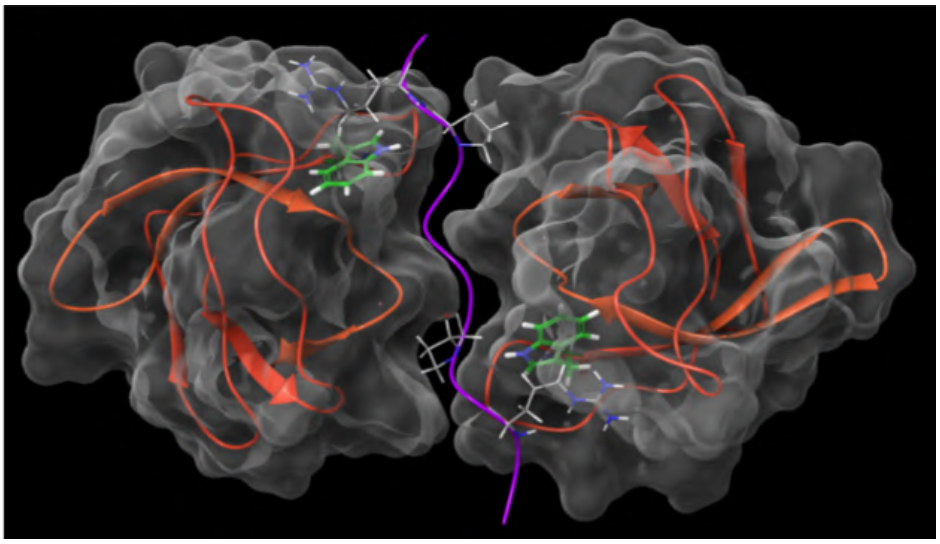


Figure 2.7. 3D representation of CIN85-Cbl-b complex (PDB ID: 2BZ8). Orange structures embedded into grey transparent surfaces are the CIN85 SH3 domains in complex with the short purple peptide that is a proline-arginine-rich fragment of Cbl-b protein [28].

Cbl-b fragment consists of eleven amino acids, 902-PARPPKPRPRR-912, arranged in a polyproline II (PPII) helix conformation. This protein is a ubiquitin ligase playing a key role in receptor downregulation by mediating multiple monoubiquitinations of the receptors and promoting their sorting for lysosomal degradation [28]. The structure analysis of the PDB 2BZ8 unveiled a pseudo-symmetrical orientation of the peptide, as depicted in figure 2.8, with a stoichiometry 1:0.57. Indeed, the N-terminal portion of Cbl-b fragment is engaged in the Type I orientation, while the C-terminal portion is involved in the Type II orientation [26].

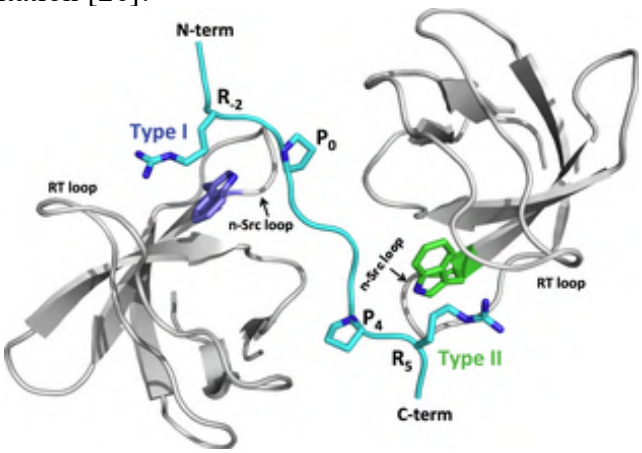


Figure 2.8. Pseudo-symmetrical orientation of Cbl-b peptide in complex with two CIN85 SH3 domains. The N-terminal region of the peptide is involved in the Type I orientation, while the C-terminus is engaged in the Type II orientation [26]

Furthermore, the analysis of Cbl-b binding mode to two SH3 domains of CIN85 allowed to identify the interactions and the involved residues for both proteins, that are shown in figure 2.9 including a 2D interaction diagram.

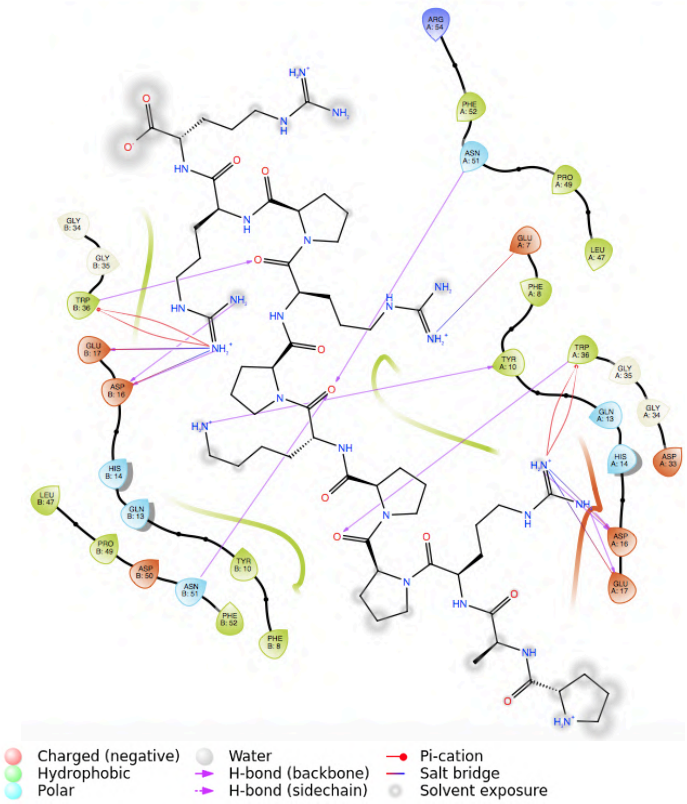


Figure 2.9. 2D interaction diagram of Cbl-b peptide and the established interactions with CIN85 SH3 domains amino acids. Purple arrows stand for hydrogen bonds, red arrows are salt bridges and red lines pi-cation interactions

The observed interactions were mainly H-bonds, where the side chains of Cbl-b amino acids Arg904 and Arg911 interact with Asp16 and Glu17 belonging to SH3 domains, creating H-bonds and salt bridges. The Cbl-b residues Lys907 and Arg909 backbone carbonyl groups establish hydrogen bonds with Asn51 and Trp36 residues of the two CIN85 domains, respectively. Moreover, Arg904 and Arg911 are also involved in pi-cation interaction with Trp36. Finally, some hydrophobic contacts were detected between Cbl-b Pro906, Pro908, Pro910 and CIN85 Trp36 and Phe52. All these interactions are grouped in table 2.1.

Table 2.1. Residues of Cbl-b protein and CIN85 SH3 domains involved in interactions.

Interaction type	Cbl-b Peptide	CIN85 SH3 domains
H-Bond	Arg904, Arg911, Lys907, Arg909	Asp16, Asp17, Asn51, Trp36
Salt bridge	Arg904, Arg911	Asp16, Asp17
Pi-Cation	Arg904, Arg911	Trp36
Hydrophobic	Pro906, Pro908, Pro910	Trp36, Phe52

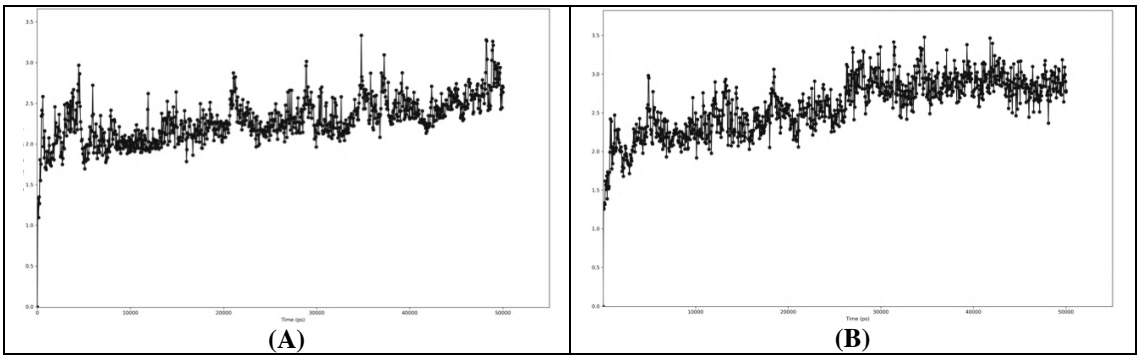
Furthermore, in order to determine the functional role of the two arginine residues of Cbl-b peptide (Arg904 and Arg911) for binding CIN85 and the receptor downregulation, Jozic *et al*. performed mutational studies, where Arg904 and Arg911 of Cbl-b were mutated to alanine (R904A and R911A, respectively). The amount of co-precipitated CIN85 together with the protein partner was quantified by reporting that the mutation R911A reduced the interaction with CIN85 by approximately 60%, while the mutation R904A reduced the binding by about 25%. Finally, mutations of both sites in Cbl-b abolished the co-precipitation with CIN85 [28]. Moreover, mutation of Lys907 did

not affect the formation of a trimeric complex [26]. These data were also consistent with the NMR titration experiments performed by Ceregido *et al*. that calculated different K_D values for Type I and Type II orientations, whereas the first orientation showed $K_D = 46.9$ μM, while Type II provided $K_D = 2.0$ μM, showing a preferential behaviour for Type II orientation. Once again this data demonstrated that Arg911, that is involved in Type II Cbl-b peptide orientation, is more important in terms of trimeric complex formation, in comparison with Arg904 that is instead involved in the Type I orientation [26]. Therefore, it should be notable that mutation of Arg904 does not appreciably change the apparent affinity of the peptide for CIN85, but changes the relative enthalpic and entropic contributions to ΔG of the complex [28]. All this information was crucial for the several steps of this project and was processed in order to guide the computational studies. Moreover, as above mentioned, experimental assays suggested that MUC1 VNTR is able to bind the dimeric form of CIN85 [27], and this assumption was investigated as described in the next steps.

2.2 Results and discussion

2.2.1 Molecular Dynamics simulation of CIN85-Cbl-b complex

The first step of this work was the analysis of the PDB structure of CIN85 in a dimeric form bound to Cbl-b peptide (PDB ID: 2BZ8 with resolution 2.0 Å) [28]. The only available data about crucial interactions between CIN85 and the protein partner were retrieved from the above-mentioned mutational studies and from a single PDB structure (PDB 2BZ8), where two SH3 domains of CIN85 bind to the proline-arginine-rich peptide of Cbl-b. However, a PDB structure accounts only static positions and conformations of amino acid side chains or protein backbone, hence it could be likely that some interaction cannot be visible in that precise conformation of the complex. Therefore, two short Molecular Dynamics simulations of the PDB 2BZ8 were run on CIN85-Cbl-b peptide complex to explore the most stable and frequent interactions established during the trajectories. For this purpose, the complex structure was first optimised at pH 7.0 ± 2.0 and the fluoroproline present on the peptide was mutated to a natural proline. The simulation time was set 50 ns for each system and both MD were run. The stability of the complexes was checked by visualising the RMSD plots (Plot 2.1A-B) and energy, temperature, pressure and volume were monitored during the entire trajectories (average energy values in table 2.2).

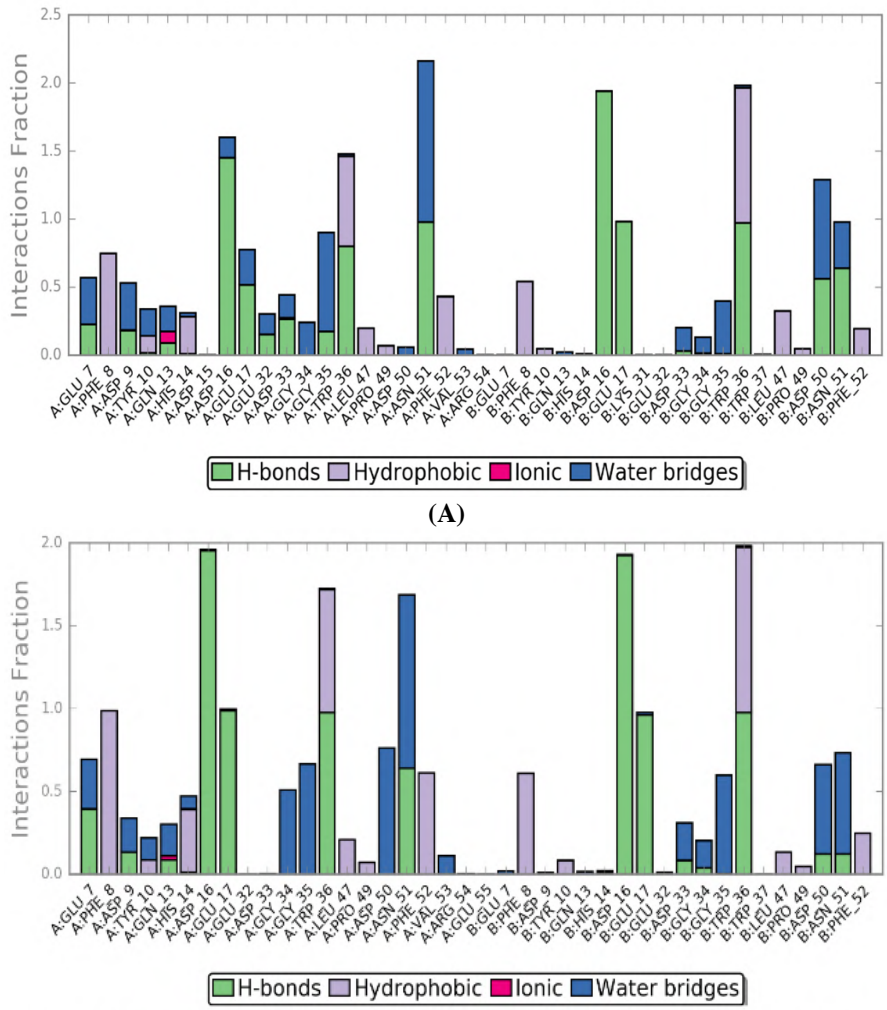


Plot 2.1. RMSD plots of the first **(A)** and the second **(B)** MD simulations of CIN85-Cbl-b peptide complex

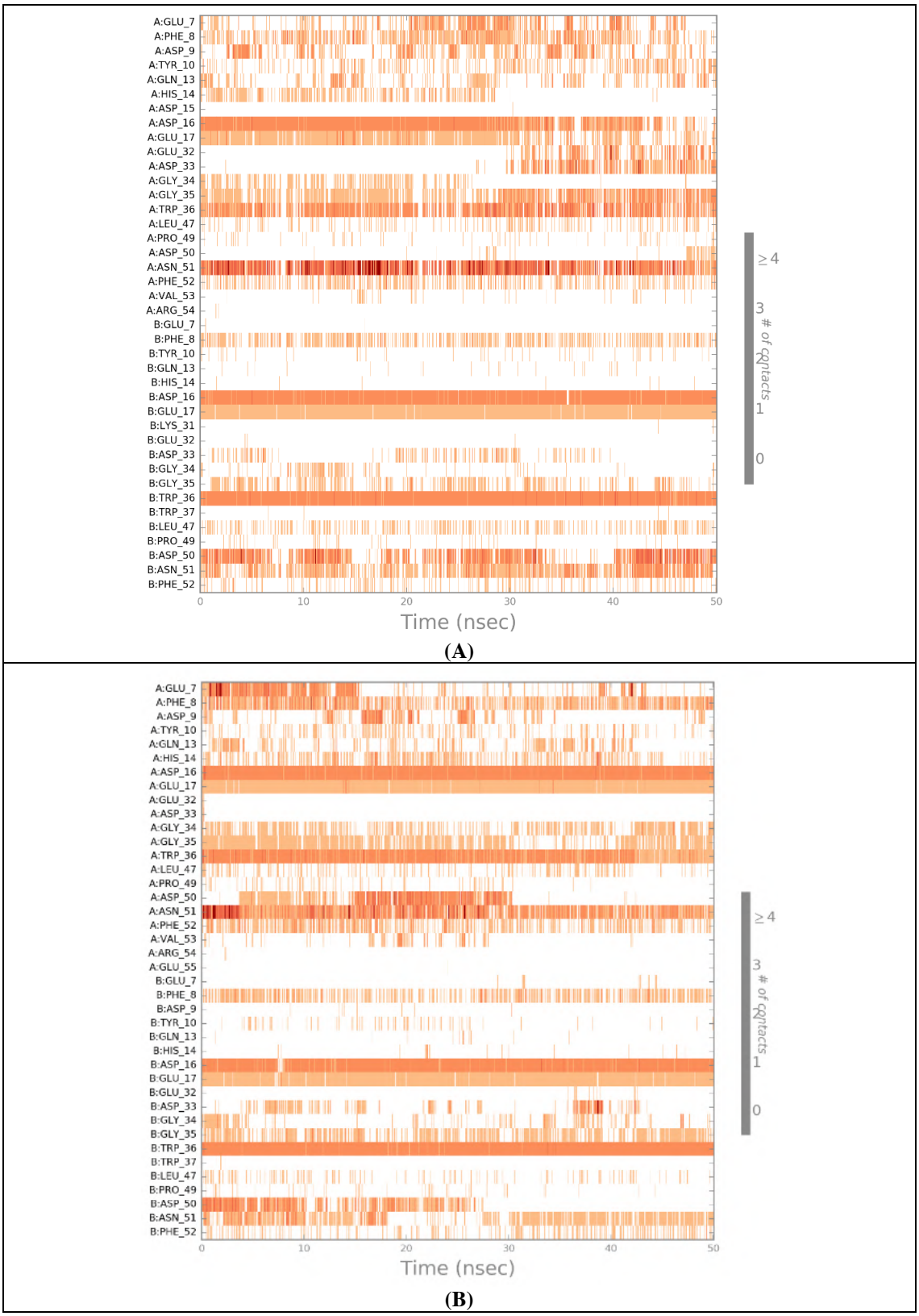
Table 2.2. Energy values monitored during the two MD trajectories of CIN85-Cbl-b peptide complex

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-55660.520	91.135	0.000	-67384.654	102.536	0.000
Potential energy (kcal/mol)	-68546.195	68.881	0.000	-82957.344	77.312	0.000

Both simulation outputs provided similar information and they retrieved mostly the same interactions already visible in the PDB structure, except for a new hydrophobic contact generated between Phe8 of CIN85 and Pro906 of Cbl-b, that was observed and very frequent in both simulations. This information was therefore used for the next steps of the workflow. The protein-ligand contacts plots for the two MD simulations are reported below (Plot 2.2A-B) together with the timeline representation of the protein-ligand contacts, that provide a measure to understand the frequency of occurrences of the interactions (Plot 2.3A-B). Finally, figure 2.10 shows the ligand interaction path including the percentage of occurrence per each during one of the two trajectories.



Plot 2.2. Protein-ligand contacts retrieved from the first **(A)** and the second **(B)** MD simulations of CIN85-Cbl-b peptide complex. The plots describe the intensity of protein interactions monitored throughout the simulations. The green bars represent the hydrogen bonds, the purple ones are for hydrophobic contacts, the pink bar is for ionic or polar interaction, and the blue ones show water bridges.



Plot 2.3. Timeline representation of the interactions and contacts established by CIN85 protein with Cbl-b peptide during the first **(A)** and the second **(B)** MD simulations.

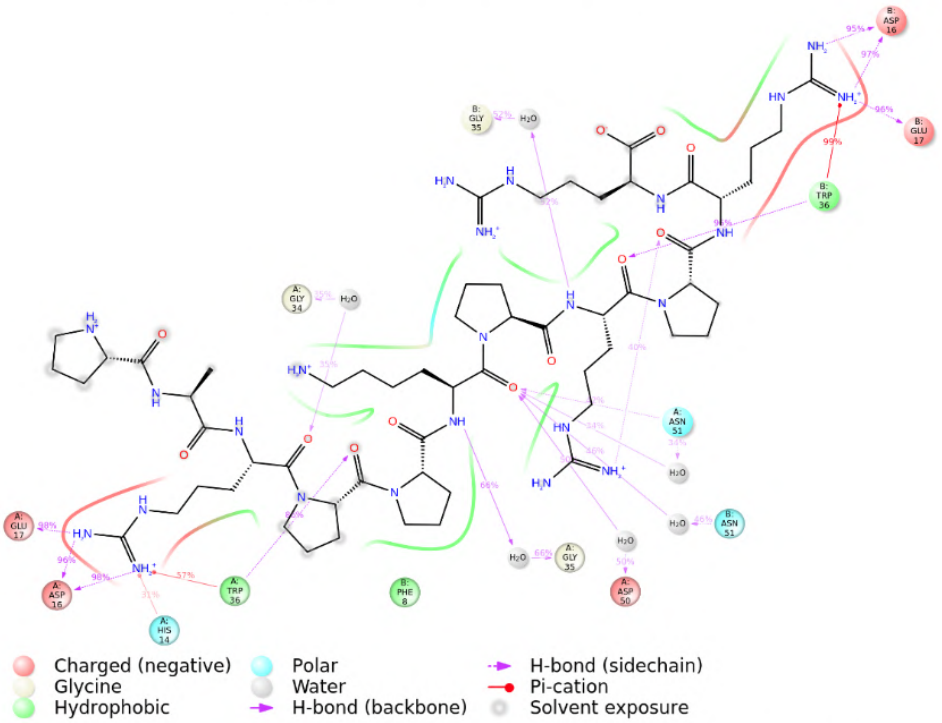


Figure 2.10. Cbl-b peptide ligand interactions with the two chains of the protein partner.

2.2.2 Protein-peptide docking of CIN85 SH3 domains and MUC1 peptide

The analysis of MD outputs provided key data for the next step of this work. As previously mentioned, this study was based on a hypothesis to be explored, *i.e.* MUC1 peptide could bind and share the same CIN85 interacting interface of Cbl-b peptide since experimental evidence demonstrated that CIN85 SH3 domains are able to bind MUC1 VNTR building a heterotrimeric complex [27]. Furthermore, both Cbl-b and MUC1 VNTR share similar proline-rich motifs PXXXP. This information is crucial to investigate the protein-protein interaction but does not provide information about specific interactions established between the three protein partners. Moreover, unfortunately, a PDB structure of CIN85-MUC1 complex is not currently available, for this reason, it was necessary to create an interaction model of this complex by running a protein-peptide docking of MUC1 VNTR to CIN85 dimer. Hence, the analysis of the PDB 2BZ8 complex and MD results were used as a benchmark to perform and validate the protein-peptide docking results.

For this purpose, the PDB structure of MUC1 hypoglycosylated VNTR peptide (GVTSAfPDT*RPAP, including a fluoroproline and a sugar moiety linked to Thr5, the 2-acetamido-2-deoxy-alpha-D-galactopyranose) in complex with two subunits of the anti-MUC1 antibody SM3 (PDB ID: 5OWP with resolution 1.85 Å) [29] was downloaded from the Protein Data Bank [30]. The protein partners of MUC1 were deleted and the peptide was prepared at pH 7.0 ± 2.0 by also mutating the fluoroproline of the peptide to a natural proline. The protein-peptide docking was performed in a standard precision mode using Glide. The calculations generated several protein-peptide combinations, where the docked peptide exhibiting the lowest docking score (-9.802 kcal/mol) was the one depicted in figure 2.11. By this binding mode MUC1 peptide maintained as much as

similar conformation to the starting one present in the PDB 5OWP. The other results showed the peptide in highly folded conformation quite different from the well-known natural PPII helix conformation, that is usually assumed by MUC1 VNTR, as reported in the literature [26]. Therefore, the other solutions were neglected and the first protein-peptide docked complex was processed to perform additional computational studies, *i.e.* Molecular Dynamics simulations.

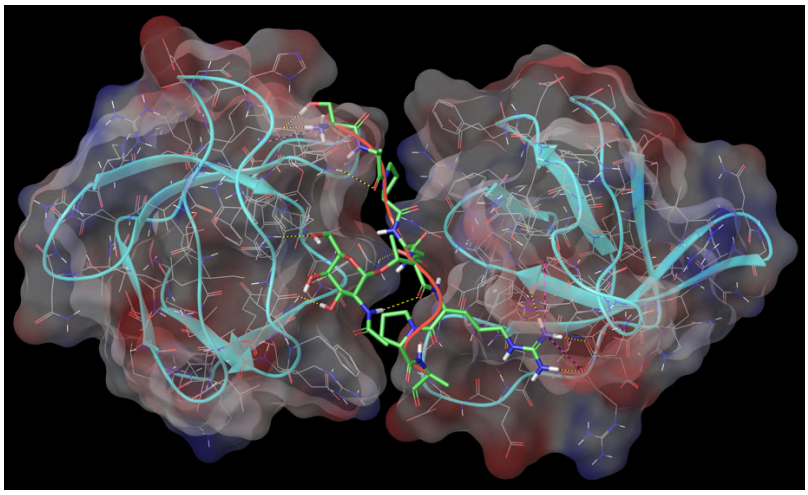


Figure 2.11. CIN85 SH3 domains (grey structures on the left and on the right) in complex with MUC1 VNTR peptide with the sequence GVTSAPDT*RPAP (red filament with green stick bonds).

Furthermore, most of the CIN85-MUC1 peptide interactions retrieved from this docking were in accordance with those identified from the PDB 2BZ8 between CIN85 SH3 domains and Cbl-b peptide, whereas the key residues of CIN85 were Asp16, Glu17 and Asn51 for the hydrogen bonds and Trp36 for both hydrophobic contacts and H-bonds. The amino acids Asp16 and Glu17 of an SH3 domain established hydrogen bonds with MUC1 Arg6, and from the other SH3 chain generated H-bonds with MUC1 Ser1 backbone. Trp36 of CIN85 created a salt bridge with MUC1 Arg6, Asn51 of CIN85 formed a hydrogen bond with MUC1 Asp4, and Phe52 established a hydrophobic contact with MUC1 Pro3, as shown in figure 2.12. In order to deeply explore these interactions and analyse their stability two Molecular Dynamics simulations were run, as described in the next section.

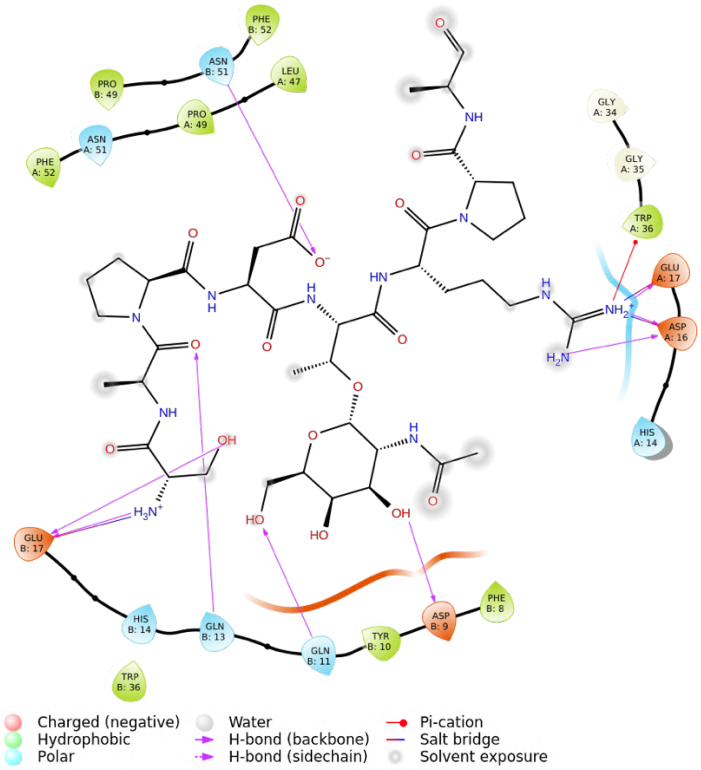
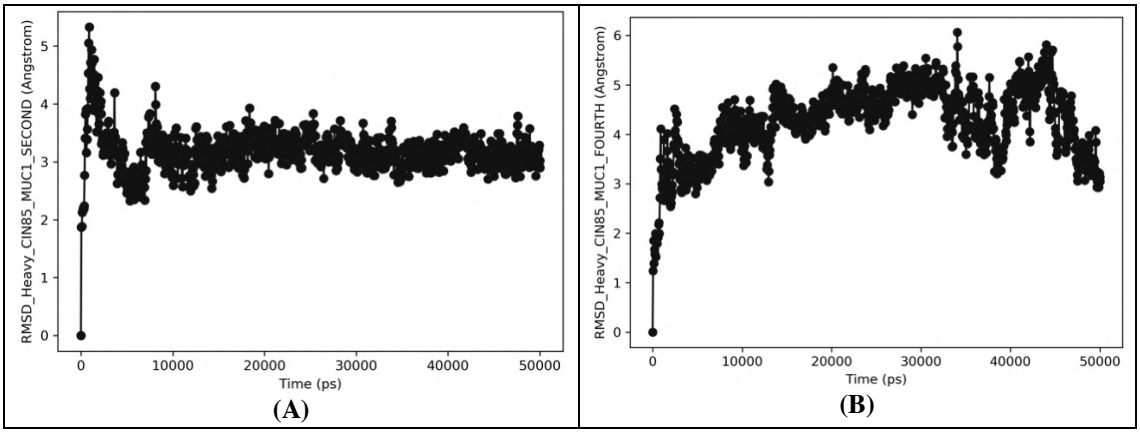


Figure 2.12. MUC1 VNTR peptide interactions with SH3 domains residues of CIN85 from first prioritised protein-peptide docked complex.

2.2.3 Molecular Dynamics simulations of CIN85-MUC1 complex

The heterotrimeric complex, SH3-SH3-MUC1 VNTR, selected from protein-peptide docking outputs was used to run two different short MD simulations of 50 ns, to minimise the system and extract the most frequent and stable interactions.

The stability of the complexes was checked by analysing the RMSD plots (Plot 2.4) and energy, temperature, pressure and volume were monitored during the entire trajectories (average energy values in table 2.3).

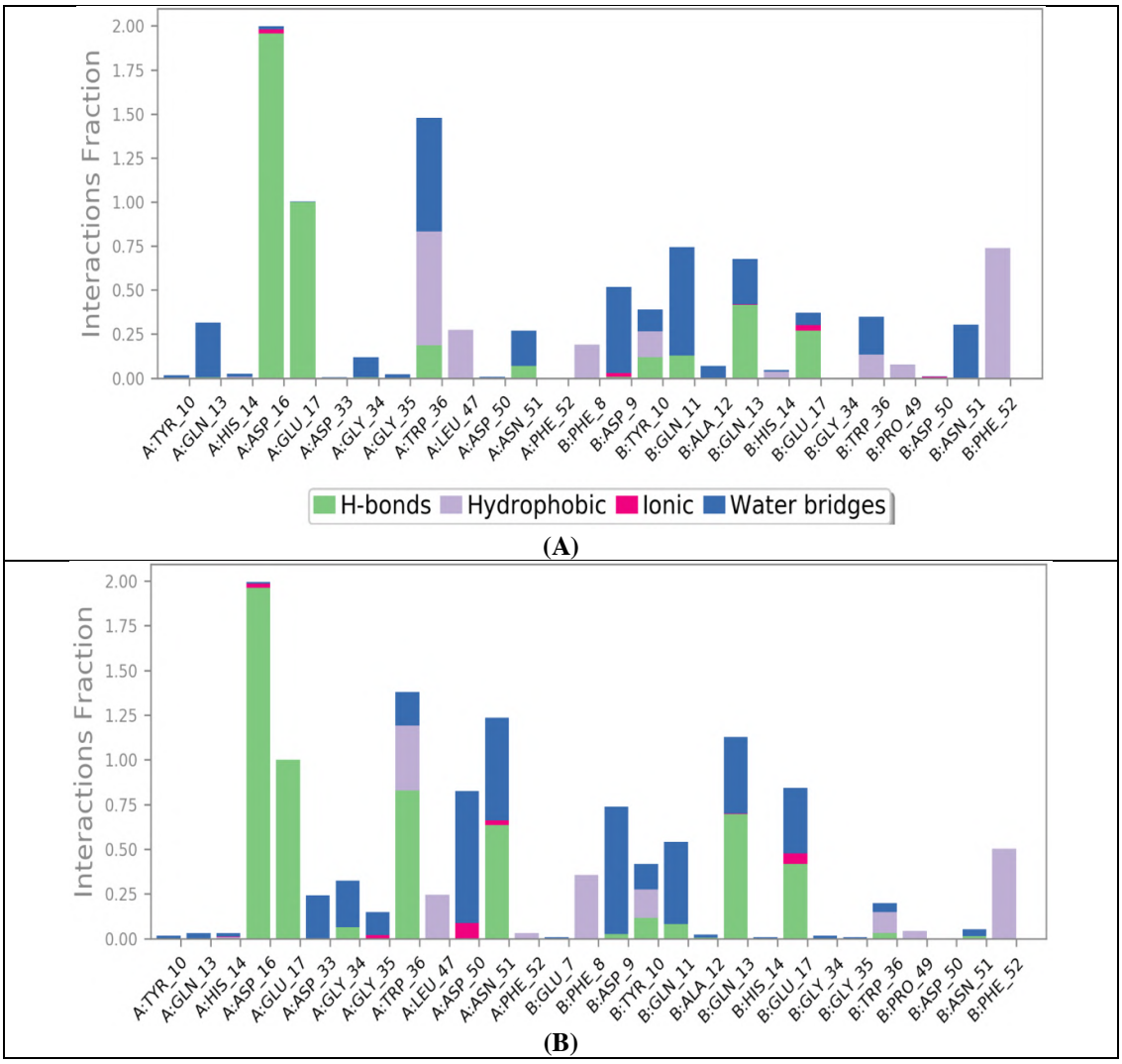


Plot 2.4. RMSD plots of the first **(A)** and the second **(B)** MD simulations of CIN85-MUC1 peptide complex

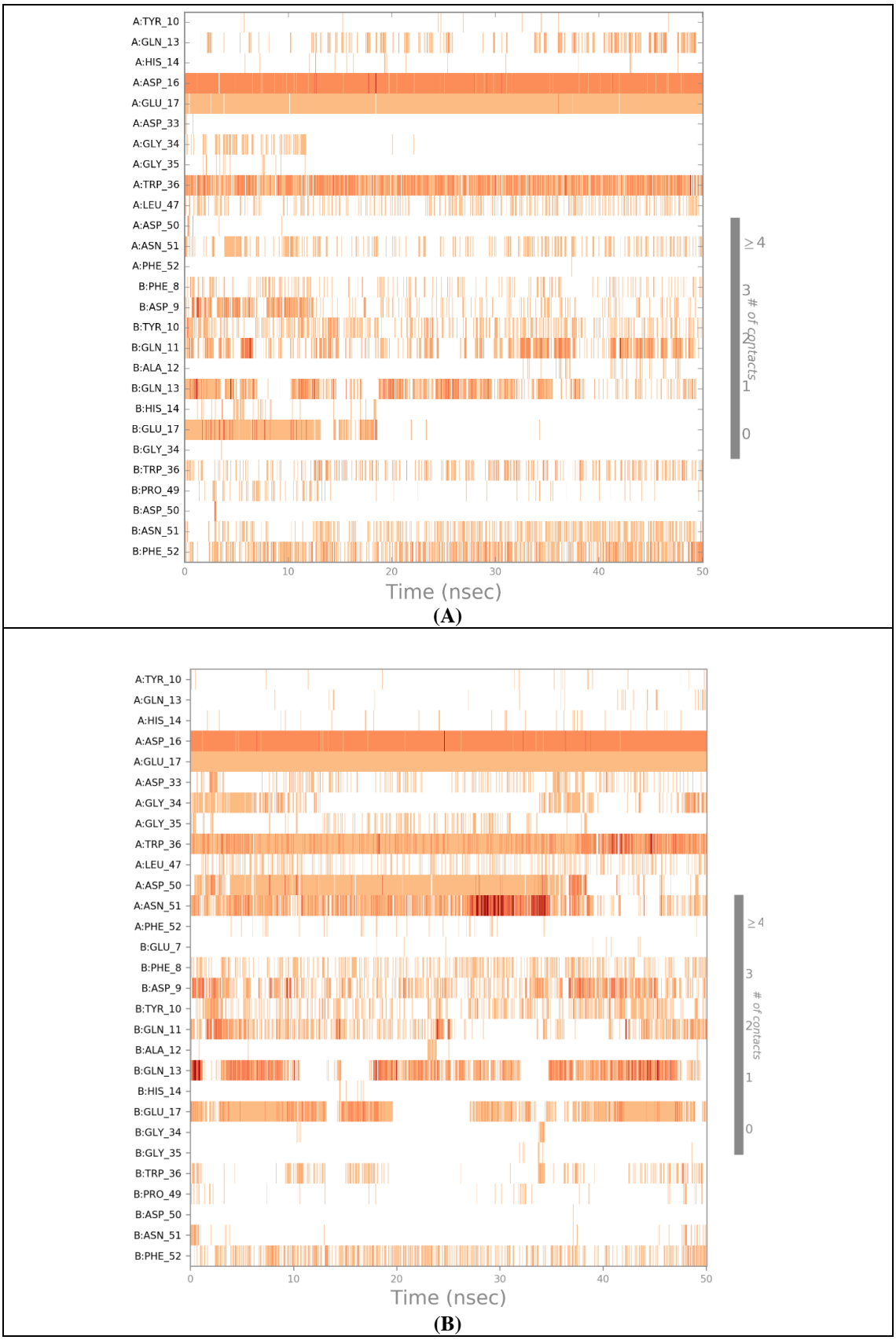
Table 2.3. Energy values monitored during the two MD trajectories of CIN85-MUC1 peptide complex

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-56873.047	93.095	0.000	-53980.081	90.896	-0.001
Potential energy (kcal/mol)	-69904.760	70.510	0.000	-66340.262	69.104	-0.001

Even for CIN85-MUC1 peptide complex the residues Asp16, Asp17 and Trp36 of CIN85 established the main interactions with MUC1 peptide, although not all the interactions with the key residues were observable as the most stable during the whole trajectories, such as contacts with Asn51 of CIN85, as visible in Plot 2.5A-B and Plot 2.6A-B.



Plot 2.5. Protein-ligand contacts retrieved from the first **(A)** and the second **(B)** MD simulations of CIN85-MUC1 complex. The plots describe the intensity of protein interactions monitored throughout the simulations. The green bars represent the hydrogen bonds, the purple ones are for hydrophobic contacts, the pink bar is for ionic or polar interaction, and the blue ones show water bridges.



Plot 2.6. Timeline representation of the interactions and contacts established by CIN85 protein with MUC1 peptide during the first **(A)** and the second **(B)** MD simulations.

Furthermore, for a better interaction analysis figure 2.13 shows the ligand and the contacts established with the protein partner including the related percentage of occurrences during one of the two trajectories.

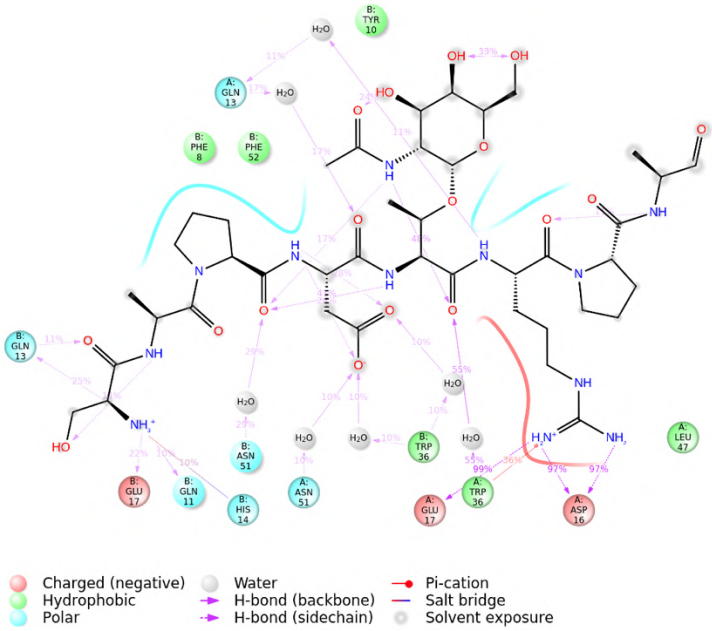


Figure 2.13. MUC1 peptide ligand interactions with the two chains of the protein partner during an MD trajectory.

All the above information from docking and MD simulations suggested that the hypothesis of MUC1 binding to CIN85 dimer could be likely, but it should be further investigated. These above-produced data were collected and used to guide the next step of this work together with the MD information retrieved from CIN85-Cbl-b peptide complex simulations, in order to address and apply a supervised structure-based consensus approach.

2.2.4 Pharmacophore map creation based on literature and MD data

All the information from the literature (mutational studies and PDB protein-protein complex structure) and MD simulations were collected and considered to guide the creation of a supervised pharmacophore map of the interactions between two CIN85 SH3 domains and Cbl-b peptide. The first step was the generation of the pharmacophore importing the protein-protein complex in LigandScout software, thus getting the 3D map illustrated in figure 2.14.

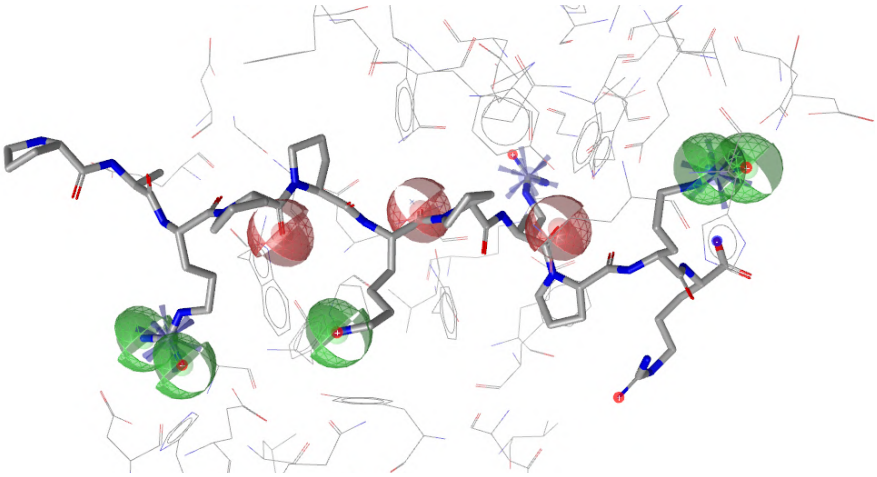
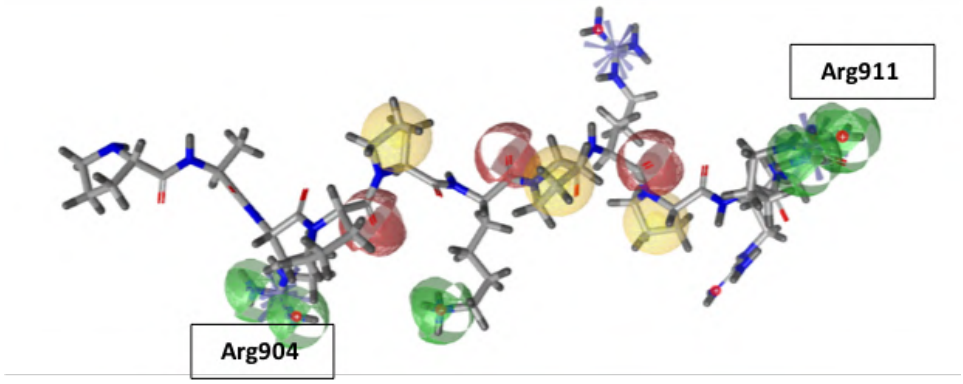


Figure 2.14. First pharmacophore map of the contacts between CIN85 SH3 domains and Cbl-b peptide from PDB 2BZ8.

This pharmacophore map did not include the hydrophobic features referring to the proline amino acids, Pro906, Pro908 and Pro910, that were identified as fundamental from experimental assays [26]. Therefore, these residues were selected and included in the map for the next screening, and the resulting pharmacophore is depicted in figure 2.15A. However, this 3D interaction map consisted of many features (13 features). Therefore, such a numerous pharmacophore map could fail in promisingly matching molecules, because small molecules might not be able to cover all these features and such a widespread interaction area. Hence, three pharmacophore features corresponding to Arg904 side chain group were deleted. This choice was done according to literature data, where mutational studies confirmed that Arg904 mutated to alanine does not significantly reduce the trimeric complex formation, and for this reason, the related features were even considered less important compared for example to the ones referring to Arg911. The resulting pharmacophore map is shown in figure 2.15B.



(A)

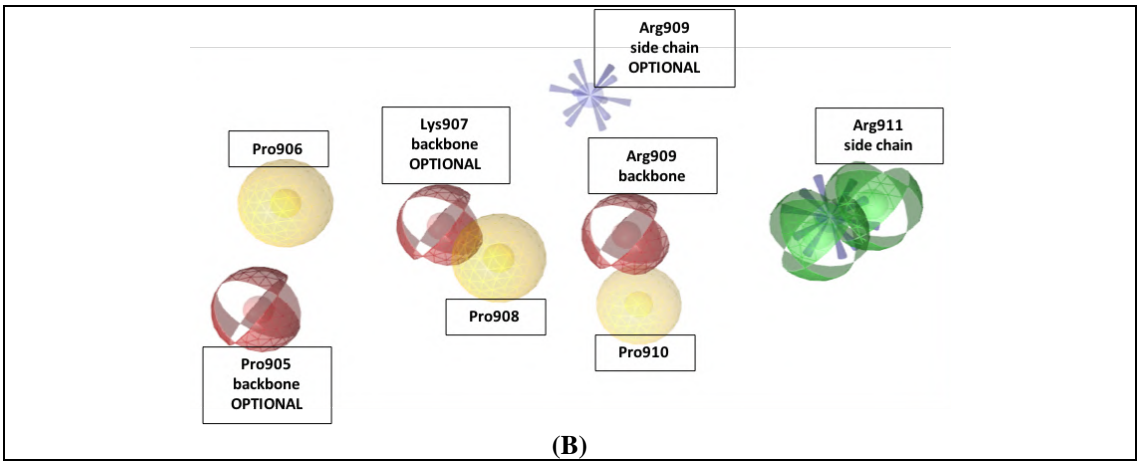


Figure 2.15. (A) Pharmacophore map updated including hydrophobic features referring to Pro906, Pro908 and Pro910 superimposed on the Cbl-b peptide. **(B)** Focus on modified pharmacophore map not including three interaction features corresponding to Arg904. Green spheres refer to hydrogen bond donors, red spheres correspond to hydrogen bond acceptors, yellow spheres are hydrophobic contacts, and blue spikes are positive ionizable features.

The last pharmacophore map was composed of ten features. However, the model was still too numerous for the pharmacophore screening, hence three pharmacophore features were marked as optional, *i.e.* the ones corresponding to Pro905 and Lys907 backbone carbonyl groups and Arg909 side chain. Then, this pharmacophore was used to perform a virtual screening including a dataset of about 110,000 PPI-targeted compounds belonging to several chemical providers, such as BioAscent and MolPort. The screening was performed allowing maximum 3 features to be omitted and the output provided 5,535 compounds matching the pharmacophore map, with a hit-rate of 5.3% compared to the initial compound dataset. Among these molecules, the first prioritised 500 were analysed through a visual check to select the most promising ones in terms of best matching the pharmacophore, especially for the features corresponding to Arg911, that from mutational studies was defined as the most important residue for protein-protein binding [28].

The selected compounds were 225, that were further processed to perform docking screening in order to apply a consensus approach, including two different computational techniques and select the best compounds for both approaches.

2.2.5 Docking screening of 225 best compounds from pharmacophore screening

From the pharmacophore screening, only 225 molecules from the initial PPI-targeted database were selected to be processed for a docking screening. For this purpose, the PDB 2BZ8 was used to create a docking grid, whereas the grid was defined selecting Cbl-b peptide. The molecule conformations were first generated using the Schrödinger tool and then docked on the receptor. All the initial compounds were retrieved from the screening and the analysis of docking results showed that most of them were able to establish the crucial interactions known from literature and MD simulations outputs. For this reason, all the compounds were considered for the next step of the workflow.

2.2.6 Cluster generation of the selected compounds

The 225 selected molecules were then filtered using PAINS and REOS structures of CANVAS [31, 32], in order to delete all those containing toxic groups or portions that could interfere with biological assays, respectively, thus reaching overall 183 compounds. Furthermore, in order to avoid to select compounds sharing the same chemotypes and avoid chemotype redundancy among the consensus compounds, a clustering process was computed identifying finally 120 compound centroids per each cluster. For this purpose, 2D radial fingerprints [31–33] were calculated for each molecule, as this fingerprint type accounted the ramified structures and therefore more radially distributed atoms within classical PPI-targeted compound structures [34]. These molecules were then grouped through a hierarchical clustering to identify and select overall 120 compound groups, that could be used as structures representative for their respective clusters (figure 2.16). Thus, the commercial availability of these molecules representative for the clusters was checked. Finally, only 56 compounds of these initial 120 were purchased and shipped to the laboratory of Dr Sandra Cascio at the Department of Immunology, University of Pittsburgh, to perform experimental assays, identify putative hit compounds among the 56 purchased molecules and validate the above described computational workflow.

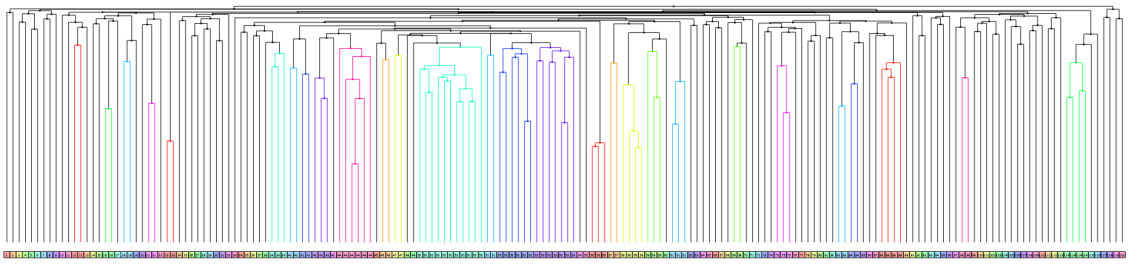
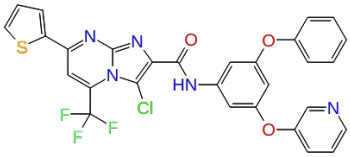
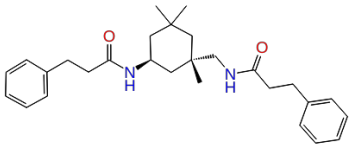
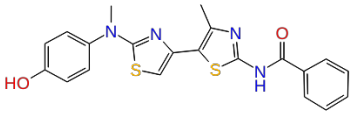
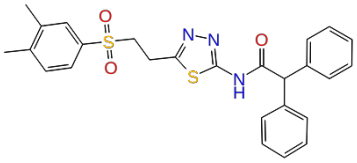
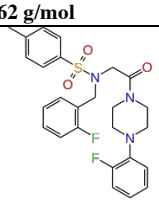
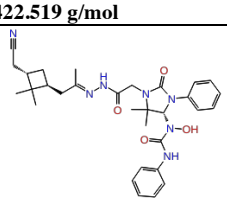
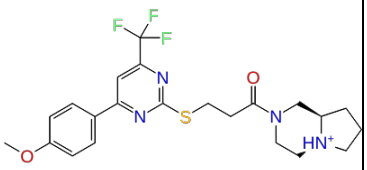
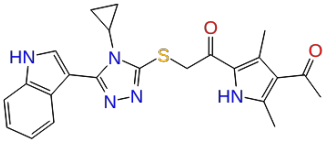
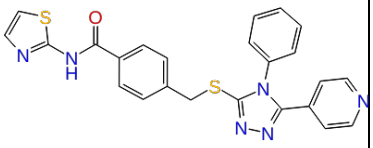
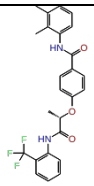
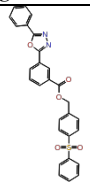
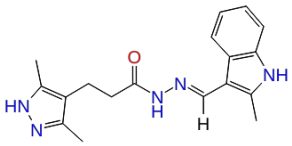
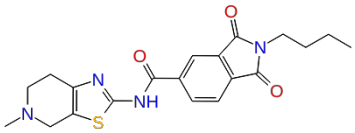
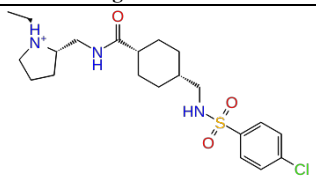
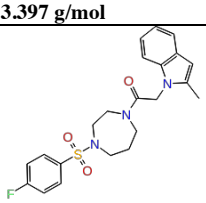
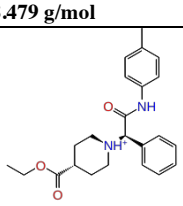
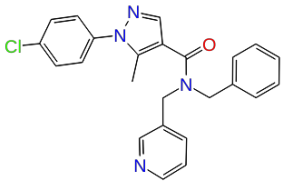
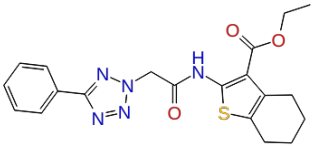
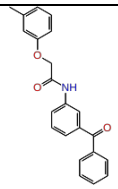
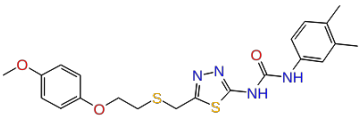
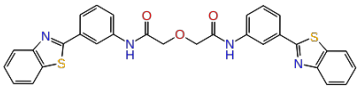
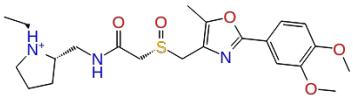
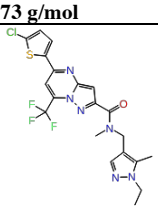
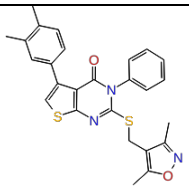
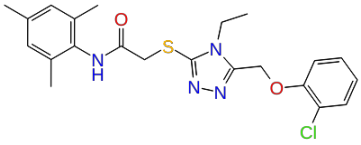
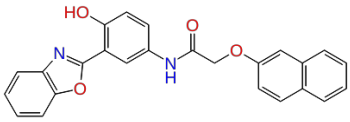
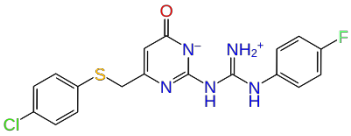
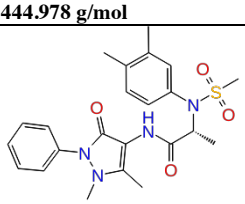
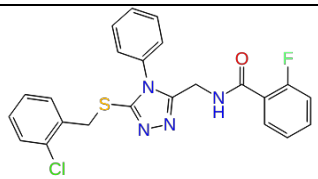
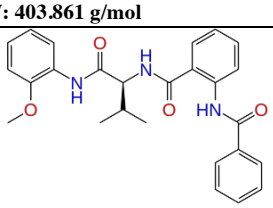
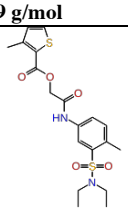
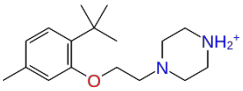
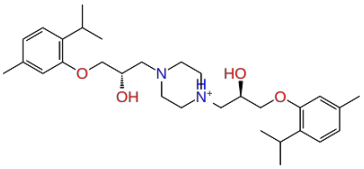
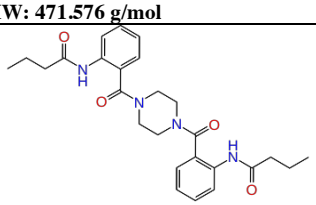


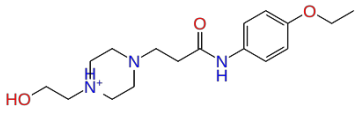
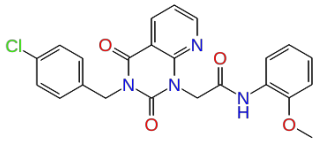
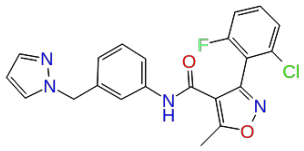
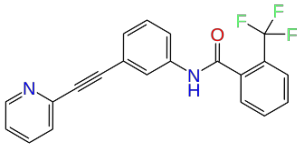
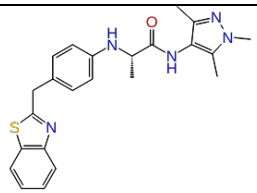
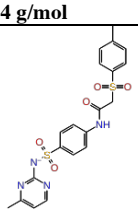
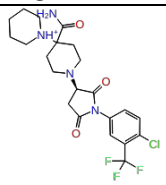
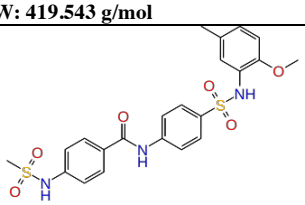
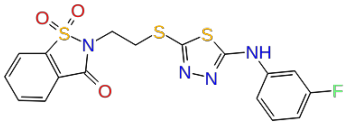
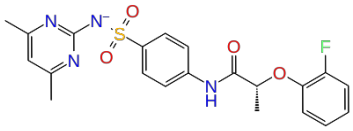
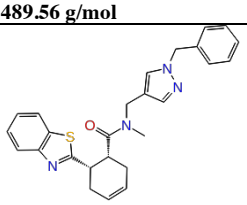
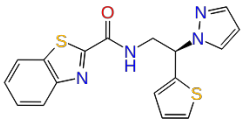
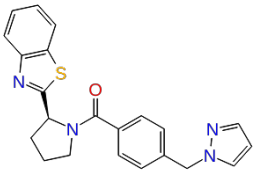
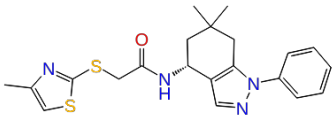
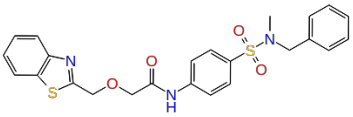
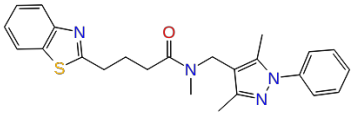
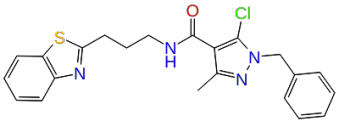
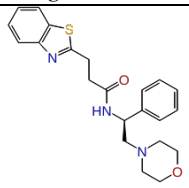
Figure 2.16. Hierarchical clusters representation of the selected compounds for CIN85 interacting surface.

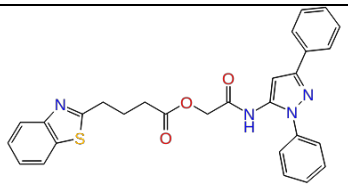
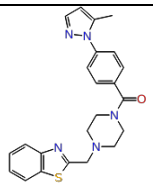
The structures of these compounds and the related properties are reported in table 2.4, except for two molecules that reported promising results for CIN85-Cbl-b co-precipitation assays and low micromolar and nanomolar IC_{50} values, as described in the next section. Some of these compounds present Rule of five violations, for example, owing to molecular weight often over 500 g/mol or logP over 5. However, even though the rule of five remains the benchmark for selecting drug-like compounds, nowadays most of the recently identified drugs do not meet these rules, and the drug-likeness rules are getting lesser and lesser restrictive [35]. Moreover, this project deals with a protein-protein interaction that is a well-known challenging target, where the binding surface is very widespread and shallow and usually a designed molecule that is intended to mimic a protein partner portion should cover a wide area of the target by violating, in this way, some Lipinski's rules [36]. For this reason, the physicochemical properties of the selected compounds were all considered acceptable and the molecules were purchased and tested.

Table 2.4. Consensus molecules purchased and assayed on CIN85 protein in the homodimeric form.

 **RIM410** **logP$_{o/w}$: 7.021** **PSA: 90.308** **MW: 607.993 g/mol**	 **RIM411** **logP$_{o/w}$: 5.524** **PSA: 72.663** **MW: 434.62 g/mol**	 **RIM412** **logP$_{o/w}$: 4.197** **PSA: 79.286** **MW: 422.519 g/mol**
 **RIM413** **logP$_{o/w}$: 4.893** **PSA: 99.763** **MW: 491.622 g/mol**	 **RIM414** **logP$_{o/w}$: 3.707** **PSA: 62.854** **MW: 499.574 g/mol**	 **RIM415** **logP$_{o/w}$: 4.209** **PSA: 150.049** **MW: 573.694 g/mol**
 **RIM416** **logP$_{o/w}$: 3.887** **PSA: 62.549** **MW: 466.52 g/mol**	 **RIM417** **logP$_{o/w}$: 4.875** **PSA: 108.931** **MW: 433.527 g/mol**	 **RIM418** **logP$_{o/w}$: 4.249** **PSA: 91.321** **MW: 470.566 g/mol**
 **RIM419** **logP$_{o/w}$: 5.663** **PSA: 71.175** **MW: 456.464 g/mol**	 **RIM420** **logP$_{o/w}$: 4.395** **PSA: 106.681** **MW: 496.536 g/mol**	 **RIM421** **logP$_{o/w}$: 3.531** **PSA: 93.217** **MW: 323.397 g/mol**
 **RIM422** **logP$_{o/w}$: 2.144** **PSA: 114.275** **MW: 398.479 g/mol**	 **RIM423** **logP$_{o/w}$: 2.213** **PSA: 86.434** **MW: 442.015 g/mol**	 **RIM424** **logP$_{o/w}$: 2.912** **PSA: 67.322** **MW: 429.508 g/mol**
 **RIM425** **logP$_{o/w}$: 4.295** **PSA: 67.589** **MW: 380.486 g/mol**	 **RIM426** **logP$_{o/w}$: 5.152** **PSA: 53.457** **MW: 416.909 g/mol**	 **RIM427** **logP$_{o/w}$: 3.421** **PSA: 112.226** **MW: 411.478 g/mol**

 **RIM428** **$logP_{o/w}$: 4.292** **PSA: 70.953** **MW: 345.397 g/mol**	 **RIM429** **$logP_{o/w}$: 4.671** **PSA: 91.4** **MW: 444.566 g/mol**	**Confidential structure** **RIM430** **$logP_{o/w}$: 4.759** **PSA: 96.099** **MW: 439.373 g/mol**
 **RIM431** **$logP_{o/w}$: 5.601** **PSA: 104.931** **MW: 550.649 g/mol**	 **RIM432** **$logP_{o/w}$: 1.81** **PSA: 93.684** **MW: 449.564 g/mol**	 **RIM433** **$logP_{o/w}$: 5.768** **PSA: 71.669** **MW: 482.91 g/mol**
 **RIM434** **$logP_{o/w}$: 5.887** **PSA: 65.202** **MW: 473.607 g/mol**	 **RIM435** **$logP_{o/w}$: 5.913** **PSA: 73.202** **MW: 444.978 g/mol**	 **RIM436** **$logP_{o/w}$: 4.573** **PSA: 85.783** **MW: 410.428 g/mol**
 **RIM437** **$logP_{o/w}$: 3.51** **PSA: 102.848** **MW: 403.861 g/mol**	 **RIM438** **$logP_{o/w}$: 3.346** **PSA: 100.736** **MW: 456.559 g/mol**	 **RIM439** **$logP_{o/w}$: 5.638** **PSA: 65.72** **MW: 452.933 g/mol**
 **RIM440** **$logP_{o/w}$: 5.549** **PSA: 103.287** **MW: 445.517 g/mol**	 **RIM441** **$logP_{o/w}$: 2.956** **PSA: 106.478** **MW: 424.529 g/mol**	**Confidential structure** **RIM442** **$logP_{o/w}$: 4.709** **PSA: 87.787** **MW: 471.576 g/mol**
 **RIM443** **$logP_{o/w}$: 2.91** **PSA: 22.006** **MW: 276.421 g/mol**	 **RIM444** **$logP_{o/w}$: 5.142** **PSA: 63.301** **MW: 498.704 g/mol**	 **RIM445** **$logP_{o/w}$: 4.429** **PSA: 123.57** **MW: 464.563 g/mol**

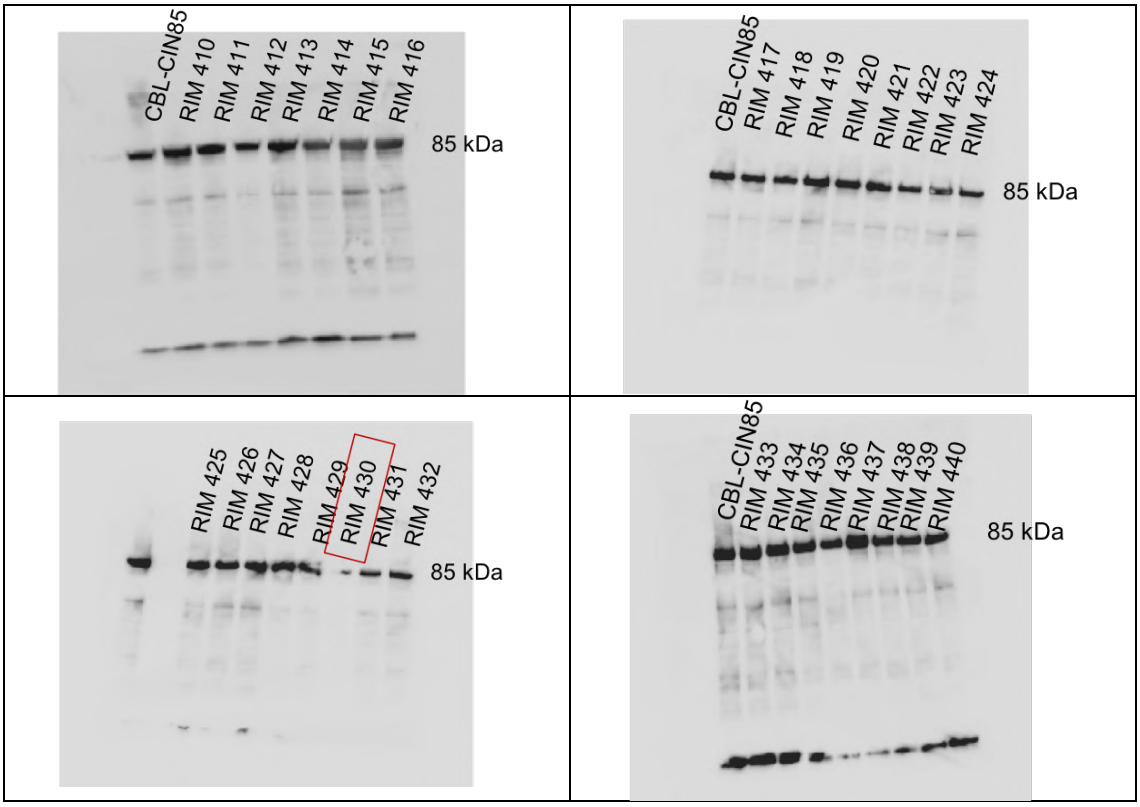
 **RIM446** **logP$_{o/w}$: 1.066** **PSA: 77.496** **MW: 321.419 g/mol**	 **RIM447** **logP$_{o/w}$: 4.394** **PSA: 103.726** **MW: 450.88 g/mol**	 **RIM448** **logP$_{o/w}$: 4.792** **PSA: 74.063** **MW: 410.834 g/mol**
 **RIM449** **logP$_{o/w}$: 5.388** **PSA: 43.574** **MW: 366.342 g/mol**	 **RIM450** **logP$_{o/w}$: 5.23** **PSA: 74.842** **MW: 419.543 g/mol**	 **RIM451** **logP$_{o/w}$: 1.726** **PSA: 147.265** **MW: 460.522 g/mol**
 **RIM452** **logP$_{o/w}$: 2.13** **PSA: 110.083** **MW: 486.921 g/mol**	 **RIM453** **logP$_{o/w}$: 1.835** **PSA: 143.054** **MW: 489.56 g/mol**	 **RIM454** **logP$_{o/w}$: 3.275** **PSA: 108.714** **MW: 436.493 g/mol**
 **RIM455** **logP$_{o/w}$: 3.303** **PSA: 110.802** **MW: 444.48 g/mol**	 **RIM456** **logP$_{o/w}$: 5.358** **PSA: 49.576** **MW: 442.578 g/mol**	 **RIM457** **logP$_{o/w}$: 4.12** **PSA: 62.385** **MW: 354.444 g/mol**
 **RIM458** **logP$_{o/w}$: 4.692** **PSA: 56.519** **MW: 388.486 g/mol**	 **RIM459** **logP$_{o/w}$: 5.012** **PSA: 63.003** **MW: 412.567 g/mol**	 **RIM460** **logP$_{o/w}$: 3.948** **PSA: 96.641** **MW: 481.583 g/mol**
 **RIM461** **logP$_{o/w}$: 5.138** **PSA: 56.909** **MW: 418.556 g/mol**	 **RIM462** **logP$_{o/w}$: 5.786** **PSA: 62.705** **MW: 424.947 g/mol**	 **RIM463** **logP$_{o/w}$: 2.986** **PSA: 60.576** **MW: 395.518 g/mol**

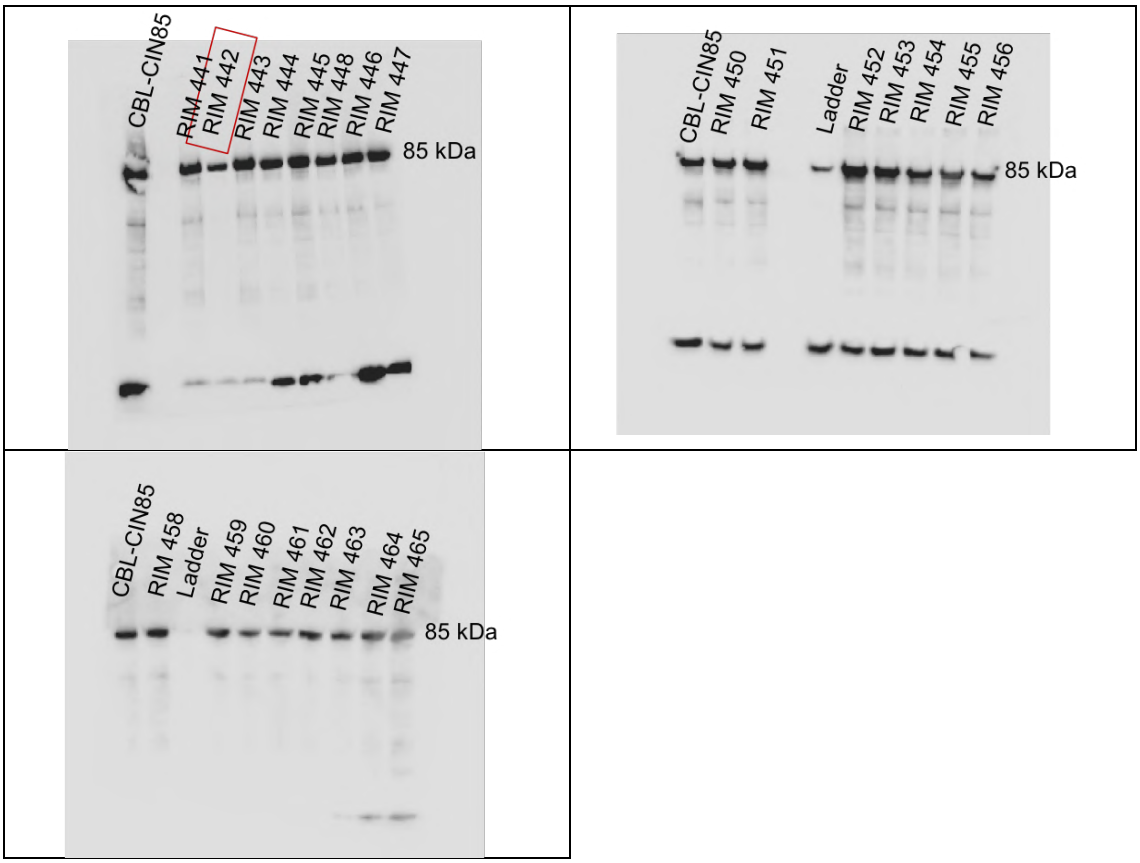
 **RIM464** **logP$_{o/w}$: 6.275** **PSA: 100.161** **MW: 496.583 g/mol**	 **RIM465** **logP$_{o/w}$: 3.875** **PSA: 62.068** **MW: 417.528 g/mol**

2.2.7 Experimental assays results

The experimental assays of the 56 purchased compounds were conducted by Dr Sandra Cascio and her research group at the Department of Immunology of the University of Pittsburgh. The performed assays used semi-quantitative methods, meaning that they were surely able to provide reliable data about the compound ability to inhibit CIN85-Cbl-b and CIN85-MUC1 complexes, but they did not provide a quantification of the biological activity. For the pull-down assays, the inhibition of CIN85-Cbl-b complex formation was investigating, by detecting the co-precipitation of the complex using SDS phage methods to check the intensity of the bands including each compound with a concentration of 10 μM. The related outputs are reported below in table 2.5.

