Università degli Studi di Messina

Dipartimento di Scienze Chimiche, Biologiche, Farmaceutiche e Ambientali

Dottorato di Ricerca in
"Biologia Applicata e Medicina Sperimentale"
XXXIII Ciclo
Curriculum: Scienze del Farmaco
SSD CHIM/08

STUDY AND DESIGN OF MODULATORS OF PROTEIN-PROTEIN INTERACTIONS IMPLICATED IN TUMOR AND NEURODEGENERATIVE DISEASES

Doctoral thesis:

Serena Vittorio

Supervisor:

Prof. Laura De Luca

Coordinator of the Ph.D. course:

Prof. Nunziacarla Spanò

Ph.D. course 2017-2020

Index

Abbreviation List

PPI	Protein-protein interaction
α-syn	α-synuclein
PD	Parkinson's disease
VS	Virtual screening
HTS	High throughput screening
BBB	Blood-brain-barrier
NMR	Nuclear magnetic resonance
MD	Molecular dynamics
SBDD	Structure-based drug design
LBDD	Ligand-based drug design
QSAR	Quantitative Structure - Activity Relationship
MSA	Multiple sequence alignment
FFT	Fast-Fourier transform
CAPRI	Critical Assessment of Predicted Interactions
DUD-E	Directory of Useful Decoys-Enhanced
AUC	Area under the curve
ROC	Receiver Operating Characteristic
MACCS	Molecular ACCess System
ECFP4	Extended connectivity fingerprint with bond diameter four
Tc	Tanimoto coefficient
MM	Molecular mechanics
FF	Force field
CNS	Central nervous system
DA	Dopamine
MAO-B	Monamino oxidase-B
COMT	Catechol-o-methyltransferase
HSP	Heat shock proteins
PCA	Protocatechuic acid
GA	Gallic acid
HCA	Hydroxycinnamic acid

PcTs	Phthalocyanine tetrasulfonate
DOX	Doxycycline
CMTs	Chemically modified tetracyclines
SC-D	SynuClean-D
TS	Training set
TEM	Transmission electron microscopy
MUC1	Mucin 1
SEA	Sea urchin sperm protein enterokinase and agrin
MUC1-N	N-terminal domain MUC1
MUC1-C	C-terminal domain MUC1
VNTR	Variable number of tandem repeats
GalNAc	N-acetylglucosamine
STn	Sialyl-Tn antigen
ECD	Extracellular domain
TMD	Transmembrane domain
CT	Cytoplasmatic tail
CIN85	Cbl-interacting protein of 85 kDa
SH3	Src Homology 3
SHIP2	SH2-containing inositol phosphatase 1
CMS	Clathrin-mediated endocytic
RTKs	Tyrosine kinases receptors
APCs	Antigen-presenting cells
NK	Natural killer
RMSD	Root mean square deviations
ITC	Isothermal titration calorimetry

Preface

The PhD research project was focused on the study of two different protein-protein interactions (PPIs) involved in Parkinson's disease and cancer progression.

This doctoral dissertation is divided in three sections.

Section 1 provides a general introduction about PPIs as biological targets for the development of new therapeutic agents, presenting the challenges and the recent advances in this research field (***Chapter 1***). Moreover, in the same section, the computational methods employed in the drug discovery process are discussed (***Chapter 2***).

Section 2 is focused on the identification of novel α-synuclein (α-syn) aggregation inhibitors as potential therapeutics for the cure of Parkinson's disease (PD). Firstly, in ***Chapter 3*** I described the role of α-syn in the pathogenesis of PD and the growing interest in the modulation of its amyloidogenic aggregation as viable strategy to halt the neurodegenerative process. In the same chapter, an overview of the different approaches used to target α-syn are illustrated. In ***Chapter 4***, the discovery of a novel class of α-syn aggregation inhibitors by combining pharmacophore-based virtual screening (VS) and synthetic approaches is reported. To achieve this scope, two collaborations with Professor Rosaria Gitto (University of Messina) and Professor Salvador Ventura (University of Barcelona) allowed us to respectively synthesize and test the designed compounds. Instead, in ***Chapter 5*** the use of a similarity-based VS workflow to identify new modulators of α-syn amyloid aggregation is illustrated.

Section 3 furnished new structural insights about MUC1-CIN85 PPI which is involved in the formation of tumor metastasis thus paving the way for the development of new anticancer agents. Specifically, **Chapter *6*** introduces the two protein partners, describing their structural features, activity and their role in the invasiveness of cancer cells. ***Chapter 7*** reports the research activity performed during my six-months internship at the Pharmaceutical Chemistry Department of the University of Vienna under the supervision of Professor Thierry Langer. It concerns the use of *in silico* methods to identify druggable ligand-binding sites on CIN85 SH3 domain involved in the binding with MUC1. Finally, in ***Chapter 8***, I documented the work accomplished during my six-months research experience at

Ri.MED Foundation of Palermo, under the supervision of Dr. Ugo Perricone. This study regards the exploration of the molecular contacts of MUC1 at CIN85 interface thus unveiling useful hints for the structure-based design of MUC1-CIN85 modulators.

Aim of the work

PPIs regulate many biological processes providing new targets for pharmacological intervention. Despite the challenges associated with the modulation of PPIs by small molecules, research efforts led to the discovery of PPIs inhibitors and some of them were approved for clinical usage. In this scenario, the research work performed during my PhD was focused on the study of two PPI interactions, α-syn aggregation and MUC1-CIN85, involved respectively in neurological disorders and cancer progression.

The amyloid aggregation of α-syn plays a pivotal role in the pathogenesis of PD. To date, the therapies available for the treatment of PD are addressed to reduce the related motor symptoms; therefore, there is an increasing interest in the development of effective therapeutic agents. The inhibition of α-syn aggregation by small molecules has emerged as promising new disease-modifying strategy to slow or block the neurodegenerative process, thus offering new opportunities for drug discovery. In the last decades, several inhibitors have been reported in literature some of which are under clinical investigation. The aim of my research was to design novel α-syn aggregation inhibitors by using *in silico* methods.

Despite extraordinary progresses have been made, the cure for most cancers is still a long way from the reality. This prompted the scientists to deeply investigate the biological pathways involved in malignancy thus allowing to identify novel targets of pharmaceutical interest. Recently, MUC1-CIN85 PPI, implicated in metastasis formation, has been validated as new target for the development of anticancer drugs. In this context, the scope of my research was the application of different structure-based approaches to gain useful structural insights for the design of MUC1-CIN85 PPI inhibitors.

Keywords: Protein-Protein interactions; α-synuclein; MUC1-CIN85; computer aided drug discovery; Parkinson's disease; cancer.

Section 1

Chapter 1

Targeting protein-protein interactions for the development of new therapeutic agents

1.1 Protein-protein interactions: challenges and opportunities

To date, the drug discovery process is focused on the identification of new therapeutics able to selectively bind a target macromolecule modulating its biological activity [1]. In this scenario, protein-protein interactions (PPIs) have been widely recognized as valuable tools for the treatment of various illnesses including cancer, neurodegenerative diseases, inflammation, viral and bacterial infections [2, 3]. PPIs are defined as physical contacts established between proteins that occur *in vivo* in a cell or in a living organism [4]. They play a pivotal role in the regulation of many cellular processes such as cellular growth, DNA replication, transcriptional activation, translation and transmembrane signal transduction [5]. PPIs are often dysregulated in disease thus offering new targets for pharmacological intervention [6]. It has been estimated that in human body, the number of PPIs lies between 130000 and 650000 [7] but only a small percentage of these (~2%) have been targeted [1]. The modulation of PPIs by small molecules presents several challenges related to the contact surfaces involved in the interaction between the proteins, which are larger (1500-3000 $Å^2$) than those of traditional targets like enzymes active sites (300-500 $Å^2$) [8]. Furthermore, the contact surfaces are usually flat and lack the grooves and the pockets that could accommodate a small molecule. Differently from the traditional targets such as enzymes and protein G coupled receptors, PPIs do not have natural small molecules binders or substrates that could serve as starting point for the design of new ligands [9]. Another issue concerning the druggability of PPIs is that some of them occur via phosphorylation. In these systems, the protein partner binds only if a key residue in its binding epitope is phosphorylated, therefore, they do not represent ideal targets since a small molecule mimicking the charged species of a phosphorylated site would be too polar with a poor bioavailability [10].

Considering all the above reported aspects, PPIs have been considered as "undruggable" targets for decades [1]. Nevertheless, in 90's, small molecules modulators of PPIs such as Tirofiban and taxanes were approved for clinical usage [8, 11, 12, 13]. As reported in Figure 1, over the years we assisted to a growing interest in the development of new PPI modulators and the number of PPI inhibitors approved or that have reached clinical trials has increased [8].

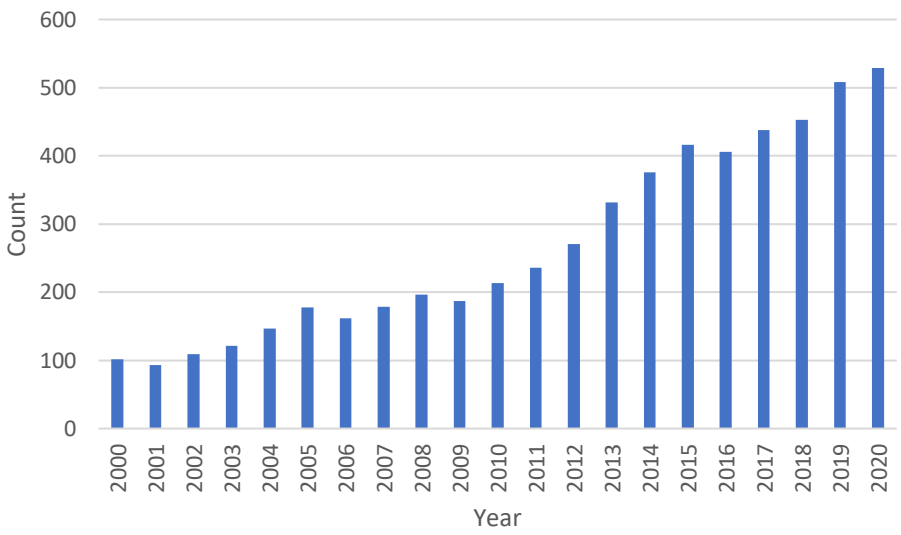


Figure 1. Published papers about PPIs inhibitors in the latest 20 years according to PubMed (https://pubmed.ncbi.nlm.nih.gov/).

The advances in this research field are especially due to some observations related to the characteristics of the interface between the two protein partners. Indeed, despite the large binding surface, mutational studies demonstrated that only a subset of the residues involved in the interaction accounts for most of the binding free energy. These residues, named "hotspots", represent less than half of the contact surface and they could be targeted by a low molecular weight compound [9,14]. An amino-acid residue is defined as hotspot if its mutation to alanine, by alanine scanning mutagenesis, leads to an increase of the binding free energy greater than 2 kcal mol^{-1} [14]. Therefore, a small molecule able to interact with hotspots residues might prevent the interaction between proteins (Figure2) because most of the binding energy is concentrated in this area [15].

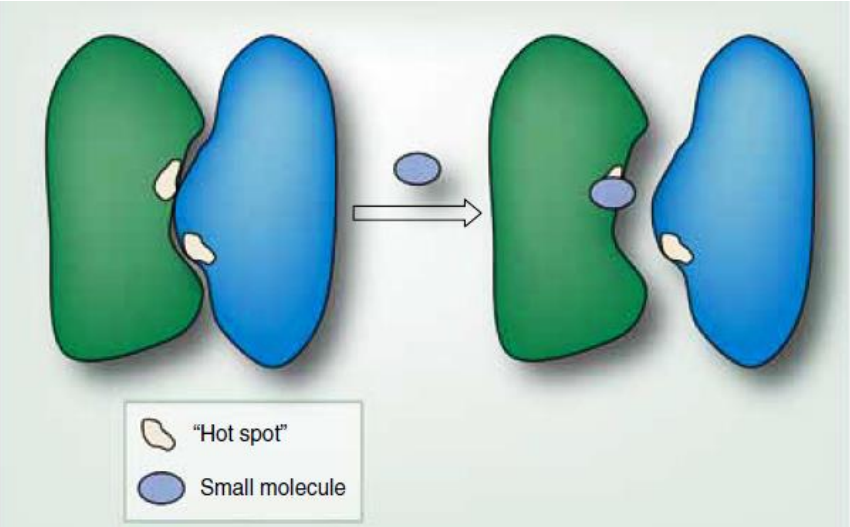


Figure2. Targeting hot spots to modulated PPI. This figure is reproduced from Zinzalla G. *et al.* [5].

The hotspot region is usually divided in a more accessible rim region and a more buried core region. The rim, characterised by an amino acid composition similar to the rest of the protein surface, surrounds the core which contains aromatic residues [16]. The hotspot amino acids found more frequently in PPIs are Trp, Tyr and Arg. In particular, Tyr and Trp can engage stacking and hydrophobic interactions due to their aromatic side chain allowing at the same time the possibility to form hydrogen bonds through the hydroxyl group and the indole nitrogen. Instead, the Arg residue, being a polar amino acid carrying a charged guanidium group, can establish hydrogen bonds and salt bridges [14]. The aminoacidic composition of hotspots reflect the fact that PPIs are principally driven by hydrophobic interactions, with hydrogen bonds and electrostatic interactions having a supporting but crucial role [5]. Sequential and structural analysis revealed that hotspot residues are conserved and clustered in tightly packed regions that are distributed not homogenously over the surface [17, 18]. These multiple clusters are involved in the interaction with the protein partner and their contribution to the stability of the complex is additive [18, 19]. *In vivo*, PPIs can form obligate or non-obligate complexes. In obligate PPIs, the protomers are not stable their own *in vivo*, while in non- obligate-complexes the protein partners may exist in the bound or dissociated forms under different conditions [14,20]. Obligatory associations are usually tighter, larger and more hydrophobic compared with the non-obligate interactions characterized by a more polar/charged interface [21]. Furthermore, PPIs

may also be classified into transient or permanent basing on their half-lives [15]; in the first case the two proteins partners associate and disassociate *in vivo*, while in the second case the interaction is very stable and the subunits usually exists in its complexed form [22]. The stability of a PPIs highly depends on physiological conditions, changes in protein expression and localization, or on the presence of effector molecules like GTP. Indeed, an interaction can be mainly transient *in vivo* but it might become permanent under particular cellular conditions [15,20]. Targeting a permanent PPIs is very difficult because the only way to hamper this kind of interaction is to identify molecules able to interfere during protein synthesis and the folding process [15].

Notwithstanding the above reported hurdles that make targeting PPIs a challenging task, advances in structural biology, mutational studies, biophysical methods and computational techniques along with the development of suitable high throughput screening (HTS) processes for PPIs inhibitors, led to successful outcomes resulting in the identification of PPIs modulators approved or entered in clinical studies.

1.2 Mode of actions of PPIs inhibitors

PPIs can be modulated by disrupting or stabilizing the binding between the protein partners. The disruptors act by inhibiting the interaction, while stabilizers increase the affinity of a PPI by forming a ternary complex which is blocked into an inactive conformation [23,24]. These two categories can be further divided in orthosteric and allosteric inhibitors (Figure3). The former bind to the PPI interfaces inhibiting the formation of the complex in a competitive manner. Instead, allosteric inhibitors interact with sites distinct from protein interface; upon ligand binding, the target protein undergoes conformational changes hampering the interaction [24].

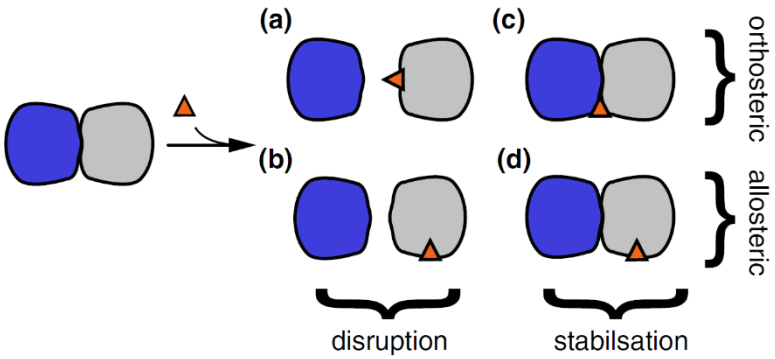


Figure3. Mechanism of action of PPI modulators (orange triangle). a) Orthosteric inhibition; b) Allosteric modulation; c) Stabilization at a site formed by the protein complex; d) Stabilization at allosteric site. This figure is reproduced from Fisher G. *et al.* [23].

Most PPI modulators belong to orthosteric disruptors. A successful example of this class of PPI inhibitors is ABT-199 **(1)**, a small molecule that binds to the anti-apoptotic protein Bcl2 (Figure4A) preventing its interaction with the pro-apoptotic protein partners thus favouring the apoptosis of tumor cells. ABT-199 was approved by FDA in 2016 for the treatment of chronic lymphocytic leukemia [25,26]. Orthosteric inhibitors interact with hotspots residues present at the interface and the major challenge in their development is related to the flat and large surface [27]. To overcome this issue, the allosteric modulation could represent a valid strategy. Targeting PPIs with allosteric inhibitors offers several advantages such as a higher subtype receptor selectivity, its effect is saturable and less prone to overdose [28]. An example of allosteric modulators is BIO8898 **(2)**, a synthetic organic molecule, which occupies an allosteric site of CD40L trimer (Figure4B) blocking its binding to CD40 [6] and inhibiting in this way CD40L-dependent cell apoptosis in cellular assays [29]. The main problem with this class of inhibitors is related to the identification of suitable allosteric sites [5]. For this purpose, several approaches like HTS followed by crystallography, phage display combined with crystallography, tethering and computational methods have been applied [28].

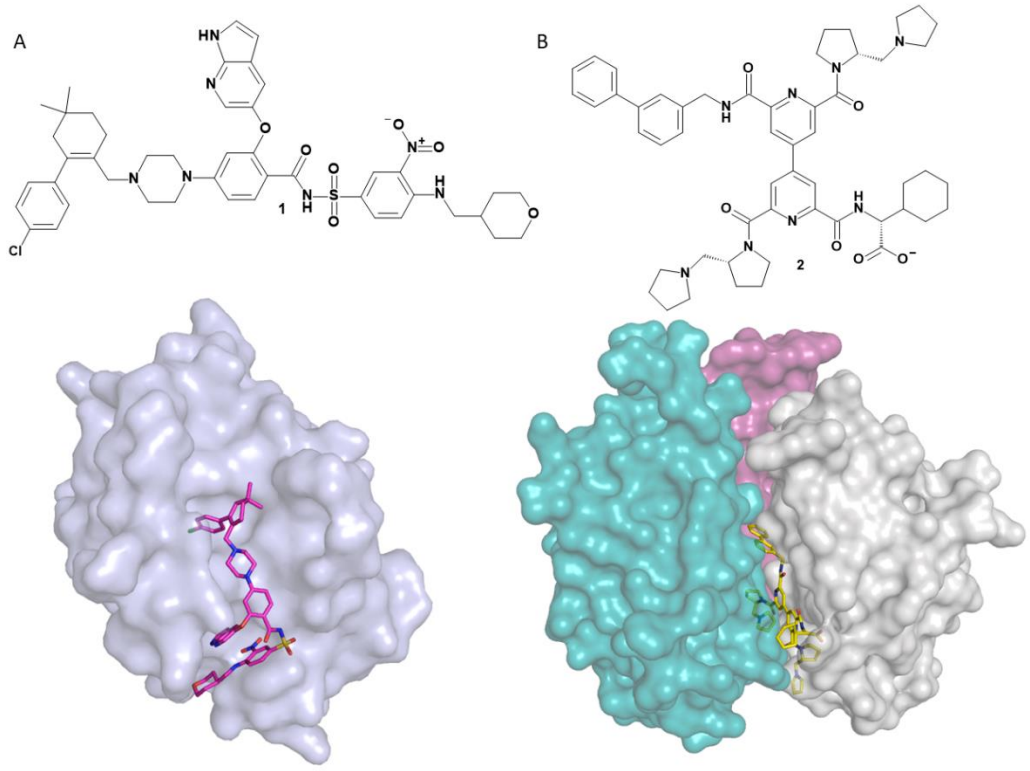


Figure 4. A) Crystal structure of the orthosteric inhibitor ABT-199 (pink stick) bound to Bcl2 interface (PDB code 6O0K). B) Crystal structure of the allosteric inhibitor BIO8898 (yellow stick) bound to an allosteric site of the CD40L trimer (PDB code 3LKJ). This image was created by using PyMOL software (www.pymol.org)

Orthosteric and allosteric inhibitors show different pharmacological properties; usually, allosteric modulators are more hydrophobic because of the more hydrophobic character of allosteric sites compared to the PPI interface, they possess a lower molecular weight and greater structural rigidity when compared to orthosteric inhibitors [27].

Overall, when the protein surface areas are small and the interfacial conformation are relatively simple, the development of orthosteric modulators might represent the right strategy, on the contrary for PPIs characterized by a large interface and complex interfacial conformations, the design of allosteric ligands may prove to be a valid approach [27].

1.3 Classes of PPI modulators

PPIs modulators can be classified in three main classes: monoclonal antibodies, peptides and peptidomimetics, and small molecules.

Monoclonal antibodies usually compete with the protein partner acting through an orthosteric mechanism [30]. The advantages of developing monoclonal antibodies for the modulation of PPIs consist in their high target selectivity and affinity [3,5]. In the last years, several antibodies started clinical trials or were approved by FDA for therapeutic or diagnostic applications [30]. Despite the higher specificity, antibodies present several limitations such as their expensive manufacturing cost, low cell and blood-brain-barrier (BBB) permeability and they are not orally bioavailable [5].

Peptides and peptidomimetics represent the second class of PPI modulators. Generally, peptides are design starting from the amino acid residues present at the interface of PPIs mimicking the natural interactions [8]. Peptides are also used as molecular probes to study the mechanism of PPI modulation or for the identification and validation of new PPIs as potential new targets. From a pharmacological point of view, they possess poor cell permeability and low stability *in vivo* being metabolised by proteases [31]. Furthermore, short peptide sequences are generally unstructured in water, while a specific conformation is required for the recognition and binding processes [32]. To improve the pharmacokinetic properties of peptides, several chemical strategies have been applied like backbone modifications, amino acid substitution using non-standard residues and macrocyclization [33]. In particular, macrocyclization constraints the peptide in its bound conformation, increases proteolytic stability and improves cell permeability [34]. Amongst the cyclization methods, the use of hydrocarbon-stapled peptides revealed to be promising for the design of PPIs inhibitors. Considering that helical motifs are common in most of protein-protein interfaces [35], with this approach it is possible to stabilize the α-helical secondary structure of a peptide by introducing an olefinic side chains at the α-carbon of a non-natural amino acid by using metathesis chemistry. Hydrocarbon-stabled peptides proved to be more stable with an improved membrane permeability [5,36]. This strategy has been successfully employed to design inhibitors of PPI of therapeutic interest [34]. An interesting result in this research field is the identification of the dual selective stapled peptide ALRN-6924 which inhibits p53-Mdm2/Mdm4 interactions; it is under clinical evaluation for the treatment of haematological and solid tumors [37].

The third class of PPIs inhibitors is represented by small molecules. Targeting PPIs with low molecular weight compounds has been considered for decades a difficult task because of the large and flat protein-protein interface. In a PPI, hotspot residues are usually concentrated at the centre of the binding interface covering an area of 250-900 Å, facilitating their interaction with a small molecule [30]. Furthermore, proteins are dynamic systems subjected to conformational changes that could result in the formation of cavities able to accommodate a small ligand. Techniques like nuclear magnetic resonance (NMR) and molecular dynamics (MD) simulation are useful to study proteins structures in solution and their dynamic behaviour [9,14]. For example, Eyrisch et al. applied a computational protocol to detect transient pockets on the surface of three systems BCL-XL, MDM2 and IL-2 in their apo-conformations. Since the native inhibitor binding site was not present in the unbound state, these crystal structures were subjected to MD simulation. The use of binding pocket detection algorithms revealed that transient pockets opened within 2,5 ps and that pockets of similar size to that of the known inhibitors at the native binding site were detected as well [38]. These reported features make targeting PPIs by small molecules feasible. Small compounds have many advantages such as metabolic stability, low manufactory cost, high bioavailability and cell permeability [8].

Small molecules could modulate PPI interactions by competing with the cognate protein partner, by occupying an allosteric site or by stabilizing the interaction [24]. An emergent strategy to inhibit PPIs involving small molecules is the covalent modification of a nucleophilic residue, such as cysteine, methionine and lysine located close to the interface, which resulted in high affinity inhibitors. Moreover, covalent inhibitors that binds to non-conserved residues showed a better selectivity over other related members of the same protein family [39].

Small molecule modulators of PPIs can be identified by HTS of synthetic and natural compounds libraries or applying *in silico* methods. Usually, they are characterised by a greater molecular weight (>400 Da) and a higher hydrophobicity compared to typical drugs due to the large and hydrophobic protein-protein interface [22, 40]. Despite some of them fall outside Lipinski's rule of five, they showed oral bioavailability [5]. Furthermore, small modulators of PPIs present more complex

structures than traditional drug-like compounds present in chemical libraries. Such characteristics highlighted the necessity to develop databases of molecules for PPI screening characterised by a higher chemical complexity and diversity than those used for classical drug design. In this context natural products offer a rich source of different chemotypes representing a good starting point for the realization, through chemical manipulation, of compound libraries for the discovery of PPIs modulators [41]. Despite the hurdles in targeting PPI by small molecules, several modulators have been successfully identified [1] leading to the approval by FDA of five small molecules for the treatment of cancer, dry eye syndrome, autoimmune diseases and as immune suppression agents. Moreover, more than 27 small molecules PPI modulators are being investigated at various stages of clinical trials as potential anticancer, antiviral and immune suppression drugs [42,43].

Chapter2

Computationally driven drug discovery methods

2.1 Introduction

The main goal of the early stages of the drug discovery process is the identification of a lead compound showing activity against a biological target followed by the optimization of its pharmacological properties and potency. For this purpose, HTS campaigns of large chemical libraries were performed by pharmaceutical industries as means to discover new lead compounds. However, HTS is time-demanding and characterised by high costs [44]. In recent years, computer aided drug design approaches became powerful tools to help researchers in identifying new drug candidates. The use of computational methods allows to decrease the number of compounds to screen in biological assays by prioritizing those showing promising results *in silico*, leading to a reduction of time and costs associated with the drug discovery process [45].

Computer aided drug design techniques are generally classified in two broad categories: structure-based and ligand-based methods.

Structure-based drug design (SBDD) approaches rely on the knowledges derived by the 3D structure of the biological target usually obtained through X-ray crystallography, NMR or homology modelling. The RCSB Protein Data Bank (www.rcsb.org) is the main repository of experimentally derived 3D structures of macromolecules representing an important source for molecular modelling studies. SBDD methods include VS and *de novo* design. In structure-based VS, libraries of drug-like compounds are computationally screened against the target of interest and the hits are ranked based on their predicted binding affinity or complementarity to the binding site [42]. Structure based VS methods include docking and pharmacophore based VS. Instead, *de novo* design methods exploit receptor structure to build up new molecules by using ligand growing programs [44,46]. In recent years, MD simulation became an important tool in SBDD due to possibility to account for receptor flexibility [47].

Ligand-based drug design (LBDD) approaches are usually employed when the 3D structure of the target is not available and are based on the information extrapolated from known active ligands. LBDD methods includes similarity and substructure searching methods, quantitative structure-activity relationship (QSAR), pharmacophore modelling and three-dimensional shape matching. In

ligand- based VS strategies, 2D or 3D chemical structures or molecular descriptors of known active compounds are set as reference to retrieve molecules with chemical or conformational similarity from a database [22,42].

Recently, both SBDD and LBDD methods have been successfully applied to design small molecules PPI modulators revealing the potentialities of these techniques as tools to overcome the challenges associated with the discovery of this class of inhibitors [40,48]. In this chapter, an overview of the computational methods employed in my research work is presented.

2.2 Ligand-binding site prediction

Drugs exert their biological effects by specifically interacting with their macromolecular target modulating its function. The interaction involves specific residues of the receptor, named hotspots, that are usually clustered in one region constituting the ligand binding site [49]. The identification of suitable ligand binding sites represents a key step in SBDD approaches such as docking [50]. When experimental information about hotspot region is not available, *in silico* techniques can be exploited to detect proper ligand binding sites. For this purpose, many computational methods have been developed that can be classified in structure-based and sequence-based approaches. The first ones are based on the 3D structure of the protein target while the latter exploit information derived from the evolutionary conservation and sequence similarity between homologous proteins [51]. Once the binding site has been detected, the next step regards the evaluation of its druggability, which is defined as the ability of a protein to bind a drug-like compound with a high affinity and specificity. For this purpose, by using binding free energy analysis and different geometrical descriptors, a druggability index is calculated providing a measure of the druggability of a given binding site [49]. In the next paragraphs, the main binding site identification approaches are discussed.

2.2.1 Structure-based methods

Structure-based methods relies on the 3D structural coordinates of the studied target to identify potential ligand binding sites. The main limitation of this approach is that it requires the 3D structure of the receptor. However, when the 3D coordinates are not available, homology modelling can be used to predict the three-dimensional protein structure [51]. Structure-based procedures can be further categorised in three types: geometric-based, energetic-based and template-based [50]. Geometric-based approaches use geometric parameters such as depth and surface area, to detect cavities and clefts on the protein surface [49]. Energy-based methods search for energetically favourable areas by calculating the interaction energy between protein atoms and a chemical probe. Finally, template-based methods predict putative ligand binding site starting from known protein templates. This approach relies on the assumption that structurally similar proteins share a similar activity and exploits global or local alignment to evaluate the similarity [50].

2.2.2 Sequence-based methods

Sequence-based methods identify ligand-binding sites by employing information derived from evolutionary conservation or sequence similarity of homologous proteins. They rely on the assumption that ligand binding site residues are conserved in a protein family [50]. Sequence-based methods require related-proteins with high sequence identity to the studied targets. In these approaches, homologous sequences to the query protein sequence are collected and a multiple sequence alignment (MSA) is generated; subsequently, conserved residues are identified among all the sites in the MSA [52].

Sequence-based methods are less computationally expensive and accurate compared to structure-based methods [50].

2.3 Molecular docking

Molecular docking is a widely used structure-based method aimed to predict the binding pose of a ligand within a target binding site. Docking algorithms estimate binding energies and rank the ligands by using different scoring functions [53].

Molecular docking simulation consists of two mains steps: the conformational searching followed by pose ranking by means of scoring functions [48].

Docking revealed to be a valid technique to study PPIs thanks to the development of protein-protein and protein-peptide docking methods that allow to investigate the interactions between the protein partners giving useful insights for the design of PPIs modulators [40].

2.3.1 Conformational search methods

According to the degree of flexibility of the molecules involved in the calculation, docking can be classified into rigid, semi-flexible and flexible [54].

In rigid docking both ligand and protein are treated as rigid and three translational and three rotational degrees of freedom are used for sampling [55].