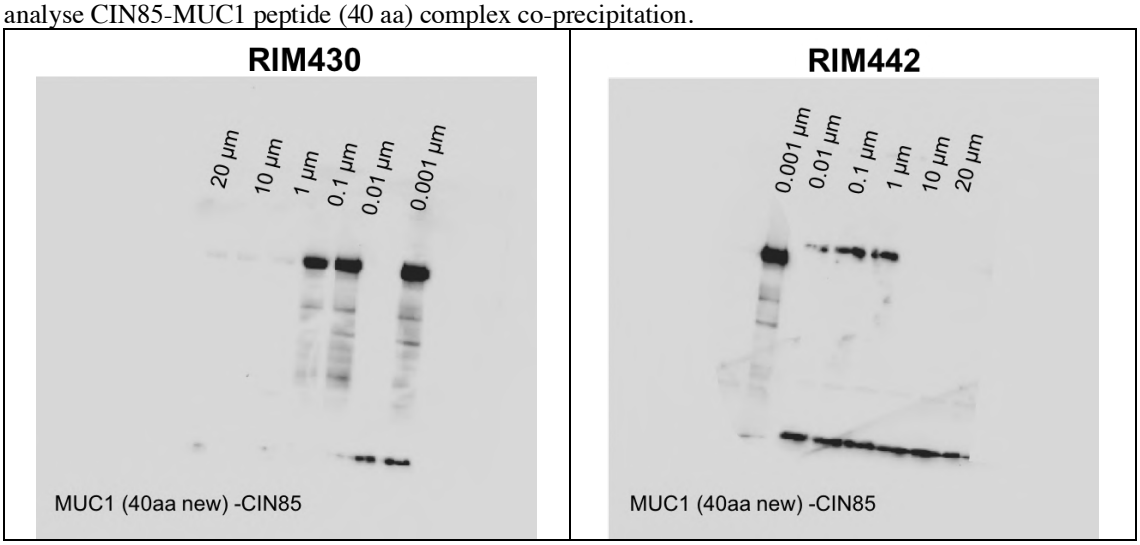
Table 2.5. Pull-down assays to detect the co-precipitation of CIN85-Cbl-b complex in presence of the selected compounds. Compounds squared in red showed the best results in terms of complex formation inhibition.

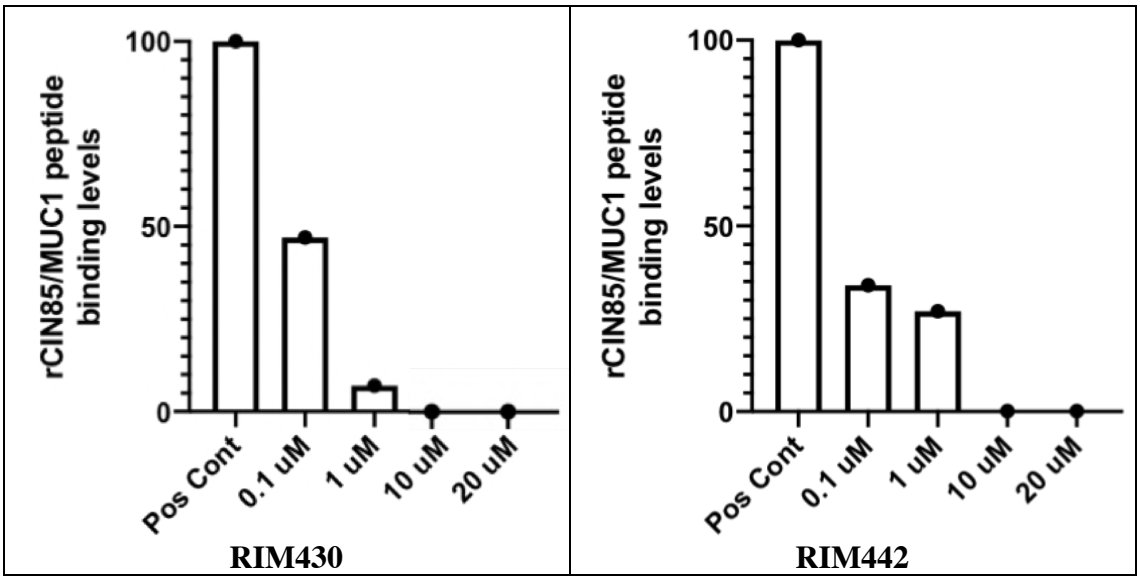




As can be seen, compounds RIM430 and RIM442 showed the best results. They were able to inhibit the protein-protein complex formation with higher intensity compared to the other molecules. For this reason, these two chemical entities were further processed in order to get IC_{50} values to explore if these molecules were also able to inhibit the formation of CIN85-MUC1 complex. For this purpose, a 40-mer MUC1 peptide was used and interestingly both molecules showed a good inhibitory capacity thus provided for RIM430 nanomolar IC_{50} value and for RIM442 low micromolar IC_{50} value (tables 2.6).

Table 2.6. Pull-down assay results showing IC_{50} values of compounds RIM430 and RIM442 in order to analyse CIN85-MUC1 peptide (40 aa) complex co-precipitation.





2.3 Methods

2.3.1 Preparation of PDB structures

The PDB structures used in this work were first prepared and optimised by using "Protein preparation wizard" [37] tool of Schrödinger suite (release 2016-4). For this purpose, the bond orders for untemplated residues and known HET groups were assigned and hydrogens were added. Bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds were created. Water molecules beyond 5 Å from HET groups were deleted. For ligands, cofactors and metals het states were generated at pH 7.0 ± 2.0 using Epik [38]. Finally, H-bonds were optimised by using PROPKA [39] at pH 7.0.

2.3.2 MD simulations of CIN85 in complex with MUC1 and Cbl-b

MD simulations were run twice for the two systems, CIN85-Cbl-b peptide and CIN85-MUC1 peptide complexes by using Schrödinger suite. The systems were first tuned through "System builder" tool. The solvent model TIP3P [40] and the orthorhombic box shape were selected, the box side distances were set 12 Å and the system was neutralized by adding Na^+ ions. Then these systems were used to run MD calculations [41] of 50 ns per each trajectory. The number of atoms, pressure and temperature were maintained constant, whereas pressure was set 1.01325 bar and temperature 300.0 K. Finally, the model systems were relaxed before simulation, and the force field was set as OPLS3 [42].

2.3.3 Ligand preparation for docking screening

In order to perform docking and pharmacophore screening, BioAscent and MolPort PPI-targeted compound libraries were considered including over 110,000 molecules. The virtual libraries were filtered through KNIME platform [43] using the SMART alerts, in order to delete those compounds containing carcinogenic, mutagenic, chelating, reactive, unstable, toxic and skin sensitising groups [44]. The resulting compounds were prepared using "LigPrep" tool of Schrödinger suite. The selected force field was OPLS3 [42] and the protonation states were generated at pH 7.4 ± 0.2 using Epik [38]. The molecules were desalted and tautomers were generated retaining compound specific chirality. Finally, no more than 32 different conformations were generated per ligand.

2.3.4 Ligands preparation for pharmacophore screening

All the compounds already prepared by using "LigPrep" tool were further processed through "Create screening database" tool of LigandScout software (version 4.3 - released by Inte:Ligand GmbH) [45–48], specifying "iCon Best" [49] as conformer generation type to create high-quality ligand conformations. All the other settings were applied as default.

2.3.5 Pharmacophore map creation and screening

The pharmacophore map of CIN85 in complex with Cbl-b peptide was generated by importing PDB 2BZ8 into the structure-based perspective of LigandScout. Then the pharmacophore map was created on Cbl-b peptide resulting in 11 features (figure 2.14). Four of these were deleted because considered not crucial for protein-protein binding, *i.e.* the pharmacophore features referring to Arg904 and Lys907. On the other hand, three more hydrophobic features were included on Pro906, Pro908 and Pro910, and three pharmacophore features were marked as optional, thus getting a pharmacophore map composed by 10 features (figure 2.15B). The resulting pharmacophore map was used to perform screening of MolPort and BioAscent PPI-targeted database. For this purpose, the pharmacophore-fit was set as scoring function, for the screening mode, all query features must be matched except for maximum three pharmacophore features that could be omitted. Finally, for the retrieval mode, the best matching conformations were retained. Pharmacophore screening was performed and the outputs were collected and analysed.

2.3.6 Receptor grids generation of CIN85 and docking screenings

In order to perform peptide and ligand docking, two grids were generated on PDB 2BZ8 including CIN85 in complex with Cbl-b peptide. The binding region was defined by selecting Cbl-b peptide and in detail for the peptide docking, the option to create a grid suitable for peptide docking was flagged. The VdW radii scaling factor for non-polar atoms were set by 1.0 with partial charge cut-off 0.25. For both grids, the applied force field was OPLS3 [42]. Then, the docking screenings were performed by using "ligand

docking" tool of Schrödinger suite [50, 51]. The selected protocol was standard precision and the selected ligand sampling method was flexible. Finally, the VdW radii scaling factor for non-polar atoms was set 0.8 with partial charge cut-off 0.15. All the other settings were maintained as default. Docking screening was run and results were compared with pharmacophore outputs.

2.3.7 Cluster generation of the selected compounds

The overall 183 consensus molecules were clustered in order to reduce chemotype redundancy. For this purpose, these compounds were imported in CANVAS software [31, 32] released by Schrödinger and the "Binary fingerprints from structures" tool was used. The fingerprint type was chosen as radial (ECFP) [33] and the "Hierarchical clustering" tool was run by applying Tanimoto similarity as metric and the cluster linkage method as average.

2.4 Conclusions

The workflow herein reported describes the steps involving the performed computational techniques to rationally design potential inhibitors of MUC1-CIN85 interaction. This work has demonstrated to be effective, as it led to the identification of compounds that were able first to inhibit CIN85-Cbl-b interaction at a concentration of 10 μM and then CIN85-MUC1 interaction reporting an inhibiting activity falling in a nanomolar range. Therefore, twenty close structural analogues of RIM430 and RIM442 were selected and purchased. These compounds will be sent to the Department of Immunology of the University of Pittsburgh, where Dr Sandra Cascio and her research group will conduct further biological assays in order to confirm the preliminary hit compounds and guide next structure-activity relationship process.

These findings should provide crucial information about unveiling the contacts between MUC1 and CIN85 proteins and identifying unprecedently found inhibitors of this interaction. Finally, the effective inhibiting activity of these compounds will allow to tackle the associated metastatic process and improve the pathological conditions of patients affected by cancer diseases.

This work has allowed to prepare a research article submitted to a scientific peer-reviewed journal.

References – Chapter Two

1. Cancer incidence - WHO. https://www.who.int/cancer/PRGlobocanFinal.pdf. Accessed 20 May 2020
2. Porquet N, Gout S, Huot J (2010) The Metastatic Process: An Overview. pp 1–31
3. Cascio S, Farkas AM, Hughey RP, Finn OJ (2013) Altered glycosylation of MUC1 influences its association with CIN85: the role of this novel complex in cancer cell invasion and migration. Oncotarget 4:. https://doi.org/10.18632/oncotarget.1265
4. Hattrup CL, Gendler SJ (2008) Structure and Function of the Cell Surface (Tethered) Mucins. Annu Rev Physiol 70:431–457. https://doi.org/10.1146/annurev.physiol.70.113006.100659
5. Chang J-F, Zhao H-L, Phillips J, Greenburg G (2000) The Epithelial Mucin, MUC1, Is Expressed on Resting T Lymphocytes and Can Function as a Negative Regulator of T Cell Activation. Cell Immunol 201:83–88. https://doi.org/10.1006/cimm.2000.1643
6. Gendler SJ, Spicer AP (1995) Epithelial Mucin Genes. Annu Rev Physiol 57:607–634. https://doi.org/10.1146/annurev.ph.57.030195.003135
7. Yolken RH, Peterson JA, Vonderfecht SL, et al (1992) Human milk mucin inhibits rotavirus replication and prevents experimental gastroenteritis. J Clin Invest 90:1984–1991. https://doi.org/10.1172/JCI116078
8. Kufe DW (2009) Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer 9:874–885. https://doi.org/10.1038/nrc2761
9. Levitin F, Stern O, Weiss M, et al (2005) The MUC1 SEA Module Is a Self-cleaving Domain. J Biol Chem 280:33374–33386. https://doi.org/10.1074/jbc.M506047200
10. Hanisch F-G, Muller S (2000) MUC1: the polymorphic appearance of a human mucin. Glycobiology 10:439–449. https://doi.org/10.1093/glycob/10.5.439
11. Nath S, Mukherjee P (2014) MUC1: a multifaceted oncoprotein with a key role in cancer progression. Trends Mol Med 20:332–342. https://doi.org/10.1016/j.molmed.2014.02.007
12. Nabavinia MS, Gholoobi A, Charbgoo F, et al (2017) Anti-MUC1 aptamer: A potential opportunity for cancer treatment. Med Res Rev 37:1518–1539. https://doi.org/10.1002/med.21462
13. Whitehouse C, Burchell J, Gschmeissner S, et al (1997) A Transfected Sialyltransferase That Is Elevated in Breast Cancer and Localizes to the medial/trans-Golgi Apparatus Inhibits the Development of core-2–based O-Glycans. J Cell Biol 137:1229–1241. https://doi.org/10.1083/jcb.137.6.1229
14. Brockhausen I, Yang J-M, Burchell J, et al (1995) Mechanisms Underlying Aberrant Glycosylation of MUC1 Mucin in Breast Cancer Cells. Eur J Biochem 233:607–617. https://doi.org/10.1111/j.1432-1033.1995.607_2.x
15. Kowanetz K, Szymkiewicz I, Haglund K, et al (2003) Identification of a Novel Proline-Arginine Motif Involved in CIN85-dependent Clustering of Cbl and Down-regulation of Epidermal Growth Factor Receptors. J Biol Chem 278:39735–39746. https://doi.org/10.1074/jbc.M304541200
16. Havrylov S, Jolanta Redowicz M, Buchman VL (2010) Emerging Roles of Ruk/CIN85 in Vesicle-Mediated Transport, Adhesion, Migration and Malignancy. Traffic 11:721–731. https://doi.org/10.1111/j.1600-0854.2010.01061.x
17. Dikic I (2002) CIN85/CMS family of adaptor molecules. FEBS Lett 529:110–115. https://doi.org/10.1016/S0014-5793(02)03188-5
18. Li Q, Yang W, Wang Y, Liu W (2018) Biochemical and Structural Studies of the Interaction between ARAP1 and CIN85. Biochemistry 57:2132–2139. https://doi.org/10.1021/acs.biochem.8b00057
19. Cascio S, Finn O (2016) Intra- and Extra-Cellular Events Related to Altered Glycosylation of MUC1 Promote Chronic Inflammation, Tumor Progression, Invasion, and Metastasis. Biomolecules 6:39. https://doi.org/10.3390/biom6040039
20. Schmidt MHH (2003) SETA/CIN85/Ruk and its binding partner AIP1 associate with diverse cytoskeletal elements, including FAKs, and modulate cell adhesion. J Cell Sci 116:2845–2855. https://doi.org/10.1242/jcs.00522
21. Johnson RI, Seppa MJ, Cagan RL (2008) The Drosophila CD2AP/CIN85 orthologue Cindr regulates junctions and cytoskeleton dynamics during tissue patterning. J Cell Biol 180:1191–1204. https://doi.org/10.1083/jcb.200706108
22. Lynch DK, Winata SC, Lyons RJ, et al (2003) A Cortactin-CD2-associated Protein (CD2AP)

Complex Provides a Novel Link between Epidermal Growth Factor Receptor Endocytosis and the Actin Cytoskeleton. J Biol Chem 278:21805–21813. https://doi.org/10.1074/jbc.M211407200
23. Gout I, Middleton G, Adu J, et al (2000) Negative regulation of PI 3-kinase by Ruk, a novel adaptor protein. EMBO J 19:4015–4025. https://doi.org/10.1093/emboj/19.15.4015
24. Nam J-M, Onodera Y, Mazaki Y, et al (2007) CIN85, a Cbl-interacting protein, is a component of AMAP1-mediated breast cancer invasion machinery. EMBO J 26:647–656. https://doi.org/10.1038/sj.emboj.7601534
25. Ababou A, Pfuhl M, Ladbury JE (2009) Novel Insights into the Mechanisms of CIN85 SH3 Domains Binding to Cbl Proteins: Solution-Based Investigations and In Vivo Implications. J Mol Biol 387:1120–1136. https://doi.org/10.1016/j.jmb.2009.02.061
26. Ceregido MA, Garcia-Pino A, Ortega-Roldan JL, et al (2013) Multimeric and differential binding of CIN85/CD2AP with two atypical proline-rich sequences from CD2 and Cbl-b*. FEBS J 280:3399–3415. https://doi.org/10.1111/febs.12333
27. Cascio S, Sciurba J, Hughey R, et al (2014) Abstract 3151: Muc1/Cin85 complex is a new molecular target for control of cancer invasion and metastasis. In: Tumor Biology. American Association for Cancer Research, pp 3151–3151
28. Jozic D, Cárdenes N, Deribe YL, et al (2005) Cbl promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol 12:972–979. https://doi.org/10.1038/nsmb1000
29. Somovilla VJ, Bermejo IA, Albuquerque IS, et al (2017) The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer. J Am Chem Soc 139:18255–18261. https://doi.org/10.1021/jacs.7b09447
30. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
31. Duan J, Dixon SL, Lowrie JF, Sherman W (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. J Mol Graph Model 29:157–170. https://doi.org/10.1016/j.jmgm.2010.05.008
32. Sastry M, Lowrie JF, Dixon SL, Sherman W (2010) Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. J Chem Inf Model 50:771–784. https://doi.org/10.1021/ci100062n
33. Glem RC, Bender A, Arnby CH, et al (2006) Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. IDrugs 9:199–204
34. Villoutreix BO, Labbé CM, Lagorce D, et al (2012) A leap into the chemical space of protein-protein interaction inhibitors. Curr Pharm Des 18:4648–67. https://doi.org/10.2174/138161212802651571
35. Benet LZ, Hosey CM, Ursu O, Oprea TI (2016) BDDCS, the Rule of 5 and drugability. Adv Drug Deliv Rev 101:89–98. https://doi.org/10.1016/j.addr.2016.05.007
36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/S0169-409X(00)00129-0
37. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221–234. https://doi.org/10.1007/s10822-013-9644-8
38. Shelley JC, Cholleti A, Frye LL, et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691. https://doi.org/10.1007/s10822-007-9133-z
39. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
40. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
41. (2006) SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Association for Computing Machinery, New York, NY, USA
42. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864
43. Berthold MR, Cebron N, Dill F, et al (2008) KNIME: The Konstanz Information Miner. pp 319–326
44. Sushko I, Salmina E, Potemkin VA, et al (2012) ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. J Chem Inf Model 52:2310–

2316. https://doi.org/10.1021/ci300245q
45. Wolber G, Langer T (2005) LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 45:160–169. https://doi.org/10.1021/ci049885e
46. Steindl TM, Schuster D, Wolber G, et al (2007) High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. J Comput Aided Mol Des 20:703–715. https://doi.org/10.1007/s10822-006-9066-y
47. Steindl TM, Schuster D, Laggner C, et al (2007) Parallel Screening and Activity Profiling with HIV Protease Inhibitor Pharmacophore Models. J Chem Inf Model 47:563–571. https://doi.org/10.1021/ci600321m
48. Krovat EM, Frühwirth KH, Langer T (2005) Pharmacophore Identification, in Silico Screening, and Virtual Library Design for Inhibitors of the Human Factor Xa. J Chem Inf Model 45:146–159. https://doi.org/10.1021/ci049778k
49. Poli G, Seidel T, Langer T (2018) Conformational Sampling of Small Molecules With iCon: Performance Assessment in Comparison With OMEGA. Front Chem 6:
50. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430
51. Halgren TA, Murphy RB, Friesner RA, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem 47:1750–1759. https://doi.org/10.1021/jm030644s

CHAPTER THREE

NLRP3$_{PYD}$-ASC$_{PYD}$ INTERACTION – Supervised structural prediction of protein-protein complexes

3.1 Introduction

The NACHT, LRR and PYD domains-containing 3 (NALP3 or NLRP3) is a cytosolic pattern recognition receptor (PRR) belonging to the nucleotide-binding oligomerization domain (NOD)-like receptor family, that senses exogenous and endogenous danger signals. NLRP3 is expressed predominantly in macrophages and is a component of the inflammasome, a cytosolic machinery consisting of multiprotein oligomers of the innate immune system responsible for the activation of inflammatory responses [1, 2].

Among members of this PRR family, the most explored are NLRP1 (NLR family, pyrin domain- containing 1), the NLRC4 (NLR family, CARD domain-containing 4) and the AIM2 (absent in melanoma 2) inflammasomes [3, 4].

NLRP3 is composed by three domains: a nucleotide-binding oligomerisation domain containing a CARD (caspase activation and recruitment domain) (NACHT), a leucine-rich repeat domain (LRR) and a pyrin domain (PYD).

NLRP3 inflammasome molecular activation still remains to be fully defined. However, recent studies have greatly enhanced to understand the mechanism.

Upon activation, NLRP3 triggers an oligomerisation process producing a helical fibrillar assembly of the adapter apoptosis-associated speck-like protein containing a CARD (ASC) via PYD–PYD interactions. ASC fibrils assemble into the so-called ASC speck structures and recruit pro-caspase-1, that undergoes an autoproteolytic activation, by releasing the activated caspase 1 (figure 3.1) [5–7]. The latter is able to cleave pro-interleukin(IL)-1β and pro-IL-18 to generate the inflammatory cytokines IL-1β and IL-18 [3, 4, 8]. Furthermore, activated caspase-1 is also able to cleave gasdermin-D (GSDMD) to the active form (GSDMD-N), thus triggering the pyroptosis, that is the programmed cell death [3, 9].

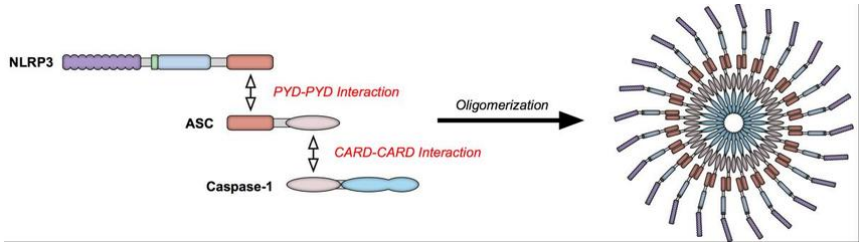


Figure 3.1. NLRP3 oligomerisation complex established by NLRP3$_{PYD}$ and ASC$_{PYD}$ interactions and by ASC and Caspase 1 CARD domains contacts [10].

NLRP3 inflammasome acts as a highly sensitive surveillance mechanism that works against any type of perturbation that damages plasma membrane integrity and the associated K^+ gradient across the membrane [5, 11–13]. Its activation usually requires

two different signals, which are generally referred to as the priming signal (signal 1) and the activation signal (signal 2). The priming step (signal 1) is induced by inflammatory stimuli such as TLR4 agonists which trigger NF-κB-mediated NLRP3 and pro-IL-1β expression, on the other hand, the activation step (signal 2) is induced by PAMPs (pathogen-associated molecular patterns) and DAMPs (damage-associated molecular patterns), that promote NLRP3 inflammasome assembly and caspase-1-mediated IL-1β and IL-18 secretion and pyroptosis [14].

The events that usually provoke NLRP3 inflammasome activation (figure 3.2) are the following:

1. Ion fluxes, such as K^+ efflux, Ca^{2+} signaling, Na^+ influx, and Cl^- efflux [13, 15–20];
2. Reactive oxygen species (ROS) generation, especially from the mitochondria, has also been identified as promoting NLRP3 inflammasome activation; whereas chemical inhibitors preventing ROS production abrogate numerous stimuli-induced NLRP3 inflammasome activation [20–23];
3. Amyloid β (Aβ) protein, a pathogenic misfolded protein expressed in neurodegenerative disease [20, 24];
4. Post-translational modifications of NLRP3, including phosphorylation and ubiquitination [20, 25, 26].

Furthermore, other routes for the activation of the NLRP3 inflammasome, namely non-canonical and alternative inflammasome activation have been recently proposed [27]; whereas the non-canonical NLRP3 inflammasome activation is strictly correlated with gram-negative infection or *in vitro* lipopolysaccharide (LPS)-stimulation.

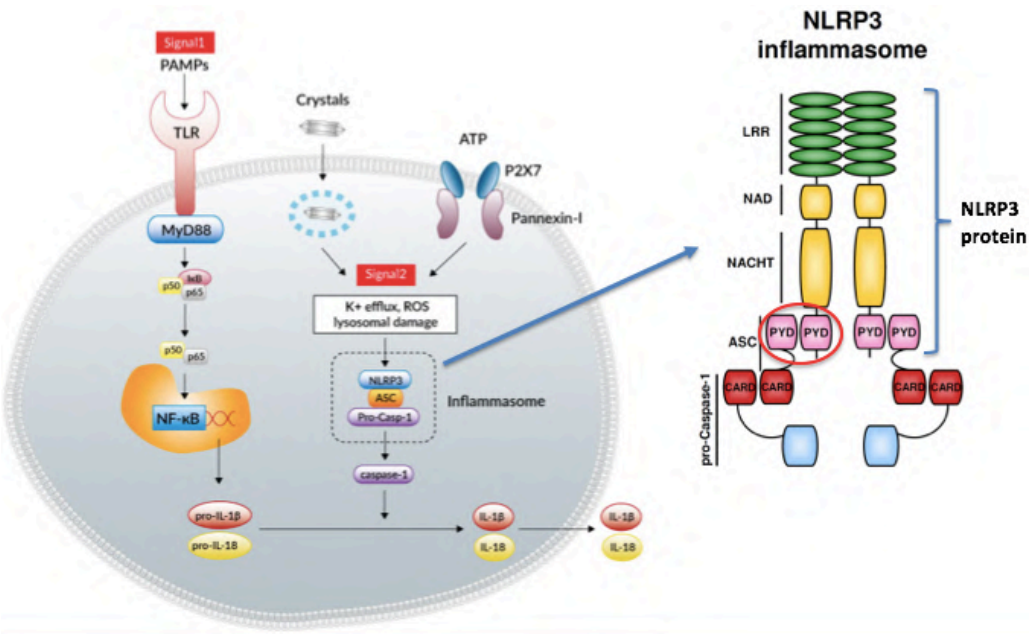


Figure 3.2. NLRP3 Inflammasome composition and cell function.

However, enhanced NLRP3 activity followed by a consequent increase in IL-1β release (and in some cases pyroptotic cell death) has been associated to a large array of diseases, such as chronic inflammatory and autoimmune diseases (such as inflammatory

bowel disease, ulcerative colitis and Crohn's disease [28–31], psoriasis [32], autoimmune encephalomyelitis and multiple sclerosis [33]), several rheumatic diseases and crystal arthropathies (such as systemic lupus erythematosus, rheumatoid arthritis, systemic-onset juvenile idiopathic arthritis, etc. [34–36]).

Moreover, in recent years NLRP3 overexpression and the activation of the NLRP3-mediated inflammatory response has been demonstrated in various neurodegenerative diseases [37, 38] (e.g. amyotrophic lateral sclerosis [39], Alzheimer's and Parkinson's diseases [40–44]), and metabolic and vascular diseases (such as Type 2 diabetes and atherosclerosis [45–49]).

Finally, the NLRP3 inflammasome acts on the pathogenesis of cancer by modulating innate and adaptive immune responses, cell death, proliferation and the gut microbiota. Indeed, excessive inflammation promotes breast cancer, fibrosarcoma, gastric carcinoma and lung metastasis in a context-dependent manner [50, 51].

Therefore, inhibiting NLRP3 inflammatory activity represents a therapeutic strategy that has been demonstrated in several studies, registering inflammatory responses in animal models of myocardial infarction using a variety of experimental NLRP3 inhibitors [52].

Indeed, the aim of this project was to first understand the structural basis of NLRP3 inflammasome assembly, and in detail NLRP3-ASC contacts via PYD-PYD interaction; and then identify putative modulators of this homotypic interaction between the pyrin domains of NLRP3 and ASC protein, in order to prevent the multimeric complex formation.

3.1.1 Pyrin domains characteristics

Experimental evidence has extensively demonstrated that NLRP3 inflammasome assembly depends on a protein interacting domain belonging to the death domain superfamily, which is composed by the following four subfamilies: 39 death domains (DDs), 8 death effector domains (DEDs), 33 caspase-recruitment domains (CARDs) and 22 PYDs belonging to several proteins [53, 54]. The death domain superfamily is one of the biggest families of protein domains and highly prevalent in apoptotic and inflammatory signalling proteins [55, 56].

The canonical globular structure of pyrin domains is characterised by six amphipathic α-helices arranged in an antiparallel α-helical bundle with Greek-key topology (figure 3.3).

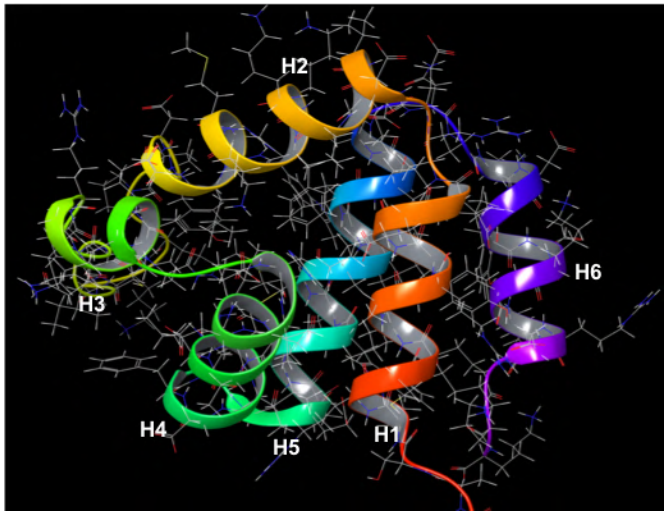


Figure 3.3. Globular structure of a pyrin domain composed of six amphipathic α-helices (H1, H2, H3, H4, H5 and H6).

These six α-helices can be spatially arranged and combined into three distinct types of asymmetric homotypic interactions, involving six patches, whereas the helices positions are illustrated in figure 3.4. In type I interaction, patch Ia, including helices 1 and 4, interacts with helices 2 and 3 from patch Ib. A type II interaction is formed when residues from helix 4 and the loop between helices 4 and 5 composing patch IIa interact with residues of helices 5 and 6 loop (patch IIb). Finally, the type III interaction is established when residues from helix 3 (patch IIIa) interact with the two loops between helices 1 and 2 and helices 3 and 4 (patch IIIb).

In particular, the Type I interaction is the most commonly observed and abundant among the PYDs homotypic complexes [57].

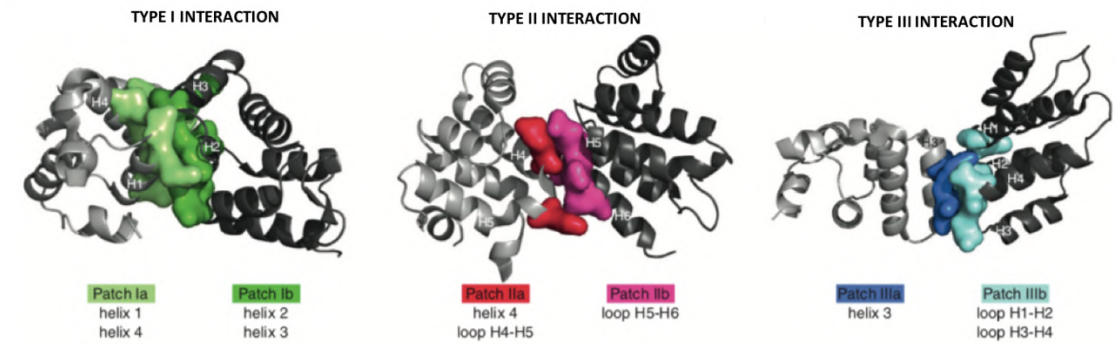


Figure 3.4. The three different asymmetric interaction types established by PYDs [57].

Each PYD exhibits its own surface polarity, as it can be seen in figure 3.5, where significantly widespread blue and red areas at the PYD surfaces are depicted, that refer to partially positively charged and partially negatively charged residue side chains, respectively [58]. The distribution of the polarity at the PYD surface could suggest that these domains should prefer electrostatic or charge-charge interactions, providing important information to help to unveil PYD-PYD interaction.

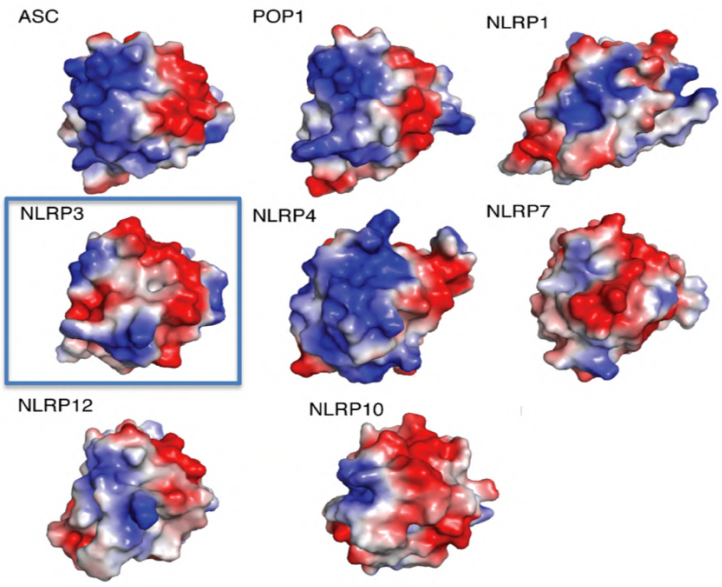


Figure 3.5. Surface polarity of some PYDs, among which NLRP3 pyrin domain (blue-squared structure). The red-coloured areas correspond to the position of partially negatively charged amino acids, while the blue-coloured portions are partially positively charged residues and the white regions correspond to neutral amino acids.

3.1.2 NLRP3$_{PYD}$ interacting region

As a member of death fold domain superfamily, NALP3$_{PYD}$ exhibits the canonical six α-helices structure, H1 to H6, where the helix bundle is tightly packed by a central hydrophobic core. The latter includes Leu10, Ala11, Tyr13, and Leu14 from H1; Phe25 and Leu29 from H2; Leu54, Ala55, and Met58 from H4; Ala67, Ile74, and Phe75 from H5; and Ala87 from H6 (figure 3.6A). A second hydrophobic patch of NALP3 PYD can be detected on residues Phe32, Ile39, Pro40, Leu41, Pro42, Leu57, and Phe61, that take part in the formation of H3 and anchor it to H2 (Figure 3.6B).

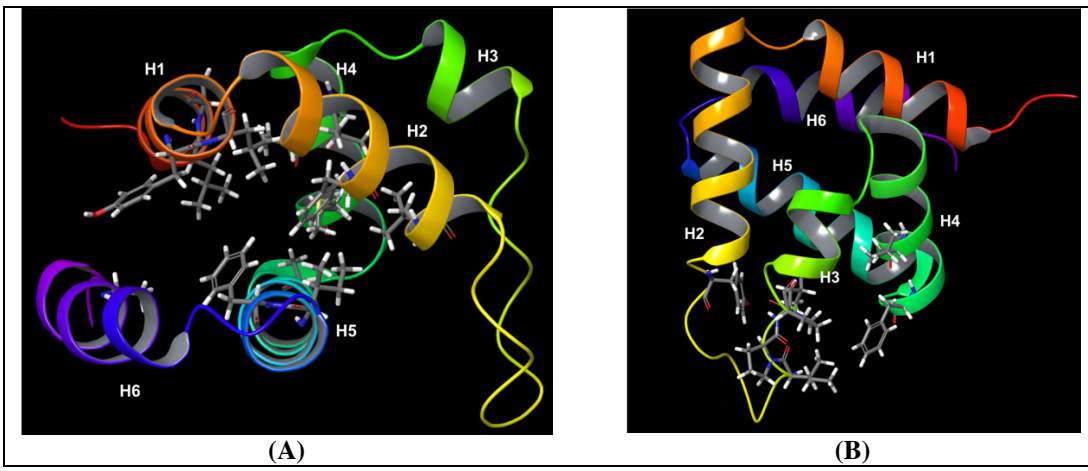


Figure 3.6. 3D structures of NLRP3$_{PYD}$ including residues composing **(A)** the hydrophobic core and **(B)** the hydrophobic patch between H3 and H6.

All these amino acids were found conserved among different PYDs, thus suggesting a potential role for ASC interaction and the overall protein folding [59].

Indeed, NALP3 PYD is highly similar to other PYDs, whereas Bae *et al*. [59] reported the top eight PYD matches identified by a structured-based sequence alignment: ASC, ASC2, NALP7, NALP10, MNDA, NALP1, vFLIP and FADD DEDs.

Furthermore, a number of residues on NALP3 surface are abundantly conserved among PYDs, including the hydrophobic residues Leu17, Leu22, Pro33, Pro34, His51, Val52, Ile59, Gly63, Ile78, and Tyr84. In addition, 10 surface-charged residues, Glu15, Asp21, Lys23, Lys24, Lys48, Asp53, Glu64, Glu65, Arg81, and Lys89 are also conserved among different PYDs [59].

In 2017, through a mass spectrometry analysis Stutz *et al*. discovered that NLRP3 inflammasome assembly is regulated by phosphorylation on three Serine sites, Ser5 located in PYD, Ser161 positioned between PYD and NOD domains, and Ser728 identified in LRR domain [60]. Among these, Ser5 was located at the PYD charge-charge interacting interface, suggesting that phosphorylation likely disrupts the PYD-PYD interaction. Indeed, this insight was confirmed by further assays on the dephosphorylation performed by the protein phosphatase 2A (PP2A), that allowed to restore the downstream signalling.

Therefore, Ser5 phosphorylation is a natural mechanism of NLRP3 inactivation disrupting its assembly. Ser5 is located in a polybasic patch formed by three positively charged residues in Helix 1 (Arg7, Lys9 and Arg12) and three other positively charged amino acids in Helix 6 (Lys86, Arg89 and Lys93).

Therefore, the suggested mechanism of NLRP3 inactivation is based on a negative charge insertion upon Ser5 phosphorylation within the positively charged patch; whereas the addition of the phosphate negative charge neutralises the positive neighbourhood, causing the inactivation of the protein (figure 3.7A-B).

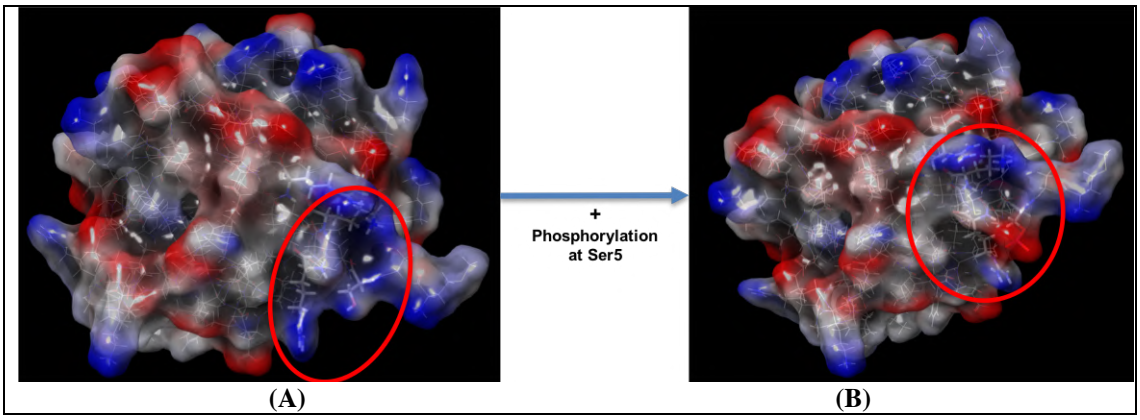


Figure 3.7. NLRP3$_{PYD}$ phosphorylation at Ser5 introduces a negative charge within the positively charged patch by neutralizing it. **(A)** NLRP3$_{PYD}$ surface including the positively charged patch highlighted by the red circle. **(B)** The addition of a phosphate group within the positively charged patch neutralises the area. Blue regions at the surface stand for positive residues, red regions for negative amino acids and white/grey regions for neutral amino acids.

This mechanism was also confirmed by mutational studies performed by Stutz *et al*., where Ser5 was mutated to a neutral amino acid (S5A) and to a positive amino acid (S5R), where these mutations did not disrupt the interaction with ASC$_{PYD}$. On the contrary, the phosphomimetic mutants of NLRP3$_{PYD}$, *i.e*. S5E and S5D, prevented the interaction with ASC pyrin domain. Furthermore, alanine mutations were conducted on Arg7, Lys9 and Arg12 belonging to H1, and Lys86, Arg89 and Lys93 referring to H6. Mutations on helix 1 generated the complete abrogation of NLRP3$_{PYD}$ filament

formation, while mutations on helix 6 did not result in any NLRP3 inflammasome disruption. These data provided crucial information about $NLRP3_{PYD}$ interacting surface, identifying Arg7, Lys9 and Arg12 as the responsible residues for PYD-PYD interaction, and confirming that this latter is driven by electrostatic contacts [60]. Furthermore, NMR titration data shed light on two opposing interfaces on $NLRP3_{PYD}$ that should be directly involved in association with ASC: one interface involves H1 and the N-terminus of H2 and H4 (including residues Arg9, Tyr10, Glu13, Asp14, Val18, Asp19, Leu20, Ala47, Asp48, Val50, Asp51, Lys84 and Asp88), and the other formed by H5 (with residues Thr4, Gly35, Ile37, Phe59, Gly61, Glu63, Thr66, Ala67, Val70, Trp71, Ala74, Glu89 and Lys91) [61].

3.1.3 ASC_{PYD} interacting regions

ASC protein together with NLRP3 plays a key role in the regulation of apoptosis and inflammation through self-association and protein-protein interactions mediated by PYD domains. The pyrin domain of ASC protein is known to self-associate by involving two different interacting surfaces: one consisting of H1, H4 and H5 N-terminus, while the other surface composed by H2, H3 and H5 C-terminus [62]. Indeed, these two interacting surfaces suggest that ASC_{PYD} is able to establish the most common and usual interaction type among PYDs, *i.e.* the type I; that is properly established between helices 1 and 4 of a pyrin domain and helices 2 and 3 of another pyrin domain. In detail, NMR titration data on alanine mutants highlighted that the nature of the interactions between the protein partners is electrostatic. In fact, these two regions show a polar character, where the interface composed by H1, H3-H4 loop, and H4 is negatively charged involving three identified hot spot residues (Glu13, Asp48 and Asp51). On the other hand, the interface generated by H2 and H3 is positively charged with two key residues, Lys21 and Arg41 [62]. Moreover, in the latter interface other key amino acids were identified, Leu25, Val30 and Leu45, establishing hydrophobic interactions (Figure 3.8). Therefore, ASC_{PYD} interactions appear to be mainly dominated by electrostatic contacts and hydrogen bonding with a smaller contribution from hydrophobic side chains. Furthermore, Vajjhala *et al.* suggested that ASC_{PYD} can use these two interacting surfaces to simultaneously engage in self-association and interaction with NLRP3, resembling a type I interaction [63].

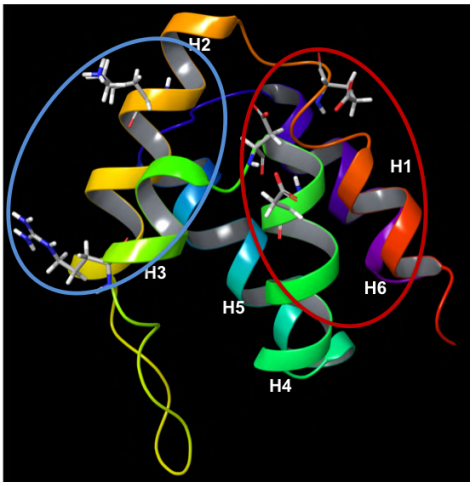


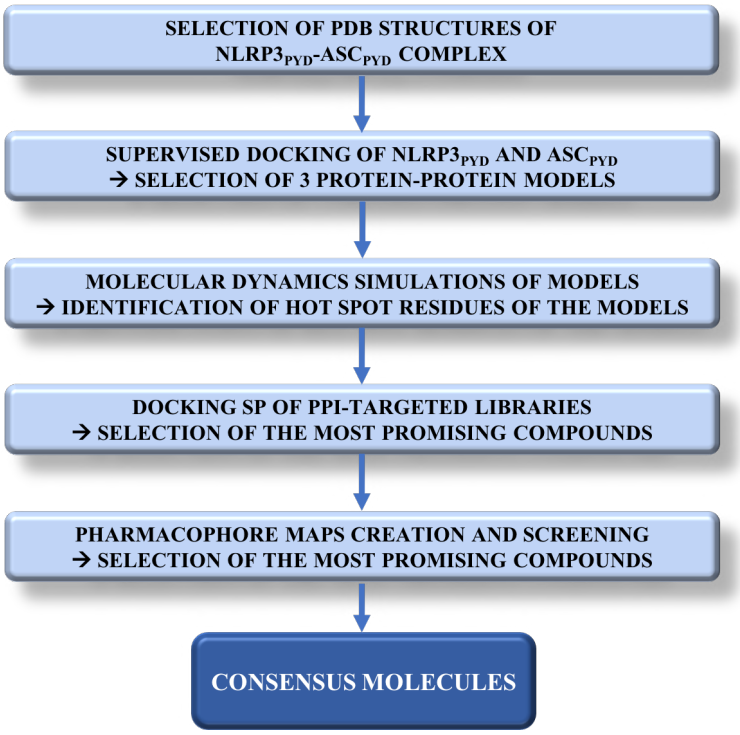
Figure 3.8. ASC_{PYD} including a positively charged interface highlighted by the blue circle and a negatively charged interface identified by the red circle.

Indeed, experimental assays by NMR and analytical ultracentrifugation reported by Oroz *et al*. confirmed that ASC_{PYD}-ASC_{PYD} and ASC_{PYD}-$NLRP3_{PYD}$ share the same interacting regions, but ASC_{PYD} showed a preference for $NLRP3_{PYD}$, as demonstrated by a lower value of the dissociation constant (K_D = 22 μM) for ASC_{PYD}-$NLRP3_{PYD}$ complex compared to ASC_{PYD}-ASC_{PYD} complex (K_D = 40-100 μM) [61].

3.2 Results and discussion

3.2.1 Virtual screening workflow

The above-described data from experimental assays were crucial to design the computational study of this project. Indeed, a crystal structure of ASC_{PYD}-$NLRP3_{PYD}$ complex is currently not available, thus making more complex the study of this protein-protein interaction, and eventually the design and identification of putative modulators. Therefore, in this work, the first step was to build a complex of PYD-PYD interaction by applying a supervised protein-protein docking between ASC_{PYD} and $NLRP3_{PYD}$. This docking provided several results among which three PYD-PYD interaction models were selected as the best ones in terms of docking score and the most similar structures to Type I interaction. Then, these three models were used to perform Molecular Dynamics simulations, setting the simulation time 50 ns, in order to identify the most stable interactions and, in this way, to retrieve the hot spot residues per each complex. Data retrieved from MD calculations were used to run ligand docking of PPI-targeted libraries setting constraints for each model grid. Then, the most promising compounds were selected according to docking scores and the key interactions established with NLRP3 hot spot residues. These molecules were further processed to run a pharmacophore screening, where the pharmacophore maps were built for each $NLRP3_{PYD}$-ASC_{PYD} model. Finally, the related screenings identified consensus molecules according to both computational techniques, ligand docking and pharmacophore screenings. Scheme 3.1 summarises the above-mentioned steps of this work, that are deeply described below in the next sections.



Scheme 3.1. Overview of the computational workflow performed for building ASC_{PYD}-$NLRP3_{PYD}$ complex and screening the PPI-targeted compound libraries.

3.2.2 Selection of the PDB structures

In order to perform the protein-protein docking, the PDB structures of $NLRP3_{PYD}$ and ASC_{PYD} were selected and downloaded from the Protein Data Bank [64]. For ASC protein the PDB 3J63 was used, where several PYD structures are combined to create a unified assembly. For $NLRP3_{PYD}$ the PDB 2NAQ was downloaded, that collects twenty 3D NMR solution structures. Therefore, in order to guide the decision process about $NLRP3_{PYD}$ structure to be used among the twenty, the conformation of ASC_{PYD} PDB 3J63 structure (*i.e.* chain A) was used as a reference due to its bounded state to other PYD structures. Thus, a superimposition was run including ASC_{PYD} (chain A) and each of the twenty $NLRP3_{PYD}$ structures and RMSD values were calculated. The aim was to selected and use for the next steps of this work the $NLRP3_{PYD}$ structure with the lower RMSD value, that could reproduce spatial conformation most similar to ASC_{PYD} (PDB 3J63 chain A). Indeed, the latter already exhibits a binding conformation and besides the pyrin domains essentially share the same interaction type (type I). Therefore, each of the twenty $NLRP3_{PYD}$ structures was superimposed on Chain A of PDB 3J63 using the superposition tool of Schrödinger, considering firstly the C-α of the backbone and then the AA side chains. Although the RMSD values calculated for these structures were similar among them, the best $NLRP3_{PYD}$ entry was the number 6 reporting the lowest RMSD values considering both backbone C-α and AA side chains. Therefore, entry 6 was further processed to perform protein-protein docking. Table 3.1 reports the RMSD values for both backbone C-α and AA side chains.

Table 3.1. RMSD values calculated on the twenty 3D NMR solution structures of $NLRP3_{PYD}$ from PDB 2NAQ, by using Chain A of ASC_{PYD} assembly of PDB 3J63.

PDB 2NAQ – Entry number	RMSD (Å) – Backbone C-α	RMSD (Å) – Side chains
Entry 1	3.1107	7.6229
Entry 2	3.2339	7.8069
Entry 3	2.9078	7.6338
Entry 4	3.1872	7.9247
Entry 5	3.0677	7.8981
Entry 6	**2.8620**	**7.4860**
Entry 7	3.1439	7.6654
Entry 8	3.0968	7.6607
Entry 9	3.2437	7.6734
Entry 10	3.1454	7.7077
Entry 11	3.1172	7.7956
Entry 12	3.1016	7.8481
Entry 13	3.2598	7.7021
Entry 14	3.0752	7.6129
Entry 15	3.1583	7.6243
Entry 16	3.2182	7.7137
Entry 17	3.0192	7.8918
Entry 18	3.2094	7.7054
Entry 19	3.2992	7.9603
Entry 20	3.1053	7.8364

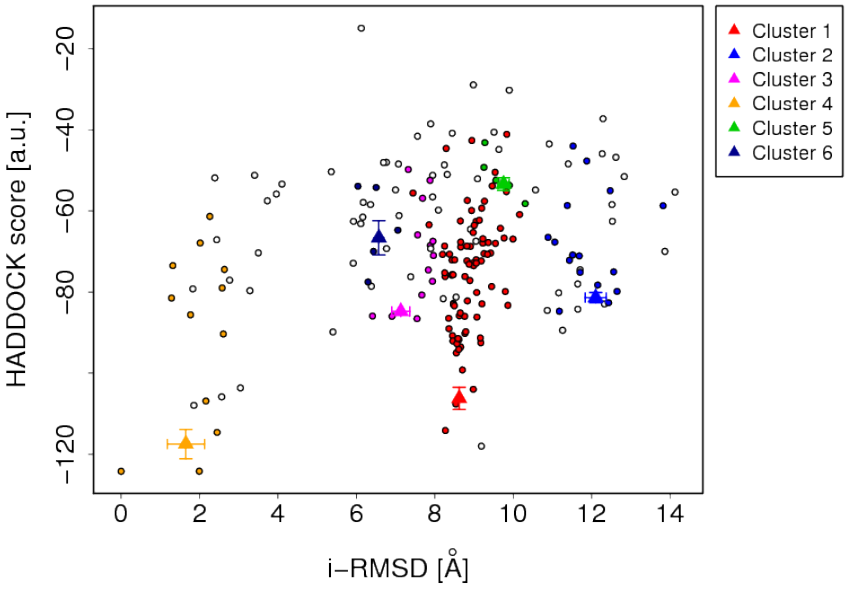
3.2.3 Supervised protein-protein docking

As previously mentioned, ASC_{PYD} is able to establish a Type I interaction with its PYD partner. Therefore, the aim of the protein-protein docking was to reproduce this protein-protein binding mode for ASC_{PYD}-$NLRP3_{PYD}$ complex. For this purpose, HADDOCK software (v2.2) [65] was used and two different docking protocols were applied.

For the first protocol, all the hot spot residues for both proteins as reported in literature were defined as "active" in the software, *i.e.* Arg7, Lys9 and Arg12 (involved in H1-H4) were indicated for $NLRP3_{PYD}$, while Lys21, Leu25, Arg41, Leu45 (involved in H1-H4), and Glu13, Asp48, Asp51 (involved in H2-H3) were selected for ASC_{PYD}. HADDOCK software (v2.2) generated 133 complexes that were grouped into 6 clusters, whereas the related data are reported in table 3.2 and the principal component analysis in plot 3.1.

Table 3.2. ASC_{PYD}-$NLRP3_{PYD}$ complex clusters sorted by HADDOCK scores, from lowest to highest.

	ASC_{PYD}-$NLRP3_{PYD}$ COMPLEX CLUSTER NUMBER					
	Cluster 4	**Cluster 1**	**Cluster 3**	**Cluster 2**	**Cluster 6**	**Cluster 5**
HADDOCK score	-117.5 ± 7.3	-106.2 ± 5.4	-84.8 ± 2.4	-81.4 ± 2.5	-66.6 ± 8.5	-53.4 ± 3.2
Cluster size	12	82	13	16	5	5
RMSD from the overall lowest-energy structure	2.0 ± 1.2	8.4 ± 0.2	7.0 ± 0.5	12.5 ± 0.7	6.6 ± 0.3	10.1 ± 0.3
Van der Waals energy	-31.7 ± 2.4	-34.1 ± 3.3	-32.2 ± 3.1	-28.1 ± 2.9	-17.1 ± 7.5	-22.4 ± 5.0
Electrostatic energy	-427.0 ± 60.3	-282.2 ± 25.7	-244.1 ± 54.9	-217.7 ± 64.4	-194.5 ± 28.1	-141.8 ± 33.9
Desolvation energy	-1.3 ± 4.6	-19.0 ± 6.3	-4.0 ± 7.9	-10.6 ± 12.0	-12.9 ± 8.0	-4.3 ± 7.6
Restraints violation energy	9.9 ± 13.54	33.0 ± 32.26	2.9 ± 0.49	8.7 ± 14.69	22.1 ± 22.56	16.9 ± 16.85
Buried surface area	1348 ± 54.6	1228.0 ± 29.1	1365.6 ± 73.4	933.4 ± 45.9	897.2 ± 16.3	877.9 ± 168.5
Z-Score	-1.5	-1.0	0.0	0.2	0.8	1.4



Plot 3.1. HADDOCK scores of first docking protocol are plotted vs i-RMSD of each PYD-PYD generated complex; where i-RMSD is the interface-RMSD calculated on the backbone (CA, C, N, O, P) atoms of all residues involved in intermolecular contact using a 10 Å cut-off; a.u. are arbitrary units. The cluster averages and standard deviations are indicated by coloured dots with associated error bars. The average values are calculated on the best four structures of each cluster (based on the HADDOCK score).

Cluster 4 and Cluster 1 were considered the most promising according to the lowest HADDOCK scores, Z-scores, the most abundant cluster size and the capacity in reproducing type I interaction. Indeed, the first PYD-PYD complex, hereby Model 1 presented the lowest HADDOCK score (-117.5 ± 7.3) and Z-score (-1.5), and good cluster size including 12 PYD-PYD complexes. On the contrary, in terms of helices binding modes similarity to type I interaction, the helices involved in the protein-protein contacts were not exactly matching to Type I helices spatial arrangement. Indeed, according to Model 1, H1-H6 of NLRP3$_{PYD}$ established interactions with H1-H4 of ASC$_{PYD}$. Figure 3.9 depicts the Model 1.

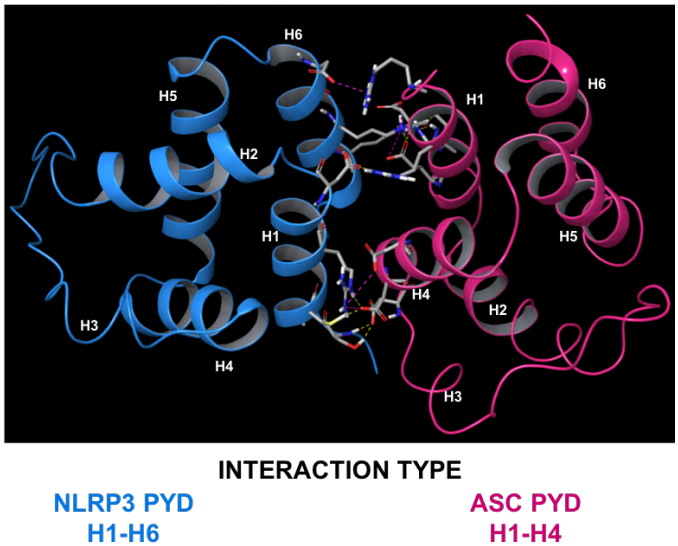


Figure 3.9. Model 1 of ASC$_{PYD}$-NLRP3$_{PYD}$ complex, where the blue structure is NLRP3$_{PYD}$ and the pink one is ASC$_{PYD}$. The protein-protein interaction is established between H1-H6 of NLRP3$_{PYD}$ and H1-H4 of ASC$_{PYD}$.

According to Model 1 complex, the main interactions were essentially hydrogen bonds and some salt bridges. Table 3.3 reports this information.

Table 3.3. Main interactions established between $NLRP3_{PYD}$ and ASC_{PYD} in Model 1.

MODEL 1		
$NLRP3_{PYD}$ amino acid	**ASC_{PYD} amino acid**	**Interaction**
Ser5	Asp48	2 H-bonds
Cys8	Asp51	1 H-bond
Arg12	Asp51 Asp54	1 H-bond + 1 salt bridge 1 salt bridge
Asp16	Arg5	1 H-bond
Asp80	Arg3	1 salt bridge
Lys86	Asp6 Asp10 Glu13	1 H-bond + 1 salt bridge 1 salt bridge 1 H-bond + 1 salt bridge
Arg89	Glu13	1 H-bond + 1 salt bridge

The $NLRP3_{PYD}$-ASC_{PYD} complex was also explored performing a computational alanine scanning by using PPCheck tool [66] to unveil the hot spot amino acids for both protein partners of Model 1. All those amino acids that showed high differences between the complex total energies after mutation and before mutation were considered hot spots for the Model 1. Table 3.4 shows the results including the total energies of the wild-type (before mutation) and the mutated complexes (after mutation) for those amino acids providing high $\Delta\Delta G$ difference value.

Table 3.4. Computational alanine scanning results of Model 1.

		Total energy of the complex (kJ/mol)		
	Amino acid	**Before mutation**	**After mutation**	**Difference**
ASC	Asp10	-252.59	-202.40	50.19
	Glu13	-252.59	-201.79	50.80
	Arg3	-252.59	-230.41	22.18
	Leu50	-252.59	-240.51	12.08
	Asp51	-252.59	-228.74	23.85
	Arg5	-252.59	-224.56	28.03
NLRP3	Arg12	-252.59	-188.00	64.59
	Lys86	-252.59	-191.43	61.16
	Arg89	-252.59	-179.46	73.13

The second PYD-PYD complex selected from the docking showed a slightly lower HADDOCK score (-106.2 ± 5.4) compared to Model 1, but better similarity to the type I interaction, where H1-H4 of $NLRP3_{PYD}$ interacted with H1-H2 loop, H2, H3-H4 loop, and H4 of ASC_{PYD}, as depicted in figure 3.10. This complex herein refers to Model 2 and considered for further analysis.

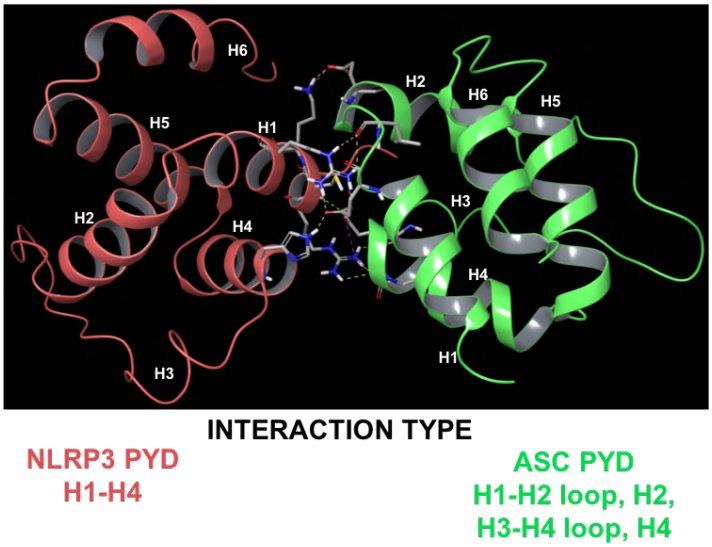


Figure 3.10. Model 2 of ASC$_{PYD}$-NLRP3$_{PYD}$ complex, where the orange structure is NLRP3$_{PYD}$ and the green one ASC$_{PYD}$. The protein-protein interaction is established between H1-H4 of NLRP3$_{PYD}$ and H1-H2 loop, H2, H3-H4 loop and H4 of ASC$_{PYD}$.

Even for Model 2, the analysis of the interactions established between the two PYDs highlighted mainly hydrogen bonds and even salt bridges and VdW contacts, as listed in table 3.5.

Table 3.5. Main interactions established between NLRP3$_{PYD}$ and ASC$_{PYD}$ in Model 2.

MODEL 2		
NLRP3$_{PYD}$ amino acid	**ASC$_{PYD}$ amino acid**	**Interaction**
Arg7	Asp51 Asp48 Ser46	2 H-bonds +1 salt bridge 1 salt bridge 1 H-bond
His51	Glu13	1 H-bond
Arg12	Glu13 Leu15	2 H-bonds + 1 salt bridge 2 H-bonds
Cys8	Asp48	1 H-bond
Lys9	Glu18	1 H-bond + 1 salt bridge
Val52	Leu50	VdW contacts
Met3	Leu25 Leu45	VdW contacts VdW contacts

This second model was also used to perform computational alanine scanning calculations in order to focus the attention on specific hot spots of this second PYD-PYD complex, and the results are reported in the following table 3.6.

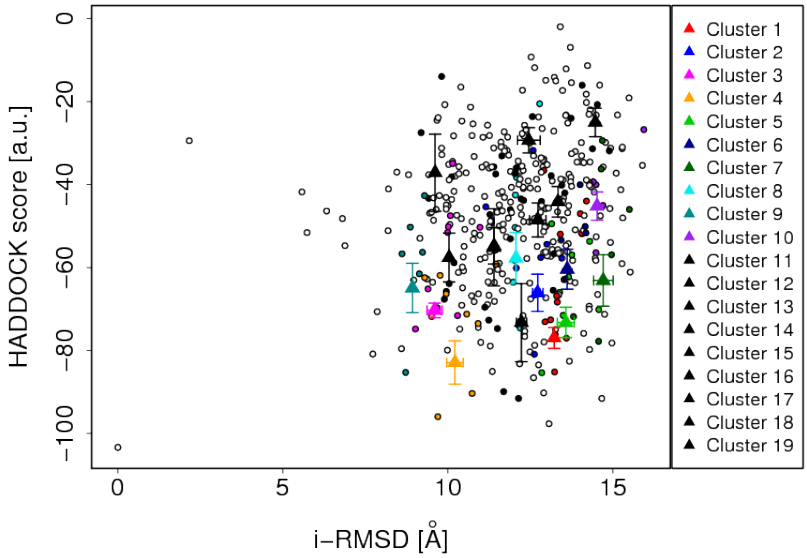
Table 3.6. Computational alanine scanning results of Model 2.

		Total energy of the complex (kJ/mol)		
	Amino acid	**Before mutation**	**After mutation**	**Difference**
ASC	Glu13	-213.07	-143.49	69.58
	Asp51	-213.07	-208.63	38.73
NLRP3	Arg12	-213.07	-146.64	66.43
	His51	-213.07	-190.50	22.57
	Arg7	-213.07	-148.78	64.29
	Lys9	-213.07	-189.16	23.91

Finally, a second protein-protein docking protocol was applied, but in this case not all the hot spots of ASC_{PYD} were defined as "active" during setting the docking parameters. The aim was to force the software in reproducing the Type I interaction for the $NLRP3_{PYD}$-ASC_{PYD} complex. Indeed, two considerations were done: 1) Type I interaction is established by H1-H4 of a pyrin domain and H2-H3 of another PYD, and 2) the NLRP3 identified hot spots residues belong to H1-H4 interface (Arg7, Lys9 and Arg12). Therefore, for ASC_{PYD} only the hot spots referring to H2-H3 interacting surface (Lys21, Arg41, Leu25 and Leu45) were defined as active. Hence, the best solution of this docking was the third model (Model 3) with HADDOCK score of -77.0 ± 5.0. This second protein-protein docking built 115 PYD-PYD structures clustered into 19 groups and, only for the best ten, data are reported in table 3.7. In plot 3.2 HADDOCK scores for each protein-protein complex are plotted vs i-RMSD.

Table 3.7. ASC_{PYD}-$NLRP3_{PYD}$ complex clusters sorted by HADDOCK scores, from lowest to highest.

	ASC_{PYD}-$NLRP3_{PYD}$ COMPLEX CLUSTER NUMBER									
	Clust. 4	**Clust. 1**	**Clust. 5**	**Clust. 11**	**Clust. 3**	**Clust. 2**	**Clust. 9**	**Clust. 7**	**Clust. 6**	**Clust. 8**
HADDOCK score	-82.9 ± 10.5	-77.0 ± 5.0	-73.3 ± 7.3	-73.3 ± 18.8	-70.3 ± 3.5	-66.1 ± 9.0	-65.0 ± 11.9	-63.1 ± 12.5	-60.4 ± 9.7	-57.7 ± 12.3
Cluster size	9	12	8	5	9	9	6	6	6	6
RMSD from the overall lowest-energy structure	10.2 ± 0.4	11.9 ± 0.4	11.7 ± 0.3	11.7 ± 0.4	9.2 ± 1.1	11.1 ± 0.4	8.7 ± 0.8	12.9 ± 0.4	11.2 ± 0.2	11.6 ± 0.4
Van der Waals energy	-19.6 ± 4.2	-17.2 ± 6.0	-11.5 ± 4.7	-14.8 ± 7.4	-21.8 ± 2.5	-13.6 ± 4.9	-22.0 ± 5.6	-4.9 ± 3.0	-8.9 ± 4.1	-14.9 ± 4.7
Electrostatic energy	-232.0 ± 29.9	-247.3 ± 37.0	-254.4 ± 24.2	-236.0 ± 74.7	-174.2 ± 23.9	-219.6 ± 76.8	-127.0 ± 59.2	-195.1 ± 43.0	-197.6 ± 32.5	-201.2 ± 82.2
Desolvation energy	-19.9 ± 12.7	-10.4 ± 11.0	-12.7 ± 2.0	-12.1 ± 5.9	-15.9 ± 4.8	-8.8 ± 10.8	-21.0 ± 7.7	-20.5 ± 3.9	-12.6 ± 1.1	-3.3 ± 5.6
Restraints violation energy	30.1 ± 16.36	0.9 ± 0.89	17.9 ± 119.2	8.9 ± 13.3	22.5 ± 18.0	2.6 ± 2.11	34.5 ± 33.0	12.7 ± 18.10	6.1 ± 9.47	7.8 ± 13.29
Buried surface area	941.6 ± 14.8	881.2 ± 174.0	703.3 ± 119.2	698.2 ± 129.5	857.4 ± 88.8	788.6 ± 107.8	741.0 ± 129.8	493.5 ± 134.5	594.9 ± 133.3	805.2 ± 29.8
Z-Score	-1.9	-1.1	-0.6	-0.6	-0.2	0.4	0.5	0.8	1.1	1.5



Plot 3.2. HADDOCK scores of second docking protocol are plotted vs i-RMSD of each PYD-PYD generated complex, where i-RMSD is the interface-RMSD calculated on the backbone (CA, C, N, O, P)

atoms of all residues involved in intermolecular contact using a 10 Å cut-off; a.u. are arbitrary units. The cluster averages and standard deviations are indicated by coloured dots with associated error bars. The average values are calculated on the best four structures of each cluster (based on the HADDOCK score).

Model 3, illustrated in figure 3.11, was also analysed to identify the main interactions between the two PYDs highlighted by this model. Table 3.8 reports the most important interactions, that are mainly hydrogen bonds and salt bridges.

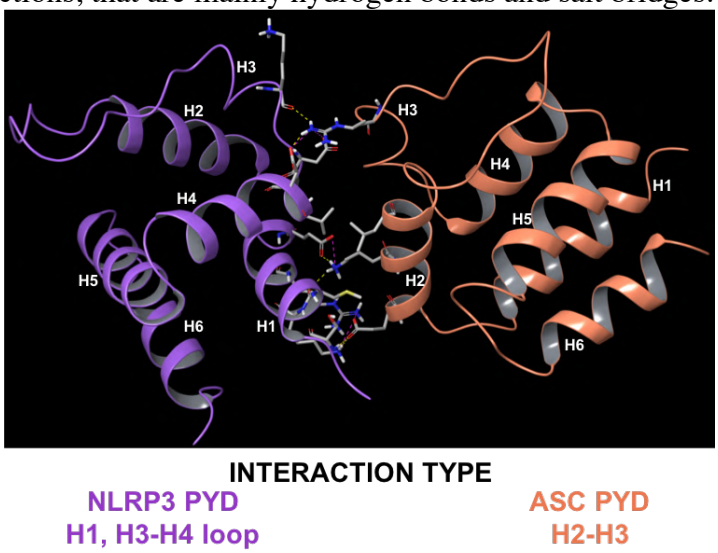


Figure 3.11. Model 3 of ASC_{PYD}-$NLRP3_{PYD}$ complex, where the purple structure is $NLRP3_{PYD}$ and the orange one ASC_{PYD}. The protein-protein interaction is established between H1 and H3-H4 loop of $NLRP3_{PYD}$ and H2-H4 of ASC_{PYD}.

Table 3.8. Main interactions established between $NLRP3_{PYD}$ and ASC_{PYD} in Model 3.

MODEL 3		
$NLRP3_{PYD}$ amino acid	**ASC_{PYD} amino acid**	**Interaction**
Ser5	Glu18	1 H-bond
Cys8	Asp48	1 H-bond
Lys9	Glu18	1 H-bond + 1 salt bridge
Arg12	Glu18	1 salt bridge
Glu15	Lys21	1 H-bond + 1 salt bridge
Asp50	Arg41	1 salt bridge
Asp53	Arg41	1 H-bond + 1 salt bridge
Lys48	Arg41	1 H-bond
His51	Leu45	VdW contacts
Val52	Leu25	VdW contacts

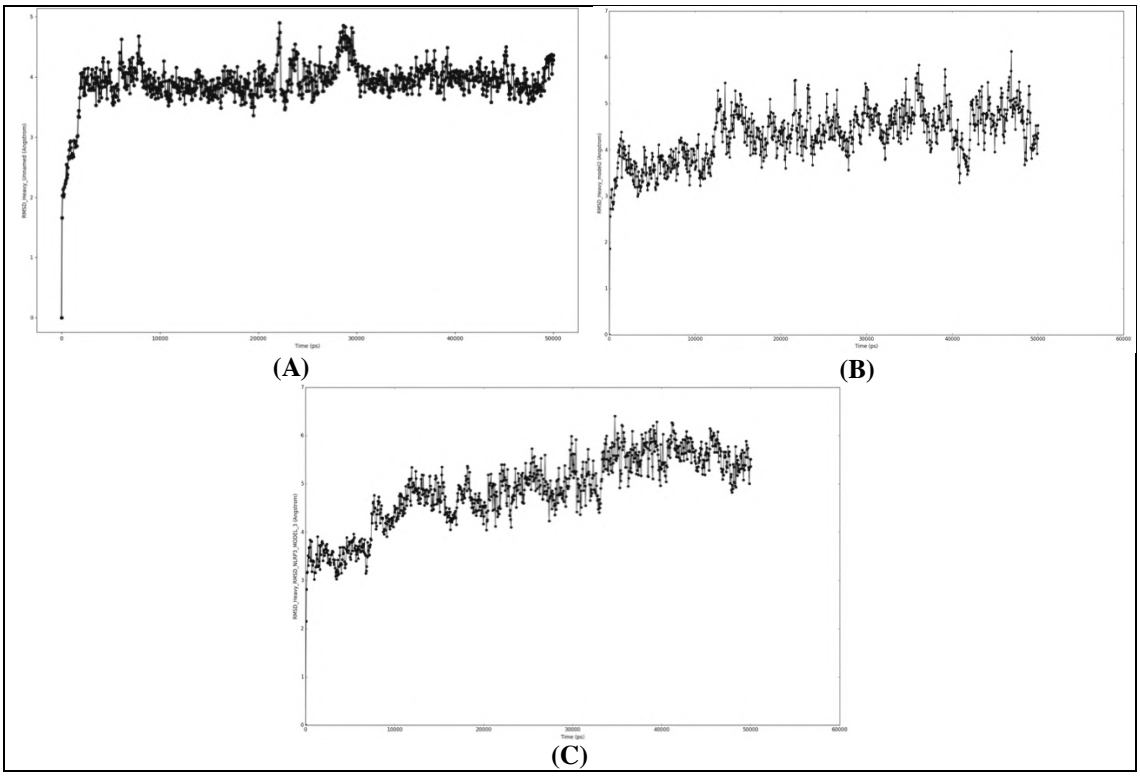
Finally, as for the previous two models, the computational alanine scanning was run to identify the key residues for Model 3, and the results are collected in table 3.9.

Table 3.9. Computational alanine scanning results of Model 3.

		Total energy of the complex (kJ/mol)		
	Amino acid	**Before mutation**	**After mutation**	**Difference**
ASC	Glu18	-184.53	-138.97	45.56
	Lys21	-184.53	-156.81	27.72
	Leu25	-184.53	-175.62	8.91
	Arg41	-184.53	-110.05	74.48
	Leu45	-184.53	-172.29	12.24
NLRP3	Glu15	-184.53	-160.93	23.60
	Asp50	-184.53	-149.17	35.36
	Asp53	-184.53	-143.20	41.33
	Lys9	-184.53	-162.56	21.97

3.2.4 Molecular Dynamics simulations of the three models

In order to better minimise the PYD-PYD complexes generated by running the protein-protein docking, and even more to confirm and/or identify the most stable interactions for both proteins, a short-time all-atom Molecular Dynamics simulation (t = 50 ns) was performed for each model. The analysis of the trajectories demonstrated that the systems were stable during the whole simulations, as shown by the plots 3.3A-C below reported.



Plot 3.3. RMSD plots of MD trajectories of **(A)** Model 1, **(B)** Model 2, and **(C)** Model 3.

MD frames of all the three models were clustered and the frames representative for the resulting clusters were analysed, in order to retrieve the most stable contacts between the two PYD domains.

For Model 1, up to ten frame clusters were generated, where the most representative frames were:

- Frame 350 at 17.5 ns → representative for 5 frames;
- Frame 800 at 41 ns → representative for 23 frames;
- Frame 270 at 13.5 ns → representative for 19 frames;
- Frame 490 at 24.5 ns → representative for 16 frames.

Therefore, the analysis of the PYD-PYD binding for each of these frames highlighted the most abundant and frequent interactions among the two proteins, as reported in table 3.10.

Table 3.10. Analysis of the most abundant and frequent interactions between NLRP3$_{PYD}$-ASC$_{PYD}$ complexes of the most representative MD frames of Model 1.

MD FRAMES OF MODEL 1		
NLRP3$_{PYD}$ amino acid	**ASC$_{PYD}$ amino acid**	**Interaction**
Lys9	Glu13	1 H-bond + 1 salt bridge
Arg12	Asp54	2 H-bonds + 1 salt bridge
Glu15	Arg5	1 salt bridge
Asp16	Arg3	1 H-bond + 1 salt bridge
Lys86	Asp10	1 H-bond + 1 salt bridge
Arg89		2 H-bonds + 1 salt bridge

For the second model, the MD frames were clustered considering up to ten clusters to retrieve. Hence, the most representatives were:

- Frame 40 at 2.0 ns → representative for 46 frames;
- Frame 930 at 46.5 ns → representative for 25 frames;
- Frame 690 at 34.5 ns → representative for 30 frames.

These three frames were analysed and the most abundant and frequent interactions were collected and listed in the following table 3.11.

Table 3.11. Analysis of the most abundant and frequent interactions between NLRP3$_{PYD}$-ASC$_{PYD}$ complexes of the most representative MD frames of Model 2.

MD FRAMES OF MODEL 2		
NLRP3$_{PYD}$ amino acid	**ASC$_{PYD}$ amino acid**	**Interaction**
Arg7	Asp51	2 H-bonds + 1 salt bridge
	Asp48	1 H-bond + 1 salt bridge
Ser5	Asp48	1 H-bond
Arg12	Glu13	1 H-bond + 1 salt bridge
His51	Glu13	VdW contacts
Val52	Leu50	VdW contacts
Cys8	Ala49	VdW contacts

For the third model, ten clusters were identified and the most populated were:

- Frame 970 at 48.5 ns → representative for 61 frames;
- Frame 340 at 17.0 ns → representative for 25 frames;
- Frame 60 at 3.0 ns → representative for 13 frames.

The interactions of the above frames of the PYD-PYD complexes were explored and the most stable ones were identified. Table 3.12 summarises them.

Table 3.12. Analysis of the most abundant and frequent interactions between NLRP3$_{PYD}$-ASC$_{PYD}$ complexes of the most representative MD frames of Model 3.

MD FRAMES OF MODEL 3		
NLRP3$_{PYD}$ amino acid	**ASC$_{PYD}$ amino acid**	**Interaction**
Asp53	Arg41	2 H-bonds + 1 salt bridge
Asp50		1 H-bond + 1 salt bridge
Ser5	Glu18	1 H-bond
Val52	Leu25	VdW contacts
Cys8	Leu25	VdW contacts
His51	Gly42	VdW contacts

These data were further processed to perform supervised ligand docking of PPI-targeted libraries by building docking grids including H-bond constraints.

3.2.5 Supervised ligand docking

All the data from literature and computational techniques were collected and processed to tune ligand docking protocols for each model. Indeed, considering the hot spot residues identified by NMR titration data performed by Varjjhala *et al*. [62] and the computational alanine scanning results obtained by using PPCheck software [66], docking grids for each model were generated including H-bond constraints. In detail, for Model 1 three constraints as hydrogen bond donors were defined on the side chain groups of Lys9, Arg12 and Lys86; for Model 2 three H-bond donor constraints on the side chain groups of Arg7, Arg12 and His51; and finally, for Model 3 only two constraints were set indicating hydrogen bond acceptors on the side chain groups of Asp50 and Asp53. In order to run the molecular docking calculations PPI-targeted libraries of commercially available compounds were downloaded from Asinex, ChemDiv, Enamine and Life Chemicals databases, that included α-helix mimetics, peptidomimetics, non-peptide peptidomimetics, shape-helix mimetics and PPI-enriched libraries together with an in-house library of chemical entities designed and synthesised by the Medicinal Chemistry group of STEBICEF Department (University of Palermo). These molecules were filtered deleting all those compounds containing reactive, toxic or carcinogenic groups, thus getting overall about 175K of compounds, and then were optimised at pH 7.0. Moreover, the related outputs were used to perform ligand docking screenings on the three grids, filtering results including only molecules matching at least 2 of the defined constraints. For Model 1 overall about 95K compounds were retrieved, while for Model 2 the outputs turned out about 120K molecules, and for Model 3 about 22K compounds were able to dock $NLRP3_{PYD}$. Examples of ligand binding modes are depicted in figure 3.12A-C.

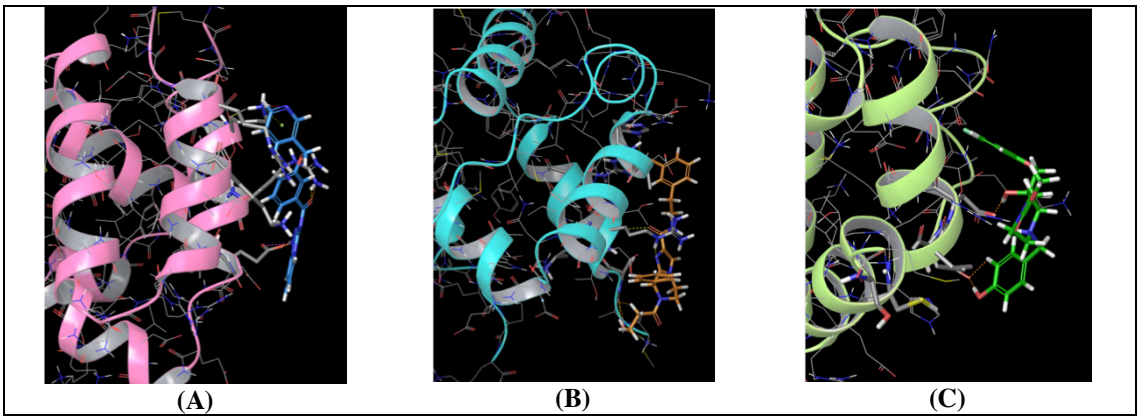


Figure 3.12. Examples of binding modes of PPI-targeted compounds to $NLRP3_{PYD}$ based on **(A)** Model 1, **(B)** Model 2 and **(C)** Model 3.

3.2.6 Pharmacophore maps creation and screenings

The docking outputs showed to be not selective in terms of number of molecules retrieved from the initial ones. Therefore, running another computational technique, such as the pharmacophore approach, could be useful to deeply explore the binding surface of NLRP3 pyrin domain and address the results to identify consensus molecules. This process should increase the success rate in identifying potential hit compounds or modulators of the PYD-PYD interaction.

Therefore, for each complex model a merged pharmacophore was created based on ASC_{PYD}-$NLRP3_{PYD}$ complex structure generated by HADDOCK software (v2.2) [65] and the related MD representative frames that were above analysed. Only for Model 3, two different pharmacophores were generated as herein described.

First, for Model 1 a supervised pharmacophore map was created. For this purpose, five different pharmacophore maps were built based on PYD-PYD complex generated by HADDOCK software (v2.2) [65] and the four most representative MD frames (frame 350, 800, 270 and 490). All these 3D maps were generated based by defining ASC_{PYD} as ligand and $NLRP3_{PYD}$ as the receptor. The five pharmacophore maps differed for some features, hence it was necessary to create a unique merged pharmacophore map that was further modified according to data from the literature and the computational studies. Therefore, a hydrophobic feature on ASC Leu45 and a hydrogen bond acceptor feature on ASC Ser46 were deleted, because they were only present in PYD-PYD complex structure from docking. Then, hydrogen bond acceptor and negative ionisable features on ASC Glu13 belonging to all the pharmacophore maps were interpolated because they showed close 3D spatial positions. The same procedure was applied for ASC Asp51 and Asp54, whose hydrogen bond acceptor and negative ionisable features were interpolated and the sphere tolerance was increased by 0.30 Å. For the same reasons, the hydrogen bond donor and positive ionisable features of ASC Arg3 and ASC Arg5 were interpolated and the sphere tolerance was increased as well by 0.30 Å. The resulting pharmacophore map was composed of 13 interaction features. However, such a copious pharmacophore could be too selective for virtual screening, hence some features were marked as optional for screening purposes. For this reason, the hydrophobic feature on ASC Leu50, the hydrogen bond acceptor and the negative ionisable features on ASC Asp6, and the hydrogen bond donor and positive ionisable features referring to ASC Arg3 and Arg5 were defined as optional, because considered less important according to literature data and computational studies outputs. Figure 3.13A shows the Model 1 pharmacophore map. This latter was used to perform a virtual screening including molecules obtained from docking outputs. For this purpose, the maximum number of omitted features to be permitted was 4, thus getting 113 consensus molecules.

For the second PYD-PYD model, four different pharmacophore maps were created, considering the $NLRP3_{PYD}$-ASC_{PYD} complex structure from HADDOCK docking, and the MD representative frames 930, 690 and 40. The 3D interaction maps were merged and the resulting pharmacophore was further refined according to data from literature and computational studies, but also considering the features mostly shared by the four maps. Therefore, the hydrophobic features corresponding to ASC Leu50 and the hydrogen bond acceptor and negative ionisable features referring to Asp48 were interpolated among the four maps and for the resulting features the tolerance was increased by 0.30 Å. Finally, the features related to ASC Leu9 and Glu18 were marked as optional. The final pharmacophore map was composed of overall 10 features and is depicted in figure 3.13B. It was used to perform another pharmacophore screening with the molecules retrieved from the docking on Model 2. The screening produced 77 consensus molecules.

For the third model, two different final pharmacophore maps were created and used for virtual screenings, due to specific positional differences for two features referred to ASC Glu18 and Lys21. Both maps were based on merging the pharmacophore map built on PYD-PYD complex from HADDOCK docking and an MD frame, *i.e.* frame 970 (figure 3.13C) and frame 340 (figure 3.13D). Therefore, for both final maps, the two

hydrogen bond donor features corresponding to ASC Arg41 were interpolated and the related tolerance was increased by 0.30 Å. For frame 970, a hydrophobic feature on Leu28 was deleted, while for frame 340 a hydrophobic, a negative ionisable and a hydrogen bond donor features were deleted as well, marking the hydrophobic feature on Leu45 as optional for both pharmacophores. These two 3D maps were used to run separately two screening including molecules selected from the ligand docking previously run on Model 3. The overall consensus molecules obtained from both screenings were 76.

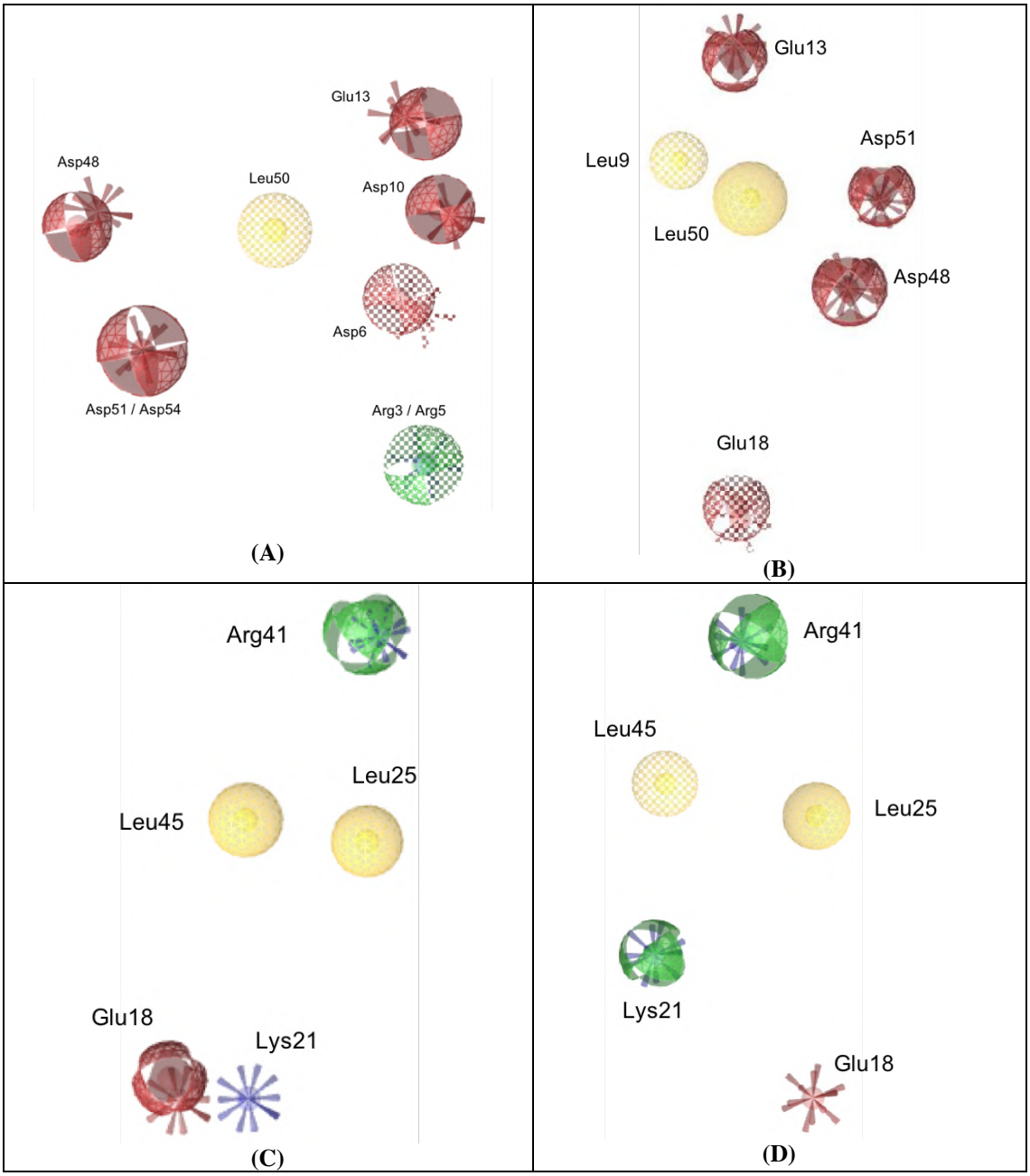
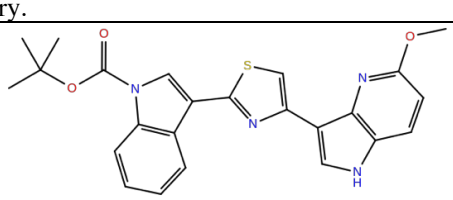
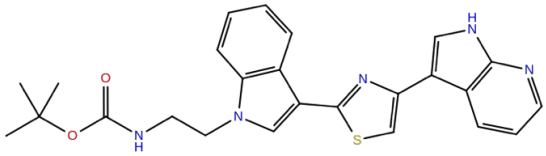
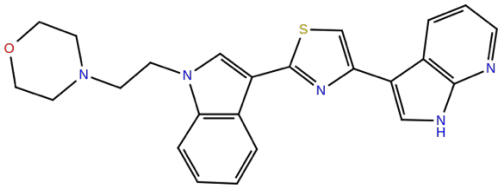
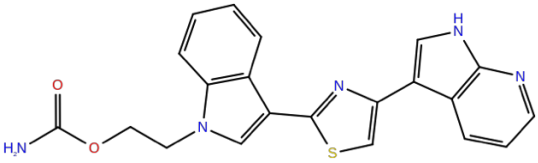
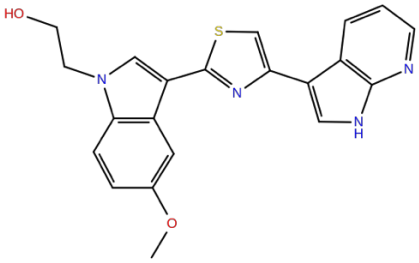


Figure 3.13. Supervised pharmacophore maps of **(A)** Model 1, **(B)** Model 2, **(C)** Model 3 MD frame 970 and **(D)** Model 3 MD frame 340. Yellow spheres are hydrophobic interactions; red spheres are hydrogen bond acceptors; green spheres are hydrogen bond donors; blue spikes are positive ionisable features; red spikes are negative ionisable features; and dotted spheres are features marked as optional.

The consensus molecules for the three models were then filtered to delete all the duplicates between the virtual screening outputs, identifying overall 28 molecules from the in-house compound library (the main scaffolds of these compounds are depicted in table 3.13) and 229 compounds from the PPI-targeted libraries. The latter were further filtered through CANVAS software [67, 68] released by Schrödinger, by using REOS and PAINS structures in order to delete all those molecules containing reactive groups or groups able to interfere with the biological assays, respectively. The resulting compounds were overall 28 in-house compounds, designed by Medicinal Chemistry group of STEBICEF Department, and 193 unique chemical entities from PPI-targeted libraries. These selected molecules have been assaying in order to validate which of these three models is the most reliable and similar to the actual ASC_{PYD}-$NLRP3_{PYD}$ complex. Table 3.13 shows the main representative scaffolds of the 28 in-house consensus compounds. Furthermore, the identification of potential hit can pave the way to the discovery of modulators of the PYD-PYD interaction between NLRP3 and ASC proteins, thus representing a potential strategy to tackle the diseases associated to the overactivation of the inflammasome.

Table 3.13. Main representative scaffolds of the 28 consensus compounds from STEBICEF in-house library.

 **AC185** **MW: 446.523 g/mol** **$logP_{o/w}$: 6.171**	 **ACM56** **MW: 459.565 g/mol** **$logP_{o/w}$: 5.934**
 **MF195** **MW: 429.539 g/mol** **$logP_{o/w}$: 4.413**	 **MF33** **MW: 403.458 g/mol** **$logP_{o/w}$: 3.569**
 **MF89** **MW: 390.459 g/mol** **$logP_{o/w}$: 4.166**	

3.3 Methods

3.3.1 Protein preparation

The PDB structures of $NLRP3_{PYD}$ (PDB ID: 2NAQ) and ASC_{PYD} (PDB ID: 3J63) were downloaded from PDB database [64] and were pre-processed by using "Protein Preparation Wizard" tool [69] by Schrödinger suite. Bond orders for untemplated residues and known HET groups were assigned. Hydrogens were added to the structures. Bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds were created. Water molecules beyond 5 Å from het groups were deleted. For ligands, cofactors and metals, het states were generated at pH 7.0 ± 2.0 using Epik [70]. Finally, H-bonds were optimised by using PROPKA [71] at pH 7.0.

3.3.2 Protein-protein docking using HADDOCK and computational alanine scanning using PPCHECK

HADDOCK software (v2.2) [65] was used to perform two protein-protein dockings between $NLRP3_{PYD}$ (PDB 2NAQ entry 6) and ASC_{PYD} (PDB 3J63). For the first docking, the following residues were defined as active: Glu13, Lys21, Leu25, Arg41, Leu45, Asp48 and Asp51 for ASC_{PYD}, and Arg7, Lys9 and Arg12 for $NLRP3_{PYD}$. For the second docking, $NLRP3_{PYD}$ active residues were the same used for the first docking, while for ASC_{PYD}, only residues involved in H2 and H3 were included as active, that is Lys21, Leu25, Arg41 and Leu45. For both protein-protein dockings, passive residues were automatically defined by the software by flagging the related box. The three selected protein-protein models were processed by using PPCheck software [66], whereas residues at the binding interface were detected, including the following ones for the models, and the computational alanine scanning was performed.

- Model 1
  ASC_{PYD} → Asp10, Ala11, Leu12, Glu13, Asn14, Leu20, Lys21, Lys24, Gly2, Arg3, Leu44, Leu45, Ser46, Met47, Asp48, Ala49, Leu50, Asp51, Leu52, Thr53, Asp54, Lys55, Val57, Arg5, Asp6, Ala7, Ile 8 and Leu9
  $NLRP3_{PYD}$ → Arg12, Tyr13, Glu15, Asp16, Leu17, Met3, Ala4, Ser5, His51, Thr6, Val52, Arg7, Cys8, Arg80, Lys9, Asp82, Leu83, Glu85, Lys86, Ala87, Lys88, Arg89, Asp90, Glu91, Leu10 and Ala10
- Model 2
  ASC_{PYD} → Asp10, Ala11, Leu12, Glu13, Asn14, Leu15, Thr16, Ala17, Glu18, Glu19, Leu20, Lys21, Lys22, Lys24, Leu25, Arg41, Gly42, Ala33, Leu44, Leu45, Ser46, Met47, Asp48, Ala49, Leu50, Asp51, Leu52 and Leu9
  $NLRP3_{PYD}$ → Arg12, Tyr13, Glu15, Asp16, Met3, Ala4, Ser5, Asp50, His51, Thr6, Val52, Asp53, Ala55, Thr56, Arg7, Cys8, Lys9, Arg89, Asp90, Leu10 and Ala11
- Model 3
  ASC_{PYD} → Thr16, ala17, Glu18, Glu19, Lys21, Lys22, Lys24, Leu25, Leu28, Pro40, Arg41, Gly42, Ala43, Leu44, Leu45 and Ser46

$NLRP3_{PYD}$ → Arg12, Tyr13, Glu15, Met3, Leu22, Ala4, Ser5, Gln45, Glu47, Lys48, Ala49, Asp50, His51, Thr6, Val52, Asp53, Leu54, Thr56, Arg7, Cys8, Lys9, Leu10 and Ala11

3.3.3 MD simulations of ASC_{PYD}-$NLRP3_{PYD}$ docking models

MD simulations were run for the three selected ASC_{PYD}-$NLRP3_{PYD}$ docking models, by using Schrödinger suite. The systems were first tuned through "System builder" tool. The solvent model TIP3P [72] and the orthorhombic box shape were selected. The box side distances were set 12 Å and the system was neutralized by adding Na^+ ions. Then these systems were used to run MD calculations [73] of 50 ns per each trajectory. Number of atoms, pressure and temperature were maintained constant, whereas pressure was set 1.01325 bar and temperature 300.0 K. Finally, the OPLS3 force field was set [74] and the model systems were relaxed before simulation.

3.3.4 MD trajectory clustering

In order to analyse the most stable and frequent interactions during the three MD simulations, it was necessary to cluster MD frames for the three models. For this purpose, "Desmond trajectory clustering" tool [75] released by Schrödinger suite was applied. RMSD matrix was based on protein backbone. The step frequency at which the frames were analysed was 10. The hierarchical cluster linkage method applied was average and the clusters to be generated were set by 10.

3.3.5 Ligand preparation for docking screening

PPI-targeted compound libraries were downloaded from Asinex, ChemDiv, Enamine and Life Chemicals databases, including α-helix mimetics, peptidomimetics, non-peptide peptidomimetics, shape-helix mimetics and PPI-enriched libraries together with an in-house library of chemical entities designed and synthesised by the Medicinal Chemistry group of STEBICEF Department (University of Palermo). The virtual libraries were filtered through KNIME platform [76] using the SMART alerts, in order to delete those compounds containing carcinogenic, mutagenic, chelating, reactive, unstable, toxic and skin sensitising groups [77]. All these compounds were prepared using "LigPrep" tool of Schrödinger suite. The selected force field was OPLS3 [74] and the protonation states were generated at pH 7.4 ± 0.2 using Epik [70]. The molecules were desalted and tautomers were generated retaining compound specific chirality. Finally, no more than 32 different conformations were generated per ligand.

3.3.6 Receptor grids generation on the three models and docking screenings

In order to perform ligand docking screening on the three $NLRP3_{PYD}$-ASC_{PYD} models, three grids were generated including constraints. The binding region was centred

by selecting $NLRP3_{PYD}$ interface residues per each model as reported below and ASC_{PYD} was deleted from the complex.

- Model 1 grid centred on Lys9, Arg12, Glu15, Lys86 and Arg89
- Model 2 grid centred on Ser5, Arg7, Lys9, Arg12, His51 and Val52
- Model 3 grid centred on Met3, Ser5, Cys8, Asp50, His51, Val52 and Asp52

For model 1 three H-bond donors were set as constraints on Lys9, Arg12 and Lys84 side chains; for model 2 three H-bond donors were defined on Arg7, His51 and Arg12 side chains; and for model 3 two H-bond acceptors were set on Asp48 and Asp51 side chains. The VdW radii scaling factor for non-polar atoms was set by 1.0 with partial charge cut-off 0.25. For all the grids the applied force field was OPLS3 [74]. Then, the docking screenings were performed by using "ligand docking" tool of Schrödinger suite [78, 79]. The selected protocol was standard precision and the selected ligand sampling method was flexible. Finally, the VdW radii scaling factor for non-polar atoms was set 0.8 with partial charge cut-off 0.15 and the constraints to be matched by ligands were set at least 2 per each model. All the other settings were maintained as default.

3.3.7 Ligands preparation for pharmacophore screening

In order to perform pharmacophore screening, the molecules retrieved from the three docking screenings were prepared to perform pharmacophore screenings. Hence, the related compound conformations were prepared through "Create screening database" tool of LigandScout software (version 4.3 - released by Inte:Ligand GmbH) [80–83], specifying "iCon Best" [84] as conformer generation type to create high-quality ligand conformations. All the other settings were applied as default.

3.3.8 Pharmacophore map creation and screening

The pharmacophore maps of the three $NLRP3_{PYD}$-ASC_{PYD} models were generated by importing the PDB structures provided by HADDOCK outputs and the selected most representative MD frames. For these structures, the pharmacophore maps were created and then merged according to each $NLRP3_{PYD}$-ASC_{PYD} model. Therefore, for model 1 five pharmacophores were generated on HADDOCK model 1 and MD frames 350, 800, 270 and 490. The first pharmacophore was composed by 18 features, the second by 13 features, the third by 14 features, the fourth by 15 features and the fifth by 12 features. These pharmacophores were merged, and the features corresponding to ASC Arg3 and Asp51 were interpolated with Arg5 and Asp54 features respectively, due to the close position, and the sphere tolerance was increased by 0.30 Å. Then, a hydrophobic feature on ASC Leu45 and a hydrogen bond acceptor feature on ASC Ser46 were deleted, because they were only present in a pharmacophore map. Finally, the features referring to ASC Arg3/Arg5, Asp6 and Leu50 were marked as optional. The resulting pharmacophore consisted of 13 features as shown in figure 3.13a.

For model 2, four pharmacophores were generated based by using HADDOCK protein-protein model and MD frames 930, 690 and 40. The first pharmacophore was composed of 15 features, the second by 7 features, the third by 8 features and the fourth by 12 features. The four 3D maps were merged and two hydrophobic features referring to ASC Leu50 were interpolated and the sphere tolerance was increased by 0.30 Å, as

well as for the two hydrogen bond acceptor features of ASC Asp48. Finally, the resulting pharmacophore map consisted of 10 features where three (hydrophobic on ASC Leu9, H-bond acceptor and negative ionisable features on ASC Glu18) were marked s optional.

For model 3, three pharmacophore maps were generated by using HADDOCK protein-protein model and MD frames 970 and 340. The first pharmacophore was composed by 7 features, the second by 9 features and the third by 11 features. By merging these pharmacophores the features corresponding to ASC Glu18 and Lys21 showed high spatial difference, hence two final pharmacophore maps were created by merging HADDOCK protein-protein model 3D map first with MD frame 970 pharmacophore and then with MD frame 340 pharmacophore. The first map was composed of 7 features, whereas one (feature on Leu45) was defined as optional, and the second consisted of the same features differently positioned. Even for this map, the hydrophobic feature on Leu45 was marked as optional.

Therefore, the above-described pharmacophore maps were used to perform the screening of compounds retrieved from docking outputs. For this purpose, the pharmacophore-fit was set as scoring function, for the screening mode all query features were matched except for maximum three pharmacophore features that could be omitted. Finally, for the retrieval mode, the best matching conformations were retained.

3.4 Conclusions

The work herein described allowed to build three different interaction models of $NLRP3_{PYD}$-ASC_{PYD} complex based on data collected from the literature and the applied computational techniques. Indeed, the Protein Data Bank [64] does not currently report an X-ray crystal structure of this PPI. Thus, these models were crucial to perform further computational studies and subsequently identify a set of compounds aimed at inhibiting the PPI under study. Overall, 193 unique chemical entities were selected according to the results from docking and pharmacophore screening. These molecules have been already purchased to perform biological assays at Ri.MED Foundation laboratories in order to test and investigate their potential activity against the $NLRP3_{PYD}$-ASC_{PYD} interaction. Their proved activity could represent a turning point for tackling several inflammatory and autoimmune diseases, such as ulcerative colitis, Crohn's disease [28–31] and multiple sclerosis [33]), affecting millions of people worldwide.

References – Chapter Three

1. Mariathasan S, Newton K, Monack DM, et al (2004) Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. Nature 430:213–218. https://doi.org/10.1038/nature02664
2. Broz P, Dixit VM (2016) Inflammasomes: mechanism of assembly, regulation and signalling. Nat Rev Immunol 16:407–420. https://doi.org/10.1038/nri.2016.58
3. Bertinaria M, Gastaldi S, Marini E, Giorgis M (2019) Development of covalent NLRP3 inflammasome inhibitors: Chemistry and biological activity. Arch Biochem Biophys 670:116–139. https://doi.org/10.1016/j.abb.2018.11.013
4. Guo H, Callaway JB, Ting JP-Y (2015) Inflammasomes: mechanism of action, role in disease, and therapeutics. Nat Med 21:677–687. https://doi.org/10.1038/nm.3893
5. Gaidt MM, Hornung V (2018) The NLRP3 Inflammasome Renders Cell Death Pro-inflammatory. J Mol Biol 430:133–141. https://doi.org/10.1016/j.jmb.2017.11.013
6. Jo E-K, Kim JK, Shin D-M, Sasakawa C (2016) Molecular mechanisms regulating NLRP3 inflammasome activation. Cell Mol Immunol 13:148–159. https://doi.org/10.1038/cmi.2015.95
7. Próchnicki T, Mangan MS, Latz E (2016) Recent insights into the molecular mechanisms of the NLRP3 inflammasome activation. F1000Research 5:1469. https://doi.org/10.12688/f1000research.8614.1
8. Dinarello CA, Simon A, van der Meer JWM (2012) Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. Nat Rev Drug Discov 11:633–652. https://doi.org/10.1038/nrd3800
9. He W, Wan H, Hu L, et al (2015) Gasdermin D is an executor of pyroptosis and required for interleukin-1β secretion. Cell Res 25:1285–1298. https://doi.org/10.1038/cr.2015.139
10. Yaron J (2015) Ion Flux Regulates Inflammasome Signaling
11. Martín-Sánchez F, Diamond C, Zeitler M, et al (2016) Inflammasome-dependent IL-1β release depends upon membrane permeabilisation. Cell Death Differ 23:1219–1231. https://doi.org/10.1038/cdd.2015.176
12. Rivers-Auty J, Brough D (2015) Potassium efflux fires the canon: Potassium efflux as a common trigger for canonical and noncanonical NLRP3 pathways. Eur J Immunol 45:2758–2761. https://doi.org/10.1002/eji.201545958
13. Muñoz-Planillo R, Kuffa P, Martínez-Colón G, et al (2013) K+ Efflux Is the Common Trigger of NLRP3 Inflammasome Activation by Bacterial Toxins and Particulate Matter. Immunity 38:1142–1153. https://doi.org/10.1016/j.immuni.2013.05.016
14. Bauernfeind FG, Horvath G, Stutz A, et al (2009) Cutting Edge: NF-κB Activating Pattern Recognition and Cytokine Receptors License NLRP3 Inflammasome Activation by Regulating NLRP3 Expression. J Immunol 183:787–791. https://doi.org/10.4049/jimmunol.0901363
15. Murakami T, Ockinger J, Yu J, et al (2012) Critical role for calcium mobilization in activation of the NLRP3 inflammasome. Proc Natl Acad Sci 109:11282–11287. https://doi.org/10.1073/pnas.1117765109
16. Green JP, Yu S, Martín-Sánchez F, et al (2018) Chloride regulates dynamic NLRP3-dependent ASC oligomerization and inflammasome priming. Proc Natl Acad Sci 115:E9371–E9380. https://doi.org/10.1073/pnas.1812744115
17. Lemasters JJ, Theruvath TP, Zhong Z, Nieminen A-L (2009) Mitochondrial calcium and the permeability transition in cell death. Biochim Biophys Acta - Bioenerg 1787:1395–1401. https://doi.org/10.1016/j.bbabio.2009.06.009
18. Camello-Almaraz C, Gomez-Pinilla PJ, Pozo MJ, Camello PJ (2006) Mitochondrial reactive oxygen species and Ca 2+ signaling. Am J Physiol Physiol 291:C1082–C1088. https://doi.org/10.1152/ajpcell.00217.2006
19. Csordás G, Hajnóczky G (2009) SR/ER–mitochondrial local communication: Calcium and ROS. Biochim Biophys Acta - Bioenerg 1787:1352–1362. https://doi.org/10.1016/j.bbabio.2009.06.004
20. Yang Y, Wang H, Kouadir M, et al (2019) Recent advances in the mechanisms of NLRP3 inflammasome activation and its inhibitors. Cell Death Dis 10:128. https://doi.org/10.1038/s41419-019-1413-8
21. Zhou R, Yazdi AS, Menu P, Tschopp J (2011) A role for mitochondria in NLRP3 inflammasome activation. Nature 469:221–225. https://doi.org/10.1038/nature09663
22. Sorbara MT, Girardin SE (2011) Mitochondrial ROS fuel the inflammasome. Cell Res 21:558–560. https://doi.org/10.1038/cr.2011.20
23. Heid ME, Keyel PA, Kamga C, et al (2013) Mitochondrial Reactive Oxygen Species Induces

NLRP3-Dependent Lysosomal Damage and Inflammasome Activation. J Immunol 191:5230–5238. https://doi.org/10.4049/jimmunol.1301490
24. Halle A, Hornung V, Petzold GC, et al (2008) The NALP3 inflammasome is involved in the innate immune response to amyloid-β. Nat Immunol 9:857–865. https://doi.org/10.1038/ni.1636
25. Py BF, Kim M-S, Vakifahmetoglu-Norberg H, Yuan J (2013) Deubiquitination of NLRP3 by BRCC3 Critically Regulates Inflammasome Activity. Mol Cell 49:331–338. https://doi.org/10.1016/j.molcel.2012.11.009
26. Zhang Z, Meszaros G, He W, et al (2017) Protein kinase D at the Golgi controls NLRP3 inflammasome activation. J Exp Med 214:2671–2693. https://doi.org/10.1084/jem.20162040
27. He Y, Hara H, Núñez G (2016) Mechanism and Regulation of NLRP3 Inflammasome Activation. Trends Biochem Sci 41:1012–1021. https://doi.org/10.1016/j.tibs.2016.09.002
28. Kanneganti T-D (2017) Inflammatory Bowel Disease and the NLRP3 Inflammasome. N Engl J Med 377:694–696. https://doi.org/10.1056/NEJMcibr1706536
29. Lazaridis L-D, Pistiki A, Giamarellos-Bourboulis EJ, et al (2017) Activation of NLRP3 Inflammasome in Inflammatory Bowel Disease: Differences Between Crohn's Disease and Ulcerative Colitis. Dig Dis Sci 62:2348–2356. https://doi.org/10.1007/s10620-017-4609-8
30. Pellegrini C, Antonioli L, Lopez-Castejon G, et al (2017) Canonical and Non-Canonical Activation of NLRP3 Inflammasome at the Crossroad between Immune Tolerance and Intestinal Inflammation. Front Immunol 8:. https://doi.org/10.3389/fimmu.2017.00036
31. Mridha AR, Wree A, Robertson AAB, et al (2017) NLRP3 inflammasome blockade reduces liver inflammation and fibrosis in experimental NASH in mice. J Hepatol 66:1037–1046. https://doi.org/10.1016/j.jhep.2017.01.022
32. Carlström M, Ekman A-K, Petersson S, et al (2012) Genetic support for the role of the NLRP3 inflammasome in psoriasis susceptibility. Exp Dermatol 21:932–937. https://doi.org/10.1111/exd.12049
33. Barclay W, Shinohara ML (2017) Inflammasome activation in multiple sclerosis and experimental autoimmune encephalomyelitis (EAE). Brain Pathol 27:213–219. https://doi.org/10.1111/bpa.12477
34. McCoy SS, Stannard J, Kahlenberg JM (2016) Targeting the inflammasome in rheumatic diseases. Transl Res 167:125–137. https://doi.org/10.1016/j.trsl.2015.06.006
35. Zhang Y, Zheng Y, Li H (2016) NLRP3 Inflammasome Plays an Important Role in the Pathogenesis of Collagen-Induced Arthritis. Mediators Inflamm 2016:1–9. https://doi.org/10.1155/2016/9656270
36. Mathews RJ, Robinson JI, Battellino M, et al (2014) Evidence of NLRP3-inflammasome activation in rheumatoid arthritis (RA); genetic variants within the NLRP3-inflammasome complex in relation to susceptibility to RA and response to anti-TNF treatment. Ann Rheum Dis 73:1202–1210. https://doi.org/10.1136/annrheumdis-2013-203276
37. Heneka MT, Kummer MP, Latz E (2014) Innate immune activation in neurodegenerative disease. Nat Rev Immunol 14:463–477. https://doi.org/10.1038/nri3705
38. Freeman LC, Ting JP-Y (2016) The pathogenic role of the inflammasome in neurodegenerative diseases. J Neurochem 136:29–38. https://doi.org/10.1111/jnc.13217
39. Debye B, Schmülling L, Zhou L, et al (2018) Neurodegeneration and NLRP3 inflammasome expression in the anterior thalamus of SOD1(G93A) ALS mice. Brain Pathol 28:14–27. https://doi.org/10.1111/bpa.12467
40. Heneka MT (2017) Inflammasome activation and innate immunity in Alzheimer's disease. Brain Pathol 27:220–222. https://doi.org/10.1111/bpa.12483
41. White CS, Lawrence CB, Brough D, Rivers-Auty J (2017) Inflammasomes as therapeutic targets for Alzheimer's disease. Brain Pathol 27:223–234. https://doi.org/10.1111/bpa.12478
42. Saresella M, La Rosa F, Piancone F, et al (2016) The NLRP3 and NLRP1 inflammasomes are activated in Alzheimer's disease. Mol Neurodegener 11:23. https://doi.org/10.1186/s13024-016-0088-1
43. Lee E, Hwang I, Park S, et al (2019) MPTP-driven NLRP3 inflammasome activation in microglia plays a central role in dopaminergic neurodegeneration. Cell Death Differ 26:213–228. https://doi.org/10.1038/s41418-018-0124-5
44. Zhou Y, Lu M, Du R-H, et al (2016) MicroRNA-7 targets Nod-like receptor protein 3 inflammasome to modulate neuroinflammation in the pathogenesis of Parkinson's disease. Mol Neurodegener 11:28. https://doi.org/10.1186/s13024-016-0094-3
45. McGettrick AF, O'Neill LAJ (2013) NLRP3 and IL-1 β in macrophages as critical regulators of

metabolic diseases. Diabetes, Obes Metab 15:19–25. https://doi.org/10.1111/dom.12169
46. Masters SL, Dunne A, Subramanian SL, et al (2010) Activation of the NLRP3 inflammasome by islet amyloid polypeptide provides a mechanism for enhanced IL-1β in type 2 diabetes. Nat Immunol 11:897–904. https://doi.org/10.1038/ni.1935
47. Lee H-M, Kim J-J, Kim HJ, et al (2013) Upregulated NLRP3 Inflammasome Activation in Patients With Type 2 Diabetes. Diabetes 62:194–204. https://doi.org/10.2337/db12-0420
48. Duewell P, Kono H, Rayner KJ, et al (2010) NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. Nature 464:1357–1361. https://doi.org/10.1038/nature08938
49. Grebe A, Hoss F, Latz E (2018) NLRP3 Inflammasome and the IL-1 Pathway in Atherosclerosis. Circ Res 122:1722–1740. https://doi.org/10.1161/CIRCRESAHA.118.311362
50. Kantono M, Guo B (2017) Inflammasomes and Cancer: The Dynamic Role of the Inflammasome in Tumor Development. Front Immunol 8:. https://doi.org/10.3389/fimmu.2017.01132
51. Karki R, Man SM, Kanneganti T-D (2017) Inflammasomes and Cancer. Cancer Immunol Res 5:94–99. https://doi.org/10.1158/2326-6066.CIR-16-0269
52. Toldo S, Abbate A (2018) The NLRP3 inflammasome in acute myocardial infarction. Nat Rev Cardiol 15:203–214. https://doi.org/10.1038/nrcardio.2017.161
53. Park HH, Lo Y-C, Lin S-C, et al (2007) The Death Domain Superfamily in Intracellular Signaling of Apoptosis and Inflammation. Annu Rev Immunol 25:561–586. https://doi.org/10.1146/annurev.immunol.25.022106.141656
54. Park HH (2011) Structural analyses of death domains and their interactions. Apoptosis 16:209–220. https://doi.org/10.1007/s10495-010-0571-z
55. Damiano J, Reed J (2004) CARD Proteins as Therapeutic Targets in Cancer. Curr Drug Targets 5:367–374. https://doi.org/10.2174/1389450043345470
56. Park HH (2012) PYRIN domains and their interactions in the apoptosis and inflammation signaling pathway. Apoptosis 17:1247–1257. https://doi.org/10.1007/s10495-012-0775-5
57. Kersse K, Verspurten J, Berghe T Vanden, Vandenabeele P (2011) The death-fold superfamily of homotypic interaction motifs. Trends Biochem Sci 36:541–552. https://doi.org/10.1016/j.tibs.2011.06.006
58. Su M-Y, Kuo C-I, Chang C-F, Chang C-I (2013) Three-Dimensional Structure of Human NLRP10/PYNOD Pyrin Domain Reveals a Homotypic Interaction Site Distinct from Its Mouse Homologue. PLoS One 8:e67843. https://doi.org/10.1371/journal.pone.0067843
59. Bae JY, Park HH (2011) Crystal Structure of NALP3 Protein Pyrin Domain (PYD) and Its Implications in Inflammasome Assembly. J Biol Chem 286:39528–39536. https://doi.org/10.1074/jbc.M111.278812
60. Stutz A, Kolbe C-C, Stahl R, et al (2017) NLRP3 inflammasome assembly is regulated by phosphorylation of the pyrin domain. J Exp Med 214:1725–1736. https://doi.org/10.1084/jem.20160933
61. Oroz J, Barrera-Vilarmau S, Alfonso C, et al (2016) ASC Pyrin Domain Self-associates and Binds NLRP3 Protein Using Equivalent Binding Interfaces. J Biol Chem 291:19487–19501. https://doi.org/10.1074/jbc.M116.741082
62. Vajjhala PR, Kaiser S, Smith SJ, et al (2014) Identification of Multifaceted Binding Modes for Pyrin and ASC Pyrin Domains Gives Insights into Pyrin Inflammasome Assembly. J Biol Chem 289:23504–23519. https://doi.org/10.1074/jbc.M114.553305
63. Vajjhala PR, Mirams RE, Hill JM (2012) Multiple Binding Sites on the Pyrin Domain of ASC Protein Allow Self-association and Interaction with NLRP3 Protein. J Biol Chem 287:41732–41743. https://doi.org/10.1074/jbc.M112.381228
64. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
65. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014
66. Sukhwal A, Sowdhamini R (2015) PPCheck: A Webserver for the Quantitative Analysis of Protein-Protein Interfaces and Prediction of Residue Hotspots. Bioinform Biol Insights 9:BBI.S25928. https://doi.org/10.4137/BBI.S25928
67. Sastry M, Lowrie JF, Dixon SL, Sherman W (2010) Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. J Chem Inf Model 50:771–784. https://doi.org/10.1021/ci100062n
68. Duan J, Dixon SL, Lowrie JF, Sherman W (2010) Analysis and comparison of 2D fingerprints:

Insights into database screening performance using eight fingerprint methods. J Mol Graph Model 29:157–170. https://doi.org/10.1016/j.jmgm.2010.05.008
69. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221–234. https://doi.org/10.1007/s10822-013-9644-8
70. Shelley JC, Cholleti A, Frye LL, et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691. https://doi.org/10.1007/s10822-007-9133-z
71. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
72. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
73. (2006) SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Association for Computing Machinery, New York, NY, USA
74. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864
75. Bowers KJ, Chow DE, Xu H, et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: ACM/IEEE SC 2006 Conference (SC'06). IEEE, pp 43–43
76. Berthold MR, Cebron N, Dill F, et al (2008) KNIME: The Konstanz Information Miner. pp 319–326
77. Sushko I, Salmina E, Potemkin VA, et al (2012) ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. J Chem Inf Model 52:2310–2316. https://doi.org/10.1021/ci300245q
78. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430
79. Halgren TA, Murphy RB, Friesner RA, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem 47:1750–1759. https://doi.org/10.1021/jm030644s
80. Wolber G, Langer T (2005) LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 45:160–169. https://doi.org/10.1021/ci049885e
81. Steindl TM, Schuster D, Wolber G, et al (2007) High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. J Comput Aided Mol Des 20:703–715. https://doi.org/10.1007/s10822-006-9066-y
82. Steindl TM, Schuster D, Laggner C, et al (2007) Parallel Screening and Activity Profiling with HIV Protease Inhibitor Pharmacophore Models. J Chem Inf Model 47:563–571. https://doi.org/10.1021/ci600321m
83. Krovat EM, Frühwirth KH, Langer T (2005) Pharmacophore Identification, in Silico Screening, and Virtual Library Design for Inhibitors of the Human Factor Xa. J Chem Inf Model 45:146–159. https://doi.org/10.1021/ci049778k
84. Poli G, Seidel T, Langer T (2018) Conformational Sampling of Small Molecules With iCon: Performance Assessment in Comparison With OMEGA. Front Chem 6:

CHAPTER FOUR

SPIKE RBD-ACE2 PD INTERACTION – Identification of a protein interacting region to select putative modulators

4.1 Introduction

In December 2019, a virus-transmitted flu epidemic spread out worldwide, whereas first cases of the severe acute respiratory syndrome caused by the novel coronavirus (SARS-CoV-2) were detected [1]. Due to a really high rate of virulence associated with morbidity and mortality (36.9 million infected people and more than 1 million deaths globally) [2] affecting 216 countries, in March 2020 COronaVIrus Disease 2019 (COVID-19) was considered a health emergency of international concern, thus it was stated as a pandemic by the World Health Organization (WHO) [3, 4]. Since the beginning of 2000s, coronaviruses already were associated with disease outbreaks: SARS-CoV emerged in Guangdong (China) in 2002 [5, 6], and MERS-CoV (Middle East respiratory syndrome coronavirus) affecting the Arabian Peninsula in 2012 [7, 8]. The nature of these viruses was originally zoonotic, but over the time they crossed the species barrier through bats, for SARS-CoV and SARS-CoV-2, and dromedary camels for MERS [9–11]. These viruses are grouped in four genera, whereas SARS-CoV and SARS-CoV-2 belong to the same β-CoV genus [12]. When aligned SARS-CoV and SARS-CoV-2 proteins exhibit a strong correlation with about 76% of sequence identity. Coronaviruses are positive-strand RNA viruses containing a membrane coated by Spike (S) glycoproteins that provide their characteristic crown aspect [13], as depicted in figure 4.1.

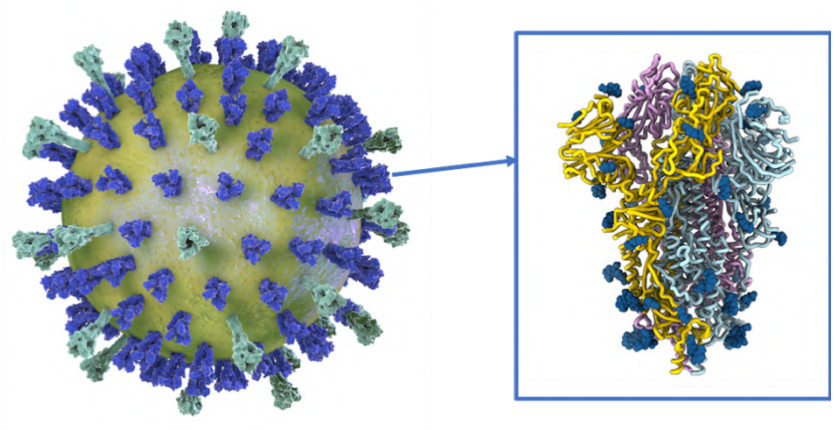


Figure 4.1. SARS-CoV-2 membrane coating through Spike glycoprotein [14, 15]

S glycoprotein promotes the interactions with the host cell by mediating receptor recognition and membrane fusion [16, 17]. S protein consists of two functional subunits, S_1 and S_2, which are non-covalently bound in the pre-fusion state. S_1 bears the receptor-binding domain (RBD), also termed as domain B, able to bind the host cell receptor, while S_2 subunit presents the fusion machinery and is responsible for the membrane fusion. During the viral infection procedure, S_1 binds the host cell receptor, thus S protein undergoes a cleavage event at the boundary site, between S_1 and S_2 subunits (S_1-S_2 cleavage site), and it is converted to the post-fusion conformation. After this first step, a

second cleavage occurs through the host protease at a different cleavage site, generating conformational changes in S_2 and causing membrane fusion [18–20]. The binding to the host receptor and the proteolytic events produce a synergistic effect inducing conformational changes and helping the coronavirus to enter the host cell [21]. The first entry step for both coronaviruses (SARS-CoV-2 and SARS-CoV) is mediated by the interaction with the host Angiotensin-converting enzyme 2 (ACE2).

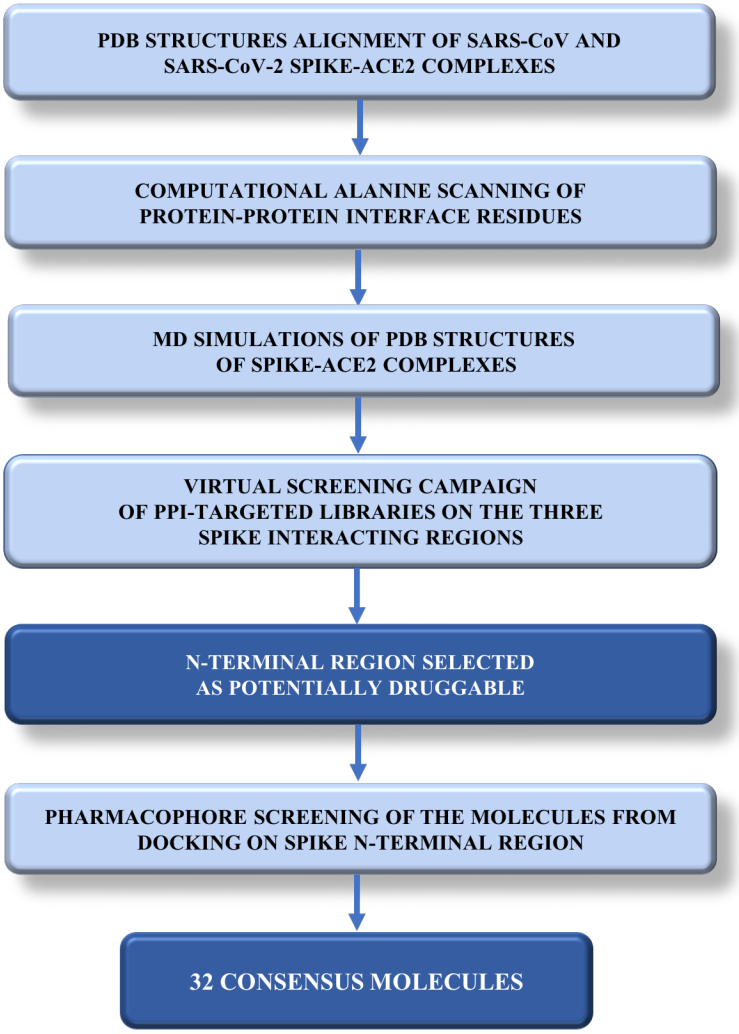
ACE2 is a protein essentially present in type II alveolar lung, oesophagus, heart and kidney cells [22, 23]. This protein contains an N-terminal peptidase domain (PD), that can bind the virion S glycoprotein, and a C-terminal Collectrin-like domain (CLD) [24, 25]. SARS-CoV-2 S protein establishes contacts with ACE2 at the surface of type II pneumocytes reporting similar affinity to SARS-CoV S glycoprotein [26, 27]. The crystal structures of both coronavirus types components shed light on the N-linked glycans overlay at the surface of S, and it was proposed as a mechanism developed by the viruses to elude the immune system [28, 29]. Indeed, in a first stage, the host immune system fails to recognise the pathogens due to the presence of S glycoproteins showing several sugar moieties covering the membrane. Thus, the immune system exchanges the virus with a normal sugar-coated host cell and does not attack it. In this way, viral S glycoproteins can bind the host cell receptor allowing the virus to enter [30]. The glycosylation region mainly refers to S_1 subunit which contains the RBD [31].

Based on this information, several research efforts have been done to tackle SARS-CoV-2 disease outbreak to identify potential therapeutics [21, 32–47]. The RBD region of SARS-CoV-2 could be a suitable region to address the design of novel therapeutics [48]. For the past SARS-CoV outbreak, S glycoprotein was indicated as the site of action of neutralising antibodies, such as S230 antibody, isolated from human survivors [49–51]. In 2019, Walls *et al*. showed that this antibody was able to prevent the virus-host receptor binding and caused a conformational modification to the fusion machinery with a ratcheting mechanism. The SARS-CoV S glycoprotein structure was characterised in complex with S230 in both closed and open states (PDB IDs: 6NB6 and 6NB7, respectively). The analysis of this antibody-S protein complex highlighted that the S230 epitope is located near Leu443 residue, whereas Tyr442 and Tyr475 are residues involved in the interaction. These amino acids have been shown to take part in the binding with ACE2 strengthening the hypothesis that S230 antibody works as a competitive inhibitor of coronavirus-host receptor interaction [52].

The work herein described has been published in the peer-reviewed journal *ChemMedChem* [53]. This project was based on a computational analysis of the interaction between S protein and ACE2 peptidase domain, in order to 1) identify a potential druggable area on the RBD interface, and 2) target this identified site with an *in silico* high-throughput screening campaign. The main goal was to select small molecules with the potential to modulate the virus entry. Putative drugs able to inhibit this PPI could work as fusion-blocking agents, representing an interesting strategy already employed for viruses outbreaks [54]. Indeed, inhibitors of the viral entry should prevent viral spreading and decrease the viral load at the very beginning of the infection.

It is noteworthy that the molecular contacts between S glycoprotein and ACE2 are widespread within the protein interfaces, hence the design of small molecule modulators represents a challenging strategy to inhibit the RBD domain. Furthermore, unlike the most common hydrophobic nature of PPIs, the S glycoprotein RBD-ACE2 molecular recognition is essentially mediated by hydrophilic interactions [55, 56]. To date, the design of monoclonal antibodies targeting the virus-human interface is the election

strategy together with the development of vaccines registering a robust preclinical and clinical pipeline [57, 58]. On the contrary, this work was based on a different strategy, that is a computational methodology consisting of 1) alanine scanning to explore the contribution of each amino acid to the protein-protein interactions, and identify potential druggable hot spots or binding pockets; 2) MD simulations to identify the most stable and frequent interactions between the two proteins; and 3) docking and pharmacophore screening campaigns to select potential hit compounds in order to provide the other research groups with useful tools for drug discovery programmes concerning COVID-19. All the steps herein described are summarised in Scheme 4.1.



Scheme 4.1. Overview of the computational workflow aimed at determining a potential binding region at S glycoprotein interface and identify putative modulators of S RBD-ACE2 PD interacting interface

4.1.1 Overlap of PDB structures highlighting open and closed states of S glycoprotein

The X-ray crystal structures currently present in the Protein Data Bank [59] highlight that SARS-CoV-2 S glycoprotein folds into a long trimeric complex with a triangular cross-section, including the receptor-binding motif (RBM), responsible for the recognition pattern of ACE2 [60]. The RBD domain consists of a core and an extensive

loop, the so-called RBM [61], and presents two different conformations, *i.e.* the *up-state* visible in PDB 6VYB and the *down-state* observable in PDB 6VXX.

In the first X-ray crystal structure, two of the three S chains show the down-state and the third one exhibits the up-state, whereas the other trimeric structure presents all S chains in down-state [62]. When the S glycoprotein is in up conformation, the RBD loop surrounds the ACE2 interface, while in the down conformation is buried into the interface between S_1 and S_2 subunits. The alignment of SARS-CoV-2 S trimeric structures in both open and closed states shed light on the exposition of the RBD loop on the virion surface to the host receptor. Furthermore, previous works on SARS-CoV and MERS-CoV unveiled that the open conformation of S protein is necessary to establish the interaction with ACE2, triggering the infection mechanism, and causing the conformational changes and the membrane fusion [62, 63]. Indeed, in the open state of S_1 loop, the residues of the receptor-binding motif are exposed to the solvent to bind ACE2 (figure 4.2), indicating that this opening process is required to form interactions with ACE2 PD.

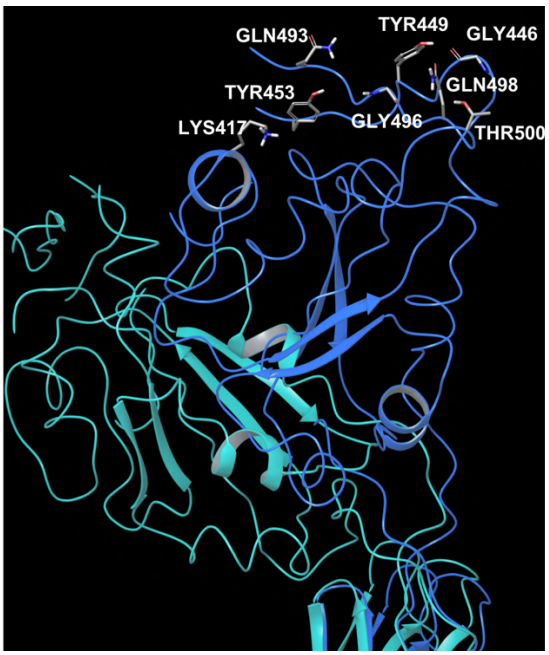


Figure 4.2. A close-up of SARS-CoV-2 PDB structures superposition that unveiled the RBD dynamic behaviour. The light blue structure shows PDB 6VXX S protein in the closed conformation, while the blue chain exhibits the open state of PDB 6VYB S protein [53].

Moreover, the superposition of SARS-CoV and SARS-CoV-2 X-ray crystal structures in open conformations revealed a similar dynamic behaviour for both B domains (figure 4.3), where the RBM loop exposes the key residues for the interaction with ACE2.

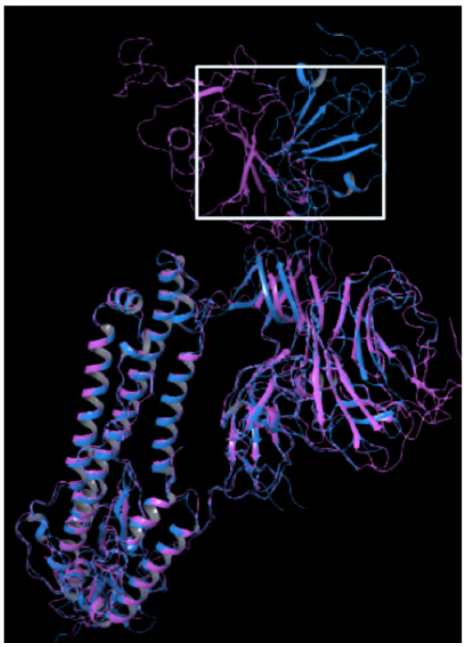
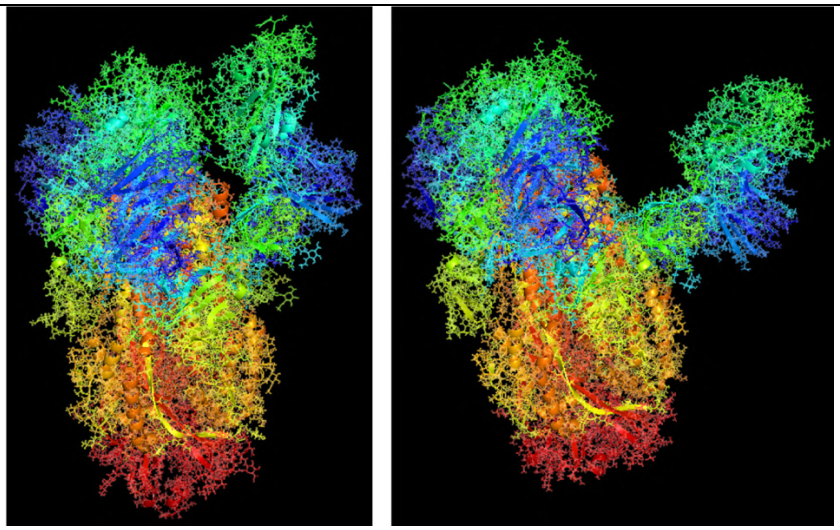


Figure 4.3. Superposition of SARS-CoV (purple chain, PDB 6ACD) and SARS-CoV-2 (blue chain, PDB 6VBY) RBDs, where both exhibit open conformations of the receptor-binding motifs [53].

This high flexibility detected in RBD from the above structural comparative analysis was extensively demonstrated by MD simulations performed by D. E. Shaw research group [64], that performed two MD simulations of 10 microseconds on PDB 6VXX in the closed conformation, and PDB 6VYB in the open state. The related trajectories were made available for the scientific community on the website. During the first MD simulation including the trimer in a closed conformation, RDB interface kept a buried state towards S_2 subunit. In the second MD simulation, RBD initially exhibited a partially open conformation (figure 4.4A), but after about 2 microseconds it exhibited a displacement, and finally drifted apart from S_2 subunit (figure 4.4B). These insights can provide crucial information about understanding the molecular events occurring as a first step of the interaction with host ACE2.



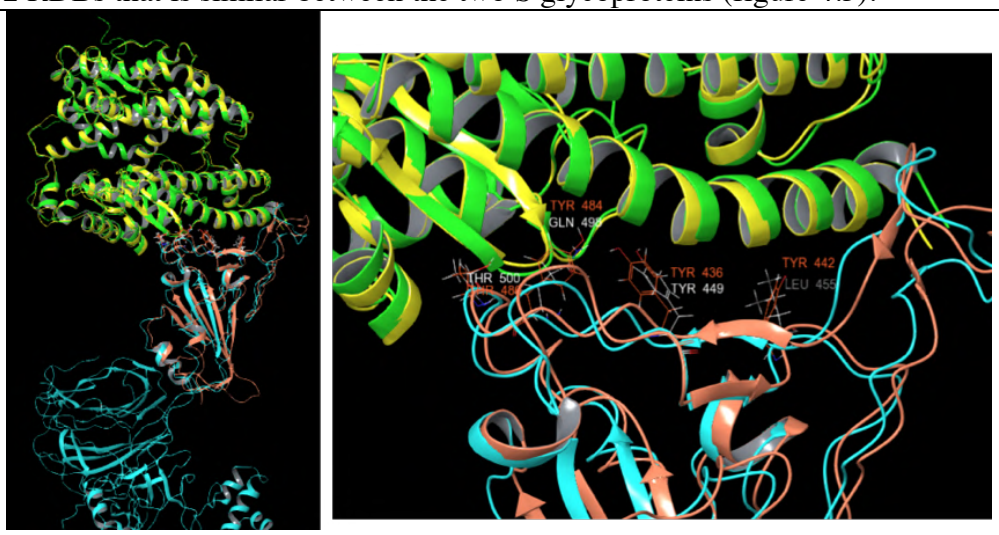
(A) **(B)**

Figure 4.4. **(A)** Frame 0 of the second MD simulation performed by D. E. Shaw group exhibiting the initial partially open state of RBD; **(B)** Frame 2003 corresponding to 2 μs of the same MD simulation presenting the open state of RBD [64]

4.2 Results and Discussion

4.2.1 Similarity analysis of SARS-CoV and SARS-CoV-2 S proteins

As above reported, SARS-CoV and SARS-CoV-2 S glycoproteins share 76% of amino acid sequence identity and 50% of identity within the RBM, in B domain [27]. This fact is also in accordance with the structural alignment of SARS-CoV and SARS-CoV-2 RDBs that is similar between the two S glycoproteins (figure 4.5).



(A) **(B)**

Figure 4.5. (A) Structure alignment of SARS-CoV (PDB 6CS2: light blue chain is Spike and yellow chain is ACE2) and SARS-CoV-2 (PDB 6M0J: orange chain is RBD Spike and green chain is ACE2); **(B)** Close-up on sequence alignment at ACE2-S proteins interface

Overall 18 residues of RDB region in SARS-CoV-2 take part in the interaction with ACE2. Amongst these, nine amino acids are equivalent in SARS-CoV and SARS-CoV-2 and include Tyr436-Tyr449, Tyr440-Tyr453, Asn473-Asn487, Tyr475-Tyr489, Gly482-Gly496, Thr486-Thr500, Gly488-Gly502, Tyr491-Tyr505, respectively. Five amino acids present side chains with similar physicochemical properties, that is Leu443-Phe456, Leu472-Phe486, Asn479-Gln493, Thr487-Asn501 and Tyr442-Leu455 (Table 4.1).

Table 4.1. Interfacial amino acids comparison of SARS-CoV and SARS-CoV-2 Spike protein. On the left, the conserved amino acids for both SARS-CoV and SARS-CoV-2 S proteins are reported; in the middle, residues with similar physicochemical properties for the two proteins are shown; and on the right, residues that differ between the two proteins are listed. Residues marked with * are involved in the interaction with ACE2.

CONSERVED AMINO ACIDS		SIMILAR AMINO ACIDS		DIFFERENT AMINO ACIDS	
SARS-CoV	SARS-CoV-2	SARS-CoV	SARS-CoV-2	SARS-CoV	SARS-CoV-2
Tyr436*	Tyr449	Leu443*	Phe456	Val404	Lys417
Tyr440*	Tyr453	Leu472*	Phe486	Thr433	Gly446
Asn473*	Asn487	Asn479*	Gln493	Pro462*	Ala475
Tyr475*	Tyr489	Thr487*	Asn501	Tyr484*	Gln498
Tyr481	Tyr495	Tyr442*	Leu455		
Gly482*	Gly496				
Thr486*	Thr500				
Gly488*	Gly502				
Tyr491*	Tyr505				

Even though the RBMs of both viruses are very similar, a few modifications on residue positions should influence the binding affinity between the two protein partners, S glycoprotein and ACE2. The main mutation affects Val404 residue in SARS-CoV, that is substituted with Lys417 in SARS-CoV-2, in the middle portion of RBD. The Lys417 side chain establishes a salt-bridge with the carboxyl group of ACE2 Asp30, probably producing a strengthening effect within the bimolecular interaction between SARS-CoV-2 RBD and ACE2. On the contrary, the amino acid Val404 does not produce interactions. The presence of Asn439 amino acid in SARS-CoV-2 as a replacement of Arg426 in SARS-CoV removes two salt-bridges with ACE2 Asn329 residue, probably weakening the protein-protein complex [63, 65]. However, several contact residues have high conservation thus explaining the overall similar binding affinity, as known from literature data, where SARS-CoV-2 exhibits K_D values of 1.2 nM while SARS-CoV 5 nM [62]. This comparison between residues can provide crucial information about the putative anchor amino acids for the interaction between RDB and ACE2 proteins.

4.2.2 Computational alanine scanning on SARS-CoV-2 – ACE2 interaction interface

The analysis of the interactions between ACE2 and SARS-CoV-2 S glycoprotein shed light on issues in the design of small molecule modulators of this PPI. It should be noticed that this PPI does not represent a traditional example of protein-protein interaction, where the protein interfaces are often shallow and do not exhibit deep pockets able to accommodate a canonical ligand. Moreover, the hot spot amino acids of the protein partners are usually mainly hydrophobic [66] and widely dislocated along the whole protein surfaces, creating a discontinuous epitope [67–69]. Indeed, analysing the currently available PDB structures of ACE2-S protein interaction (PDB IDs: 6M17 and 6M0J), the complex shows an interaction pattern with a 1:1 ratio, where the contacts between the proteins are essentially based on hydrogen bonds, some salt bridges and a few vdW forces. As this PPI is wide, three regions of interaction were defined, that is the N-terminal, the central and the C-terminal regions (figure 4.6A-C).