In semi-flexible docking the ligand is flexible while the protein is kept rigid. In this approach, the methods used for the conformational search are divided into systematic procedures, stochastic search and simulation methods. In the systematic methods, all the degrees of freedom in a molecule are considered to place the ligand in the binding site [44]. Stochastic search algorithms randomly modify the values of degrees of freedom of a single ligand or a population of ligands; the applied modifications can be accepted or rejected based on probability functions like genetic algorithm methods and the Monte Carlo method [53].

Simulation approaches exploit various techniques including MD simulation and energy minimization [56]. Lastly, in flexible docking both protein and ligand are considered flexible during the calculation. This approach relies on the concept that protein undergoes conformational changes upon ligand binding and for this reason, several attempts have been made to introduce protein flexibility in docking simulations. The strategies used to accomplish this scope include side-chain flexibility based on rotamer libraries or the use of multiple receptor conformations that can be obtained experimentally or via computational techniques [55,56].

2.3.2 Scoring functions

In docking simulation, ligand conformations are ranked by means of scoring functions that estimate the binding affinity of a ligand to the protein target [44].

Scoring functions can be divided in three classes: force field-based, empirical and knowledge-based scoring functions.

Force field-based scoring functions predict the binding free energy of the ligand-protein complex by summing intermolecular interactions like electrostatic and van der Waals forces [53]. Instead, empirical scoring functions are based counting different empirical energy terms such as van der Waals, electrostatic, hydrophobicity that are weighted by a coefficient obtained by regression analysis from experimental binding affinity data or from a training set of experimental protein-ligand complexes [55,56].

Finally, knowledge-based scoring functions assume that statistically more frequent ligand-protein contacts are energetically more favourable. In this method known structures are used to calculate the frequencies of ligand-receptor atom pairs contacts that are subsequently converted into an energy term. During docking calculation, these energy terms are summed up for all ligand-receptor atom pairs resulting in the score of the pose [55].

It is widely accepted that often scoring functions fail in ranking docking poses correctly. Indeed, many physical events that take part to molecular recognition like entropy or solvation effects are ignored in the most used scoring functions. An alternative to overcome this limitation is the use of consensus scoring that consist in combining multiple scoring functions in order to compensate for errors deriving from single scoring functions thus increasing the probability to find the correct solution [56].

2.3.3 Protein-protein and protein-peptide docking

The 3D structure of protein-protein or protein-peptide complexes is important to study protein interactions, to predict mutation effects and to provide a structural base for drug design. To this aim, docking revealed to be a useful technique for predicting protein-protein and protein-peptide complexes when experimentally solved 3D structures are not available. As protein-ligand docking, also protein-protein and protein-peptide docking consists of two stages: sampling and scoring/ranking. In sampling, potential binding orientations are generated by using global or local searches [57]. In the first case, one protein is kept fixed and is named

"receptor" while the other protein or the peptide, called "ligand" molecule, is moved around the receptor. The global search stops when all the possible orientations between the two proteins in the 3D space are explored. This method is computationally expensive as many translations and rotations are executed during the calculation; to reduce the docking complexity, several algorithms, like Fast-Fourier transform (FFT) have been developed. In the local search, local features such as pockets and solvent excluded areas on the target surface are matched to achieve a good complementarity. A third sampling approach consist of integrating prior biophysical, biochemical and chemical information about the interaction in the sampling stage in order to reduce the search space thus improving the accuracy of the docking prediction. Furthermore, the sampling can be rigid or flexible. In rigid-body sampling, all the structural features like bond length, bond angles, backbone orientation are not modified; instead, in the flexible sampling, conformational changes are considered [58].

The sampling process results in tens to thousands putative complexes that are ranked basing on the score [58].The scoring used to evaluate the putative complexes combines different terms like solvation energy, dielectric constants, electrostatics, van der Waals interactions, hydrogen bonds and clashes [59]. The growing interest in the study of protein interactions through protein-protein and protein-peptide docking led to the launch of CAPRI (Critical Assessment of Predicted Interactions), a communitywide experiment aimed to assess the performance of various docking methods [60].

2.4 Pharmacophore modelling

The use of 3D pharmacophore models is one of the most popular approaches for the fast and accurate VS of large databases [61]. A pharmacophore model is defined as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or block) its biological response" [62]. According to this definition, pharmacophore models represent an abstract concept describing the steric and electronic properties necessary for the molecular recognition of an active molecule by the target macromolecule [54]. The chemical functionalities, named "features",

involved in ligand-protein interactions and represented by the pharmacophore model are hydrogen-bond donors, hydrogen-bond acceptors, positively and negatively charged groups, aromatic rings and hydrophobic regions. Other important information that must be included into a pharmacophore model is the 3D location of the features and their spatial orientation if the interaction is directed like hydrogen bonding [61]. It is possible to distinguish two types of pharmacophore models basing on the data used for their construction: structure- and ligand-based. If the experimental structure of the ligand-receptor complex is available, the atomic coordinates and ligand interactions can be exploited to correctly place the pharmacophoric features and also to incorporate information about the shape of the binding site, obtaining a structure-based pharmacophore model [63]. If there are not experimental data about the 3D structure of the ligand-receptor complex, a set of active compounds, that binds to the same target at the same binding site, can be used for the generation of a ligand-based pharmacophore model. Because the active conformations of the training set molecules are usually not known, different conformers must be generated so that at least one represents approximatively the active conformation. Afterwards, 3D structures of the conformers are aligned and the pharmacophoric features located at a specific position, that are common to all the ligands, are extrapolated to yield multiple pharmacophore models ranked according to suitable fitness functions [63,64].

Once the pharmacophore model has been generated, if experimental data about binding ligands are available, datasets of known active and inactive compounds are exploited to assess the quality of the model. When only a limited number of experimentally tested inactive compounds are known, decoys could be used [65]. A decoy is a compound with unknown biological activity presumed to be inactive against the studied target and characterized by physicochemical properties similar to the active ligands [64]. The Directory of Useful Decoys-Enhanced (DUD-E) provides a free web-based tool (http://dude.docking.org), for the generation of decoys giving the smile codes of active compounds as input [66].

The quality of a pharmacophore model depends on its ability to discriminate between active and inactive compounds and on how many active molecules it is able to retrieve from a dataset in VS. For this purpose, various parameters can be

calculated such as the enrichment factor that measures the enrichment of actives in a hit list compared to a pure random selection and the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve where the true positive rate is plotted against the false positive rate thus highlighting the sensitivity (the ability to pick true positive hits) and the specificity (the capability to avoid inactive compounds) of the pharmacophore model [67]. Once the model has been validated and optimized, it can be used as filter to virtually screen commercial or in-house libraries. Different conformations for each molecule must be computed and stored in the database. The database search is generally implemented as a multistep filtering process consisting of a pre-filtering step and a 3D alignment procedure. In the pre-filtering process, all the compounds that, basing on feature types, features count and quick distance checks, cannot be matched by the pharmacophore model are removed. Subsequently, the conformations of the molecules that might fit the query must be examined to probe their ability to match the spatial arrangement of the pharmacophoric features by 3D alignment algorithms [61].

Beside the identification of new leads, other applications of pharmacophore models include target fishing, modelling metabolism, structure-activity relationships, scaffold hopping and selectivity profiling studies [64,65].

Recent developments in pharmacophore modelling derives from its combination with MD simulation which enable to consider the flexibility of the ligand-target complex, allowing the exploration of features not detected in the static representation of the crystal structure [68,69,70].

2.5 Similarity methods

Molecular similarity principle assumes that molecules structurally similar are more likely to share similar physical properties and biological activity [71].

The similarity concept has been widely employed in the early stages of drug discovery process to identify new bioactive compounds. Any similarity search method implies the use of one or more known active ligands as reference compounds and comprises three components: the molecular representation, the search algorithm and the quantitative measurement of the molecular similarity between the so-represented structures. Hence, both reference and database

molecules are converted into an appropriate representation format that will be used for pair-wise molecular comparison and similarity computations [72]. Different types of structural representation have been exploited to estimate molecular similarity, such as physicochemical properties, topological indices, molecular fields, molecular shapes, pharmacophore features and molecular graphs. Concerning the quantification of the similarity between molecules, various metrics have been developed, among these Tanimoto coefficient is the most used [73].

Molecular similarity techniques can be divided into 2D or 3D similarity methods basing on the structural representation. 2D approaches, like substructure search, fingerprint similarity and 2D descriptor-based methods, are the most popular. They rely on the 2D structural information without performing structural alignment to calculate the similarity between two compounds. 3D methods consider the 3D conformations of the molecules in the similarity search. They comprise shape similarity, pharmacophore modelling, 3D fingerprints and molecular field-based methods [73].

2.5.1 2D Similarity methods

Originally, 2D similarity methods were based on the comparison of 2D molecular graph in which atoms and bonds are represented as nodes and edges respectively. Substructure searching is one of the first approaches for database mining. It consists of specify a molecular fragment as query and all compounds bearing this portion are retrieved from a library. Advances in chemical languages led to the development of SMILES and SMARTS formats that convert molecular graphs into strings of ASCII characters, thus making the substructure searching computationally more efficient than graph matching. Substructure searching is not limited to single queries, but it is possible to perform a simultaneous search for multiple substructures by merging individual queries that represent sets of defined structural fragments. It has been recognized that this type of molecular representation could be used not only for substructure searching, but also to establish the structural similarity of different molecules, thus leading to calculation of structural fragment-based fingerprints which became the most popular tools for similarity searching [72].

Molecular fingerprints are defined as bit-strings encoding representative information about molecular structures and properties [74].

Each bit generally accounts for the presence or absence of a definite feature. The bit is set to "1" if the molecule contains the feature, conversely if the feature is not present, the bit is set to "0" [72]. This type of molecular representation can be easily compared allowing similarity calculation on large scale [75].

It is possible to classify fingerprints in 2D fingerprints which are calculated from molecular graphs and in 3D fingerprints calculated from molecular conformations [72]. Among the most popular 2D molecular fingerprints in medicinal chemistry, there are the Molecular ACCess System (MACCS) and extended connectivity fingerprint (ECFP) with bond diameter four (ECFP4). The MACCS fingerprint, called also MACCS structural keys, is a fragment-based fingerprint consisting of 166 structural fragments based on SMARTS patterns. Each bit position is assigned to a given structural fragment and its presence or absence in a molecule is detected. The ECFP4 fingerprint is a circular fingerprint which consider the local bond topologies, that specify the connectivity of atoms, in the proximity of each non-hydrogen atom in a molecule. The extent of the neighbourhood relies on the so-called bond diameter defined by the maximum number of bonds considered [75].

In fingerprints-based similarity search, the fingerprints of the reference and database molecules are compared in a pair-wise way and the bit string overlay is quantified giving a measure of the similarity. Through this procedure, database compounds are ranked in the order of descending similarity to the reference molecule [72]. To quantify the bit string overlap, various similarity metrics or coefficients have been developed. Among these, one of the most popular is the Tanimoto coefficient (Tc). Tc is calculated as follow, where a and b represent the number of features contained in compounds A and B respectively, and c is the number of features shared by A and B:

$$Tc\ (A,B) = \frac{\mathrm{c}}{\mathrm{a} + \mathrm{b} - \mathrm{c}}$$

Tc values are comprised between 0 and 1, where 0 means that there is no fingerprint overlap, while 1 correspond to identical fingerprints that does not necessarily coincide with identical molecules [75].

2D similarity based virtual screening methods often retrieve close chemical analogues to the reference compound and lack the information related to the 3D structure of ligands and protein, which is important considering that ligand binding to the target protein is governed by atomic interactions in the 3D space. To overcome these limitations, 3D similarity methods have been developed [76].

2.5.2 3D Similarity methods

3D similarity methods rely on the assumption that compounds with similar conformational features are likely to have similar biological activity [77].

Usually, 3D representations employ 3D features like atomic distances, fields, volume and surfaces and they can be classified into five categories basing on how the 3D structure is represented: atomic distance, Gaussian function-, surface-, field- and pharmacophore-based methods [76].

In the atomic-based distance methods, molecules are described by the relative positions of its atoms and the distance between atom pairs is computed to estimate the similarity between two compounds [73,76]. This method is simple and fast since it does not require molecules alignment [78].

Gaussian function-based approaches assume that two compounds with similar volume share a similar shape which is determinant for the recognition by the target macromolecule [73,79]. In this method, the similarity is calculated as volume overlap of two molecules after the alignment [76]. The two most used models to represent molecules are hard spheres and Gaussian spheres. The former considers each atom in the molecule as a sphere and the volume molecule is computed based on the union and the intersection of their volumes. Instead, the latter describes a molecule as a set of overlapping Gaussian spheres. In this approach, the inclusion-exclusion principle is applied in order to measure the volume of a molecule obtained by calculating the volume of all overlapping Gaussian spheres [78].

Another way to estimate shape similarity is the comparison of the surfaces of the molecules. To this aim, the more practical and much easier ways to describe molecules surface are the solvent-accessible surface and van der Waals surface [73]. Field-based methods relies on quantum mechanical calculations. Compounds characterised by different structures but possessing similar fields might be able to bind to the same target site thus showing the same biological activity. In this approach, the similarity score is associated to the electron density of the molecule which implicate the computation of steric fields and electrostatic fields [80].

Pharmacophore-based methods usually begin with the identification of pharmacophoric points representing features like hydrogen bonding, hydrophobic and aromatic groups, positive and negative ionizable groups. Basing on the used description method, 3-points or 4-points pharmacophore keys, the pharmacophoric points are connected to form respectively a triangle or a tetrahedron. These representations are encoded into strings storing the information of the selected features and the related inter-feature distance range. Tc or other metrics are employed to quantify the similarity between the query and database molecules [76].

Beyond the evaluation of similarity between molecules, 3D similarity approaches can be used for protein structures comparison. For this purpose, traditional methods are based on protein alignment which requires extensive rotational and translational sampling limiting their use on large scale. Shape similarity has been exploited to probe local or global similarity between protein structures. In this scenario, an important application of shape similarity regards the measure of similarity between protein binding sites. Usually, sequence and structural alignments are not useful to compare binding pockets of proteins with diverse fold. Ligand binding sites are more conserved than protein structures, therefore the comparison between binding pockets might be useful to predict the biological activity of a molecule or for drug repurposing assuming that similar binding site recognize chemically similar molecules. The most used 3D methods for protein structure comparison are Gaussian function- and surface-based methods [73].

2.6 Molecular dynamics simulation

It has been widely recognised that ligand binding process to its macromolecular target is a dynamic process involving conformational changes of the receptor structure overcoming the initial "lock and key" theory model according to which a rigid receptor can accommodate a small molecule without experiencing any structural rearrangements [81]. In this scenario, MD simulation is a widely applied computational methods used to calculate the time-dependent motion of biological molecules [57].

MD exploits Newton's equation of motion to compute the net force and acceleration of each atom. Every atom *i*th of the system is treated as a separate point particle with a mass m_i and a fixed charge q_i. The time evolution of a set of interacting atoms is simulated by solving the Newton's second law, where F is the force exerted on *i*th atom, at time *t* and position *r*, and *a* is the corresponding acceleration [46]:

$$F_i(t) = m_i a_i(t) = m_i \frac{d^2 r_i(t)}{dt^2}$$

During MD simulation, through the integration of Newton's equation, consecutive configurations of the evolving system are produced resulting in trajectories in which the positions and velocities of the particles over the time are specified [82].

For dynamic simulations involving biological macromolecules, numerical methods are employed to split the integration of Newton's second law into discrete time intervals, named time-step (δt) [82]. The velocity-Verlet integrator is one of the most used integration algorithms [46]. It calculates the position and velocity of an atom *i* at the time-step *t+δt*, starting from the step *t*, according to the following equation, where $r_i(t)$, $v_i(t)$ and $a_i(t)$ are respectively position, velocity and acceleration of atom *i* at time t, while $r_i(t+\delta t)$, $v_i(t+\delta t)$ and $a_i(t+\delta t)$ represent position, velocity and acceleration of atom *i* at time t+δt:

$$r_i(t+\delta t) = r^i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2$$

$$v_i(t+\delta t) = v_i(t) + \frac{1}{2}[a_i(t) + a_i(t+\delta t)]\delta t$$

Acceleration is calculated according to Newton's equation of motion from the forces acting on atom *i* [55]. For drug discovery purposes, a classical mechanics description of the forces is generally employed, where motions of nuclei are considered disregarding the presence of electrons. This simplified representation is noted as molecular mechanics (MM) of force field (FF). Therefore, forces are computed as follow, by introducing an empirical potential energy function *V(r)* [82]:

$$F_i(t) = -\frac{dV(r(t))}{dr_i(t)}$$

In FF, the potential energy of the system is computed by the sum of bonded and nonbonded energy terms. The former describes intramolecular interactions of the atoms and comprise change in the potential energy as function of bond stretching, bending and torsions involving atoms connected to each other. The latter regard van der Waals and Coulomb electrostatic interactions between atoms [55,82]. Many FFs have been developed and currently used for MD simulations such as AMBER [83, 84], OPLS [85,86], CHARMM [87,88] and GROMOS [89,90].

MD is a statistical mechanics method in which physical quantities are represented by averages of microscopic states of the system, called configurations, distributed basing on a statistical ensemble [91]; among these, the most used in MD simulation are the Microcanonical Ensemble (NVE), the Canonical ensemble (NVT), and the Isotherma-isobaric ensemble (NPT) [46]. Newtonian dynamics implies the conservation of energy, thus sampling microstates in NVE ensemble, characterised by a constant number of particles (N), volume (V) and energy (E). However, it is possible also to control the temperature and pressure during the simulation by employing thermostat and barostat algorithms to better reproduce the effective macroscopic behaviour. In NVT ensemble the temperature is maintained at the desired value by appropriate alterations of Newton's second law, instead, in NPT ensemble the pressure is kept constant by properly scaling system volume [82].

Macroscopic systems are large and therefore computationally expensive to compute by molecular simulation. To overcome this issue, periodic boundary conditions (PBCs) are employed in MD simulation to minimize surface effects and mimic an infinite system. With PBCs, the system is confined in a unit cell that is

replicated throughout the space to form an infinite lattice. During the simulation, when a molecule moves from the central box, its periodic image in the adjacent cells moves with the same orientation and in the same way. In other words, when a particle cross one boundary of the cell, one of its images appears on the opposite side. FF based MD simulation presents several limitations such as the lack of polarizability, since in molecular mechanics the charge of each atom is fixed. To address this point, polarizable FF have been developed but their application is computational expensive, and parametrization is not user-friendly [46,82]. Moreover, FF based simulations cannot be used to study chemical reactivity since bond forming and breaking cannot be modelled in classical MD simulation. This is especially important when transition metals are involved in the binding process. For this purpose, quantum mechanical (QM) calculations have been introduced into classical MD FFs. Specifically, in this case, the motions of the protein binding site, where the reactions occur, are modelled according to QM whereas the rest of the system is simulated using classical MD [47].

The main advantage of the use of MD simulation in drug discovery is the possibility to account for receptor plasticity by generating different conformations that can be exploited to perform docking studies. Furthermore, MD can be applied as post-processing docking tool to optimize the obtained protein-ligand complex, to evaluate the stability of the predicted docking poses and to explore additional ligand-protein interactions. Furthermore, MD based methods are widely exploited to calculate the ligand-protein binding free energy for example by using MM-PB/GBSA or free energy perturbation (FEP) allowing to overcome the limits of the scoring functions implemented in docking software in predicting the correct binding pose and the correct ranking of the resulting poses [92]. MD methods are also applied to investigate kinetics and thermodynamic of drug binding to its biological target, to detect and characterize binding site otherwise not detectable from the crystal structure and to investigate the role of water molecules in ligand binding process [82,93]. MD-based approaches find application even in the rational design of small molecules able to modulate PPIs. In this context, MD revealed to be a useful tool to identify hotspot residues through computational alanine scanning (CAS) [94], to unveil the presence of druggable binding site [81] and to

investigate protein-protein association process [95]. In recent years, thanks to the development of new and more-user friendly software and of more efficient computer hardware like graphics processing units (GPUs), that allow to speed up the calculations, MD simulation and related methods have become more powerful and accessible tools for guiding the rational design and optimization of bioactive molecules [93].

Section 2

Chapter3

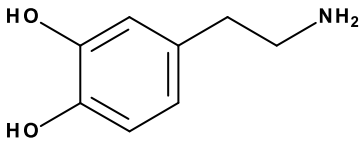
α-synuclein as target for pharmacological intervention in Parkinson's disease

3.1 Parkinson's disease: main features and current drug therapy

Parkinson's disease (PD) is one of the most common neurodegenerative disorders in elderly age.

It has been estimated that 7 to 10 million of people worldwide are affected by PD. The main risk factors comprise age, male gender, genetic factors and environmental factors like rural living, exposure to herbicides and heavy metals [96,97].

PD is characterised by the loss of dopaminergic neurons in the substantia nigra pars compacta of the central nervous system (CNS) resulting in the reduction of dopamine (DA) **(3)** levels [98].



3

DA deficiency is associated with the motor symptoms of PD, such as bradykinesia, rigidity, rest tremor and postural instability [99]. Human post-mortem and animal model studies revealed that other regions of the peripheral and CNS are involved in the neurodegenerative process, probably before the substantia nigra. The neurodegeneration in these areas leads to the non-motor symptoms of PD including rapid eye movements, sleep disorders, depression, olfactory disfunctions and constipation. Usually non motor features occur before the appearance of the typical motor symptoms [100]. Beside the loss of dopaminergic neurons, another hallmark of PD is represented by the presence of intra-cytoplasmatic inclusion bodies known as Lewy bodies which are mainly constituted by the misfolded, non-soluble and aggregated form of the pre-synaptic protein α-synuclein (α-syn). Lewy bodies can be found not only in the CNS but also in the spinal cord and peripheral nervous system [98].

To date, the management of PD includes the administration of drugs able to restore DA levels thus improving the motor features related to the disease. In particular, the clinical treatment includes the use of L-Dopa **(4)**, monamino oxidase-B (MAO-B) inhibitors, catechol-o-methyltransferase (COMT) inhibitors, DA agonists and anti-cholinergic drugs [96]. In Chart 1, some of the compounds

belonging to the before mentioned drug classes currently adopted in PD therapy are displayed.

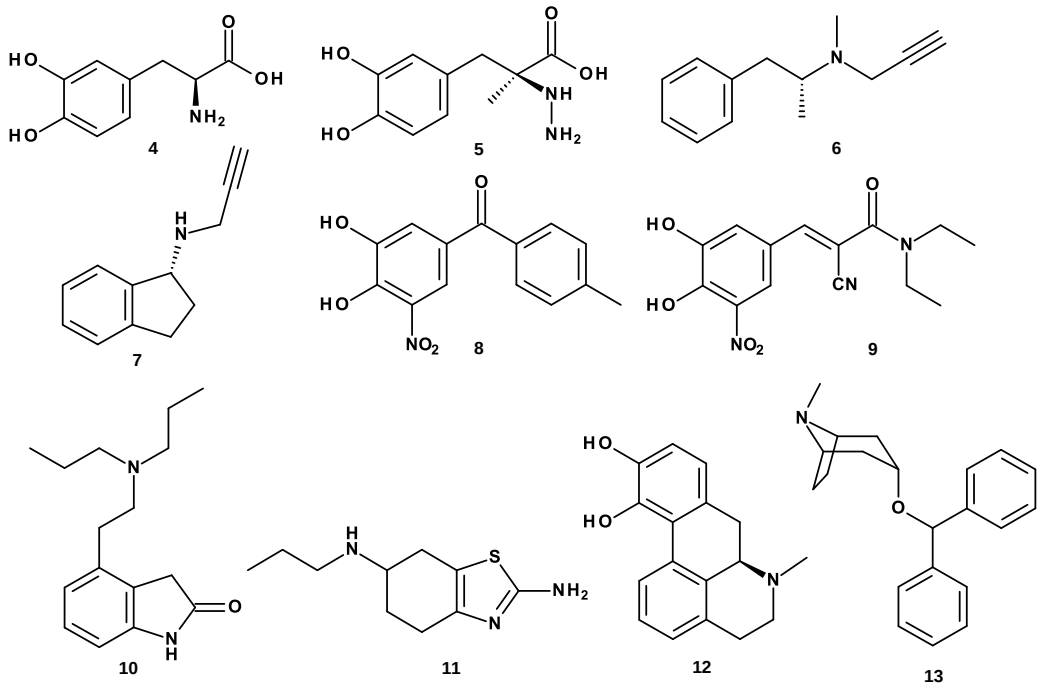


Chart1. Chemical structures of some of the compounds used in PD therapy: L-dopa **4** and carbidopa **5**, selegiline **6**, rasagiline **7**, tolcapone **8**, entacapone **9**, ropinirole **10**, pramipexole **11**, apomorphine **12** and benztropine **13.**

L-dopa, the precursor of DA, is the most prescribed therapeutic for PD. It is a prodrug converted in DA in CNS. It is associated with carbidopa **(5)**, a peripheral decarboxylase inhibitor, in order to reduce some of its side effects such as nausea and to prevent its conversion in DA in the bloodstream, allowing to get more in the CNS [101,102]. MAO-B inhibitors, such as selegiline **(6)** and rasagiline **(7)**, and COMT inhibitors, like tolcapone **(8)** and entacapone **(9)**, act by blocking the enzymes responsible for dopamine catabolism thus increasing DA concentration [102].

DA agonists mimic DA effect by activating its receptors [96]. Among these there are ropinirole **(10)**, pramipexole **(11)** and apomorphine **(12)**. Finally, anti-cholinergic such as benztropine **(13)**, bind muscarinic acetylcholine receptor re-establishing the equilibrium between DA and acetylcholine; the inhibition of the activity of acetylcholine alleviates the motor symptoms of PD. The main disadvantage associated with the use of these drugs is the reduction of their efficacy with the advancement of the disease. Furthermore, they cause several side effects like

nausea, vomit, abdominal pains, and dizziness [102]. More serious complications like hallucinations and confusion could occur with the assumption of DA agonists [103].
The main challenge in the treatment of PD regards the lack of therapeutics able to slow or halt the neurodegenerative process [98]. In this scenario, the modulation of α-syn aggregation by small molecules proved to be a promising disease-modifying strategy for the development of novel therapeutic tools for the cure of PD [104].

3.2 α-synuclein

α-synuclein is a presynaptic protein highly expressed in the CNS [105]. Although its physiological function is not fully understood, α-syn is involved in vesicle trafficking, vesicle fusion, neurotransmitter release and axonal transport [100].
From a structural point of view, α-syn is composed by 140 amino acid residues organised in three different domains: N-terminal domain (aa 1-60), central NAC domain (aa 61-95) and C-terminal domain (aa 96-140). The amphipathic N-terminal domain contains 11-aa repeats with a conserved motif KTKE/QGV [106]. Five missense mutations (A30P, E46K, H50Q, G51D and A53T) correlated with familiar forms of PD have been identified in this protein portion and it has been demonstrated that they contribute to accelerate the rate of α-syn fibrillation. The central hydrophobic NAC domain is involved in self aggregation because of its high tendency to form β-sheet rich structures. The C-terminal domain is rich of proline, aspartate, and glutamate residues and has no recognized structural elements [107,108].

In the monomeric soluble form, α-syn is an intrinsically disordered protein [109] while it adopts an alpha helical conformation upon binding with acidic phospholipids of membranes [110]. In PD, α-syn misfolded into insoluble and toxic aggregates composed by parallel hydrogen-bonded β-sheets that constitute the main component of Lewi bodies (Figure5) [98,111].

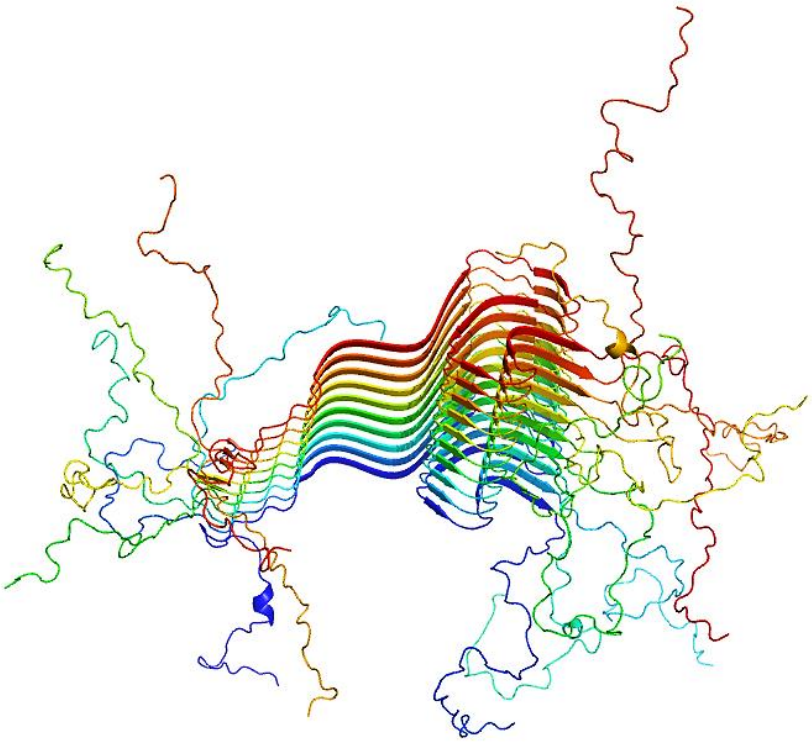


Figure5. Structural model of α-syn fibrils (PDB code 2N0A). The image was created by using PyMOL software (www.pymol.org).

In vitro aggregation studies, performed by using recombinant α-syn, revealed that the monomeric form of the protein assemblies into stable fibrils through the formation of metastable oligomers structures, which are reported to be the most toxic species (Figure6) [108,112]. However, the relationship between the different α-syn oligomeric forms and their inter-conversion mechanism has not been clearly elucidated [106]. Moreover, α-syn aggregates spread in a prion-like manner into other brain region contributing to the progression of neurodegeneration [113]. Several experimental studies demonstrated that the injection of synthetic α-syn fibrils in different brain areas of transgenic mouse overexpressing α-syn, resulted in the generation of cellular inclusion bodies located nearly and distant from the injection sites [114,115,116].

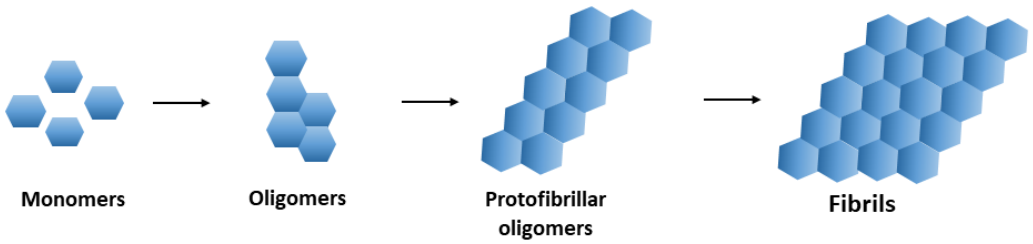


Figure6. Simplified aggregation scheme of α-syn.