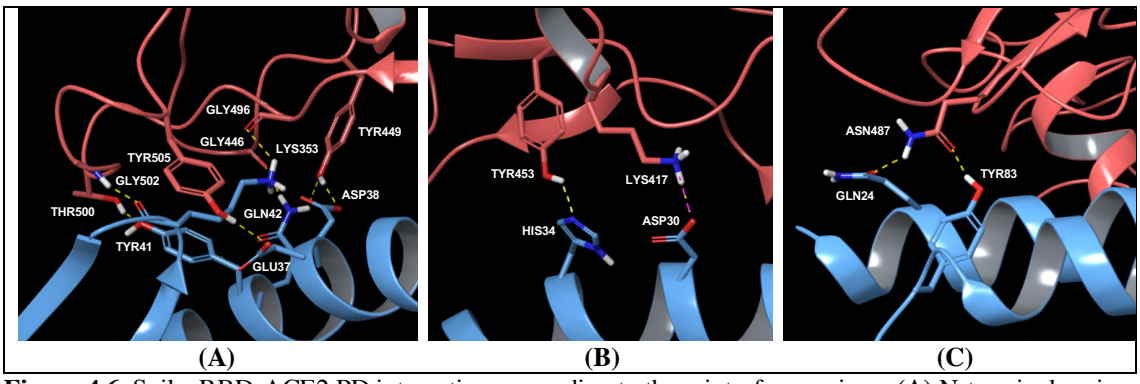


Figure 4.6. Spike RBD-ACE2 PD interactions according to three interface regions: **(A)** N-terminal region, **(B)** central region and **(C)** C-terminal region. PDB ID 6M0J – the light blue chain is ACE2 PD, while the orange chain is Spike RBD.

Both PDB structures share mostly the same crucial contacts between ACE2 PD and S glycoprotein. At the N-terminal region, the main interactions involve the following residues belonging respectively to ACE2 PD and SARS-CoV-2 RBD: Tyr41 and Thr500

hydroxyl groups establish a hydrogen bond; the backbone carbonyl of Lys353 interacts with Gly502 backbone NH group; Gln42 side chain makes contact with Gly446 carbonyl; Asp38 and Gln42 side chains contact with Tyr449 hydroxyl side chain; Glu37 side chain shows an interaction with Tyr505 hydroxyl group; and Lys353 side chain interacts with Gly496 backbone carbonyl group. In the central region of the interaction interface, three main interactions are reported: a hydrogen bond between the side chain of His34 and the aromatic hydroxyl side chain of Tyr453, while Asp30 of ACE2 peptidase domain establishes a salt bridge and hydrogen bond with Lys417. Finally, at the C-terminal region, ACE2 Gln24 and Tyr83 side chains form H-bond interactions with Asn487 side chain of S protein [65].

To quantitatively investigate the importance of these above-mentioned interacting amino acids at the protein interfaces, a computational alanine scanning was computed on both PDB structures. The complexes were prepared at pH 7.4 and the outputs were used to perform the alanine scanning calculation per each complex. After performing substitutions to alanine for those residues participating in the PPI interface, the tool provided $\Delta\Delta G_{affinity}$ values, measuring the difference between the calculated free energy of the mutated complex and the $\Delta G_{affinity}$ of the wild-type complex. Thus, positive $\Delta\Delta G_{affinity}$ values pointed out a decrease in protein-protein binding affinity for the complex, and provide information about the contribution of each amino acid to the binding affinity. According to Beard *et al.*, the results of the computational alanine scanning performed using Schrödinger suite have shown a connection with the experimental data, whereas a residue can be considered a hot spot if its mutation to alanine causes a $\Delta\Delta G_{affinity}$ increase over 3.0 kcal/mol [70]. Table 4.2 report the $\Delta\Delta G_{affinity}$ values for the residues belonging to both proteins that present values ≥ 3.0 kcal/mol subdivided according to the three interface regions (figure 4.7).

Table 4.2. Computational alanine scanning results including only ACE2 and SARS-CoV-2 S protein interface residues with $\Delta\Delta G_{affinity}$ values over 3 kcal/mol, for PDB 6M17 and PDB 6M0J according to the three interface regions (N-terminal, central and C-terminal portions). The last row of the table provides the total number of hot spots per region considering both PDB structures.

PDB ID: 6M17											
N-TERM				**MIDDLE**				**C-TERM**			
ACE2	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**	**ACE2**	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**	**ACE2**	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**
Tyr41	12.82	Thr500	7.65	His34	8.91	Phe456	4.80	Tyr83	9.28	Phe486	11.93
		Gly496	5.54			Leu455	4.26	Gln24	7.09	Tyr489	7.92
		Asn501	5.29			Gln493	4.23	Met82	5.19	Asn487	3.45
		Tyr505	3.88			Lys417	3.83				
		Gly502	3.81								
		Gln498	3.38								
PDB ID: 6M0J											
N-TERM				**MIDDLE**				**C-TERM**			
ACE2	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**	**ACE2**	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**	**ACE2**	**$\Delta\Delta G_{aff}$**	**Spike**	**$\Delta\Delta G_{aff}$**
Tyr41	15.18	Tyr505	9.51	Lys31	7.65	Lys417	10.44	Tyr83	8.86	Asn487	10.75
Gln42	7.70	Asn501	7.95	His34	4.18	Gln493	7.96	Gln24	7.67	Phe486	9.90
Lys353	7.03	Thr500	7.53			Leu455	6.73	Thr27	5.97	Tyr489	6.46
Asp38	4.95	Gln498	6.75			Phe456	6.42				
Glu37	4.35	Gly502	6.69								
		Gly496	3.75								
TOTAL HOT SPOTS = 11				**TOTAL HOT SPOTS = 6**				**TOTAL HOT SPOTS = 7**			

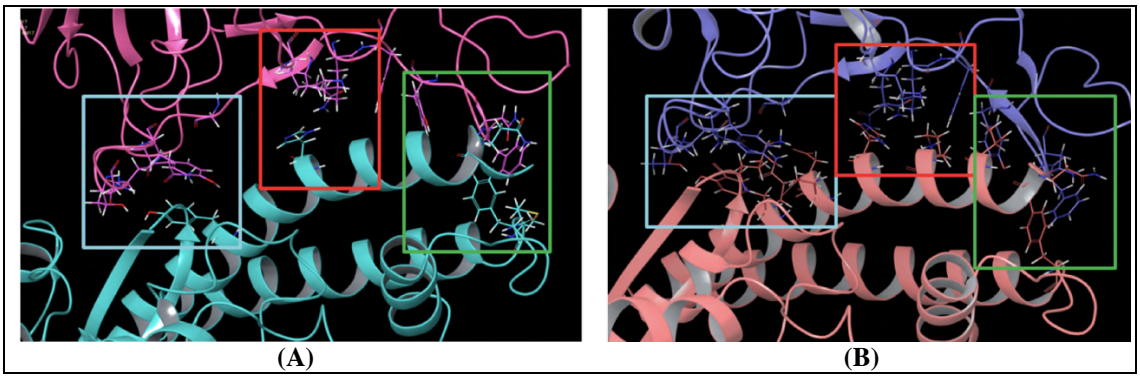


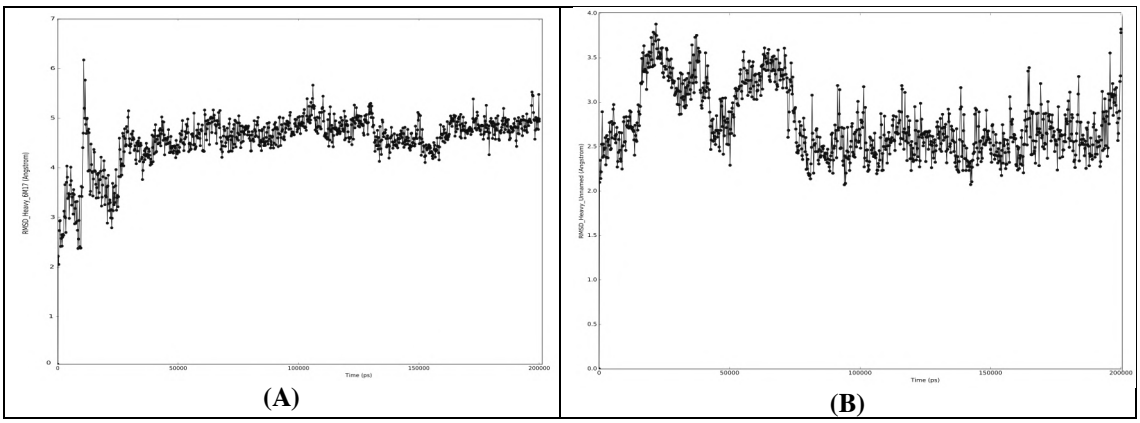
Figure 4.7. (A) Hot spots residues at the three Spike RBD-ACE2 PD interaction regions for PDB 6M17 (the dark pink chain is ACE2, while the light blue chain is S protein); **(B)** Hot spots residues at the three Spike RBD-ACE2 PD interactions regions for PDB 6M0J (the purple chain is ACE2, and the pink chain is S protein). In both pictures, the light blue square highlights the N-term, the red one the central region, and the green square indicates the C-term at ACE2-S protein interaction interface.

Analysing both alanine scanning results, the identified hot spots of S glycoproteins are shared between the two PDB structures, while for ACE2 some hot spot residues differ among the two PDBs. Therefore, only the common hot spots were considered to proceed with for the study, neglecting those non-shared hot spots. Moreover, the $\Delta\Delta G_{affinity}$ values were ranked in terms of contribution-to-binding, to discriminate the most important hot spots from the less relevant amino acids. In this context, for ACE2 peptidase domain (from PDB 6M17), the key residues were the following ones: Tyr41 within the recognition pattern at the N-terminal region, Tyr83 in the C-terminal region and His34 in the central region. Considering the other PDB structure 6M0J, the most valued hot spots were in the N-terminal region, involving Tyr41, Gln42, Gln24 and Lys353, and other two residues, Tyr83 and Lys31, that belonged to the C-terminal and middle regions, respectively. At the same time, the residues Phe486 and Tyr489 in the C-terminal region of S glycoprotein in PDB 6M17 were found crucial, and while the amino acids Thr500, Gly496, and Asn501 in the N-terminal region were mostly contributing to the ACE2-binding. On the other hand, from PDB 6M0J, the most valued residues were Phe486 and Asn487 belonging to the C-terminal region, Gln493 and Lys417 in the central region, and Thr500, Asn501 and Tyr505 in the N-terminal portion. In the light of the above, most of the identified hot spots for both proteins were mainly involved in the N-terminal portion (Table 4.2), reporting overall 11 hot spots. It suggests that this part of the protein-protein interface could be crucial for the PPI in comparison with the central and the C-terminal regions. Furthermore, the hydroxyl and aromatic side chain of Tyr41, the backbone carbonyl group of Lys353 and the side chain of ACE2 Gln42 appear to be the key recognition features in the interaction with S glycoprotein and may guide the selection process and design of novel RBD S small molecule inhibitors.

4.2.3 Molecular dynamics simulations on SARS-CoV-2 S protein in complex with ACE2

The above-described analysis provided crucial information about key interactions but only from a static point of view. Hence, in order to get more information about crucial contacts, the two PDB 6M17 and 6M0J of ACE2-Spike protein complex were used to perform two MD simulations of 200 nanoseconds per each by using Desmond [71] to

investigate interactions frequency and stability during the trajectories. For both MD simulations, the RMSD plots were generated to check the stability of the complexes during the simulation, and monitoring also energy, temperature and pressure of the systems. The stationary shape for PDB 6M17 was obtained at about 30 nanoseconds of simulation, while for PDB 6M0J it was reached at about 80 nanoseconds of the trajectory (Plot 4.1A-B).



Plot 4.1. **(A)** RMSD plot of MD on PDB 6M17 is depicted, where the system reaches the stationary shape after about 30 ns of simulation; **(B)** RMSD plot of MD on PDB 6M0J is showed, where the system achieves a stationary shape at about 80 ns.

The trajectories turned out to be reliable for further analysis. Thus, the frames of both trajectories were grouped into ten clusters for both MD simulations by applying average as a hierarchical cluster linkage method. According to the RMSD plots, only the frames referring to the stable trajectory portion – after 30 nanoseconds for PDB 6M17 and after 80 nanoseconds for PDB 6M0J – were considered to retrieve the most abundant and frequent interactions amongst the clusters, as reported in table 4.3. The analysis of the MD simulation on PDB 6M17 (cryo-EM) provided a fewer number of interactions than PDB 6M0J (X-ray). This should be ascribed to the different starting points for MD simulations and the different resolution methods employed for resolving the 3D structures, whereas PDB 6M17 is a cryo-EM complex while PDB 6M0J is an X-ray crystal structure. However, the two MD simulations present ten common interactions, that hence were considered the most important. Furthermore, for those different contacts, the involved residues were also highlighted for both MD. All this data was essentially in accordance with the information retrieved from literature and computational alanine scanning approaches. Therefore, these results were collected and processed for guiding the selection of putative modulators of ACE2-S protein interaction.

Table 4.3. On the left, key interactions between ACE2 PD and SARS-CoV-2 RBD from MD results on PDB 6M17; on the right, key residues involved into interactions between ACE2 PD and SARS-CoV-2 RBD from MD results on PDB 6M0J.

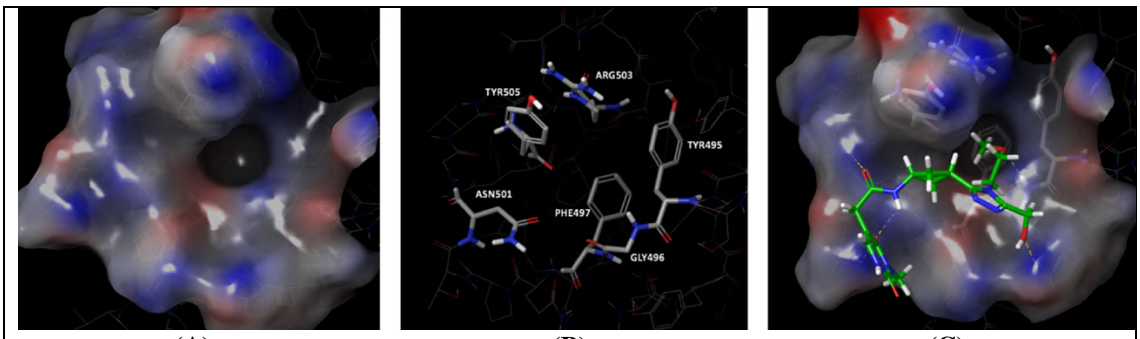
PDB ID: 6M17			PDB ID: 6M0J		
ACE2 PD	**Spike RBD**	**Interaction type**	**ACE2 PD**	**Spike RBD**	**Interaction type**
Gln24	Gln474	VdW	Gln24	Asn487	1 H-bond
Thr27	Phe456	VdW	Thr27	Phe456	VdW
Phe28	Tyr489	VdW	Phe28	Ty489	VdW
Asp30	Lys417	1 H-bond + 1 salt bridge	Phe28	Phe486	VdW
Lys31	Gln493	1 H-bond	Asp30	Lys417	1 H-bond + 1 salt bridge
His34	Tyr453	1 H-bond	Lys31	Gln493	1 H-bond
His34	Leu455	VdW	Lys31	Tyr489	VdW
Tyr41	Thr500	1 H-bond	His34	Tyr453	1 H-bond
Tyr83	Ala475	1 H-bond	His34	Leu455	VdW
Tyr83	Gly476	VdW	Tyr41	Thr500	VdW
Lys353	Gly502	1 H-bond	Tyr41	Gln498	VdW
Lys353	Asn501	VdW	Phe79	Gln486	VdW
Lys353	Tyr505	VdW	Tyr83	Asn487	VdW
			Tyr83	Phe486	VdW
			Lys353	Gly502	1 H-bond
			Lys353	Asn501	VdW
			Lys353	Tyr505	VdW

4.2.4 Supervised molecular docking to identify potential compounds able to bind N-terminal region

In the light of the previous data, a knowledge-based and data-driven docking screening was performed on at the N-terminal, central and C-terminal regions of ACE2 PD and SARS-CoV-2 RBD interacting interface. For this purpose, PDB 6M0J was used due to its better resolution (2.45 Å) compared to the PDB 6M17 (2.9 Å), and three different docking grids were built on S RBD, one per each interface region.

Two different compound libraries were used for docking screening, that is the in-stock MolPort library and a PPI-targeted compound library composed by Asinex, ChemDiv, Enamine and Life Chemicals databases. Owing to the large number of molecules, high-throughput virtual screenings were performed and the first 10,000 molecules prioritised were re-docked by applying docking standard precision (SP) through Schrödinger suite [72, 73]. The best 1,000 molecules of the docking outputs were selected according to those establishing the interactions retrieved from the literature, the computational alanine scanning, the MD simulations results and the docking scores. The analysis of these results highlighted that the N-terminal region of Spike RBD interface was able to accommodate ligands better than the other two regions, the middle and C-terminal ones. Indeed, when analysing docking outputs from these two regions, the compounds showed a high difference in binding poses among them, while the N-terminal region of S glycoprotein exhibited a small pocket able to accommodate functional groups

of the docked compounds. Interestingly, several ligands showed a complementarity with the RBD S pocket composed by the amino acids Tyr495, Gly496, Phe497, Asn501, Arg503 and Tyr505, as illustrated in figure 4.8. This data was used as a good starting point to deeply investigate this N-terminal region and allowed to consider it as the most potentially druggable region at the protein-protein interface compared to the other two. Therefore, compounds forming contacts with key amino acids at the N-terminal portion were taken into account for the next pharmacophore screenings.



(A) **(B)** **(C)**

Figure 4.8. RBD N-terminal binding region description. **(A)** Protein surface of N-terminal region; **(B)** Residues composing the cavity; **(C)** Example of ligand binding pose at N-terminal region.

4.2.5 Pharmacophore screening of selected compounds from docking screenings

In order to select a small representative group of the most promising compounds retrieved from docking outputs, the molecules were further processed through a pharmacophore screening. For this purpose, two different pharmacophore maps for PDB structures 6M17 and 6M0J were built on the N-terminal region of SARS-CoV-2 RBD-ACE2 PD interface. As above mentioned, the N-terminal region was detected as the most potentially druggable, thus the other two regions were neglected within this work. Hence, for PDB 6M17, the three-featured pharmacophore map was generated (figure 4.9A) including two hydrogen bond acceptors, that is one on Glu37 side chain of ACE2 PD and another on Lys353 backbone carbonyl, and a hydrogen-bond donor on Tyr41 side chain hydroxyl. On the other hand, for PDB 6M0J, the pharmacophore map consisted of six features (figure 4.9B), showing four hydrogen-bond acceptors on the backbone carbonyl of Lys353, carboxylic groups of Glu37 and Asp38 and the side chain hydroxyl of Tyr41, a hydrogen-bond donor on the amine side chain group of Gln42, and a negative ionisable feature on Asp38 side chain of ACE2 PD.

When comparing the two pharmacophore maps, it was observable that both PDB structures shared two comparable features corresponding to residues Lys353 and Glu37, while the features referring to the Tyr41 side chain hydroxyl were different. Indeed, in PDB 6M0J, the hydrogen of the hydroxyl group of Tyr41 accepted an H-bond from Thr500 side chain of S protein, while in PDB 6M17, the oxygen atom of the same hydroxyl group formed an H-bond to Asn501 side chain. It means that the same hydroxyl group of Tyr41 side chain could exhibit two different behaviours, H-bond acceptor or donor. Thus, the information provided by these pharmacophore maps were considered equally important, creating a shared pharmacophore (figure 4.9C), including overall the seven features from both PDB complexes.

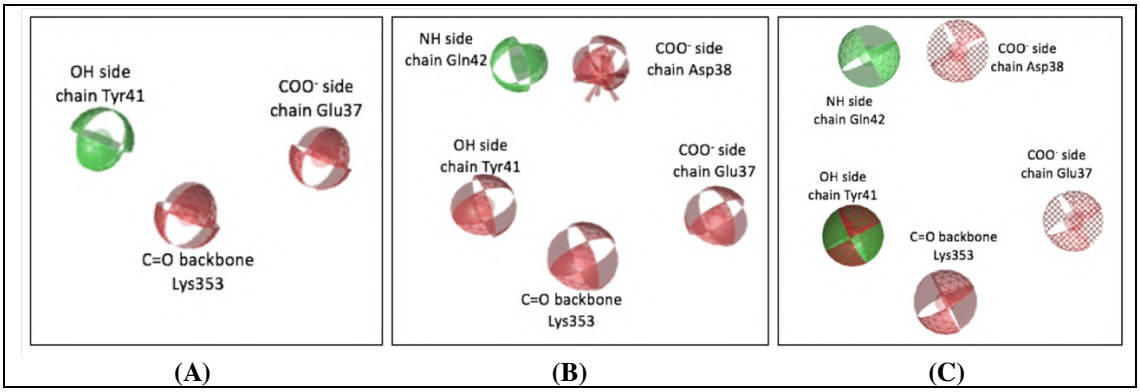
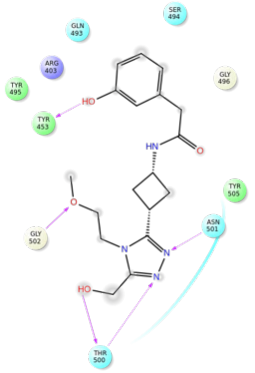
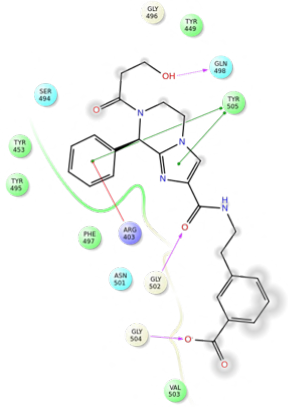
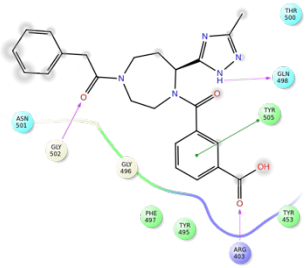
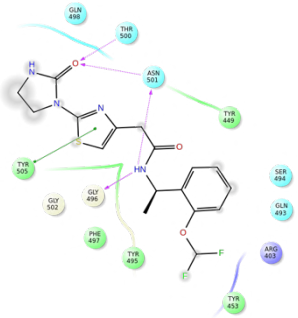
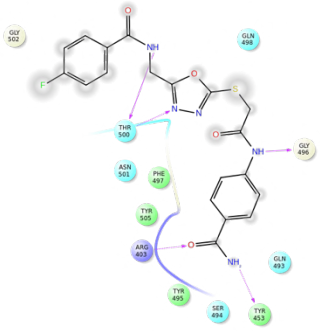
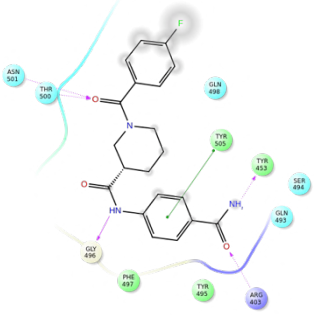


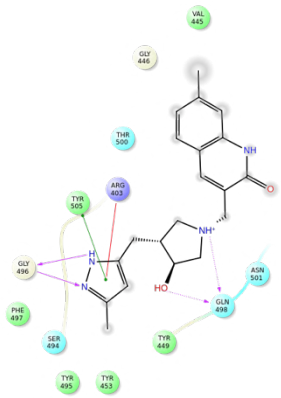
Figure 4.9. Pharmacophore maps built on RBD N-terminal region of **(A)** PDB 6M17 and **(B)** PDB 6M0J; and **(C)** shared pharmacophore map. Red spheres are hydrogen-bond acceptors, green spheres are hydrogen bond donors, the green-red sphere is both hydrogen-bond donor and acceptor, the red spike is a negative ionisable feature and dotted spheres are features marked as optional.

The high number of pharmacophore features may be too restrictive for a preliminary virtual screening, thus decreasing the possibility of identifying potentially promising compounds even not exactly matching all the pharmacophore features. For this reason, the resulting shared pharmacophore was modified considering the alanine scanning $\Delta\Delta G_{affinity}$ values, whereas Glu37 and Asp38 were included as the less valued hot spots compared to Tyr41, Gln42 and Lys353. Therefore, two H-bond acceptor features corresponding to Glu37 and Asp38 were marked as optional, because considered less important for the screening purposes, and the negative ionisable feature corresponding to Asp38 was deleted. Subsequently, the pharmacophore screening was run setting no omitted features allowed. From the outputs from the initial 1,000 PPI-targeted compound databases, 22 molecules were retrieved, while from the initial 1,000 of MolPort library 19 compounds were obtained.

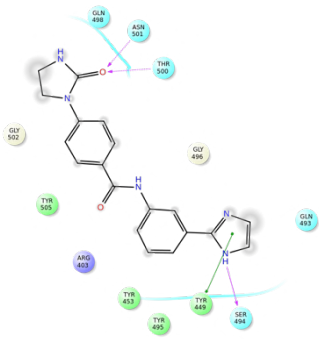
Analysing the outputs some resulting molecules were nucleoside analogues as a consequence of the highly hydrophilic nature of the binding site. However, these nucleoside analogues were neglected from the list of consensus molecules, because they were considered not relevant to enter a hit-to-lead optimization in a drug discovery process. On the contrary, 8 molecules were chosen as the most promising considering docking and pharmacophore outcomes, physicochemical properties and ease of chemical scaffold elaboration. Table 4.4 depicts the 2D structures of these 8 selected molecules that were able to establish interactions with key amino acids of S glycoprotein, while Table 4.5 illustrates the structures and some related physicochemical properties of the final 32 compounds identified from docking and pharmacophore approach. These compounds may provide crucial information about the modulation and the druggability of the N-terminal region. Moreover, they would be further validated via biophysical or biological screening before undergoes a hit optimisation programme to identify novel anti-COVID-19 therapeutics.

Table 4.4. Ligand interaction diagrams of the eight most promising compounds among the 32 consensus molecules selected according to docking binding poses and pharmacophore matching capacity.

SELECTED COMPOUNDS	
 **Spike_RM03** **$\Delta\Delta G_{binding}$ = - 58.259 kcal/mol**	 **Spike_RM14** **$\Delta\Delta G_{binding}$ = - 56.750 kcal/mol**
 **Spike_RM15** **$\Delta\Delta G_{binding}$ = - 53.986 kcal/mol**	 **Spike_RM25** **$\Delta\Delta G_{binding}$ = - 51.707 kcal/mol**
 **Spike_RM30** **$\Delta\Delta G_{binding}$ = -51.658 kcal/mol**	 **Spike_RM29** **$\Delta\Delta G_{binding}$ = - 50.333 kcal/mol**

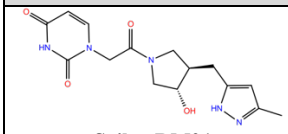
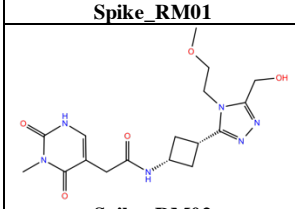
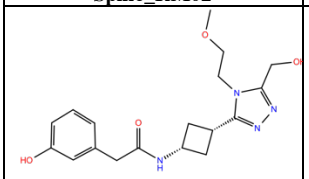
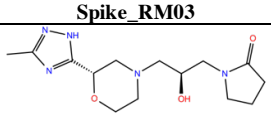
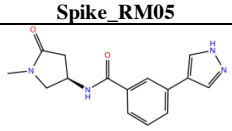


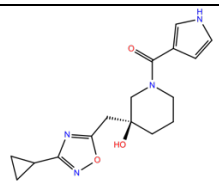
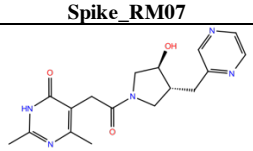
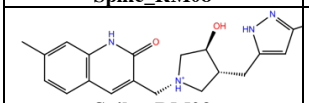
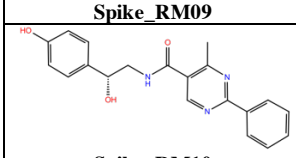
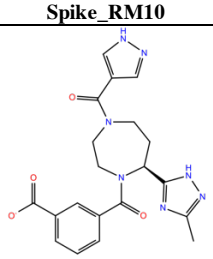
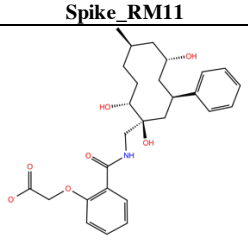
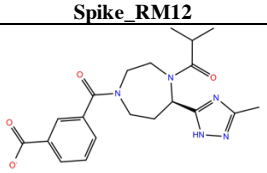
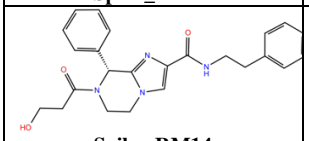
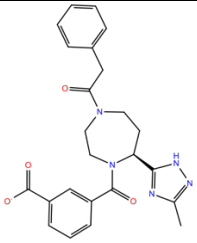
Spike_RM09
$\Delta\Delta G_{binding}$ = - 49.420 kcal/mol

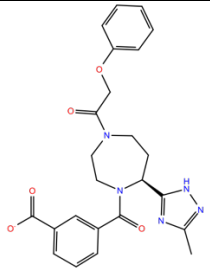
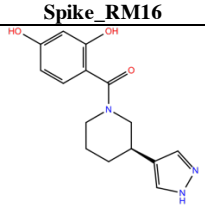
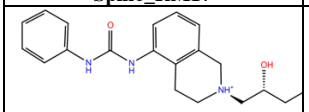
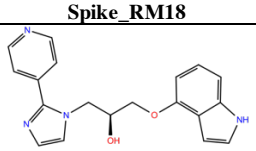
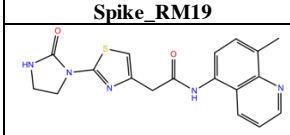
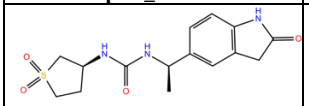
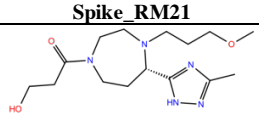
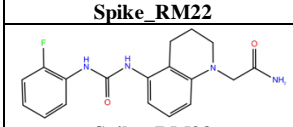
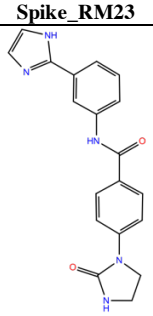
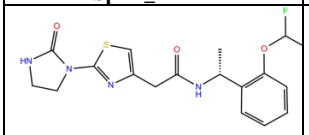


Spike_RM24
$\Delta\Delta G_{binding}$ = - 48.338 kcal/mol

Table 4.5. Physicochemical information of consensus molecules retrieved from docking and pharmacophore screenings performed on Spike RBD N-terminal region.

Consensus molecule	MW (g/mol)	cLogP[a]	PSA[a]	Rule of five violations	$\Delta G_{binding}$[b] (kcal/mol)	Docking score[a]	Pharmacophore-fit score[a]
 **Spike_RM01**	333.346	-0.554	153.327	0	-44.653	-5.318	58.730
 **Spike_RM02**	392.414	-0.237	170.248	1	-48.039	-4.460	58.580
 **Spike_RM03**	360.412	1.224	116.340	0	-58.259	-5.036	58.640
 **Spike_RM04**	309.367	-1.066	109.749	0	-43.861	-5.042	57.410
 **Spike_RM05**	324.186	-2.569	201.174	2	-36.930	-6.836	58.290
 **Spike_RM06**	284.317	0.737	99.295	0	-41.534	-5.562	56.110

 **Spike_RM07**	316.359	2.177	99.412	0	-44.845	-4.381	56.240
 **Spike_RM08**	343.385	0.027	124.934	0	-43.589	-4.290	58.700
 **Spike_RM09**	352.435	1.920	93.900	0	-49.420	-4.935	57.650
 **Spike_RM10**	349.388	2.717	97.396	0	-48.692	-4.690	55.830
 **Spike_RM11**	423.430	1.060	176.119	1	-46.155	-5.676	66.590
 **Spike_RM12**	485.576	3.518	134.235	0	-34.966	-4.281	58.070
 **Spike_RM13**	399.449	1.159	142.345	0	-41.774	-4.894	57.430
 **Spike_RM14**	462.504	2.397	152.169	0	-56.750	-5.320	57.600
 **Spike_RM15**	447.493	1.987	140.299	0	-53.986	-4.834	57.040

 **Spike_RM16**	463.492	1.894	155.876	0	-49.348	-5.003	57.260
 **Spike_RM17**	287.318	1.669	102.195	0	-42.492	-4.944	57.950
 **Spike_RM18**	391.472	2.600	86.270	0	-46.844	-4.568	57.670
 **Spike_RM19**	334.377	3.323	70.653	0	-44.880	-4.897	57.050
 **Spike_RM20**	367.425	2.704	100.627	0	-47.562	-4.494	56.300
 **Spike_RM21**	337.393	-0.002	132.701	0	-35.847	-5.296	65.190
 **Spike_RM22**	325.410	-0.027	99.560	0	-52.666	-5.164	58.490
 **Spike_RM23**	342.372	1.291	98.253	0	-45.305	-4.722	56.150
 **Spike_RM24**	347.376	2.676	106.796	0	-48.338	-5.290	56.090
 **Spike_RM25**	396.411	3.170	99.040	0	-51.707	-4.996	56.300

Spike_RM26	369.376	2.247	125.321	0	-48.653	-4.958	55.910
Spike_RM27	327.385	1.176	112.532	0	-41.896	-4.100	55.870
Spike_RM28	373.457	0.891	114.365	0	-52.603	-5.720	65.860
Spike_RM29	369.395	2.229	113.967	0	-50.333	-5.224	55.330
Spike_RM30	429.425	1.171	157.547	0	-51.658	-4.842	56.210
Spike_RM31	334.377	3.672	67.779	0	-42.773	-4.987	56.780
Spike_RM32	392.413	2.555	116.506	0	-49.262	-4.854	56.190

[a] Property computationally calculated; [b] Property computationally calculated as MM-GBSA.

4.3 Methods

4.3.1 Preparation of PDB complex structures of ACE2 PD–SARS-CoV-2 RBD

The PDB structures 6M17 and 6M0J including the protein complexes ACE2 PD–SARS-CoV-2 RBD were optimised using the "Protein preparation wizard" [74] tool (Schrödinger Release 2018-3) [72, 73, 75]. The bond orders for untemplated residues were assigned by using known HET groups based on their SMILES strings in Chemical

Component Dictionary. Hydrogens were added to the structure, eventual bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds between two sulfurs, if they were close to each other, were created and water molecules beyond 5.0 Å from any of the HET groups, including ions, were deleted. Then, protonation and metal charge states for the ligands, cofactors and metals were generated at pH 7.4 ± 0.2 using Epik [76, 77]. Finally, PROPKA [77] was run under pH 7.4 to optimise hydroxyl, Asn, Gln and His states using ProtAssign.

4.3.2 Computational Alanine Scanning on SARS-CoV-2 – ACE2 interaction interface

The residues of both PDB complexes were imported into the "Residue scanning" tool [70] released with Biologics suite to perform computational alanine scanning. The calculation type was flagged on "stability and affinity" to get $\Delta\Delta G_{affinity}$ values for each mutated residue. Only the residues from both proteins that take part in the interaction interfaces were selected for the substitutions to alanine, *i.e.* amino acids in positions 416, 417, 455, 456, 475 to 478 and 486 to 505 for S glycoprotein; and residues from position 21 to 48, from 79 to 83, and from 352 to 357 for ACE2 peptidase domain. Finally, the side-chains of the mutated residues were refined through a backbone minimization.

4.3.3 MD simulations on SARS-CoV-2 Spike protein in complex with ACE2

The optimised PDB structures 6M17 and 6M0J were also used for MD simulations through Desmond (released version 11.6) [71]. For both protein-protein complexes, a system was generated by the "System builder" tool. TIP3P [78] was chosen as a water solvent model and the orthorhombic box shape was selected to include the system. The simulation box size was calculated by using a buffer with 10 Å of distance between the solute structures and the simulation box boundary. Na^+ ions were added to neutralize both simulation boxes, and the applied force field was OPLS3 [79]. Then, the "Molecular Dynamics" tool was used to run MD simulations. The simulation time was set 200 nanoseconds for each system with a trajectory recording interval of 200 picoseconds, and the simulation seed was randomised. Finally, the number of atoms, pressure (1.01325 bar) and temperature (300 K) were maintained constant during the whole simulation. Subsequently, the MD outputs were processed to identify the most abundant and frequent interactions between SARS-CoV-2 RBD and ACE2 PD getting ten clusters per each trajectory. The backbone was chosen to set the RMSD matrix and frequency of clustering was 10, setting average as the hierarchical cluster linkage method.

4.3.4 Virtual compound libraries preparation for molecular docking screening

Several libraries of PPI-targeted compounds, such as Asinex, ChemDiv, Enamine, and Life Chemicals databases, and MolPort compound library were downloaded and prepared for the calculations. These virtual libraries were filtered through KNIME

platform [80] using the SMART alerts, in order to delete those compounds containing carcinogenic, mutagenic, chelating, reactive, unstable, toxic and skin sensitising groups [81], thus getting overall about 1.8 million molecules. Then, all the compounds were prepared using "LigPrep" tool of Schrödinger suite. The selected force field was OPLS3 [79] and the protonation states were generated at pH 7.4 ± 0.2 using Epik [82]. The molecules were desalted and tautomers were generated retaining compound specific chirality. Finally, no more than 32 different conformations were generated per ligand.

4.3.5 High-throughput virtual screening

Three different docking grids were generated by using the "Receptor grid generation" tool of Schrödinger suite on the three above mentioned interaction regions at ACE2 PD-SARS-CoV-2 RBD interface. The grid centroids were defined selecting the key amino acids according to the previously collected data from literature and computational results analysis: for N-terminal region Gly496, Gln498, Thr500, Asn501, Gly502, Tyr505, for the middle region Lys417, Leu455, Phe456, Gln493, and for the C-terminal portion Phe486, Asn487, Tyr489. The vdW radius scaling factor was set 1.0 for non-polar atoms with a partial charge cut-off of 0.25. Then, these grids were used to perform molecular docking screenings choosing a flexible protocol and the vdW radii of ligand non-polar atoms were scaled by 0.80 with partial atomic charge cut-off 0.15. The above-mentioned compound libraries were docked on these three grids, whereas high-throughput virtual screening workflows were run due to the large number of molecules for the two compound libraries. The first 10,000 compounds prioritised were re-docked using docking SP.

4.3.6 Pharmacophore screening of selected compounds from docking screening

The molecules selected from docking screening were optimised using the tool "Create screening database" of LigandScout software (version 4.3 - released by Inte:Ligand GmbH) [83–86], specifying "iCon Best" [87] as conformer generation type to create high-quality ligand conformations. The maximum number of conformations per compound was maintained 200 as default, and all other default settings were applied. Then, the pharmacophore maps for the screening were created by using PDB 6M17 and PDB 6M0J. Chain B of PDB 6M17 and chain A of PDB 6M0J corresponding to ACE2 protein were converted to ligands, in order to allow the software to define one of the two protein as a ligand. The two pharmacophore maps were generated using the "Create pharmacophore" tool, and they were transferred to the "Alignment perspective" window. From both pharmacophores those features not involved in the N-terminal region were deleted, getting three features for PDB 6M17 (figure 4.9A) and six features for PDB 6M0J (figure 4.9B). All hydrogen-bond vectors were converted into features to increase the ligand-matching capacity of pharmacophores. Then, a shared pharmacophore was generated (figure 4.9C), using the tool "Generate shared feature pharmacophore", setting 6M0J pharmacophore as reference. The resulting 3D map consisted of overall seven features, where a negative ionisable feature on Asp38 was deleted and the two H-bond acceptor features corresponding to Glu37 and Asp38 side chains were converted in

optional. This modified pharmacophore map was used to perform screening on the compound libraries previously generated. For this purpose, the scoring function was set "pharmacophore-fit", the screening mode was "match all query features", and for the retrieval mode was chosen "get best matching conformation". Finally, for the compound libraries, the maximum number of permitted omitted features was 0.

4.4 Conclusions

The aim of the work herein detailed was the identification of a potentially druggable region on the S glycoprotein RBD binding interface, and select potential hit molecules to enter drug discovery program against the current COVID-19 pandemic. Although targeting PPIs is a very challenging strategy in drug discovery programmes, in this work, using orthogonal computational techniques and investigating the S-ACE2 interaction interface, the N-terminal region of S RBD was identified as a druggable site showing a small pocket and highlighting the highest number of hot spot residues. Therefore, this region could be targeted as a therapeutic intervention point that may interfere with the host-guest recognition mechanism.

Finally, in this context, a supervised virtual screening was performed by applying a consensus strategy. Indeed, docking and pharmacophore screening yielded a list of potential modulators, that will be validated through biochemical, biophysical or cellular screening.

References – Chapter Four

1. Zhu N, Zhang D, Wang W, et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017
2. (2020) Coronavirus statics. https://www.worldometers.info/coronavirus/. Accessed 9 Oct 2020
3. Zhou P, Yang X Lou, Wang XG, et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7
4. Liu C, Zhou Q, Li Y, et al (2020) Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Cent Sci 6:315–331. https://doi.org/10.1021/acscentsci.0c00272
5. Drosten C, Günther S, Preiser W, et al (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348:1967–1976. https://doi.org/10.1056/NEJMoa030747
6. Hu B, Zeng L-P, Yang X-L, et al (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLOS Pathog 13:66–98. https://doi.org/10.1371/journal.ppat.1006698
7. Zaki AM, Van Boheemen S, Bestebroer TM, et al (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367:1814–1820. https://doi.org/10.1056/NEJMoa1211721
8. Sabir JSM, Lam TTY, Ahmed MMM, et al (2016) Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science (80-) 351:81–84. https://doi.org/10.1126/science.aac8608
9. Yang X-L, Hu B, Wang B, et al (2016) Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. J Virol 90:3253 LP – 3256. https://doi.org/10.1128/JVI.02582-15
10. Ge XY, Li JL, Yang X Lou, et al (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535–538. https://doi.org/10.1038/nature12711
11. Haagmans BL, Al Dhahiry SHS, Reusken CBEM, et al (2014) Middle East respiratory syndrome coronavirus in dromedary camels: An outbreak investigation. Lancet Infect Dis 14:140–145. https://doi.org/10.1016/S1473-3099(13)70690-X
12. Li F (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol 3:237–261. https://doi.org/10.1146/annurev-virology-110615-042301
13. Su S, Wong G, Shi W, et al (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol 24:490–502. https://doi.org/10.1016/j.tim.2016.03.003
14. The Native Antigen Company SARS-CoV-2. https://thenativeantigencompany.com/products/respiratory-syncytial-virus-a-glycoprotein-g-mouse-fc-tag/. Accessed 20 May 2020
15. UW Medicine Spike protein. https://newsroom.uw.edu/news/covid-19-coronavirus-spike-holds-infectivity-details. Accessed 20 May 2020
16. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM (2003) The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. J Virol 77:8801 LP – 8811. https://doi.org/10.1128/JVI.77.16.8801-8811.2003
17. Gallagher TM, Buchmeier MJ (2001) Coronavirus spike proteins in viral entry and pathogenesis. Virology 279:371–374. https://doi.org/10.1006/viro.2000.0757
18. Millet JK, Whittaker GR (2015) Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res 202:120–134. https://doi.org/10.1016/J.VIRUSRES.2014.11.021
19. Belouzard S, Chu VC, Whittaker GR (2009) Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci 106:5871–5876. https://doi.org/10.1073/PNAS.0809524106
20. Heald-Sargent T, Gallagher T (2012) Ready, set, fuse! the coronavirus spike protein and acquisition of fusion competence. Viruses 4:557–580. https://doi.org/10.3390/v4040557
21. Tortorici MA, Veesler D (2019) Structural insights into coronavirus entry. Adv Virus Res 105:93–116. https://doi.org/10.1016/BS.AIVIR.2019.08.002
22. Xu H, Zhong L, Deng J, et al (2020) High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci 12:1–5. https://doi.org/10.1038/s41368-020-0074-x
23. Zou X, Chen K, Zou J, et al (2020) Single-cell RNA-seq data analysis on the receptor ACE2

expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front Med 14:185–192. https://doi.org/10.1007/s11684-020-0754-0
24. Yan R, Zhang Y, Li Y, et al (2020) Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science 2762:1–10. https://doi.org/10.1126/science.abb2762
25. Prabakaran P, Xiao X, Dimitrov DS (2004) A model of the ACE2 structure and function as a SARS-CoV receptor. Biochem Biophys Res Commun 314:235–241. https://doi.org/10.1016/j.bbrc.2003.12.081
26. Li W, Zhang C, Sui J, et al (2005) Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J 24:1634–1643. https://doi.org/10.1038/sj.emboj.7600640
27. Wan Y, Shang J, Graham R, et al (2020) Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. J Virol 1–9. https://doi.org/10.1128/jvi.00127-20
28. Masters SL, Dunne A, Subramanian SL, et al (2010) Activation of the NLRP3 inflammasome by islet amyloid polypeptide provides a mechanism for enhanced IL-1β in type 2 diabetes. Nat Immunol 11:897–904. https://doi.org/10.1038/ni.1935
29. Wu K, Chen L, Peng G, et al (2011) A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. J Virol 85:5331–7. https://doi.org/10.1128/JVI.02274-10
30. Xiong X, Tortorici MA, Snijder J, et al (2018) Glycan Shield and Fusion Activation of a Deltacoronavirus Spike Glycoprotein Fine-Tuned for Enteric Infections. J Virol 92:1–16. https://doi.org/10.1128/JVI.01628-17
31. Watanabe Y, Allen JD, Wrapp D, et al (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. Science (80-) eabb9983. https://doi.org/10.1126/science.abb9983
32. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins Struct Funct Bioinforma n/a-n/a. https://doi.org/10.1002/prot.24065
33. Ekins S, Mottin M, Ramos PRPS, et al (2020) Déjà vu: Stimulating open drug discovery for SARS-CoV-2. Drug Discov Today 25:928–941. https://doi.org/10.1016/j.drudis.2020.03.019
34. Pandey P, Rane JS, Chatterjee A, et al (2020) Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. J Biomol Struct Dyn 1–11. https://doi.org/10.1080/07391102.2020.1796811
35. Walls AC, Tortorici MA, Frenz B, et al (2016) Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol 23:899–905. https://doi.org/10.1038/nsmb.3293
36. Yu X, Zhang S, Jiang L, et al (2015) Structural basis for the neutralization of MERS-CoV by a human monoclonal antibody MERS-27. Sci Rep 5:1–11. https://doi.org/10.1038/srep13133
37. Jiang L, Wang N, Zuo T, et al (2014) Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. Sci Transl Med 6:1–10. https://doi.org/10.1126/scitranslmed.3008140
38. Ying T, Prabakaran P, Du L, et al (2015) Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. Nat Commun 6:8223. https://doi.org/10.1038/ncomms9223
39. Elshabrawy HA, Coughlin MM, Baker SC, Prabhakar BS (2012) Human Monoclonal Antibodies against Highly Conserved HR1 and HR2 Domains of the SARS-CoV Spike Protein Are More Broadly Neutralizing. PLoS One 7:1–9. https://doi.org/10.1371/journal.pone.0050366
40. Gurwitz D (2020) Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug Dev Res ddr.21656. https://doi.org/10.1002/ddr.21656
41. Zhou Y, Hou Y, Shen J, et al (2020) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 6:14. https://doi.org/10.1038/s41421-020-0153-3
42. Li G, De Clercq E (2020) Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov 19:149–150. https://doi.org/10.1038/d41573-020-00016-0
43. Wu R, Wang L, Kuo H-CD, et al (2020) An Update on Current Therapeutic Drugs Treating COVID-19. Curr Pharmacol Reports. https://doi.org/10.1007/s40495-020-00216-7
44. Martinez MA (2020) Compounds with Therapeutic Potential against Novel Respiratory 2019 Coronavirus. Antimicrob Agents Chemother 64:. https://doi.org/10.1128/AAC.00399-20
45. Zhou G, Zhao Q (2020) Perspectives on therapeutic neutralizing antibodies against the Novel Coronavirus SARS-CoV-2. Int J Biol Sci 16:1718–1723. https://doi.org/10.7150/ijbs.45123
46. Dai W, Zhang B, Su H, et al (2020) Structure-based design of antiviral drug candidates targeting

the SARS-CoV-2 main protease. Science (80-) eabb4489. https://doi.org/10.1126/science.abb4489
47. Wu C, Liu Y, Yang Y, et al (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B 10:766–788. https://doi.org/10.1016/j.apsb.2020.02.008
48. Li Y, Wan Y, Liu P, et al (2015) A humanized neutralizing antibody against MERS-CoV targeting the receptor-binding domain of the spike protein. Cell Res 25:1237–1249. https://doi.org/10.1038/cr.2015.113
49. Du L, He Y, Zhou Y, et al (2009) The spike protein of SARS-CoV - A target for vaccine and therapeutic development. Nat Rev Microbiol 7:226–236. https://doi.org/10.1038/nrmicro2090
50. Traggiai E, Becker S, Subbarao K, et al (2004) An efficient method to make human monoclonal antibodies from memory B cells: Potent neutralization of SARS coronavirus. Nat Med 10:871–875. https://doi.org/10.1038/nm1080
51. Corti D, Zhao J, Pedotti M, et al (2015) Prophylactic and postexposure efficacy of a potent human monoclonal antibody against MERS coronavirus. Proc Natl Acad Sci U S A 112:10473–10478. https://doi.org/10.1073/pnas.1510199112
52. Walls AC, Xiong X, Park YJ, et al (2019) Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. Cell 176:1026–1039. https://doi.org/10.1016/j.cell.2018.12.028
53. Gulotta MR, Lombino J, Perricone U, et al (2020) Targeting SARS-CoV-2 RBD Interface: a Supervised Computational Data-Driven Approach to Identify Potential Modulators. ChemMedChem cmdc.202000259. https://doi.org/10.1002/cmdc.202000259
54. Zhou Y, Simmons G (2012) Development of novel entry inhibitors targeting emerging viruses. Expert Rev Anti Infect Ther 10:1129–1138. https://doi.org/10.1586/eri.12.104
55. Chanphai P, Bekale L, Tajmir-Riahi HA (2015) Effect of hydrophobicity on protein–protein interactions. Eur Polym J 67:224–231. https://doi.org/10.1016/j.eurpolymj.2015.03.069
56. Bettinetti L, Magnani M, Padova A (2013) Drug Discovery by Targeting Protein–Protein Interactions. In: Disruption of Protein-Protein Interfaces. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–29
57. Mustafa S, Balkhy H, Gabere MN (2018) Current treatment options and the role of peptides as potential therapeutic components for Middle East Respiratory Syndrome (MERS): A review. J Infect Public Health 11:9–17. https://doi.org/10.1016/j.jiph.2017.08.009
58. Amanat F, Krammer F (2020) SARS-CoV-2 Vaccines: Status Report. Immunity 52:583–589. https://doi.org/10.1016/j.immuni.2020.03.007
59. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
60. Lan J, Ge J, Yu J, et al (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215–220. https://doi.org/10.1038/s41586-020-2180-5
61. Li F, Li W, Farzan M, Harrison SC (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–8. https://doi.org/10.1126/science.1116480
62. Walls AC, Park Y-J, Tortorici MA, et al (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 1–12. https://doi.org/10.1016/j.cell.2020.02.058
63. Kirchdoerfer RN, Wang N, Pallesen J, et al (2018) Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Sci Rep 8:1–11. https://doi.org/10.1038/s41598-018-34171-7
64. (2020) Molecular Dynamics Simulations Related to SARS-CoV-2. https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/. Accessed 22 Apr 2020
65. Yan R, Zhang Y, Li Y, et al (2020) Structure of dimeric full-length human ACE2 in complex with B0AT1. bioRxiv 2020.02.17.951848. https://doi.org/10.1101/2020.02.17.951848
66. Mekni N, De Rosa M, Cipollina C, et al (2019) In Silico Insights towards the Identification of NLRP3 Druggable Hot Spots. Int J Mol Sci 20:4974. https://doi.org/10.3390/ijms20204974
67. Perricone U, Gulotta MR, Lombino J, et al (2018) An overview of recent molecular dynamics applications as medicinal chemistry tools for the undruggable site challenge. Medchemcomm 9:920–936. https://doi.org/10.1039/C8MD00166A
68. Cukuroglu E, Engin HB, Gursoy A, Keskin O (2014) Hot spots in protein–protein interfaces: Towards drug discovery. Prog Biophys Mol Biol 116:165–173. https://doi.org/10.1016/j.pbiomolbio.2014.06.003
69. Metz A, Ciglia E, Gohlke H (2012) Modulating Protein-Protein Interactions: From Structural

Determinants of Binding to Druggability Prediction to Application. Curr Pharm Des 18:4630–4647. https://doi.org/10.2174/138161212802651553
70. Beard H, Cholleti A, Pearlman D, et al (2013) Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. PLoS One 8:e82849. https://doi.org/10.1371/journal.pone.0082849
71. Bowers KJ, Chow DE, Xu H, et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: ACM/IEEE SC 2006 Conference (SC'06). IEEE, pp 43–43
72. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430
73. Halgren TA, Murphy RB, Friesner RA, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem 47:1750–1759. https://doi.org/10.1021/jm030644s
74. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221–234. https://doi.org/10.1007/s10822-013-9644-8
75. Friesner RA, Murphy RB, Repasky MP, et al (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. J Med Chem 49:6177–6196. https://doi.org/10.1021/jm051256o
76. Greenwood JR, Calkins D, Sullivan AP, Shelley JC (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comput Aided Mol Des 24:591–604. https://doi.org/10.1007/s10822-010-9349-1
77. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
78. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
79. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864
80. Berthold MR, Cebron N, Dill F, et al (2008) KNIME: The Konstanz Information Miner. pp 319–326
81. Sushko I, Salmina E, Potemkin VA, et al (2012) ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. J Chem Inf Model 52:2310–2316. https://doi.org/10.1021/ci300245q
82. Shelley JC, Cholleti A, Frye LL, et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691. https://doi.org/10.1007/s10822-007-9133-z
83. Wolber G, Langer T (2005) LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 45:160–169. https://doi.org/10.1021/ci049885e
84. Steindl TM, Schuster D, Wolber G, et al (2007) High-throughput structure-based pharmacophore modelling as a basis for successful parallel virtual screening. J Comput Aided Mol Des 20:703–715. https://doi.org/10.1007/s10822-006-9066-y
85. Steindl TM, Schuster D, Laggner C, et al (2007) Parallel Screening and Activity Profiling with HIV Protease Inhibitor Pharmacophore Models. J Chem Inf Model 47:563–571. https://doi.org/10.1021/ci600321m
86. Krovat EM, Frühwirth KH, Langer T (2005) Pharmacophore Identification, in Silico Screening, and Virtual Library Design for Inhibitors of the Human Factor Xa. J Chem Inf Model 45:146–159. https://doi.org/10.1021/ci049778k
87. Poli G, Seidel T, Langer T (2018) Conformational Sampling of Small Molecules With iCon: Performance Assessment in Comparison With OMEGA. Front Chem 6:

CHAPTER FIVE

C3b-FH INTERACTION – Structural insights of mutated C3b protein involved in Age-Macular Degeneration disease

5.1 Introduction

The immune response is the first-line defence against a wide variety of different microorganisms including viruses, bacteria, parasites and fungi, which could cause serious problems to the health of the host organism if not cleared from the body. Patients partially immunodeficient can undergo severe infectious issues despite significant advances in supportive care [1].

There are two distinct types of the immune response, the innate and the adaptive, which aim to fight and annihilate pathogens. Among several components of the innate immune response, there are physical barriers such as the skin and mucous membranes, immune cells like neutrophils, macrophages, and monocytes, and soluble factors including cytokines and complement [2].

The complement is a group of sequentially reacting proteins that are critical components of the innate immune response. Unlike the adaptive immunity, which needs several days to trigger an effective immune response, upon infection or injury complement takes off immediately against pathogens. Furthermore, complement does not exhibit "immunological memory", a key feature of the adaptive immune system. On the contrary, complement is the main actor of the acute phase response aiming to destroy pathogens, reduce innate immune-mediated damage to the host tissues and foster a prompt restoration of the homeostasis [3]. It is widespread produced in essentially all body fluids, including tears, saliva and cerebrospinal fluid to promptly intervene everywhere in the body and prevent the infection [4, 5].

The role of complement is targeting pathogens for removal by covalent attachment of activation fragments to the cell membrane, chemoattracting and fostering to activate phagocytic cells, such as neutrophils and macrophages, to engulf pathogens and lyse them through the formation of the membrane attack complex (MAC).

Complement proteins can be divided into two broad categories: those involved in the activation, and those regulating the activation. On the other hand, the regulatory components can be subdivided into those that are soluble and those that are membrane bound. The complement activators can be grouped according to the main three activation pathways listed below:

a) Classical pathway – It is activated by complexes formed by antibodies bound to bacteria and other pathogens interacting with C1 protein;
b) Lectin pathway – It is activated by pathogen-specific carbohydrates (*e.g.* terminal D-mannose residues or *N*-acetyl-D-glucosamine) on the surface of some bacteria, fungi, viruses, parasitic protozoans, and by acetylated groups (*e.g.* *N*-acetyl-glycine, *N*-acetyl-cysteine and acetylcholine) [6, 7];
c) Alternative pathway (AP) – It provides protection from a wide variety of extracellular pathogens and is activated by pathogen-associated molecular patterns (PAMPS), chemical moieties unique to pathogens not found in the host,

such as lipopolysaccharide at the cell wall of Gram-negative bacteria and carbohydrates found in the cell wall of fungi (zymosan) [8–12]. Unlike the classical and lectin pathways, the alternative pathway has been shown to maintain an active state at a small degree all the time by a mechanism known as "tickover". Thus, the alternative pathway is always primed to respond quickly and vigorously to pathogens or injury [13–17].

Once activated, the alternative pathway triggers a cascade of cleavage events, where each proenzyme forms are cleaved to their active forms and subsequently activate the next protein in the pathway and so on. The cleavage events share the same result, *i.e.* the generation of two multi-molecular enzyme complexes termed convertases. These ones cleave either C3 or C5 proteins, the two main proteins of the complement system, producing the cleaved fragments named based on their molecular size: the larger fragment is termed "*b*" fragment and the smaller fragment is named "*a*" fragment. For example, C3 cleavage by the C3 convertase generates the smaller C3a and larger C3b fragments. These fragments mediate several host defence functions, as follows [4, 5]:

a) Chemoattraction – C3a and C5a fragments are potent chemoattractants, that is they stimulate and increase the phagocytic and killing activity of myeloid cells, such as neutrophils and macrophages, and other cell types, by recruiting them to the site of complement activation in a concentration-dependent manner [18–20];
b) Opsonisation – C3b and C4b are potent opsonins, that are able to target and mark pathogens for phagocytosis by covalently binding to the pathogen cell surface through a thioester moiety. Before complement activation, the α-chains of C3 contain a thioester bond, that is stable until the protein is cleaved to C3b. The thioester bond is then highly metastable and it can interact with hydroxyl groups on carbohydrates or amine groups on the surface of the invading pathogens, generating ester or amide linkages, respectively (figure 5.1) [14]. Thus, C3b can coat the surface of invading pathogens at a high density by masking and neutralize virulence factors required for interaction with host cells [21, 22].

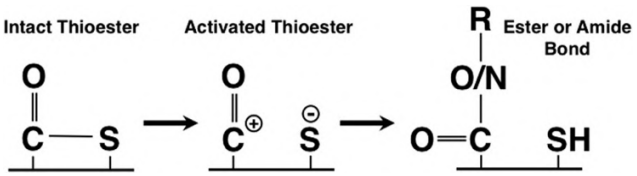


Figure 5.1. Thioester bond of C3 protein. Before alternative pathway activation, it is intact. Upon activation, it exhibits a metastable form, that interacts with nucleophile species to generate ester or amide linkages on pathogens surface [23].

c) Cell lysis – Another function of complement is the formation of the membrane attack complex by associating C5b protein, in a concentration-dependent manner, with C6, C7, C8 and multiple C9 proteins [24–26]. MAC inserts into lipid bilayers of susceptible bacteria, viruses, and parasites, leading to cellular dysfunction and osmotic lysis to disrupt cellular integrity [27].

However, an indiscriminate deposition of complement and an overactivation of the innate immune response may induce damages to neighbouring cells, causing several health issues such as the deterioration of vision in age-related macular degeneration

(AMD) and loss of long-term memory in Alzheimer's disease [28, 29]. To avoid self-injury, complement activation is controlled. An important mechanism of protection is provided by the cofactor H (FH), which belongs to a protein family of regulators of complement activity (RCA) [30]. FH recruits, in turn, the factor I (FI) that binds to C3b-FH complex and subsequently degrades C3b by cleaving it up to three times, and generating proteolytic fragments iC3b (after two cleavages) and C3c/C3dg (after three cleavages) [31]. These fragments cannot re-form C3 and C5 convertases and thereby stop further opsonisation and generation of membrane attack complexes [32].

This above-described mechanism is a protection that host cells tune through the regulators of complement activation (RCA) family. Their role is to either impair the generation of new C3b by accelerating the decay of the C3 convertases or act as a cofactor for factor I in degrading existing C3b [33]. In addition, unlike other activation regulators, the soluble and highly abundant regulator factor H provides an additional layer of protection by controlling the tickover activation state of the alternative pathway in circulation [31].

FH consists of 20 linearly distributed complement-control-protein (CCP) domains, even termed short consensus repeats (SCRP), that are composed by about 60 residues and that are common to all RCA proteins [34]. The first four domains (CCP1-4) of FH are crucial and sufficient to regulate the complement, while the other domains CCP5-20 of FH are useful to determine the specificity [34–36].

On the other hand, the structure of C3b shows an arrangement of 12 domains formed by the β chain (amino acid residues 1–645) and the α' chain (amino acid residues 727–1641) [37, 38]. The core of the structure consists of eight macroglobulin (MG) domains and a linker (LNK) domain. A complement C1r/C1s, UEGF, BMP1 (CUB) domain and a thioester-containing domain (TED) are visible between MG7 and MG8. TED domain is responsible for covalent pathogens surface attachment through the thioester moiety in Cys988-Gln991 [38].

C3b-FH complex structure retrieved from PDB database [39] reveals an extensive interacting interface, involving several domains of C3b and four contact regions of FH (CCP1–4). The bottom half of CCP1 and the CCP1-CCP2 linker bind C3b through hydrophobic interactions and salt bridges to the acidic α' N-terminal (α'NT) region (amino acid residues 727–746) and the MG7 domain [31, 40, 41]. The second binding region consists of a patch of conserved hydrophobic residues together with hydrophilic residues on CCP2 interacting with MG6 of C3b. In the third contact region, CCP3 establishes interactions with residues of MG2 and CUB domains in α' and β chains of C3b, respectively. Finally, at the fourth interacting site, CCP4 forms another bridge between MG1 and TED (Figure 5.2). The C3b domains α'NT, MG7, CUB and TED are known to arrange substantial changes during the conversion of C3 to C3b [37, 38, 42, 43], thus providing the specificity of FH for C3b rather than C3 [44].

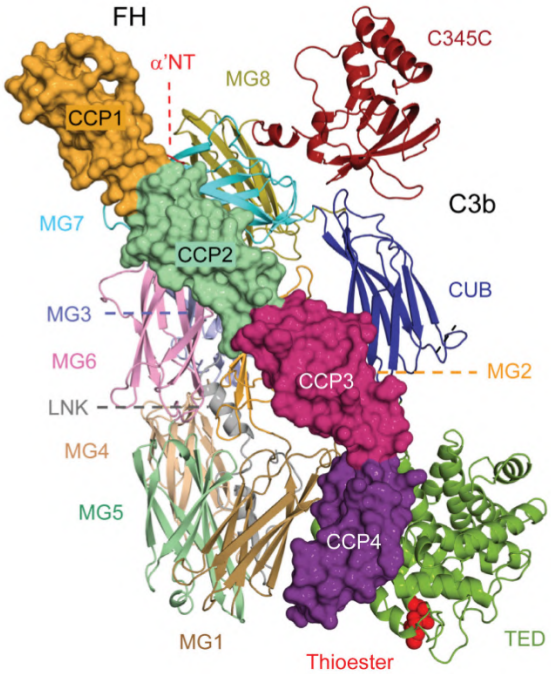


Figure 5.2. C3b-FH complex structure. FH is depicted including the surface composed by CCP1 to CCP4 (yellow, light green, pink and purple regions). C3b is depicted including the tertiary structure of each domain. In particular, MG1 and TED domains are the binding partners for FH CCP4 domain [44].

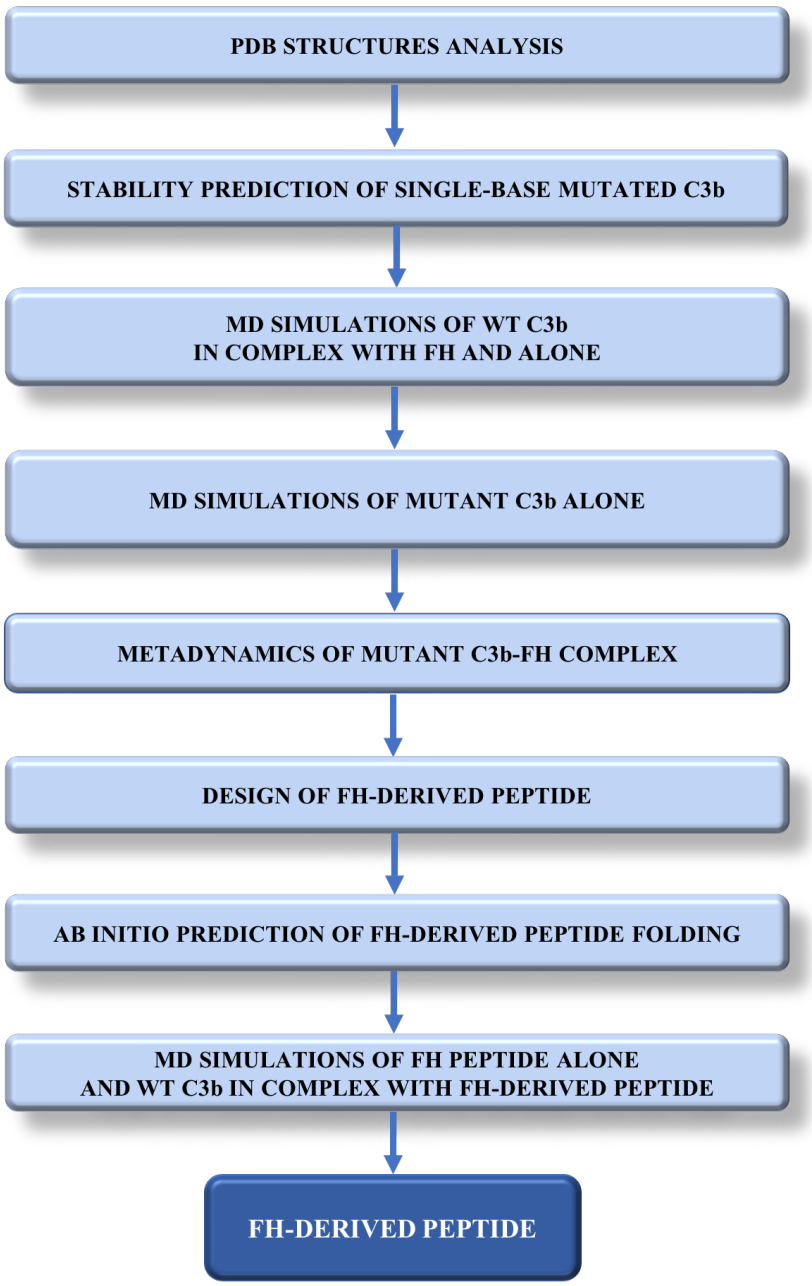
As above mentioned, C3b-FH complex formation is involved in the alternative pathway of complement activation, whereas this pathway has been characterised by C3 polymorphism recently associated with the AMD disease [45, 46]. Although this association has been identified, the molecular basis remained unknown for decades. In 2010, Heurich *et al*. shed light on $C3b_{R102G}$ mutation affecting AP activation by influencing the efficiency of regulation tuned by FH. Based on their experiments, authors found that the AMD risk variant ($C3b_{102G}$) weakly bound FH compared with wild-type $C3b_{102R}$, even causing decreased FI activity, extended convertase functions, and increased AP effects [44]. Furthermore, $C3b_{R102G}$ mutation was also associated with IgA nephropathy [47], systemic vasculitis [48], kidney allograft dysfunction [49], and dense deposit disease [50]. In detail, FH exhibited a weaker binding affinity for the mutated form ($C3b_{102G}$), with K_D $C3b_{102G}$: 1.4 μM, compared to the wild-type $C3b_{102R}$, with K_D $C3b_{102R}$: 1.0 μM [36, 51].

$C3b_{R102G}$ mutation is located in a positively charged area at the interface between the MG1 domain and TED, and CCP4 binds TED in close proximity to MG1 [44]. Therefore, it is likely that Arg substitution in position 102 with Gly can decrease overall positive charge in C3b and alter interdomain associations around the FH binding site, influencing the binding affinity and the cofactor activity.

In light of the above, unveiling mechanisms behind this mutation is essential to aid in understanding the disease aetiology. Therefore, the aim of this work was to design a peptide aimed at discriminating wild-type C3b from $C3b_{R102G}$ mutant. This peptide should be a potential instrument to be used for a diagnostic tool, whereas patients plasma should be employed to diagnose AMD associated with $C3b_{R102G}$ mutant form.

For this purpose, in this work a structural analysis of C3b-FH complex was performed first to identify the key interactions for the three protein chains of interest,

second molecular dynamics technique was exploited to investigate frequency and stability of those interactions and then they were observed and compared between wild-type and mutated system. This analysis allowed to develop a hypothesis to explain the binding affinity decrease associated with $C3b_{R102G}$ mutation. In this way, it was possible to design a peptide based on FH (CCP1-4) structure. The next sections provide a detailed description of this work leading to identify a putative FH-derived peptide that potentially should bind the wild-type C3b with higher binding affinity compared to $C3b_{R102G}$ variant overexpressed in patients affected by AMD. All the steps described in detail in the next sections are listed in Scheme 5.1.



Scheme 5.1. Overview of the computational workflow performed to identify the FH minimum active sequence composing the FH-derived peptide aimed at discriminating WT C3b from mutant $C3b_{R102G}$ form

5.2 Results and discussion

5.2.1 Analysis of PDB structures of the trimeric complex C3b-FH

The first step of this work was the analysis of the available PDB structures of C3b-FH complex from the Protein Data Bank [52], *i.e.* the PDBs 2WII (resolution 2.70 Å), 5O32 (resolution 4.21 Å) and 5O35 (resolution 4.20 Å), where C3b and FH take part in the interaction with a ratio 2:1, respectively. Due to the low-quality resolutions of the last two PDBs, they were only used to explore the key interactions observable between C3b protein and factor H, while the PDB 2WII with the best resolution was chosen to deeply investigate the protein-protein interactions in the next steps. The protein structures were firstly prepared by adding missing atoms and optimising other parameters such as chiralities and protonated or deprotonated group states at pH 7.4 ± 0.2. In particular, for PDB 2WII a crosslinking process was necessary by integrating the structure with two missing amino acids in position 98 and 99 on chain β of C3b protein (Ser98 and Glu99). The analysis of the interactions highlighted the following key residues reported in table 5.1 between C3b MG1 (chain β) and TED (chain α') domains and FH CCP4 domain.

Table 5.1. Observed interactions from PDBs 2WII, 5O32 and 5O35 between C3b domains (MG1 and TED) and FH CCP4.