α-syn aggregates induce cytotoxicity by altering membrane permeability, promoting oxidative stress, causing mitochondrial depolarization and lysosomal dispersion.

Various factors induce α-syn aggregation such as low pH, organic solvents, pesticides, dyes, heparin and other glycosaminoglycans. The interaction with metals also promotes α-syn misfolding and aggregation. In order to reduce α-syn toxicity due to metal interaction, various strategies have been developed like inhibiting its interaction with metals by employing compounds that stabilize α-syn, by reducing the levels of free metals or decreasing the oxidative stress caused by metals [106].

Considering its central role in PD, α-syn represents a viable target for the development of new therapeutic agents for the treatment of PD. Different approaches that directly target α-syn have been emerged including gene silencing, immunotherapy and aggregation modulation [117]. An overview of these different strategies is presented in the following paragraphs with a more in-depth description concerning α-syn aggregation inhibition by small molecules.

3.3 Therapeutic strategies to target α-synuclein

3.3.1 Gene silencing

Aging is one of the risk factors of PD as with human aging there is an accumulation of soluble form of α-syn that is related with a reduction of DA levels in the striatum without death of neurons in substantia nigra. The risk of α-syn aggregation increases with its overexpression, consequently decreasing of the protein levels reduces the risk of oligomerization thus preventing the formation of toxic aggregates [118].

The production of α-syn can be inhibited by employing siRNA that targets α-syn mRNA. *In vivo* experimental studies revealed that the infusion of siRNA reduce hippocampal and cortical α-syn levels [119]. However, unless there is a specific targeting towards α-syn in the brain, this strategy could induce similar effects in the peripheral tissues. α-syn is also present in peripheral nerves and red blood cells, therefore, if administered systematically, this therapy could induce peripheric adverse effects [118]. Additionally, functional deficits in the nigrostriatal

system have been observed in mice after gene silencing studies. Furthermore, the effects of gene therapy are limited to a single brain site thus not affecting the disease spreading [117]. Another strategy for decreasing α-syn expression is the use of β2-adrenergic agonists. Epidemiological studies showed that in patients treated with β2-adrenergic agonists for asthma the incidence of PD is lower. Several experiments conducted in cell cultures and animal models demonstrated that this class of drugs reduces the expression of α-syn mRNA involving an epigenetic mechanism probably due to the acetylation of Histone 3 lysine 27 [104,120].

3.3.2 Immunotherapy

Among the possible therapeutic strategies for PD, there are active and passive immunotherapies which target extracellular α-syn due to their low cell permeability. Active immunization consists in stimulating the production of antibodies against the target protein, whereas the passive immunization is reached by the direct administration of antibodies providing temporary protection against the pathology. Different immunotherapy approaches are currently being studies in clinical trials. Among these, there is AFFITOPE® PD03A, developed by the biotech company Affiris, which is a synthetic vaccine containing an α-syn mimicking peptide. Clinical studies did not highlight serious adverse effect related to administration of AFFITOPE® PD03A; furthermore, the immune response to the vaccine was dose dependent [118].

Regarding the passive immunization, two antibodies have been investigated in clinical trials. The first one is BIIB054 developed by Biogen. It is a human derived antibody which targets the aggregate form of α-syn thus preventing its spreading. This antibody presented a good tolerability in healthy volunteers and in patients with early PD, and it also displayed favourable pharmacokinetic properties. The second one is PRX002 realised by Prothena, which is a humanized monoclonal antibody targeting aggregated α-syn. Clinical trials revealed that PRX002 possesses good tolerability, safety and pharmacokinetic profiles [117,118].

Despite the promising results achieved during clinical studies, α-syn modulation through immunotherapy presents several challenges. Indeed, antibodies bind to certain forms of oligomeric and fibrillary structures of α-syn and it is not possible

to establish target engagement due to the impossibility to measure α-syn accumulation in the brain [117].

3.3.3 Inhibition of α-syn aggregation

Considering the toxicity of α-syn aggregates, the inhibition of its aggregation represents a valuable disease modifying strategy aimed to reduce the neurotoxic effect of fibrils and to slow down the prion-like spreading of the pathology [117]. For this purpose, several approaches could be exploited such as the use of heat shock proteins (HSP), intrabodies, peptides and small molecules [118].

3.3.3.1 Heat shock proteins

HSPs are molecular chaperones that prevent protein misfolding and aggregation. *In vitro* and *in vivo* studies revealed that HSPs could reduce α-syn aggregation [118]. HSPs inhibit α-syn assembly by forming transient interactions and furthermore, their effect is dependent on the kinetics of aggregation; the slower the rate of aggregation, the more effective the HSP are at preventing the formation of fibrils [121].

3.3.3.2 Intrabodies

Another approach useful for decreasing protein association inside cells, is the use of intrabodies which are small antibodies fragments able to target intracellular epitopes. The Fv regions of antibodies which confers antibodies specificity, can be expressed separately form the full-length immunoglobulin thus maintaining the high specificity and affinity target of conventional antibodies. Intrabodies can bind to different α-syn structures (monomeric, oligomeric and fibrillar) and proved to be effective in reducing α-syn soluble form thus preventing its aggregation. The main limitation concerning the clinical application of intrabodies is related to their route of administration. Indeed, they require direct delivery into CNS, through viral vectors, to achieve therapeutic concentrations in the brain for protracted periods [118].

3.3.3.3 Peptide inhibitors

A valuable strategy to design α-syn aggregation inhibitors is the development of peptides that mimic the amyloidogenic sequence of the protein α-syn [122].

Peptides present many advantages such as high selectivity, good tolerability, low accumulation in tissues, predictable metabolism, and low toxicity. However, their application as therapeutic agents is restricted by their chemical and physical instability, the high sensitivity to proteolytic degradation, the low membrane permeability, the short-half life and the generation of immune response. Several approaches could be employed to enhance peptides stability like replacing L-amino acid with D-amino acid and applying chemical modifications such as N-methylation of the constituting residues or cyclization [107].

El-Agnaf *et al.* synthesised an overlapping library of synthetic 7-mer peptides covering the entire sequence of α-syn with the aim to inhibit α-syn aggregation. The best result was obtained with peptides reproducing the NAC sequence of α-syn [122]. Another approach is represented by β-sheet breaking peptide designed starting from the α-syn mutant sequences. Also in this case, the substitutions in the NAC portion yielded the best results with the T72P-6 mer peptide (PGTAV) being the most effective in halting α-syn fibrillation. Despite the mechanism of action of these peptides is not clear; it seems that they inhibit α-syn fibrillation by interacting with the monomeric α-syn preventing the formation of the toxic β-sheets-rich aggregates [123].

Cyclic D,L- α-peptides showed to be able to self-assemble to form supramolecular structures that mimic structure and function of amyloids. These peptides might act by cross-reacting with the misfolded proteins thus modulating their assembly because of their high similarity. Among these, the cyclic D,L- α-hexapeptide CP-2 **14** (Chart 2) was able to inhibit α-syn aggregation by interacting with the N-terminal and the NAC regions of the soluble form of α-syn thus forming CP-2- α-syn aggregates that showed no toxicity in PC12 cells [124].

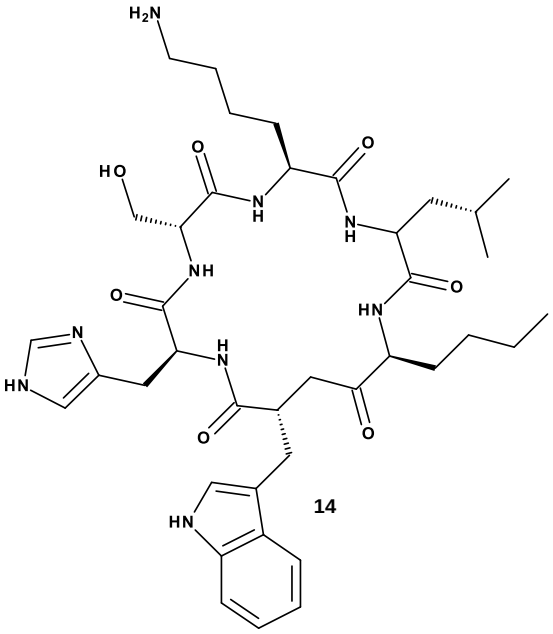


Chart 2: Structure of CP-2

Another class of peptides-based inhibitors of α-syn aggregation, that presents no sequence homology to α-syn, are β-Hairpins, small protein structure motifs characterised by two β-strands, linked by a short loop or a turn, fold to establish H-bonds with each other. β-Hairpins carrying cross strands Trp-Trp and Tyr-Tyr pairs exhibited the capability to inhibit α-syn assembly probably by interacting with hydrophobic sites of the protein interface causing the formation of non-amyloid aggregates that are non-toxic [125].

3.3.3.4 Small molecule inhibitors

Small molecules are relatively low molecular weight compounds that can interact with the monomeric form of α-syn or its aggregation intermediates thus halting the amyloidogenic process. It is possible to classify small molecules α-syn aggregation inhibitors in two main categories: polyphenols and non-polyphenols [106].

3.3.3.4.1 Polyphenols

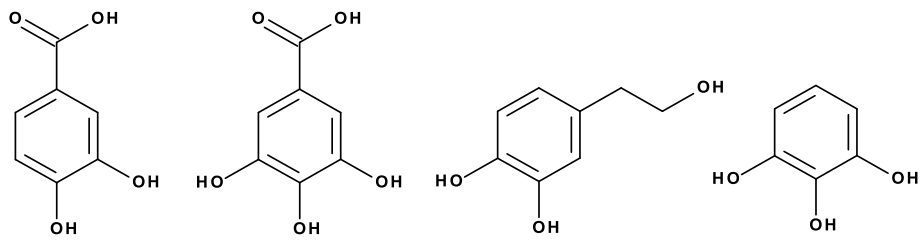
Many polyphenols compounds have been identified as α-syn aggregation inhibitors and their activity on amyloid aggregation is probably related to their antioxidant properties. Polyphenolics compounds comprehend: phenolic acids and catechols,

flavonoids, curcumin and hydroxycinnamic acids derivatives, catecholamines and orcein derivatives [106].

- *Phenolics acids and catechols*

Phenolic acids are widespread in plants and are characterised by various biological properties such as antioxidant, anti-inflammatory, antiviral and anticancer activity [106]. Among them, protocatechuic acid (PCA) and gallic acid (GA) were identified as inhibitors of α-syn aggregation (Chart 3).

PCA **(15)** is a metabolite of anthocyanins able to cross BBB. It proved to be able to inhibit α-syn assembly and to destabilize its fibrillar structures by probably binding to the oligomers or preformed fibrils of α-syn [126]. GA **(16)**, which differs from PCA for a hydroxyl group, can hamper both α-syn and A53T α-syn fibrilization by interacting with α-syn oligomers [127]. Structure-activity relationship studies revealed that the total number of phenolics groups as well as the position of hydroxyl groups are crucial for the activity. It has been demonstrated that phenolic compounds with OH groups at consecutive positions are more active than those having OH groups at non-consecutive positions. Furthermore, the catechol moiety seems to be essential for the inhibitory activity of α-syn aggregation. Indeed, several cathecols like hydroxytyrosol **(17)**, pyrogallol **(18)**, entacapone **(9)** and tolcapone **(8)** are inhibitors of α-syn fibrillization (Chart3) [128,129]. Interestingly, as reported above, tolcapone and entacapone are two inhibitors of COMT, and are used in the therapy of PD. To date, it is not clear if their inhibitory activity on α-syn association contributes to their therapeutic effect but surely their scaffold represents a good starting point for the design of new potential anti-PD drugs because of their dual mechanism of action [128].



15 **16** **17** **18**

- **Chart 3:** Structures of Phenolics acids and Catechols **15**-**18**

- *Flavonoids*

Flavonoids are a subclass of polyphenols showing various biological effects like antioxidant, antibacterial and anticancer activities. Several flavonoids with different substituents, have been tested to probe their ability to modulate α-syn fibrillogenesis. The activity of flavonoids on α-syn assembly is related to their antioxidant properties, however their exact mechanism of actions has not been fully clarified [130,131,132]. Among anti-amyloidogenic flavonoids (Chart4), there is baicalein **(19)**, the major bioactive component of the traditional Chinese herbal medicine *Scutellaria baicalensis*. Baicalein reduces neurotoxicity induced by α-syn probably by forming stable oligomers of α-syn able to inhibit its fibrillization [133]. The most active species as fibrillization inhibitors are its glucuronide baicalin **(20)** and the oxidized form of baicalein **(21)**. In particular, the quinone species reacts with a Lys of the protein by forming a Shiff base near a Tyr residue producing an oligomeric, soluble and non-toxic form of α-syn [134]. Baicalein is also able to disaggregate α-syn fibrils by intercalating its relatively planar structure into the fibrils [135]. Despite these properties, the therapeutic use of baicalein is limited by its poor water solubility and its rapid hepatic metabolism by glucuronidation and sulfation reactions after administration [136,137].

Among natural flavonoids, also quercetin **(22)** showed the ability both to block α-syn aggregation and to disassembly preformed fibrils. In analogy to baicalein **(19)**, the active species of quercetin are its oxidation derivatives. Quercetin interacts covalently with α-syn forming adducts that binds to protein oligomers or monomers hindering fibril growth. Quercetin can cross the BBB, therefore it represents a good candidate for PD therapy [138,139]. Another flavonoid compound with promising anti-amyloidogenic activity, is epigalloctechin-3-gallate (EGCG) **(23)**, a component of green tea, which binds to α-syn producing spherical not toxic oligomers [140,141]. Furthermore, EGCG reduces the affinity of α-syn oligomers towards cell membranes, thus reducing α-syn cytotoxicity and it also able to convert mature fibrils into smaller unstructured adducts with lower cellular toxicity [142,143].

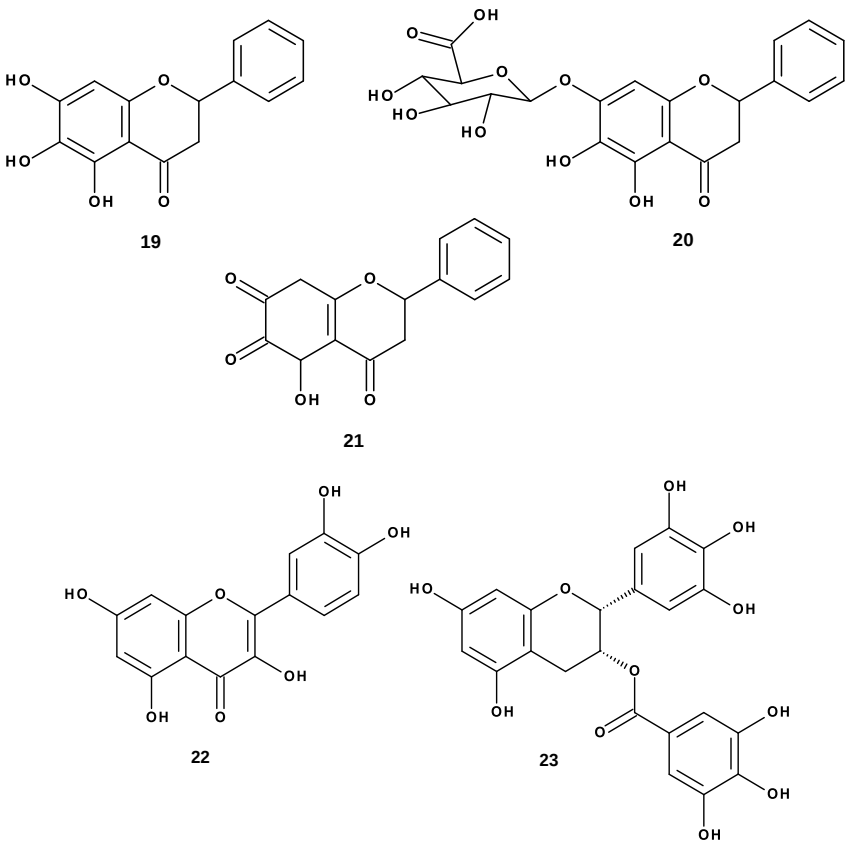


Chart 4: Structures of Flavonoids **19**-**23**

- *Curcumin and hydroxycinnamic acids derivatives*

Curcumin **(24)** is a natural polyphenolic compound isolated from the roots of *Curcuma Longa*. It possesses many biological properties, such as antioxidant, anti-inflammatory, hepatoprotective, antitumor, antifungal, iron-chelation, and neuroprotective properties. Curcumin inhibits the toxic aggregation of α-syn and showed the ability to disaggregate preformed fibrils. Its mechanism of action is still unclear and moreover, its use as drug is limited by its low water solubility and therefore poor bioavailability [106]. In this context, hydroxycinnamic acid (HCA) derivatives (Chart5), that are structurally similar to "half" of curcumin molecule, represents a valid alternative. Indeed, HCA derivatives are soluble in water thus avoids curcumin disadvantages. From a structural point of view, HCA derivatives are characterised by a 4-hydroxyphenylpropanoid skeleton and differ from each other by the substituents present on the phenol ring. Among the natural HCA compounds, the most active in reducing α-syn amyloid aggregation is ferulic acid **(25)** which could be extracted from the raw materials of several plants like *Angelica*

sinensis, *Cimicifuga heracleifolia* and *Lignsticum chuangxiong*. Interestingly, it is also a natural metabolite that can be retrieved in the human blood at concentration up to 140 nM after coffee assumption [144,145]. Experimental data showed that ferulic acid does not interact with the monomeric form of α-syn suggesting that it could bind to the oligomeric or fibril forms. Moreover, ferulic acid proved to be non-cytotoxic allowing to perform more in-depth studies about its potential use for the cure of PD [145]. Additionally, the anti-aggregation activity of some synthetic HCA derivatives bearing different substituents on the aromatic ring, was also investigated. Among them, the 3-methoxy-4acetamidoxycinnamic acid **(26)** proved to be the most active despite it resulted less potent than ferulic acid. Compound **(26)** inhibits α-syn fibrillation in dose-dependent manner probably by interacting with oligomeric or fibrils species [145]. Structure-activity relationship studies concerning this class of compounds revealed that the presence of the aromatic ring is important to establish π-π and hydrophobic interactions with α-syn which hinders the accessibility of the protein for the interactions with other monomers. Furthermore, the presence of hydroxyl groups seems to be important for the anti-aggregation activity probably due to their ability to destabilize the aggregated protein structures [144].

Rosmarinic acid **(27)** is another natural occurrent polyphenol found in various plants such as *Rosmarinus officinalis* (rosemary) from which it was previously isolated [146]. Rosmarinic acid inhibits α-syn oligomerization by interacting with the N-terminal portion of α-syn [147].

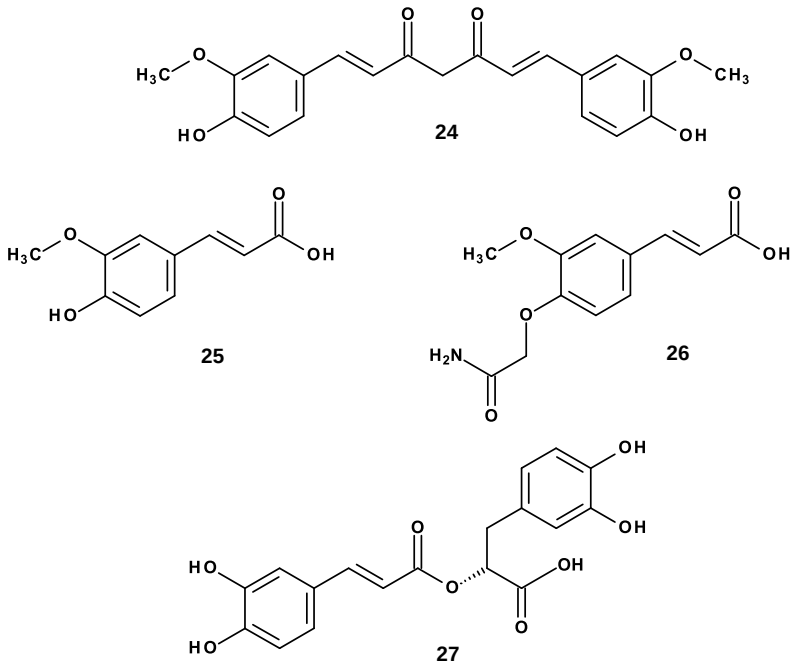


Chart 5: Structures of Curcumin **24** and hydroxycinnamic acids derivatives **25**-**27**

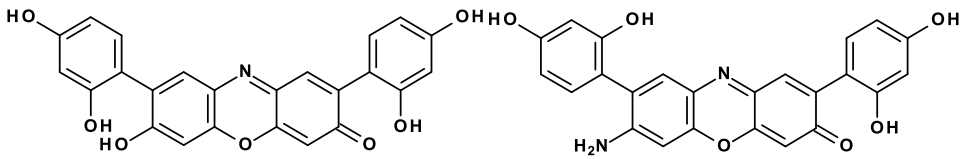
- *Catecholamines*

It has been reported that DA **(3)** is able to modulate α-syn aggregation, but its mechanism is still unclear. It has been proposed that covalent adducts might be formed between a small fraction of α-syn and DA or its oxidized form while the rest of the protein established non-covalent interactions with DA. An indirect mechanism was also proposed according to reactive oxygen species produced by DA, promote the oxidation of α-syn methionine residues by addition of an oxygen atom to sulfur yielding the sulfoxide that is then converted to sulfone. This oxidized form of α-syn is not able to self-associate causing the inhibition of fibrillization of the non-oxidized protein [106].

- *Orcein derivatives*

An alternative strategy to reduce α-syn toxicity is the stabilization of the mature fibrils with the depletion of the toxic oligomeric intermediate species. In this context, the orcein derivative O4 **(28)** (Chart 6) proved to be able to stabilize α-syn fibrils and moreover to inhibit the spreading of misfolded α-syn by seeding [148]. Lacmoid **(29)** (Chart6), another orcein related compound, affects α-syn fibrillation by interacting with the N-terminal region of the protein. This compound can self-

associate forming oligomeric species that, according to experimental data, are responsible for the binding to α-syn [149].



28 29

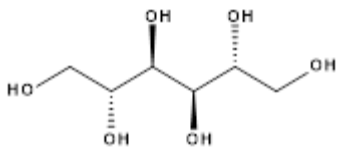
Chart 6: Structures of Orcein derivatives **28**-**29**

3.3.3.4.2 Non-polyphenols

The second class of α-syn aggregation inhibitors comprises non-polyphenols compounds that include molecules belonging to different chemotypes. Most of them are characterised by the presence of at least one aromatic ring or by a planar structure which seems to be important for the inhibitory activity [106].

- *Sugar alcohols*

Mannitol **(30)** (Chart 7) is a sugar alcohol present in fruits and vegetables. It is an osmotic diuretic, and it is also considered as molecular chaperone due to its ability to stabilize protein conformations. Mannitol proved to inhibit the formation of α-syn fibrils and to affect the secondary structure of α-syn oligomers. The α-syn anti-aggregation activity of mannitol was also demonstrated *in vivo* in PD Drosophila or human α-syn transgenic mouse models [150].



30

Chart 7: Structure of mannitol **30**

- *Tetrapyrrole macrocycles*

Tetrapyrrolic systems showed the capability to modulate amyloid aggregation of different proteins including α-syn and to prevent the cytotoxic effect of amyloid

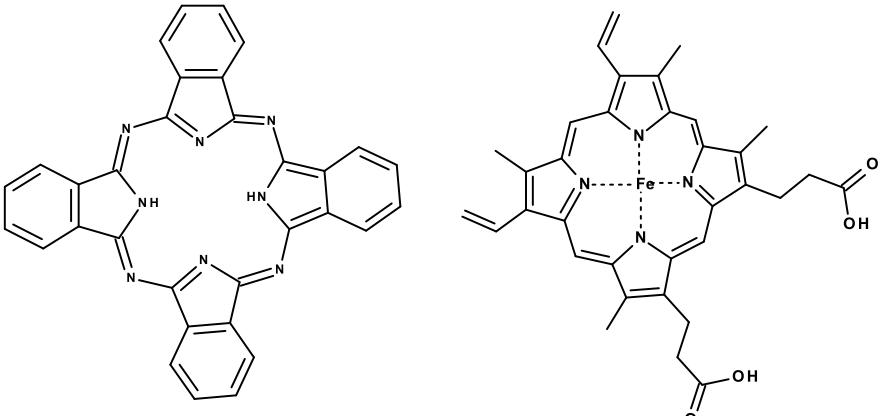
aggregates. Compounds containing these systems are characterized by planarity and π-electron delocalized ring systems. In this context, phthalocyanines and porphyrins are the most investigated as α-syn aggregation inhibitors. Specifically, phthalocyanines interact with the amyloidogenic proteins by establishing aromatic interactions between their aromatic ring system and the aromatic residues of the protein. Structure-activity relationships studies of phthalocyanines, revealed that their anti-aggregation activity relies on the type of metal that is coordinated by the tetrapyrrolic rings; moreover, the number of charged substituents, that mediate the formation of electrostatic interactions with protein charged residues, does not influence the anti-amyloidogenic activity of these derivatives. Phthalocyanines oligomerize through the formation of stacking interactions and their tendency to self-aggregate is correlated with their capability to bind to the target protein exerting the anti-amyloidogenic effect [151].

Phthalocyanine tetrasulfonate (PcTs) **(31)** (Chart 8) is one of the most studied α-syn aggregation inhibitors belonging to this class of compounds. It affects α-syn aggregation by forming non-fibrillar aggregated with α-syn, that are not toxic in cellular studies [152]. NMR studies showed that PcTs binds to the N-terminal region of α-syn monomers at residues Phe4 and Tyr39, and in lesser extent to the NAC domain at Phe94. The binding affinity is higher for the Tyr39 binding site compared to the others [153,154]. These data clearly demonstrated that aromatic interactions are the driving force of ligand-protein binding. Additionally, electrostatic interaction between the sulfonate groups and positively charged amino acid residues contribute to stabilize the α-syn-PcTs interaction [155]. It has been shown that, above critical concentrations, PcTs generates supramolecular species that binds to α-syn. Upon this interaction α-syn changes its conformation forming off-pathway not-pathological oligomers [154].

PcTs presents several advantages as therapeutic agents for the cure of PD: the ability to cross BBB and cellular membranes, low toxicity, chemical versatility and easy functionalization to gain optimized compounds, and the capability to coordinate metal ions that influence its activity. In this respect, the inhibitory activity of α-syn aggregation of PcTs metal complexes have been probed revealing that the nature of the metal ions and consequently the coordination geometry and

the residual positive charge of the metal ions, are important for the activity. Among the studied metal complexes, there are the Cu^{2+}, Ni^{2+}, Zn^{2+} and Al^{3+} complexes. The first three systems bind to the N-terminal region of α-syn while no interaction with α-syn was observed for the PcTs [Al(III)] complex [106,156].

Concerning porphyrins systems, the effect of heme **(32)** (Chart 8) on α-syn amyloid association has been evaluated. In this regard, heme reduces the fibrillation process by forming alternative soluble oligomeric structures which proved to be not toxic in SH-SY5Y cells [157].



31 **32**

Chart 8: Structures of Tetrapyrrole macrocycles **31**-**32**

- *Terpenoids*

Some terpenoids derivatives (Chart 9) such as retinoids and triterpenoids affect α-syn aggregation.

Among the retinoids, vitamin A **(33)** and β-carotene **(34)** inhibit α-syn fibrillation in dose-dependent manner. These compounds possess a high hydrophobic character and probably their activity on α-syn is related to their interaction with the NAC hydrophobic region of α-syn [158].

Concerning triterpenoids, some ginsenosides revealed to be able to modulate α-syn aggregation. Ginsenosides are active ingredients contained in the medicinal herb Ginseng. They are triterpenoid saponins, constituted by a steroid scaffold and their biological effect strongly depends on the type and number of sugar molecules attached at C-3 and C-6 positions. Among the ginsenosides screened as α-syn aggregation inhibitors, ginsenoside Rb1 **(35)** is the most active. According to

spectroscopic studies, this compound stabilizes soluble oligomeric form of α-syn that seems to be non-toxic [159].

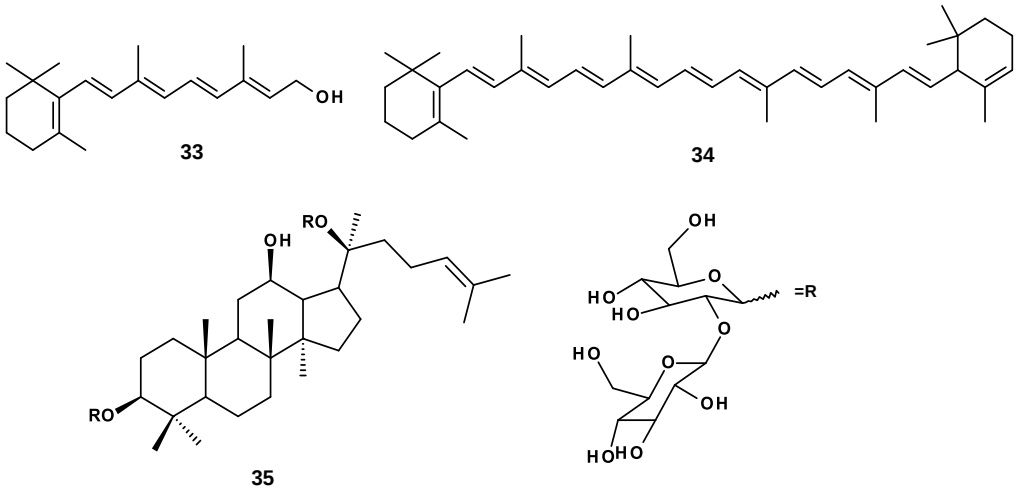


Chart 9: Structures of Terpenoids **33-35**

- *Alkaloids*

Several epidemiological studies showed that smoking is associated with a low incidence of PD. This is related to the ability of the alkaloid nicotine **(36)** (Chart 10), a component of cigarette smoke, to hamper α-syn association. Nicotine interacts with the monomeric form of α-syn and its binding is stereospecific. In particular, the (-)-nicotine binds to the C- and N-terminal domains while the (+)-nicotine interacts with the N-terminal region [160,161].

Epidemiological evidence outlined also that drinking coffee is related to a low risk of developing PD. Caffeine **(37)** (Chart 10), the major constituent of coffee, affects α-syn aggregation through the formation of low toxic aggregates thus explaining its beneficial effect on PD [162].

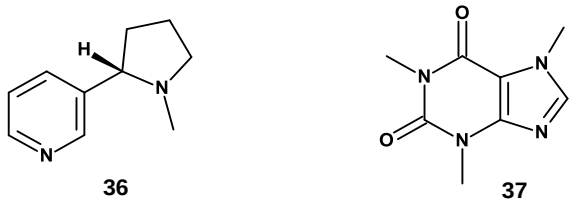


Chart 10: Structures of Alkaloids **36-37**

- *Diazo dyes*

Congo red **(38)** (Chart11), a dye used to detect the presence of amyloid fibrils, is an inhibitor of α-syn fibrillation interacting with N-terminal domain of α-syn. Probably, its mechanism of action is related to its ability to self-aggregate forming small oligomeric species that seem to be responsible of the inhibition of toxic α-syn fibril formation [149].