PDB ID	C3b MG1	FH CCP4	Interaction type	C3b TED	FH CCP4	Interaction type	C3b MG1	C3b TED	Interaction type
2WII	Gly64	Arg246	1 H-bond	Val1068	Arg232	VdW contacts	Phe62	Arg1020	VdW contacts
	Lys65	Glu245	1 H-bond + 1 salt bridge	Asn1069	Arg246	1 H-bond	Pro63	Trp1012	VdW contacts
	Lys66	Arg246 Arg257 Pro258	VdW contacts	Ala1072	Glu245	VdW contacts	Ser98	Glu1010	1 H-bond
	Leu67	Arg257	VdW contacts	Ser1075	Gln234	1 H-bond	Lys100	Asp1266	1 H-bond + 1 salt bridge
	Arg94	Glu253	1 H-bond + 1 salt bridge	Gln1076	Ile221	VdW contacts	Arg102	Glu1010	VdW contacts
	Arg94	Glu188	VdW	Asp1134	Arg232	VdW contacts	Lys104	Glu1010	1 salt bridge
	Thr162	Glu189	1 H-bond	Ile1135	Arg232	VdW contacts	Phe105	Leu1017	VdW contacts
	Asp178	Gln172	VdW contacts	Glu1138	Gln223	1 H-bond	Lys119	Glu1018	1 salt bridge
	Ser179	Gln172	1 H-bond				Val120	Leu1017	VdW contacts
	Leu180	Gly171	VdW contacts						
	Leu180	Ala173	VdW contacts						
	Ser181	Gly171	1 H-bond						
	Gln185	Phe170	VdW contacts						
	Glu211	Arg175	2 H-bonds + 1 salt bridge						
5O32	Lys65	Glu245	1 Salt bridge	Asn1091	Arg232	1 H-bond	Phe62	Trp1034	VdW contacts
	Lys66	Arg246	VdW contacts	Ile1093	Glu245	VdW contacts	Pro63	Trp1034	VdW contacts
	Lys66	Arg257	VdW contacts	Ala1094	Tyr243	VdW contacts	Arg102	Thr1031	VdW contacts
	Lu67	Arg257	VdW contacts	Asp1096	Tyr243	VdW contacts	Lys104	Glu1032	1 H-bond + 1 salt bridge
	Arg94	Glu188	1 H-bond	Ser1097	Gln234	VdW contacts	Phe105	Glu1035	VdW contacts
	Glu95	Arg257	1 Salt bridge	Ile1157	Arg232	VdW contacts	Glu118	Gln1043	VdW contacts
	Glu95	Ser254	VdW contacts	Glu1160	Ser222	1 H-bond	Val120	Leu1039	VdW contacts
	Thr162	Glu189	1 H-bond				Leu198	Arg979	1 H-bond
	Asp178	Gln172	1 H-bond						
	Ser179	Ala173	1 H-bond						
	Ser179	Gln172	VdW contacts						
	Leu180	Gly171	VdW contacts						
	Leu180	Gln172	VdW contacts						
	Leu180	Ala173	VdW contacts						
	Ser181	Gly171	1 H-bond						
	Ser181	His191	VdW contacts						

	Gln183	His191	VdW contacts						
	Gln185	Glu116	1 H-bond						
	Gln185	Phe170	VdW contacts						
	Glu211	Arg175	2 H-bonds + 1 salt bridge						
	Glu211	Glu189	VdW contacts						
5O35	Lys65	Glu245	1 Salt bridge	Val1090	Arg232	VdW contacts	Phe62	Leu1039	VdW contacts
	Lys66	Arg246	VdW contacts	Asn1091	Arg246	1 H-bond	Pro63	Arg1042	VdW contacts
	Lys66	Arg257	VdW contacts	Ile1093	Glu245	VdW contacts	Arg102	Glu1032	VdW contacts
	Lys66	Pro258	VdW contacts	Ala1094	Tyr243	VdW contacts	Asn103	Glu1035	VdW contacts
	Leu67	Arg257	VdW contacts	Ser1097	Gln234	VdW contacts	Lys104	Glu1032	1 salt bridge
	Thr162	Glu189	VdW contacts	Ile1157	Arg232	VdW contacts	Phe105	Leu1039	VdW contacts
	Asp178	Gln172	VdW contacts	Glu1160	Ser222	1 H-bond	Lys119	Glu1040	1 H-bond + 1 salt bridge
	Ser179	Ala173	1 H-bond	Leu1109	Cys1138	VdW contacts	Val120	Leu1039	VdW contacts
	Ser179	Gln172	VdW contacts				Leu198	Arg979	VdW contacts
	Leu180	Gly171	VdW contacts						
	Gln183	His191	1H-bond						
	Gln185	Phe170	VdW contacts						
	Gln185	Glu116	VdW contacts						
	Glu211	Arg175	1 H-bond +1 salt bridge						
	Glu211	Glu189	VdW contacts						

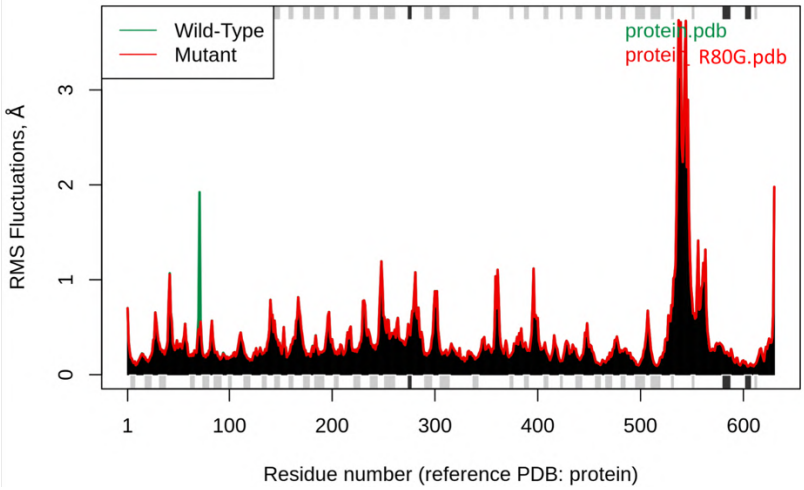
As it can be observed, Arg102 in MG1 domain does not appear interacting with factor H, on the contrary, it is involved in interactions with the other C3b chain on TED domain. Thus, these considerations suggest that Arg102 should be crucial to stabilise the overall C3b quaternary structure including chain β and chain α'. Indeed, $C3b_{R102G}$ mutation could disrupt the linking between TED and MG1. Therefore, the work herein described was based on this hypothesis, that was further investigated in the next steps below detailed.

5.2.2 Stability prediction of single-base mutated C3b

As previously mentioned, in AMD disease C3b was found mutated in position 102 of MG1 domain, where arginine residue is substituted with glycine. According to PDBs analysis, Arg102 is essentially involved in interactions with C3b TED domain without binding FH. Therefore, this fact could suggest that glycine in position 102 destabilises the binding of C3b chain β with C3b chain α'. Indeed, glycine has been extensively demonstrated to produce a destabilising effect especially for well-defined protein secondary structures (*e.g.* α-helix [53] and β-sheet [54]). In fact, glycine lacks β-carbon resulting in an unusual flexibility, that can take on polypeptide backbone conformations not allowed by other amino acids. Therefore, mutation to glycine should cause flexibility and possible conformational changes convoluted with the effects of removing the side chain atoms. Furthermore, specifically, for C3b single-base mutation, glycine is a neutral amino acid that in fact lacks the positive charge usually present in arginine side chain in physiological conditions. Therefore, these considerations highlight that the presence of glycine in position 102 could disrupt the interactions established by Arg102 of C3b MG1 domain with its protein partner. In order to deeply explore and quantitatively analyse this hypothesised destabilisation, the computational residue scanning was performed by using Schrödinger suite and DynaMut [55], two prediction tools.

For the "Residue Scanning" tool of Schrödinger, the PDB 2WII was used including the trimeric complex C3b chain β and chain α' and FH CCP1-4. The

calculations provided $\Delta\Delta G_{stability}$ = 20.47 kcal/mol and $\Delta\Delta G_{affinity}$ = 0.59 kcal/mol. These values confirmed the destabilising effect of $C3b_{R102G}$ mutation. These results were also consistent with the DynaMut outcomes. For this tool, a different PDB structure was processed, that is the PDB 2I07 (where R80 corresponds to R102) including only chain β and chain α'. Even in this case, the prediction outcome suggested that $C3b_{R102G}$ mutation produces a destabilising effect on C3b. This information was retrieved from the Δ vibrational entropy energy values ($\Delta\Delta S_{Vib}$) between wild-type and mutant systems with $\Delta\Delta S_{Vib}$ = 4.117 kcal/mol·K, confirming an increase of molecule flexibility. Finally, wild-type and mutant sequence of C3b protein were extracted from their respective 3D structures and then aligned. The results of normal mode analysis data for each sequence are displayed below in plot 5.1, that reports differences in atom fluctuations comparing wild-type and mutant forms.



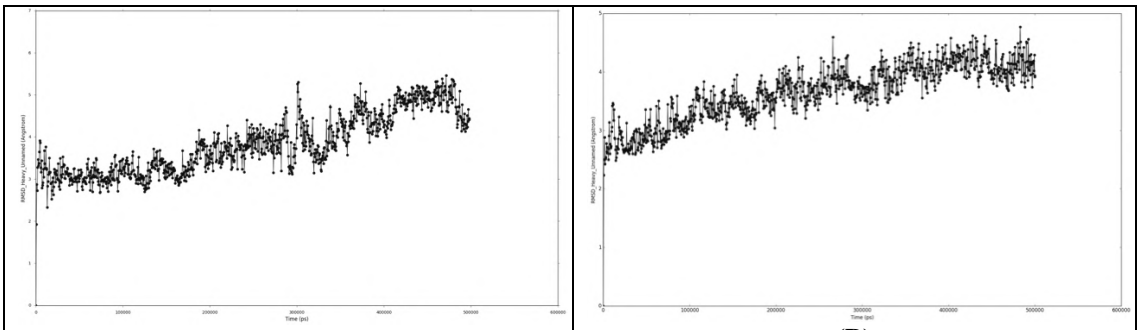
Plot 5.1. RMS fluctuations of wild-type and R80G mutant C3b (R80 of the PDB used corresponds to R102). The secondary structure types on each region of the sequence are added to the top and bottom margins of the plot (black stands for helices, while grey for strands).

5.2.3 Molecular Dynamics simulation of wild-type C3b-FH complex

In order to deeply explore the most stable and frequent interactions of C3b residues towards factor H, two MD simulations of 500 ns were performed on wild-type C3b-FH complex by using PDB 2WII. The total energy of the systems and the RMSD plots were checked during the entire trajectories to ensure the outcomes reliability. The related data are shown in table 5.2 (average energy values) and Plot 5.2A-B (RMSD plots).

Table 5.2. Energy values monitored during the two MD trajectories on C3b-FH complex

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-599319.5	317.462	-0.001	-609656.7	315.624	0.000
Potential energy (kcal/mol)	-742199.2	248.268	-0.001	-754902.7	244.986	0.000



(A) **(B)**

Plot 5.2. RMSD plots of **(A)** first and **(B)** second MD simulation performed on C3b-FH complex.

The RMSD plots highlighted the overall acceptable reliability and stability of the systems. Thus, all the frames of these simulations were clustered generating 10 groups of frames per each and the MD frames representative of the most abundant clusters were further analysed:

- First MD → Frame 730 (32 frames), frame 260 (20 frames), frame 90 (12 frames), frame 490 (24 frames) and frame 600 (6 frames);
- Second MD → Frame 710 (46 frames), frame 620 (6 frames), frame 250 (25 frames) and frame 150 (9 frames).

The analysis of these frames shed light on the most frequent and stable interactions extracted from the two MD trajectories, as shown in table 6.3.

Table 5.3. Interactions between MG1, TED and CCP4 extracted from MD simulations performed on C3b-FH complex (PDB 2WII).

First MD			Second MD		
C3b MG1	**FH CCP4**	**Interaction type**	**C3b MG1**	**FH CCP4**	**Interaction type**
Lys65	Glu245	1 H-bond + 1 salt bridge	Lys65	Glu245	1 H-bond + 1 salt bridge
Lys66	Pro258	VdW	Lys66	Pro258	VdW
Leu67	Arg257	VdW	Arg94	Glu188	2 H-bonds +1 salt bridge
Lys73	Glu264	1 H-bond	Glu95	Arg257	2 H-bonds + 1 salt bridge
Arg94	Glu188	1 H-bond	Asp178	Arg166	2 H-bonds + 1 salt bridge
Glu95	Arg257	2 H-bonds + 1 salt bridge	Ser179	Ser160	1 H-bond
Thr162	Glu189	1 H-bond	Ser179	Ala173	1 H-bond
Ser179	Ala173	1 H-bond	Ser181	Gly171	1 H-bond
Leu180	Gly171	1 H-bond	Glu211	Arg175	2 H-bonds + 1 salt bridge
Ser181	Gly171	1 H-bond	Glu211	Ser159	1 H-bond
Gln185	Phe170	VdW			
Glu211	Arg175	2 H-bonds + 1 salt bridge			
C3b TED	**FH CCP4**	**Interaction type**	**C3b TED**	**FH CCP4**	**Interaction type**
Lys1028	Glu245	1 H-bond + 1 salt bridge	Lys1028	Glu245	1 H-bond + 1 salt bridge
Asp1134	Asn230	1 H-bond	Asn1069	Arg232	1 H-bond
Glu1137	Lys228	1 H-bond + 1 salt bridge	Glu1138	Lys224	1 H-bond
Glu1138	Lys224	1 H-bond + 1 salt bridge	Arg1281	Asp165	1 H-bond
Arg1281	Asp165	1 H-bond	Ile1135	Gln234	1 H-bond
			Gln1139	Gln234	1 H-bond
			Asp1134	Arg232	2 H-bonds
			Glu1292	Lys156	1 H-bond + 1 salt bridge
			Glu1138	Lys224	1 H-bond + 1 salt bridge
C3b MG1	**C3b TED**	**Interaction type**	**C3b MG1**	**C3b TED**	**Interaction type**
Lys119	Glu1018	1 H-bond	Lys100	Asp1266	1 H-bond + 1 salt bridge
Arg102	Glu1010	2 H-bonds + 1 salt bridge	Arg102	Glu1010	2 H-bonds + 1 salt bridge
Arg102	Glu1013	1 H-bond	Arg102	Glu1013	2 H-bonds + 1 salt bridge
Glu118	Gln1021	1 H-bond	Glu118	Gln1021	1 H-bond
Lys100	Glu1292	1 H-bond + 1 salt bridge	Lys119	Glu1018	1 H-bond + 1 salt bridge
Lys104	Glu1010	1 Salt bridge	Val120	Glu1018	1 H-bond
Phe62	Tro1012	1 Pi-Pi			
Val120	Glu1018	1 H-bond			

By analysing these interactions, Factor H appears to establish a higher number of interactions with C3b MG1 domain compared to TED domain. It seems to suggest a slight preference for MG1 domain.

In detail, herein the most stable interactions shared by the two MD simulations were the following ones:

- For MG1-CCP4 interactions → Lys65-Glu245, Lys66-Pro258, Leu67-Arg257, Arg94-Glu188, Glu95-Arg257, Ser179-Ala173, Ser181-Gly171 and Glu211-Arg175;
- For TED-CCP4 interactions → Lys1028-Glu245, Glu1138-Lys224 and Arg1281-Asp165;
- For MG1-TED interactions → Arg102-Glu1010, Arg102-Glu1013, Glu118-Gln1021, Lys119-Glu1018 and Val120-Glu1018.

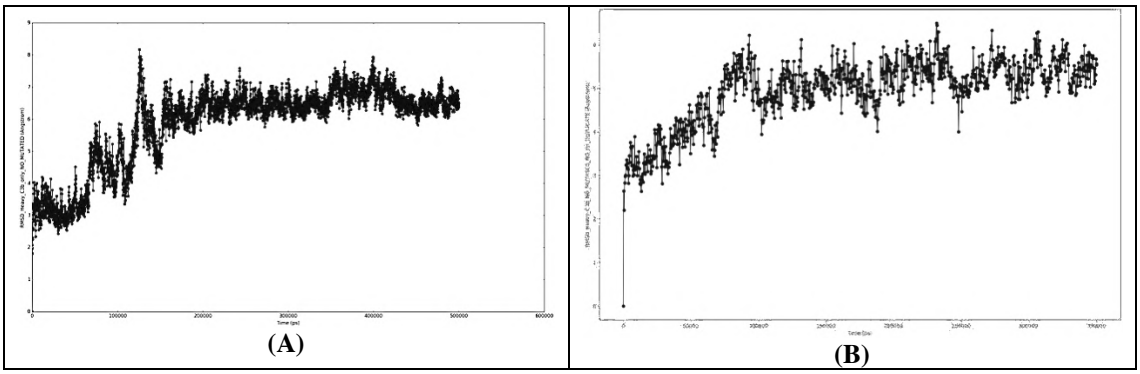
The results of these MD were in accordance with the preliminary PDBs analysis, where Arg102 was not involved in binding with FH CCP4, but it was strongly observed interacting with the other C3b chain involving TED domain residues, Glu1010 and Glu1013, by establishing mainly two H-bonds plus a salt bridge and an H-bond and a salt bridge, respectively. Furthermore, among the two simulations, the shared stable interactions between MG1 and TED domain were Arg102-Glu1010, Arg102-Glu1013, Glu118-Gln1021, Lys119-Glu1018 and Val120-Glu1018, respectively.

5.2.4 Molecular Dynamics simulation of wild-type C3b protein

In order to deeply investigate whether the interactions between MG1 and TED domains should expect to be stable even in absence of a third protein (FH), other two MD simulations were performed by including only C3b protein deleting FH chain from PDB 2WII. The simulations were run setting 500 ns of time per each and the related energy values are reported in table 5.4. Upon the deletion of FH, C3b chains exhibited a first stage of molecular arrangement involving the amino acids at the protein-protein interface that should move on from bound to unbound state. This fact was also observable from the RMSD plots that presented a first unstable portion during the trajectories until about 150 ns for the first MD and 180 ns for the second MD (plot 5.3A-B). Therefore, the frames falling into these unstable regions of the MD simulations were not included in the following analysis.

Table 5.4. Energy values monitored during the two MD trajectories on the wild-type C3b

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-5342255.0	295.364	0.000	-533885.3	297.63	-0.001
Potential energy (kcal/mol)	-661645.44	228.510	0.000	-661257.0	232.350	-0.001



Plot 5.3. RMSD plots of **(A)** first, **(B)** second MD simulation performed on wild-type C3b

On this regard, the trajectory frames were clustered by generating 10 frame groups per MD, whereas the most abundant clusters were represented by the following frames:

- First MD → Frame 350 (28 frames), frame 280 (37 frames), frame 410 (22 frames);

Second MD → Frame 270 (45 frames), frame 170 (25 frames), frame 380 (17 frames) and frame 320 (12 frames).

The analysis of the above-listed frames allowed to identify the key residues for MG1 and TED domains, as illustrated in table 5.5. The analysis of the most stable interactions of these MD was consistent with the previously shown results from the two MD simulations on C3b-FH complex.

Table 5.5. Intramolecular interactions between MG1 and TED domains extracted from MD simulations performed on WT C3b protein without protein partner.

First MD			Second MD		
C3b MG1	**C3b TED**	**Interaction type**	**C3b MG1**	**C3b TED**	**Interaction type**
Lys119	Glu1018	1 H-bond	Lys119	Glu1018	1 H-bond + 1 salt bridge
Arg102	Glu1010	2 H-bonds + 1 salt bridge	Arg102	Glu1010	2 H-bonds + 1 salt bridge
Glu118	Gln1021	1 H-bond	Glu118	Gln1021	1 H-bond
Val120	Glu1018	1 H-bond	Val120	Glu1018	1 H-bond

As it can be observed, even in these simulations the two hydrogen bonds and the salt bridge established between Arg102 and Glu1010 were retrieved among the most stable and frequent interactions between MG1 and TED. It demonstrates that these interactions should be crucial to stabilise the complex and their stability could be considered independent from FH binding.

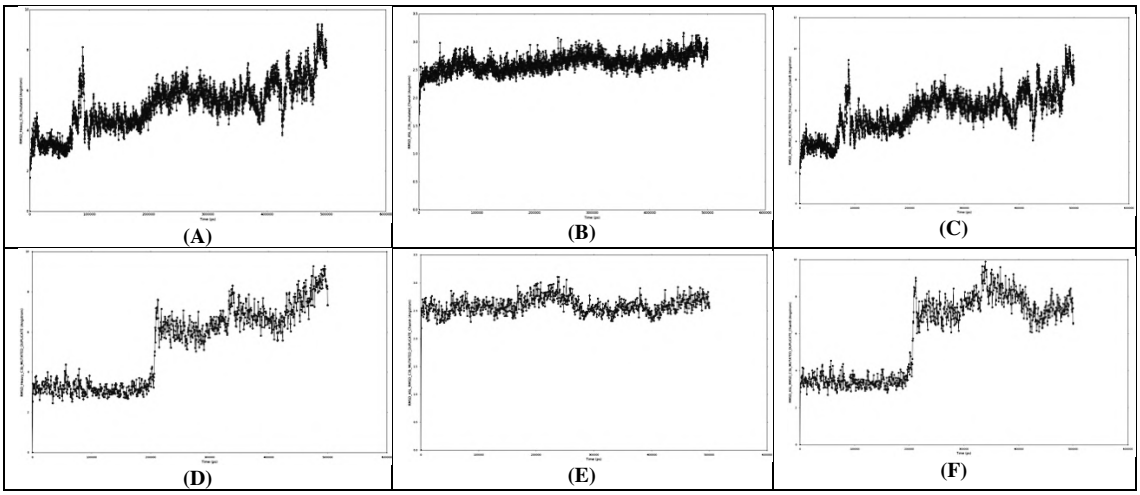
5.2.5 Molecular Dynamics simulation of mutant C3b protein

After the above described MD simulations, it was interesting to further investigate the previously mentioned hypothesis, *i.e.* $C3b_{R102G}$ mutation should provoke a destabilisation of the interaction between the two domains of C3b and consequently disrupt the contacts with FH CCP1-4 domains. Therefore, in order to deeply investigate the $C3b_{R102G}$ mutation-triggered destabilising effect only focusing the attention on MG1 and TED domains binding, C3b protein was mutated in position 102 to glycine and two MD simulations were performed not including FH. For this purpose, PDB 2WII was used deleting FH. The simulation of both trajectories was 500 ns. Herein, the average energy

values are reported in table 5.6 and the RMSD values for each chain are plotted in Plot 5.4A-F.

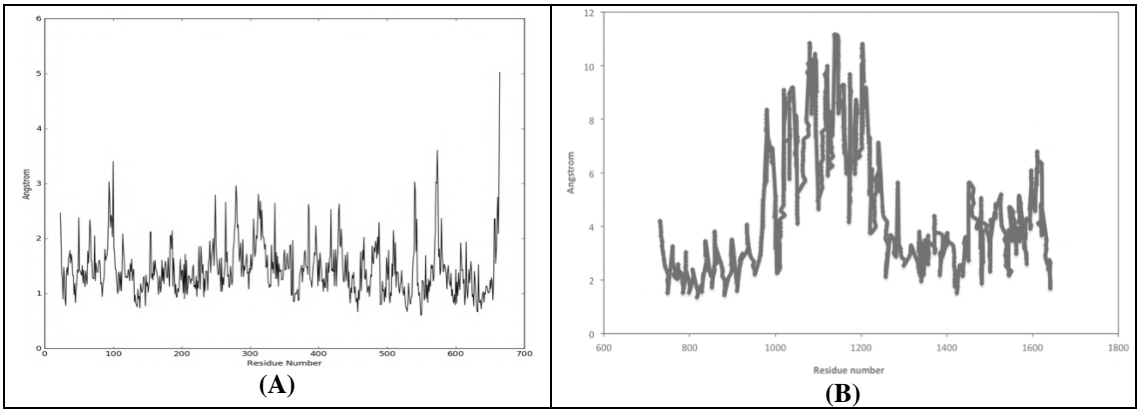
Table 5.6. Energy values monitored during the two MD trajectories on C3b mutant

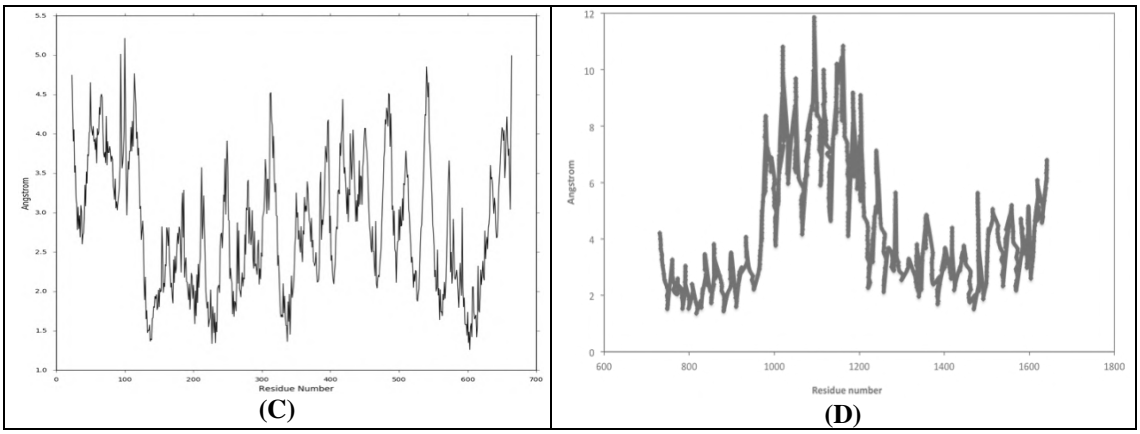
	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-584502.1	308.721	0.000	-534307.0	291.141	0.000
Potential energy (kcal/mol)	-723394.3	239.011	0.000	-661655.0	224.101	0.000



Plot 5.4. RMSD plots of MD on C3b mutant. RMSD plot of $C3b_{R102G}$ chain α'-chain β complex during the first **(A)** and the second **(D)** MD; RMSD plot of $C3b_{R102G}$ chain β during the first **(B)** and the second **(E)** MD; RMSD plot of $C3b_{R102G}$ chain α' during the first **(C)** and the second **(F)** MD

The RMSD plots of the protein-protein complex are not very stable and looking at the RMSD plot of chain β and chain α' is deducible that the unreached RMSD stationary shape for the protein-protein complex is due to chain α' (involving TED domain), whom RMSD plot highlights a certain instability. In order to deeply explore which residues mainly contribute to the molecular motions highlighted from the RMSD plot, the RMSF plot was computed [56] and analysed for both chain β and chain α', and they are reported below (plot 5.5A-B). As expected, the highest fluctuations ranging from 3 to 12 Å can be identified involving residues in position about 960 to 1280. This residue range, in fact, corresponds to the TED domain, that is the region of chain α' that normally interacts with MG1 domain in chain β.





Plot 5.5. RMS Fluctuations plots on C3b mutant. RMS fluctuation plot of chain β for the first **(A)** and the second **(C)** MD; and RMS fluctuation plot of chain α' for the first **(B)** and the second **(D)** MD

Indeed, during the trajectories, the TED domain slightly moved away from MG1 domain probably due to the loss of stabilising interactions such as those formed between Arg102 and Glu1010.

In the light of the above analysis, the interactions between the two chains of C3b mutant were also explored in order to investigate whether the key interactions identified from previous computational studies were retrieved even for these two MD. For this purpose, the MD frames were clustered by getting 10 representative frames per each MD, where the most abundant ones were considered for the interaction analysis, thus selecting the following frames:

- First MD → Frame 4430 (11 frames), frame 2430 (38 frames), frame 1720 (24 frames), frame 520 (8 frames) and frame 4940 (10 frames);
- Second MD → Frame 820 (38 frames), frame 370 (6 frames), frame 260 (35 frames) and frame 910 (10 frames).

As expected by exploring the interaction interface of chain β and chain α', most of the crucial contacts extracted from the previous analyses were not found in both MD, confirming that single mutation of C3b could destabilise the quaternary structure of this protein, thus consequently disrupting the interaction with FH (table 5.7).

Table 5.7. Analysis of stability of the crucial interactions for C3b MG1-TED extracted from MD simulations on C3b mutant

MUTATED C3b		
Crucial Interaction MG1 - TED	**First MD**	**Second MD**
Gly102 – Glu1010	✗	✗
Lys119 – Glu1018	✓	✗
Val120 – Glu1018	✗	✗
Glu118 – Gln1021	✗	✗

As expected, glycine in position 102 is not able to reproduce H-bond contacts established by Arg102 with Glu1010. This fact paves the way to a higher flexibility of the loop involving residue 102 resulting in new interactions not occurring in normal conditions, that could disrupt protein-protein binding. In fact, these considerations were

demonstrated for example by analysing the first MD simulation of mutated C3b protein, where Arg94 established intramolecular contacts with Glu211 that has been extensively identified as a key residue by tuning two H-bonds and one salt bridge with FH Arg175. This event is depicted in figure 5.3A-B.

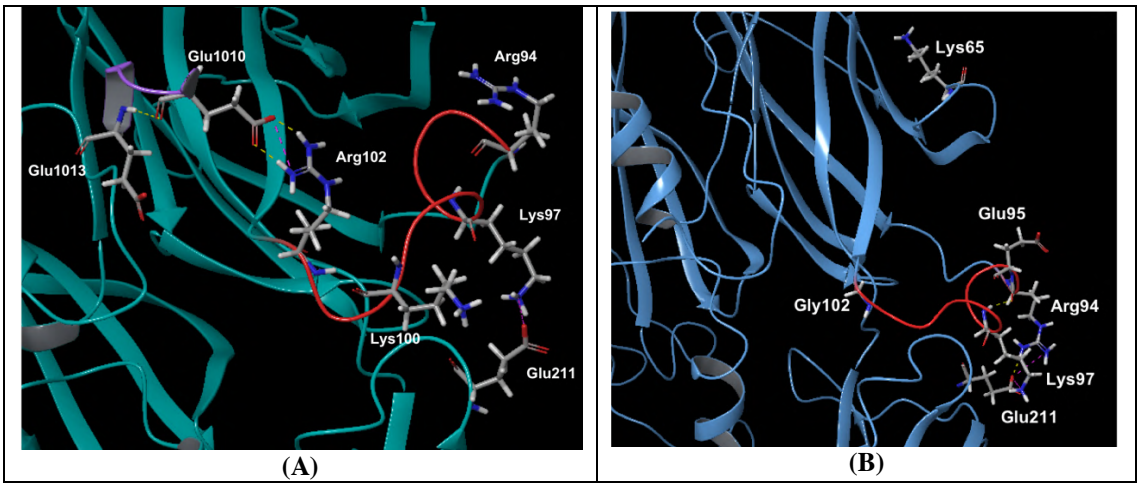
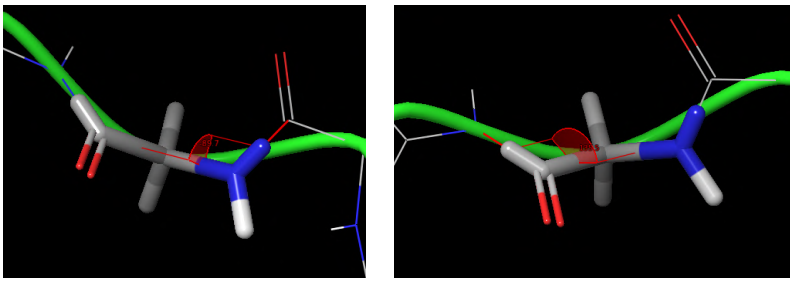


Figure 5.3. Comparison between MD frame of WT C3b **(A),** where Arg102 interacts with Glu1010 and Arg94 does not establish contacts with Glu211, and mutated C3b **(B),** where Gly102 is not able to bind Glu1010 and probably confers more flexibility to the red loop that can move towards Glu211 and allow it to interact with Arg94.

Therefore, these findings suggested that the punctual mutation in position 102 does not seem directly responsible for the decrease in binding affinity between C3b and its cofactor FH; on the contrary, it could be the result of the low stability at the expense of C3b quaternary assembly by involving MG1 and TED interactions. Hence, this structural destabilisation could, in turn, generate a lower affinity of C3b for Factor H. Thus, this insight was further explored by performing a Metadynamics simulation described in the next section including the mutated C3b in complex with FH.

5.2.6 Metadynamics simulation on mutant $C3b_{R102G}$ in complex with FH CCP1-4 domains

In order to further explore the $C3b_{R102G}$ mutation effect in presence of FH protein, an enhanced sampling technique was employed, that is the metadynamics simulation [57]. For this purpose, Arg102 was substituted to glycine to reproduce the AMD-associated mutation and two collective variables (CVs) were chosen to stimulate the system in potentially simulating the disease-associated behaviour of C3b and FH proteins. The selected CVs were the dihedral angles φ and ψ (figure 5.4A-B) of Gly102 that were exploited to investigate potential destabilising effects borne to the contacts between C3b MG1 domain, C3b TED domain and FH CCP4 domain in presence of the mutation R102G.



(A) **(B)**

Figure 5.4. Dihedral angles **(A)** φ and **(B)** ψ of Gly102 were selected as collective variables of the metadynamics simulation

The simulation time was set 100 ns and the energy, temperature, pressure and volume were monitored during the entire simulations, whereas the average energy values are reported in table 5.8. Figure 5.5 depicts the free-energy minima landscape of the two selected collective variables.

Table 5.8. Energy values monitored during the metadynamics trajectory on C3b mutant

	Average	Std Dev	Slope (ps^{-1})
Total energy (kcal/mol)	-608192.697	327.450	-0.003
Potential energy (kcal/mol)	-753399.632	259.432	-0.003

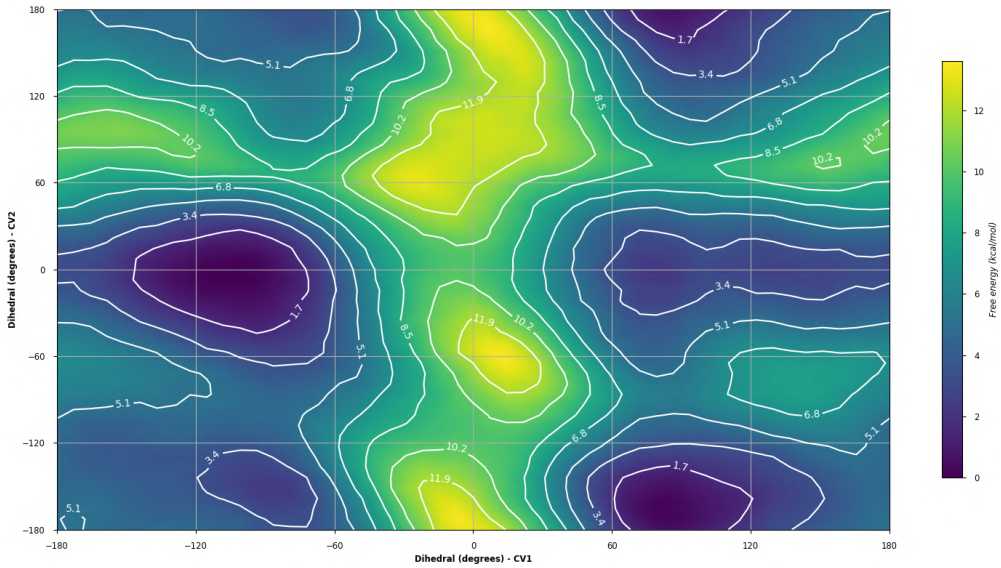
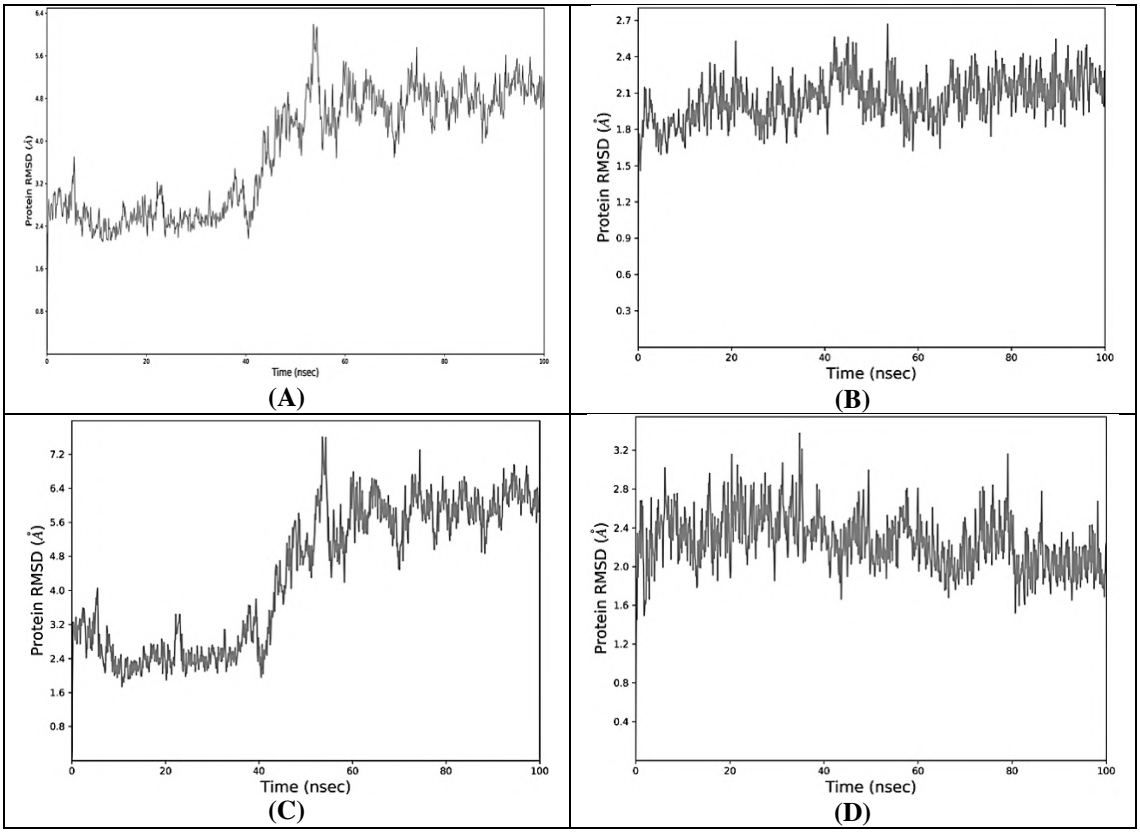


Figure 5.5. Free-energy minimum landscape of the two selected collective variables, the dihedral angles φ and ψ of Gly102; from blue shades that are the lowest free energy values to yellow shades that are the highest free energy values

The RMSD plots of the complex and the individual chains were generated and they are reported in plots 5.6A-D. The first plot depicted (plot 5.6A) shows the RMSD values representation of mutant C3b in complex with FH, that reports a certain instability after about 40 ns of simulation. Analysing the other three plots, that illustrated respectively the RMSD values of chain β including $C3b_{R102G}$ mutation (plot 5.6 B), chain α' (plot 5.6C) and FH CCP1-4 domains (plot 5.6D) during the metadynamics trajectory, it is immediately visible that the instability of the RMSD plot of the complex has to be ascribed to chain α' (including TED domain). Therefore, it seems that $C3b_{R102G}$ on MG1 indirectly affects TED domain stability. This data was further explored by investigating

the most frequent interactions visible during the trajectory in presence of C3b$_{R102G}$ mutation.



Plot 5.6. RMSD plots of Cα of **(A)** mutated C3b-FH complex, **(B)** chain β including C3b$_{R102G}$ mutation, **(C)** chain α' and **(D)** FH CCP1-4 domains during the metadynamics trajectory

Thus, the frames obtained from the metadynamics simulation were clustered to select 10 groups and the frames representative for the most frequent and abundant clusters were the following ones: frame 870 (19 frames), frame 310 (18 frames), frame 720 (12 frames) and frame 100 (11 frames). These latter were analysed and the shared visible interactions were collected and registered in table 5.9.

Table 5.9. Comparison between the stable interactions extracted from the previous MD simulations and the stable interactions extracted from the metadynamics simulation

Stable interactions extracted from previous MD simulations			**Stable interactions extracted from Metadynamics simulation**		
C3b MG1	**FH CCP4**	**Interaction type**	**C3b MG1**	**FH CCP4**	**Interaction type**
Lys65	Glu245	1 H-bond + 1 salt bridge	Lys65	Glu245	1 H-bond + 1 salt bridge
Lys66	Pro258	VdW	Glu95	Arg257	2 H-bonds + 1 salt bridge
Leu67	Arg257	VdW	Glu95	Ser254	1 H-bond
Glu95	Arg257	2 H-bonds + 1 salt bridge			
C3b TED	**FH CCP4**	**Interaction type**	**C3b TED**	**FH CCP4**	**Interaction type**
Lys1028	Glu245	1 H-bond + 1 salt bridge	Asp1074	Lys236	1 salt bridge
Glu1138	Lys224	1 H-bond + 1 salt bridge	Gln1139	Gln234	1 H-bond
			Ile1135	Gln234	1 H-bond
C3b MG1	**C3b TED**	**Interaction type**	**C3b MG1**	**C3b TED**	**Interaction type**
Arg102	Glu1010	2 H-bonds + 1 salt bridge	Lys119	Glu1018	1 H-bond + 1 salt bridge
Arg102	Glu1013	1 H-bond			
Glu118	Gln1021	1 H-bond			
Lys119	Glu1018	1 H-bond			
Val120	Glu1018	1 H-bond			

As it can be noticed, some key interactions between MG1 domain and FH CCP4 domain identified from the previous MD simulations were not met during the analysis of the metadynamics outcome. Furthermore, TED domain and FH CCP4 domain did not completely establish the crucial interactions herein previously highlighted. Finally, only one interaction between MG1 and TED domain was stable during the entire trajectory and it was also observed from the previously analysed MD simulations. These results are in accordance with the instability registered from the RMSD plot of chain α' (in plot 5.6C) and strengthen the above-formulated hypothesis, that mutation in position 102 can potentially disrupt the interactions between the two domains of C3b (MG1 and TED), through a destabilisation of C3b quaternary structure. This destabilisation could eventually affect also the contacts with FH protein, in particular considering the CCP4 domain. All these considerations were collected and led the design of a peptide based on FH protein considering the region interacting both contemporarily with MG1 and TED domains.

5.2.7 Identification of FH-derived peptide and structure folding prediction

Computational studies shed light on the structural insights about the affinity decrease of single-mutated C3b for FH protein. Therefore, in order to design a peptide able to discriminate the WT C3b protein from the mutated one, a promising strategy could be the selection of a portion of FH corresponding to the CCP4 domain, that can bind not only MG1 but also TED domain in normal conditions. This strategy could provide a peptide able to recognise and bind WT C3b and probably unbind or scarcely bind the mutated $C3b_{R102G}$ detected in patients affected by AMD. For this purpose, based on the previous studies, the key interactions formed by FH CCP4 residues with WT C3b MG1 and TED domains were considered and they are reported below in table 5.10.

Table 5.10. Key residues contacting each other referring to FH CCP4 and C3b domains (MG1 and TED) as retrieved from MD simulations analyses

FH CCP4	C3b	
Key residues	**MG1 Domain**	**TED Domain**
Glu245	Lys65	Lys1028
Arg257	Glu95	-
Glu253	Arg94	-
Arg232	-	Asp1134
Gln234	-	Gln1139

Therefore, the designed peptide should contain Arg232, Gln234, Glu245, Glu253 and Arg257 of FH. The resulting peptide, hereinafter FH peptide, was composed of 34 amino acids based on FH sequence including the following residues.

231-ERFQYKCNMGYEYSERGDAVCTESGWRPLPSCEE-264

The corresponding portion of FH is characterized by β-strands and loops as illustrated in figure 5.7, and it presents a disulfide bridge between Cys237 and Cys262.

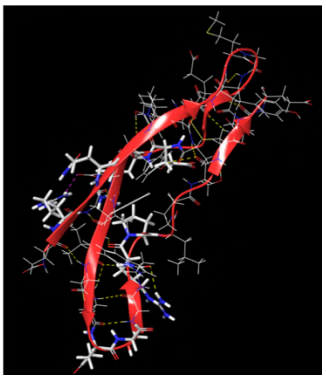


Figure 5.7. FH peptide structure

The next step was the study aimed at predicting the putative peptide folding to reproduce its native conformation. For this purpose, two software based on *ab initio* calculations were used, QUARK and PEP-FOLD3, that are usually performed when no global template information is available. Indeed, in this case, *ab initio* calculations were useful to avoid influences from available homologous templates in reproducing peptide folding.

The first software, QUARK, exploits a computer algorithm for *ab initio* protein structure prediction and peptide folding, which aims to construct the correct protein 3D model from amino acid sequence only. QUARK models are built from small fragments by replica-exchange Monte Carlo simulation under the guide of an atomic-level knowledge-based force field [58, 59].

The other software, PEP-FOLD3, applies a *de novo* approach aimed at predicting peptide structures from amino acid sequences. This method is based on a Hidden Markov Model sub-optimal conformation sampling approach, that allows generating models for peptides from 5 to 50 amino acids [60–66].

Both software generated five models that showed a succession of β-strands and loops very similar to the original structure. The models are illustrated in figure 5.8A-B.

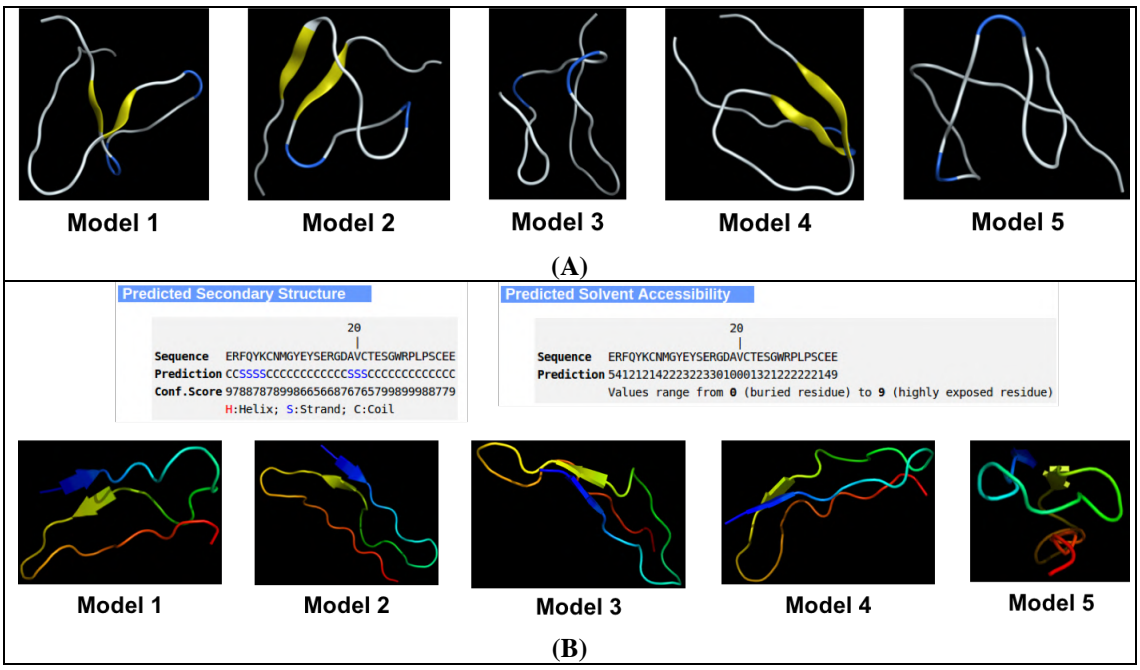


Figure 5.8. QUARK models **(A)** and PEP-FOLD3 models **(B)**.

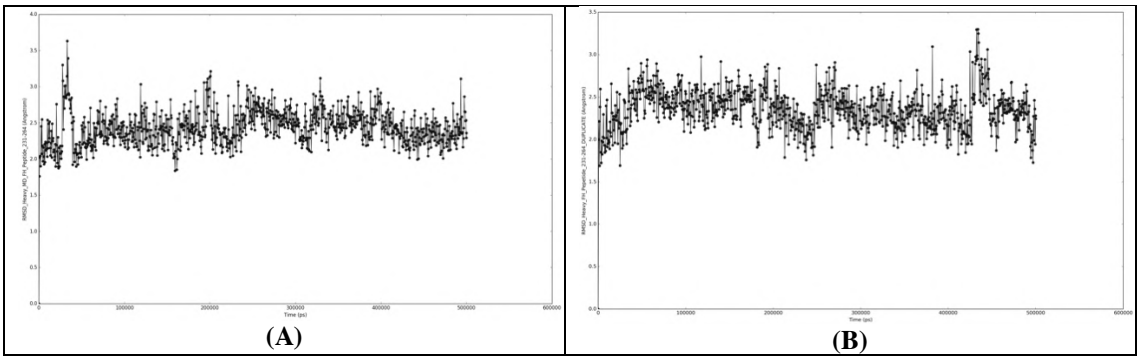
All these models are not perfectly able to match with native FH conformation but they show similar folding. However, native FH peptide sequence includes a disulfide bridge between Cys237 and Cys262, that should contribute to further stabilise the peptide conformation and reproduce the native folding.

5.2.8 Molecular Dynamics of FH peptide

In order to explore the stability of FH peptide in its folded conformation, two MD simulations were run setting a time of 500 ns by employing FH peptide in the native conformation. For both trajectories, the system energy, pressure, temperature and volume were also monitored (average energy values in table 5.11) and the RMSD plots were examined observing good stability (Plot 5.6A-B).

Table 5.11. Energy values monitored during the two MD trajectories of FH peptide

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-22464.846	58.126	0.000	-22465.720	58.070	0.000
Potential energy (kcal/mol)	-27598.547	43.959	0.000	-27599.576	44.023	0.000



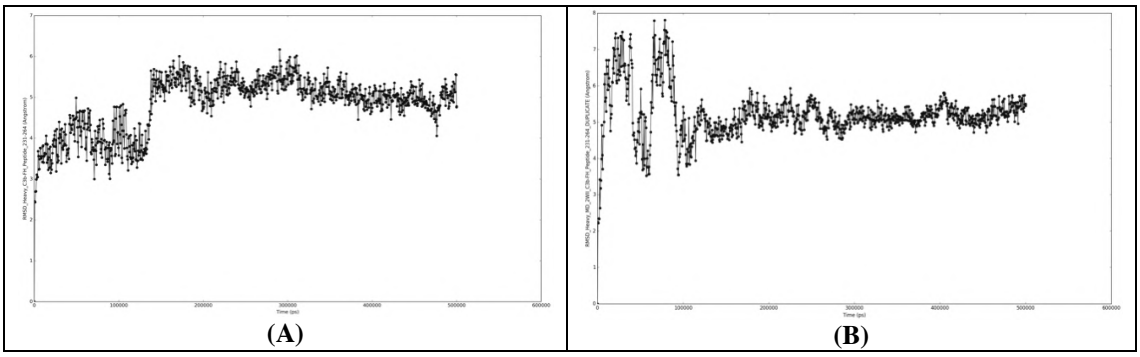
Plot 5.6. RMSD plots of first **(A)** and second **(B)** MD simulation performed on FH peptide

5.2.9 Molecular Dynamics of C3b-FH peptide complex

In order to check the stability of contacts between the designed FH peptide and C3b chains, two MD simulations of the complex were computed. The simulation time was set 500 ns and the related energy values and RMSD plots are reported respectively in table 5.12 and in plot 5.7A-B.

Table 5.12. Energy values monitored during the two MD trajectories on C3b-FH peptide complex

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-583626.391	307.453	0.000	-533943.837	297.835	0.000
Potential energy (kcal/mol)	-722449.857	236.721	0.000	-661202.901	231.987	0.000



(A) **(B)**

Plot 5.7. RMSD plots of first **(A)** and second **(B)** MD simulation performed on C3b-FH peptide complex

As it can be observed from the RMSD plots, the stationary shapes were reached after about 130 ns of trajectory for the first MD and after about 90 ns for the second MD. This initial instability of the RMSD plots should be due to the lack of CCP1-3 domains of FH proteins that were deleted from the original structure to include only the residues 231-264 of the designed FH peptide. Therefore, the analysis of the most frequent and stable interactions between C3b domains and FH peptide was based on the MD frames referring to the stable portions of the RMSD plots. As made for the previous MD simulations, even in this case the frames were clustered into 10 groups. Only the frames falling into the stable portion of the RMSD plots (after about 130 ns for the first simulation and after about 90 ns for the second MD) were considered, that is:

- First MD → Frame 660 (12 frames), frame 940 (22 frames), frame 550 (29 frames) and frame 180 (15 frames);
- Second MD → Frame 820 (75 frames), frame 110 (18 frames), frame 150 (14 frames).

Hereby the most stable and frequent interactions retrieved from the MD trajectories are reported in table 5.13.

Table 5.13. The most stable and frequent interactions between C3b domains and FH peptide during the two MD simulations

First C3b-FH peptide MD			Second C3b-FH peptide MD		
FH	**C3b**		**FH**	**C3b**	
Key residues	**MG1 Domain**	**Interaction type**	**Key residues**	**MG1 Domain**	**Interaction type**
Glu245	Lys65	1 H-bond + 1 salt bridge	Glu245	Lys65	1 H-bond + 1 salt bridge
Glu245	Gln109	1 H-bond	Glu245	Gln109	1 H-bond
Glu264	Lys73	1 H-bond + 1 salt bridge	Glu263	Lys73	1 H-bond + 1 salt bridge
Glu253	Arg94	2 H-bonds + 1 salt bridge	Glu264	Lys73	1 H-bond + 1 salt bridge
			Arg257	Glu95	2 H-bonds + 1 salt bridge
			Ser254	Glu95	1 H-bond
Key residues	**TED Domain**	**Interaction type**	**Key residues**	**TED Domain**	**Interaction type**
Arg232	Gln1131	1 H-bond	Gln234	Ser1075	1 H-bond
Arg232	Asp1134	2 H-bonds	Arg232	Asp1134	1 H-bond
Tyr243	Asp1074	1 H-bond	Arg232	Asn1069	1 H-bond
Arg246	Val1068	1 H-bond			
Lys236	Asp1074	1 H-bond + 1 salt bridge			

As it can be noticed from the analysis of these MD simulations, FH peptide fulfils the crucial interactions with MG1 domain and only one key contact, Arg232-Asp1134, with TED domain. Therefore, based on this data it seems that FH peptide shows a preference for MG1 domain of C3b. However, this information should be further investigated through biological assays.

5.3 Methods

5.3.1 Crosslinking of C3b protein in PDB 2WII

The PDB structure of C3b-FH complex (PDB ID: 2WII) presented two missing amino acids in positions 98 and 99 of C3b chain α', that is serine and glutamate, respectively. For this purpose, "Crosslink proteins" tool by Schrödinger suite was used. Therefore, Lys97 and Lys100 were set as connection residues. All the other settings, such as inter-residue distance and average chosen monomer length, were maintained by default. The linker conformation prediction was based on PDB database by selecting "loop lookup from curated PDB", and for energy calculation the implicit solvent was applied.

5.3.2 Preparation of PDB structures

In order to analyse the interactions between C3b domains and FH CCP1-4 domains, it was necessary to prepare and optimise the three PDB structures available of C3b-FH complex, that is 2WII, 5O32 and 5O35. The bond orders were assigned for untemplated residues and known HET groups and hydrogens were added. Bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds were created and the water molecules beyond 5 Å from the het groups were deleted. The protonation and metal charge states were generated for ligands, cofactors and metals by using Epik [67] at pH 7.0 ± 2.0. Finally, the H-bonds were optimised by using PROPKA [67] at pH 7.0.

5.3.3 Computational residue scanning on $C3b_{R102G}$

The PDB 2WII including the trimeric complex (C3b chain α' and chain β and FH CCP1-4) was used to perform a computational residue scanning by using the "Residue Scanning" tool of Schrödinger suite. For this purpose, Arg102 of C3b chain α' was chosen for exploring its mutation to Gly, as for the mutant form of C3b. The stability and affinity of the protein complex were computed and the resulting structures were refined by selecting side-chain prediction with backbone minimization.

5.3.4 Molecular Dynamics simulations performed on C3b and FH systems

In this work, ten MD simulations were run to explore the interactions stability and frequency within the systems under evaluation, by using the PDB 2WII. Indeed, the systems listed below were processed to perform MD simulations of 500 ns in duplicate:

- WT C3b-FH complex;
- WT C3b without FH protein;
- Mutant $C3b_{R102G}$ without FH;
- FH peptide without C3b protein;
- WT C3b-FH peptide complex.

For each system, the "system builder" tool by Schrödinger was applied. TIP3P [68] was used as a solvent model and the box shape was set as orthorhombic. The box size calculation method was selected as buffer and the box side distances were set as 10.0 Å, except for an MD simulation of WT C3b-FH complex where the box side distances were indicated by 12 Å. Finally, the force field OPLS3 [69] was applied. Each system was then used to run Molecular Dynamics simulations [70]. During the whole trajectories the number of molecules, the pressure and the temperature were maintained constant, whereas temperature and pressure were set 300 K and 1.01325 bar, respectively. Finally, the systems were relaxed before simulations.

5.3.5 Metadynamics simulation on mutant $C3b_{R102G}$ in complex with FH CCP1-4 domains

The system for the metadynamics simulation was built by using Desmond [70] for Schrödinger suite and choosing TIP3P [68] as a solvent model. The orthorhombic shape was selected for the simulation box, and the box size calculation method was selected as buffer with box side distances set by 10.0 Å. The system was neutralised by adding Na^+ ions and the force field applied was OPLS3 [69].

The metadynamics simulation was run by using Desmond as well and the collective variables selected were the dihedral angles Gly102 in C3b MG1 domain with a width of 5.0 degrees. During the trajectory, the number of molecules, the pressure and the temperature were maintained constant, whereas temperature and pressure were set 300 K and 1.01325 bar, respectively. The height of the Gaussian potential was set 0.03 kcal/mol with an interval of 0.09 ps. The simulation time was set 100 ns and the system was relaxed before running the simulation.

5.3.6 Clustering of MD frames

In order to retrieve the key contacts between the protein partners during the entire simulations, for each trajectory, the MD frames were clustered to identify the most abundant and representative frames to be analysed. Therefore, "Desmond trajectory clustering" tool by Schrödinger was used. For the RMSD matrix calculation the protein backbone was used, the frequency of frames analysis was set 10 and the hierarchical cluster linkage method as average. Finally, for each MD trajectory, 10 clusters were generated.

5.4 Conclusions

The above-described work was performed at the School of Pharmacy and Pharmaceutical Sciences of Cardiff University (UK) under the supervision of Professor Andrea Brancale. All the collected and analysed data extracted from MD simulations suggested that the initially formulated hypothesis about the destabilising effect of $C3b_{R102G}$ mutation could be likely. Indeed, all the data above reported showed a certain instability at the expenses of TED domain suggesting that R102G mutation in MG1 domain could decrease the stable contacts with TED domain. This fact could also cause movements of the TED region, that could indirectly affect and impair also the FH binding.

The next step of this work will be a further investigation of this hypothesis at Cardiff University laboratories by synthesising the designed FH peptide and assaying it by performing biological assays. The results will allow validating the above-mentioned considerations and will provide crucial information about the potential activity of the designed peptide to discriminate the wild-type form of C3b from the single-base mutated $C3b_{R102G}$ associated to AMD disease.

Finally, this work has been considering for a research article and it will be submitted to a scientific peer-reviewed journal.

References – Chapter Five

1. Meidani M, Naeini AE, Rostami M, et al (2014) Immunocompromised patients: Review of the most common infections happened in 446 hospitalized patients. J Res Med Sci Off J Isfahan Univ Med Sci 19:S71-3
2. Gordon S (2002) Pattern Recognition Receptors: Doubling Up for the Innate Immune Response. Cell 111:927–930. https://doi.org/10.1016/S0092-8674(02)01201-1
3. de Jong HK, van der Poll T, Wiersinga WJ (2010) The Systemic Pro-Inflammatory Response in Sepsis. J Innate Immun 2:422–430. https://doi.org/10.1159/000316286
4. Merle NS, Church SE, Fremeaux-Bacchi V, Roumenina LT (2015) Complement System Part I: Molecular Mechanisms of Activation and Regulation. Front Immunol 6:. https://doi.org/10.3389/fimmu.2015.00262
5. Merle NS, Noe R, Halbwachs-Mecarelli L, et al (2015) Complement System Part II: Role in Immunity. Front Immunol 6:. https://doi.org/10.3389/fimmu.2015.00257
6. Bajic G, Degn SE, Thiel S, Andersen GR (2015) Complement activation, regulation, and molecular basis for complement-related diseases. EMBO J 34:2735–2757. https://doi.org/10.15252/embj.201591881
7. Fujita T, Matsushita M, Endo Y (2004) The lectin-complement pathway - its role in innate immunity and evolution. Immunol Rev 198:185–202. https://doi.org/10.1111/j.0105-2896.2004.0123.x
8. Ting JP-Y, Lovering RC, Alnemri ES, et al (2008) The NLR Gene Family: A Standard Nomenclature. Immunity 28:285–287. https://doi.org/10.1016/j.immuni.2008.02.005
9. Pasare C, Medzhitov R Toll-Like Receptors: Linking Innate and Adaptive Immunity. In: Mechanisms of Lymphocyte Activation and Immune Regulation X. Springer US, Boston, MA, pp 11–18
10. Motta V, Soares F, Sun T, Philpott DJ (2015) NOD-Like Receptors: Versatile Cytosolic Sentinels. Physiol Rev 95:149–178. https://doi.org/10.1152/physrev.00009.2014
11. Kim YK, Shin J-S, Nahm MH (2016) NOD-Like Receptors in Infection, Immunity, and Diseases. Yonsei Med J 57:5. https://doi.org/10.3349/ymj.2016.57.1.5
12. Brubaker SW, Bonham KS, Zanoni I, Kagan JC (2015) Innate Immune Pattern Recognition: A Cell Biological Perspective. Annu Rev Immunol 33:257–290. https://doi.org/10.1146/annurev-immunol-032414-112240
13. Oppermann M, Würzner R (2010) Modern Determination of Complement Activation. Semin Thromb Hemost 36:611–619. https://doi.org/10.1055/s-0030-1262882
14. Nilsson B, Ekdahl KN (2012) Complement Diagnostics: Concepts, Indications, and Practical Guidelines. Clin Dev Immunol 2012:1–11. https://doi.org/10.1155/2012/962702
15. Bergseth G, Ludviksen JK, Kirschfink M, et al (2013) An international serum standard for application in assays to detect human complement activation products. Mol Immunol 56:232–239. https://doi.org/10.1016/j.molimm.2013.05.221
16. Thurman JM, Holers VM (2006) The Central Role of the Alternative Complement Pathway in Human Disease. J Immunol 176:1305–1310. https://doi.org/10.4049/jimmunol.176.3.1305
17. Harboe M, Ulvund G, Vien L, et al (2004) The quantitative role of alternative pathway amplification in classical pathway induced terminal complement activation. Clin Exp Immunol 138:439–446. https://doi.org/10.1111/j.1365-2249.2004.02627.x
18. Pasupuleti M, Walse B, Nordahl EA, et al (2007) Preservation of Antimicrobial Properties of Complement Peptide C3a, from Invertebrates to Humans. J Biol Chem 282:2520–2528. https://doi.org/10.1074/jbc.M607848200
19. Nordahl EA, Rydengard V, Nyberg P, et al (2004) Activation of the complement system generates antibacterial peptides. Proc Natl Acad Sci 101:16879–16884. https://doi.org/10.1073/pnas.0406678101
20. Malmsten M, Davoudi M, Walse B, et al (2007) Antimicrobial peptides derived from growth factors. Growth Factors 25:60–70. https://doi.org/10.1080/08977190701344120
21. Pedersen MB, Zhou X, Larsen EKU, et al (2010) Curvature of Synthetic and Natural Surfaces Is an Important Target Feature in Classical Pathway Complement Activation. J Immunol 184:1931–1945. https://doi.org/10.4049/jimmunol.0902214
22. Vorup-Jensen T, Boesen T (2011) Protein ultrastructure and the nanoscience of complement activation. Adv Drug Deliv Rev 63:1008–1019. https://doi.org/10.1016/j.addr.2011.05.023
23. Barnum SR (2017) Complement: A primer for the coming therapeutic revolution. Pharmacol

Ther 172:63–72. https://doi.org/10.1016/j.pharmthera.2016.11.014
24. Serna M, Giles JL, Morgan BP, Bubeck D (2016) Structural basis of complement membrane attack complex formation. Nat Commun 7:10587. https://doi.org/10.1038/ncomms10587
25. Aleshin AE, Schraufstatter IU, Stec B, et al (2012) Structure of Complement C6 Suggests a Mechanism for Initiation and Unidirectional, Sequential Assembly of Membrane Attack Complex (MAC). J Biol Chem 287:10210–10222. https://doi.org/10.1074/jbc.M111.327809
26. Aleshin AE, DiScipio RG, Stec B, Liddington RC (2012) Crystal Structure of C5b-6 Suggests Structural Basis for Priming Assembly of the Membrane Attack Complex. J Biol Chem 287:19642–19652. https://doi.org/10.1074/jbc.M112.361121
27. Morgan BP (2016) The membrane attack complex as an inflammatory trigger. Immunobiology 221:747–751. https://doi.org/10.1016/j.imbio.2015.04.006
28. Brouwers N, Van Cauwenberghe C, Engelborghs S, et al (2012) Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. Mol Psychiatry 17:223–33. https://doi.org/10.1038/mp.2011.24
29. Lambert J-C, Heath S, Even G, et al (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094–1099. https://doi.org/10.1038/ng.439
30. Zipfel PF, Skerka C (2009) Complement regulators and inhibitory proteins. Nat Rev Immunol 9:729–740. https://doi.org/10.1038/nri2620
31. Lambris JD, Lao Z, Oglesby TJ, et al (1996) Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component. J Immunol 156:4821–32
32. Xue X, Wu J, Ricklin D, et al (2017) Regulator-dependent mechanisms of C3b processing by factor I allow differentiation of immune responses. Nat Struct Mol Biol 24:643–651. https://doi.org/10.1038/nsmb.3427
33. Ross G, Newman S, Lambris J, et al (1983) Generation of three different fragments of bound C3 with purified factor I or serum. II. Location of binding sites in the C3 Fragments for Factors B and H, complement receptors , and bovine conglutinin. J Exp Med 158:334–352. https://doi.org/10.1084/jem.158.2.334
34. Schmidt CQ, Herbert AP, Hocking HG, et al (2008) Translational mini-review series on complement factor H: structural and functional correlations for factor H. Clin Exp Immunol 151:14–24. https://doi.org/10.1111/j.1365-2249.2007.03553.x
35. Schmidt CQ, Herbert AP, Kavanagh D, et al (2008) A new map of glycosaminoglycan and C3b binding sites on factor H. J Immunol 181:2610–9. https://doi.org/10.4049/jimmunol.181.4.2610
36. Gordon DL, Kaufman RM, Blackmore TK, et al (1995) Identification of complement regulatory domains in human factor H. J Immunol 155:348–56
37. Wiesmann C, Katschke KJ, Yin J, et al (2006) Structure of C3b in complex with CRIg gives insights into regulation of complement activation. Nature 444:217–220. https://doi.org/10.1038/nature05263
38. Janssen BJC, Christodoulidou A, McCarthy A, et al (2006) Structure of C3b reveals conformational changes that underlie complement activity. Nature 444:213–216. https://doi.org/10.1038/nature05172
39. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
40. Becherer JD, Alsenz J, Esparza I, et al (1992) Segment spanning residues 727-768 of the complement C3 sequence contains a neoantigenic site and accommodates the binding of CR1, factor H, and factor B. Biochemistry 31:1787–1794. https://doi.org/10.1021/bi00121a029
41. Oran AE, Isenman DE (1999) Identification of Residues within the 727–767 Segment of Human Complement Component C3 Important for Its Interaction with Factor H and with Complement Receptor 1 (CR1, CD35). J Biol Chem 274:5120–5130. https://doi.org/10.1074/jbc.274.8.5120
42. Gros P, Milder FJ, Janssen BJC (2008) Complement driven by conformational changes. Nat Rev Immunol 8:48–58. https://doi.org/10.1038/nri2231
43. Janssen BJC, Huizinga EG, Raaijmakers HCA, et al (2005) Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437:505–11. https://doi.org/10.1038/nature04005
44. Wu J, Wu Y-Q, Ricklin D, et al (2009) Structure of complement fragment C3b–factor H and implications for host protection by complement regulators. Nat Immunol 10:728–733. https://doi.org/10.1038/ni.1755
45. Yates JRW, Sepp T, Matharu BK, et al (2007) Complement C3 Variant and the Risk of Age-

Related Macular Degeneration. N Engl J Med 357:553–561. https://doi.org/10.1056/NEJMoa072618
46. Spencer KL, Olson LM, Anderson BM, et al (2008) C3 R102G polymorphism increases risk of age-related macular degeneration. Hum Mol Genet 17:1821–1824. https://doi.org/10.1093/hmg/ddn075
47. Rambausek M, van den Wall Bake AW, Schumacher-Ach R, et al (1987) Genetic polymorphism of C3 and Bf in IgA nephropathy. Nephrol Dial Transplant 2:208–11
48. Finn JE, Zhang L, Agrawal S, et al (1994) Molecular analysis of C3 allotypes in patients with systemic vasculitis. Nephrol Dial Transplant 9:1564–7
49. Andrews PA, Finn JE, Mathieson PW, Sacks SH (1995) Molecular analysis of C3 allotypes related to transplant outcome in human renal allografts. Transplantation 60:1342–1346
50. Finn JE, Mathieson PW (2008) Molecular analysis of C3 allotypes in patients with nephritic factor. Clin Exp Immunol 91:410–414. https://doi.org/10.1111/j.1365-2249.1993.tb05917.x
51. Kühn S, Zipfel PF (1996) Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. Eur J Immunol 26:2383–2387. https://doi.org/10.1002/eji.1830261017
52. (2020) RCSB Protein Data Bank. https://www.rcsb.org. Accessed 18 May 2020
53. Chakrabartty A, Schellman JA, Baldwin RL (1991) Large differences in the helix propensities of alanine and glycine. Nature 351:586–588. https://doi.org/10.1038/351586a0
54. Merkel JS, Regan L (1998) Aromatic rescue of glycine in β sheets. Fold Des 3:449–456. https://doi.org/10.1016/S1359-0278(98)00062-5
55. Rodrigues CH, Pires DE, Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res 46:W350–W355. https://doi.org/10.1093/nar/gky300
56. Kuzmanic A, Zagrovic B (2010) Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. Biophys J 98:861–871. https://doi.org/10.1016/j.bpj.2009.11.011
57. Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. WIREs Comput Mol Sci 1:826–843. https://doi.org/10.1002/wcms.31
58. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins Struct Funct Bioinforma n/a-n/a. https://doi.org/10.1002/prot.24065
59. Xu D, Zhang Y (2013) Toward optimal fragment generations for ab initio protein structure assembly. Proteins Struct Funct Bioinforma 81:229–239. https://doi.org/10.1002/prot.24179
60. Camproux A., Gautier R, Tufféry P (2004) A Hidden Markov Model Derived Structural Alphabet for Proteins. J Mol Biol 339:591–605. https://doi.org/10.1016/j.jmb.2004.04.005
61. Maupetit J, Derreumaux P, Tufféry P (2009) A fast method for large-scale De Novo peptide and miniprotein structure prediction. J Comput Chem NA-NA. https://doi.org/10.1002/jcc.21365
62. Maupetit J, Tuffery P, Derreumaux P (2007) A coarse-grained protein force field for folding and structure prediction. Proteins Struct Funct Bioinforma 69:394–408. https://doi.org/10.1002/prot.21505
63. Harpreet Kaur, Aarti Garg, G.P.S. Raghava (2007) PEPstr: A de novo Method for Tertiary Structure Prediction of Small Bioactive Peptides. Protein Pept Lett 14:626–631. https://doi.org/10.2174/092986607781483859
64. Wang Z, Eickholt J, Cheng J (2011) APOLLO: a quality assessment service for single and multiple protein models. Bioinformatics 27:1715–1716. https://doi.org/10.1093/bioinformatics/btr268
65. Beaufays J, Lins L, Thomas A, Brasseur R (2012) In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. J Pept Sci 18:17–24. https://doi.org/10.1002/psc.1410
66. Shen Y, Maupetit J, Derreumaux P, Tufféry P (2014) Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. J Chem Theory Comput 10:4745–4758. https://doi.org/10.1021/ct500592m
67. Shelley JC, Cholleti A, Frye LL, et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 21:681–691. https://doi.org/10.1007/s10822-007-9133-z
68. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w

69. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864
70. (2006) SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Association for Computing Machinery, New York, NY, USA

CHAPTER SIX

HOX-PBX-DNA COMPLEX – A methodology to design peptides potentially preventing HOX-PBX-DNA complex formation

6.1 Introduction

Development, evolution and physiopathological processes in bilaterian animals are very complex events and involve several key actors, such as HOX genes. These latter have been shown actively involved in the control of the final morphology of bilaterian animals [1, 2]. Indeed, both decrease and increase of HOX genes activity can often produce homeotic transformations resulting in the formation of a structure or organ in erroneous locations within the animal organism. HOX genes have reported three different levels of evolutionary conservation: 1) at a structural level, Hox genes are usually organised in complexes, reflecting their phylogeny and regulatory aspects of their expression [3, 4]; 2) at a molecular level, they all encode homeodomain transcription factors [5]; and 3) at a functional level, they trigger similar effects in most animals and can work in substitution of an orthologue in other species [6].