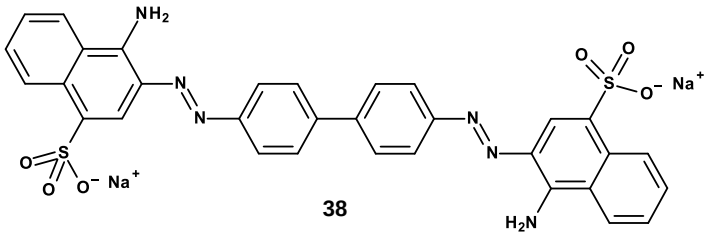
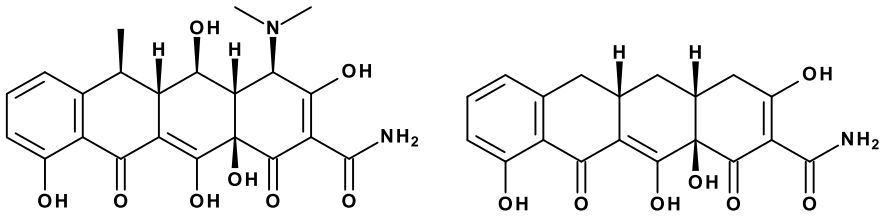


Chart 11: Structures of Congo red **38**

- *Tetracyclines*

Drug repurposing is a widely used strategy to identify new active compounds against the studied target. Recently, preclinical studies revealed that tetracyclines possess neuroprotective properties. In particular, doxycycline (DOX) **(39)** (Chart 12) has shown inhibitory activity against α-syn aggregation and the capability to reduce the loss of dopaminergic neurons in animal models. DOX interferes with α-syn association by converting toxic oligomers in strains characterised by low toxicity and seeding ability. Nonetheless, the antibiotic activity of DOX limits its usage in PD therapy encouraging the design of chemically modified tetracyclines (CMTs) with reduced antibiotic properties. González-Lizárraga *et al.* reported the inhibition of α-syn amyloid aggregation exerted by CMT-3 **(40)** (Chart 12) that have a reduced antimicrobial activity. Similarly to DOX, CMT-3 reshapes α-syn oligomers into not toxic strains that are unable to act as seeds. Differently from DOX, CMT-3 is also able to disrupt preformed α-syn fibrils [163].



39 **40**

Chart 12: Structures of tetracyclines **39**-**40**

- *Synthetic inhibitors*

In the last years, several synthetic compounds bearing different scaffolds were reported in literature as inhibitors of α-syn amyloidogenic aggregation; some of them are displayed in Chart 13.

Through a high throughput screening campaign followed by medicinal chemistry optimization, Wagner *et al.* identified the 3,5 diphenil-pyrazole derivative Anle138b **(41)** as potent α-syn aggregation inhibitor. This compound interacts with the oligomeric species of α-syn, *in vitro* and *in vivo*, modulating their aggregation. Anle138b reduced the accumulation of toxic oligomers, neurodegeneration, and disease progression in three different PD mouse models. It showed no toxicity at therapeutic concentration, good bioavailability and it is also able to penetrate the BBB [164]. Interestingly, this compound proved to be efficacious also after the onset of PD symptoms which is an essential prerequisite for its usage as disease-modifying therapy [165].

The iso-indole pyrimido pyrizine derivative NPT100-18A **(42)** was developed through *de novo in silico* design starting from the C-terminal domain of α-syn. This compound proved to reduce neuronal accumulation of α-syn reducing its cellular toxicity. Its mechanism of action is probably related to its ability to displace α-syn from the membrane, which represents an important step in the oligomerization process, thus haltering the formation of toxic oligomers. *In vivo* experiments on three α-syn transgenic mouse models revealed that NPT100-18A improved motor impairments in a dose-dependent manner [166]. Despite these promising results, the potential therapeutic use of NPT100-18A is limited by its low oral bioavailability and poor BBB penetration. Therefore, lead optimization strategies were applied leading to the development of NPT200-11 characterised by a better

pharmacokinetic profile compared to the lead compound NPT100-18A, while retaining its beneficial effects in α-syn-based animal models. NPT200-11 was effective in reducing α-syn pathology, neurodegeneration and CNS inflammation. A Phase I clinical trial was recently completed for this compound revealing a good tolerability [167].

Pujols *et al.* developed an efficient high-throughput virtual screening strategy that allowed to the authors to efficiently screen a library containing more than 14000 molecules bearing different chemical scaffolds. As results, they identified three small molecules as α-syn aggregation inhibitors, SynuClean-D **(43)** (SC-D) [168], ZPD-2 **(44)** [169] and ZPDm **(45)** [170] which showed the 50%, 80% and 60% of aggregation inhibition respectively, employing a concentration of 70 μM α-syn with 100 μM of compound. From a structural point of view ZPDm could be considered a building block of ZPD2 sharing the 2-nitro-4-(trifluoromethyl)phenyl moiety. All the compounds are characterised by one or two aromatic rings, that probably establish contacts with the hydrophobic regions of α-syn, and the presence of cyano, trifluoromethyl and nitro groups which might be involved in the formation of H-bonds. Despite these similarities, these inhibitors displayed a different mechanism of action. SC-D mainly targets α-syn fibrils whereas ZPD-2 acts on oligomeric species. A still different mechanism was observed for ZPDm which act in the latest stages of the fibrillation process by disrupting large fibrillar species into shorter fibrils or amorphous agglomerate. All the compounds also proved inhibition of the aggregation of α-syn mutants H50Q and A30P responsible for familial PD; furthermore, they prevent α-syn seeded polymerization and showed to reduce α-syn aggregates in *in vivo* experiments performed by using *Caenorhabditis elegans* models of PD.

Toth et al. employed a biophysical screening against monomeric α-syn to screen a library containing more than 90000 small molecules and 23000 fragment molecules resulting in the identification of more than 500 compounds binding α-syn. Among them the most potent was derivative 576755 **(46)** which hampered the formation of oligomeric species protecting the dopaminergic neurons from α-syn toxicity [171].

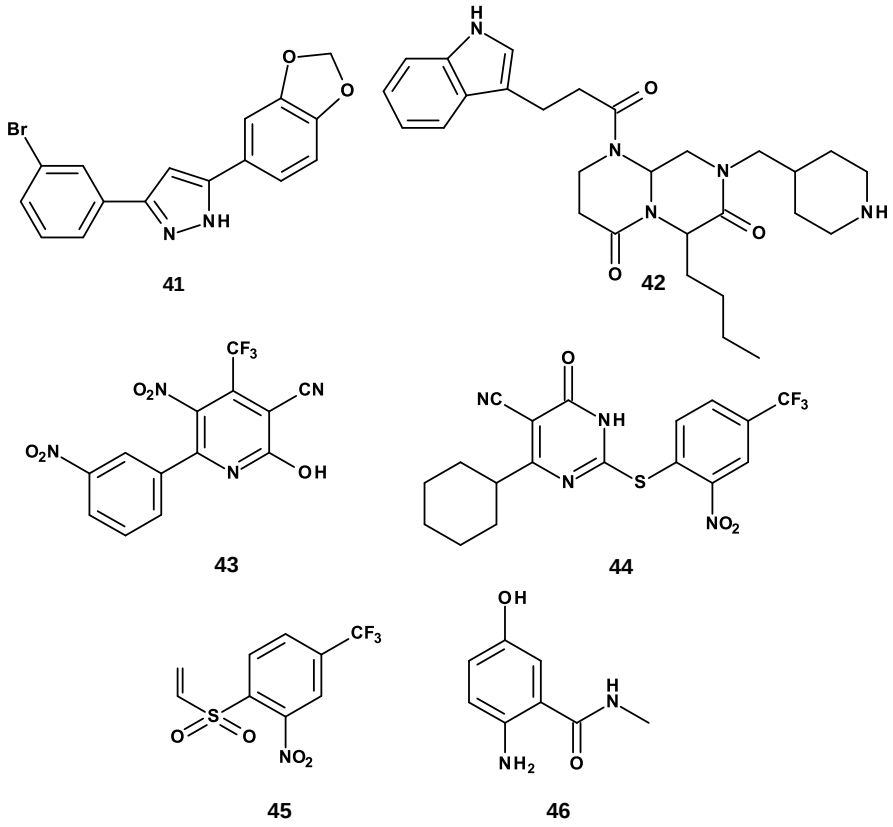


Chart 13: Structures of some Synthetic inhibitors **41-46**

Maqbool at. al reported the synthesis of a small library of diphenyl triazine derivatives for which the inhibition of α-syn aggragation was evaluated *in vitro.* The results showed that four compounds, **(47)**, **(48)**, **(49)** and **(50)** significantly blocked α-syn amyloid aggregation by disassembling preformed fibrils. Cell viability studies revealed that all the compounds were not toxic, except A2 for which a moderate cytotoxicity was observed [172].

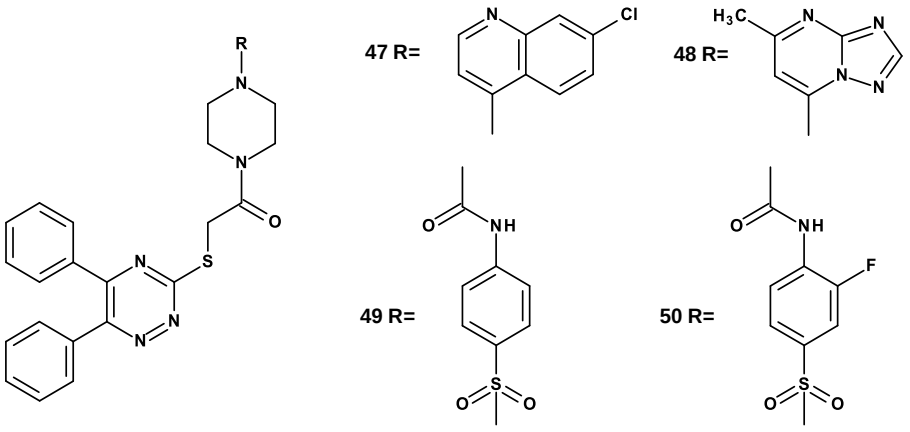


Chart 14: Structures of diphenyl triazine derivatives **47-50**

Chapter 4

Pharmacophore-based discovery of new α-syn aggregation inhibitors

4.1 Introduction

The inhibition of α-syn aggregation represents an attractive strategy for the development of novel anti-Parkinson's agents. In this chapter, I described the generation of a ligand-based pharmacophore model used as query for the VS of two 3D chemical libraries. The capability of the hits selected from the VS to halt α-syn aggregation was probed *in vitro* leading to the identification of a new hit compound **53** that was used for the design of a new series of small molecules. The resulting derivatives were synthesised and their activity was evaluated. Finally, the binding mode of this new class of α-syn aggregation inhibitors was elucidated by docking studies.

The content of this chapter has been object of one publication and several communications:

- ***Serena Vittorio**, Ilenia Adornato, Rosaria Gitto, Samuel Peña-Díaz, Salvador Ventura and Laura De Luca. Rational design of small molecules able to inhibit α-synuclein amyloid aggregation for the treatment of Parkinson's disease. Journal of Enzyme Inhibition and Medicinal Chemistry, 2020, 35, 1727-1735.*

- ***Serena Vittorio**, Rosaria Gitto, Samuel Peña-Díaz, Salvador Ventura and Laura De Luca., Pharmacophore based discovery of new α-synuclein aggregation inhibitors. Workshop of Sicily Section of the Italian Chemical Society (SCI), 3 December 2020 (Poster).*

- ***Serena Vittorio**, Ilenia Adornato, Rosaria Gitto, Samuel Peña-Díaz, Salvador Ventura and Laura De Luca. Computational and synthetic approaches for the development of α-synuclein aggregation inhibitors for the treatment of Parkinson's disease. Italian Young Medicinal Chemistry Virtual Meeting, 22-24 July 2020 (Poster).*

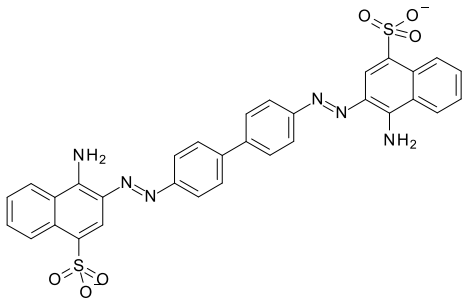
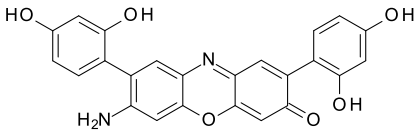
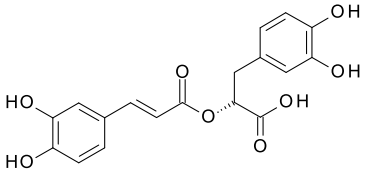
- ***Serena Vittorio**, Rosaria Gitto, Samuel Peña-Díaz, Salvador Ventura and Laura De Luca. In silico screening for the discovery of new α-synuclein aggregation inhibitors. Merck Young Chemists' Symposium 2019, 25-27 November 2019, Rimini, Italy (Flash + poster presentation). **This communication was awarded by Merck as best flash presentation.***

4.2 Results and discussion

4.2.1 Pharmacophore generation, virtual screening and biological evaluation

A ligand-based pharmacophore model was built by means of LigandScout V4.4 [173] software in order to retrieve small molecules able to inhibit α-syn aggregation. For this purpose, we employed as training set (TS) three compounds (Table1) whose interaction with the N-terminal domain of α-syn was experimentally validated. These compounds were selected from literature [106] and include both polyphenolic and non-polyphenolics inhibitors.

Table1. Chemical structure of the α-syn aggregation inhibitors employed as TS with the corresponding fit-score values.

NAME	STRUCTURE	FIT-SCORE
Congo Red **38**		56.49
Lacmoid **29**		57.23
Rosmarinic Acid **27**		57.36

To construct the model, we account only for the features common to the TS molecules thus obtaining a shared features pharmacophore model. Ten pharmacophoric hypothesis were generated by LigandScout and the one with the highest score was selected for the following virtual screening studies. This model was composed by one aromatic features, two H-bond acceptors, two hydrophobic features and forty-five excluded volumes (Figure7).

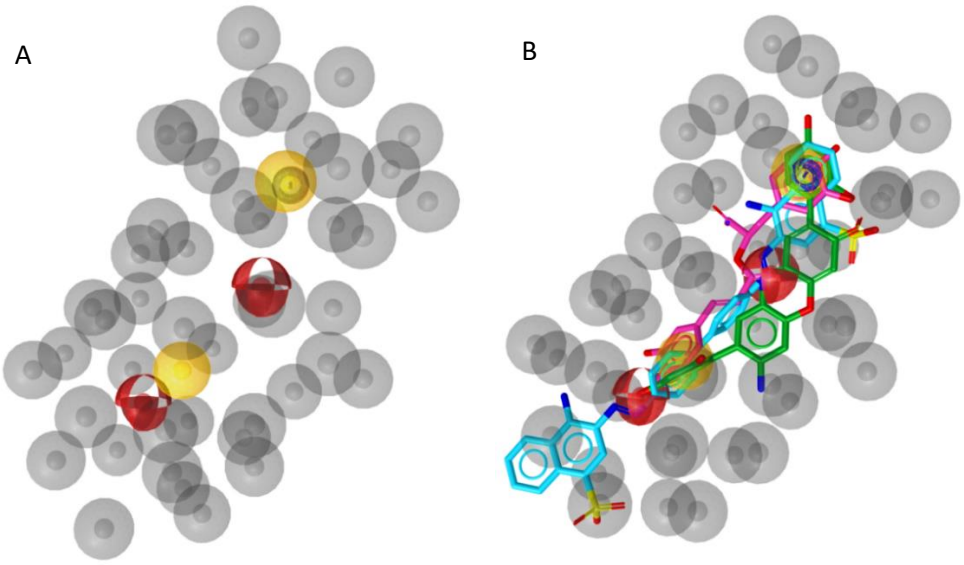


Figure 7. A) Selected ligand-based pharmacophore model. The pharmacophoric features are colour coded: yellow spheres represent hydrophobic features; red spheres represent H-bond acceptors while blue circles represent aromatic features. Excluded volumes are represented as grey spheres. B) TS molecules superimposed with the ligand-based pharmacophore model. Compound **38** is displayed as cyan sticks, compound **29** as green sticks whereas **27** as magenta sticks.

The so generated model was employed as filter to screen two 3D databases: i) the *in-house* 3D library CHIME, which collects 1329 small molecules designed and synthesized by our research group over the years, and ii) the MyriaScreen Diversity Library II, containing 10,000 drug-like compounds (https://www.sigmaaldrich.com/chemistry/chemistryservices/high-throughput-screening/screening-compounds.html).

The search performed on CHIME database resulted in three hits **51a-c** sharing a very similar structure (Figure8); therefore, we selected only 2-[4-[(4-fluorophenyl)methyl]-1-piperidyl]-1-(6-methoxy-1H-indol-3-yl)ethenone **51a**, having the highest pharmacophore fit-score (56.11), as representative hit retrieved from the virtual screening of CHIME library.

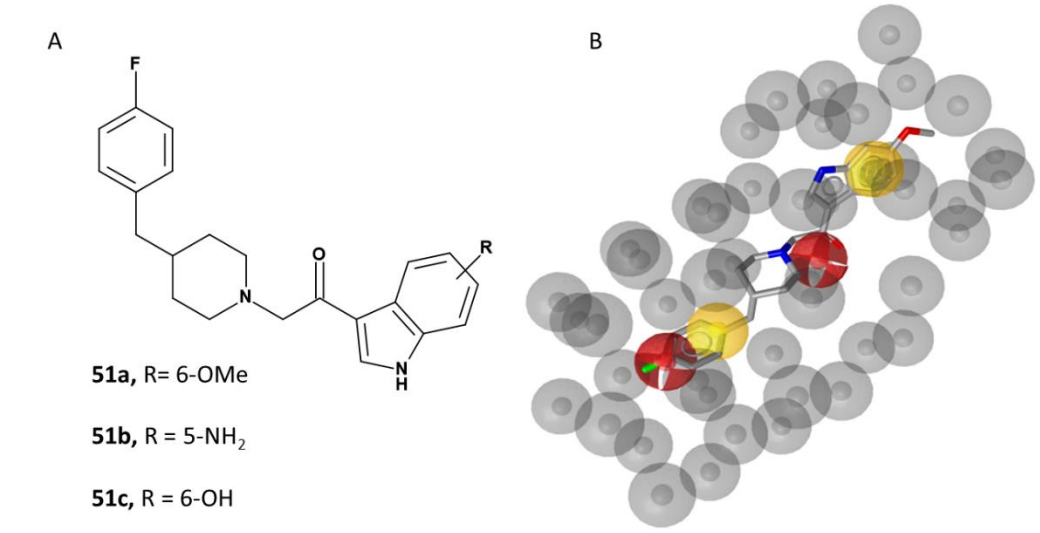


Figure8. A) Chemical structures of compounds **51a-c** obtained from the virtual screening of CHIME database. B) Compound **51a** superimposed to the pharmacophore model. Compound 1a is represented as grey sticks.

The virtual screening carried out on MaryaScreen Diversity library II conducted to the identification of 113 hits. Among them, three compounds (Figure9), 3-[5-[(4-methoxyphenyl)methylsulfanyl]-4-methyl-1,2,4-triazol-3-yl]pyridine **(52)**, 3-(cinnamylsulfanyl)-5–(4-pyridinyl)-1,2,4-triazol-4-amine **(53)** and 3-(3-chloro-4-fluoro-anilino)-1-(2-naphthyl)propan-1-one **(54)**, were selected on the basis of the following criteria: i) pharmacophore fit-score>57 and ii) commercial availability from the supplier. Specifically, **(52)**, **(53)** and **(54)** showed a fit-score value respectively of 57.53, 57.30 and 57.83.

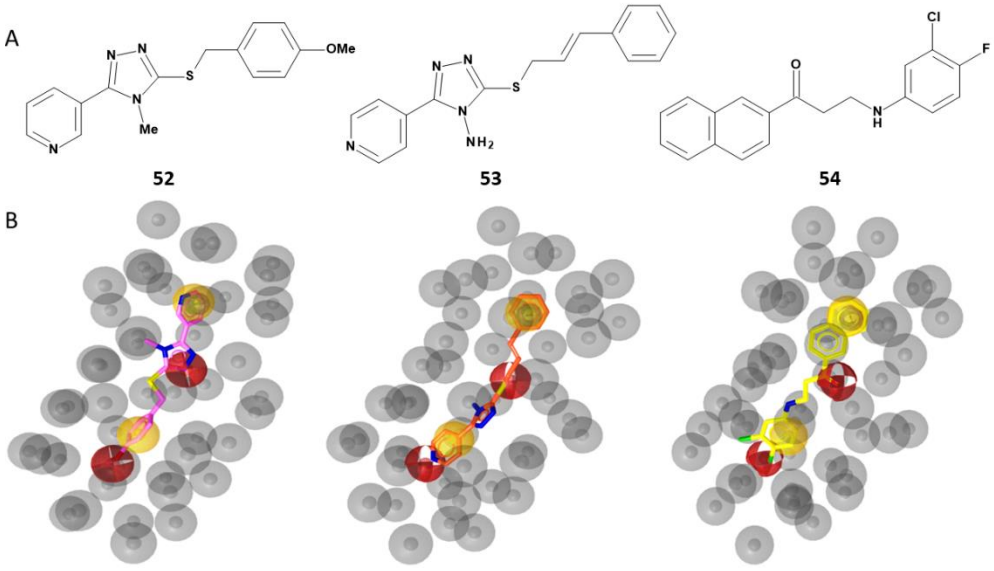


Figure9. A) Chemical structures of compounds **52**, **53** and **54** selected from the MaryaScreen Diversity Library II. B) Compounds **52** (pink sticks), **53** (orange sticks) and **54** (yellow sticks) aligned to the ligand-based pharmacophore model.

For each compound, the conformity to Lipinski's rule and the presence of PAINS were computationally predicted by using the online web tool SwissADME (http://www.swissadme.ch/). The results revealed that all the selected hits respect Lipinski's rule and no PAINS were detected. Encouraged by the *in silico* results, compound **(51a)** was synthesized in the laboratory of Prof. Gitto of University of Messina by using a previously reported synthetic route [174] while compounds **(52-54)** were purchased from Sigma Aldrich (https://www.sigmaaldrich.com/italy.html).

A collaboration with Prof. Salvador Ventura of the University of Barcelona allowed to test compounds **(51a**-**54)**. Specifically, the α-syn anti-aggregation activity of these compounds was probed *in vitro* applying a previously reported robust screening protocol [175] employing the inhibitor SC-D **(43)** as reference. The incubation of 70 μM of α-syn in presence of 100 μM of the compounds led to a decrease of Thioflavin-T (Th-t) fluorescence emission up to 10, 26, 31 and 10% for derivatives **(51a), (52), (53)** and **(54)**, respectively. The results showed that all compounds affected the kinetic constants (Figure10A). In particular, derivative **(52)** decreased the homogeneous nucleation rate constant (k_b= 0.01591) if compared to the control (k_b= 0.03234). Conversely, compound **(53)** affected the autocatalytic rate constant (k_a= 0.1892 h^{-1}) when compared to the untreated sample (k_a= 0.2037 h^{-1}). Light scattering measurements were carried out at 300 nm at the end-point of the reaction highlighting a diminution of a 13 and 43% of aggregates in presence of derivatives **(52)** and **(53)**, respectively (Figure10B). Moreover, the samples of the most active inhibitor **(53)** were also analysed by using transmission electron microscopy (TEM) to further evaluate its inhibitory activity on α-syn aggregation. As a result, TEM images showed a substantial reduction of the quantity of fibrils species in the presence of compound **(53)** (Figure10C) compared to the control (Figure10D).

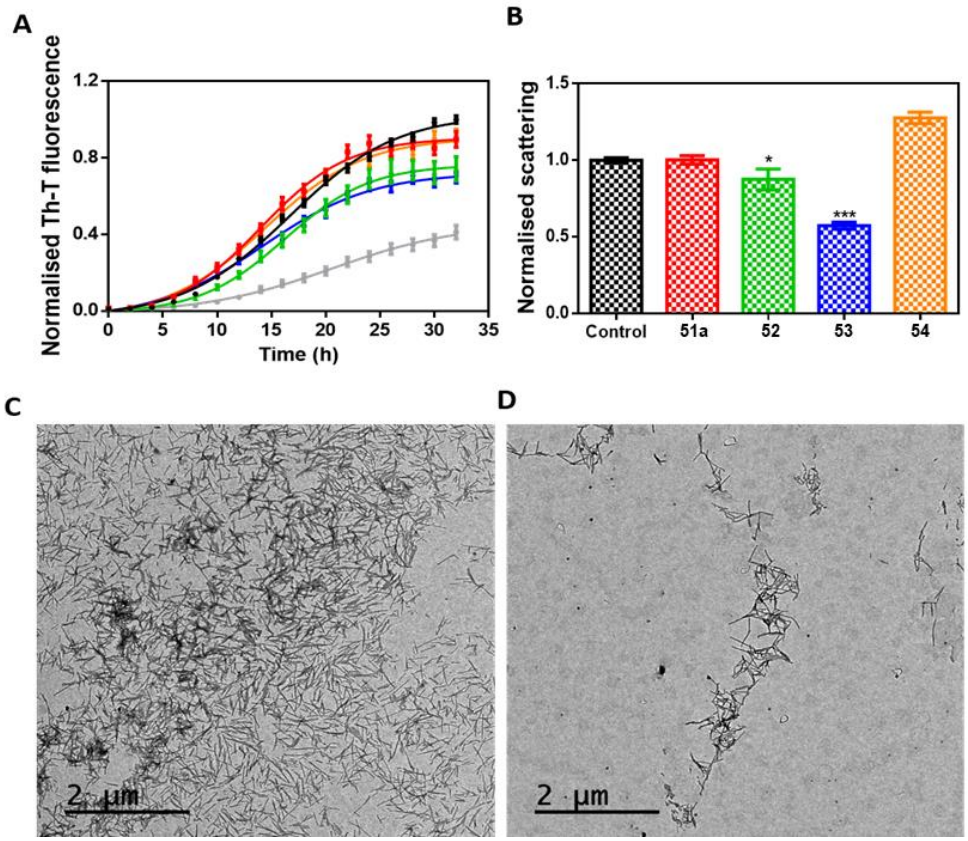


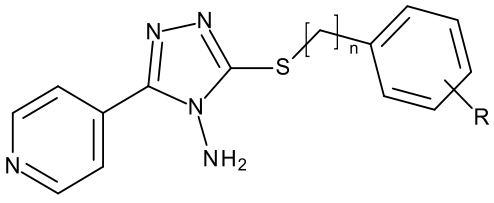
Figure10. A) Aggregation kinetic of a-syn in absence (black) and presence of 100 mM of compound **51a** (red), compound **52** (green), compound **53** (blue), compound **54** (orange) and and SynuClean-D **43** (grey), followed by Th-T fluorescence emission. B) Light-scattering measurements at 300 nm in the absence (black) and presence of 100 mM of compound **51a** (red), compound **52** (green), compound **53** (blue) and compound **54** (orange). Representative TEM images of untreated (C) and compound **53** treated samples (D). Th-T fluorescence is plotted as normalised means. Final points were obtained at 48 h. Error bars are represented as SE of mean values; ** $p<0.01$ and *** $p<0.001$.

Among the studied hits, compound **(53)** proved to be the most active compounds as displayed by *in vitro* studies. Compound **(53)** bears a pyridinyl-triazole moiety like compound **(52)** in which a nitrogen of the triazole ring is bound to a methyl group whereas an amino group is present at the same position in **(53)**. On the basis of the biological screening results, we hypothesized that the amino group might be involved in crucial interactions with α-syn, significant for the inhibition of its aggregation.

4.2.2 Design and synthesis of new derivatives of compound 53

Prompted by the above reported resulted, compound **(53)** was considered as lead for the design of a new series of potential inhibitors of α-syn association

(Scheme1). Specifically, the pyridinyl-triazole portion was maintained whereas the length of the linker connecting the sulphur atom to the phenyl ring was shortened. Furthermore, we introduced in meta and in para positions of the phenyl ring, various electron withdrawing (F and Cl) and electron donating substituents (CH_3) to probe their influence on α-syn amyloid aggregation.



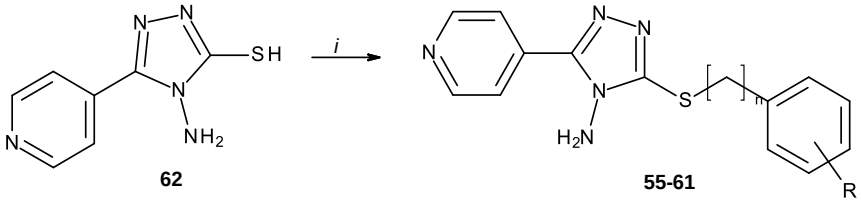
55-61

	n	R
55	1	H
56	2	H
57	1	4-F
58	1	3-F
59	1	4-CH_3
60	1	3-CH_3
61	1	4-Cl

Scheme1. New pyridinyl-triazole derivatives inspired by compound **53**

Each designed compound proved to respect Lipinski's rule and no PAINS were detected as predicted by the online platform SwissADME.

Compounds **(55-61)** were obtained by following a simple synthetic approach as displayed in Scheme2. The 4-amino-5–(4-pyridinyl)-4H-1,2,4-triazole-3-thiol **(62)** was coupled with the suitable benzyl bromide in alkaline medium at room temperature.



62 **55-61**

Scheme2. *Reagents and conditions*: *i*) $Ar(CH_2)_nBr$, NaOH, MeOH, rt.

4.2.3 Biological screening of derivatives 55-61

The ability of the synthesized compounds **(55-61)** to inhibit α-syn aggregation was studied *in vitro* following the previously used protocol. The incubation of α-syn in presence of the different derivatives at 0,70:1 α-syn: compound molar ratio, showed that compounds **(55), (58), (59)** and **(61)** reduced, at the end of the reaction, the formation of Th-t positive structures by 33, 15, 19 and 29%

respectively (Figure11A and B). Light scattering measurements, executed at 300 nm, revealed that **(55)**, **(56)**, **(57)**, **(58)**, **(59)** and **(60)** decreased the level of aggregates up to a 23, 16, 26, 14, 24, 17%, respectively (Figure11C). TEM analysis confirm that only compounds **(55)** (Figure11E), **(58)** (Figure11F), **(59)** (Figure9G) and **(61)** (Figure9H) reduced the amount of amyloid fibrils if compared to the control.

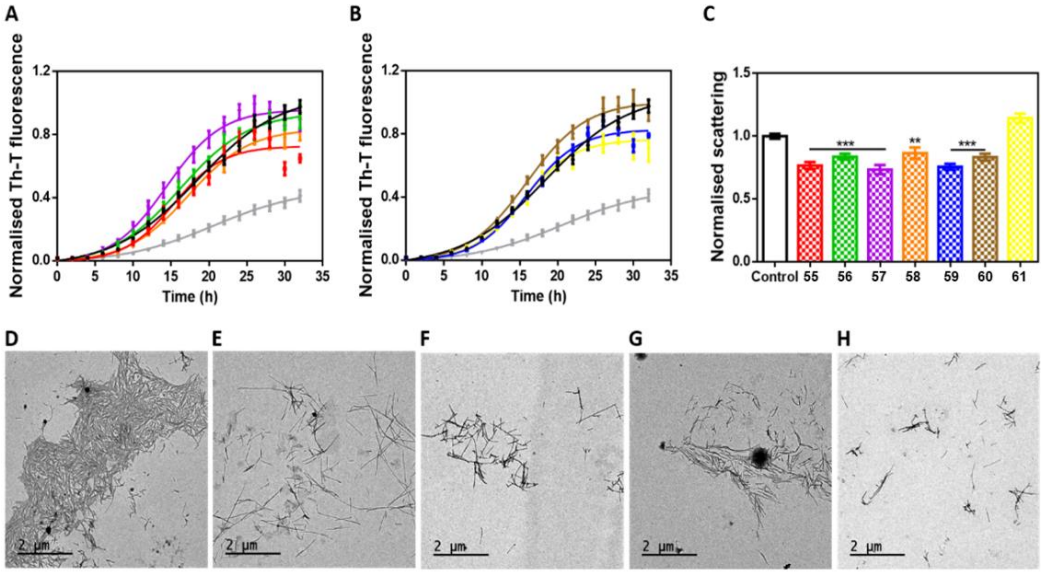


Figure 11. A and B) Aggregation kinetics of α-syn in the absence (black) and presence of 100 µM of compounds **55** (red), **56** (green), **57** (violet), **58** (orange), **59** (blue), **60** (brown), **61** (yellow) and SynuClean-D **43** (grey), followed by Th-T fluorescence emission. C) Light-scattering measurements at 300 nm in the absence (black) and presence of 100 µM of compounds **55** (red), **56** (green), **57** (violet), **58** (orange), **59** (blue), **60** (brown) and **61** (yellow). D to G) Representative TEM images of control (D) and **55** (E), **58** (F), **59** (G) and **61** (H) treated samples. Th-T fluorescence is plotted as normalized means. Final points were obtained at 48 h. Error bars are represented as SE of mean values; ** $p < 0.01$ and *** $p < 0.001$.

4.2.4 Docking studies

The binding mode of the active compounds **(53)**, **(55)**, **(58)**, **(59)** and **(61)** was investigated by molecular docking studies performed by means of Autodock 4.2 software [176], by using the NMR structure of α-syn fibrils (PDB code 2N0A) as three dimensional coordinates [177]. Taking into account that ligands that bind to the N-terminal domain of α-syn were used for building the pharmacophore model, we speculated that these new inhibitors could interact with the same protein portion. Therefore, this region of the protein was defined as search space in our docking simulation. The results highlighted that all the compounds bind to the same

binding site situated between the N-terminal portion and the NAC domain and formed by residues Ala53, Val55, Thr59, Glu61, Thr72, Gly73 and Val74 (Figure12A). Interestingly, the same binding region was retrieved by docking studies for the TS compound **(29)** (Figure12B).

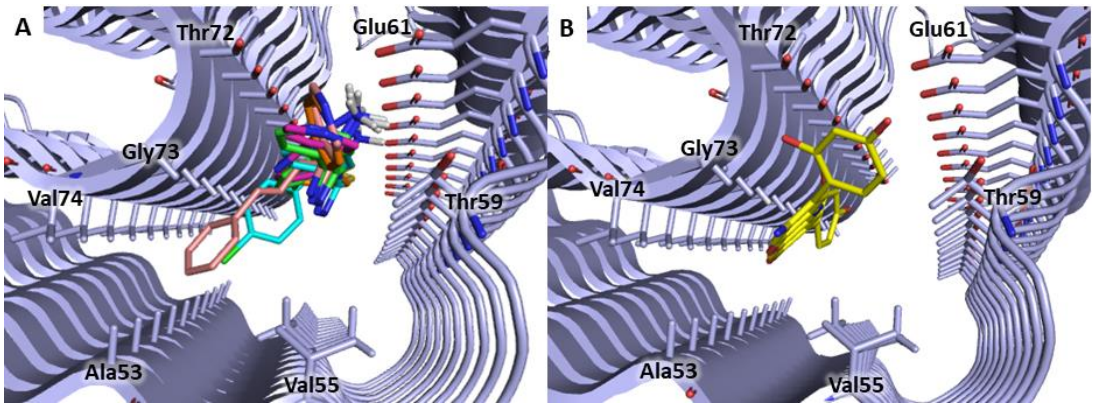


Figure12. Binding site identified by molecular docking simulation. A) Compounds **(53)**, **(55)**, **(58)**, **(59)** and **(61)** into the identified site. B) Docking pose of the TS compound **(29)**. The image was created by using PyMOL software (https://pymol.org).

As displayed in Figure13, compound **(53)** might interact with α-syn by forming a salt bridge with Glu61 through the amino group and π-anionic interactions between Glu61B and the pyridine ring. The cinnamyl portion is accommodated between Ala53, Gly73 and Val74 and could establish hydrophobic contacts with Ala53. Moreover, van der Waals interactions might occur with Val55, Thr59, Thr72, Gly73 and Val74.

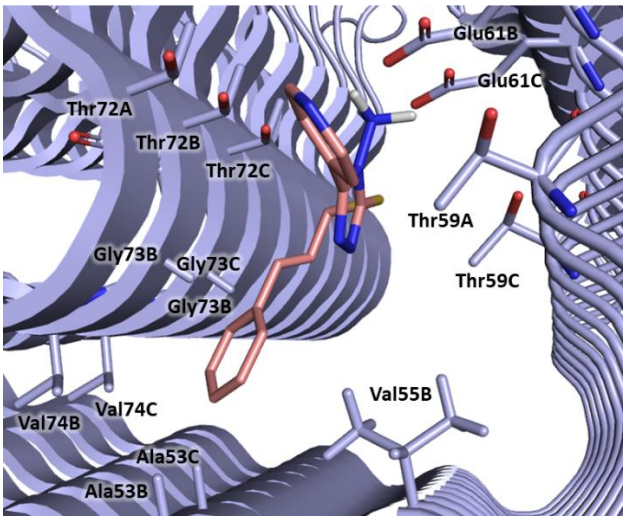


Figure13. Plausible binding mode of compound **53** (pink sticks). The amino acid residues involved in the interactions with the inhibitor are represented as light blue sticks. The image was created by using PyMOL software (https://pymol.org).

Compounds **(55)**, **(58)**, **(59)** and **(61)** displayed a slightly different binding orientation if compared to the lead compound **(53)**, probably because of the shorter and more flexible chain between the sulphur atom and the phenyl ring. The docking studies (Figure14) highlighted that all the compounds might engage a salt bridge with Glu61. A pi-anion interaction between Glu61B and the pyridine moiety was detected for compounds **(55)**, **(58)** and **(59)**. In addition, inhibitor **(55)** could form a H-bond between the sulphur atom and the side chain of Thr72C and pi-sigma interactions with Thr72C (Figure14A). Instead, compound **(58)** might establish a halogen bond through the fluorine atom and Thr72D (Figure14B). Finally, derivative **(59)** might establish pi sigma interactions with Thr72C (Figure14C), whereas compound **(61)** could engage van der Waals interactions with Val55C (Figure14D).

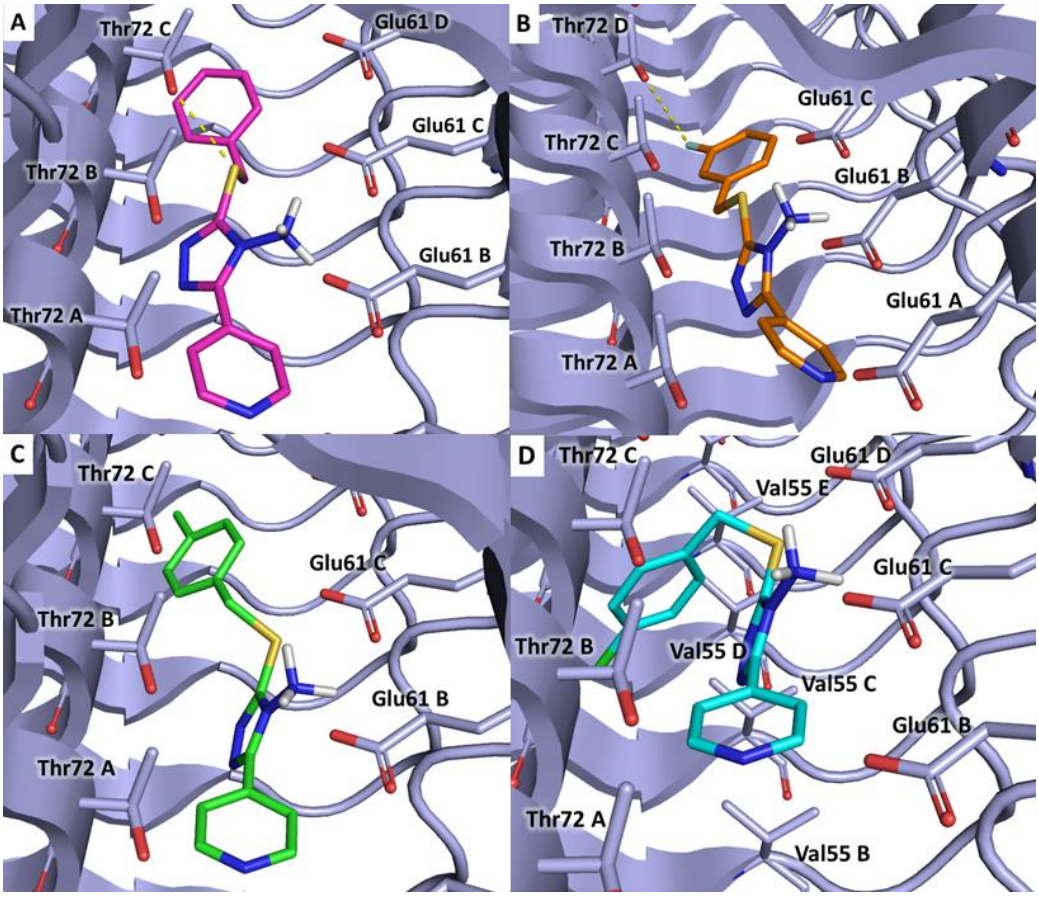


Figure 14. Plausible binding modes for compound **55** (magenta stick, panel A), **58** (orange stick, panel B), **59** (green stick, panel C) and **61** (cyan stick, panel D). The interacting residues of the binding site are highlightes as light blues stick. The images were created by PyMOL software (https://pymol.org).

4.3 Materials and methods

4.3.1 Pharmacophore modelling and virtual screening

LigandScout V4.4 [173] was employed for pharmacophore construction and virtual screening. Three small molecules that bind the N-terminal domain of α-syn were selected from literature [106] and used as TS. A shared-featured pharmacophore model was built applying the default settings.

All the virtual screening experiments were performed by using the "Get the best matching conformation" option as retrieval mode.

4.3.2 Molecular docking

Molecular docking studies were carried out by Autodock 4.2 [176] software employing the solid-state NMR of α-syn fibrils (PDB code 2N0A) [177]. The structure presents a central part with a greek-key topology and terminal flexible loops. A grid box with dimension of 126 x 126 x 126 Å and center x= 97.487, y= 148.695 and z= -34.111, was set to include the amino acid residues of the N-terminal domain, excluding those present in the unstructured flexible loops. Ligand structures were constructed by Vega ZZ software and energy minimized by following a conjugate gradient minimization by AMMP calculation as implemented in VEGAZZ program [178]. Autodock force field (based on Amber) was used for the calculation and Gasteiger charges were assigned to the protein. The Lamarckian Genetic Algorithm was used to generate ten protein–ligand binding poses for each molecule by using the default settings. The highest scored docking pose was chosen for analysis and representation. The visualisation of docking results was performed by using PyMOL software while the analysis of the putative ligand-protein interactions was performed by Discovery Studio Visualizer [179].

Chapter 5

Identification of α-syn aggregation inhibitors by similarity-based VS

5.1 Introduction

In this Chapter, another ligand-based VS approach was applied to identify small molecules able to inhibit α-syn aggregation. In particular, by using the web-tool Swiss Similarity, we performed a similarity-based VS employing as query compound SC-D **43**, identified by Pujols *et al.,* which showed a good inhibitory activity on α-syn aggregation both *in vitro* and *in vivo*. The resulting molecules were filtered according to their druglikeness properties evaluated in silico by using SwissADME. Their putative binding mode was investigated by docking studies by means of Autodock and Gold software. Specifically, docking runs were initially performed in order to identify potential binding sites on α-syn. Then the compounds whose binding site was predicted in consensus by both software were re-docked in the predicted site by using both Autodock and Gold. The hits showing similar binding modes in the results obtained from the two docking tools were chosen as potential inhibitors of α-syn amyloid aggregation. The selected compounds were purchased and their biological evaluation is in progress.

5.1 Results and discussion

5.2.1 Similarity-based virtual screening

In order to identify compounds able to inhibit the amyloid aggregation of α-syn, a similarity-based virtual campaign was performed by means of the online tool SwissSimilarity [180] (http://www.swisssimilarity.ch/).

Among the inhibitors reported in literature, compound SC-D **43** (Chart 13 in paragraph 3.3.3.4.2) showed a good inhibitory activity on α-syn aggregation both *in vitro* and in *in vivo* models, thus representing a promising candidate as therapeutic agents for the treatment of PD [168]. For this reason, SC-D was chosen as query for the similarity-based VS. The search was performed on Sigma-Aldrich database containing 65,000 small molecules; three methods were applied to evaluate the similarity: FP2 fingerprints, electroshape and spectrophores.

The VS runs led to the identification of 370 molecules: one obtained through the FP2 fingerprints similarity measure, one by electroshape similarity and 368 by spectrophores method. For each resulted compound, the ADME profile and the druglikeness properties were evaluated by SwissADME platform. Particularly, to assess the druglikeness, the following criteria were considered: Lipinski's rules [181], Veber [182], Brenk [183] and the absence of PAINS. The application of these filters led to the selection of 47 compounds. Among them, some hits share a very similar structure, therefore, we decided to choose those characterised by the highest score thus obtaining 38 compounds.

5.2.2 Molecular docking studies

The putative binding mode of the selected molecules on α-syn fibrils, was probed by docking simulation using two software, Autodock 4.2 [176] and Gold V2020.2.0 [184]. Considering that α-syn inhibitors could interact at different protein domains, we selected as search space the region comprised between Lys34 and Lys96 which covers the major β-sheet structures (for more details see Material and Methods section, par. 5.3.3).

The docking analysis performed with Autodock suite allowed us to identify 8 different potential bind sites on α-syn depicted in Figure15.

- Binding site 1 (Figure15A) is located in the NAC domain and is lined by Ala53, Val55, Thr59, Glu61, Thr72 and Val74.
- Binding site 2 is situated in the N-terminal region and contains Met1, Phe4, Lys6, Lys43 and Lys45. (Figure15B)
- Binding sites 3 and 4 are both in the C-terminal region of α-syn; the first is lined by Glu83, Gly84 and Ala85 (Figure15C) while the second is formed by Gly86, Ile88, Phe94 and Lys86 (Figure15D).
- Binding site 5 is situated in the N-terminal domain and comprehend residues Tyr39, Ser42 and Thr44 (Figure15E),
- Binding site 6 is in the NAC domain lined by Val49, Gly51, Val74, Thr75, Ala76 and Val77 (Figure15F).
- Binding sites 7 and 8 are both located in the N-terminal region of α-syn; the former is formed by Lys34, Glu35, Gly36 and Val37 (Figure15G) while the latter comprises Met1, Asp2, Val48 and His50 (Figure15H).

Instead, the docking studies performed by Gold highlighted four putative binding sites that correspond to binding sites 1, 2, 3 and 4 described previously.

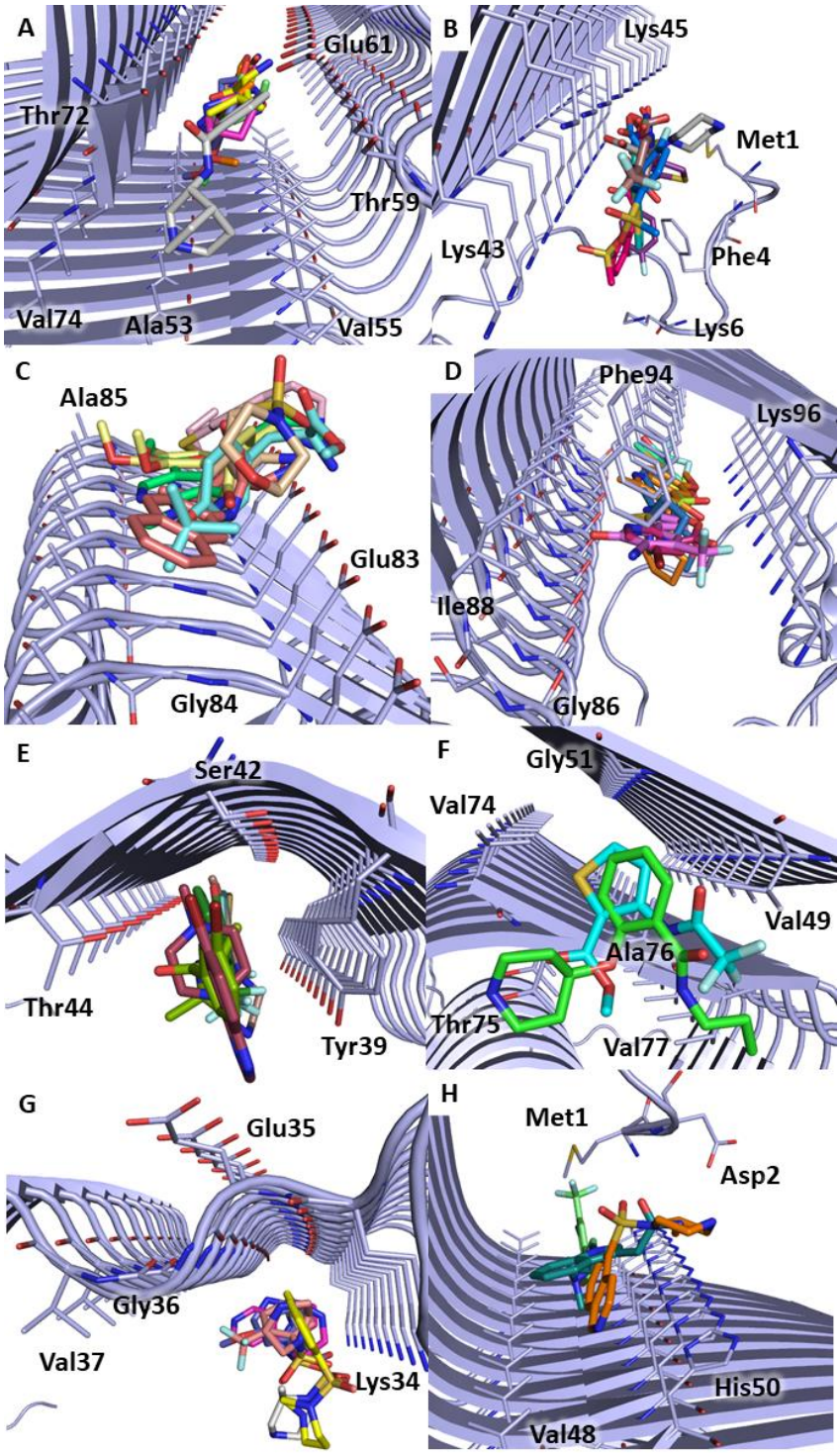


Figure15. Depiction of the binding sites identified through docking studies. The amino acid residues forming the binding site are highlighted as lightblue sticks while the ligands binding to the identified site, are represented in sticks. A) Binding site1, B) Binding site 2, C) Binding site 3, D) Binding site 4, E) Binding site 5, F) Binding site 6, G) Binding site 7, H) Binding site 8. The image was created by PyMOL software (www.pymol.org).

The results obtained from the two docking studies executed with Autodock and Gold were combined leading to the identification of eleven molecules (Chart15) whose binding sites were predicted in consensus by both software. In particular,

we chose compounds **(62-70)** that showed the ability to occupy binding site 1 and **(71-72)** which bind to site 2.

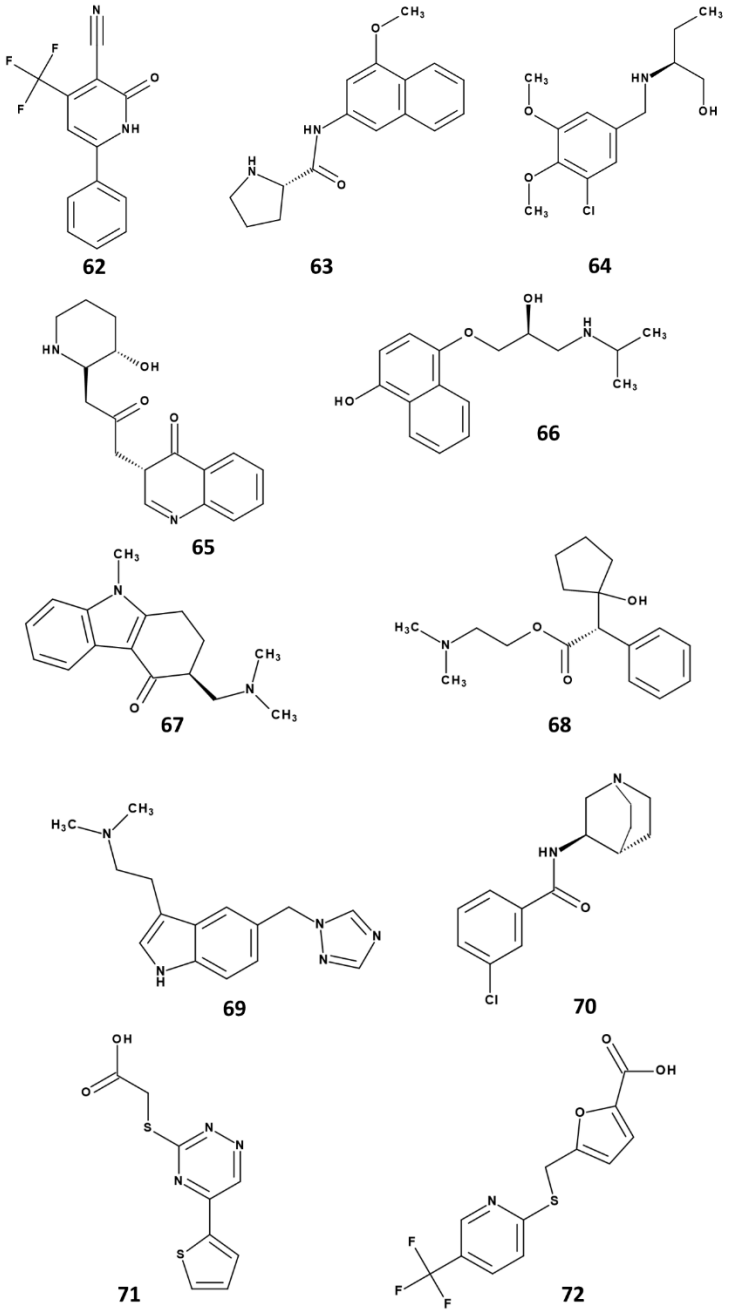


Chart15. Compounds whose binding site was predicted in consensus by Autodock and Gold docking studies.

In order to refine the docking procedure, these molecules were redocked restricting the search space to the binding site identified in the previous docking simulations. The obtained poses were analysed and five compounds, **(64)**, **(65)**, **(67)**, **(69)** and **(72)**, that showed similar binding modes in the results gained from

Autodock and Gold were selected as potential inhibitors of α-syn amyloid aggregation. Specifically, compounds **(64)**, **(65)**, **(67)** and **(69)** were selected for the binding site 1, whereas only compound **(72)** for the binding site 2.

In Figure16, the plausible binding modes of compounds **(64)**, **(65)**, **(67)** and **(69)** in the binding site 1 are displayed.

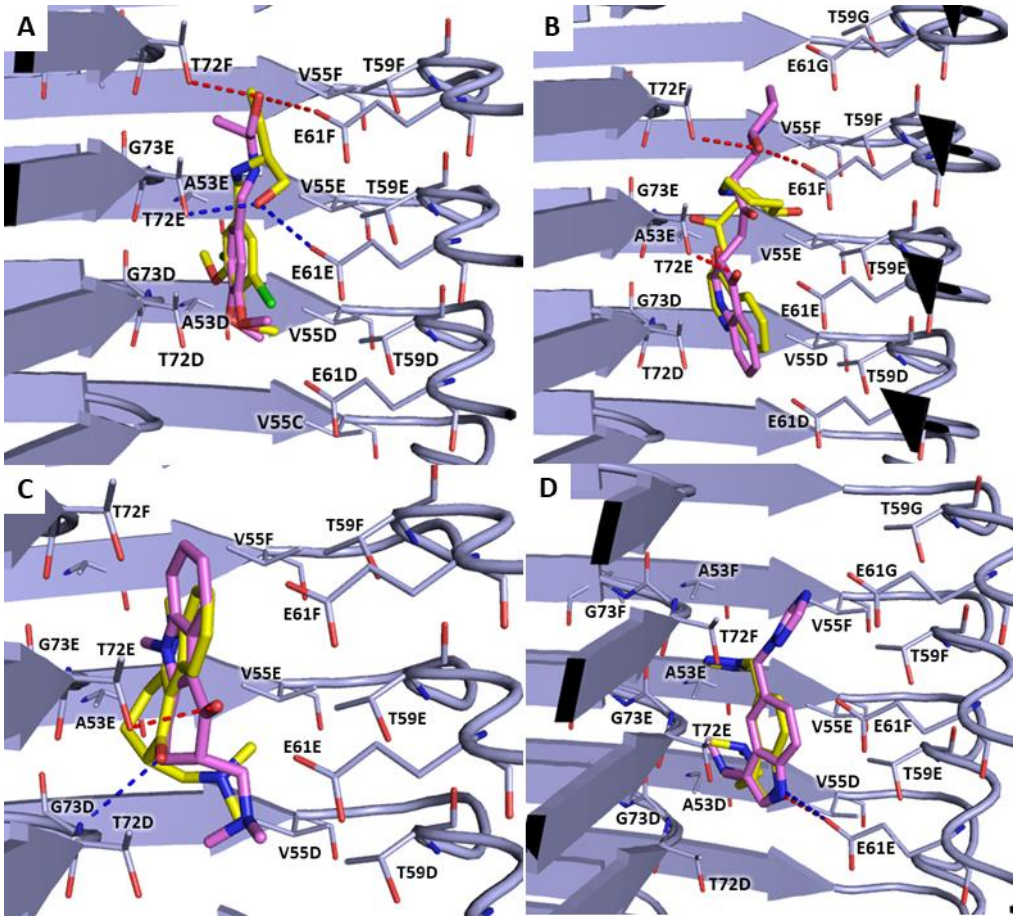


Figure16 Plausible binding poses of compounds **64** (A), **65** (B), **67** (C) and **69** (D) predicted by Gold (violet sticks) and Autodock (yellow sticks). Hydrogen bond interactions are represented as red dotted lines for the poses resulting from Gold and blue dotted lines for those resulting from Autodock. The residues oof the binding site are represented as lightblue sticks. The image was created by PyMOL software (www.pymol.org).

The binding poses obtained both from Autodock (yellow sticks) and Gold (violet sticks) revealed that all compounds could interact with α-syn by establishing van der Waals interaction with Ala53, Val55, Thr59, Thr72 and Gly73 (Figure16). Compound **(64)** might engage hydrophobic contacts between the methoxy groups and Ala53 and Val55. Furthermore, in both the poses a H-bond mediated by the hydroxyl group could be formed with Glu61 and Thr72 (Figure16A). In the binding pose predicted by Gold software, compounds **(65)** might establish H-bonds with

Thr72E, Thr72F and Glu61F (Figure16B), while no H-bonds were observed in the pose obtained from Autodock. Instead, compound **(67)** could form H-bonds through its carbonyl group with Thr72E, in the pose predicted by Gold, and G73D, in the one calculated by Autodock (Figure16C). Finally, concerning compound **(69)**, a H-bond between the nitrogen of the indole moiety and the side chain of Glu61E was observed in the poses predicted by both software (Figure16D).

Differently to the other virtual hits, compound **(72)** could inhibit α-syn aggregation by accommodating in binding site2, by establishing a H-bond with Lys45I through its carboxylic group and π-cation interaction with Lys43G through the pyridine ring, as showed in Figure17 for both Autodock and Gold poses. Moreover, in the Autodock pose we observed a salt bridge between the carboxylic group and Lys45H, while in the putative binding pose predicted by Gold, π-cation interaction might be formed between Lys43H and the furan ring.

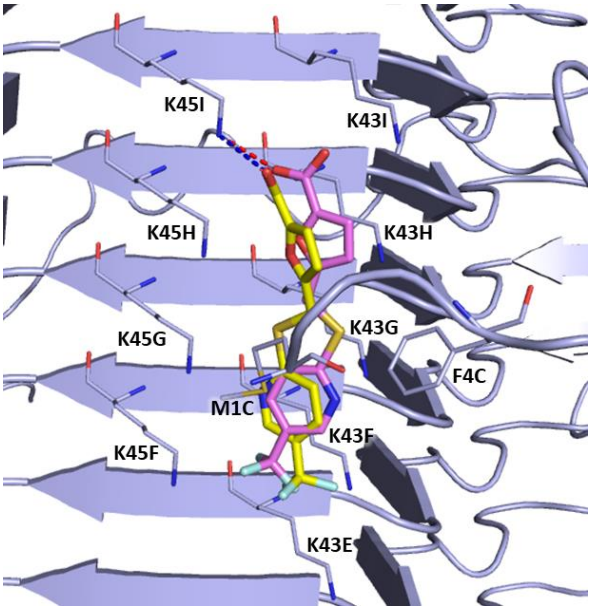


Figure17. Plausible binding poses of compound **72** predicted by Gold (violet sticks) and Autodock (yellow stick) software. Hydrogen bond interactions are represented as red dotted lines for the poses resulting from Gold and blue dotted lines for those resulting from Autodock. The residues oof the binding site are represented as lightblue sticks.

Basing on the results obtained from this computational study, compounds **(64)**, **(65)**, **(67)**, **(69)** and **(72)** could be considered as potential inhibitors of α-syn amyloid aggregation.

All the selected compounds were purchased and their *in vitro* screening is in progress.