In the past, HOX genes were known as developmental genes, since in general mammals present 39 HOX genes that operate at a very early stage of the embryonic development by patterning the main embryonic anterior to posterior axis [7]. However, they also play key roles in adults [8], whereas the most known Hox gene functions in adults are the following ones: the maintenance of the hematopoietic stem cells (HSCs) [9], the specification of different blood cell lineages [10] and the regulation of tissue identity during implantation and the menstrual cycle [11]. Therefore, when the Hox genes become highly dysregulated and overexpressed, they have been reported associated with a wide range of both solid and haematological cancers [12].

In the last decades, the processes modulated by HOX genes have been extensively studied providing a substantial, although not exhaustive, analysis of them. For example, HOX genes have been shown to define cellular territories and establishing boundaries. Recent studies highlighted that HOX genes also contribute to organogenesis [13] by influencing a huge number of cellular functions such as differentiation, proliferation, migration or death [14]. Indeed, HOX proteins also regulate the transcriptional events, although the related molecular mechanisms are still poorly characterised, and a few HOX-dependent gene regulatory networks have been understood.

HOX proteins present two different and highly conserved portions: the hexapeptide (HX) motif and the homeodomain (HD). The HX motif establishes interactions with protein members of the PBC class, such as Pre-B-cell Leukemia Homeobox (PBX) proteins in humans and Extradenticle (Exd) in Drosophila [15], while the HD motif is the DNA-binding domain. The HD folds into a triple-helix structure, including the N-terminal arm binding the minor groove of DNA, and helix 3 (also named the recognition helix) contacting the DNA in the major groove. Amino acids in HOX HDs helices 1 and 3 have shown to be the most conserved, and some residues of the N-terminal arm and loops between the helices have been reported well conserved. Furthermore, the

conservation of HD sequences is highly shared among HOX proteins, raising the issue of how they employ functional specificity [16–18].

Indeed, the homeodomain of HOX proteins does not exhibit high specificity for DNA, as it takes part in the molecular recognition by involving only five amino acids [19]. In this context, functional studies in the field of cancer and developmental biology highlighted that PBX proteins function as HOX co-factors [20–22], where PBX family members bind to HOX proteins 1-11 [23–25]. These proteins may establish a cooperative binding to DNA [26, 27] (figure 6.1), indicating that the interaction of HOX proteins to PBX proteins modifies the DNA-binding of HOX and contribute to a greater specificity [28]. Furthermore, the HOX co-factors play other key roles influencing transcriptional events, by recruiting the RNA polymerase II and III or transcriptional inhibitors like HDAC, and post-translational events, by fostering the entry of HOX proteins into the nucleus.

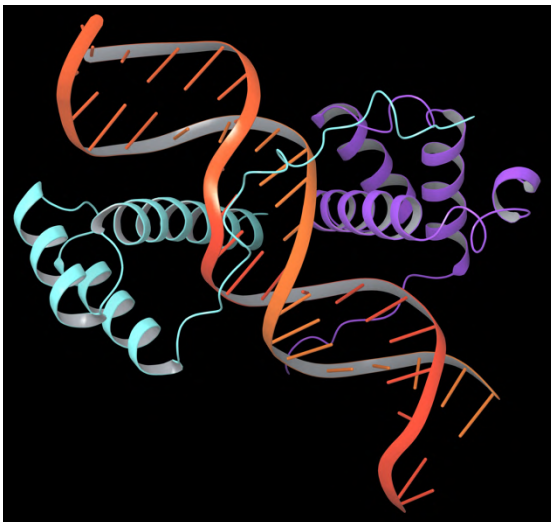


Figure 6.1. HOXA9-PBX1-DNA complex retrieved from PDB 1PUF [29], where the orange structure is DNA bound to the Homeobox protein HOXA9 (light blue chain) and Pre-B-cell leukaemia transcription factor-1 PBX1 (purple chain)

PBX genes are homologues of the Drosophila Extradenticle gene (Exd) and four different types of genes are encoded in the human genome (PBX1-4). Like the HOX genes, PBX genes also encode evolutionarily conserved homeodomains and other highly conserved regions [21]. PBX proteins also show two nuclear localization signals (NLSs) in the homeodomain and a nuclear export sequence (NES) [30–32].

PBX proteins may participate in a DNA binding consensus through the formation of strong complexes with HOX1-11 proteins [26, 33, 34]. HOX-PBX interactions were shown to involve a highly conserved interaction mode between the HX motif of HOX and the three-amino acid loop extension (TALE) or three-amino acid insertion peptide of PBX, that lies between helices 1 and 2 of the homeodomain [23, 29, 34–38]. Experimental evidence highlighted that a full-length PBX1 protein alone is not able to trigger the transcription, but PBX1 amino acids 39 to 232 can specifically stop transcriptional activation [39, 40]. Furthermore, upon binding to a HOX protein, PBX switches from a transcriptional repressor to a transcriptional activator [41].

6.1.1 Hox minimal DNA-binding sequence

In 1995, Knoepfler and Kamps [19] identified the minimal sequences that enable certain HOX proteins to bind cooperatively PBX1 protein. The deletion mutagenesis of HOXB8 and HOXA5 proteins allowed to identify the minimal required sequences to the PBX1 homeodomain and a short N-terminal homeodomain section. This minimal sequence was the conserved pentapeptide motif Y/F-P-W-M-R/K.

Point mutational analysis of this pentapeptide sequence in HOX proteins highlighted that the residues tryptophan and methionine are crucial for the cooperative binding to PBX proteins but dispensable for DNA binding. Experimental evidence revealed that synthetic peptides including the wild-type pentapeptide abrogated the cooperative binding between multiple HOX proteins and PBX1, while a peptide with a mutation at the tryptophan residue did not produce binding abrogation. Moreover, peptides including the pentapeptide fostered the DNA binding of PBX1.

In particular, mutations of Trp135 to phenylalanine (W135F) or alanine (W135A) performed on HOXB8 did not alter the DNA binding but completely abolished the cooperativity of HOXB8 with PBX1. On the other hand, Met136 had also been reported as an important but not essential amino acid, owing to its substitution to isoleucine (M136I) or alanine (M136A) did not alter the DNA binding properties but strongly disrupted the cooperativity of HOXB8 with PBX1. Therefore, both residues, Trp135 and Met136 were highlighted as crucial for the pentapeptide motif, although Trp135 showed greater importance [37]. This data was also enforced by the fact that tryptophan is the only shared amino acid among all Hox proteins pentapeptide sequences. Finally, Pro134 of HOXB8 was found dispensable, because its substitution to leucine (P134L) did not alter both the DNA-binding and the cooperativity with PBX1.

Thus, Knoepfler and Kamps assumed that the HOX pentapeptide motif stabilises the trimeric complex HOX-PBX1-DNA by bearing a portion of HOX protein surface that contacts with PBX1, and enhancing the DNA-binding in presence of PBX1 [19].

However, the X-ray crystallographic structures of HOXB1–PBX1 and HOXA9–PBX1 in presence of DNA revealed that the protein-protein-DNA contacts are stabilised by the interaction between HOX and PBX mediated no more by a pentapeptide sequence but a conserved hexapeptide sequence in Hox proteins [5, 23, 29, 36].

Indeed, in 1999 Piper *et al.* [23] compared PBX1-binding motifs in 80 HOX proteins yielding a consensus sequence that integrated the above-mentioned pentapeptide with a hydrophobic residue. Thus, the consensus identified hexapeptide motif of HOX proteins was

φ-Y/F-P-W-M-K/R

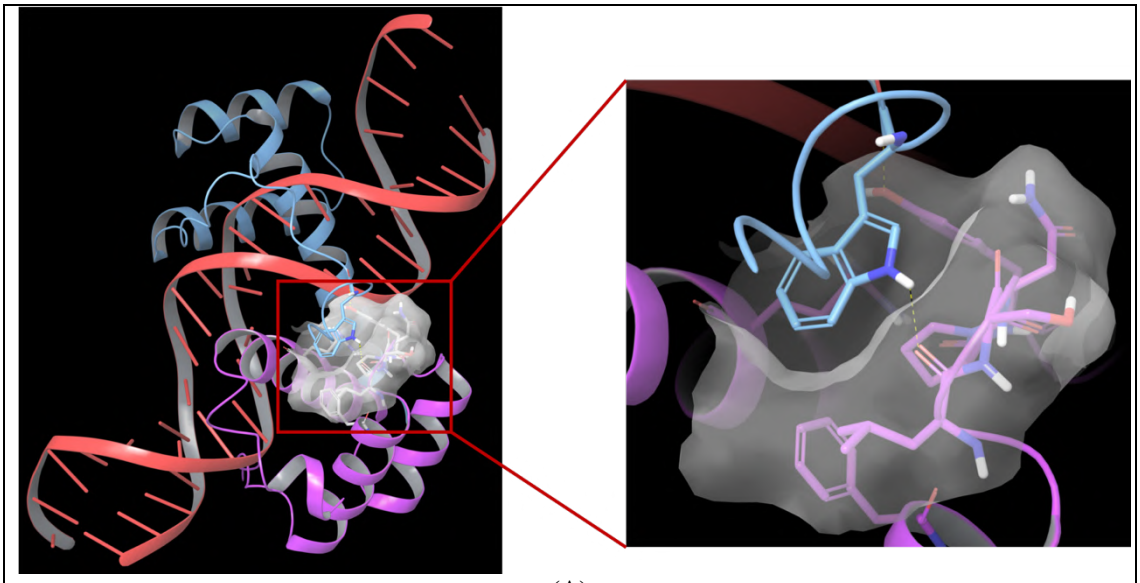
where φ stands for a hydrophobic residue. As above reported, tryptophan and methionine showed strict conservation that for example in the HOXB1–PBX1 structure can be due to several contacts mediated by these amino acids.

This hexapeptide is joined to the N-terminal arm of the homeodomain by a linker that exhibits different length and sequence among the HOX proteins and species. The authors demonstrated that this minimal Hox portion containing the hexapeptide and the homeodomain was able to cooperatively stabilise the DNA binding with PBX1. It could be also assumed that differences among HOX proteins in terms of PBX1-mediated DNA binding should work as a regulation mechanism to define different affinities of HOX-PBX1 complexes for DNA.

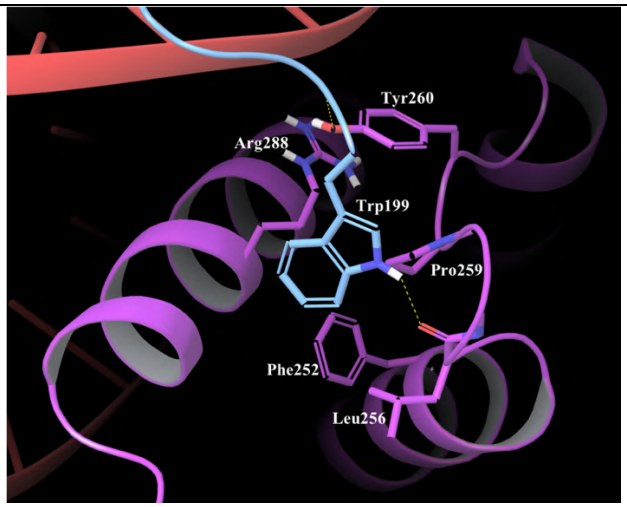
The X-ray crystallographic structures of the ternary complex, HOX–PBX1–DNA (*e.g.* HOXA9 in PDB 1PUF, resolution: 1.90 Å; and HOXB1 in PDB 1B72, resolution: 2.35 Å) revealed that the HOX protein and PBX1 establish contacts with opposite faces of the DNA, burying 2400 $Å^2$ of protein and DNA surface area. The HOX hexapeptide mediates contacts with PBX1 within a hydrophobic pocket that is located between the three-amino acid insertion and helix 3 of the PBX1 homeodomain.

PCR site-selection experiments performed by Piper *et al.* [23] allowed to identify the optimal HOXB1–PBX1 binding site on the 20 bp duplex DNA oligonucleotide, that is 5'-ATGATTGATCG-3' [42]. The PDB structure solved by La Ronde-Le Blanc and Wolberger [29] revealed that the interactions between HOXA9 and PBX1 are mediated by the hexapeptide of HOXA9, consisting of the residues 196 to 201 with the sequence AANWLH bound to the PBX1 homeodomain. Amino acids 196 to 199 of HOXA9 hexapeptide motif form a 3_{10} helix, and this hexapeptide is connected to the HOXA9 homeodomain through a flexible linker including 20 amino acids that exhibit a disordered structure in the trimeric complex.

The main intramolecular interactions established by the hexapeptide residues are hydrophobic. In detail, the side chains of HOXA9 Trp199 inserts into a hydrophobic pocket of PBX1 consisting of the C terminus of helix 3, a handle between helices 3 and 4, and the three–amino acid insertion (figure 6.2.A). The main interactions observed are established between the following amino acids of HOXA9 and PBX1: HOXA9 Trp199 with its indole ring that forms van der Waals contacts with several PBX1 residues, such as Phe252 side chain in helix 1, Leu256 within the three–amino acid insertion, Pro259 and Tyr260 following the TALE peptide, and Arg288 in helix 3. Furthermore, Trp199 is highly buried into the PBX1 binding pocket by forming a hydrogen bond between the indole nitrogen and the backbone carbonyl of PBX1 Leu256 (figure 6.2B).



(A)



(B)

Figure 6.2. (A) Binding pocket surface (grey surface) of PBX1 protein (purple chain) surrounding Trp199 of HOXA9 (light blue chain) hexapeptide retrieved from PDB 1PUF [29]; **(B)** Amino acids composing PBX1 binding pocket (purple residues and chain) surrounding HOXA9 Trp199

Leu200 backbone nitrogen of the HOXA9 hexapeptide establishes van der Waals contacts in the binding pocket of PBX1 with Ly292 and a hydrogen bond with Tyr260 hydroxyl group. Finally, His201 of HOXA9 hexapeptide forms a hydrogen bond with Lys292 of PBX1.

Furthermore, mutational studies performed by Piper *et al*. [23] on PBX1 were performed on the hexapeptide-contacting residues Leu252 and Pro259, that were substituted to alanine resulting in the disruption of the interactions with the hexapeptide *in vitro* and in a yeast two-hybrid assay [43]. Moreover, deletion assays involving the three–amino acid insertion, that should destabilise the binding pocket, abrogated the cooperative binding of PBX1 with HOX proteins [44]. On the other hand, the deletion of the HOX hexapeptide caused the disappearance of cooperative interactions between PBX1 and HOX proteins [19, 34, 45].

Although the 3D ternary structures HOX-PBX-DNA are available, the high level of functional redundancy among the several HOX proteins and some issues associated with the design of effective small molecule inhibitors of HOX-PBX interaction have made difficult the drug discovery process for this PPI. An accepted strategy is to target HOX and PBX binding interface, due to the presence of highly conserved residues especially in HOX proteins hexapeptide and for the hydrophobic nature of PBX proteins binding pocket. In the last decades, a small molecule inhibitor of this interaction was identified. However, its K_D was in the micromolar range (65 μM) and it was neglected for further experimental assays or clinical trials [46]. In the last years, several peptides have been designed based on the hexapeptide consensus motif of HOX proteins, to act as a competitive antagonist of HOX-PBX binding [47]. The most frequently used peptide among these is HXR9, an 18-amino acid peptide containing the hexapeptide sequence together with nine arginine residues, that promote cellular uptake by endocytosis with the sequence WYKWMKKAARRRRRRRRR.

The peptide HXR9 was first shown to be cytotoxic to melanoma cell lines and primary melanoma cells and registered a reduction of B16F10 murine melanoma tumours growth in an orthotropic model [48]. Other experimental studies reported that HXR9 was

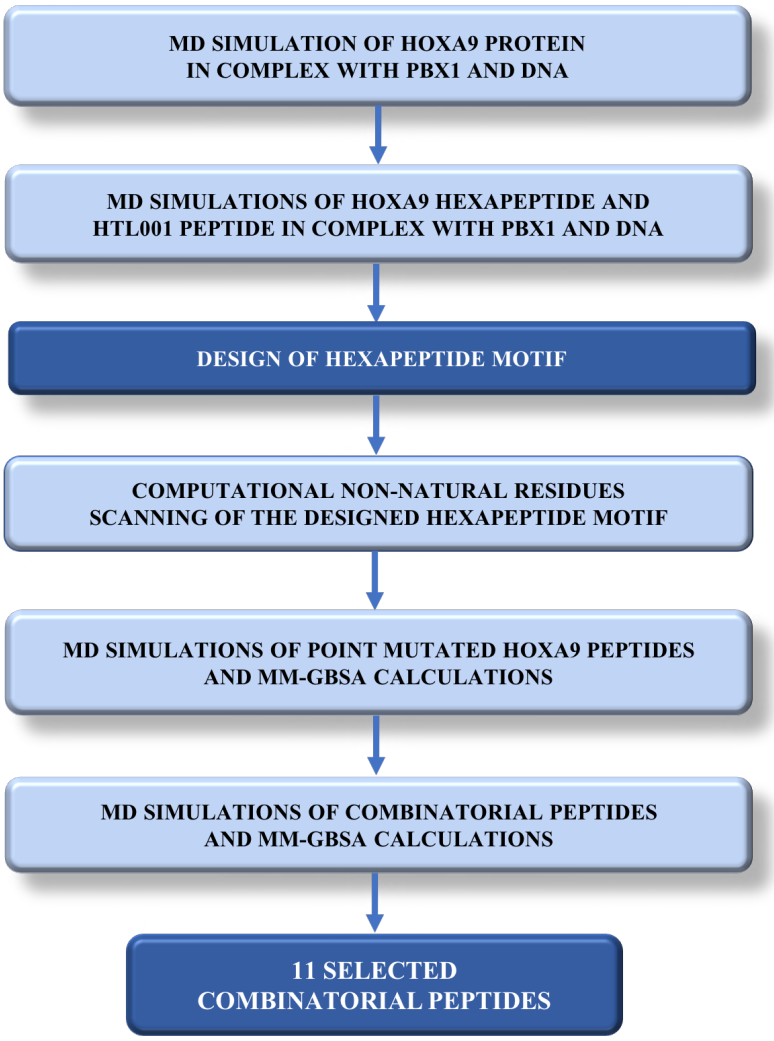
able to inhibit the growth of several tumour types in mouse xenograft models, including non-small cell lung [49], breast [50], ovarian [50], and prostate cancer [51], and mesothelioma [52], melanoma [53], and meningioma [54].

Recently, another peptide was identified based on modifications of HXR9 sequence, *i.e.* the HTL001 peptide [55] with the sequence **WYPWMKKHHRRRRRRRRR,** that was tested in cancer cells representative of 14 different malignancies, in human and animal cancer tissues. This peptide demonstrated selective toxicity for cancer cells and safety for normal cells. To date, human clinical trials are ongoing to test the efficacy and safety of this novel peptide.

Although the two above-mentioned peptides showed efficacy to inhibit this PPI under study, the mechanism associated with HOX-PBX inhibition and the resulting cell death is still to be fully elucidated. Generally, in most solid tumours cell death is mediated by apoptosis [48, 50–52, 56].

The work herein described was based on the design of novel peptides including non-standard amino acids potentially binding PBX1 and inhibit HOX-PBX1 interaction. For this purpose, Molecular Dynamics simulations and MM-GBSA (Molecular Mechanics – Generalised Born Surface Area) calculations were exploited and computational non-standard residue scanning of the consensus hexapeptide motif was performed. First of all, an MD simulation of 200 ns was performed on the ternary complex HOXA9-PBX1-DNA with high-quality resolution (PDB ID: 1PUF, resolution: 1. 90 Å) to retrieve the key interactions and residues. Then, other two MD were run, one including HOXA9 hexapeptide (196-AANWLH-201) and another including the patented core peptide HTL001 without polyarginine coil in complex with PBX1-DNA complex to calculate $\Delta G_{binding}$ average values that were used as a reference for the next MD calculations.

Then, HOXA9 hexapeptide was processed by a point mutational scanning using a non-natural amino acids database populated by the Swiss Institute of Bioinformatics [57, 58]. The mutations were selected according to $\Delta\Delta G_{affinity}$ and $\Delta\Delta G_{stability}$ and were further explored by applying MD simulations and MM-GBSA calculations. All those residues reporting $\Delta G_{binding}$ average value lower compared to those of HOXA9 hexapeptide and HTL001 peptide were chosen for the next steps of the work. Thus, the selected mutations were combined providing overall twelve combinatorial peptides. These peptides in turn were used to run other MD simulations and MM-GBSA calculations. Finally, eleven of the initial twelve peptides presented lower $\Delta G_{binding}$ values in comparison with HOXA9 hexapeptide and HTL001 peptide. Therefore, these eleven peptides are been considering to follow-up this study through peptide synthesis at Cardiff University to be experimentally assayed in order to check their putative ability of HOX-PBX1 interaction inhibition. All the steps of the above-described workflow are summarized in Scheme 6.1.

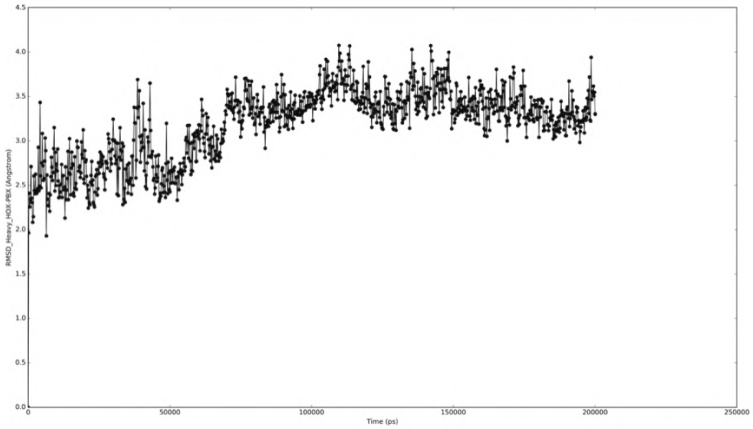


Scheme 6.1. Overview of the computational workflow performed to identify the eleven combinatorial peptides potentially inhibiting HOX-PBX1 cooperative binding

6.2 Results and discussion

6.2.1 Molecular Dynamics simulation of HOXA9-PBX1-DNA complex

The first step of this work was a molecular dynamics simulation of the trimeric complex HOXA9-PBX1-DNA through Desmond in Schrödinger suite [59]. For this purpose, the currently available PDB structure with the lowest resolution of 1.90 Å was used (PDB ID: 1PUF), in order to analyse and identify the most stable interactions ant involved residues for both proteins. The simulation time was set 200 ns and the complex stability was then investigated by observing the RMSD plot (plot 6.1), that showed a stable behaviour of the system.



Plot 6.1. RMSD plot of HOXA9-PBX1-DNA MD simulation of 200 ns

Then energy, temperature, pressure and volume were monitored during the whole trajectory and table 6.1 reports the related average energy values.

Table 6.1. Energy values of HOXA9-PBX1-DNA complex monitored during the MD simulation

	Average	Std. Dev.	Slope (ps^{-1})
Total energy (kcal/mol)	-95014.430	116.847	0.000
Potential energy (kcal/mol)	-116315.750	89.630	0.000

The system showed to be stable, hence it was further processed to retrieve the most frequent interactions established between the two proteins HOX and PBX. Therefore, MD frames were clustered into 10 groups, whereas the frames representative for the most abundant clusters were: frame 880 (representative for 63 frames), frame 60 (representative for 34 frames), frame 540 (representative for 28 frames), frame 360 (representative for 22 frames), and frame 270 (representative for 15 frames).

These frames were analysed to identify the most stable and frequent interactions during the trajectory and table 6.2 reports the residues involved for both proteins considering HOXA9 hexapeptide residues and PBX1 contacting region.

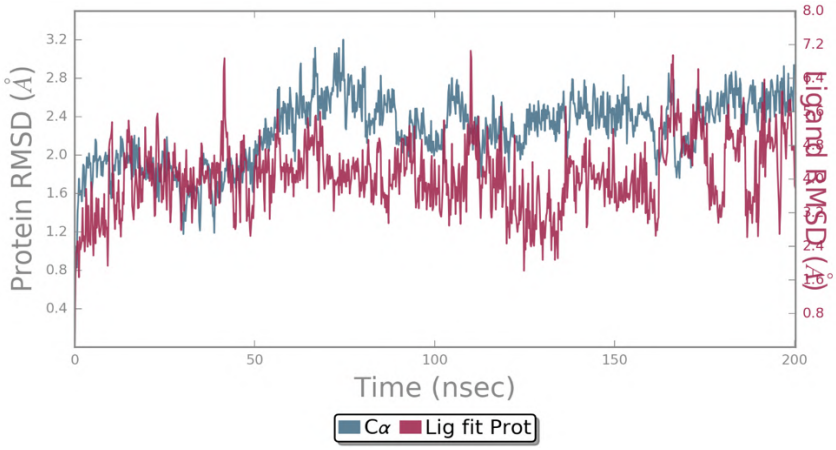
Table 6.2. Most stable and frequent interactions between HOXA9 hexapeptide and PBX1 homeodomain proteins retrieved from MD simulation

HOXA9 residue	PBX1 residue	Interaction type
Trp199	Ser257	1 H-bond
Trp199	Leu256	1 H-bond
Trp199	Tyr291	Pi-Pi stacking
Trp199	Tyr260	Pi-Pi stacking
Leu200	Tyr260	1 H-bond
Ala197	Asn258	1 H-bond

As it can be observed, Trp199 showed the majority of the interactions with PBX1 residues. This data was consistent with information from literature highlighting this tryptophan [19] as the fundamental residue.

6.2.2 MD simulations of HOXA9 hexapeptide and HTL001 peptide core in complex with PBX1

In order to collect more data about the binding mode of PBX1 in complex with the minimal active HOXA9 hexapeptide sequence, an MD simulation of 200 ns was performed by using the PDB 1PUF including HOXA9, PBX1 and DNA, where HOX protein was modified by deleting all those amino acids not included into the hexapeptide 196-AANWLH-201. Protein and ligand RMSD plot of the trajectory was analysed by registering the trend illustrated in plot 6.2 together with the interaction diagram and the bar chart in figure 6.3A-B.



Plot 6.2. PBX1 protein and HOXA9 hexapeptide RMSD plot during MD trajectory

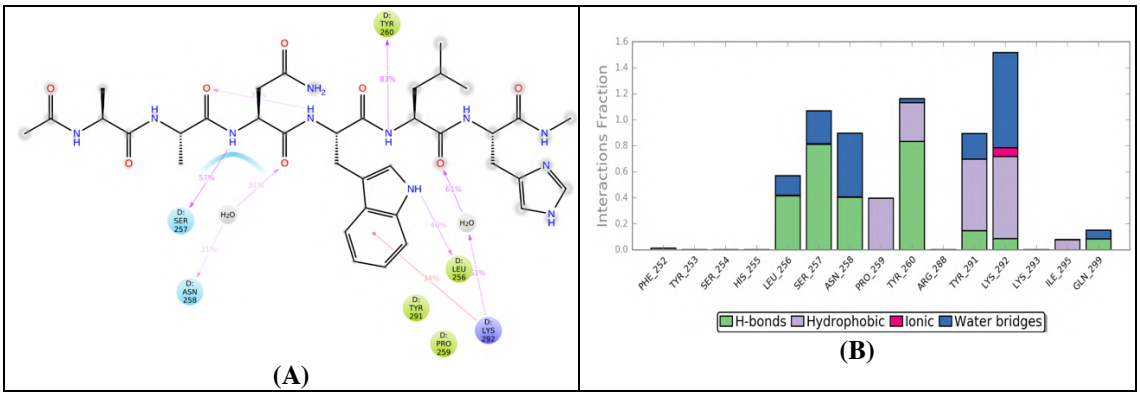
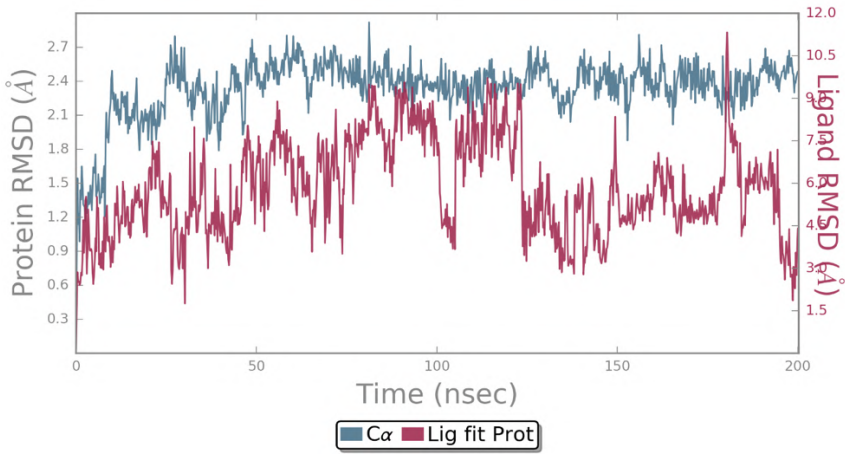


Figure 6.3. (A) HOXA9 hexapeptide interaction diagram during MD simulation; **(B)** Bar chart of protein-ligand interaction occurrences during MD simulation

To date, the patented peptide clinically employed HTL001 has been shown to act against PBX1 protein by preventing HOX-PBX cooperative binding [55]. In detail, HTL001 peptide sequence incorporates the hexapeptide WYKWMK responsible for the binding affinity with PBX proteins.

Therefore, PDB 1PUF was used for another MD simulation of 200 ns, whereas HOXA9 hexapeptide was substituted and minimised to reproduce the HTL001 hexapeptide sequence (WYKWMK) in complex with PBX1 and DNA. Then PBX1 protein and HTL001 hexapeptide RMSD values were monitored during the entire trajectory and they are reported in plot 6.3. Finally, figure 6.4A-B shows the ligand interaction diagram and the bar chart of protein-ligand interaction occurrences.



Plot 6.3. PBX1 protein and HTL001 hexapeptide RMSD plot during MD trajectory

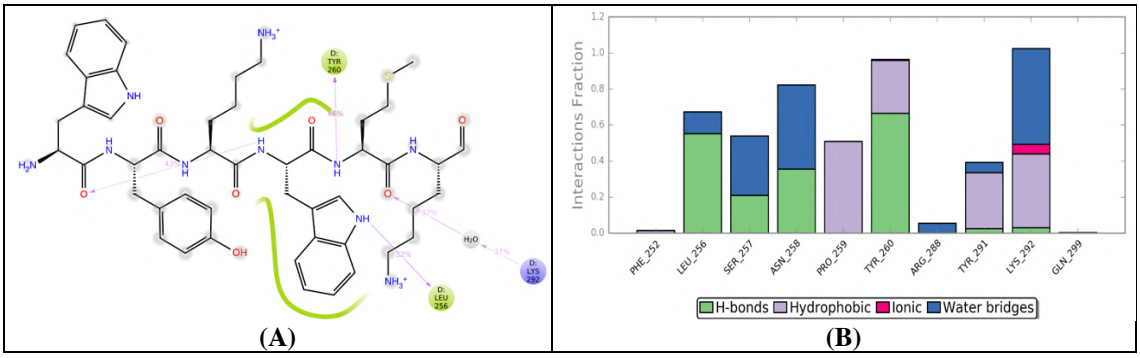


Figure 6.4. (A) HTL001 hexapeptide interaction diagram during MD simulation; **(B)** Bar chart of protein-ligand interaction occurrences during MD simulation

In order to quantitatively evaluate the interaction between HOXA9 and HTL001 hexapeptides with PBX1 and DNA, MM-GBSA were calculated for both MD simulations by reporting the $\Delta G_{binding}$ average values presented in table 6.3. Thus, these results were used as a reference for the next steps of this work to compare MM-GBSA outputs of the designed peptides below described.

Table 6.3. MM-GBSA calculation results of MD simulations performed on HOXA9 and HTL001 hexapeptides in complex with PBX1 protein and DNA

	HOXA9 HEXAPEPTIDE	HTL001 HEXAPEPTIDE
$\Delta G_{binding}$ average	-58.1922 kcal/mol	-53.6882 kcal/mol
$\Delta G_{binding}$ Std. Dev.	8.99	8.53
$\Delta G_{binding}$ range	-84.6286 to -34.1107 kcal/mol	-78.0904 to -28.9169 kcal/mol

6.2.3 Design of a potential non-standard PBX1-binding hexapeptide

As above described, experimental evidence [19, 23] highlighted the consensus HOX hexapeptide sequence φ-Y/F-P-W-M-R/K (where φ is a hydrophobic residue) [23], necessary for the cooperative binding to PBX proteins and to increase specificity for DNA. Based on this information from literature and from the above described computational data, it was possible to design a peptide motif including the key amino acid tryptophan as follows:

$$X_1\text{-}X_2\text{-}X_3\text{-}W\text{-}X_4\text{-}X_5$$

where X stands for any non-standard amino acid in order to generate peptides different from those patented (HXR9 and HTL001) including standard amino acids.

For this purpose, a database of amino acids consisting of non-natural residues, "SwissSidechain", was downloaded from the Swiss Institute of Bioinformatics website [57]. SwissSidechain is a structural and molecular mechanics database of 200 non-natural amino-acid side chains (both D and L conformations), that can be used to study *in silico* their insertion into natural peptides or proteins. Non-natural side chains were useful in this work in order to potentially increase ligand binding affinity. Indeed, HOXA9 hexapeptide and HTL001 peptide were used as a reference to identify peptides with ameliorated PBX-binding affinity by mutating standard amino acids with non-natural residues. This non-standard amino acids database designed by Gfeller *et al.* [58] demonstrated very good reliability based on a comparison between predicted and experimental binding free-energies for a BCL9 peptide targeting beta-catenin. These results indicated that such non-natural residues can be used to design novel protein-protein inhibitors. During the design process of this database, Gfeller *et al.* focused their attention on amino acid side chains with structural information known from the Protein Data Bank (PDB) [60] as well as commercially available amino acids. Non-natural amino acids that could cause modifications of the backbone (such as β-homo, cyclic or aromatic backbones, or proline derivatives) were neglected, since they are more likely to perturb the overall conformation of peptides or proteins and are therefore less amenable to molecular modelling studies. This resulted in a total of 200 non-natural side chains, among which 141 residues were present in the PDB. Hence all these non-standard residues were collected and downloadable from a *.nsr* file including parameter and topology data [58].

Therefore, this database was downloaded and uploaded in Schrödinger tool "Manage non-standard amino acids" and joined with the non-natural residue library already available in the Schrödinger suite, achieving overall 220 non-standard amino acids. Then point mutations were performed on HOXA9-PBX1-DNA complex (PDB 1PUF) by running "Residue scanning" tool for each of the five X amino acids present in the designed peptide motif and corresponding to Ala196, Ala197, Asn198, Leu200 and His201 of HOXA9 protein, while maintaining Trp199 due to its relevance for PBX binding [37].

After running mutational calculations, for each X residue of the designed peptide four non-standard amino acids were selected for further analysis. These amino acids were chosen according to the following criteria:

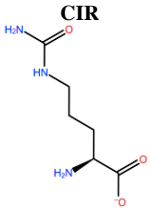
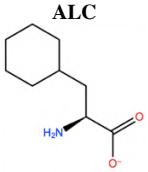
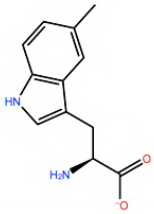
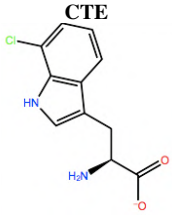
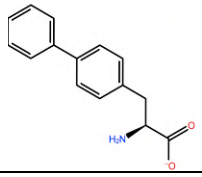
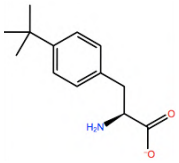
1. Lowest difference values between mutated and wild-type complexes free energies of affinity ($\Delta\Delta G_{affinity}$) tuning a cut-off of at least -3.0 kcal/mol;
2. Lowest difference values between mutated and wild-type complexes free energies of stability ($\Delta\Delta G_{stability}$);
3. Commercial availability of the non-standard amino acids.

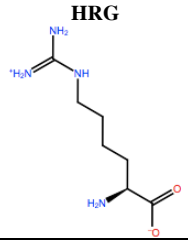
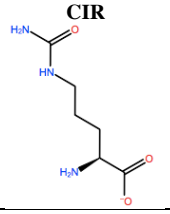
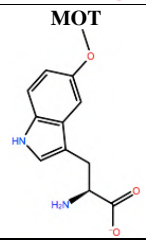
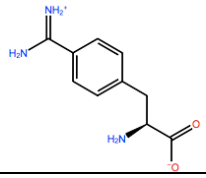
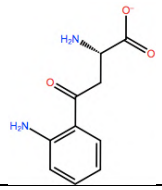
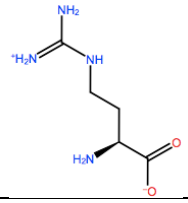
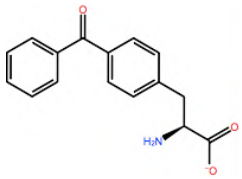
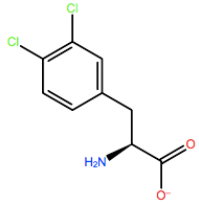
The first criterion was applied because according to Beard *et al.* [61] the predicted affinities computed with Schrödinger suite can be considered reliable when they report a difference of at least 3 kcal/mol between wild-type and mutant. On the other hand,

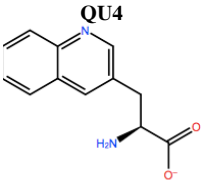
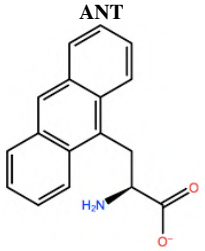
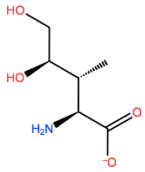
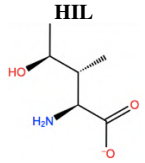
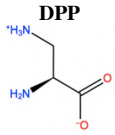
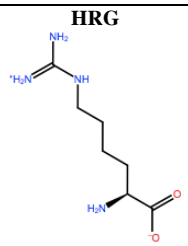
$\Delta\Delta G_{stability}$ was considered reliable if negative, without setting a cut-off because of the lack of a well-defined secondary structure for the wild-type HOXA9 hexapeptide. Hence, it should be expected that mutations should not significantly affect the $\Delta\Delta G_{stability}$ of the peptide.

In table 6.4 the selected non-standard amino acids are depicted, whereas X_1, X_2, X_3, X_4 and X_5 amino acids provided respectively 119, 10, 52, 35 and 5 acceptable mutations, but for simplicity only the best four were selected to proceed with according to the three above listed criteria owing to the other acceptable mutations reported edge $\Delta\Delta G_{affinity}$ values.

Table 6.4. Non-standard amino acids selected from residue scanning calculations according to the above-described selection criteria

Corresponding HOXA9 aa	Substitution	$\Delta\Delta G_{affinity}$	$\Delta\Delta G_{stability}$
ALA196	**CIR** 	-45.128 kcal/mol	-0.816 kcal/mol
	ALC 	-18.704 kcal/mol	-4.095 kcal/mol
	MTR 	-17.088 kcal/mol	-3.208 kcal/mol
	CTE 	-15.008 kcal/mol	-5.778 kcal/mol
ALA197	**BIF** 	-12.892 kcal/mol	-3.615 kcal/mol
	TBP 	-8.133v	-3.556 kcal/mol

	HRG 	-6.681 kcal/mol	-9.341 kcal/mol
	CIR 	-5.075 kcal/mol	-2.606 kcal/mol
ASN198	**MOT** 	-7.505 kcal/mol	-13.303 kcal/mol
	0BN 	-6-051 kcal/mol	-10.151 kcal/mol
	KYN 	-6.041 kcal/mol	-5.511 kcal/mol
	GBU 	-5.867 kcal/mol	-0.688 kcal/mol
LEU200	**PBF** 	-51.368 kcal/mol	-3.302 kcal/mol
	CP3 	-11.929 kcal/mol	-1.106 kcal/mol

	QU4 	-11.415 kcal/mol	-4.372 kcal/mol
	ANT 	-11.134 kcal/mol	-1.353 kcal/mol
HIS201	**ILX** 	-11.562 kcal/mol	-1.371 kcal/mol
	HIL 	-10.330 kcal/mol	-2.198 kcal/mol
	DPP 	-4.195 kcal/mol	-1.029 kcal/mol
	HRG 	-4.018 kcal/mol	-5.686 kcal/mol

6.2.4 MD simulations of point mutated HOXA9 peptides and MM-GBSA calculations

Each of these point mutated HOXA9 peptide-PBX1-DNA complexes were further processed to run MD simulation of 200 ns per each to explore the binding stability. The RMSD plots were observed for all the complexes but they are not reported in this manuscript due to spatial needs. These systems showed good stability and low oscillations of the RMSD plots. Therefore, MM-GBSA were computed during the whole trajectories of each protein-protein complex in order to compare the $\Delta G_{binding}$ average values to the $\Delta G_{binding}$ of the wild-type system, *i.e.* HOXA9 hexapeptide ($\Delta G_{binding\text{-}HOXA9}$ = -58.1922 kcal/mol), and HTL001 core peptide ($\Delta G_{binding\text{-}HTL001}$ = -53.6882 kcal/mol). All those

mutations presenting $\Delta G_{binding}$ average values lower than the above-mentioned ones were further processed for the next steps of the work. In table 6.5 MM-GBSA values are listed for each mutated peptide.

Table 6.5. Data results from MM-GBSA calculations of MD trajectories performed on point mutated HOXA9 peptides in complex with PBX1 and DNA

	X_1 = ALA196 MUTATION		X_2 = ALA197 MUTATION		X_3 = ASN198 MUTATION	
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**CIR**	-68.1064 kcal/mol 9.19 -91.4424 to -33.6024 kcal/mol	**BIF**	-52.5114 kcal/mol 8.23 -77.3444 to -25.9087 kcal/mol	**MOT**	-51.3925 kcal/mol 9.77 -79.7187 to -21.3466 kcal/mol
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**MTR**	-58.3419 kcal/mol 7.37 -85.6739 to -35.5824 kcal/mol	**TBP**	-59.0603 kcal/mol 9.34 -80.7314 to -31.1314 kcal/mol	**0BN**	-59.1051 kcal/mol 8.71 -82.4310 to -27.3523 kcal/mol
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**ALC**	-59.6952 kcal/mol 7.39 -78.3738 to -26.9511 kcal/mol	**HRG**	-54.3707 kcal/mol 9.34 -85.1600 to -24.8071 kcal/mol	**KYN**	-56.4406 kcal/mol 7.35 -75.4170 to -28.6690 kcal/mol
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**CTE**	-56.3011 kcal/mol 7.55 -80.4156 to -29.7472 kcal/mol	**CIR**	-57.6592 kcal/mol 8.88 -81.7311 to -26.2628 kcal/mol	**GBU**	-55.1339 kcal/mol 7.97 -77.5916 to -30.5304 kcal/mol
	X_4 = LEU200 MUTATION		**X_5 = HIS201 MUTATION**			
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**PBF**	-68.1857 kcal/mol 8.44 -95.1687 to -39.7321 kcal/mol	**ILX**	-48.9087 kcal/mol 8.81 -74.3783 to -20.2857 kcal/mol		
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**CP3**	-64.6802 kcal/mol 9.49 -89.3192 to -34.0239 kcal/mol	**HIL**	-50.3148 kcal/mol 9.03 -79.2197 to -22.2265 kcal/mol		
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**QU4**	-61.8016 kcal/mol 10.53 -88.4144 to -33.9895 kcal/mol	**DPP**	-55.6169 kcal/mol 11.48 -91.1993 to -22.9813 kcal/mol		
$\Delta G_{binding}$ average **$\Delta G_{binding}$ Std. Dev.** **$\Delta G_{binding}$ range**	**ANT**	-63.3043 kcal/mol 8.50 -87.5841 to -35.9672 kcal/mol	**HRG**	-57.0861 kcal/mol 9.46 -86.8681 to -27.0434 kcal/mol		

All these peptides showed to establish frequently the crucial interactions with Trp199, although not all of them presented $\Delta G_{binding}$ average values lower than those of HOXA9 and HTL001 hexapeptides in complex with PBX1 and DNA. Indeed, for position X_1 of the designed hexapeptide the first three amino acids, CIR, MTR and ALC, showed better $\Delta G_{binding}$ average values, for position X_2 only residue TBP reported good $\Delta G_{binding}$ average value, for position X_3 only the second non-standard amino acid 0BN was considered for further analysis, for X_4 all the four amino acids showed good $\Delta G_{binding}$ average values, and finally, for X_5 none of the four non-natural amino acids was suitable to be used for the next steps.

6.2.5 HOXA9 combinatorial peptides generation and related MD simulations and MM-GBSA calculations

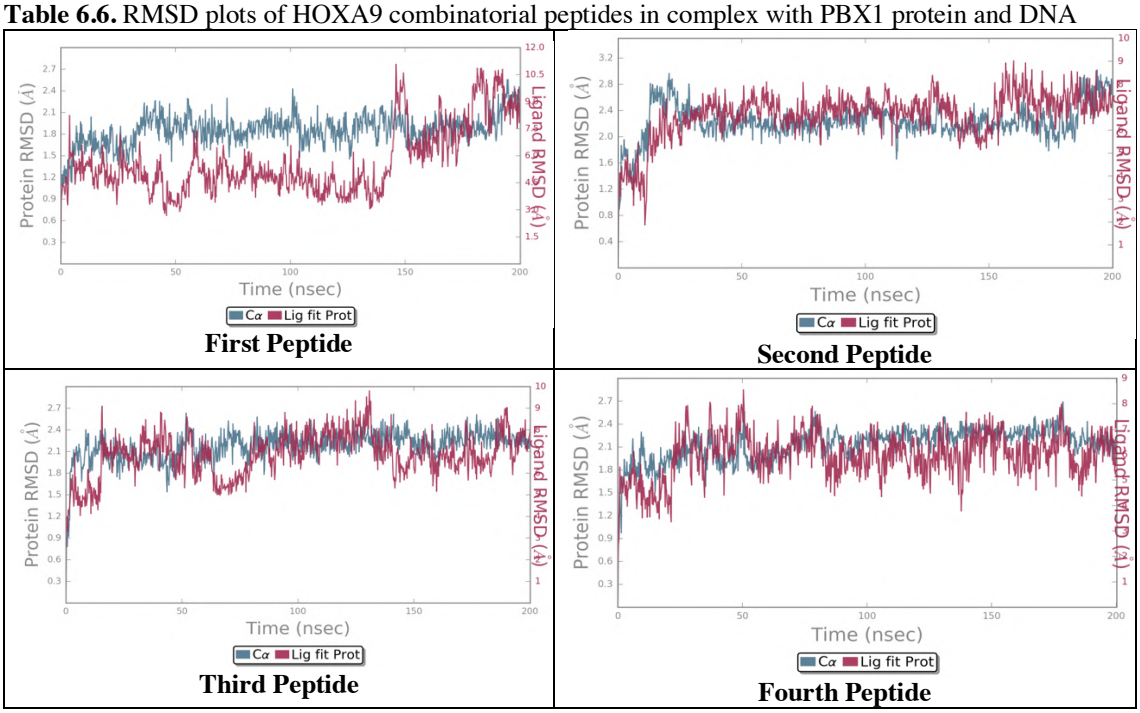
The above-described MM-GBSA calculations were used to select only those non-standard amino acids mutations that reported $\Delta G_{binding}$ average values lower compared to the reference ones ($\Delta G_{binding\text{-}HOXA9}$ average: -58.1922 kcal/mol and $\Delta G_{binding\text{-}HTL001}$ average: -53.6882 kcal/mol). Therefore, these non-natural residues were combined based

on the designed peptide motif (X_1-X_2-X_3-W-X_4-X_5) by employing HOXA9 hexapeptide scaffold, thus getting twelve combinatorial peptides, as listed below. Only tryptophan and histidine were maintained.

1. CIR – TBP – 0BN – *Trp* – PBF – *His*
2. CIR – TBP – 0BN – *Trp* - CP3 – *His*
3. CIR – TBP – 0BN – *Trp* - QU4 – *His*
4. CIR – TBP – 0BN – *Trp* – ANT – *His*
5. ALC – TBP – 0BN – *Trp* – PBF – *His*
6. ALC – TBP – 0BN – *Trp* – CP3 – *His*
7. ALC – TBP – 0BN – *Trp* – QU4 – *His*
8. ALC – TBP – 0BN – *Trp* – ANT – *His*
9. MTR – TBP – 0BN – *Trp* – ANT – *His*
10. MTR – TBP – 0BN – *Trp* – CP3 – *His*
11. MTR – TBP – 0BN – *Trp* – QU4 – *His*
12. MTR – TBP – 0BN – *Trp* – ANT – *His*

These combinatorial peptides in complex with PBX1 protein and DNA were processed to create systems in order to perform MD simulations of 200 ns per each complex. Indeed, twelve MD were run and the RMSD values were plotted for each ternary DNA-protein-peptide complex, as depicted in table 6.6, while table 6.7 shows the bar charts of protein-ligand interactions and the plots illustrating the frequency of interactions occurrences during the trajectories. The analysis of these RMSD plots confirmed the stability of the simulated systems. Furthermore, the bar charts and the interaction frequency plots highlighted that the twelve designed combinatorial peptides met most of the key interactions previously identified from PDB structure analysis and MD simulations of HOXA9 protein, HOXA9 hexapeptide and HTL001 peptide with PBX1.

Table 6.6. RMSD plots of HOXA9 combinatorial peptides in complex with PBX1 protein and DNA



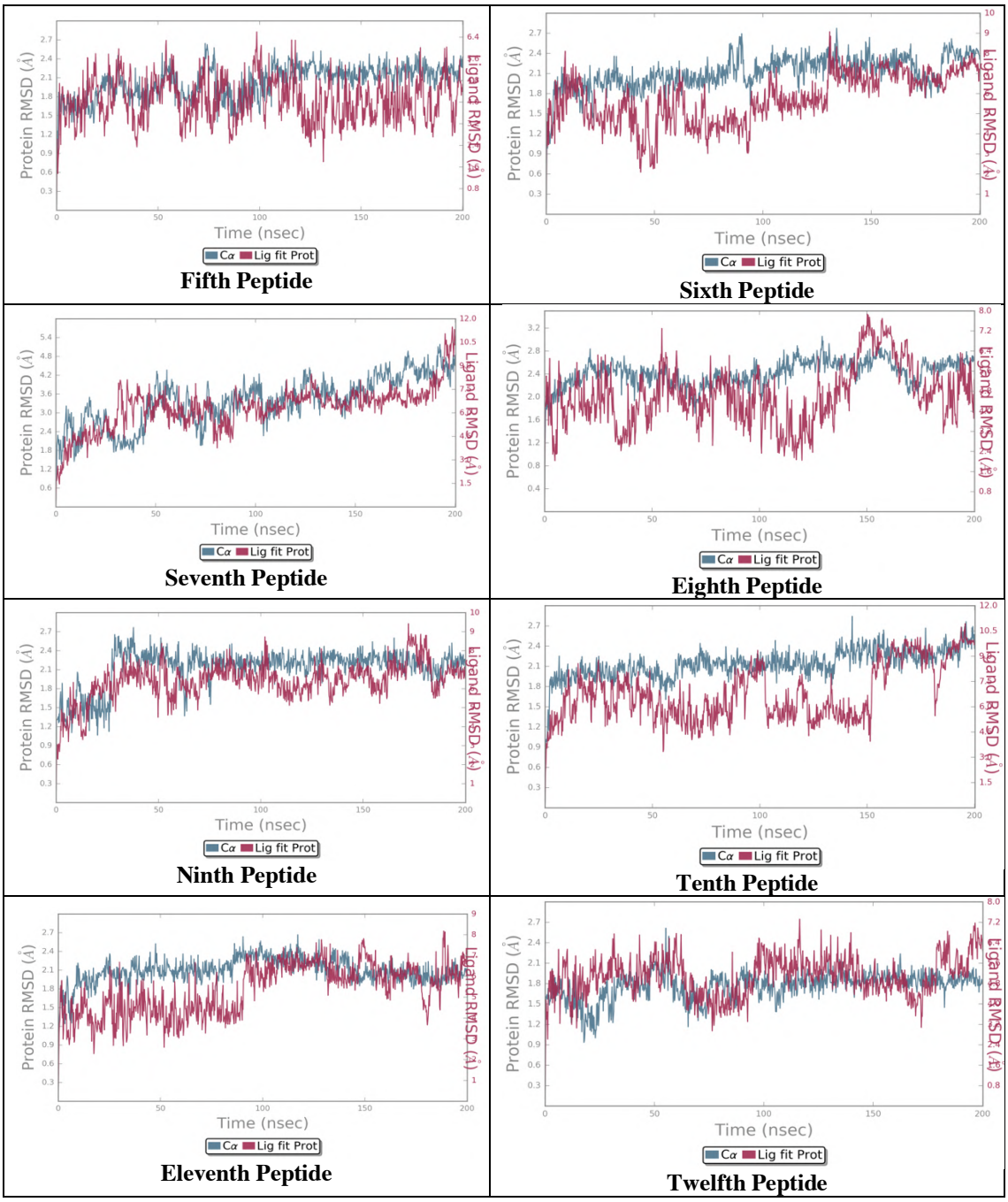
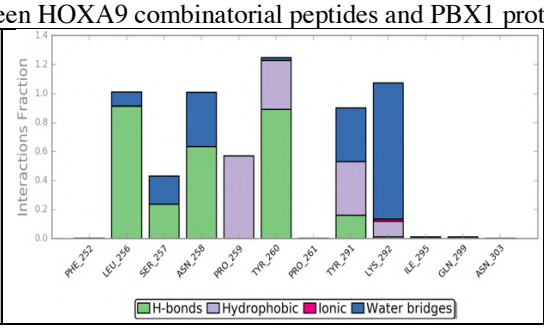
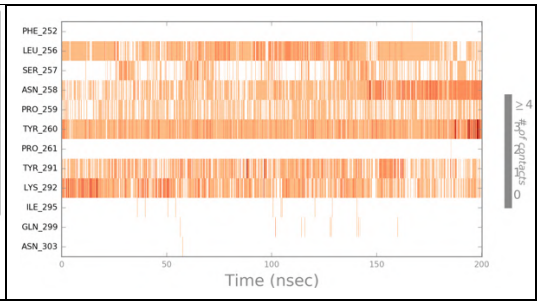
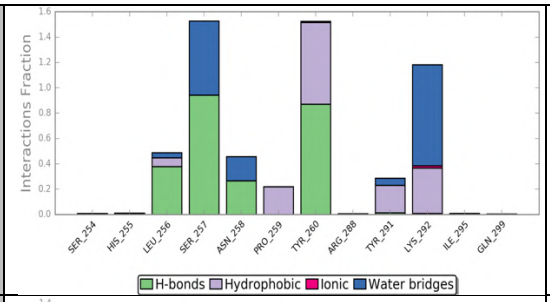
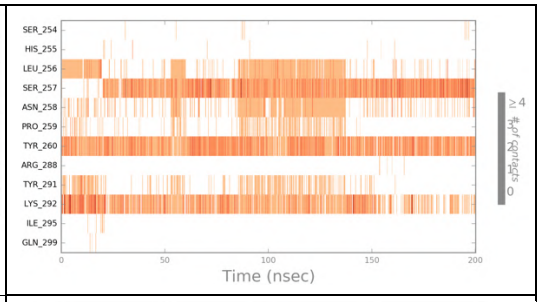


Table 6.7. On the left column, the bar charts of protein-ligand interactions for the twelve HOXA9 combinatorial peptides; on the right column, the plots illustrating the frequency of interactions occurrences between HOXA9 combinatorial peptides and PBX1 protein

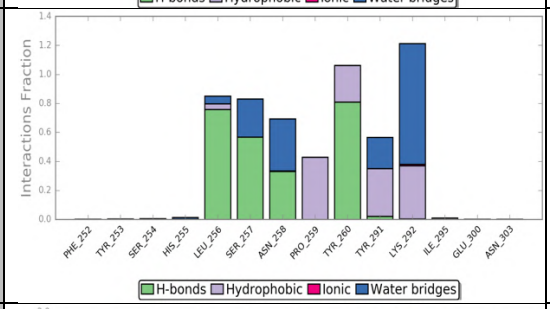
First Peptide		

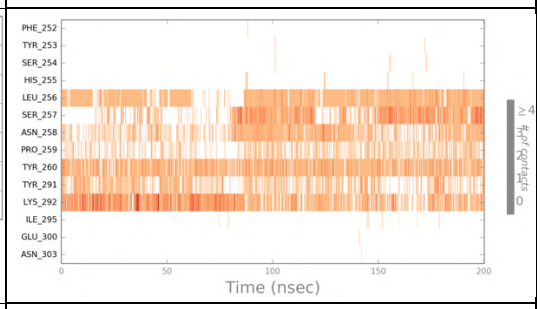
Second Peptide



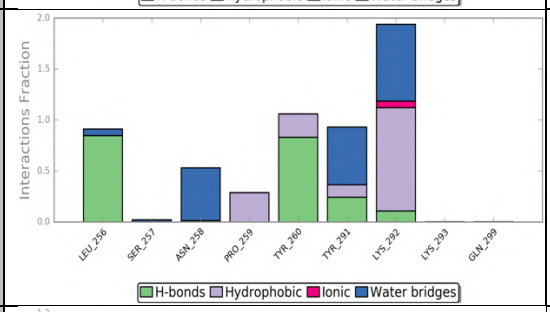


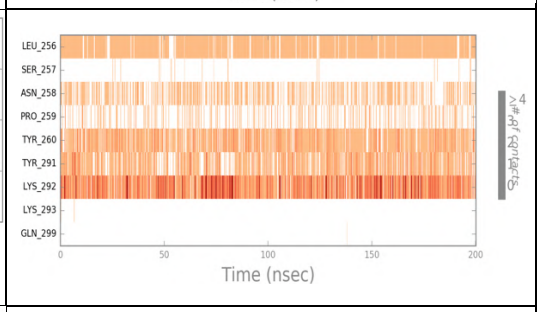
Third Peptide



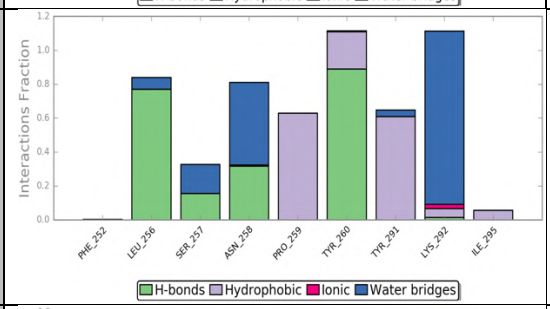


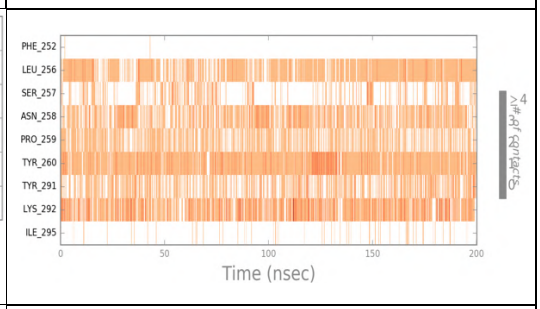
Fourth Peptide



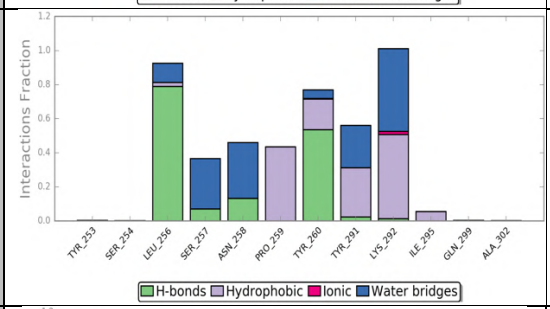


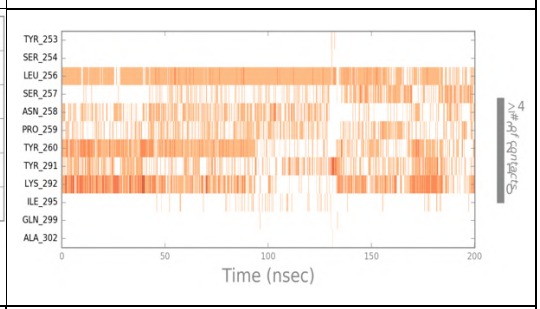
Fifth Peptide



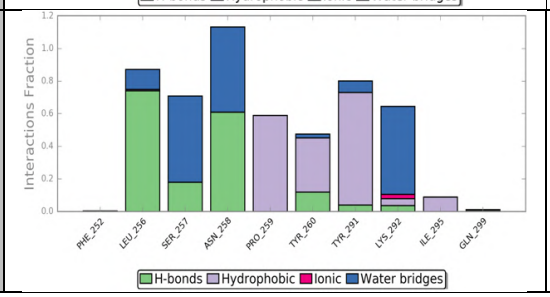


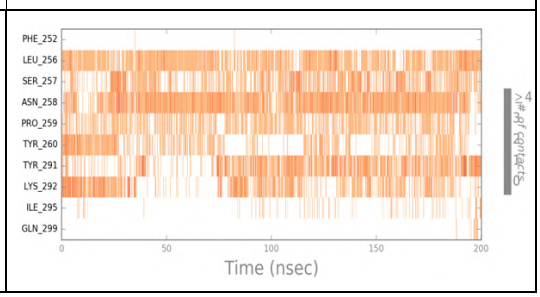
Sixth Peptide

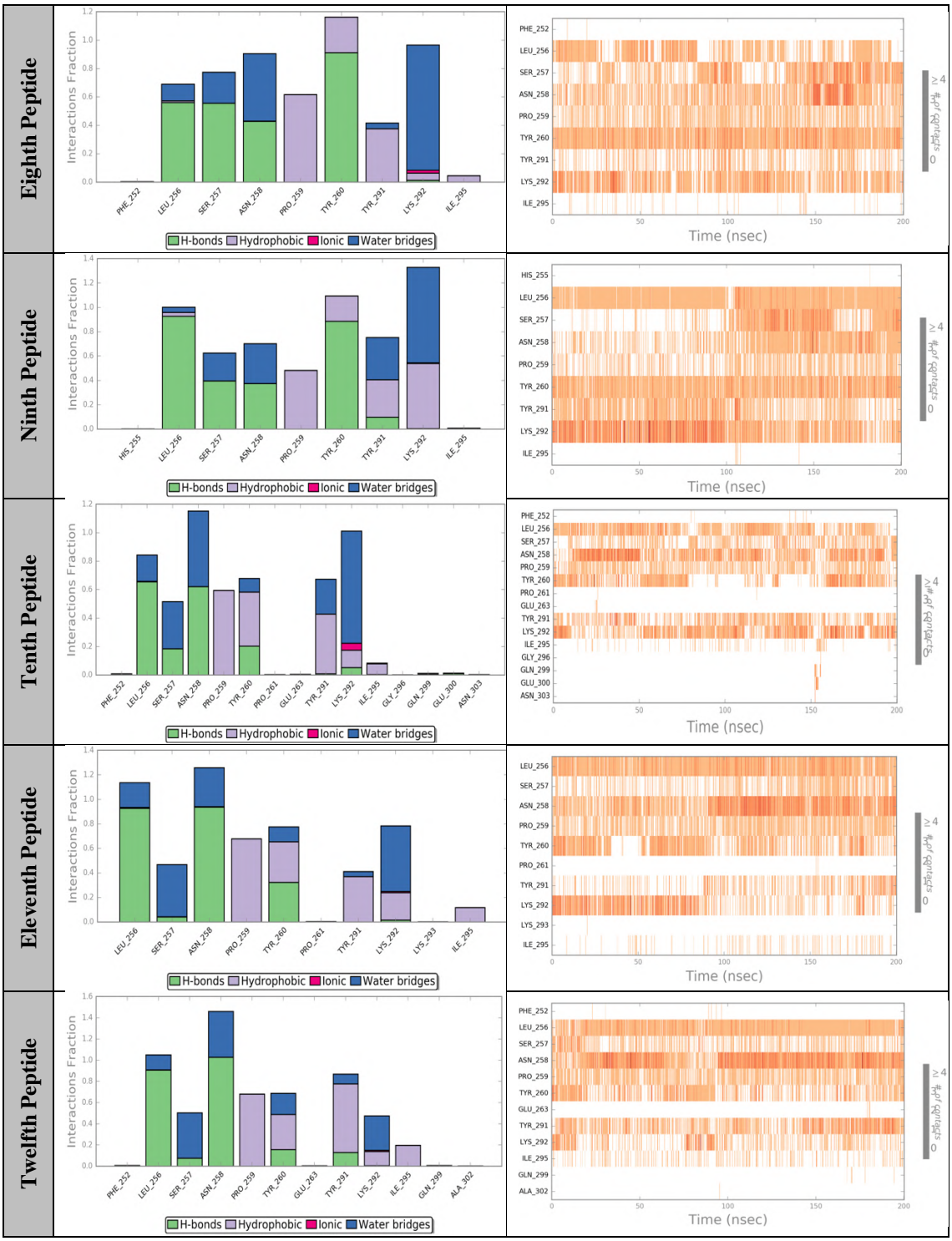




Seventh Peptide







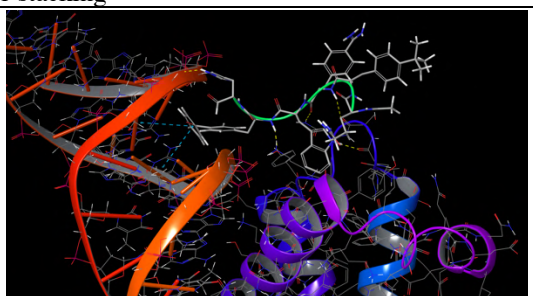
Finally, MM-GBSA calculations were performed and the results are reported in table 6.8. Even for these peptides, the resulting $\Delta G_{binding}$ average values were compared to those retrieved from MD simulations of HOXA9 and HTL001 hexapeptides in complex with PBX1 and DNA. Only the sixth peptide showed a higher $\Delta G_{binding}$ average value, hence it was not considered for further studies.

Table 6.8. Data results from MM-GBSA calculations of combinatorial peptides MD trajectories

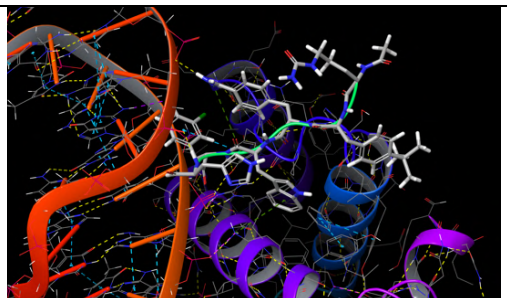
	First Peptide	Second Peptide	Third Peptide
$\Delta G_{binding}$ average $\Delta G_{binding}$ Std. Dev. $\Delta G_{binding}$ range	-79.6771 kcal/mol 10.18 -104.585 to -38.2615 kcal/mol	-61.8602kcal/mol 12.72 -99.0013 to -30.7190 kcal/mol	-68.0795 kcal/mol 10.63 -98.3690 to -29.5313 kcal/mol
	Fourth Peptide	**Fifth Peptide**	**Sixth Peptide**
$\Delta G_{binding}$ average $\Delta G_{binding}$ Std. Dev. $\Delta G_{binding}$ range	-64.6664 kcal/mol 7.53 -87.5689 to -30.1013 kcal/mol	-81.8766 kcal/mol 7.44 -101.5164 to -45.3623 kcal/mol	-55.1927 kcal/mol 10.09 -85.5158 to -22.5652 kcal/mol
	Seventh Peptide	**Eighth Peptide**	**Ninth Peptide**
$\Delta G_{binding}$ average $\Delta G_{binding}$ Std. Dev. $\Delta G_{binding}$ range	-62.8885 kcal/mol 9.19 -89.1247 to -19.4438 kcal/mol	-71.9163 kcal/mol 9.19 -101.6790 to -44.4808 kcal/mol	-74.0909 kcal/mol 11.42 -105.5444 to -32.2303 kcal/mol
	Tenth Peptide	**Eleventh Peptide**	**Twelfth Peptide**
$\Delta G_{binding}$ average $\Delta G_{binding}$ Std. Dev. $\Delta G_{binding}$ range	-60.2167 kcal/mol 9.56 -89.0633 to -28.6783 kcal/mol	-65.0198 kcal/mol 8.24 -89.3709 to -36.2812 kcal/mol	-68.3222 kcal/mol 8.13 -95.3406 to -42.1593 kcal/mol

An interesting aspect is that these peptides showed a specific binding behaviour during the trajectories, whereas the tryptophan amino acid was kept stuck within the hydrophobic pocket of PBX1 during all the twelve MD simulations. Moreover, substitutions in position X_1 of the designed peptides included non-standard amino acids with hydrophobic side chains, as required from the consensus HOX hexapeptide. These amino acids in position X_1 showed also to stack their side chain into the DNA minor groove by establishing pi-stacking contacts with the nitrogenous bases. This fact suggested that the contacts between combinatorial peptides and DNA should contribute to stabilise the complex PBX1-peptide-DNA. Table 6.9 depicts the combinatorial peptides binding modes with PBX1 protein and DNA.