5.3 Material and methods

5.3.1 Similarity-based VS

The similarity-based VS was performed by using the web tool SwissSimilarity [180] (http://www.swisssimilarity.ch/) employing the 2D structure of the α-syn aggregation inhibitor SC-D **(41)** as input and Sigma Aldrich database as screening library. Three different VS runs were executed using three different methods to estimate the similarity between the query and the database molecules: FP2 fingerprints, electroshape and spectrophores. FP2 fingerprint belongs to 2D similarity methods and is a path-based fingerprint that analyses a molecule based on the presence of linear fragment whose length is up to 7 atoms [185]. Instead, Electroshape and Spectrophores are 3D methods that do not require geometric alignment thus allowing to speed up the VS when compared to 3D similarity methods that rely on conformational superimpositions [180]. A spectrophore is a descriptor calculated from the 3D atomic properties of a molecules including molecular shape, atomic partial charges and atomic lipophilicity indices [186], while electroshape methods is a descriptor-based tool that combines shape and electrostatic information into a single framework [187]. When these two methods are employed 20 conformers of the reference molecule are compared with 20 conformers of each compound present in the database. The conformers are calculated by molconvert program. The molecules obtained from the virtual screening were ranked according to a similarity score ranging from 0, which indicated totally dissimilar compounds, to 1 for identical molecules. The provided score corresponds to Ts for FP2 fingerprints method and to Manhattan-based score for spectrophores and electoshape methods [180].

5.3.2 Leadlikeness assessment

SwisseADME was employed to evaluate the ADME profile and the leadlikeness properties of the molecules retrieved by VS. In particular, to assess the leadlikeness, the following criteria were used: Lipinski's rule, Brenk, Veber and the absence of PAINS.

According to the Lipinski's "rule of five", a molecule must have less than 5 H-bond donors, less than 10 H-bond acceptors, a molecular weight lower than 500 and a LogP less than 5, to have good absorption and permeation properties [181].

The selection criteria used by Brenk *et al.* [183] consist of: i) a number of heavy atoms between 10 and 27, ii) less than 4 H-bond donors and than 7 H-bond acceptors, iii) ClogP/ClogD between 0 and 4, iv) less than 8 rotatable bonds, v) a number of ring lower than 5 and vi) the absence of unwanted chemical functionalities like reactive groups o groups characterised by unfavourable pharmacokinetic properties.

Veber *et al*. [182] establish that a molecule possessing 10 or less rotatable bonds and a polar surface area equal or less than 140 Å should have a good oral bioavailability.

Finally, compounds that interferes with the common biochemical high throughput screening methods, PAINS, were also discarded.

5.3.3 Molecular docking with Autodock

The docking simulation performed by Autodock 4.2 [176] was executed as described previously in par. 4.3.2 with slight modifications. Initially, the search space was set in order to contain residues from Lys34 to Lys96 located in the β-sheet region with parts of the N- and C-terminus disorder residues nearby to the β-sheet area. A grid box of size 56x60x70 Å and centre x=113.169, y=145.695 and z=-17.985 was applied to cover the mentioned protein region.

Subsequently, in the re-docking procedure, we restricted the space for the docking calculations according to the binding sites identified from the previous docking runs. To this aim, a grid box with size 10x10x10 Å and centre x=118.599, y=153,499 and z=-31.192 was used to define site 1 while a grid box with the same dimension but centre x=97.025, y=159.608 and z=-7.302 was employed to define the binding site 2. Ligand preparation and the docking procedure were performed as reported in par. 4.3.2.

The docking poses representative of the most populated cluster were chosen for analysis and representations. The analysis of the putative ligand-protein interactions was carried out by means of Discovery Studio Visualiser [179].

5.3.4 Molecular docking with Gold

The putative binding mode of the selected compounds was also probed by molecular docking by using the software Gold V 2020.2.0 [184] and employing the solid-state NMR structure of full length α-syn fibrils (PDB code 2N0A) [177]. The search space was defined to cover the same protein region used in the docking performed with Autodock. Therefore, we set as centre the same coordinates used in the previous docking software and all atoms within 40 Å were included in the search space. During the re-docking procedure, the search space was restricted employing as centres the coordinates provided in Autodock and including the residues within 10Å. The standard default settings were used in all the calculations and the ligands were subjected to 10 genetic algorithm runs. The "allow early termination" option was unchecked. The results whose RMSD differs by less than 0.75 Å were clustered together. ChemPLP was chosen as fitness function. The docking poses representative of the most populated cluster were chosen for analysis and representation. The analysis of the putative ligand-protein interactions was carried out by means of Discovery Studio Visualiser [179].

Conclusions

Conclusions

The search for a cure of PD represents a great challenge in the pharmaceutical research field. In this scenario, the inhibition of α-syn aggregation by small molecules has emerged as valuable strategy to block the neurodegenerative process.

During my PhD, two VS screening strategies were exploited to identify novel α-syn aggregation inhibitors: i) pharmacophore-based and ii) similarity-based VS.

In the first case, a ligand-based pharmacophore model was built and used as filter to screen two 3D libraries leading to the selection of four compounds. The ability of the selected molecules to inhibit α-syn aggregation was evaluated *in vitro,* resulting in the identification of a new hit compound **53.** Structural modifications were carried out on inhibitor **53** yielding four interesting derivatives. Lastly, the binding mode of this new class of α-syn aggregation inhibitors was clarified by docking studies. This work allowed to discover a new class of α-syn aggregation inhibitors bearing a pyridinyl-triazole scaffold, providing new information for the design of novel modulators.

In the second computational workflow, we used the inhibitor SC-D as query to perform a similarity-based VS, employing three different similarity methods. The drug-like properties of the obtained molecules were investigated *in silico* and the compounds reporting alerts were discarded. The binding mode of the hits selected from the VS was probed by docking studies performed through two docking tools, Autodock and Gold. The compounds that showed a similar binding mode in the outputs obtained from the two software, were chosen as potential α-syn aggregation modulators. Therefore, the selected molecules were purchased and their biological activity on α-syn is under investigation.

Section 3

Chapter 6

MUC1-CIN85 PPI as novel pharmacological target for anticancer therapy

6.1 MUC1

6.1.1 MUC1 structure

Mucin 1 (MUC1) is a transmembrane glycoprotein expressed in the apical and basolateral surface of the glandular epithelial cells of almost all tissues, including the mammary gland, lungs, stomach, oesophagus, duodenum, pancreas, uterus, prostate and hematopoietic cells [188].

MUC1 gene encodes for a single polypeptide chain that, after translation, undergoes to autoproteolytic cleavage at the GVSVVV motif present in the Sea urchin sperm protein enterokinase and agrin (SEA) domain thus forming two subunits: a N-terminal domain (MUC1-N) and a shorter C-terminal domain (MUC1-C) linked by stable H-bonds (Figure18) [189].

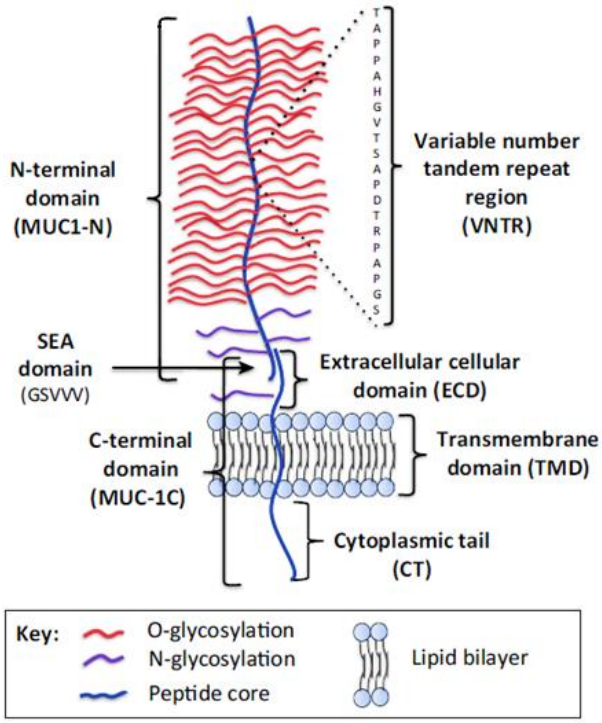


Figure18. Schematic representation of N-terminal and C-terminal subunits of MUC1. The N-terminal domain comprehends a variable number of tandem repeats region (VNTR) highly glycosylated in physiological conditions. The C-terminal domain is composed by an extracellular domain, a transmembrane domain and a cytoplasmatic tail. The two subunits, N- and C-terminal, are linked by non-covalent interactions at SEA domain. Adapted from Nath S. *et al.* [189].

The N-terminal subunit is located on the extracellular side and contains a variable number of tandem repeats (VNTR) composed by a sequence of 20 amino acids

(PDTRPAPGSTAPPAHGVTSA) rich in Ser, Thr and Pro residues [190]. The VNTR region is flanked by a short degenerate sequence similar to that of VNTR [189].

In physiological condition the extracellular domain is highly glycosylated through O-glycosylation and in less extent N-glycosylation. The O-glycosylation occurs at Thr and Ser residues present in the VNTRs and is important for the biological activity of MUC1. Indeed, the hyperglycosylated branches oligomerize forming a mucinous gel that protects epithelia from pathological agents, pH changes and desiccation. The O-glycosylation starts by adding N-acetylglucosamine (GalNAc) to Thr and Ser residues by GalNAc transferases leading to the formation of GalNAcαSer/Thr structures also known as Tn antigen. Subsequently, GalNAc residue are subjected to further modifications by various glycosyl transferases resulting in the construction of different glycan moieties such as i) core 1 formed by addition of galactose (T or TF antigen), ii) core 3 glycan obtained by adding GlcNAcβ1-3GalNAcα, and iii) Sialyl-Tn antigen (STn) yielded by addition of sialic acid. The glycosylation proceeds by extension and chain termination occurs via the addition of carbohydrates such as sialic acid. Instead, N-glycosylation takes place at Asn residues located in the degenerate sequence and is crucial for MUC1 folding, sorting, apical expression and secretion. Glycosylation protects the peptide backbone from the cleavage by proteolytic enzymes and from undergoing to endocytosis clathrin-mediated [188].

Instead, MUC1-C subunit consists of an extracellular domain (ECD) composed by 58 amino acids, a transmembrane domain (TMD) containing 28 amino acids and a cytoplasmatic tail (CT) formed by 72 amino acids. This last region shows seven Tyr residues, 4 of which can be phosphorylated by kinases thus promoting signal transduction cascades [191].

6.1.2 Biological functions of MUC1

The physiological functions of MUC1 include the lubrication and hydration of cell surface and the protection from microorganism and proteolytic enzymes [190]. Moreover, MUC1 explicates an integral role in the regulation of inflammatory responses through its CT domain.

Experimental studies highlight that MUC1-CT act as negative regulator of Toll-like-receptors (TLRs) signalling. TLRs are involved in the innate immune system response through the early recognition of pathogens that promotes host defensive responses. Infections caused by *P. Aeruginosa* and respiratory syncytial virus lead to the activation of TLRs thus triggering the production of inflammatory mediators. In turn, MUC1 is upregulated and inhibits TLR signalling reducing inflammation.

MUC1 is also involved in the regulation of the NLRP3 inflammasome complex whose activation induces the production of inflammatory cytokines and later the enrolment of effector cells. The induction of inflammasome complex requires the activation of TLRs and NF-kB pathway that leads to the upregulation of pro-IL-1β, pro-IL-18, pro-caspase-1 and of the inflammasome components which comprise a sensor protein like NLRP. MUC1 acts as inhibitor of the inflammasome pathway by reducing the activation of TLR and consequently blocking the upregulation of inflammasome constituents. Recently, it has been discovered that the expression of MUC1 during *Helicobacter pylori* infection in mouse and *in vitro* inhibits the production of gastric IL-1 β and, as consequence, a reduction of the inflammation caused by *H. pylori* was observed [191].

6.1.3 Tumor-associated MUC1

The tumor-associated MUC1 (TA-MUC1) presents different biochemical characteristics and cellular distribution respect to the protein expressed in normal cells. In cancers, MUC1 is overexpressed and it is distributed both on the surface and in the cytoplasm of epithelial cells due to the loss of cell polarity [192]. Furthermore, the VNTRs of TA-MUC1 are hypo-glycosylated with the long-branched glycans truncated and often capped by sialic acid because of the overexpression of α2,6- and α2,3-sialyl transferases that cause premature termination of chain elongation [193]. Several studies showed that TA-MUC1 exerts a pivotal role in the transcription of genes associated with tumor invasion, angiogenesis, cell proliferation, apoptosis, inflammation and drug resistance. MUC1-CT associates with the epidermal growth factor receptor (EGFR) and translocates to the nucleus starting a signal cascade that stimulate cell proliferation. In pre-clinical models, TA-MUC1 favours the formation of metastasis

through the disruption of normal cell-cell and cell-matrix adhesion and increase of cell-endothelial adhesion. In this regard, it has been shown that TA-MUC1 interacts with cell adhesion molecules ICAM-1 and E-selectin promoting cell migration and vascular invasion.[188] The overexpression of TA-MUC1 have been linked with poor prognosis in colon, pancreas, bladder and breast cancers [194,195].

6.1.4 Anti-MUC1 monoclonal antibodies

TA-MUC1 represents an attractive target for immunotherapy due to its expression on the surface of tumor cells. The truncated carbohydrate antigens Tn and TF localised in the VNTR can be recognised by antibodies thus representing a valid therapeutic strategy for the cure of cancer. Within this context, several antibodies targeting different domains of MUC1 have been developed and some of them are investigated in clinical trials [188, 196].

Among the studied antibodies, there is the human milk fat globule 1 (HMFG1) which is an IgG1 with a k light chain murine antibody which recognise the PDTR epitope present in the VNTR of the extracellular domain of MUC1. The radiolabeled form of HMFG1 with ^{90}Y was developed for the treatment of ovarian cancer. The humanised form of this antibody, AS1402, showed a good tolerability and safety profile in Phase I trial. In Phase II study it was administered in combination with the endocrine therapy; however, this combination did not lead to positive results despite no safety related problems were observed [197, 198,199].

TAB004 is a humanised IgG1 mAB that recognised the sequence STPPVHNV located in the VNTR region of TA-MUC1. This immunoglobulin proved to be effective in targeting pancreatic cancer cells, but not normal cells. Furthermore, TAB004 is a diagnostic marker able to determine the various stages of pancreatic tumor by identifying circulating MUC1 in patient's serum. It has been shown that TAB004 halt the anti-apoptotic activity of MUC1 and increase the anti-tumor effect of chemotherapy drugs [200,201,202].

Another anti-MUC1 monoclonal antibody is MUC1-014E which targets the intracellular 19 amino acid sequence (RYVPPSSTDRSPYEKVSAG) of the MUC1-CT. Immunohistochemical staining of 107 gastrectomy samples revealed a high rate of positive staining when compared to other anti-MUC1 antibodies. MUC1-014E is

also able to recognise isolated tumor cells of signet-ring cell carcinoma and non-solid type poorly differentiated stomach adenocarcinoma indicating that this antibody can be exploited for the detection of these cells that are frequently encountered in scirrhous gastric tumor [203].

The above reported antibodies recognize non-glycopeptide epitopes of TA-MUC1. In addition, monoclonal antibodies interacting with glycopeptide epitopes of TA-MUC1, such as PankoMab, have also been developed.

PankoMab is a murine IgG1, kappa light chain antibody that targets TA-MUC1 glycopeptide. It showed high specificity and affinity, high antibody-dependent cell cytotoxicity (ADCC) activity and fast internalization thus proving to be a potential anti-cancer agent. Different chimeric and humanized forms of PankoMab have been developed which are under investigation in clinical studies as potential therapeutic and diagnostic tools. Among them, PankoMab-GEX™ (PMG) is a glycol-optimized humanised antibody with an improved affinity towards TA-MUC1, increased ADCC and ability to induce apoptosis in cancer cells expressing MUC1. PMG proved to be safe and well tolerated in phase I clinical trials [204,205,206].

6.2 CIN85

6.2.1 Structural features of CIN85

The Cbl-interacting protein of 85 kDa (CIN85) is a multifunctional adaptor protein involved in many biological functions. Adaptor proteins are non-catalytic peptides characterised by the presence of one or more domain able to engage other proteins thus forming multimeric complexes implicated in the transmission of intracellular signals [207].

From a structural point of view, CIN85 is composed at the N-terminus by three Src Homology 3 (SH3) domains (usually referred to as SH3A, SH3B and SH3C), a central proline-rich region and a C-terminal coiled-coil (Cc) domain (Figure19).

Figure19. Schematic representation of CIN85 structure consisting of three SH3 domains, a central Pro-rich region and a coiled-coil C-termina domain.

The three SH3 domains share a high similarity among themselves compared to any other SH3 domain and mediate PPI by interacting with Pro-rich motifs. Between the second and the third SH3 domains, there is a region rich in Ser and Thr residues that could undergo to phosphorylation. Moreover, in the N-terminal domain there are three FXDXF sequences that might work as binding site for the clathrin adaptor protein AP2. The central Pro-rich region serves as potential binding site for SH3 domains of other different proteins and is located close to the third SH3 domain. Finally, the C-terminal domain is characterised by a high propensity to form coiled-coil structures that are involved in CIN85 oligomerization with the resulting formation of multiprotein CIN85-associated complexes [207,208].

6.2.2 SH3 domains

SH3 domains are small protein interactions modules composed by approximately 60 aminoacidic residues. Structurally, SH3 domains consist of five β-strands arranged in two anti-parallel β-sheets connected by three loops named RT, distal and n-Src loops and a short 3_{10} helix (Figure20) [209,210].

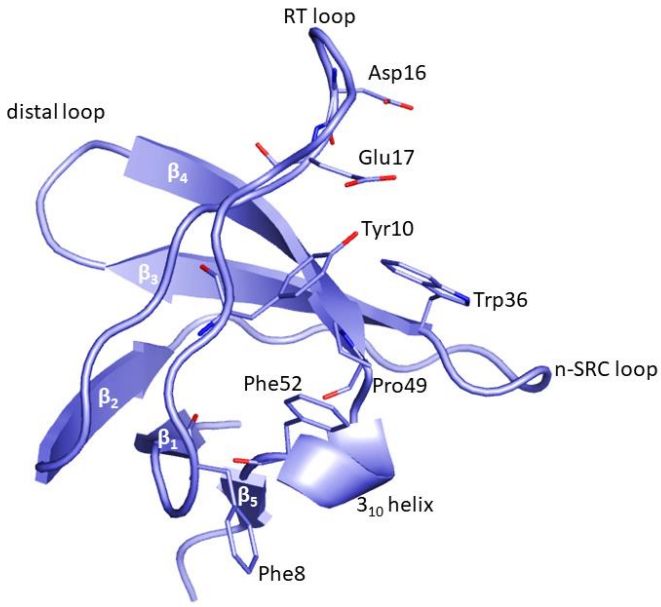


Figure20. Structure of SH3 domain. The amino acids residues involved in the binding with the Pro-rich binding sequences are highlighted in sticks. The image was created by means of PyMOL software (www.PyMOL.org)

SH3 domains are known to bind proline rich sequences containing the consensus motif PXXP where X can be any amino acid. Due to the small size of the recognised

sequence, the binding specificity is given by the nearby residues [211]. However, an increasing number of peptides that bind SH3 domains with atypical binding motifs have been identified over the years [210].

The binding surfaces of SH3 domains is quite conserved and comprehend the flat valley above the β3 and β4 strands, flanked by the terminal part of strand β2/n-Src loop and the tip of the RT loop. The binding surface includes hydrophobic residues such as Tyr, Phe, Pro and Trp mainly involved in the binding of Pro-rich sequences and an acidic pocket that can establish ionic interactions with basic residues adjacent to the binding motif [212]. SH3 domains usually bind the protein partner with a 1:1 stoichiometry and proteins having multiple SH3 domains can engage contacts with multiple proteins or ligands and therefore increase the affinity and specificity of the interaction [213]. However, in literature there are also different examples in which the Pro-rich motif binds simultaneously two SH3 domains resulting in the formation of heterotrimeric complexes [214,215].

When bound to SH3 domains, the ligand usually assumes an extended left-handed helical conformation, named polyproline-2 (PPII) helix which is composed by three residues per turn [216].

Pro-rich sequences can interact with SH3 domains assuming two opposite orientations, type I and type II, according to the position of a basic residue that can precedes or follow the PXXP motif. In the type I orientation, the basic residues N-terminal to the binding motif occupy the hydrophilic pocket, while in the type II the same pocket accommodated basic residues C-terminal to the motif [217].

CIN85 SH3 domains recognise an atypical PXXXPR binding motif present in the ubiquitin ligases c-Cbl and Cbl-b, and in other CIN85 effectors [218]. Crystallographic studies of CIN85 SH3 N-terminal (SH3A) domain with a 11-mer proline-arginine peptide from Cbl-b (PDB code 2BZ8) revealed the formation of a heterotrimeric complex in which Cbl-b is sandwiched between two SH3 domains (Figure21). In the above reported complex, one SH3 domain bound the Pro-rich peptide adopting a type I orientation while the other SH3 domain engages Cbl-b in a type II orientation [213].

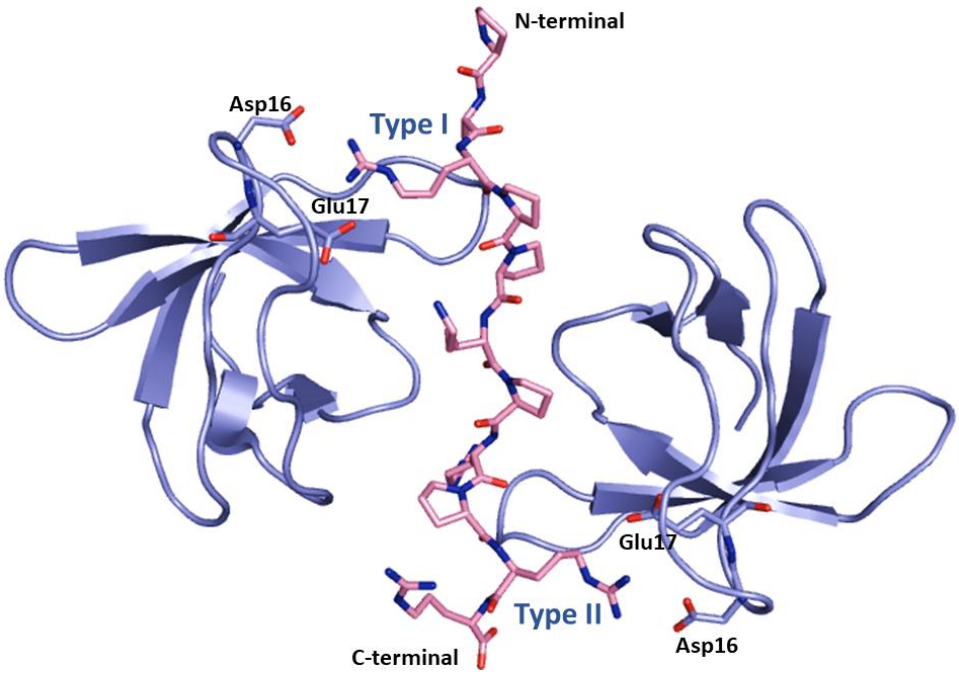


Figure21. Crystal structures of CIN85 SH3A domain in complex with Cbl-b peptide. SH3 domains are represented as lightblue cartoon while Cbl-b peptide as pink sticks. The residues forming the hydrophilic pocket are highlighted in sticks. The image was created by PyMOL software (www.pymol.org).

It has been also observed that SH3 domains of CIN85 can bind the protein partner in their monomeric form, as in the case of SH2-containing inositol phosphatase 1 (SHIP-1) [219], Clathrin-mediated endocytic (CMS) protein [220] and others [221].

6.2.3 Biological activity of CIN85

CIN85 interacts with various proteins forming multimeric complexes that are involved in the regulation of many biological processes such as tyrosine kinases receptors (RTKs) down-regulation, immunological synapse and malignancy of cancer cells.

6.2.3.1 Down-regulation of RTKs

CIN85 is implicate in the down-regulation of RTKs mediated by the ubiquitin ligase Cbl through clathrin-mediated endocytosis.

Specifically, growth factor binds to RTKs inducing receptor auto-phosphorylation and phosphorylation of multiple substrates. Activated RTKs changes their localization from the cell surface to the endosomal compartment where they can be recycled to the plasma membrane or sorted for lysosomal degradation. These

events are regulated by a several PPIs, receptor phosphorylation and ubiquitination.

Cbl proteins, in particular Cbl-b and c-Cbl, are involved in the ubiquitination followed by lysosomal degradation of several RTKs. Moreover, Cbl proteins modulate RTKs endocytosis by binding with CIN85. In details, Cbl binds to the phosphorylated RTKs and itself is tyrosine-phosphorylated. Cbl-RTK complex interact with SH3 domains of CIN85 whose Pro-rich region is constitutively associated with endophilins. Endophilins are regulatory proteins of clathrin-coated vesicle that promote membrane invagination during RTKs endocytosis [207,208].

6.2.3.2 Receptor clustering and formation of immunological synapse

CIN85 is implicated in CD2 receptor clustering and cytoskeletal polarity in the specialised junction between T-cells and antigen-presenting cells (APCs), noted also as "immunological synapse". CD2 is an adhesion molecule expressed on the surface of T lymphocytes and natural killer (NK) cells. In humans, CD2 mainly interacts with CD58, a cell adhesion molecule expressed on APCs promoting T-cells activation.

The engagement of CD2 in the immunological synapse is due to its interaction with SH3 domains of the adaptor proteins CIN85 and CD2AP which occurs via the polyproline motifs present in the C-terminal tail of CD2. At the same time, the C-terminal region of CIN85 binds CAPZ, a capping protein that caps the plus end of actin filaments, providing a direct connection between CD2 and the cytoskeleton. Therefore, the linkage of CIN85 to CAPZ is important for the regulation of cytoskeleton activity in T cell activation, fostering the alignment between the T cell and APC membranes [207,222].

6.2.3.3 Role of CIN85 in malignancy and invasiveness of cancer cells

Several experimental studies highlighted that CIN85 fosters the malignancy and regulate the invasiveness in cancer cells in different ways.

A high expression of CIN85 was observed in invasive breast cancer cells where this adaptor protein was identified as component of invadopodia, which are membrane protrusions, rich in actin, that confers to the tumor cells the ability to invade

through extracellular matrix (ECM) and neighbouring tissues. The proposed mechanism assumes that CIN85 is implicated in invadopodia biogenesis by linking Cbl proteins to ASAP1 and AMAP1 that work as effectors of oncogenic signalling downstream of several proteins that, through the instauration of PPIs, promote the biogenesis of invanopodia. Indeed, the inhibition of Cbl-CIN85-ASAP1/AMAP1 complex by siRNA-mediated silencing revealed a reduction of invasiveness in breast cancer cells *in vitro*. Additionally, CIN85 can also engage other proteins involved in invadopodia biogenesis such as N-WASP, N-WASP-interacting adaptor protein WIP, dynamin 2, synaptojanin2 and p130Cas [223].

Another mechanism through which CIN85 promotes the invasiveness of cancer cells concerns its interaction with TA-MUC1 which will be described more in details in the next paragraph.

6.3 MUC1-CIN85 PPI

In 2013, Cascio S. *et al.* reported for the first time the interaction between the hypo-glycosylated form of MUC1 and CIN85 [224].

Similarly to the PXXXPR motif found in Cbl and other CIN85 effectors, the VNTRs of MUC1 contain the conserved sequence PDTRP that is recognised by the SH3 domains of CIN85.

Both proteins are over-expressed in the early and advanced clinical stages of breast cancer and co-localised at invadopodia-like structures enhancing the invasiveness of cancer cells by probably regulating the activity of other proteins involved in the invadopodia formation.

In vivo studies, conducted on a tumor metastasis model of B16 melanoma, revealed that when CIN85 depleted melanoma cells were injected in mices, they showed few or no lung metastasis, while the overexpression of MUC1 restored the shCIN85 reduced metastatic process [224].

The formation of MUC1-CIN85 complex depends on the glycosylation degree of the VNTRs of MUC1. In particular, the glycosylation decreases as tumor progresses to later stages. The reduction of glycosylation could increase the number of MUC1-CIN85 complexes fostering the invasive ability of cancer cells [224]. This is in good agreement with previous works reporting that alterations in the length of the sugar

chains of VNTRs have a role in tumor progression and metastasis formation [225,226,227].

Experimental studies showed that in presence of a MUC1 peptide agonist able to induce dimerization, the interaction between MUC1 and CIN85 was two times higher than in its absence thus suggesting that CIN85 would bind MUC1 forming a heterotrimeric complex as observed for Cbl-b [228]. Unfortunately, no experimentally solved structure is available for MUC1-CIN85 complex.

Considering the important role that this complex plays in tumor invasion and metastasis formation, the modulation of MUC1-CIN85 PPI represents a viable strategy for pharmacological intervention in cancer therapy.

Chapter 7

In silico identification of putative ligand binding sites on CIN85 SH3 domain

7.1 Introduction

From June 2018 to December 2018, I spent a research internship at the Pharmaceutical Chemistry Department of the University of Vienna, working under the supervision of Prof. Thierry Langer.

As discussed in Chapter 1, the design of PPI modulators is a difficult task due to the large and shallow protein interface that lack well-defined binding sites for small molecules. However, due to the dynamic nature of protein interfaces, the opening of transient ligand binding sites can be observed furnishing new opportunities for drug discovery. The research activity performed during my stage was focused on the identification of putative druggable binding sites on CIN85 SH3 N-terminal (SH3A) domain, considering that ligands able to bind to CIN85 can disrupt its association with MUC1. For this purpose, we employed two distinct binding-site detection methods: i) the geometry-based algorithms F-pocket [229] and ii) the computational mapping server FTMap [230]. Both approaches rely on the atomic coordinates of the target macromolecule but, considering that proteins are dynamic systems, the use of a single experimental structure would be too reductive. Indeed, structural changes could promote the exposure of suitable site for small molecules that could be missed when using a single set of coordinates [231]. In this scenario, MD simulation represents a useful tools to study protein motions providing some representative conformations of the system [82].

The content of this chapter has been object of one publication and three poster presentations:

- ***Serena Vittorio**, Thomas Seidel, Arthur Garon, Rosaria Gitto, Thierry Langer and Laura De Luca. In Silico Identification of Potential Druggable Binding Sites on CIN85 SH3 Domain. International Journal of Molecular Science, 2021, 22(2), 534.*

-***Serena Vittorio**, Rosaria Gitto, Arthur Garon, Ugo Perricone, Thierry Langer and Laura De Luca. In silico studies for the discovery of MUC1/CIN85 protein-protein interaction inhibitors as anti-metastatic agents. Paul Erlich Euro-PhD Network & MuTaLig COST Action meeting 2019, 13-15 June 2019, Catanzaro, Italy.*

- ***Serena Vittorio**, Rosaria Gitto, Arthur Garon, Ugo Perricone, Thierry Langer and Laura De Luca. An in silico approach targeting MUC1/CIN85 protein-protein interaction for the development of novel anti-metastatic agents. XII European Workshop in Drug Design, 19-25 May 2019, Siena, Italy.*

-***Serena Vittorio**, Rosaria Gitto, Arthur Garon, Ugo Perricone, Thierry Langer and Laura De Luca. MUC1/CIN85 PPI inhibitors: binding site prediction and molecular docking studies. Workshop of Sicily and Calabria Sections of the Italian Chemical Society (SCI), 1-2 March 2019, Palermo, Italy.*

7.2 Results and discussion

7.2.1 MD simulation and clustering process

In this study, to include protein flexibility, the crystal structure of the SH3A domain of CIN85, extracted from the crystallographic complex with Cbl-b (PDB code 2BZ8)[213], was subjected to 400 ns MD simulation by employing the software CHARMM [232]. The root mean square deviations (RMSD) was calculated and plotted as shown in Figure22. Between 175 and 225 ns, we observed an increase of the RMSD mainly due to the motions of RT loop; after this interval, the system stays stable for the rest of the trajectory.

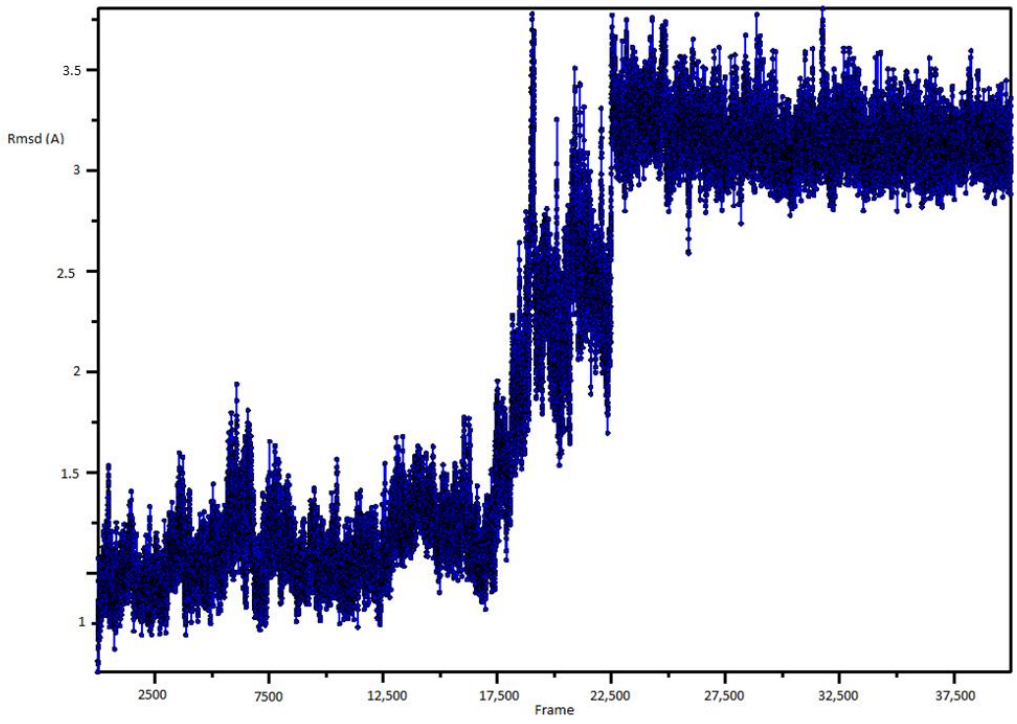


Figure22. RMSD plot of protein backbone.

The 40,000 snapshots obtained from the MD simulation were clustered basing on RMSD by means of TTClust [233] affording 32 clusters. For each cluster, a representative frame, which corresponds to the one with the lowest RMSD respect to all the other frames in the cluster, was identified thus providing 32 conformers. In Figure 23, the barplot showing the distribution of the frames within each cluster (Panel A), the 2D projection plot of the relative dustances between clusters (Panel B) and the timeline barplot displaying the appearance of each cluster during the trajectory, are depicted.

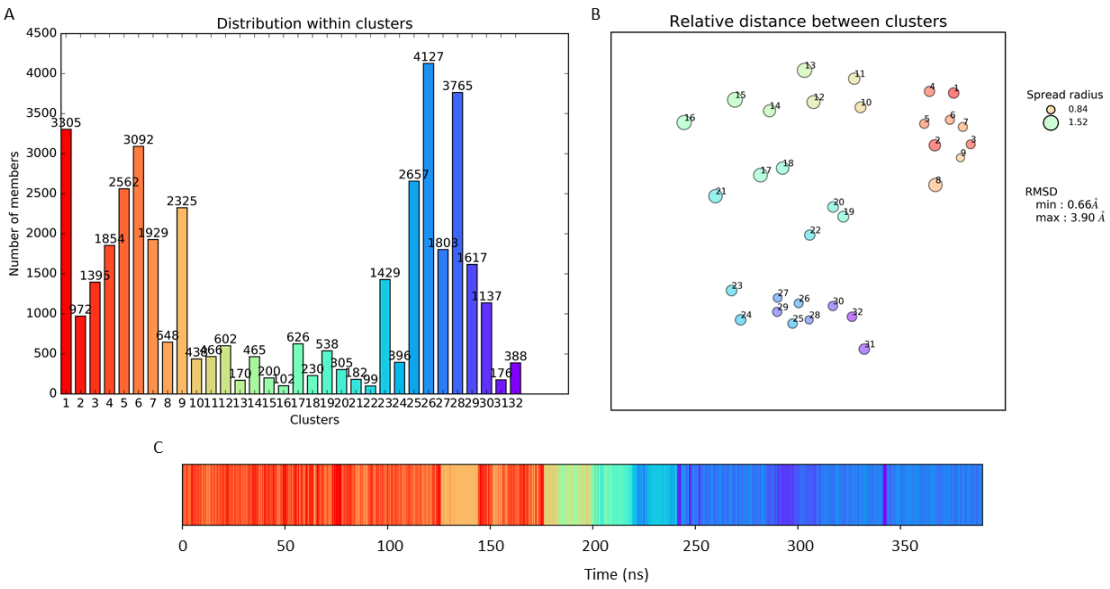


Figure23. Outcomes resulting from the clustering process. A) Distribution of the frames within each cluster; B) Relative distance between clusters according to the RMSD between representative frames; C) Timeline barplot showing the appearance of each cluster during the trajectory. Each cluster is color coded.

As depicted in Figure23A, cluster 26 (C26), which appears in the last part of the simulation, is the most populated with 4127 conformers. Basing on the RMSD between representative frames, the minimum and maximum distances between clusters are 0.66Å and 3.90Å respectively, whereas the minimum and maximum spreads (average RMSD in each cluster) are 0.84 and 1.52 as shown in Figure23B.

7.2.2 Druggable ligand binding sites prediction

The presence of druggable binding site for small molecules was probed both for the SH3A domain derived from the crystal structure and for each protein conformation obtained from the clustering process, by means of F-Pocket2 [229] and FTMap webserver [230]. F-pocket exploits a geometry-based algorithm relied on the concept of alpha spheres to predict ligand binding pockets. Initially, alpha spheres are determined for the whole protein structure and then clustered in order to identify any pocket. Small clusters of poor interest are discarded whereas neighboring pockets are merged. Lastly the found pockets are score and ranked according to their likeliness to accomodate small molecules [229]. To this aim, F-pocket furnishes a druggability score between 0 and 1; a score lower than 0.5 (the threshold value) indicates that drug-like compounds are not likely to bind to the

predicted pocket, while values greater than 0.5 imply that the binding could occur [234]. FTMap is a computational mapping server (https://ftmap.bu.edu/) for binding hotspots detection, that employs 16 small organic probes, having different size and polarity, to map protein surface. The probes are docked, clustered and then ranked basing on their binding energy. The top ranked clusters of different probes are merged providing consensus sites (CSs) that represent putative ligand binding sites. According to FTMap, a binding site is predicted as druggable if the number of probes in the CS is higher the 16 [230].

Regarding the crystal structure, no druggable binding sites were detected by both approaches.

The exploration of the binding sites on the MD derived structures by F-pocket led to the identification of two putative druggable binding pockets named P1 and P2 depicted in Figure24.

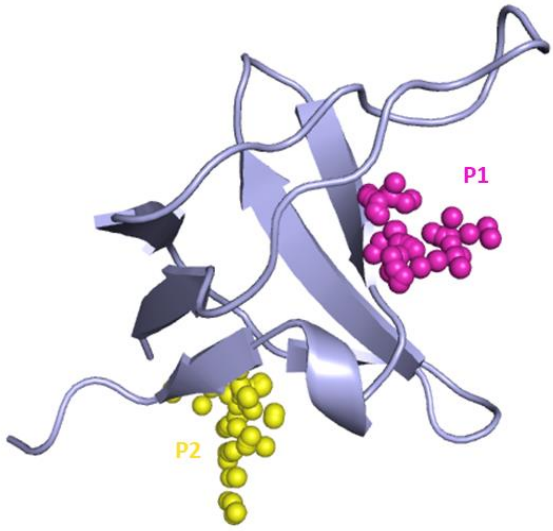


Figure24. Druggable binding pockets identified by F-Pocket. P1 is represented by magenta alpha spheres while P2 by yellow alpha spheres. The image was created by means of PyMOL software (www.pymol.org).

P1 was detected in the representative snapshots of C8, C10, C11, C13, C14 and C19 with druggability score values spanning from 0.645 to 0.934. This pocket is located at the binding surface of CIN85 SH3A doman and opened up between 170 and 230 ns. It incluedes Tyr10, Glu17, Leu18, Ile20, Trp36, Leu47, Phe48, Pro49 and Phe52 (Figure 25).

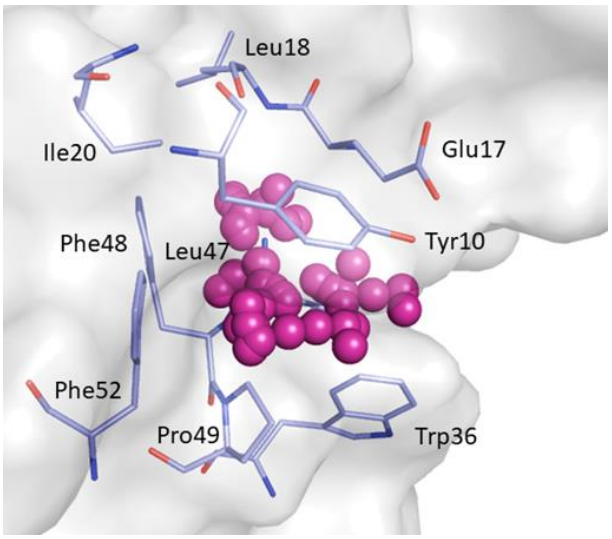


Figure25. Close view of the putative ligand binding pocket P1 predicted by F-pocket. P1 is represented by magenta alpha spheres. The residues belonging to P1 are shown as lightblue sticks. The image was created by means of PyMOL software (www.pymol.org).

The second binding pocket predicted by F-pocket, P2, was detected in a region between in the representative frames of C12 and C16 with druggability score values of 0.6150 and 0.577, respectively. P2 is located in the region between strands β1, β5 and the 3_{10} helix and is lined by Val2, Glu3, Ala4, Ile29, Trp37, Asp50, Val53, Arg54 and Glu55 (Figure 26).

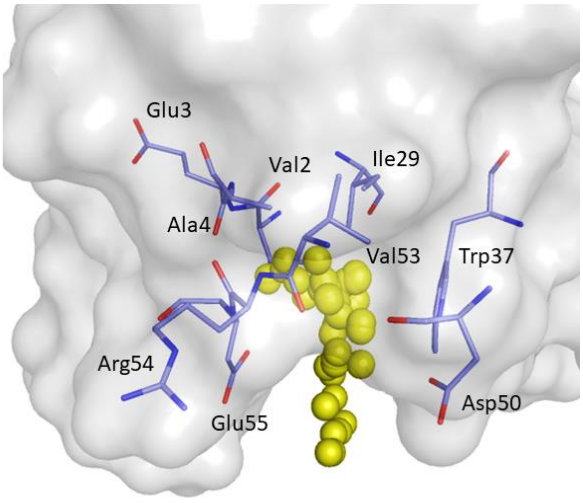


Figure26. Close view of binding pocket P2 predicted by F-pocket. P2 is shown as yellow alpha spheres while the residues lininig the pocket are displayed as lightblue sticks. The image was created by means of PyMOL software (www.pymol.org).

The binding site analysis executed by means of FTMap webserver on the MD derived protein conformations led to the identification of seven druggable CSs that are displayed in Figure27.

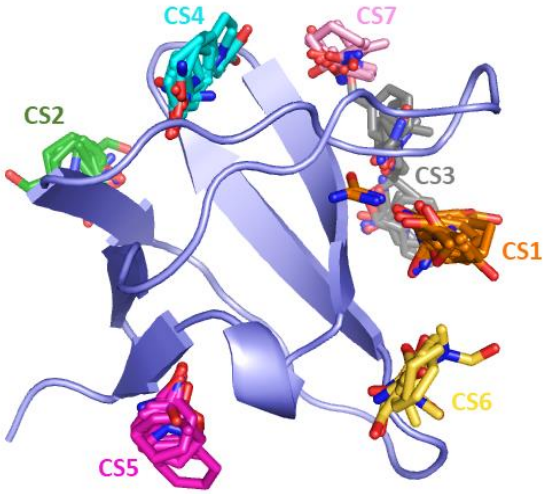


Figure27. Consensus sites (CS) predicted by FTMap. Probes clusters, that identified the hotspots regions, are depicted as sticks. Probes belonging to the same CS show the same color. The image was created by means of PyMOL software (www.pymol.org).

The most occurring binding site was CS1 that was detected in 20 clusters (C10-C13, C15-C22, C24, C26-32). It is situated at the binding surface of CIN85 SH3A domain and includes Tyr10, Asp16, Glu17, Leu18, Ile20, Trp36, Leu47, Phe48, Pro49 and Phe52 (Figure28). Interestingly, the location of CS1 matches quite well the region where pocket P1, predicted by F-pocket, is located.

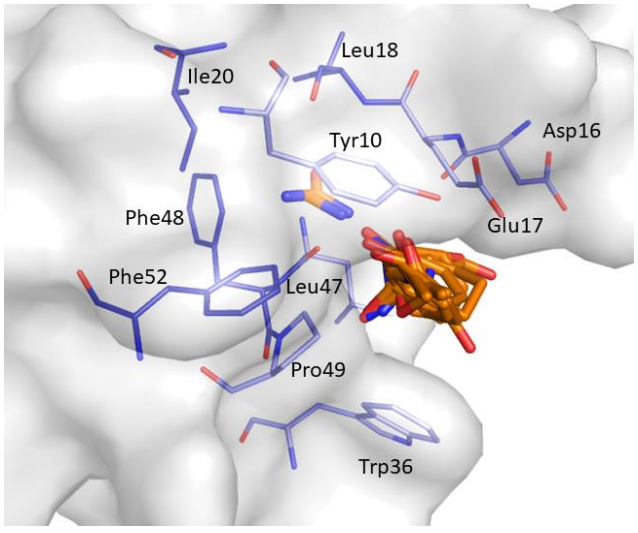


Figure28. Consensus site CS1 found most frequently in the MD snapshots according to FTMap prediction. The representative probes that defined the binding site are represented as orange sticks. The amino acid residues interacting with the probes are displayed as lightblue sticks. The image was created by means of PyMOL software (www.pymol.org).

CS2 was retrieved for 9 clusters (C7, C13, C22, C24, C26, C29-C32). It was identified in the space between the distal loop, β1 and β2 strands and comprehend residues Glu3, Ala4, Ile25, Ile26, Thr27, Gln40, Ile41, Asn42 and Lys57 as shown in Figure29.

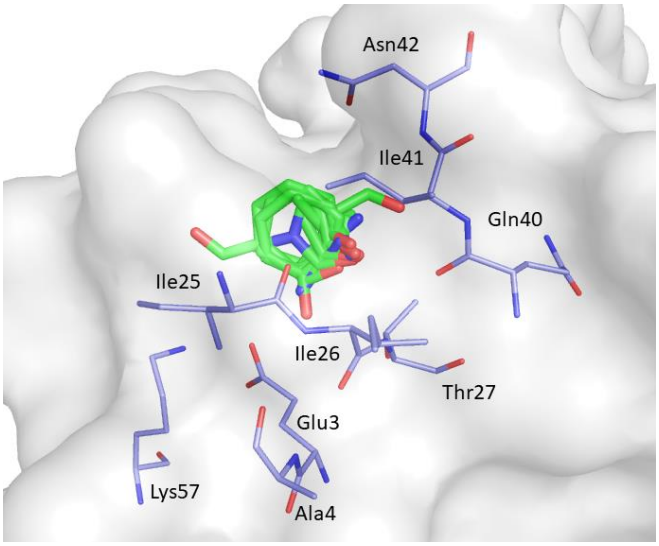


Figure29. Consensus site CS2 identified by FTMap. Clustered probes are depicted as green sticks. The image was created by means of PyMOL software (www.pymol.org).

CS3 was identified in the representative frames in 5 clusters (C10, C25-C26 and C29-C30). It is located between the RT loop and the β4 strand and is lined by Asp16, Glu17, Leu18, Glu38, Gly39, Arg44, Arg45, Gly46 and Leu47 (Figure30).

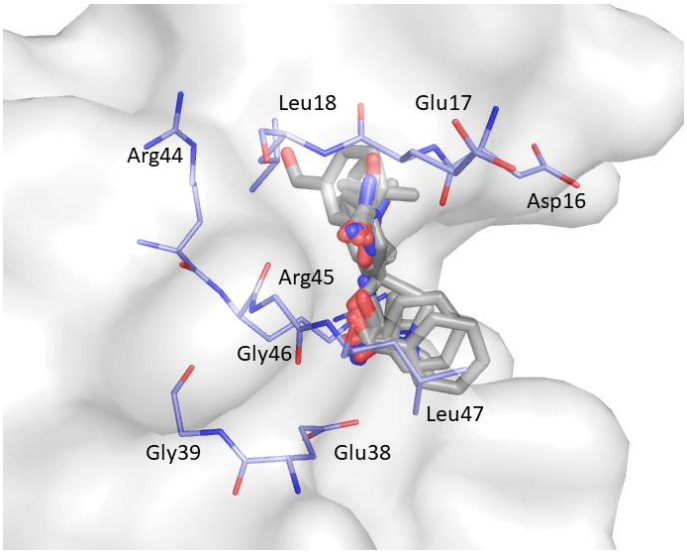


Figure30. Consensus site CS3 detected by FTMap. Clustered probes are shown as grey sticks. The image was created by means of PyMOL software (www.pymol.org).

CS4 was predicted in 4 clusters (C1, C3, C12, C24). It was detected in the region between β2 strand, distal loop, and the RT loop, and comprises residues Leu18, Thr19, Ile20, Ser21, Glu24, Ile41, Asn42 and Arg44 (Figure31).

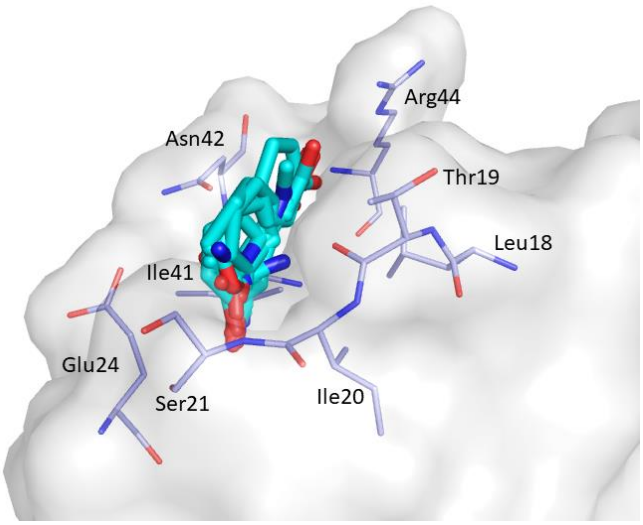


Figure31. Consensus site CS4 found by FTMap. Clustered probes are displayed as cyan sticks. The image was created by means of PyMOL software (www.pymol.org).

CS5 was found in 4 clusters (C1, C6, C8, C12). It is situated close to β2 strand and includes residues Glu7, Phe8, Trp37, Asp50, Asn51, Phe52, Val53, Arg54 and Glu55 as shown in Figure 32.

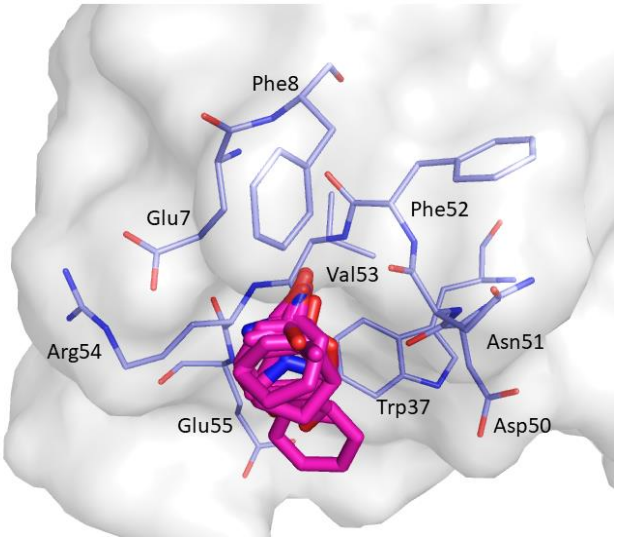


Figure32. Consensus site CS5 identified by FTMap. Clustered probes are represented as magenta sticks. The image was created by means of PyMOL software (www.pymol.org).

CS6, displayed in Figure33, was identified in 2 clusters (C2, C9) in the space between the n-Src loop, 310 helix, β3 and β4 strands. It is lined by Gly34, Gly35, Trp36, Pro49, Asp50 and Asn51.

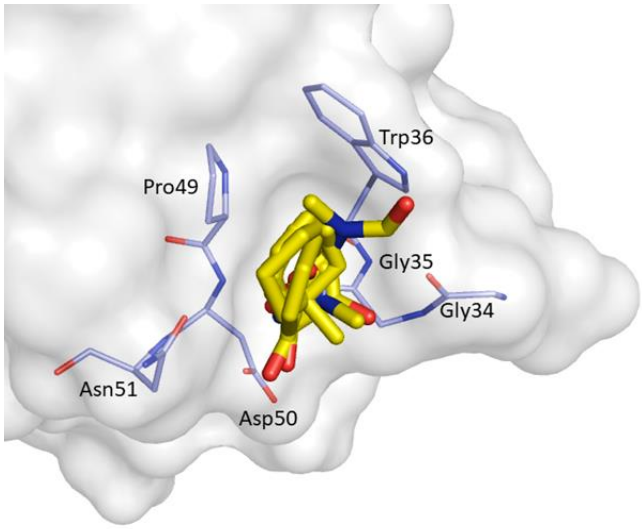


Figure33. Consensus site C6 found by FTMap. Clustered probes are highlighted as yellow sticks. The image was created by means of PyMOL software (www.pymol.org).

Lastly, the seventh binding site predicted by FTMap, CS7, was found in 2 clusters (C4, C7) in the region between the RT loop and the β4 strand. The clustered probes are surrounded by Asp15, Glu17, Leu18, Thr19, Arg44 and Arg45 as depicted in Figure34.

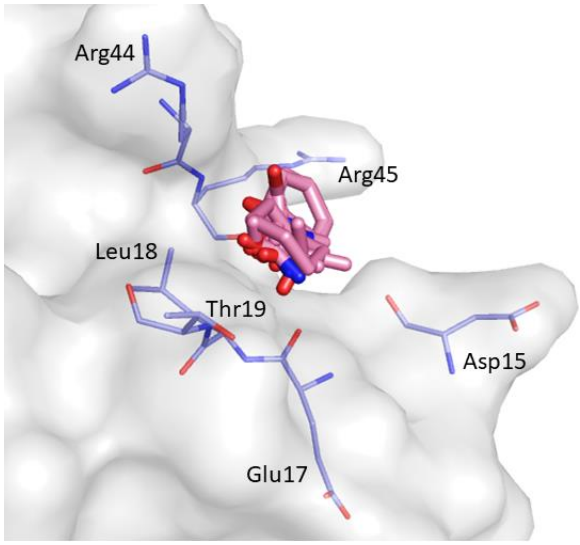


Figure34. Consensus site C7 detected by FTMap. Clustered probes are represented by pink sticks. The image was created by means of PyMOL software (www.pymol.org).

The visual inspection of the probes clustered by FTMap gives useful hints about potential ligand binding sites that can be employed for the structure-based design of novel ligands. Moreover, this tool furnishes a list of residues implicated in hydrogen bonds and other non-bonded interactions with the probes thus providing additional insights concerning the residues that should be targeted. In Table 2, the residues involved in more than the 10% of all probe-residue interactions over the 32 mapped protein conformations are reported. Some of these amino acids, *e.g*. Tyr10 and Trp36, are conserved among SH3 domains and are implicated in the recognition of Pro-rich sequences. Hence, the binding to these residues by a small molecule, could disrupt the PPI through an orthosteric mechanism.

Table 2. Residues interacting with the mapping probes that account for at least the 10% of all the interactions over the 32 clusters.

Non-bonded interactions	H-bonding interactions
Phe8, Tyr10, Gln13, Asp16, Glu17, Leu18, Glu24, Ile25, Trp36, Trp37, Glu38, Asn42	Tyr10, Gln11, Gln13, Glu17 Thr19, Ser21, Ile25, Thr27, Asn42, Arg44, Arg45, Leu47, Asp50, Val53, Glu55

The inclusion of protein flexibility led to the identification of several potential druggable ligand binding sites on CIN85 SH3A domain, revealing how structural rearrangements could favour the exposure of binding site otherwise not appreciable in the static representation of a single crystal structure. A binding site located on the binding surface of CIN85 SH3 domain (P1 for F-pocket and CS1 for FTMap) was predicted in consensus by both software. Interestingly, the mapping by FTMap of the representative frames in which P1 was predicted by F-pocket, revealed that several small probes are able to occupy this pocket resulting in a druggable consensus cluster. Compounds able to interact with this pocket might act as orthosteric inhibitors of MUC1-CIN85 PPI.

The other binding sites detected by F-pocket and FTMap could be considered as viable regions to accommodate allosteric modulators.

Overall, this study provides new structural insights that can be exploited for the design of new ligands through the application of structure- and fragment-based strategies.

7.2.3 Docking studies

Basing on the above reported results, docking studies were performed by employing P1/CS1 as binding site, as it was predicted in consensus by F-pocket and FTMap. The virtual screening was conducted by using the Asinex PPI library (Non-Macrocyclic) containing 11,780 molecules that were filtered, before docking, according to the results gained from Grail approach (GRids of phArmacophore Interaction fieLds) developed by the research group of Prof. Langer [235]. This method describes favourable interaction sites by combining grid-based approaches with the pharmacophore concept. In Figure 35, the results obtained for the SH3A domain of CIN85 are displayed.

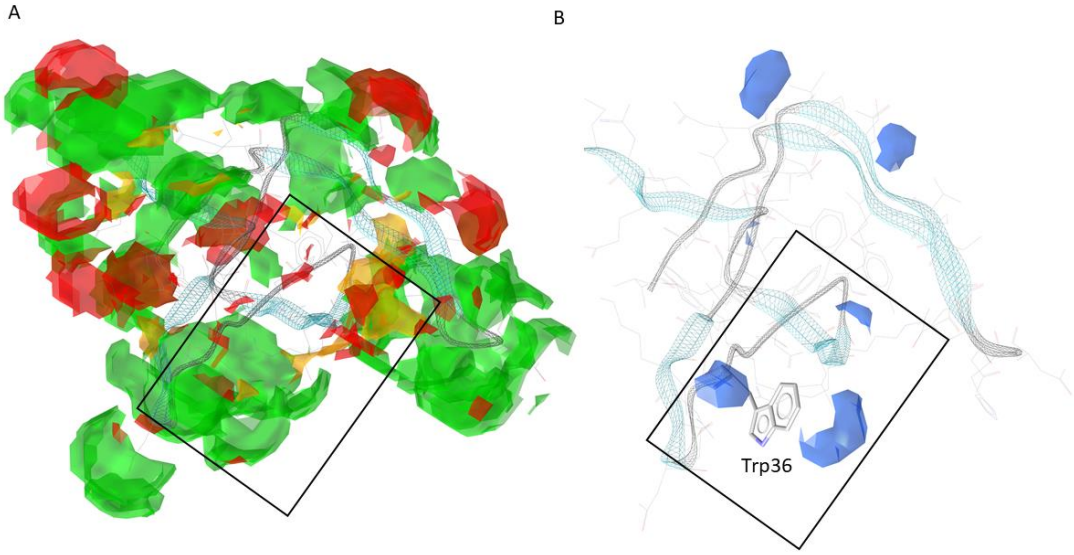


Figure35. A) Grids related to H-bond donors (green) and acceptors (red). These grids highlighted regions in which the presence of H-bond donors or acceptors is favourable for the interaction with the protein. B) Grids related to π-π interactions: blue grids represent regions where the presence of aromatic rings is favourable for the interaction with the target. The black boxes indicate the binding site.

The region that includes the binding site, delimited by the black box, showed mostly sites in which the presence of a H-bond donor (green grids) is favourable for the interaction with the target protein. Concerning the grid related to the aromatic interactions (Panel B), we observed a favourable interaction site at Trp36. Considering all these findings, we filtered the Asinex PPI library applying the following filters: i) aromatic ring ≥1 and ii) H-bond donors ≥1. The resulting 9242 compounds were docked into the identified pocket by using Autodock Vina software [236]. The thirty top ranked compounds were chosen for further analysis.

After the visual inspection and the analysis of the protein-ligand interactions, eight compounds (Chart16) were selected for biological evaluation.