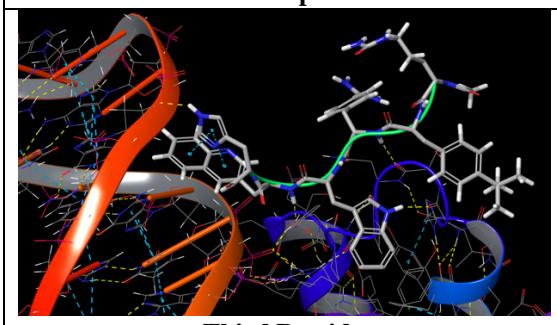
Table 6.9. Binding mode of the twelve combinatorial peptides in complex with PBX1 protein and DNA during MD simulations of 200 ns. Yellow-dotted lines stand for hydrogen bonds and blue-dotted lines are pi-stacking



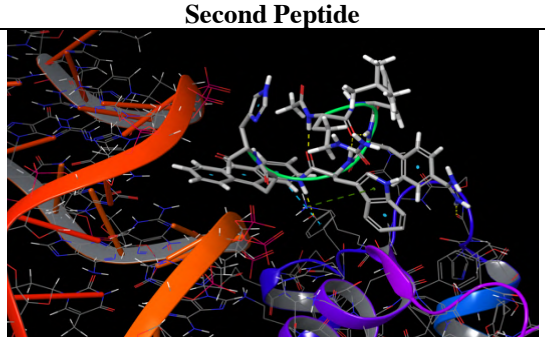
First Peptide



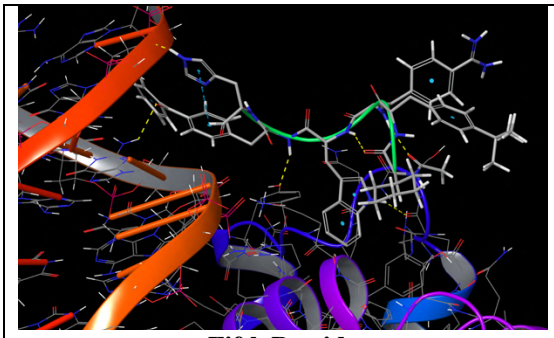
Second Peptide



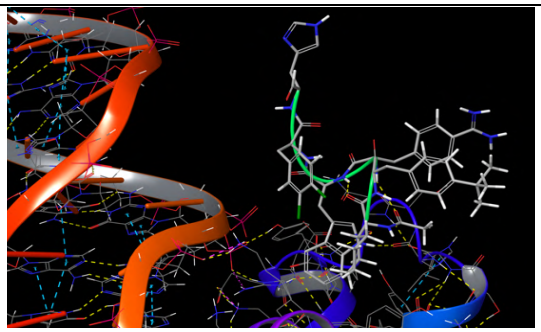
Third Peptide



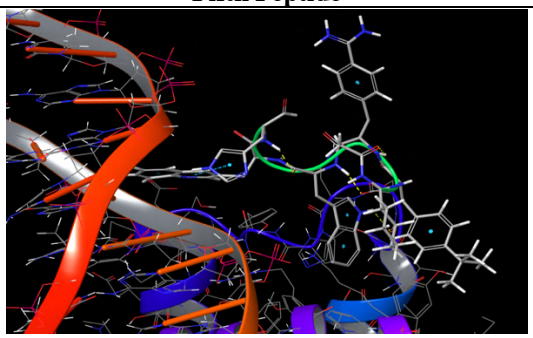
Fourth Peptide



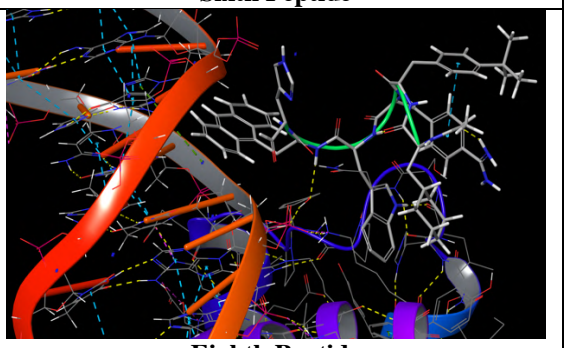
Fifth Peptide



Sixth Peptide

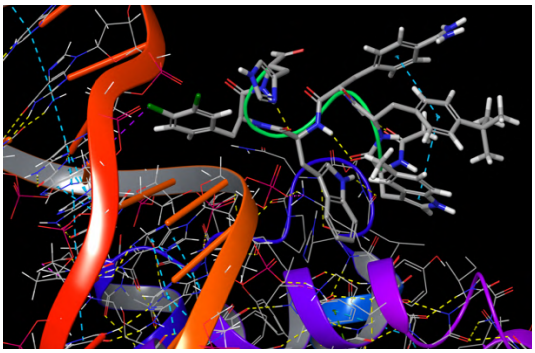


Seventh Peptide

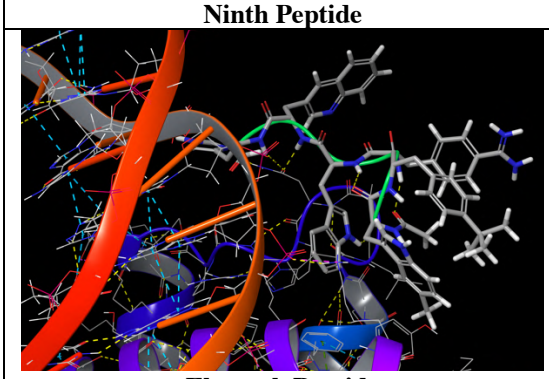


Eighth Peptide

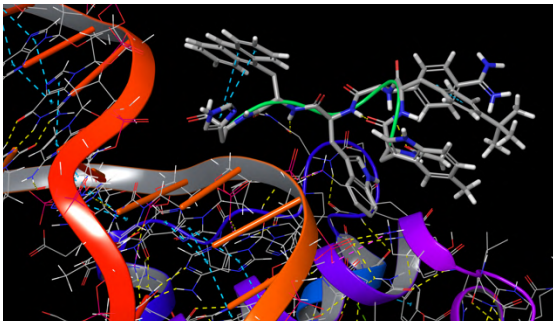
Ninth Peptide



Tenth Peptide



Eleventh Peptide



Twelfth Peptide

Finally, it was considered that both proteins, HOX and PBX, are involved into transcriptional events, therefore physicochemical properties of these combinatorial peptides, such as PSA and $\log P_{o/w}$, were computed and table 6.10 reports the related values about the twelve designed peptides, to check their ability to permeate cells.

Table 6.10. PSA and $logP_{o/w}$ values of the twelve combinatorial peptides

	PEPTIDES	PSA	$logP_{o/w}$
1	CIR-TBP-0BN-TRP-PBF-HIS	302.15	3.2
2	CIR-TBP-0BN-TRP-CP3-HIS	310.38	2.6
3	CIR-TBP-0BN-TRP-QU4-HIS	322.80	2.0
4	CIR-TBP-0BN-TRP-ANT-HIS	307.47	3.3
5	ALC-TBP-0BN-TRP-PBF-HIS	300.04	6.1
6	ALC-TBP-0BN-TRP-CP3-HIS	284.52	6.0
7	ALC-TBP-0BN-TRP-QU4-HIS	297.08	5.4
8	ALC-TBP-0BN-TRP-ANT-HIS	250.48	5.9
9	MTR-TBP-0BN-TRP-PBF-HIS	276.41	4.3
10	MTR-TBP-0BN-TRP-CP3-HIS	268.43	5.4
11	MTR-TBP-0BN-TRP-QU4-HIS	280.80	4.8
12	MTR-TBP-0BN-TRP-ANT-HIS	260.87	6.0

6.3 Methods

6.3.1 Preparation of HOXA9-PBX1-DNA complex

The 3D trimeric complex of HOXA9-PBX1-DNA was downloaded from the Protein Data Bank [60] (PDB ID: 1PUF) and imported in Schrödinger suite to optimise the structure by using "Protein preparation" tool [62]. The bond orders for untemplated residues were assigned by using known HET groups based on their SMILES strings in Chemical Component Dictionary. Hydrogens were added to the structure, eventual bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds between two sulfurs, if they were close to each other, were created and water molecules beyond 5.0 Å from any of the HET groups, including ions, were deleted. Then, protonation and metal charge states for the ligands, cofactors and metals were generated [63, 64]. Finally, PROPKA [64] was run under pH 7.0 to optimise hydroxyl, Asn, Gln and His states using ProtAssign.

6.3.2 HOXA9 hexapeptide residues scanning using non-standard "SwissSidechain" amino acids

The "SwissSidechain" database of non-natural amino acids was downloaded from the Swiss Institute of Bioinformatics website [57], imported into the "Residue Scanning" tool of Schrödinger suite and, together with 20 non-standard amino acids already present in this tool, was used to perform point mutations on HOXA9 residues (Ala196, Ala197, Asn198, Trp199, Leu200 and His201) by using PDB 1PUF. The stability and affinity were computed for each of these mutations and the resulting structures were refined by selecting side-chain prediction with backbone minimization.

6.3.3 MD simulations of PBX1-DNA in complex with HOXA9 protein, HOXA9 hexapeptide, HTL001 core peptide, point mutated peptides and combinatorial peptides

In this work overall thirty-five Molecular Dynamics simulations of 200 ns per each were performed. All of them were run by applying the following settings. The systems were created using the "System builder" tool of Schrödinger suite. TIP3P [65] was selected as a solvent model and the orthorhombic shape box was chosen. The selected box size calculation method was buffer and the box side distances were set 10 Å. The force field OPLS3 [66] was applied and the system was neutralized by adding Na^+ ions. The outputs were further processed by performing MD simulations of 200 ns using Desmond [65]. The number of atoms, the pressure and the temperature were maintained constant for the entire trajectories. Pressure and temperature were set 1.01325 bar and 300 K, respectively. Finally, the systems were relaxed before starting simulations.

6.3.4 MM-GBSA calculations of the complexes used to perform MD simulations

The MD outputs of HOXA9 hexapeptide, HTL001 core peptide, point mutated HOXA9 peptides and combinatorial peptides in complex with PBX1 protein and DNA were used to compute MM-GBSA calculations through the command line. For this purpose, the Python script "thermal_mmgbsa.py" was run.

Overall thirty-four MM-GBSA calculations were performed and data are reported in the "Results and discussion" section.

6.4 Conclusions

This work was conducted at the School of Pharmacy and Pharmaceutical Sciences of Cardiff University under the supervision of Professor Andrea Brancale. The above-described steps performed by applying computational tools led to the identification of eleven peptides whose design was based on experimental data [5, 19, 23, 29, 36, 37] and computational results. The related MD simulations of these peptides in complex with PBX1 protein and DNA reported promising results analysing the binding mode of these combinatorial peptides, the predicted $\Delta G_{binding}$ average values and the physicochemical properties. These peptides will be synthesised at Cardiff University laboratories and assayed to investigate their potential efficacy about preventing HOX-PBX1 cooperative binding in cancer cells to tackle several kinds of malignancies associated in some cases with decrease or increase of HOX gene activity.

Finally, this work has been considering for a research article that soon will be sent to a scientific peer-reviewed journal.

References – Chapter Six

1. Pearson JC, Lemons D, McGinnis W (2005) Modulating Hox gene functions during animal body patterning. Nat Rev Genet 6:893–904. https://doi.org/10.1038/nrg1726
2. Krumlauf R (1994) Hox genes in vertebrate development. Cell 78:191–201. https://doi.org/10.1016/0092-8674(94)90290-9
3. Duboule D (2007) The rise and fall of Hox gene clusters. Development 134:2549–2560. https://doi.org/10.1242/dev.001065
4. Mallo M, Alonso CR (2013) The regulation of Hox gene expression during animal development. Development 140:3951–3963. https://doi.org/10.1242/dev.068346
5. Gehring WJ, Qian YQ, Billeter M, et al (1994) Homeodomain-DNA recognition. Cell 78:211–223. https://doi.org/10.1016/0092-8674(94)90292-5
6. McGinnis N, Kuziora MA, McGinnis W (1990) Human Hox-4.2 and Drosophila Deformed encode similar regulatory specificities in Drosophila embryos and larvae. Cell 63:969–976. https://doi.org/10.1016/0092-8674(90)90500-E
7. Holland PW, Booth HAF, Bruford EA (2007) Classification and nomenclature of all human homeobox genes. BMC Biol 5:47. https://doi.org/10.1186/1741-7007-5-47
8. Morgan R (2006) Hox genes: a continuation of embryonic patterning? Trends Genet 22:67–69. https://doi.org/10.1016/j.tig.2005.11.004
9. Lebert-Ghali C-E, Fournier M, Dickson GJ, et al (2010) HoxA cluster is haploinsufficient for activity of hematopoietic stem and progenitor cells. Exp Hematol 38:1074-1086.e5. https://doi.org/10.1016/j.exphem.2010.07.006
10. Alharbi RA, Pettengell R, Pandha HS, Morgan R (2013) The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia 27:1000–1008. https://doi.org/10.1038/leu.2012.356
11. Xu B, Geerts D, Bu Z, et al (2014) Regulation of endometrial receptivity by the highly expressed HOXA9, HOXA11 and HOXD10 HOX-class homeobox genes. Hum Reprod 29:781–790. https://doi.org/10.1093/humrep/deu004
12. Quinonez SC, Innis JW (2014) Human HOX gene disorders. Mol Genet Metab 111:4–15. https://doi.org/10.1016/j.ymgme.2013.10.012
13. Hombría JC-G, Lovegrove B (2003) Beyond homeosis—HOX function in morphogenesis and organogenesis. Differentiation 71:461–476. https://doi.org/10.1046/j.1432-0436.2003.7108004.x
14. Sánchez-Herrero E (2013) Hox Targets and Cellular Functions. Scientifica (Cairo) 2013:1–26. https://doi.org/10.1155/2013/738257
15. Mann RS, Chan S-K (1996) Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet 12:258–262. https://doi.org/10.1016/0168-9525(96)10026-3
16. Hayashi S, Scott MP (1990) What determines the specificity of action of Drosophila homeodomain proteins? Cell 63:883–894. https://doi.org/10.1016/0092-8674(90)90492-W
17. Berger MF, Badis G, Gehrke AR, et al (2008) Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell 133:1266–1276. https://doi.org/10.1016/j.cell.2008.05.024
18. Noyes MB, Christensen RG, Wakabayashi A, et al (2008) Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. Cell 133:1277–1289. https://doi.org/10.1016/j.cell.2008.05.023
19. Knoepfler PS, Kamps MP (1995) The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. Mol Cell Biol 15:5811–5819. https://doi.org/10.1128/MCB.15.10.5811
20. Moens CB, Selleri L (2006) Hox cofactors in vertebrate development. Dev Biol 291:193–206. https://doi.org/10.1016/j.ydbio.2005.10.032
21. Longobardi E, Penkov D, Mateos D, et al (2014) Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev Dyn 243:59–75. https://doi.org/10.1002/dvdy.24016
22. Peifer M, Wieschaus E (1990) Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. Genes Dev 4:1209–1223. https://doi.org/10.1101/gad.4.7.1209
23. Piper DE, Batchelor AH, Chang C-P, et al (1999) Structure of a HoxB1–Pbx1 Heterodimer Bound to DNA. Cell 96:587–597. https://doi.org/10.1016/S0092-8674(00)80662-5
24. Brendolan A (2005) A Pbx1-dependent genetic and transcriptional network regulates spleen

ontogeny. Development 132:3113–3126. https://doi.org/10.1242/dev.01884
25. Allen TD, Zhlp Y-X, Hawley TS, Hawley RG (2000) TALE Homeoproteins as HOX11-Interacting Partners in T-cell Leukemia. Leuk Lymphoma 39:241–256. https://doi.org/10.3109/10428190009065824
26. Chan S-K, Jaffe L, Capovilla M, et al (1994) The DNA binding specificity of ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell 78:603–615. https://doi.org/10.1016/0092-8674(94)90525-8
27. van Dijk MA, Murre C (1994) extradenticle Raises the DNA binding specificity of homeotic selector gene products. Cell 78:617–624. https://doi.org/10.1016/0092-8674(94)90526-6
28. Phelan ML, Rambaldi I, Featherstone MS (1995) Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Mol Cell Biol 15:3989–3997. https://doi.org/10.1128/MCB.15.8.3989
29. LaRonde-LeBlanc NA (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev 17:2060–2072. https://doi.org/10.1101/gad.1103303
30. Abu-Shaar M, Ryoo HD, Mann RS (1999) Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. Genes Dev 13:935–945. https://doi.org/10.1101/gad.13.8.935
31. Stevens KE, Mann RS (2007) A Balance Between Two Nuclear Localization Sequences and a Nuclear Export Sequence Governs Extradenticle Subcellular Localization. Genetics 175:1625–1636. https://doi.org/10.1534/genetics.106.066449
32. Saleh M, Huang H, Green NC, Featherstone MS (2000) A Conformational Change in PBX1A Is Necessary for Its Nuclear Localization. Exp Cell Res 260:105–115. https://doi.org/10.1006/excr.2000.5010
33. Merabet S, Hudry B, Saadaoui M, Graba Y (2009) Classification of sequence signatures: a guide to Hox protein function. BioEssays 31:500–511. https://doi.org/10.1002/bies.200800229
34. Johnson FB, Parker E, Krasnow MA (1995) Extradenticle protein is a selective cofactor for the Drosophila homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. Proc Natl Acad Sci 92:739–743. https://doi.org/10.1073/pnas.92.3.739
35. Joshi R, Passner JM, Rohs R, et al (2007) Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. Cell 131:530–543. https://doi.org/10.1016/j.cell.2007.09.024
36. Passner JM, Ryoo HD, Shen L, et al (1999) Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. Nature 397:714–719. https://doi.org/10.1038/17833
37. Shen W-F, Chang C-P, Rozenfeld S, et al (1996) Hox Homeodomain Proteins Exhibit Selective Complex Stabilities with Pbx and DNA. Nucleic Acids Res 24:898–906. https://doi.org/10.1093/nar/24.5.898
38. Joshi R, Sun L, Mann R (2010) Dissecting the functional specificities of two Hox proteins. Genes Dev 24:1533–1545. https://doi.org/10.1101/gad.1936910
39. Knoepfler PS, Lu Q, Kamps MP (1996) Pbx1-Hox Heterodimers Bind DNA on Inseparable Half-Sites That Permit Intrinsic DNA Binding Specificity of the Hox Partner at Nucleotides 3' to a TAAT Motif. Nucleic Acids Res 24:2288–2294. https://doi.org/10.1093/nar/24.12.2288
40. Lu Q, Kamps MP (1997) Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specificities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain – demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo. Oncogene 14:75–83. https://doi.org/10.1038/sj.onc.1200799
41. Saleh M, Rambaldi I, Yang X-J, Featherstone MS (2000) Cell Signaling Switches HOX-PBX Complexes from Repressors to Activators of Transcription Mediated by Histone Deacetylases and Histone Acetyltransferases. Mol Cell Biol 20:8623–8633. https://doi.org/10.1128/MCB.20.22.8623-8633.2000
42. Chang CP, Brocchieri L, Shen WF, et al (1996) Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol Cell Biol 16:1734–1745. https://doi.org/10.1128/MCB.16.4.1734
43. Lu Q, Kamps MP (1996) Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex. Mol Cell Biol 16:1632–1640. https://doi.org/10.1128/MCB.16.4.1632
44. Peltenburg LT, Murre C (1997) Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. Development 124:1089–98

45. Chang CP, Shen WF, Rozenfeld S, et al (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes Dev 9:663–674. https://doi.org/10.1101/gad.9.6.663
46. Ji T, Lee M, Pruitt SC, Hangauer DG (2004) Privileged scaffolds for blocking protein–protein interactions: 1,4-disubstituted naphthalene antagonists of transcription factor complex HOX–PBX/DNA. Bioorg Med Chem Lett 14:3875–3879. https://doi.org/10.1016/j.bmcl.2004.05.068
47. Morgan R, Sohal J (2007) Field of the Invention The present invention relates to molecules which impair PBX-dependent. 2007:
48. Morgan R, Pirard PM, Shears L, et al (2007) Antagonism of HOX/PBX Dimer Formation Blocks the In vivo Proliferation of Melanoma. Cancer Res 67:5806–5813. https://doi.org/10.1158/0008-5472.CAN-06-4231
49. Plowright L, Harrington KJ, Pandha HS, Morgan R (2009) HOX transcription factors are potential therapeutic targets in non-small-cell lung cancer (targeting HOX genes in lung cancer). Br J Cancer 100:470–475. https://doi.org/10.1038/sj.bjc.6604857
50. Morgan R, Boxall A, Harrington KJ, et al (2012) Targeting the HOX/PBX dimer in breast cancer. Breast Cancer Res Treat 136:389–398. https://doi.org/10.1007/s10549-012-2259-2
51. Morgan R, Boxall A, Harrington KJ, et al (2014) Targeting HOX transcription factors in prostate cancer. BMC Urol 14:17. https://doi.org/10.1186/1471-2490-14-17
52. Morgan R, Simpson G, Gray S, et al (2016) HOX transcription factors are potential targets and markers in malignant mesothelioma. BMC Cancer 16:85. https://doi.org/10.1186/s12885-016-2106-7
53. Errico MC, Felicetti F, Bottero L, et al (2013) The abrogation of the HOXB7/PBX2 complex induces apoptosis in melanoma through the miR-221&222-c-FOS pathway. Int J Cancer 133:879–892. https://doi.org/10.1002/ijc.28097
54. Ando H, Natsume A, Senga T, et al (2014) Peptide-based inhibition of the HOXA9/PBX interaction retards the growth of human meningioma. Cancer Chemother Pharmacol 73:53–60. https://doi.org/10.1007/s00280-013-2316-5
55. Primon M, Hoffman E, Pandha HS, Morgan R (2016) HTL001, a novel inhibitor of HOX/PBX binding, is highly cytotoxic to prostate and breast cancer cells. Eur J Cancer 69:S133. https://doi.org/10.1016/S0959-8049(16)32994-X
56. Shears L, Plowright L, Harrington K, et al (2008) Disrupting the Interaction Between HOX and PBX Causes Necrotic and Apoptotic Cell Death in the Renal Cancer Lines CaKi-2 and 769-P. J Urol 180:2196–2201. https://doi.org/10.1016/j.juro.2008.07.018
57. Database of non-standard amino acids. https://www.swisssidechain.ch/. Accessed 27 Jul 2020
58. Gfeller D, Michielin O, Zoete V (2012) Expanding molecular modeling and design tools to non-natural sidechains. J Comput Chem 33:1525–1535. https://doi.org/10.1002/jcc.22982
59. (2006) SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Association for Computing Machinery, New York, NY, USA
60. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
61. Beard H, Cholleti A, Pearlman D, et al (2013) Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. PLoS One 8:e82849. https://doi.org/10.1371/journal.pone.0082849
62. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221–234. https://doi.org/10.1007/s10822-013-9644-8
63. Greenwood JR, Calkins D, Sullivan AP, Shelley JC (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comput Aided Mol Des 24:591–604. https://doi.org/10.1007/s10822-010-9349-1
64. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
65. Mark P, Nilsson L (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. J Phys Chem A 105:9954–9960. https://doi.org/10.1021/jp003020w
66. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864

CHAPTER SEVEN

Ras-RasGRF1 INTERACTION – Insights for the optimisation of a patented α-helix-shaped peptide

7.1 Introduction

Drug addiction is a chronic disease affecting the brain and associated with high relapse rates and compulsive drug use. Among several psychotropic drugs, cocaine is a powerfully addictive stimulant drug derived from the leaves of the coca plant native in South America. Although cocaine can be used for medical purposes, such as local anaesthesia for some surgeries, it causes strongly negative consequences due to large use. Indeed, in 2020 it has been registered the incidence of 16.8% of cocaine abuse in 26-aged people and over in the US [1]. Exposure to abuse substances such as cocaine provokes intense and long-lasting memories of well-being, that are crucial in the transition from recreational drug-taking to compulsive and uncontrolled drug use [2, 3]. It has been demonstrated that drug relapse following drug abstinence depends on learned associations between drug-paired cues and the rewarding effects of these drugs, that persistently elicit drug-seeking behaviours [4, 5].

Increasing evidence demonstrates that cocaine alters dopamine (DA) levels of neurotransmission in brain circuits related to the control of movement and reward to exert their molecular and behavioural effects [6–8].

Normally, dopamine recycles back into the dopaminergic neurons, interrupting the signal between neurons. In this context, cocaine prevents dopamine from being recycled, causing cocaine heap formation within the space between two nerve cells, blocking their normal communication. These large amounts of dopamine generate a vicious cycle in the brain's reward circuit, by strongly reinforcing drug-taking behaviours. These latter are caused by an adaptation mechanism to the excess of dopamine generated by cocaine, that makes nerve cells less sensitive to dopamine. As a result, people take stronger and more frequent doses in an attempt to feel the same effect and to obtain rewarding feeling.

Cocaine short-term effects include extreme happiness and energy, mental alertness, hypersensitivity to light, sound, and touch, irritability and paranoia. On the other hand, cocaine long-term effects depend on the method of use: by snorting it appears loss of smell, nosebleeds, frequent runny nose, and problems with swallowing; by smoking cocaine causes cough, asthma, respiratory distress, and a higher risk of infections like pneumonia; by consuming by mouth it generates severe bowel decay from reduced blood flow; finally by needle injection there is a higher risk for contracting HIV, hepatitis C, and other bloodborne diseases, skin or soft tissue infections, as well as scarring or collapsed veins [9].

In the brain, addictive drugs exploit cellular mechanisms and signalling pathways involved in normal learning and memory processes contributing to high rates of relapse [10–13]. In the last two decades, Ras-ERK signalling pathway has been identified as involved in both the acute and long-term effects of cocaine by performing different

experiments, that mimic drug addiction in humans, resulting in the alteration of ERK-mediated signalling in specific brain regions. Although the molecular mechanism on basis of the effects of most abused substances on ERK signalling and the drug-mediated behavioural changes are still unclear, they span across locomotor activity/sensitization, drug self-administration, and conditioned place preference (CPP). The CPP refers to a learning procedure in which a biologically potent stimulus is paired with a previously neutral stimulus used to measure the motivational effects of objects or experiences [14, 15, 24–27, 16–23].

Intracellular ERK (extracellular signal-regulated kinases) signalling has been characterised to respond to extracellular signals and regulate cell proliferation, differentiation, survival, and death [28–30]. For example, once activated by growth factors or neurotrophins, the tyrosine kinase receptors recruit Ras (*rat sarcoma* protein) family G-proteins by sequentially triggering the activation of Raf (*rat fibrosarcoma* protein), MEK (mitogen-activated protein kinase) and ERK (figure 7.1). Thus, the phosphorylated activated ERK form can translocate to the nucleus [31], and phosphorylate a ternary complex factor (Elk-1, ETS like-1 protein) [32, 33]. This complex in turn can associate with serum response factor and foster the transcription of the immediate early gene (IEG) related to neuroadaptation [34–36].

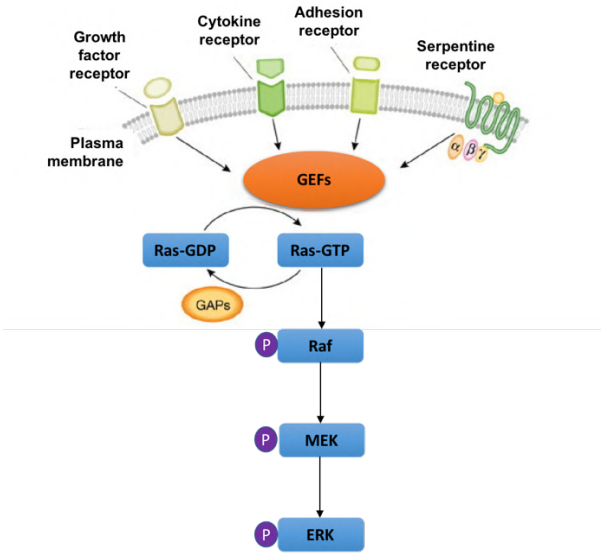


Figure 7.1. Ras-ERK activation cascade [37]

To date, there is no effective treatment for drug addiction. Therefore, understanding the neurobiological aspects behind substance abuse effects can provide crucial insights for developing potential therapeutic strategies tackling drug addiction. Indeed, the designed pharmacological inhibitors of MEK, the kinase upstream of ERK, played a crucial role in investigating lasting experience- and drug-dependent alterations in behavioural plasticity associated with ERK cascade.

The only MEK inhibitor able to penetrate the blood-brain barrier was SL327, but it was not administered to humans, owing to the toxicity and the relatively low IC_{50} (0.18-0.22 μM toward MEK1 and MEK2, respectively) [11, 14, 43–49, 15–17, 38–42]. It has been demonstrated that MEK inhibition prevents conditioned place preference (CPP) to cocaine and amphetamine [11, 14, 16, 50]. In addition, other studies highlighted Ras-

ERK cascade dependence in drug-seeking during exposure to cocaine-associated cues/contexts following the acquisition of self-administration [15, 49, 51–56].

In this context, Ras proteins act as binary switches in signalling pathways by cycling between inactive GDP- and active GTP-bound states [57]. Kinetic studies highlighted that the activation of Ras protein, proceeding from the conversion of Ras-GDP to Ras-GTP, initiates through the recruitment of the guanine nucleotide exchange factors (GEFs), such as Son of sevenless protein (Sos) and Ras guanine nucleotide-releasing factor 1 (RasGRF1) [58–63], that catalyse GDP release and allow its replacement by GTP [64–69]. Then, the GTP molecule binds to this complex promoting the release of GEF protein [70].

The region of Sos and RasGRF1 proteins, that is required for Ras-specific nucleotide exchange activity, exhibits a Ras exchanger motif (Rem) domain of about 450 amino acids and a Cdc25 homology domain [71–75]. In addition, Sos requires allosteric activation through a second Ras-binding site that bridges the Rem and Cdc25 domains [76, 77]. When Sos is activated, the helical hairpin of its Cdc25 domain inserts between two flexible regions of Ras, Switch 1 and Switch 2, causing Ras conversion to the transient state by opening the nucleotide-binding site of Ras for GDP release [72] (figure 7.2). After this event, Ras can promptly accommodate and bind GTP into the nucleotide-binding site, thus exhibiting its activate state. Therefore, a potential strategy to inhibit Ras-GEF interaction should be targeting the open – or transient – state of Ras protein by designing peptides able to bind the nucleotide-exchange region.

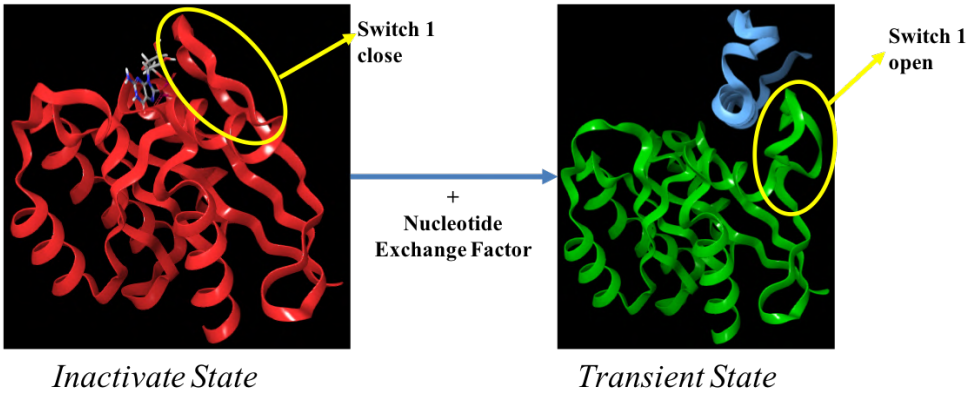


Figure 7.2. Inactive and transient state of Ras protein. On the left, Ras protein (red chain; PDB ID: 1XD2) bound to GDP (grey ligand in stick format) exhibits its inactivate state, where Switch 1 region (highlighted by a yellow circle) is close; on the right, after binding to a guanine nucleotide-exchange factor (Sos helical hairpin fragment in this picture, blue chain) Ras (green chain; PDB ID: 1XD2) experiences a transient state, where the Switch 1 region is open to accommodate the GEF α-helix

In 2006, Freedman *et al*. [78] performed nucleotide-exchange assays to monitor the release rate of fluorescently labelled GDP from Ras in presence and absence of nucleotide exchange factor [62, 79]. The authors built a construct of RasGRF1 that spans residues 1028 to 1262, *i.e*. RasGRF1 Cdc25 domain, by using computational techniques to predict the secondary structure and aligning the RasGRF1 sequence to Sos [70, 72, 80, 81]. The rate of GDP release from Ras in the presence of RasGRF1 Cdc25 domain was $50 \pm 10 \times 10^{-4}$ s^{-1} for 1 μM exchange factor [78] and was significantly higher than the intrinsic rate of nucleotide release by isolated Ras, *i.e*. $1.8 \pm 0.2 \times 10^{-4}$ s^{-1}.

In contrast to Sos, which requires Ras binding to the allosteric site for activity, the Cdc25 domain of RasGRF1 is active on its own [62, 70, 76, 77]. The structure of the Cdc25 domain of RasGRF1 is very similar to that of Sos, registering 30% of sequence identity between the two Cdc25 domains. The orientation and conformation of the RasGRF1 helical hairpin resemble that of Sos in its active form with an RMSD value of 2.3 Å for the helical hairpins after superposition on the Cdc25 domain core. Moreover, distance difference matrices demonstrated that the main differences between RasGRF1 and Sos in its inactive form have been identified into the helical hairpin, even in this case confirming that RasGRF1 Cdc25 domain is more similar to active Sos [78].

7.1.1 Mutational studies on Sos

Due to the lack of a Ras-RasGRF1 complex structure, it was useful starting from the analysis of Ras-Sos complex data available in literature.

Performed mutagenesis studies shed light on three regions of Ras crucial for the activation of the protein, *i.e.* the switch 1 region (amino acids 25–40), the switch 2 region (amino acids 57–75), and a short region spanning amino acids 100–110 [82–91].

In 1998, Boriack-Sjodin *et al.* [72] determined the crystal structure of human Ras in complex with Sos Cdc25 domain, highlighting more than 30 amino acids involved in interactions. These contacts with Sos are mainly mediated by the switch 1 and switch 2 regions of Ras [72] and are essentially hydrophobic, polar and charge-charge bonds. Hall *et al.* [92] performed site-directed mutagenesis to deeply investigate these contacts. The results shed light on the hydrophobic pocket consisting of residues Ile825, Leu872 and Phe929 of Sos protein that embed the side chain of Tyr64 of Ras into a hydrophobic contact. Indeed, the contribution of Tyr64 of Ras was explored by applying a mutation to alanine (Y64A). The result was a reduction of at least 50-fold in the apparent binding affinities of Ras for Sos, but did not provide significant nucleotide dissociation. Then, the authors performed another binding assay by using wild-type Ras and mutated Phe929 of Sos to alanine (F929A). The Sos mutant reported a decrease of more than 50-fold in binding affinity for Ras. This data indicated that Tyr64 and Phe929 mediate hydrophobic contacts crucial for the formation of a stable Ras-Sos complex.

On the other hand, polar and charged interactions showed to be not so essential for the binding affinity of Ras for Sos. Indeed, alanine mutations on Sos residues Glu1002, Thr935 and Arg826 exerted a low impact on Ras binding and activation. Furthermore, the mutation of Ala59 of Ras to glycine (A59G) did not significantly affect the GDP-dissociation rate, displaying more than 50% of the inhibitory effect on Sos-catalysed GTP dissociation.

Finally, the contribution of two Ras amino acids involved into the region switch 1 were investigated: Tyr32 of Ras that established hydrophobic contacts with Lys939 of Sos, and Tyr40 of Ras that mediated stacking interaction with His911 of Sos.

Tyr32 and Tyr40 of Ras were mutated to Ser (Y32S) and Ala (Y40A), respectively. Both mutations decreased the binding of Sos to Ras and accelerated the rate of intrinsic GDP/ GTP exchange, suggesting that these residues are important for Ras-Sos recognition and the nucleotide stabilization. Consistent with these results, mutations of Sos Lys939 and His911 to alanine (K939A and H911A, respectively) also caused a reduction Ras-Sos binding. Furthermore, the Y40A mutation had no significant effect on

Sos-catalysed guanine nucleotide exchange, whereas the disruption of the contact between Tyr32 of Ras and Lys939 of Sos reduced the sensitivity of Ras to the exchange activity of Sos [72].

7.1.2 Targeting Ras-GEFs interaction by using hydrogen-bond surrogates

Several efforts have been reported in the literature to design and identify Ras inhibitors to block the nucleotide exchange. However, to date, the scientific knowledge on ERK signalling in drug addiction has not been translated into clinical treatments due to the lack of drugs with relatively low IC_{50} values, toxicity and ability to efficiently cross the blood-brain barrier [93].

One of the first compounds identified to inhibit Ras activity was the Sulindac together with its derivatives [94, 95]. Sulindac is an NSAID that demonstrated to directly inhibit Raf activation mediated by Ras protein *in vitro* [96]. However, relatively high concentrations of Sulindac and its derivatives (10-50 μM) were necessary to achieve the desired biological effects, thus indicating a lack of potency.

Another identified compound was MCP1 [97], that together with its derivatives was able to inhibit Ras-Raf interaction and Ras activation of Raf [98–100]. However, as for Sulindac and derivatives, these compounds also lacked potency and need further modifications to overcome this issue.

Several computational approaches on Ras protein highlighted novel potential ligand-binding pockets that allowed to identify lead Ras inhibitory compounds for further development [101, 102]. An example of these lead compounds is Kobe0065 identified through an *in silico* screening approach performed by Shima *et al*. [103]. The compound Kobe0065 and its derivative Kobe2602 demonstrated to bind the inactive state of Ras and inhibit Ras-Raf interaction with a rough IC_{50} value of 10 μM and the Sos-catalysed nucleotide exchange.

In 2016, Papale and colleagues at Cardiff University designed and generated an active cell-penetrating peptide, the peptide RB3, based on the interaction between Ras and a GEF protein, *i.e*. RasGRF1, able to attenuate the activation of Ras-ERK signalling cascade *in vivo* [104].

The cell-penetrating peptides have been shown to be promising for the treatment of neuropsychiatric disorders, especially for the low reported toxicity and tolerability [105, 106]. Although their biological activity spans a micromolar range, they usually show a potential advantage, due they are able to partially disrupt protein-protein interactions without preventing the enzymatic activity. Furthermore, cell-penetrating peptides can dissolve rapidly in water solvents, unlike several molecules with high molecular weights that need organic solvents.

The peptide RB3 was designed by using molecular graphics tools, on the basis of the ternary complex consisting of Ras in its transient state bound to Sos Cdc25 domain and Ras in its inactive state complexed with a GDP molecule (PDB ID: 1XD2) [77]. The Cdc25 domain of Sos involved in this ternary complex was compared to the crystal structure of RasGRF1 Cdc25 domain (PDB ID: 2IJE). The peptide sequence (portion 1173–1203 of the CDC25 domain) includes an α-helix – from Met1181 to Glu1191 – crucial for the GDP exchange activity on Ras proteins, linked to two loops – from

Pro1173 to Gly1180 and from Gly1192 to Asn1203. Below the sequence of peptide RB3 is shown and Figure 7.3 depicts the 3D structure of the following peptide RB3 sequence.

GRKKRRQRRR – PPCVPYLGMYLTDLVFIEEGTPNYTEDGLVN

TAT sequence ***RasGRF1 interacting region***

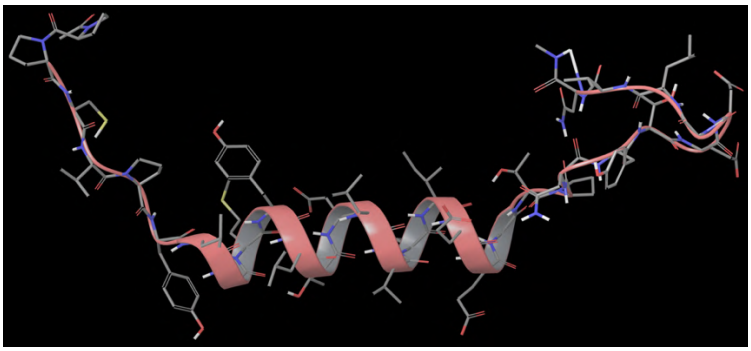


Figure 7.3. Peptide RB3 structure including an α-helix – from Met1181 to Glu1191 – linked to a loop – from Pro1173 to Gly1180 – and another loop – from Gly1192 to Asn1203.

Moreover, the research group of Cardiff University added to the peptide RB3 sequence a portion of the HIV TAT protein known to exhibit a translocating behaviour [107]. In this way, the final structure of the cell-penetrating peptide was created able to cross the cell membranes and the blood-brain barrier [108].
Then, the peptide RB3 was tested in an ex-vivo model of acute striatal brain slices to investigate its inhibitory potential on ERK phosphorylation after stimulation with 100 μM glutamate. The result was a significant reduction of ERK activity with an IC_{50} of 6 μM.

To deeply explore the effect of peptide RB3 on Ras-ERK signalling pathway, Papale and colleagues investigated whether RB3 may also affect the phosphorylation of two well-characterised ERK substrates, (Ser10)-acetylated (Lys14) histone H3 (pAc-H3) and S6 ribosomal protein, (pS6, Ser235/236 specific site) [109–111]. Even in this case, the peptide was effective in decreasing the phosphorylation of Ac-H3 with an IC_{50} of 5.2 μM and pS6 levels with an IC_{50} of 3.69 μM [104].

In the light of the above, this peptide was patented [112] and it was selected to enter a further compound optimisation process, to get stronger biological activity. In the next steps, the adopted approaches are described.

First, MD simulations were performed on Ras-Sos, Ras-RasGRF1 and Ras-peptide RB3 complexes to identify the most stable and frequent interactions between protein partners. In this context, the peptide RB3 exhibited a misfolding behaviour, whereas the helical hairpin corresponding to RasGRF1 interacting region lost helicity propensity generating instability within the system. For this purpose, the peptide HBS3, a compound reported in the literature [93, 113], was used as a reference to collect more data addressing the modifications to peptide RB3 structure.

The peptide HBS3 is a synthetic α-helix, that was designed to target the Sos-interacting region of Ras. Indeed, its structure was essentially based on Sos sequence able to bind Ras protein. Interestingly, this peptide showed a preference for nucleotide-free Ras with K_D = 28 μM compared to Ras-GDP reporting K_D = 158 μM. Moreover, the

peptide HBS3 reduced the nucleotide exchange *in vitro* and modestly ERK activation in cells [93, 113].

The peptide HBS3 structure depicted in figure 7.4 has been designed to mimic the α-helix conformation of Sos-interacting region by applying a hydrogen bond surrogate (HBS) approach. This approach has been developed especially for modulating biomolecular interactions, such as protein-protein contacts, through small molecular weight protein secondary structure mimetics, when designing small molecules could be a very challenging strategy [114–120].

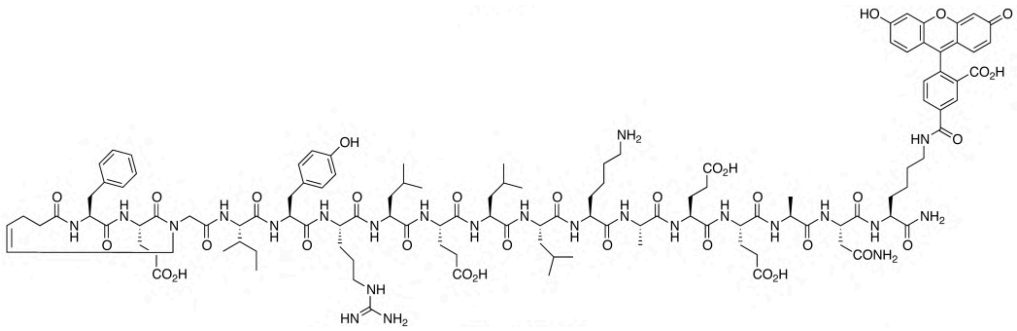


Figure 7.4. Structure of peptide HBS3 [114]

The HBS approach is based on the helix-coil transition theory for peptides, whereas α-helices composed of a few amino acids are expected to be essentially unstable due to a low nucleation probability [121, 122]. This approach is expected to overwhelm the intrinsic nucleation propensities of the amino acids by providing upstream a preorganization of the residues, that causes the helix formation initialization [123, 124]. Indeed, in a general α-helix, a hydrogen bond between the carbonyl group of the *i*th amino acid residue as acceptor and the amine group of the *(i + 4)*th amino acid residue as donor stabilises and nucleates the helical structure. Based on this evidence, the HBS strategy for artificial α-helices generation involves the replacement of one of the main chain hydrogen bonds with a covalent linkage [114, 125]. Indeed, to mimic the C=O · · · H-N hydrogen bond as closely as possible, a covalent bond of the type C=X–Y–N is included, where X and Y are usually carbon atoms, that would be part of the *i*th and the (*i* + *4*)th residues.

In detail, the analysis of peptide RB3 α-helix highlighted that the first amino acid implicated into the helix H-bonds ensemble does not establish a traditional hydrogen bond with the (*i*+*4*)th amino acid, while it forms a contact with the (*i*+*3*)th amino acid, by creating the so-called 3_{10}-helix illustrated in figure 7.5 [126].

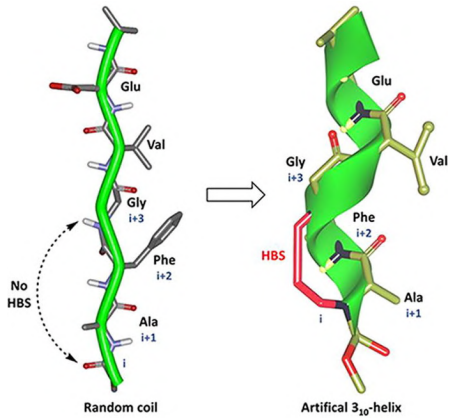
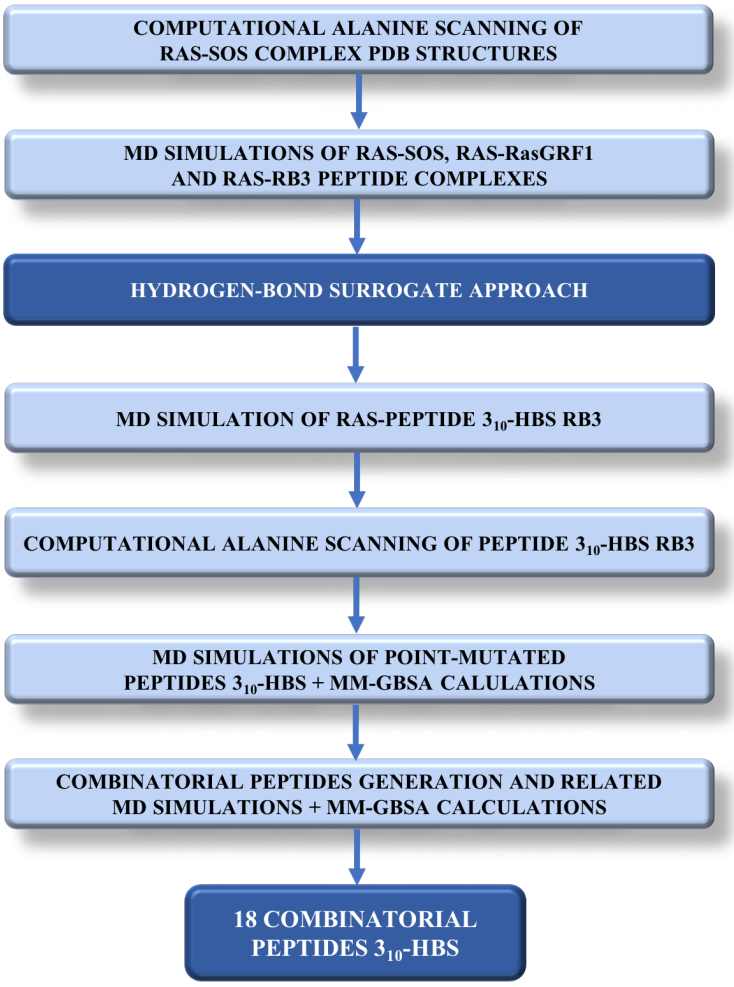


Figure 7.5. Example of 3_{10}-helix formation between *i*th amino acid and (*i+3*)th amino acid [126]

A 3_{10}-helix is a type of secondary structure observed in proteins and polypeptides. It is more tightly wound than the traditional α-helix with hydrogen bonds within the backbone formed between residues *i*th and (*i+3*)th. The term 3_{10}-helix is owed to its structure, consisting of three residues per turn and 10 atoms between hydrogen bond donor and acceptor. The 3_{10}-helix is usually observed at the termini of α-helices and play important roles as nucleation sites for helix formation during protein folding [126, 127].

Therefore, peptide RB3 was modified by creating a C-C bond between the first amino acid (Tyr1178) and the fourth residue (Met1181), hereafter peptide 3_{10}-HBS RB3, and the related performed MD simulation highlighted a stable peptide helical conformation during the whole trajectory.

Then, computational alanine scanning was run to identify and analyse the most promising mutations to be considered. These mutations were further combined into the peptide 3_{10}-HBS RB3 structure to get several combinatorial peptides. Thus, these latter were investigated to retrieve the calculated ΔG values for each couple Ras-combinatorial peptide. Finally, eighteen combinatorial peptides were selected and they will be synthesised and tested at Cardiff University laboratories. Scheme 7.1 lists the steps of the workflow in detail described in the next sections.



Scheme 7.1. Overview of the computational workflow performed to identify 18 combinatorial peptides 3_{10}-HBS potentially inhibiting Ras-GEF interaction

7.2 Results and discussion

7.2.1 Computational alanine scanning of Sos binding interface and RasGRF1 and Sos binding regions alignment

In order to deeply explore and identify the hot spot residues of Sos Cdc25 domain (*i.e.* from position 924 to 957), all the six available PDB structures of Ras-Sos complex (PDB IDs: 1XD2, 1BKD, 1NVW, 1NVV, 1NVU and 1NVX) were examined by using Robetta computational interface alanine scanning tool [128, 129]. Robetta is a tool (released by the Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical Research of the University of California in San Francisco) that was specifically created for predicting energetically important amino acid residues in protein-protein interfaces. In table 7.1 the highest ΔΔG of the alanine mutations on Sos binding interface amino acids are reported.

Table 7.1. Predicted ΔΔG values of the alanine mutations on Sos binding interface amino acids retrieved by Robetta Computational Interface Alanine Scanning

	ΔΔG (kcal/mol)					
Sos aa	**1XD2**	**1BKD**	**1NVW**	**1NVV**	**1NVU**	**1NVX**
Phe929	1.54	1.48	1.64	3.00	1.53	1.47
Thr935	2.97	1.59	1.11	n.a.	3.13	3.11
Lys939	n.a.	n.a.	n.a.	4.21	n.a.	n.a.
Glu942	n.a.	n.a.	1.10	n.a.	n.a.	n.a.
Asn944	2.51	2.63	2.35	2.66	2.63	2.70

n.a. = not available

These predicted data were in accordance with mutational studies performed by Hall *et al*. [92], whereas Phe929, Thr935 and Lys939 were highlighted as Sos interacting hot spots. In detail, from the computational alanine scanning Phe929 and Thr935 were shared by most of the six PDB complexes as hot spots, while Lys939 resulted from the PDB 1NVV analysis. As it can be seen, at the same time, this computational tool pointed out another Sos hot spot not previously identified by Hall *et al*., Asn944, that was shared from all the six PDB complexes. Another hot spot on Glu942 was retrieved from PDB 1NVW.

These identified hotspots both from biological assays [92] and computational alanine scanning were considered equally important for the following steps and were used for comparison to RasGRF1 amino acids in order to investigate similarities between the two GEFs sequences (RasGRF1 and Sos). Therefore, for this purpose PDB 1XD2 including Sos in complex with Ras protein was chosen for the low resolution, while the only available PDB structure of RasGRF1 Cdc25 domain (PDB ID: 2IJE) was used. However, the latter PDB was from *mus musculus* as organism. Hence, before proceeding with the protein structure alignment between Sos and RasGRF1 Cdc25 domains, a FASTA alignment was performed considering human and murine RasGRF1 sequences through Protein BLAST sequence alignment tool [130, 131] (released by the National Center for Biotechnology Information in Bethesda, MD). The resulted overall identity was 83.22%, whereas within the RasGRF1 region involved into Ras binding (*i.e*. from residue 1173 to 1203 of mouse sequence) the only detected difference was between Ala1198 for human and Val1187 for mouse, as illustrated in figure 7.6. These two amino acids exhibit side chains with very similar chemical properties, thus the PDB 2IJE was considered suitable for proceeding with this study.

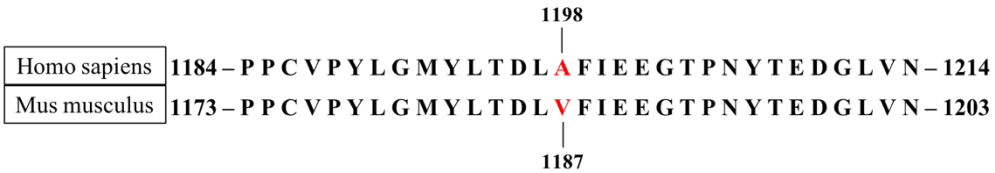


Figure 7.6. FASTA sequence alignment between RasGRF1 interacting region from two different organisms (*homo sapiens* and *mus musculus*)

Therefore, both PDB protein structures (2IJE and 1XD2) were aligned through the "Protein Structure Alignment" tool of Schrödinger suite and the result is depicted in figure 7.7. As can be seen, the two binding regions of Sos and RasGRF1 are perfectly aligned.

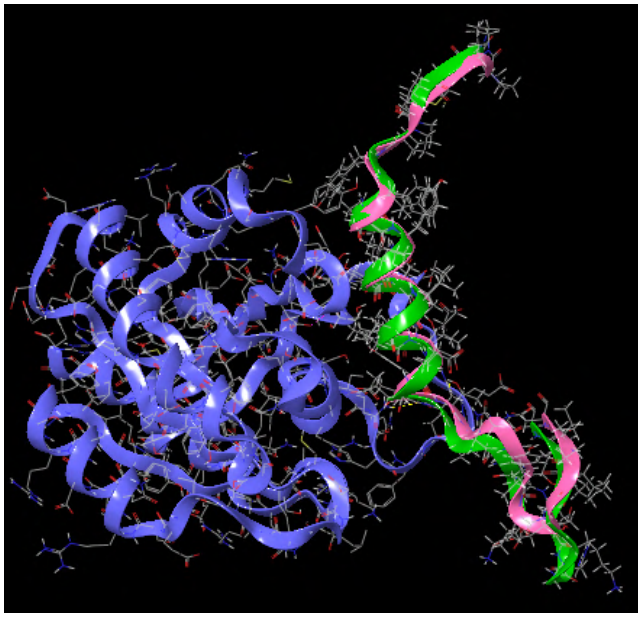


Figure 7.7. Superposition of PDB 2IJE, including RasGRF1 binding region (pink chain), and PDB 1XD2, including Sos binding region (green chain) in complex with Ras protein (purple chain)

Below a portion of this sequence alignment of Cdc25 domains involved in Ras binding is illustrated in figure 7.8.

929 935 939 942 944

Sos 924 – P P C V P F F G I Y L T N I L K T E E G N P E V L K R H G K E L I N – 957

RasGRF1 1173 – P P C V P Y L G M Y L T D L V F I E E G T P N Y T E - - - D G L V N – 1203

1178 1184 1188 1191 1193

Figure 7.8. FASTA sequence alignment between Sos and RasGRF1 regions able to bind Ras protein

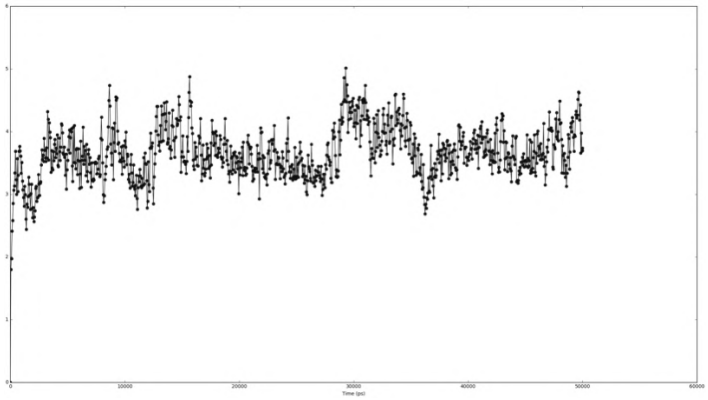
The amino acids for RasGRF1 corresponding to Sos hot spots are reported in table 7.2. As it can be seen, the pairs Thr935-Thr1184 and Glu942-Glu1191 share the same amino acid, while Phe929 and Tyr1178 both present hydrophobic side chains, and Asn944 and Thr1193 share polar uncharged side chains. Only Lys939 and Phe1188 are very different amino acids, whereas lysine shows an electrically charged side chain, while phenylalanine exhibits a hydrophobic side chain. These correspondingRasGRF1 amino acids were then considered for the next steps of the analysis.

Table 7.2. Correspondences of Sos hot spot residues identified through biological assays [92] and computational alanine scanning to RasGRF1 amino acids highlighted by performing protein structures alignment

Sos aa	Corresponding RasGRF1 aa
Phe929	Tyr1178
Thr935	Thr1184
Lys939	Phe1188
Glu942	Glu1191
Asn944	Thr1193

7.2.2 MD simulations of Ras-Sos complex

The first step of the work was a molecular dynamics simulation of Ras-Sos complex. For this purpose, PDB 1XD2 was chosen and the simulation was run for a short time of 50 ns to observe and identify the most stable interactions and investigate the importance of the computationally predicted hot spots not reported in the literature, *i.e.* Glu942 and Asn944. The stability of the system was monitored during the entire trajectory, thus registering the RMSD plot depicted in Plot 7.1 and the average energy values reported in table 7.3.



Plot 7.1. RMSD plot of MD simulation performed on Ras-Sos complex (PDB 1XD2)

Table 7.3. Energy values monitored during the MD simulation performed on Ras-Sos complex

	Average	Std Dev	Slope (ps^{-1})
Total energy (kcal/mol)	-245540.879	202.747	-0.003
Potential energy (kcal/mol)	-304042.948	157.237	-0.003

Then, the MD frames were clustered and only the frames representative of the most abundant clusters were considered for further analysis, *i.e.* frame 780 (15 frames), frame 820 (13 frames), frame 660 (12 frames) and frame 120 (11 frames).

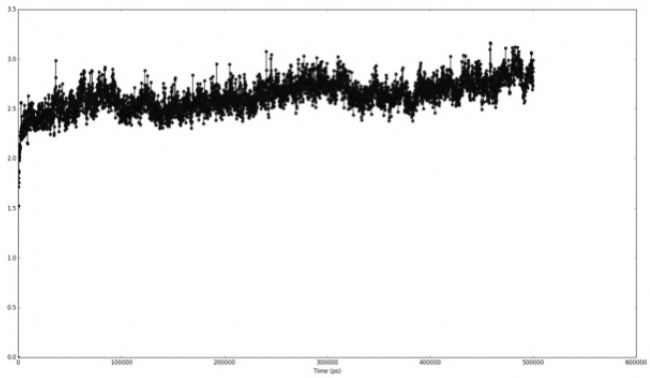
These four frames were analysed to identify the interactions between Ras and Sos proteins, that were further observed during the whole trajectory to investigate the related stability. Finally, table 7.4 shows the most stable interactions between the two proteins retrieved from MD analysis. Among the above mentioned five Sos hot spots shown in table 7.2, four of these residues established stable interactions with Ras during the simulation, *i.e.* Phe929, Thr935, Glu942 and Asn944.

Table 7.4. Stable interactions between Ras and Sos protein retrieved from MD simulation

Ras aa	Sos aa	Interaction type
Tyr64	Phe929	Hydrophobic
Gln61	Thr935	1 H-bond
Ala59	Thr935	1 H-bond
Ser17	Glu942	1 H-bond
Ala18	Glu942	1 H-bond
Tyr32	Asn944	2 H-bonds

7.2.3 MD simulations of Ras-RasGRF1 complex

As previously described, the comparison between Sos interacting region and RasGRF1 Cdc25 domain highlighted five amino acids of RasGRF1 protein as putative key amino acids to be investigated, due to their correspondence to Sos hot spots positions. Therefore, in order to explore this insight, an MD simulation of Ras-RasGRF1 complex was performed by setting 500 ns as simulation time through Schrödinger suite [132]. The protein-protein complex was generated by using the previous aligned structures between PDBs 1XD2 and 2IJE. In this way, RasGRF1 interacting region (residues 1173 to 1203) was positioned within the binding pocket of Ras protein through performing a superimposition on Sos chain that was subsequently deleted. The complex was coarsely minimised and the MD was run. The output was analysed and the stability of the system was checked through the RMSD plots (plot 7.2) and the energy, temperature, pressure and volume values, that were monitored during the entire trajectories (average energy values in table 7.5).



Plot 7.2. RMSD plot of MD simulation performed on Ras-RasGRF1 complex

Table 7.5. Energy values monitored during the MD simulation of Ras-RasGRF1 complex

	Average	Std Dev	Slope (ps^{-1})
Total energy (kcal/mol)	-104242.151	27.814	0.000
Potential energy (kcal/mol)	-129311.366	98.456	0.000

The trajectory was clustered using "Desmond trajectory clustering" tool of Schrödinger suite by setting 10 clusters to be generated. Then only the MD frames representative for the most abundant clusters were considered for further analysis, that is frame 2800 (322 frames), frame 650 (54 frames), frame 1110 (57 frames), frame 3620 (31 frames) and frame 40 (30 frames). The most stable interactions between Ras and RasGRF1 were observed and they are reported in table 7.6. As it can be observed, most of these interactions are similarly established between Ras and the corresponding Sos amino acids (refer to table 7.2) during the MD simulation. It provides interesting information to go forward with this work.

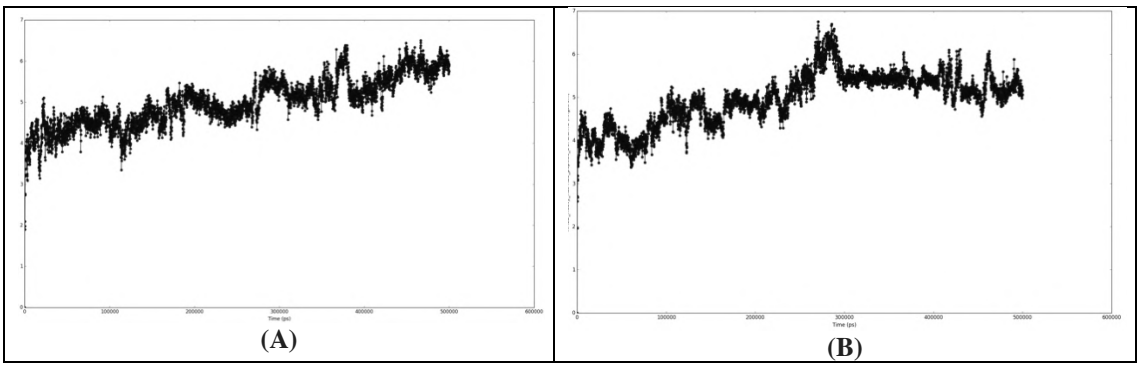
Table 7.6. The most stable interactions between Ras and RasGRF1 highlighted from MD trajectory analysis

Ras aa	RasGRF1 aa	Interaction type
Ser17	Glu1191	1 H-bond
Ala18	Glu1191	1 H-bond
Tyr32	Gly1192	1 H-bond
Ala59	Thr1184	1 H-bond
Gln61	Tyr1178	1 H-bond
Gln61	Thr1184	1 H-bond
Tyr64	Tyr1178	Pi-Pi stacking

7.2.4 MD simulations of Ras-peptide RB3 complex

After collecting information about interactions between Ras and its GEFs (Sos and RasGRF1) other MD simulations were performed to explore the binding mode and the established contacts between Ras and the patented peptide RB3.

The core sequence of this peptide (without TAT portion) refers to RasGRF1 residues from 1173 to 1203, that is PPCVPYLGMYLTDLVFIEEGTPNYTEDGLVN. Therefore, the complex Ras-RasGRF1 was used and all the residues not belonging to RB3 sequence were deleted. Thus, this new complex Ras-peptide RB3 was processed by running two MD simulations of 500 ns each. The RMSD plots (plot 7.3A-B) were generated and even energy, temperature, pressure and volume were monitored during the trajectories (average energy values in table 7.7) to check the stability of the systems.



Plot 7.3. RMSD plot of first **(A)** and second **(B)** MD simulations performed on Ras-peptide RB3 complex

Table 7.7. Energy values monitored during the MD simulations of Ras-peptide RB3complex

	First MD			Second MD		
	Average	**Std Dev**	**Slope (ps^{-1})**	**Average**	**Std Dev**	**Slope (ps^{-1})**
Total energy (kcal/mol)	-174665.918	157.936	0.000	-173436.77	105.470	0.000
Potential energy (kcal/mol)	-212782.659	119.298	0.000	-90540.589	80.692	0.000

The RMSD plots revealed a certain instability of the systems, ranging from about 2.7 to 6. On the contrary energy, temperature, pressure and volume exhibited stable values. Then, the MD frames of both simulations were grouped into ten clusters for each and the frames representative for the most abundant clusters were considered for further analysis, as follows:

- First MD → frame 390 (114 frames), frame 2340 (160 frames), frame 1400 (69 frames), frame 3590 (19 frames), frame 3720 (17 frames) and frame 4270 (117 frames);
- Second MD → frame 3610 (176 frames), frame 4800 (30 frames), frame 2660 (51 frames), frame 620 (54 frames), frame 120 (22 frames), frame 1280 (135 frames) and frame 2240 (23 frames).

The interactions were analysed and observed during both trajectories to retrieve the most frequent and stable ones and the results are reported in table 7.8.

Table 7.8. The most stable interactions between Ras and peptide RB3 highlighted from MD trajectories analysis

First MD			Second MD		
Ras aa	**RB3 aa**	**Interaction type**	**Ras aa**	**RB3 aa**	**Interaction type**
Tyr40	Asp1185	1 H-bond	Tyr40	Asp1185	1 H-bond
Tyr40	Phe1188	Pi-Pi stacking	Tyr40	Phe1188	Pi-Pi stacking
Tyr32	Gly1192	1 H-bond	Glu31	Gly1192	1 H-bond
Gln61	Tyr1182	1 H-bond	Arg149	Glu1198	1 H-bond + 1 salt bridge
			Gln25	Gly1192	1 H-bond

The two MD simulations shared only two interactions, *i.e.* one hydrogen bond between Tyr40 of Ras and Asp1185 of peptide RB3 and a pi-pi stacking between the aromatic ring of Tyr40 side chain of Ras and the other aromatic ring of Phe1188 side chain of peptide RB3. Furthermore, all the other interactions retrieved from the MD simulations were not very stable during the entire trajectories. Thus, by visually exploring both simulations an important behaviour of peptide RB3 α-helix appeared: a portion of the α-helix (from Met1181 to Thr1184) started to lose helicity propensity after about 50-60 ns of simulation resulting in a misfolding behaviour. As a result, the two above mentioned most frequent and stable interactions shared by both simulations from MD analysis were involved into the folded region of the peptide, *i.e.* the residues from Asp1185 to Glu1191 not exhibiting the misfolding. Figure 7.9 illustrates the misfolding of RB3 α-helix after 50 ns of the first MD simulation.

Figure 7.9. Frame from first MD simulation after 50 ns depicting peptide RB3 losing helicity propensity in the portion from Met1181 to Thr1184; the purple chain stands for Ras protein, while the pink chain is peptide RB3

Therefore, a strategy to overcome this issue was implemented by applying the hydrogen-bond surrogate approach, that had already provided experimental evidence of success [93, 113–120]. Thus. the peptide RB3 was processed by modifying the structure. First, two portions of the peptide were deleted, *i.e.* the residues belonging to the two loops of the peptide, owed they showed to lack crucial interactions according to literature data (*e.g.* mutational studies [72, 92]) and the MD simulations of Ras-RasGRF1, as it can be seen from table 7.6. Thus, residues 1171 to 1177 and 1194 to 1203 were considered not important and deleted from the structure. Then, the analysis of the α-helix highlighted that Tyr1178, the first amino implicated into the helix H-bonds ensemble, does not

establish a traditional hydrogen bond with the (*i+4*)th amino acid, while it forms a contact with the (*i+3*)th amino acid, by creating the so-called 3_{10}-helix [126].

Hence, an MD simulation of 500 ns was performed on Ras protein in complex with the peptide RB3 modified by deleting the two loops and creating a covalent C-C bond between the carbonyl oxygen of Tyr1178 backbone and the amine hydrogen of Met1181 backbone, hereafter termed peptide 3_{10}-HBS RB3 (figure 7.10).

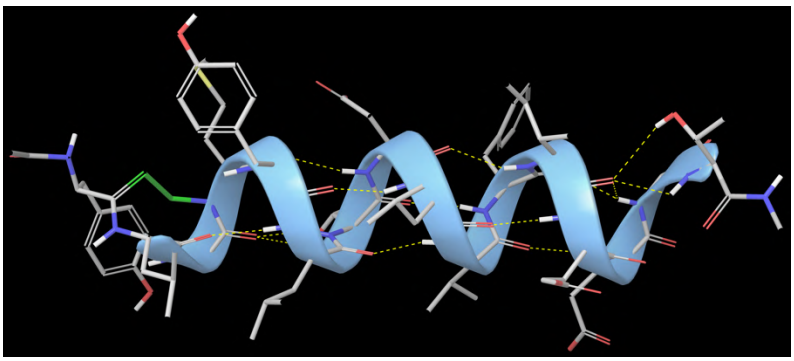
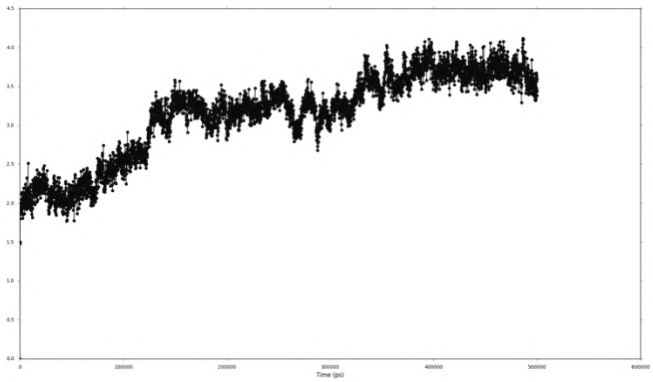


Figure 7.10. Peptide 3_{10}-HBS RB3 including a covalent C-C bond (green bond) between the carbonyl oxygen of Tyr1178 backbone and the amine hydrogen of Met1181 backbone

The analysis of output revealed a stable trend for the α-helicity issue of the peptide, that held its folded conformation. Even the RMSD plot (plot 7.4) showed certain stability of the system, whereas only the first portion (until about 120 ns of the simulation) was not so stable, thus it was not considered for the analysis. On the other hand, the second part of the plot showed a stable trend, so those frames were analysed to retrieve information about the most stable interactions. Then, energy, temperature, pressure and volume of the system were monitored during the whole trajectory (average energy values in table 7.9).



Plot 7.4. RMSD plot of MD simulation on Ras protein in complex with peptide 3_{10}-HBS RB3

Table 7.9. Energy values monitored during the MD simulation of Ras in complex with peptide 3_{10}-HBS RB3

	Average	Std Dev	Slope (ps^{-1})
Total energy (kcal/mol)	-62709.783	97.463	0.000
Potential energy (kcal/mol)	-77348.359	74.132	0.000

The MD frames were then grouped into ten clusters, where the frames representative for the most abundant clusters are the following ones: frame 4560 (107 frames), frame 3790 (33 frames), frame 3110 (199 frames), frame 3460 (35 frames),

frame 380 (73 frames) and frame 1170 (30 frames). The interactions of these frames were observed and collected into table 7.10 to identify the most stable and frequent contacts between Ras protein and peptide 3_{10}-HBS RB3.

Table 7.10. The most stable interactions between Ras and peptide 3_{10}-HBS RB3 highlighted from MD trajectory analysis

Ras aa	Peptide 3_{10}-HBS RB3 aa	Interaction type
Gln61	Gly1180	1 H-bond
Gln61	Thr1184	1 H-bond
Ala59	Thr1184	1 H-bond
Tyr40	Asp1185	1 H-bond
Tyr40	Phe1188	Pi-Pi stacking
Tyr32	Glu1191	1 H-bond
Tyr32	Thr1193	1 H-bond
Tyr32	Gly1192	1 H-bond
Ser17	Glu1191	1 H-bond

This new designed peptide RB3 showed to establish some of the key interactions identified from the previous MD simulation between Ras and RasGRF1 proteins (table 7.6) and other contacts with Ras amino acids highlighted as key residues from mutational studies [92], showing the potential for binding Ras protein.

Finally, MM-GBSA calculations were performed to get the $\Delta G_{binding}$ of the complex Ras- peptide 3_{10}-HBS RB3 that was measured by -79.6995 kcal/mol. In this way, this value was used as a reference for the peptide optimisation process, as it will be described in the next sections.

7.2.5 Computational residue scanning of the peptide 3_{10}-HBS RB3 and MD simulations of point-mutated peptides

In order to optimise the structure of the peptide 3_{10}-HBS RB3, a computational residue scanning was performed on the amino acids composing the peptide by using the "Residue scanning" tool of Bioluminate [133]. The aim was to identify the most promising mutations in terms of $\Delta\Delta G_{affinity}$ and $\Delta\Delta G_{stability}$. For this purpose, only mutations reporting both $\Delta\Delta G_{affinity}$ and $\Delta\Delta G_{stability}$ values below -3 kcal/mol were considered according to Beard *et al*. work [133]. Table 7.11 reports sixteen identified mutations to be considered for the next steps of this work.

Table 7.11. Computational residue scanning results on the peptide 3_{10}-HBS RB3 highlighting sixteen promising mutations

Peptide 3_{10}-HBS RB3 aa	Mutation	$\Delta\Delta G_{affinity}$	$\Delta\Delta G_{stability}$
Thr1184	Arg	-19.17	-3.67
	Met	-8.17	-3.18
Asp1185	Trp	-12.56	-7.64
	Tyr	-9.19	-3.49
	Phe	-8.26	-4.26
	Leu	-7.77	-12.61
Phe1188	Arg	-9.55	-4.12

Phe1188	His	-8.25	-8.75
Ile1189	Met	-3.59	-4.08
Glu1190	His	-3.29	-4.13
Glu1191	Ile	-6.97	-4.68
	Leu	-5.16	-3.79
	Val	-4.71	-4.08
	Thr	-4.32	-4.27
Thr1193	Arg	-3.85	-5.84
	Gln	-3.56	-5.86

These sixteen mutations were used to create as many point-mutated peptides complexed with Ras protein that underwent MD simulations. The trajectory time was set 100 ns per each system, since this timeframe was considered suitable to detect potential misfolding of the peptides. Indeed, the previously described MD simulations on the wild-type peptide RB3 exhibited misfolded conformation by losing α-helicity after about 50-60 ns of simulation. All the point-mutated peptides were able to keep the helical conformation, thus MM-GBSA calculations were computed and the related $\Delta G_{binding}$ values are reported below in table 7.12. The stability of the systems was investigated by analysing the RMSD plots per each complex, resulting in suitable stationary shape for each system. However, the plots are not reported herein for spatial needs. Furthermore, energy, temperature, pressure and volume were ensured to be stable during the whole trajectories.

Table 7.12. MM-GBSA calculation results based on MD simulations of sixteen point-mutated peptides 3_{10}-HBS RB3 in complex with Ras protein

	First Peptide	**Second Peptide**	**Third Peptide**
Point mutation	T1184R	T1184M	D1185W
$\Delta G_{binding}$ average	-89.5064 kcal/mol	-92.7657 kcal/mol	-103.5006 kcal/mol
$\Delta G_{binding}$ Std. Dev.	12.50	15.53	8.90
$\Delta G_{binding}$ range	-128.5163 to -55.496 kcal/mol	-134.2373 to -39.5118 kcal/mol	-126.2946 to -73.3538 kcal/mol
	Fourth Peptide	**Fifth Peptide**	**Sixth Peptide**
Point mutation	D1185Y	D1185F	D1185L
$\Delta G_{binding}$ average	-102.5021 kcal/mol	-94.8370 kcal/mol	-82.0725 kcal/mol
$\Delta G_{binding}$ Std. Dev.	22.23	8.34	9.26
$\Delta G_{binding}$ range	-145.438 to -40.6732 kcal/mol	-120.1250 to -52.3853 kcal/mol	-109.1876 to -32.6282 kcal/mol
	Seventh Peptide	**Eighth Peptide**	**Ninth Peptide**
Point mutation	F1188R	F1188H	I1189M
$\Delta G_{binding}$ average	-87.4886 kcal/mol	-69.5804 kcal/mol	-83.1178 kcal/mol
$\Delta G_{binding}$ Std. Dev.	11.60	15.23	10.89
$\Delta G_{binding}$ range	-120.3205 to -57.755 kcal/mol	-111.448 to -24.9385 kcal/mol	-122.6298 to -44.9273 kcal/mol
	Tenth Peptide	**Eleventh Peptide**	**Twelfth Peptide**
Point mutation	E1190H	E1191I	E1191L
$\Delta G_{binding}$ average	-73.3614 kcal/mol	-78.6476 kcal/mol	-95.1115 kcal/mol
$\Delta G_{binding}$ Std. Dev.	12.64	11.67	12.89
$\Delta G_{binding}$ range	-110.7906 to -39.815 kcal/mol	-115.3326 to -43.7585 kcal/mol	-140.5009 to -49.6646 kcal/mol
	Thirteenth Peptide	**Fourteenth Peptide**	**Fifteenth Peptide**
Point mutation	E1191V	E1191T	T1193R
$\Delta G_{binding}$ average	-94.4207 kcal/mol	-84.5016 kcal /mol	-90.1769 kcal/mol
$\Delta G_{binding}$ Std. Dev.	10.69	12.71	10.85
$\Delta G_{binding}$ range	-121.235 to -58.7509 kcal/mol	-116.5146 to -47.4814 kcal/mol	-119.8724 to -59.7230 kcal/mol
	Sixteenth Peptide		
Point mutation	T1193N		
$\Delta G_{binding}$ average	-97.1463 kcal/mol		
$\Delta G_{binding}$ Std. Dev.	11.88		
$\Delta G_{binding}$ range	-127.145 to -53.4326 kcal/mol		

As previously mentioned, the $\Delta G_{binding}$ of the complex between Ras and the wild-type peptide 3_{10}-HBS RB3 (-79.6995 kcal/mol) was used as a reference to select the most promising mutations associated to $\Delta G_{binding}$ values lower than the reference one. In the light of the above, from MM-GBSA results (reported in table 7.12) only three mutated peptides showed higher $\Delta G_{binding}$ values. Hence, the related mutations (F1188H, E1190H and E1191I) were neglected. On the contrary, the other thirteen mutations were considered for creating combinatorial peptides, as described in the next section.

7.2.6 Combinatorial peptides 3_{10}-HBS RB3 creation and MD simulations

The most promising mutations on peptide 3_{10}-HBS RB3 were combined each other to get overall 48 mutated peptides as below listed in table 7.13.

Table 7.13. Combinatorial peptides designed based on computational alanine scanning performed on peptide 3_{10}-HBS RB3

Combinatorial peptides	
1. YLGMYLRWLVRMELGR	25. YLGMYLRWLVRMELGN
2. YLGMYLMWLVRMELGR	26. YLGMYLMWLVRMELGN
3. YLGMYLRYLVRMELGR	27. YLGMYLRYLVRMELGN
4. YLGMYLMYLVRMELGR	28. YLGMYLMYLVRMELGN
5. YLGMYLRFLVRMELGR	29. YLGMYLRFLVRMELGN
6. YLGMYLMFLVRMELGR	30. YLGMYLMFLVRMELGN
7. YLGMYLRLLVRMELGR	31. YLGMYLRLLVRMELGN
8. YLGMYLMLLVRMELGR	32. YLGMYLMLLVRMELGN
9. YLGMYLRWLVRMEVGR	33. YLGMYLRWLVRMEVGN
10. YLGMYLRYLVRMEVGR	34. YLGMYLRYLVRMEVGN
11. YLGMYLRFLVRMEVGR	35. YLGMYLRFLVRMEVGN
12. YLGMYLRLLVRMEVGR	36. YLGMYLRLLVRMEVGN
13. YLGMYLMWLVRMEVGR	37. YLGMYLMWLVRMEVGN
14. YLGMYLMYLVRMEVGR	38. YLGMYLMYLVRMEVGN
15. YLGMYLMFLVRMEVGR	39. YLGMYLMFLVRMEVGN
16. YLGMYLMLLVRMEVGR	40. YLGMYLMLLVRMEVGN
17. YLGMYLRWLVRMETGR	41. YLGMYLRWLVRMETGN
18. YLGMYLRYLVRMETGR	42. YLGMYLRYLVRMETGN
19. YLGMYLRFLVRMETGR	43. YLGMYLRFLVRMETGN
20. YLGMYLRLLVRMETGR	44. YLGMYLRLLVRMETGN
21. YLGMYLMWLVRMETGR	45. YLGMYLMWLVRMETGN
22. YLGMYLMYLVRMETGR	46. YLGMYLMYLVRMETGN
23. YLGMYLMFLVRMETGR	47. YLGMYLMFLVRMETGN
24. YLGMYLMLLVRMETGR	48. YLGMYLMLLVRMETGN

The resulting peptides were complexed with Ras protein and the systems were processed by running MD simulations of 100 ns each to investigate helix conformational stability. All the trajectories were then observed by generating RMSD plots and energy, temperature, pressure and volume were monitored to ensure the outputs reliability. The related data are not reported herein for spatial needs. Finally, MM-GBSA calculations of the MD simulations were computed. Thus, all those combinatorial peptides not responding to the following criteria were neglected: 1) $\Delta G_{binding}$ value higher than the reference one (-79.6995 kcal/mol), and 2) loses of helical conformation. Finally, overall eighteen combinatorial peptides resulted in promising $\Delta G_{binding}$ values and exhibited

helicity. Table 7.14 reports the MM-GBSA results of the most promising combinatorial peptides. Hence, they were considered to be processed through biological assays.

Table 7.14. MM-GBSA calculation results based on MD simulations of 3_{10}-HBS combinatorial peptides not misfolded during the simulations in complex with Ras protein and with $\Delta G_{binding}$ values lower than the reference one (-79.6995 kcal/mol)

	First Peptide	**Third Peptide**	**Eleventh Peptide**
Peptide sequence	YLGMYLRWLVRMELGR	YLGMYLRYLVRMELGR	YLGMYLRFLVRMEVGR
$\Delta G_{binding}$ average	-83.4644 kcal/mol	-96.8031 kcal/mol	-91.3915 kcal/mol
$\Delta G_{binding}$ Std. Dev.	9.00	9.84	11.55
$\Delta G_{binding}$ range	-117.5361 to -55.065 kcal/mol	-123.0108 to -52.3826 kcal/mol	-119.6650 to -61.3426 kcal/mol
	Twelfth Peptide	**Fifteenth Peptide**	**Sixteenth Peptide**
Peptide sequence	YLGMYLRLLVRMEVGR	YLGMYLMFLVRMEVGR	YLGMYLMLLVRMEVGR
$\Delta G_{binding}$ average	-92.4942 kcal/mol	-79.7912 kcal/mol	-92.5403 kcal/mol
$\Delta G_{binding}$ Std. Dev.	10.15	14.46	7.93
$\Delta G_{binding}$ range	-129.871 to -56.4907 kcal/mol	-112.9407 to -41.4205 kcal/mol	-116.2758 to -65.178 kcal/mol
	Eighteenth Peptide	**Nineteenth Peptide**	**Twentieth Peptide**
Peptide sequence	YLGMYLRYLVRMETGR	YLGMYLRFLVRMETGR	YLGMYLRLLVRMETGR
$\Delta G_{binding}$ average	-100.3350 kcal/mol	-102.6262 kcal/mol	-88.7131 kcal/mol
$\Delta G_{binding}$ Std. Dev.	14.21	11.01	13.50
$\Delta G_{binding}$ range	-137.525 to -65.4195 kcal/mol	-130.8724 to -62.6852 kcal/mol	-125.0219 to -50.9362 kcal/mol
	Twenty-third Peptide	**Twenty-fourth Peptide**	**Twenty-fifth Peptide**
Peptide sequence	YLGMYLMFLVRMETGR	YLGMYLMLLVRMETGR	YLGMYLRWLVRMELGN
$\Delta G_{binding}$ average	-85.5260 kcal/mol	-82.3122 kcal/mol	-97.2416 kcal /mol
$\Delta G_{binding}$ Std. Dev.	11.35	10.37	14.17
$\Delta G_{binding}$ range	-117.7259 to -42.841 kcal/mol	-116.0259 to -40.5661 kcal/mol	-133.3342 to -60.1542 kcal/mol
	Twenty-ninth Peptide	**Forty-second Peptide**	**Forty-third Peptide**
Peptide sequence	YLGMYLRFLVRMELGN	YLGMYLRYLVRMETGN	YLGMYLRFLVRMETGN
$\Delta G_{binding}$ average	-86.5622 kcal/mol	-89.5885 kcal/mol	-123.4975 kcal/mol
$\Delta G_{binding}$ Std. Dev.	11.67	10.32	20.97
$\Delta G_{binding}$ range	-124.698 to -56.4707 kcal/mol	-128.7509 to -60.0016 kcal/mol	-161.4939 to -75.9053 kcal/mol
	Forty-fourth Peptide	**Forty-fifth Peptide**	**Forty-eighth Peptide**
Peptide sequence	YLGMYLRLLVRMETGN	YLGMYLMWLVRMETGN	YLGMYLMLLVRMETGN
$\Delta G_{binding}$ average	-96.3137 kcal/mol	-86.0366 kcal/mol	-91.7996 kcal/mol
$\Delta G_{binding}$ Std. Dev.	17.98	13.60	9.51
$\Delta G_{binding}$ range	-137.9689 to -57.812 kcal/mol	-124.0792 to -53.760 kcal mol	-122.7787 to -60.8432 kcal/mol

7.3 Methods

7.3.1 Protein preparation

The 3D structures of Ras-Sos complex (PDB IDs: 1XD2, 1BKD, 1NVW, 1NVV, 1NVU and 1NVX) and RasGRF1 protein (PDB ID: 2IJE) were downloaded from Protein Data Bank [134] and imported in Schrödinger suite to optimise the structure by using "Protein preparation" tool [135]. The bond orders for untemplated residues were assigned by using known HET groups based on their SMILES strings in Chemical Component Dictionary. Hydrogens were added to the structure, eventual bonds to metals were broken, zero-order bonds between metals and nearby atoms were added and formal charges to metals and neighbouring atoms were corrected. Disulfide bonds between two sulfurs, if they were close to each other, were created and water molecules beyond 5.0 Å from any of the HET groups, including ions, were deleted. Then, protonation and metal charge states for the ligands, cofactors and metals were generated [136, 137]. Finally, PROPKA [137] was run under pH 7.0 to optimise hydroxyl, Asn, Gln and His states using ProtAssign.

7.3.2 MD simulations of Ras protein in complex with Sos, RasGRF1, peptide RB3 and the designed 3_{10}-HBS peptides

In this work overall sixty-nine MD simulations were performed, as follows: one MD simulation of 500 ns for Ras-Sos complex, one MD simulation of 500 ns for Ras-RasGRF1 complex, two MD simulations of 500 ns for Ras in complex with the wild-type peptide RB3, one MD simulation of 500 ns for Ras in complex with the peptide 3_{10}-HBS RB3, sixteen MD simulations of 100 ns for Ras complexed with the point-mutated 3_{10}-HBS peptides and forty-eight MD simulations of 100 ns for Ras in complex with the combinatorial 3_{10}-HBS peptides.

All of them were run by applying the same MD settings. The systems were created using the "System builder" tool of Schrödinger suite. TIP3P was selected as a solvent model and the chosen box shape was orthorhombic. The selected force field was OPLS3 [138]. The selected box size calculation method was buffer, the box side distances were set 10 Å and the system was neutralized by adding Na^+ ions. The outputs were further processed by performing MD simulations choosing the simulation time as above reported. The number of atoms, the pressure and the temperature were maintained constant for the entire trajectories. Pressure and temperature were set 1.01325 bar and 300 K, respectively. Finally, the systems were relaxed before starting the simulation.

7.3.3 MD frames clustering

In order to retrieve the key contacts between the protein partners during the entire simulations, for each above described MD simulation the frames were clustered to identify the most representative cluster centroids for the MD to be analysed. Therefore, "Desmond trajectory clustering" tool by Schrödinger was used. For the RMSD matrix calculation the protein backbone was used, the frequency of frames analysis was set 10 and the hierarchical cluster linkage method as average. Finally, for each MD trajectory ten clusters were generated.

7.3.4 Computational residue scanning of peptide 3_{10}-HBS RB3 in complex with Ras

The peptide 3_{10}-HBS RB3 in complex with Ras (PDB ID: 1XD2) was used to perform a computational residue scanning by using the "Residue Scanning" tool of Schrödinger suite to perform point mutations on the peptide residues. The stability and affinity were computed for each of these mutations and the resulting structures were refined by selecting side-chain prediction with backbone minimization.

7.3.5 MM-GBSA calculations of all the complexes used to perform MD

The MD outputs of Ras protein in complex with the peptide 3_{10}-HBS RB3, the point-mutated peptides and the combinatorial peptides were used to compute MM-GBSA calculations through the terminal. For this purpose, the Python script "thermal_mmgbsa.py" was used.

Overall thirty-five MM-GBSA calculations were performed and data are reported in the above "Results and discussion" section.

7.4 Conclusions

The workflow above reported describes the optimisation process of the patented peptide RB3 [112]. This peptide has been reported blocking the Ras-ERK signalling pathway by targeting the interaction between Ras protein and guanine nucleotide exchange factors [104]. The computational workflow was performed at the School of pharmacy and pharmaceutical sciences of Cardiff University (UK) under the supervision of Professor Andrea Brancale. The applied approaches allowed to identify eighteen peptides based on the peptide RB3 structure, including amino acids mutations and an artificial construct, the hydrogen bond surrogate, to stabilise the helical conformation of the peptides. These molecules will be synthesised and assayed by performing the same tests that allowed to identify the peptide RB3 at Cardiff University, *i.e.* the phosphorylation rate of ERK and two well-characterised ERK substrates, (Ser10)-acetylated (Lys14) histone H3 and S6 ribosomal protein [109–111], will be measured.

The results of the biological screening will provide crucial information about the potential of these designed peptides in inhibiting Ras activation, thus preventing molecular effects and drug-seeking behaviours associated with cocaine abuse.

Finally, this work has been considering for a research article that soon will be sent to a scientific peer-reviewed journal.

References – Chapter Seven

1. (2020) NIH Cocaine trends & statistics. https://www.drugabuse.gov/drug-topics/cocaine/cocaine-trends-statistics. Accessed 20 Aug 2020
2. Everitt BJ (2014) Neural and psychological mechanisms underlying compulsive drug seeking habits and drug memories - indications for novel treatments of addiction. Eur J Neurosci 40:2163–2182. https://doi.org/10.1111/ejn.12644
3. Berhow MT, Hiroi N, Nestler EJ (1996) Regulation of ERK (Extracellular Signal Regulated Kinase), Part of the Neurotrophin Signal Transduction Cascade, in the Rat Mesolimbic Dopamine System by Chronic Exposure to Morphine or Cocaine. J Neurosci 16:4707–4715. https://doi.org/10.1523/JNEUROSCI.16-15-04707.1996
4. Belin D, Belin-Rauscent A, Murray JE, Everitt BJ (2013) Addiction: failure of control over maladaptive incentive habits. Curr Opin Neurobiol 23:564–572. https://doi.org/10.1016/j.conb.2013.01.025
5. Sanchis-Segura C, Spanagel R (2006) Behavioural assessment of drug reinforcement and addictive features in rodents: an overview. Addict Biol 11:2–38. https://doi.org/10.1111/j.1369-1600.2006.00012.x
6. Nestler EJ (2001) Molecular basis of long-term plasticity underlying addiction. Nat Rev Neurosci 2:119–128. https://doi.org/10.1038/35053570
7. Kalivas P (2004) Glutamate systems in cocaine addiction. Curr Opin Pharmacol 4:23–29. https://doi.org/10.1016/j.coph.2003.11.002
8. Feltenstein MW, See RE (2009) The neurocircuitry of addiction: an overview. Br J Pharmacol 154:261–274. https://doi.org/10.1038/bjp.2008.51
9. (2020) NIH Cocaine DrugFacts. https://www.drugabuse.gov/publications/drugfacts/cocaine. Accessed 20 Aug 2020
10. Nestler EJ (2004) Molecular mechanisms of drug addiction. Neuropharmacology 47:24–32. https://doi.org/10.1016/j.neuropharm.2004.06.031
11. Miller CA, Marshall JF (2005) Molecular Substrates for Retrieval and Reconsolidation of Cocaine-Associated Contextual Memory. Neuron 47:873–884. https://doi.org/10.1016/j.neuron.2005.08.006
12. Torregrossa MM, Corlett PR, Taylor JR (2011) Aberrant learning and memory in addiction. Neurobiol Learn Mem 96:609–623. https://doi.org/10.1016/j.nlm.2011.02.014
13. Itzhak Y, Perez-Lanza D, Liddie S (2014) The strength of aversive and appetitive associations and maladaptive behaviors. IUBMB Life 66:559–571. https://doi.org/10.1002/iub.1310
14. Valjent E, Corvol J-C, Pagès C, et al (2000) Involvement of the Extracellular Signal-Regulated Kinase Cascade for Cocaine-Rewarding Properties. J Neurosci 20:8701–8709. https://doi.org/10.1523/JNEUROSCI.20-23-08701.2000
15. Lu L, Hope BT, Dempsey J, et al (2005) Central amygdala ERK signaling pathway is critical to incubation of cocaine craving. Nat Neurosci 8:212–219. https://doi.org/10.1038/nn1383
16. Valjent E, Corbille A-G, Bertran-Gonzalez J, et al (2006) Inhibition of ERK pathway or protein synthesis during reexposure to drugs of abuse erases previously learned place preference. Proc Natl Acad Sci 103:2932–2937. https://doi.org/10.1073/pnas.0511030103
17. Valjent E, Corvol J-C, Trzaskos JM, et al (2006) Role of the ERK pathway in psychostimulant-induced locomotor sensitization. BMC Neurosci 7:20. https://doi.org/10.1186/1471-2202-7-20
18. Lu L, Koya E, Zhai H, et al (2006) Role of ERK in cocaine addiction. Trends Neurosci 29:695–703. https://doi.org/10.1016/j.tins.2006.10.005
19. Ferguson SM, Fasano S, Yang P, et al (2006) Knockout of ERK1 Enhances Cocaine-Evoked Immediate Early Gene Expression and Behavioral Plasticity. Neuropsychopharmacology 31:2660–2668. https://doi.org/10.1038/sj.npp.1301014
20. Girault J, Valjent E, Caboche J, Herve D (2007) ERK2: a logical AND gate critical for drug-induced plasticity? Curr Opin Pharmacol 7:77–85. https://doi.org/10.1016/j.coph.2006.08.012
21. Fasano S, D'Antoni A, Orban PC, et al (2009) Ras-Guanine Nucleotide-Releasing Factor 1 (Ras-GRF1) Controls Activation of Extracellular Signal-Regulated Kinase (ERK) Signaling in the Striatum and Long-Term Behavioral Responses to Cocaine. Biol Psychiatry 66:758–768. https://doi.org/10.1016/j.biopsych.2009.03.014
22. Fasano S, Brambilla R (2011) Ras-ERK Signaling in Behavior: Old Questions and New Perspectives. Front Behav Neurosci 5:. https://doi.org/10.3389/fnbeh.2011.00079
23. Pascoli V, Besnard A, Hervé D, et al (2011) Cyclic Adenosine Monophosphate–Independent

Tyrosine Phosphorylation of NR2B Mediates Cocaine-Induced Extracellular Signal-Regulated Kinase Activation. Biol Psychiatry 69:218–227. https://doi.org/10.1016/j.biopsych.2010.08.031
24. Pascoli V, Turiault M, Lüscher C (2012) Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. Nature 481:71–75. https://doi.org/10.1038/nature10709
25. Pascoli V, Cahill E, Bellivier F, et al (2014) Extracellular Signal-Regulated Protein Kinases 1 and 2 Activation by Addictive Drugs: A Signal Toward Pathological Adaptation. Biol Psychiatry 76:917–926. https://doi.org/10.1016/j.biopsych.2014.04.005
26. Cahill E, Salery M, Vanhoutte P, Caboche J (2014) Convergence of dopamine and glutamate signaling onto striatal ERK activation in response to drugs of abuse. Front Pharmacol 4:. https://doi.org/10.3389/fphar.2013.00172
27. García-Pardo MP, Roger-Sanchez C, Rodríguez-Arias M, et al (2016) Pharmacological modulation of protein kinases as a new approach to treat addiction to cocaine and opiates. Eur J Pharmacol 781:10–24. https://doi.org/10.1016/j.ejphar.2016.03.065
28. Kyosseva S V (2004) Mitogen-Activated Protein Kinase Signaling. pp 201–220
29. Kim EK, Choi E-J (2010) Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta - Mol Basis Dis 1802:396–405. https://doi.org/10.1016/j.bbadis.2009.12.009
30. Seger R, Krebs EG (1995) The MAPK signaling cascade. FASEB J 9:726–35
31. Chen RH, Sarnecki C, Blenis J (1992) Nuclear localization and regulation of erk- and rsk-encoded protein kinases. Mol Cell Biol 12:915–927. https://doi.org/10.1128/MCB.12.3.915
32. Gille H, Strahl T, Shaw PE (1995) Activation of ternary complex factor Elk-1 by stress-activated protein kinases. Curr Biol 5:1191–1200. https://doi.org/10.1016/S0960-9822(95)00235-1
33. Gille H, Sharrocks AD, Shaw PE (1992) Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter. Nature 358:414–417. https://doi.org/10.1038/358414a0
34. Davis S, Vanhoutte P, Pagès C, et al (2000) The MAPK/ERK Cascade Targets Both Elk-1 and cAMP Response Element-Binding Protein to Control Long-Term Potentiation-Dependent Gene Expression in the Dentate Gyrus In Vivo. J Neurosci 20:4563–4572. https://doi.org/10.1523/JNEUROSCI.20-12-04563.2000
35. Hill CS, Marais R, John S, et al (1993) Functional analysis of a growth factor-responsive transcription factor complex. Cell 73:395–406. https://doi.org/10.1016/0092-8674(93)90238-L
36. Treisman R (1996) Regulation of transcription by MAP kinase cascades. Curr Opin Cell Biol 8:205–215. https://doi.org/10.1016/S0955-0674(96)80067-6
37. Zheng Y, Quilliam LA (2003) Activation of the Ras superfamily of small GTPases. EMBO Rep 4:463–468. https://doi.org/10.1038/sj.embor.embor831
38. Atkins CM, Selcher JC, Petraitis JJ, et al (1998) The MAPK cascade is required for mammalian associative learning. Nat Neurosci 1:602–609. https://doi.org/10.1038/2836
39. Scherle PA, Ma W, Lim H, et al (2000) Regulation of Cyclooxygenase-2 Induction in the Mouse Uterus During Decidualization. J Biol Chem 275:37086–37092. https://doi.org/10.1074/jbc.M006168200
40. Sun W-L, Quizon PM, Zhu J (2016) Molecular Mechanism: ERK Signaling, Drug Addiction, and Behavioral Effects. pp 1–40
41. Pierce RC, Pierce-Bancroft AF, Prasad BM (1999) Neurotrophin-3 Contributes to the Initiation of Behavioral Sensitization to Cocaine by Activating the Ras/Mitogen-Activated Protein Kinase Signal Transduction Cascade. J Neurosci 19:8685–8695. https://doi.org/10.1523/JNEUROSCI.19-19-08685.1999
42. Giordano TP, Tropea TF, Satpute SS, et al (2010) Molecular Switch from L-Type Cav1.3 to Cav1.2 Ca2+ Channel Signaling Underlies Long-Term Psychostimulant-Induced Behavioral and Molecular Plasticity. J Neurosci 30:17051–17062. https://doi.org/10.1523/JNEUROSCI.2255-10.2010
43. Kim S, Shin J-K, Yoon HS, Kim J-H (2011) Blockade of ERK Phosphorylation in the Nucleus Accumbens Inhibits the Expression of Cocaine-induced Behavioral Sensitization in Rats. Korean J Physiol Pharmacol 15:389–95. https://doi.org/10.4196/kjpp.2011.15.6.389
44. Liddie S, Itzhak Y (2016) Variations in the stimulus salience of cocaine reward influences drug-associated contextual memory. Addict Biol 21:242–254. https://doi.org/10.1111/adb.12191
45. Wells AM, Arguello AA, Xie X, et al (2013) Extracellular Signal-Regulated Kinase in the Basolateral Amygdala, but not the Nucleus Accumbens Core, is Critical for Context-Response-Cocaine Memory Reconsolidation in Rats. Neuropsychopharmacology 38:753–762.

https://doi.org/10.1038/npp.2012.238
46. Favata MF, Horiuchi KY, Manos EJ, et al (1998) Identification of a Novel Inhibitor of Mitogen-activated Protein Kinase Kinase. J Biol Chem 273:18623–18632. https://doi.org/10.1074/jbc.273.29.18623
47. Lu L, Wang X, Wu P, et al (2009) Role of Ventral Tegmental Area Glial Cell Line–Derived Neurotrophic Factor in Incubation of Cocaine Craving. Biol Psychiatry 66:137–145. https://doi.org/10.1016/j.biopsych.2009.02.009
48. Lu L (2004) A Single Infusion of Brain-Derived Neurotrophic Factor into the Ventral Tegmental Area Induces Long-Lasting Potentiation of Cocaine Seeking after Withdrawal. J Neurosci 24:1604–1611. https://doi.org/10.1523/JNEUROSCI.5124-03.2004
49. Whitfield TW, Shi X, Sun W-L, McGinty JF (2011) The Suppressive Effect of an Intra-Prefrontal Cortical Infusion of BDNF on Cocaine-Seeking Is Trk Receptor and Extracellular Signal-Regulated Protein Kinase Mitogen-Activated Protein Kinase Dependent. J Neurosci 31:834–842. https://doi.org/10.1523/JNEUROSCI.4986-10.2011
50. Gerdjikov T V., Ross GM, Beninger RJ (2004) Place Preference Induced by Nucleus Accumbens Amphetamine Is Impaired by Antagonists of ERK or p38 MAP Kinases in Rats. Behav Neurosci 118:740–750. https://doi.org/10.1037/0735-7044.118.4.740
51. Edwards S, Graham DL, Bachtell RK, Self DW (2007) Region-specific tolerance to cocaine-regulated cAMP-dependent protein phosphorylation following chronic self-administration. Eur J Neurosci 25:2201–2213. https://doi.org/10.1111/j.1460-9568.2007.05473.x
52. Miszkiel J, Detka J, Cholewa J, et al (2014) The effect of active and passive intravenous cocaine administration on the extracellular signal-regulated kinase (ERK) activity in the rat brain. Pharmacol Reports 66:630–637. https://doi.org/10.1016/j.pharep.2014.02.001
53. Koya E, Uejima JL, Wihbey KA, et al (2009) Role of ventral medial prefrontal cortex in incubation of cocaine craving. Neuropharmacology 56:177–185. https://doi.org/10.1016/j.neuropharm.2008.04.022
54. Edwards S, Bachtell RK, Guzman D, et al (2011) Emergence of context-associated GluR1 and ERK phosphorylation in the nucleus accumbens core during withdrawal from cocaine self-administration. Addict Biol 16:450–457. https://doi.org/10.1111/j.1369-1600.2010.00296.x
55. Sun W-L, Zelek-Molik A, McGinty JF (2013) Short and long access to cocaine self-administration activates tyrosine phosphatase STEP and attenuates GluN expression but differentially regulates GluA expression in the prefrontal cortex. Psychopharmacology (Berl) 229:603–613. https://doi.org/10.1007/s00213-013-3118-5
56. Doyle SE, Ramôa C, Garber G, et al (2014) A Shift in the Role of Glutamatergic Signaling in the Nucleus Accumbens Core With the Development of an Addicted Phenotype. Biol Psychiatry 76:810–815. https://doi.org/10.1016/j.biopsych.2014.02.005
57. Vetter IR, Wittinghofer A (2001) The Guanine Nucleotide-Binding Switch in Three Dimensions. Science (80-) 294:1299–1304. https://doi.org/10.1126/science.1062023
58. Mattingly RR, Macara IG (1996) Phosphorylation-dependent activation of the Ras-GRF/CDC25Mm exchange factor by muscarinic receptors and G-protein βγ subunits. Nature 382:268–272. https://doi.org/10.1038/382268a0
59. Baouz S, Jacquet E, Bernardi A, Parmeggiani A (1997) The N-terminal Moiety of CDC25 Mm , a GDP/GTP Exchange Factor of Ras Proteins, Controls the Activity of the Catalytic Domain. J Biol Chem 272:6671–6676. https://doi.org/10.1074/jbc.272.10.6671
60. Farnsworth CL, Freshney NW, Rosen LB, et al (1995) Calcium activation of Ras mediated by neuronal exchange factor Ras-GRF. Nature 376:524–527. https://doi.org/10.1038/376524a0
61. Schweighoffer F, Faure M, Fath I, et al (1993) Identification of a human guanine nucleotide-releasing factor (H-GRF55) specific for Ras proteins. Oncogene 8:1477–1485
62. Lenzen C, Cool RH, Wittinghofer A (1995) Analysis of intrinsic and CDC25-stimulated guanine nucleotide exchange of p21ras-nucleotide complexes by fluorescence measurements. pp 95–109
63. Tian X, Gotoh T, Tsuji K, et al (2004) Developmentally regulated role for Ras-GRFs in coupling NMDA glutamate receptors to Ras, Erk and CREB. EMBO J 23:1567–1575. https://doi.org/10.1038/sj.emboj.7600151
64. Bar-Sagi D (1994) The Sos (Son of sevenless) protein. Trends Endocrinol Metab 5:165–169. https://doi.org/10.1016/1043-2760(94)90014-0
65. Tinhofer I, Maly K, Dietl P, et al (1996) Differential Ca 2+ Signaling Induced by Activation of the Epidermal Growth Factor and Nerve Growth Factor Receptors. J Biol Chem 271:30505–30509. https://doi.org/10.1074/jbc.271.48.30505

66. Egan SE, Giddings BW, Brooks MW, et al (1993) Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation. Nature 363:45–51. https://doi.org/10.1038/363045a0
67. Gale NW, Kaplan S, Lowenstein EJ, et al (1993) Grb2 mediates the EGF-dependent activation of guanine nucleotide exchange on Ras. Nature 363:88–92. https://doi.org/10.1038/363088a0
68. Li N, Batzer A, Daly R, et al (1993) Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling. Nature 363:85–88. https://doi.org/10.1038/363085a0
69. Buday L, Downward J (1993) Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein, and Sos nucleotide exchange factor. Cell 73:611–620. https://doi.org/10.1016/0092-8674(93)90146-H
70. Lenzen C, Cool RH, Prinz H, et al (1998) Kinetic Analysis by Fluorescence of the Interaction between Ras and the Catalytic Domain of the Guanine Nucleotide Exchange Factor Cdc25 Mm †. Biochemistry 37:7420–7430. https://doi.org/10.1021/bi972621j
71. Zarich N, Oliva JL, Jorge R, et al (2000) The isoform-specific stretch of hSos1 defines a new Grb2-binding domain. Oncogene 19:5872–5883. https://doi.org/10.1038/sj.onc.1203955
72. Boriack-Sjodin PA, Margarit SM, Bar-Sagi D, Kuriyan J (1998) The structural basis of the activation of Ras by Sos. Nature 394:337–343. https://doi.org/10.1038/28548
73. Chardin P, Mattei M-G (1994) Chromosomal localization of two genes encoding human ras exchange factors: SOS1 maps to the 2p22 -> p16 region and SOS2 to the 14q21 -> q22 region of the human genome. Cytogenet Genome Res 66:68–69. https://doi.org/10.1159/000133667
74. Boguski MS, McCormick F (1993) Proteins regulating Ras and its relatives. Nature 366:643–654. https://doi.org/10.1038/366643a0
75. Liu BX, Wei W, Broek D (1993) The catalytic domain of the mouse sos1 gene product activates Ras proteins in vivo and in vitro. Oncogene 8:3081–4
76. Margarit SM, Sondermann H, Hall BE, et al (2003) Structural Evidence for Feedback Activation by Ras·GTP of the Ras-Specific Nucleotide Exchange Factor SOS. Cell 112:685–695. https://doi.org/10.1016/S0092-8674(03)00149-1
77. Sondermann H, Soisson SM, Boykevisch S, et al (2004) Structural Analysis of Autoinhibition in the Ras Activator Son of Sevenless. Cell 119:393–405. https://doi.org/10.1016/j.cell.2004.10.005
78. Freedman TS, Sondermann H, Friedland GD, et al (2006) A Ras-induced conformational switch in the Ras activator Son of sevenless. Proc Natl Acad Sci 103:16692–16697. https://doi.org/10.1073/pnas.0608127103
79. Guo Z, Ahmadian MR, Goody RS (2005) Guanine Nucleotide Exchange Factors Operate by a Simple Allosteric Competitive Mechanism †. Biochemistry 44:15423–15429. https://doi.org/10.1021/bi0518601
80. Chenna R, Sugawara H, Koike T, et al (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500. https://doi.org/10.1093/nar/gkg500
81. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic Acids Res 32:W321–W326. https://doi.org/10.1093/nar/gkh377
82. Willumsen BM, Papageorge AG, Kung HF, et al (1986) Mutational analysis of a ras catalytic domain. Mol Cell Biol 6:2646–2654. https://doi.org/10.1128/MCB.6.7.2646
83. Fasano O, Crechet JB, De Vendittis E, et al (1988) Yeast mutants temperature-sensitive for growth after random mutagenesis of the chromosomal RAS2 gene and deletion of the RAS1 gene. EMBO J 7:3375–83
84. Mistou MY, Jacquet E, Poullet P, et al (1992) Mutations of Ha-ras p21 that define important regions for the molecular mechanism of the SDC25 C-domain, a guanine nucleotide dissociation stimulator. EMBO J 11:2391–7
85. Howe LR, Marshall CJ (1993) Identification of amino acids in p21ras involved in exchange factor interaction. Oncogene 8:2583–90
86. Segal M, Willumsen BM, Levitzki A (1993) Residues crucial for Ras interaction with GDP-GTP exchangers. Proc Natl Acad Sci 90:5564–5568. https://doi.org/10.1073/pnas.90.12.5564
87. Mosteller RD, Han J, Broek D (1994) Identification of residues of the H-ras protein critical for functional interaction with guanine nucleotide exchange factors. Mol Cell Biol 14:1104–1112. https://doi.org/10.1128/MCB.14.2.1104
88. Park W, Mosteller RD, Broek D (1994) Amino acid residues in the CDC25 guanine nucleotide exchange factor critical for interaction with Ras. Mol Cell Biol 14:8117–8122. https://doi.org/10.1128/MCB.14.12.8117

89. Quilliam LA, Kato K, Rabun KM, et al (1994) Identification of residues critical for Ras(17N) growth-inhibitory phenotype and for Ras interaction with guanine nucleotide exchange factors. Mol Cell Biol 14:1113–1121. https://doi.org/10.1128/MCB.14.2.1113
90. Quilliam LA, Hisaka MM, Zhong S, et al (1996) Involvement of the Switch 2 Domain of Ras in Its Interaction with Guanine Nucleotide Exchange Factors. J Biol Chem 271:11076–11082. https://doi.org/10.1074/jbc.271.19.11076
91. Day G-J, Mosteller RD, Broek D (1998) Distinct Subclasses of Small GTPases Interact with Guanine Nucleotide Exchange Factors in a Similar Manner. Mol Cell Biol 18:7444–7454. https://doi.org/10.1128/MCB.18.12.7444
92. Hall BE, Yang SS, Boriack-Sjodin PA, et al (2001) Structure-based Mutagenesis Reveals Distinct Functions for Ras Switch 1 and Switch 2 in Sos-catalyzed Guanine Nucleotide Exchange. J Biol Chem 276:27629–27637. https://doi.org/10.1074/jbc.M101727200
93. O'Bryan JP (2019) Pharmacological targeting of RAS: Recent success with direct inhibitors. Pharmacol Res 139:503–511. https://doi.org/10.1016/j.phrs.2018.10.021
94. Thompson HJ, Jiang C, Lu J, et al (1997) Sulfone metabolite of sulindac inhibits mammary carcinogenesis. Cancer Res 57:267–71
95. Karaguni I-M, Glüsenkamp K-H, Langerak A, et al (2002) New indene-derivatives with anti-proliferative properties. Bioorg Med Chem Lett 12:709–713. https://doi.org/10.1016/S0960-894X(01)00839-3
96. Herrmann C, Block C, Geisen C, et al (1998) Sulindac sulfide inhibits Ras signaling. Oncogene 17:1769–1776. https://doi.org/10.1038/sj.onc.1202085
97. Kato-Stankiewicz J, Hakimi I, Zhi G, et al (2002) Inhibitors of Ras/Raf-1 interaction identified by two-hybrid screening revert Ras-dependent transformation phenotypes in human cancer cells. Proc Natl Acad Sci 99:14398–14403. https://doi.org/10.1073/pnas.222222699
98. Lu Y, Sakamuri S, Chen Q-Z, et al (2004) Solution phase parallel synthesis and evaluation of MAPK inhibitory activities of close structural analogues of a Ras pathway modulator. Bioorg Med Chem Lett 14:3957–3962. https://doi.org/10.1016/j.bmcl.2004.05.048
99. Skobeleva N, Menon S, Weber L, et al (2007) In vitro and in vivo synergy of MCP compounds with mitogen-activated protein kinase pathway- and microtubule-targeting inhibitors. Mol Cancer Ther 6:898–906. https://doi.org/10.1158/1535-7163.MCT-06-0602
100. González-Pérez V, Reiner DJ, Alan JK, et al (2010) Genetic and functional characterization of putative Ras/Raf interaction inhibitors in C. elegans and mammalian cells. J Mol Signal 5:2. https://doi.org/10.1186/1750-2187-5-2
101. Grant BJ, Lukman S, Hocker HJ, et al (2011) Novel Allosteric Sites on Ras for Lead Generation. PLoS One 6:e25711. https://doi.org/10.1371/journal.pone.0025711
102. Buhrman G, O'Connor C, Zerbe B, et al (2011) Analysis of Binding Site Hot Spots on the Surface of Ras GTPase. J Mol Biol 413:773–789. https://doi.org/10.1016/j.jmb.2011.09.011
103. Shima F, Yoshikawa Y, Ye M, et al (2013) In silico discovery of small-molecule Ras inhibitors that display antitumor activity by blocking the Ras-effector interaction. Proc Natl Acad Sci 110:8182–8187. https://doi.org/10.1073/pnas.1217730110
104. Papale A, Morella IM, Indrigo MT, et al (2016) Impairment of cocaine-mediated behaviours in mice by clinically relevant Ras-ERK inhibitors. Elife 5:. https://doi.org/10.7554/eLife.17111
105. Ramsey JD, Flynn NH (2015) Cell-penetrating peptides transport therapeutics into cells. Pharmacol Ther 154:78–86. https://doi.org/10.1016/j.pharmthera.2015.07.003
106. Raucher D, Ryu JS (2015) Cell-penetrating peptides: strategies for anticancer treatment. Trends Mol Med 21:560–570. https://doi.org/10.1016/j.molmed.2015.06.005
107. Gump JM, Dowdy SF (2007) TAT transduction: the molecular mechanism and therapeutic prospects. Trends Mol Med 13:443–448. https://doi.org/10.1016/j.molmed.2007.08.002
108. Brambilla R, Papale A, Fasano S (2012) WO/2012/016963 - Peptides for the treatment of brain diseases
109. Santini E, Valjent E, Usiello A, et al (2007) Critical Involvement of cAMP/DARPP-32 and Extracellular Signal-Regulated Protein Kinase Signaling in L-DOPA-Induced Dyskinesia. J Neurosci 27:6995–7005. https://doi.org/10.1523/JNEUROSCI.0852-07.2007
110. Santini E, Alcacer C, Cacciatore S, et al (2009) l-DOPA activates ERK signaling and phosphorylates histone H3 in the striatonigral medium spiny neurons of hemiparkinsonian mice. J Neurochem 108:621–633. https://doi.org/10.1111/j.1471-4159.2008.05831.x
111. Darmopil S, Martín AB, De Diego IR, et al (2009) Genetic Inactivation of Dopamine D1 but Not D2 Receptors Inhibits L-DOPA–Induced Dyskinesia and Histone Activation. Biol Psychiatry

66:603–613. https://doi.org/10.1016/j.biopsych.2009.04.025
112. Brambilla R, Papale A, Fasano S (2012) WO 2012/016963 Al. 2012
113. Maurer T, Garrenton LS, Oh A, et al (2012) Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. Proc Natl Acad Sci 109:5299–5304. https://doi.org/10.1073/pnas.1116510109
114. Patgiri A, Jochim AL, Arora PS (2008) A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in α-Helical Conformation. Acc Chem Res 41:1289–1300. https://doi.org/10.1021/ar700264k
115. Davis JM, Tsou LK, Hamilton AD (2007) Synthetic non-peptide mimetics of α-helices. Chem Soc Rev 36:326–334. https://doi.org/10.1039/B608043J
116. Cheng RP, Gellman SH, DeGrado WF (2001) β-Peptides: From Structure to Function. Chem Rev 101:3219–3232. https://doi.org/10.1021/cr000045i
117. Bautista AD, Craig CJ, Harker EA, Schepartz A (2007) Sophistication of foldamer form and function in vitro and in vivo. Curr Opin Chem Biol 11:685–692. https://doi.org/10.1016/j.cbpa.2007.09.009
118. Garner J, Harding MM (2007) Design and synthesis of α-helical peptides and mimetics. Org Biomol Chem 5:3577. https://doi.org/10.1039/b710425a
119. Goodman CM, Choi S, Shandler S, DeGrado WF (2007) Foldamers as versatile frameworks for the design and evolution of function. Nat Chem Biol 3:252–262. https://doi.org/10.1038/nchembio876
120. Murray JK, Gellman SH (2007) Targeting protein–protein interactions: Lessons from p53/MDM2. Biopolymers 88:657–686. https://doi.org/10.1002/bip.20741
121. Siedlecka M, Goch G, Ejchart A, et al (1999) α-Helix nucleation by a calcium-binding peptide loop. Proc Natl Acad Sci 96:903–908. https://doi.org/10.1073/pnas.96.3.903
122. Yang J, Zhao K, Gong Y, et al (1998) α-Helix Nucleation Constant in Copolypeptides of Alanine and Ornithine or Lysine. J Am Chem Soc 120:10646–10652. https://doi.org/10.1021/ja982319d
123. Wang D, Chen K, Kulp JL, Arora PS (2006) Evaluation of Biologically Relevant Short α-Helices Stabilized by a Main-Chain Hydrogen-Bond Surrogate. J Am Chem Soc 128:9248–9256. https://doi.org/10.1021/ja062710w
124. Kemp DS, Curran TP, Boyd JG, Allen TJ (1991) Studies of N-terminal templates for .alpha.-helix formation. Synthesis and conformational analysis of peptide conjugates of (2S,5S,8S,11S)-1-acetyl-1,4-diaza-3-keto-5-carboxy-10-thiatricyclo[2.8.1.04,8]tridecane (Ac-Hel1-OH). J Org Chem 56:6683–6697. https://doi.org/10.1021/jo00023a038
125. Cabezas E, Satterthwait AC (1999) The Hydrogen Bond Mimic Approach: Solid-Phase Synthesis of a Peptide Stabilized as an α-Helix with a Hydrazone Link. J Am Chem Soc 121:3862–3875. https://doi.org/10.1021/ja983212t
126. Pal S, Prabhakaran EN (2018) Hydrogen bond surrogate stabilized water soluble 310-helix from a disordered pentapeptide containing coded α-amino acids. Tetrahedron Lett 59:2515–2519. https://doi.org/10.1016/j.tetlet.2018.05.029
127. Schievano E, Bisello A, Chorev M, et al (2001) Aib-Rich Peptides Containing Lactam-Bridged Side Chains as Models of the 3 10 -Helix. J Am Chem Soc 123:2743–2751. https://doi.org/10.1021/ja0027261
128. Kortemme T, Kim DE, Baker D (2004) Computational Alanine Scanning of Protein-Protein Interfaces. Sci Signal 2004:pl2–pl2. https://doi.org/10.1126/stke.2192004pl2
129. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci 99:14116–14121. https://doi.org/10.1073/pnas.202485799
130. Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421
131. Protein BLAST Sequence Alignment. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Accessed 25 Aug 2020
132. (2006) SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Association for Computing Machinery, New York, NY, USA
133. Beard H, Cholleti A, Pearlman D, et al (2013) Applying Physics-Based Scoring to Calculate Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes. PLoS One 8:e82849. https://doi.org/10.1371/journal.pone.0082849
134. Protein Data Bank. https://pdb101.rcsb.org. Accessed 21 May 2020
135. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: parameters,

protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221–234. https://doi.org/10.1007/s10822-013-9644-8
136. Greenwood JR, Calkins D, Sullivan AP, Shelley JC (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comput Aided Mol Des 24:591–604. https://doi.org/10.1007/s10822-010-9349-1
137. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
138. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864

CHAPTER EIGHT

Conclusions

In the past decades, targeting protein-protein interactions (PPI) was considered a hardly challenging strategy for a drug discovery programme. Thus PPIs were often assumed as undruggable targets [1,2] because of hard problems associated with them, such as the lack of well-defined and deep pockets [3,4], shallow and flat interfaces, discontinuous interacting epitopes dislocated over the protein surfaces [5] and too wide binding regions (1200-2000 $Å^2$) [6].

Nonetheless, in recent years the Medicinal Chemistry frontiers have been changing and PPIs have gained popularity amongst the scientific community due to their key roles in such a huge number of diseases associated with protein-protein deregulations, in a wide range of medical areas, such as oncology [7–11], cancer immunotherapy [12], tropical infectious diseases [13], neurological disorders [14], heart failure [15], inflammation and oxidative stress [16].

It is believed that the overall biological complexity in higher organisms, especially in humans, is due not only to a relatively greater number of genes but particularly because human proteins generate more intricate protein networks [17,18]. Indeed, about 650,000 PPIs have been identified within the human organism [19,20], whereas only 20,000 protein genes code them [16]. A single organism may include more than one interactome representing different tissue types, biological states, etc. Thanks to recent advances in the field of genomics and proteomics and the development of large-scale high-throughput experiments, some of these networks have been discovered and characterised, resulting in the production of a large volume of data which has aided in the uncovering of PPIs [21]. The complete elucidation of all interaction networks within the cell may improve our understanding of numerous diseases, providing key information for the development of novel therapeutics with significant implications for science [22]. Furthermore, conserved protein interactomes across organisms have meant important findings associated with their evolutionary relationships, providing insights into their previously unknown dependencies. Therefore, the elucidation of PPIs especially within the interaction networks is a hot point in biological research, and it may lead to enhanced approaches for drug discovery [23,24].

Until recently, PPIs were determined by experimental evidence through techniques specifically developed to target a small group of interactions. However, recent genomic techniques have allowed to carry out high-throughput experiments, which to date may exhaustively explore all possible interactions of an entire genome. The model organism of election used for functional proteomics characterisation is the *Saccharomyces cerevisiae*, also referred to as baker's yeast, whose genomic sequence was totally elucidated in 1996 [25]. This discovery has allowed determining whole PPI maps applying several methods including yeast two-hybrid [26,27], such as affinity purification/mass spectrometric identification methods an others [28–31], and indirect large-scale approaches, like synthetic lethal analysis [32] and correlated mRNA expression profile [33].

However, these methods present several limitations in terms of high costs and labour- and time-wasting. Moreover, when comparing data generated from different

small- or large-scale studies, such as high-confidence experiments or high-throughput studies, significant interstudy discrepancies have been detected [23,34]. Furthermore, the data from biological experiments often may include false positives, where proteins are shown erroneously or not correctly correlated with each other [23,34]. Therefore, further confirmations of the interactions are often useful in applying other methods.

A large number of computational methods have been successfully applied to evaluate, validate, and deeply analyse the experimentally determined protein interactomes. In this context, a high number of computational tools and techniques have been developed, such as methods developed to construct interaction databases, automated data mining techniques to extract relevant information about potential interactions from PPI databases [23], quantum mechanics and molecular mechanics (QM/MM) to study the electronic properties, simulate chemical reactions, and calculate spectra [35–37], and all-atom, united-atom and coarse-grained MD simulations [38–43] including explicit/implicit solvent models to simulate temporal and spatial scales of inter- and intramolecular interactions [44–46]. These techniques have allowed to explore protein interaction networks and the related functional features and predict novel PPIs by learning from known interactions [21].

This PhD thesis has reported the use of some computational techniques as valuable tools to explore protein-protein interfaces, identifying their hot spot residues, selecting small molecules and designing peptides with the aim of inhibiting the studied PPIs.

Indeed, a success story of *In Silico* approaches to PPI study has been reported in Chapter 2, where MD simulations, docking and pharmacophore screenings led to the identification of a set of PPI modulators. Among these, two molecules, RIM430 and RIM442, registered good inhibitory activity with IC_{50} values even within the nanomolar range against the interaction between MUC1 and CIN85 proteins in cancer disease.

Chapter 3 describes how the computational techniques herein used are crucial for the generation and rationalisation of three interaction models of $NLRP3_{PYD}$-ASC_{PYD} complex, a host-guest system not available as solved structure and otherwise not accessible for drug design [47]. Based on these protein-protein models it was possible to identify and select potential inhibitors of the analysed PPI involved in inflammatory diseases.

Chapter 4 reports a study where computational tools allowed to identify a potentially druggable region on the surface of SARS-CoV-2 Spike protein, the N-terminal portion, registering the highest number of calculated hot spots compared to the other interacting regions over the protein surface, and highlighting a small binding pocket able to accommodate functional groups of the ligands. Therefore, based on these insights it was possible to identify potential inhibitors of the interaction between Spike protein and the host ACE2 receptor that will be assayed in the cellular environment.

In Chapter 5, MD simulations were exploited to investigate intramolecular modifications as a consequence of a point mutation on C3b protein (R102G), by exploring variations in the stability of the key interactions between C3b and FH protein in patients affected by AMD disease.

Finally, Chapter 6 and 7 describe two similar computational methodologies, based essentially on computational alanine scanning and MD simulations, to design peptides. For HOX-PBX complex non-standard amino acids were employed to optimise a patented peptide structure by including non-natural amino acids potentially able to inhibit HOX-

PBX cooperative binding to DNA against several forms of cancer. For Ras-RasGRF1 complex, the computational techniques shed light on a group of designed helix-shaped peptides embedded with the hydrogen bond surrogate approach between the amino acid *i* and the amino acid *i+3* for targeting and inhibiting the studied PPI to tackle cocaine abuse relapses.

Although all the herein exploited techniques are based on predictive calculations and need experimental evidence to confirm the findings, the results and molecular insights retrieved and collected show the potential of this field of applications in Medicinal Chemistry, guaranteeing labour- and time-saving to the research groups. On the other hand, computing ability, improved algorithms and fast-growing data sets are rapidly fostering advances in multiscale molecular modelling, providing a powerful emerging paradigm for drug discovery. It means that more and more research efforts will be done to invest in novel and more precise computational techniques and fine-tune the currently employed methodologies [46].

References – Chapter Eight

1. Ran, X.; Gestwicki, J.E. Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. *Curr. Opin. Chem. Biol.* **2018**, *44*, 75–86.
2. Perricone, U.; Gulotta, M.R.; Lombino, J.; Parrino, B.; Cascioferro, S.; Diana, P.; Cirrincione, G.; Padova, A. An overview of recent molecular dynamics applications as medicinal chemistry tools for the undruggable site challenge. *Medchemcomm* **2018**, *9*, 920–936.
3. McCormick, F. KRAS as a Therapeutic Target. *Clin. Cancer Res.* **2015**, *21*, 1797–801.
4. Whitfield, J.R.; Beaulieu, M.-E.; Soucek, L. Strategies to Inhibit Myc and Their Clinical Applicability. *Front. cell Dev. Biol.* **2017**, *5*, 10.
5. Fernandez, A.; Scheraga, H.A. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci.* **2003**, *100*, 113–118.
6. Horton, N.; Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1992**, *1*, 169–181.
7. Arrowsmith, C.H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **2012**, *11*, 384–400.
8. Nero, T.L.; Morton, C.J.; Holien, J.K.; Wielens, J.; Parker, M.W. Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* **2014**, *14*, 248–262.
9. Arkin, M.R.; Whitty, A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Curr. Opin. Chem. Biol.* **2009**, *13*, 284–290.
10. Arkin, M.R.; Tang, Y.; Wells, J.A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chem. Biol.* **2014**, *21*, 1102–1114.
11. Li, Z.; Ivanov, A.A.; Su, R.; Gonzalez-Pecchi, V.; Qi, Q.; Liu, S.; Webber, P.; McMillan, E.; Rusnak, L.; Pham, C.; et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. *Nat. Commun.* **2017**, *8*, 14356.
12. Li, Q.; Quan, L.; Lyu, J.; He, Z.; Wang, X.; Meng, J.; Zhao, Z.; Zhu, L.; Liu, X.; Li, H. Discovery of peptide inhibitors targeting human programmed death 1 (PD-1) receptor. *Oncotarget* **2016**, *7*, 64967–64976.
13. Dawidowski, M.; Emmanouilidis, L.; Kalel, V.C.; Tripsianes, K.; Schorpp, K.; Hadian, K.; Kaiser, M.; Mäser, P.; Kolonko, M.; Tanghe, S.; et al. Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. *Science (80-.).* **2017**, *355*, 1416–1420.
14. Hayes, M.P.; Soto-Velasquez, M.; Fowler, C.A.; Watts, V.J.; Roman, D.L. Identification of FDA-Approved Small Molecules Capable of Disrupting the Calmodulin–Adenylyl Cyclase 8 Interaction through Direct Binding to Calmodulin. *ACS Chem. Neurosci.* **2018**, *9*, 346–357.
15. Anand, P.; Brown, J.D.; Lin, C.Y.; Qi, J.; Zhang, R.; Artero, P.C.; Alaiti, M.A.; Bullard, J.; Alazem, K.; Margulies, K.B.; et al. BET Bromodomains Mediate Transcriptional Pause Release in Heart Failure. *Cell* **2013**, *154*, 569–582.
16. Lu, M.-C.; Tan, S.-J.; Ji, J.-A.; Chen, Z.-Y.; Yuan, Z.-W.; You, Q.-D.; Jiang, Z.-Y. Polar Recognition Group Study of Keap1-Nrf2 Protein–Protein Interaction Inhibitors. *ACS Med. Chem. Lett.* **2016**, *7*, 835–840.
17. Alm, E. Biological networks. *Curr. Opin. Struct. Biol.* **2003**, *13*, 193–202.
18. Claverie, J.-M. What If There Are Only 30,000 Human Genes? *Science (80-.).* **2001**, *291*, 1255–1257.
19. Venkatesan, K.; Rual, J.-F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K.-I.; et al. An empirical framework for binary interactome mapping. *Nat. Methods* **2009**, *6*, 83–90.
20. Stumpf, M.P.H.; Thorne, T.; de Silva, E.; Stewart, R.; An, H.J.; Lappe, M.; Wiuf, C. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci.* **2008**, *105*, 6959–6964.
21. Pitre, S.; Alamgir, M.; Green, J.R.; Dumontier, M.; Dehne, F.; Golshani, A. Computational Methods For Predicting Protein–Protein Interactions. In; 2008; pp. 247–267.
22. Price, N.D.; Papin, J.A.; Schilling, C.H.; Palsson, B.O. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol.* **2003**, *21*, 162–169.
23. Franzot, G.; Carugo, O. Computational approaches to protein-protein interaction. *J. Struct. Funct. Genomics* **2003**, *4*, 245–255.
24. Salwinski, L.; Eisenberg, D. Computational methods of analysis of protein–protein interactions. *Curr. Opin. Struct. Biol.* **2003**, *13*, 377–382.
25. Goffeau, A.; Barrell, B.G.; Bussey, H.; Davis, R.W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J.D.; Jacq, C.; Johnston, M.; et al. Life with 6000 Genes. *Science (80-.).* **1996**, *274*,

546–567.
26. Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T.A.; Judson, R.S.; Knight, J.R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; et al. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* **2000**, *403*, 623–627.
27. Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M.; Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.* **2001**, *98*, 4569–4574.
28. Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G.D.; Moore, L.; Adams, S.-L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002**, *415*, 180–183.
29. Mann, M.; Pandey, A. Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends Biochem. Sci.* **2001**, *26*, 54–61.
30. Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; et al. Global Analysis of Protein Activities Using Proteome Chips. *Science (80-.).* **2001**, *293*, 2101–2105.
31. Tong, A.H.Y.; Drees, B.; Nardelli, G.; Bader, G.D.; Brannetti, B.; Castagnoli, L.; Evangelista, M.; Ferracuti, S.; Nelson, B.; Paoluzi, S.; et al. A Combined Experimental and Computational Strategy to Define Protein Interaction Networks for Peptide Recognition Modules. *Science (80-.).* **2002**, *295*, 321–324.
32. Tong, A.H.Y.; Evangelista, M.; Parsons, A.B.; Xu, H.; Bader, G.D.; Pagé, N.; Robinson, M.; Raghibizadeh, S.; Hogue, C.W.; Bussey, H.; et al. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science (80-.).* **2001**, *294*, 2364–2368.
33. Ge, H.; Liu, Z.; Church, G.M.; Vidal, M. Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. *Nat. Genet.* **2001**, *29*, 482–486.
34. Aloy, P.; Russell, R.B. The third dimension for protein interactions and complexes. *Trends Biochem. Sci.* **2002**, *27*, 633–638.
35. Fan, J.; Lin, L.; Wang, C.-K. Excited state properties of non-doped thermally activated delayed fluorescence emitters with aggregation-induced emission: a QM/MM study. *J. Mater. Chem. C* **2017**, *5*, 8390–8399.
36. Jindal, G.; Warshel, A. Exploring the Dependence of QM/MM Calculations of Enzyme Catalysis on the Size of the QM Region. *J. Phys. Chem. B* **2016**, *120*, 9913–9921.
37. Morzan, U.N.; Alonso de Armiño, D.J.; Foglia, N.O.; Ramírez, F.; González Lebrero, M.C.; Scherlis, D.A.; Estrin, D.A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.
38. Buchete, N.-V.; Hummer, G. Peptide folding kinetics from replica exchange molecular dynamics. *Phys. Rev. E* **2008**, *77*, 030902.
39. Liu, Y.; Strümpfer, J.; Freddolino, P.L.; Gruebele, M.; Schulten, K. Structural Characterization of λ-Repressor Folding from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2012**, *3*, 1117–1123.
40. Sothiselvam, S.; Liu, B.; Han, W.; Ramu, H.; Klepacki, D.; Atkinson, G.C.; Brauer, A.; Remm, M.; Tenson, T.; Schulten, K.; et al. Macrolide antibiotics allosterically predispose the ribosome for translation arrest. *Proc. Natl. Acad. Sci.* **2014**, *111*, 9804–9809.
41. Hernández-Rodríguez, M.; C. Rosales-Hernández, M.; E. Mendieta-Wejebe, J.; Martínez-Archundia, M.; Correa Basurto, J. Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design. *Curr. Med. Chem.* **2016**, *23*, 3909–3924.
42. Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* **2015**, *48*, 3026–3035.
43. Mortier, J.; Rakers, C.; Bermudez, M.; Murgueitio, M.S.; Riniker, S.; Wolber, G. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discov. Today* **2015**, *20*, 686–702.
44. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.
45. Borhani, D.W.; Shaw, D.E. The future of molecular dynamics simulations in drug discovery. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 15–26.
46. Lin, X.; Li, X.; Lin, X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020**, *25*, 1375.
47. Protein Data Bank Available online: https://pdb101.rcsb.org (accessed on May 21, 2020).

LIST OF PUBLICATIONS

REVIEW ARTICLE published in *Medicinal Chemistry Communications* Journal
DOI: 10.1039/C8MD00166A
Date of publication: 19 April 2018

TITLE
An overview of recent Molecular Dynamics applications as medicinal chemistry tool for undruggable sites challenge.

AUTHORS
Maria Rita Gulotta, Ugo Perricone, Jessica Lombino, Barbara Parrino, Stella Cascioferro, Patrizia Diana, Girolamo Cirrincione and Alessandro Padova

ABSTRACT

Molecular Dynamics (MD) has become increasingly popular due to the development of hardware and software solutions and improvement in algorithms, that allowed researchers to scale up calculations in order to speed up them. MD simulations are usually used to address protein folding issues or protein-ligand complex stability through energy profile analysis over time. In recent years, the development of new tools able to deeply explore Potential Energy Surface (PES) allowed researchers to focus on the dynamic nature of binding recognition process and binding-induced protein conformational change. Moreover, modern approaches have demonstrated to be effective and reliable in calculating some kinetic and thermodynamic parameters behind the host-guest recognition process. Starting from all of these considerations, several efforts have been made in order to integrate MD within the virtual screening process in drug discovery. Knowledge retrieved from MD can be, in fact, exploited as a starting point to build pharmacophores or docking constraints in the early stage of the screening campaign as well as to define key features, in order to unravel hidden binding modes and help the optimisation of the molecular structure of a lead compound. Based on these outcomes, researchers are nowadays using MD as an invaluable tool to discover and target previously considered undruggable binding sites, including protein-protein interactions and allosteric sites on protein surface. As a matter of fact, the use of MD has been recognised as vital in the discovery of selective protein-protein interaction modulators. The use of a dynamic overview on how the host-guest recognition occurs and of the relative conformational modifications induced, allow researchers to optimise small molecules and small peptides capable to tightly interact within the cleft between the two proteins.

In this review we point to present the most recent applications of MD as integrated tool to be used in the rational design of small molecules or small peptides able to modulate undruggable targets, such as allosteric sites and protein-protein interactions.

RESEARCH ARTICLE published in the *International Journal of Molecular Sciences*
DOI: 10.3390/ijms20204974
Date of publication: 9 October 2019

TITLE
In Silico Insights towards the Identification of NLRP3 Druggable Hot Spots

AUTHORS
Nedra Mekni, Maria De Rosa, Chiara Cipollina, **Maria Rita Gulotta**, Giada De Simone, Jessica Lombino, Alessandro Padova and Ugo Perricone

ABSTRACT

NLRP3 (NOD-like receptor family, pyrin domain-containing protein 3) activation has been linked to several chronic pathologies, including atherosclerosis, type-II diabetes, fibrosis, rheumatoid arthritis, and Alzheimer's disease. Therefore, NLRP3 represents an appealing target for the development of innovative therapeutic approaches. A few companies are currently working on the discovery of selective modulators of NLRP3 inflammasome. Unfortunately, limited structural data are available for this target. To date, MCC950 represents one of the most promising noncovalent NLRP3 inhibitors. Recently, a possible region for the binding of MCC950 to the NLRP3 protein was described but no details were disclosed regarding the key interactions. In this communication, we present an in silico multiple approach as an insight useful for the design of novel NLRP3 inhibitors. In detail, combining different computational techniques, we propose consensus-retrieved protein residues that seem to be essential for the binding process and for the stabilization of the protein–ligand complex.

RESEARCH ARTICLE published in *ChemMedChem* Journal
DOI: 10.1002/cmdc.202000259
Date of publication: 23 July 2020

TITLE
Targeting SARS-CoV-2 RBD interface: a supervised computational data-driven approach to identify potential modulators

AUTHORS
Maria Rita Gulotta, Jessica Lombino, Ugo Perricone, Giada De Simone, Nedra Mekni, Maria De Rosa, Patrizia Diana and Alessandro Padova

ABSTRACT

Coronavirus Disease 2019 (COVID-19) has spread out as a pandemic threat affecting over 2 million people. The infectious process initiates via binding of SARS-CoV-2 Spike (S) glycoprotein to host Angiotensin-converting enzyme 2 (ACE2). The interaction is mediated by the receptor-binding domain (RBD) of S glycoprotein, promoting host receptor recognition and binding to ACE2 peptidase domain (PD), thus representing a promising target for therapeutic intervention. Herein, we present a computational study aimed at identifying small molecules potentially able to target RBD. Although targeting PPI remains a challenge in drug discovery, our investigation highlights that interaction between SARS-CoV-2 RBD and ACE2 PD might be prone to small molecule modulation, due to the hydrophilic nature of the bi-molecular recognition process and the presence of druggable hot spots. The fundamental objective is to identify, and provide to the international scientific community, hit molecules potentially suitable to enter the drug discovery process, preclinical validation and development.

RESEARCH ARTICLE published in the *Molecular Informatics* Journal
DOI: 10.1002/minf.202000148
Date of publication: 24 August 2020

TITLE

Dynamic-shared pharmacophore approach as tool to design new allosteric PRC2 inhibitors, targeting EED binding pocket.

AUTHORS

Jessica Lombino, **Maria Rita Gulotta**, Giada De Simone, Nedra Mekni, Maria De Rosa, Daniela Carbone, Barbara Parrino, Stella Maria Cascioferro, Patrizia Diana, Alessandro Padova and Ugo Perricone

ABSTRACT

The Polycomb Repressive complex 2 (PRC2) maintains a repressive chromatin state and silences many genes, acting as methylase on histone tails. This enzyme was found overexpressed in many types of cancer. In this work, we have set up a Computer-Aided Drug Design approach based on the allosteric modulation of PRC2. In order to minimize the possible bias derived from using a single set of coordinates within the protein-ligand complex, a dynamic workflow was developed. In details, molecular dynamic was used as tool to identify the most significant ligand-protein interactions from several crystallized protein structures. The identified features were used for the creation of dynamic pharmacophore models and docking grid constraints for the design of new PRC2 allosteric modulators. Our protocol was retrospectively validated using a dataset of active and inactive compounds, and the results were compared to the classic approaches, through ROC curves and enrichment factor. Our approach suggested some important interaction features to be adopted for virtual screening performance improvement.

RESEARCH ARTICLE published in the *Computation* Journal
DOI: 10.3390/computation8030077
Date of publication: 29 August 2020

TITLE
Exploring the SARS-CoV-2 Proteome in the Search of Potential Inhibitors via Structure-Based Pharmacophore Modeling/Docking Approach

AUTHORS
Giulia Culletta, **Maria Rita Gulotta**, Ugo Perricone, Maria Zappalà, Anna Maria Almerico and Marco Tutone

ABSTRACT

To date, SARS-CoV-2 infectious disease, named COVID-19 by the World Health Organization (WHO) in February 2020, has caused millions of infections and hundreds of thousands of deaths. Despite the scientific community efforts, there are currently no approved therapies for treating this coronavirus infection. The process of new drug development is expensive and time-consuming, so that drug repurposing may be the ideal solution to fight the pandemic. In this paper, we selected the proteins encoded by SARS-CoV-2 and using homology modeling we identified the high-quality model of proteins. A structure-based pharmacophore modeling study was performed to identify the pharmacophore features for each target. The pharmacophore models were then used to perform a virtual screening against the DrugBank library (investigational, approved and experimental drugs). Potential inhibitors were identified for each target using XP docking and induced fit docking. MM-GBSA was also performed to better prioritize potential inhibitors. This study will provide new important comprehension of the crucial binding hot spots usable for further studies on COVID-19. Our results can be used to guide supervised virtual screening of large commercially available libraries.

LIST OF POSTERS

- **A computational study to explore the molecular mechanisms behind the antiproliferative activity of Nortopsentin derivatives**
  M.R. Gulotta, J. Lombino, B. Parrino, S.M. Cascioferro, P. Diana, A. Padova, G. Cirrincione, U. Perricone; presented at *MedChemSicily 2018*, 17-20 July 2018, Palermo, Italy

- **PRC2 allosteric modulation. An alternative strategy in drug discovery for epigenetic diseases**
  J. Lombino, S. Cascio, **M.R. Gulotta**, U. Perricone, S.M. Cascioferro, B. Parrino, P. Diana, A. Padova; presented at *MedChemSicily 2018*, 17-20 July 2018, Palermo, Italy

- **A computational structure-based approach to address hit identification modulating TOM20/α-Synuclein Interaction**
  M.R. Gulotta, R. Di Maio, J. Lombino, G. De Simone, A. Padova, G. Cirrincione, U. Perricone; presented at *EMBO Practical Course* "Computational analysis of protein-protein interactions: Sequences, networks and diseases", 5-10 November 2018, Rome, Italy

- **A combined pharmacophore and MD approach to design putative modulators of a protein-protein interaction involved in Parkinson's Disease**
  M.R. Gulotta, R. Di Maio, N. Mekni, J. Lombino, G. De Simone, P. Diana, U. Perricone; presented at *XII European Workshop in Drug Design 2019*, 19-24 May 2019, Siena, Italy

- **A Molecular Dynamics-based virtual screening framework for the identification of EED inhibitors in the PRC2 modulation**
  J. Lombino, **M.R. Gulotta**, U. Perricone, P. Diana, A. Padova; presented at *XII European Workshop in Drug Design 2019*, 19-24 May 2019, Siena, Italy

- **A structure-based computational approach to design protein-protein inhibitors for Parkinson's Disease**
  M.R. Gulotta, N. Mekni, J. Lombino, M. De Rosa, U. Perricone; presented at *Ri.MED Research Retreat 2019*, 23 October 2019, Palermo, Italy

- **Towards the Discovery of NLRP3 Protein Druggable Hot Spots**
  N. Mekni, **M.R. Gulotta**, M. De Rosa, U. Perricone; presented at *Ri.MED Research Retreat 2019*, 23 October 2019, Palermo, Italy

- **The Ri.MED compounds database: A shared library of workflows for compound management and experimental data analysis.**
  G. De Simone, **M.R. Gulotta**, M. De Rosa, U. Perricone, C. Coronnello; presented at *Ri.MED Research Retreat 2019*, 23 October 2019, Palermo, Italy