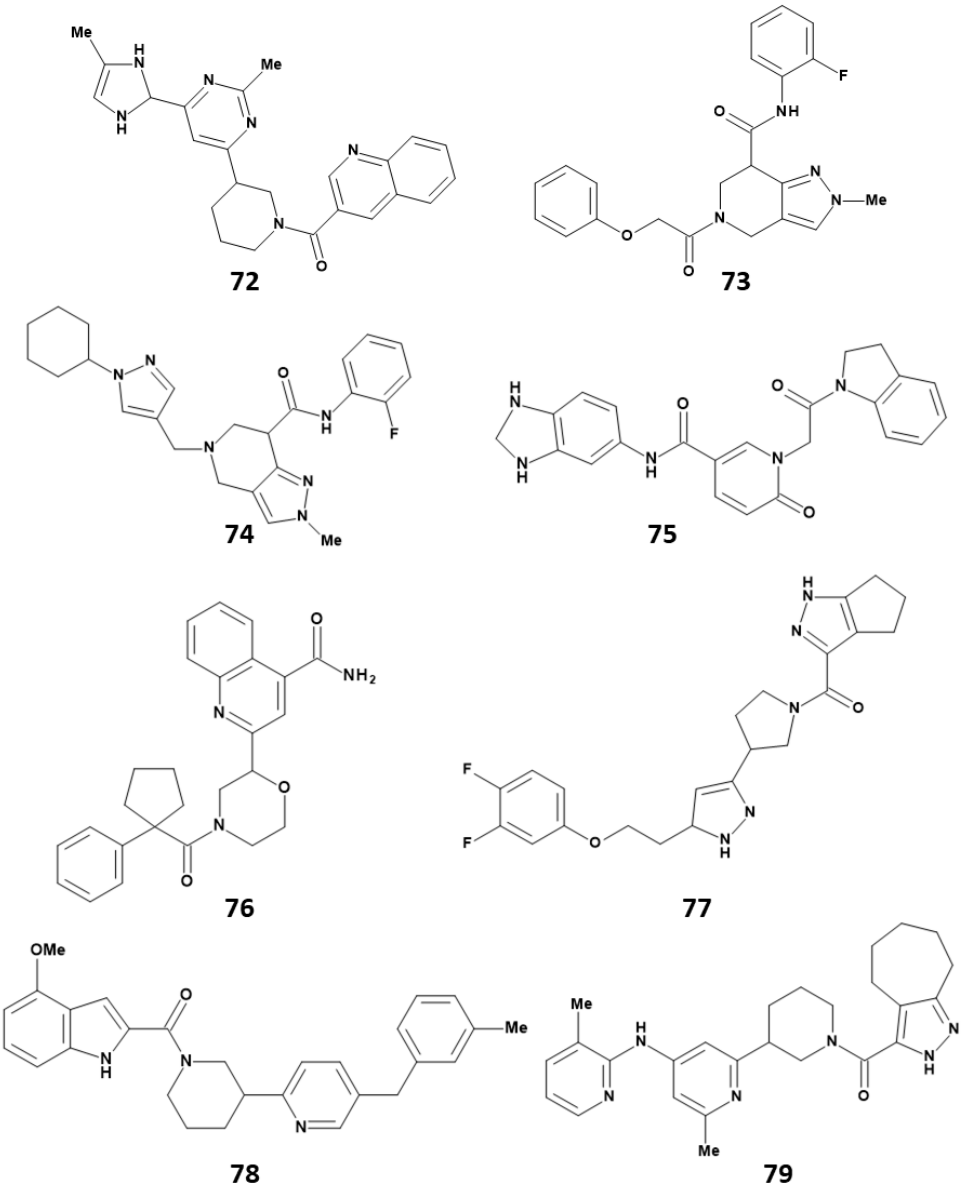


Chart16. 2D chemical structures of the eight compounds selected from the docking studies.

In Figure 36, a plausible binding mode for each of the selected molecule is depicted.

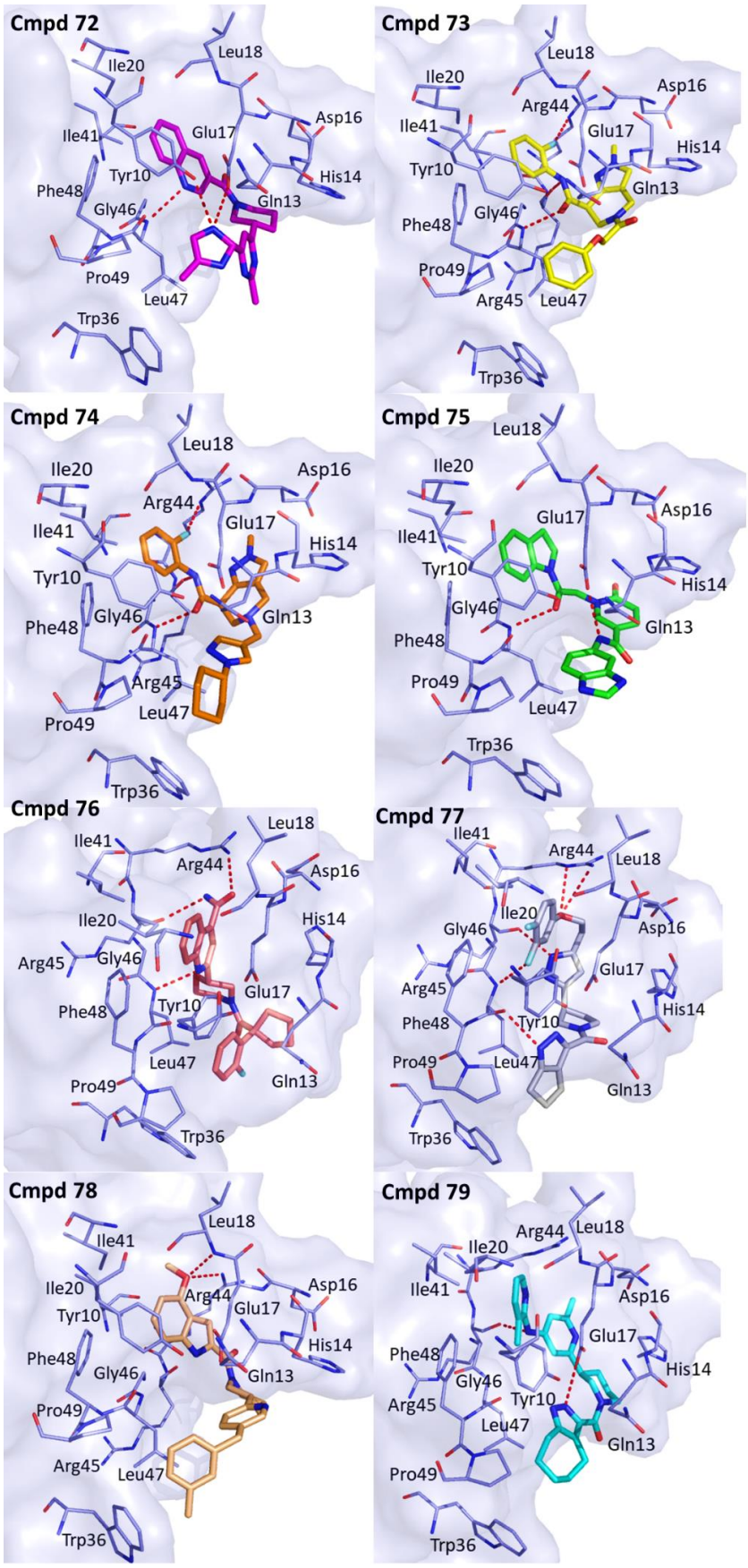


Figure36. Plausible binding mode predicted for compounds **72**-**79**. Each compound is represented in sticks whereas the amino acids residues involving in the interactions are highlighted as lightblue sticks. H-bonds are represented as red dashed lines. The image was created by means of PyMOL software (www.pymol.org).

All the compounds might bind to CIN85 SH3A domain at the previously found binding site P1/CS1, by establishing van der Waals interactions with Gln13, His14, Asp16, Leu18, Ile20, Ile41, Gly46 and Phe48. In details, compounds **72**-**77** could form a H-bond with the NH of the backbone of Leu47, while **72**, **75** and **79** might engage H-bond with the side chain of Glu17. Moreover, H-bonds involving Arg44 and Arg45 were observed for compounds **73**, **74**, **76**-**78** and **73**, **74**, **76**, **77** and **79**, respectively. Further H-bonds were detected between **72** and Tyr10 and between **78** and Leu18.

Moreover, π-π interactions were found between: i) Tyr10 and **73**-**76**, ii) His14 and **75**, and iii) Trp36 and **78**. Finally, all the compounds might engage hydrophobic contacts with residues Tyr10, His14, Trp36, Leu47, Phe48 and Pro49.

To validate the *in silico* study, compounds **72**-**79** were purchased and their biological activity on MUC1-CIN85 PPI is under investigation thanks to a collaboration with Dr. Sandra Cascio from University of Pittsburgh and Ri.MED Foundation.

7.3 Materials and Methods

7.3.1 MD simulation and frames clustering

MD simulation was carried out by using the OpenMM implementation of CHARMM [232] employing the 3D coordinates of the human SH3 N-terminal domain (SH3A) of CIN85 extracted from the crystal structure of CIN85 SH3 domain in complex with the Cbl-b peptide (PDB code 2BZ8) [213]. Specifically, chain A was maintained, while chain B and Cbl-b peptide were removed. The web-based graphical interface CHARMM-gui [237] (http://www.charmm-gui.org/) was used to set up the simulation and to generate the initial input files. The protein was solvated in a rectangular box employing the TIP3P water model and neutralized with KCl setting a salt concentration of 0.15M. The system was initially equilibrated for 25 ps in NVT ensemble using a time step of 1fs; then it was simulated in NPT ensemble for 400ns with a time step of 2fs at 310K by using Langevin dynamics whereas pressure was kept around 1 atm by Monte Carlo barostat. The coordinates were saved every 10 ps which resulted in 40,000 frames. Visualization and analysis of the trajectory were carried out by means of the Visual Molecular Dynamics (VMD) software[238].

The RMSD of the protein backbone was calculated by using the "RMSD Trajectory Tool" implemented in VMD.

The obtained snapshots were clustered according to RMSD by TTClust [233] a Python program for hierarchical clustering based on the RMSD of atomic coordinates. Firstly, the program aligns the trajectory on protein backbone. Subsequently, the RMSD between all atom pairs for each frame is computed and stored in a matrix which is then exploited to calculate a linkage matrix. Ward algorithm was selected for the linkage matrix calculation.

7.3.2 Binding site detection

The prediction of the putative druggable binding sites was performed for the x-ray solved SH3A domain of CIN85 and for the MD derived protein conformations by means of F-pocket [229] and FTMap web tool [230] (https://ftmap.bu.edu). Only the binding sites predicted as druggable by the two approaches were considered in this work. F-pocket and FTMap analyses were both executed using default parameter settings.

7.3.3 Docking studies

Docking studies were performed on the identified binding site P1/CS1 by means of Autodock Vina [236]. A grid box of dimension 22.50x20.25x20.25 Å and centre x= 7.574, y= 39.574, z= 27.609 was applied as search space in order to include all the amino acid residues belonging to the above-mentioned binding site. Maximum 9 poses per ligand were generated and docking calculations were performed by using the default settings. The visualization and the analysis of the ligand-protein interactions were executed by Discovery Studio Visualizer [179].

Chapter 8

Exploration of the molecular interactions of MUC1 at CIN85 binding interface

8.1 Introduction

From April 2020 to September 2020, I spent a research internship at Ri.MED Foundation of Palermo, working under the supervision of Dr. Ugo Perricone. The research activity performed during this stage was focused on the exploration of the molecular contacts between MUC1 and CIN85.

MUC-CIN85 PPI represents a viable target for the development of novel anti-metastatic agents. However, the lack of experimentally solved structures of the complex hampers the design of potential MUC1-CIN85 inhibitors. To address drug discovery efforts, an interaction model of this complex was generated in order to investigate the molecular contacts between the two proteins. As reported in Chapter6, experimental evidence suggested that CIN85 would engage MUC1 as dimer similarly to Cbl-b peptide; however, it is well known that SH3 domain of CIN85 can bind Pro-rich sequences both in its monomeric and dimeric forms. In light of this, a 3D model was generated both for the heterodimeric and heterotrimeric MUC1-CIN85 complex by exploiting protein-peptide docking. MD simulation was subsequently used to probe the stability of the obtained complexes and to further explore the interactions occurring between MUC1 and CIN85. For a comparison purpose MD simulation was also carried out on the co-crystal structure of CIN85-Cbl-b complex (PDB 2BZ8).

Finally, the average binding free energies of the obtained MD snapshots were computed by MM-GBSA method and the results were compared to evaluate which of the two MUC1-CIN85 complexes might be energetically favourable.

The content of this chapter has been object of one submitted manuscript:

- *Maria Rita Gulotta, **Serena Vittorio**, Rosaria Gitto, Ugo Perricone and Laura De Luca. Exploring molecular contacts of MUC1 at CIN85 binding interface to address future drug design efforts. Submitted on January 18, 2021.*

8.2 Results and discussion

8.2.1 Analysis of the crystal structure of CIN85-Cbl-b complex and MD investigation

The first step of this work consists in the analysis of the co-crystal structure (PDB code 2BZ8) of the SH3A domain of CIN85 with the 11-mer Cbl-b derived peptide (PARPPKPRPRR) containing the proline-arginine binding motif PKPRPR [213]. As described in Chapter6, in this complex, Cbl-b is sandwiched between two SH3 domains and adopts a pseudo-symmetrical orientation. Indeed, one SH3 domain binds the Pro-rich peptide adopting a type I orientation while the other SH3 domain engages Cbl-b in a type II orientation as displayed in Figure37. Moreover, isothermal titration calorimetry (ITC) experiments conducted by using the two separated complexes with a 1:1 stoichiometry, showed that Cbl-b predilects a type II orientation to bind CIN85 SH3A domain. Indeed, the K_d for the type I orientation is equal to 46.9 µM while a K_d value of 2.0 µM was obtained for the type II interaction [217].

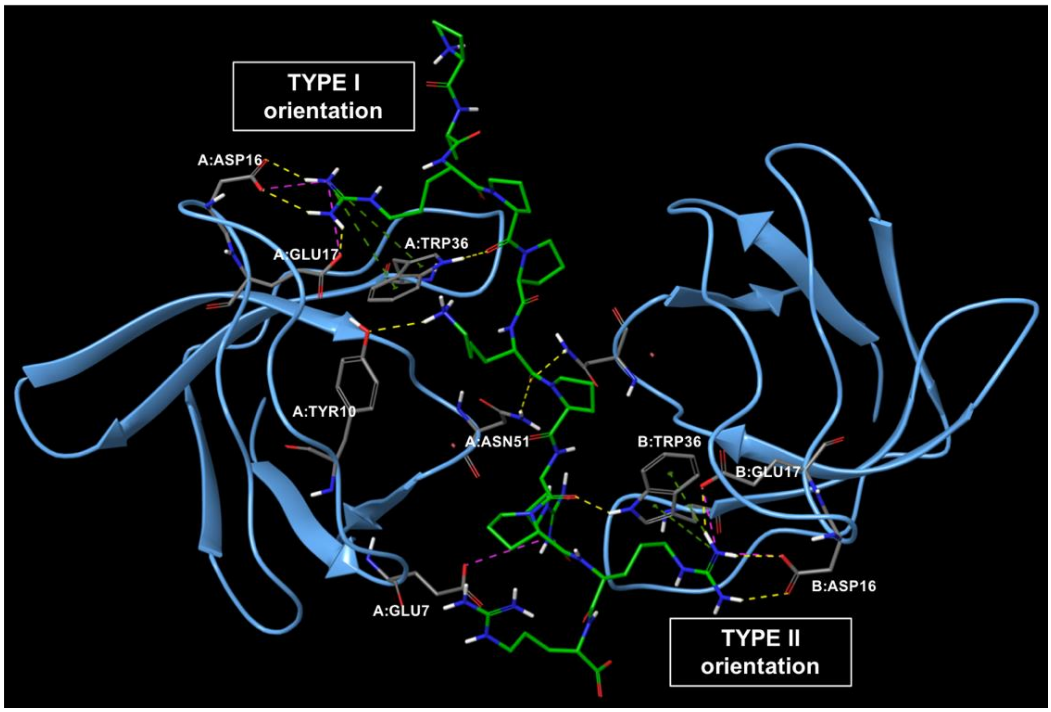


Figure37. Crystal structure of the experimentally solved SH3A domain of CIN85 in complex with Cbl-b peptide (PDB code 2BZ8). One SH3 domain binds Cbl-b with a type I orientation with the N-terminal Arg engaging contacts with the acidic pocket, while the other SH3 domain engage the peptide assuming a type II orientation with the Arg residue, C-terminal to the binding motif, occupies the acidic pocket.

As shown in Figure38, the analysis of the interactions revealed that one SH3 domain binds Cbl-b peptide by establishing five H-bonds, formed respectively between: i) the backbone of Lys907 of Cbl-b and Asn51; ii) the side chain of Lys907 and Tyr10; iii) the carbonyl backbone of Pro905 and Trp36; iv) the side chain of Arg904 and Asp16 and Glu17. Moreover, salt bridges were observed between i) Arg909 and Glu7, and ii) Arg904 and Asp16 and Glu17. Finally, Trp36 establishes π-cation interaction with Arg904. Instead, the second SH3 domain interacts with Cbl-b by forming four H-bonds between: i) the carbonyl backbone of Arg909 and Trp36; ii) the carbonyl backbone of Lys907 and Asn51; iii) the side chain of Arg911 and Asp16 and Glu17. Furthermore, salt bridges are established between Arg911 and Asp16 and Glu17 and π-cation interaction is engaged between Arg911 and Trp36. Both SH3 domains engage hydrophobic contacts with Pro906, Pro908 and Pro910 of Cbl-b, through their Trp36 and Phe52 residues.

As described, above both Arg904 and Arg911 form an extensive network of interactions with Asp16, Glu17 and Trp36 of CIN85. Previous studies had already shown the importance of Arg911 (underlined in PXXXP<u>R</u>) for the binding to CIN85, while no information have been reported regarding the role of the N-terminal Arg904 in complex formation.

In order to deeply investigate the functional role of the two arginines in full-length Cbl-b, Jozic *et al.* performed mutational studies by separately mutating the two residues to alanine. The amount of CIN85 co-precipitated with Cbl-b mutated peptides was quantified showing that mutation of Arg911 leads to a reduction of the interaction by ~60%, while mutation of Arg904 decreases the binding by ~25%. Furthermore, the mutation of both arginine residues of Cbl-b was performed as well, leading to the abolishment of its co-precipitation with CIN85. Basing on these outcomes, the authors speculated that Arg904 could be important for the multimerization of the complex [213]. Other studies present in literatures pointed out the key role of Pro906 and 911 of Cbl-b, and Trp36 of CIN85 for the binding of SH3A domain of CIN85 with Cbl-b.

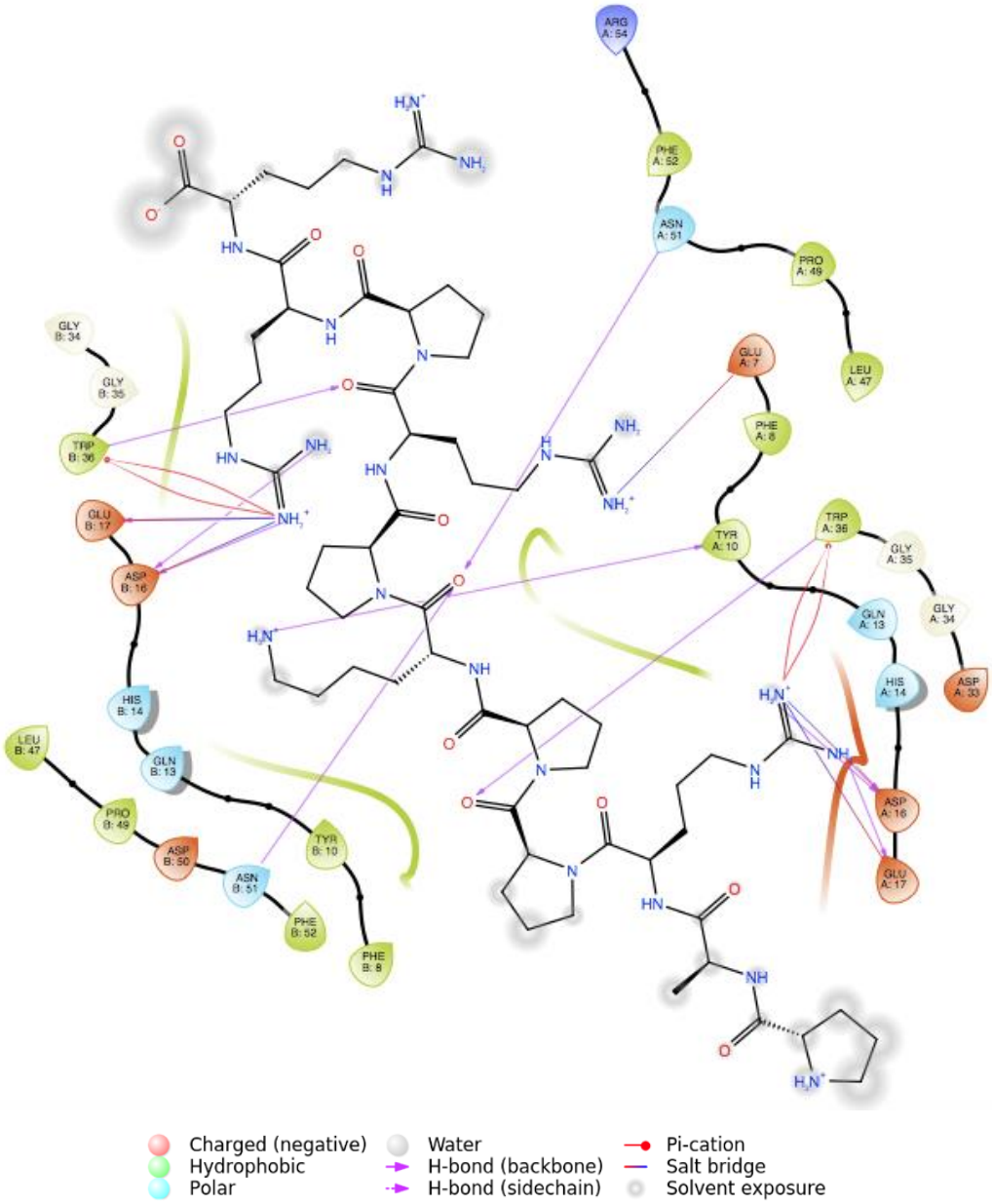


Figure38. 2D depiction of the interactions found in the crystal structure of CIN85 SH3A-Cbl-b crystallographic complex.

To probe the stability of the interactions observed in the crystal structure, two independent 50 ns MD simulations were run by using Desmond software. The RMSD of both protein and ligand was computed and plotted for the two simulations as shown in FigureS1 and S2 of Supporting Materials. The obtained trajectories furnished similar information regarding the frequency of the peptide-protein interactions. Indeed, as displayed in Figures39 and 40, the most stable interactions are those found in the crystallographic structure, except for an additional hydrophobic contact established between Phe8 of CIN85 and Pro906 of

Cbl-b which appeared and was stable during the two simulations. The 2D depiction of the interactions observed in the two trajectories are reported in Figures S3 and S4 of Supporting Materials. Overall, these data provide useful insights to guide the computational investigation aimed at clarifying the molecular recognition pattern of MUC1-CIN85 complex.

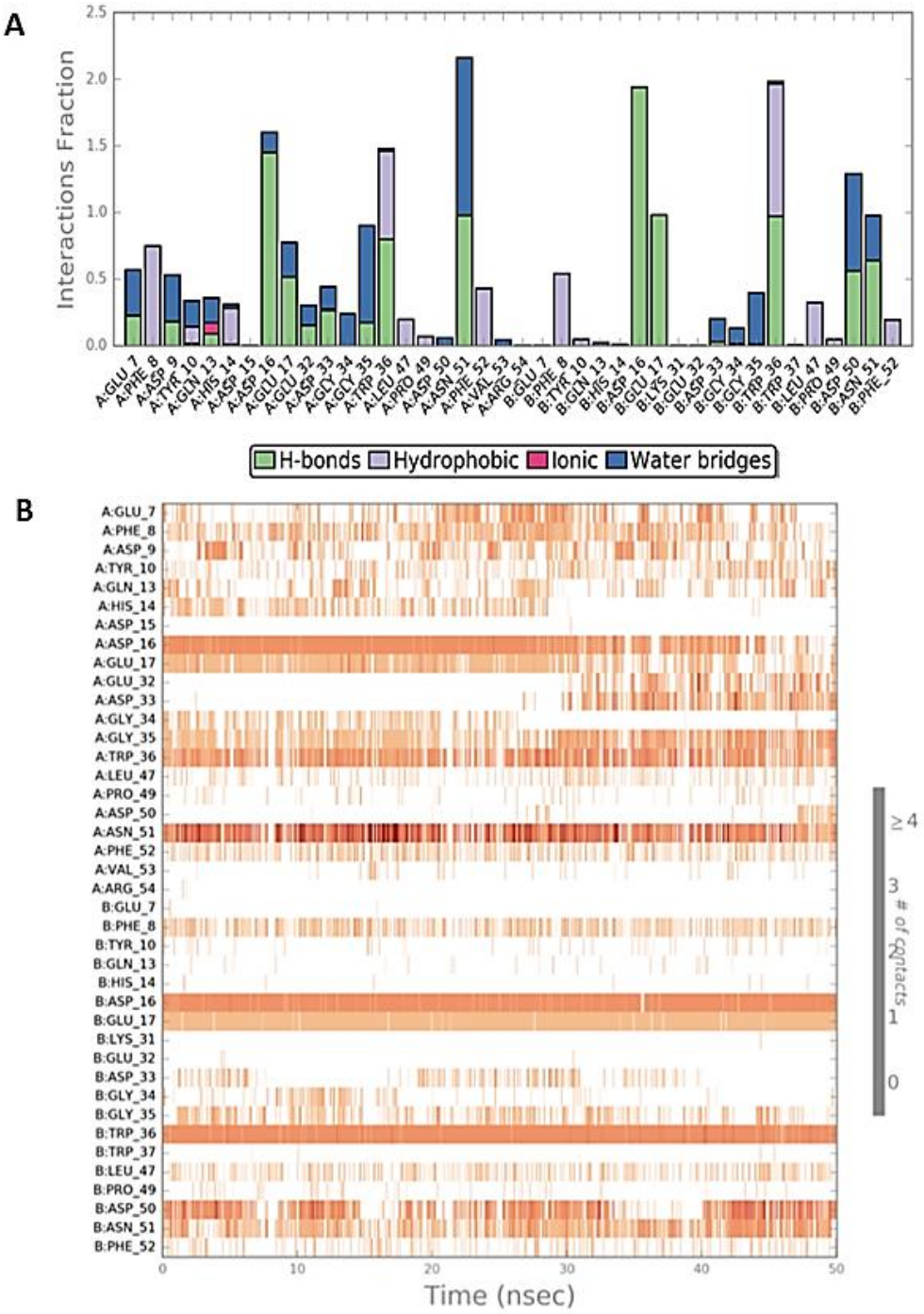


Figure39. Schematic representations of the interactions observed during the first MD simulation of the experimentally solved complex CIN85 SH3A-Cbl-b. A) Histogram plot of the protein-ligand contacts. B) Timeline representation of the interactions during the MD trajectory.

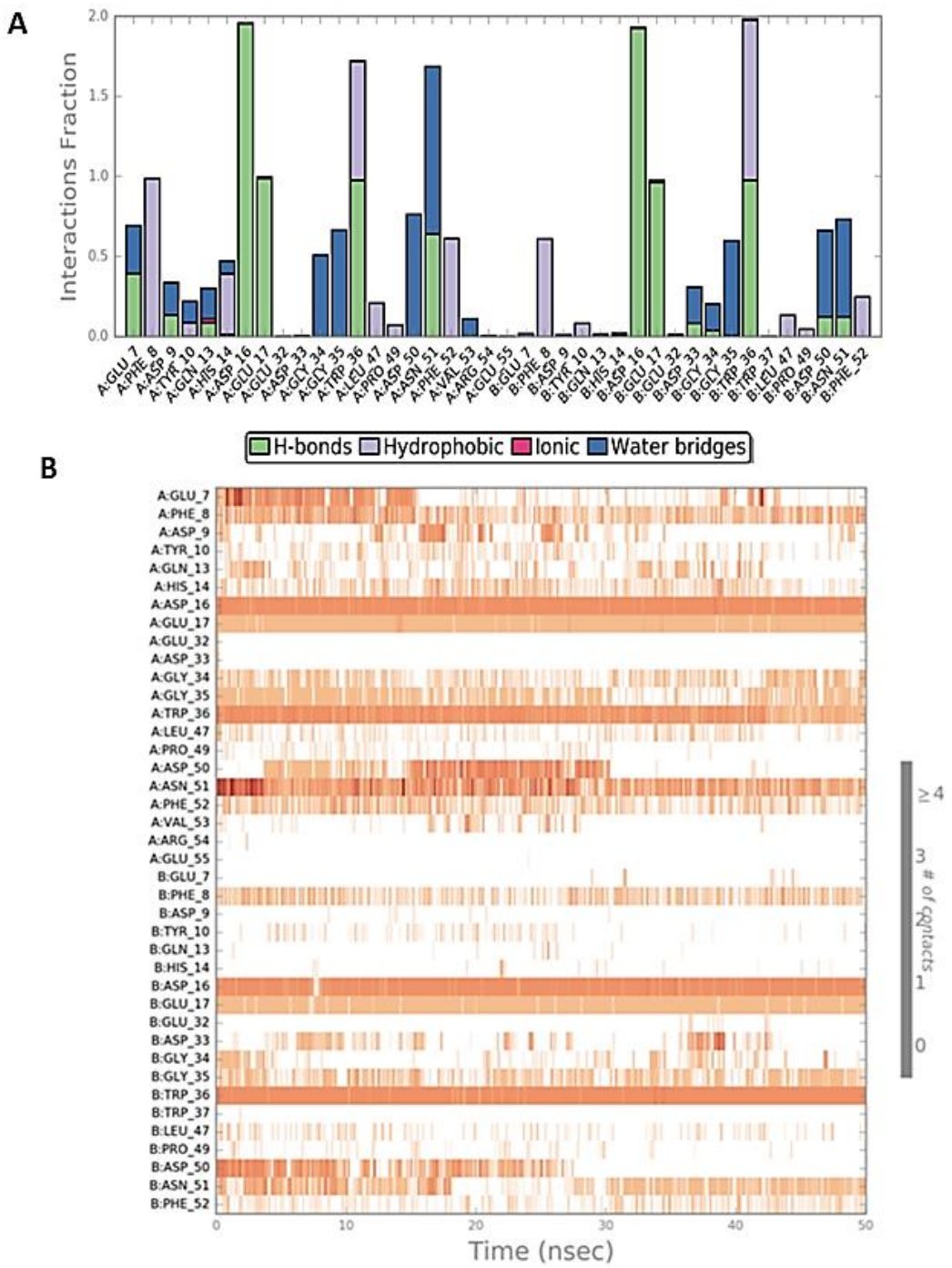


Figure40. Schematic representations of the interactions observed during the second MD simulation of the experimentally solved complex CIN85 SH3A-Cbl-b. A) Histogram plot of the protein-ligand contacts. B) Timeline representation of the interactions during the MD trajectory.

8.2.2 Protein-peptide docking of MUC1-CIN85 as heterotrimeric complex and MD exploration

As mentioned in the introduction, the scope of this study was to explore the putative ligand binding mode of MUC1 at CIN85 SH3 domain binding surface. For this purpose, protein-peptide docking was exploited by using the 3D structure of a

hypo-glycosylated MUC1 derived peptide containing the binding motif PDTRP extracted from the crystal structure of MUC1 glycopeptide in complex with the antibody SM3 (PDB code 5OWP) [239]. In particular, this peptide is characterized by the following sequence SAPDT*RPAP present in the VNTR of the extracellular domain of MUC1 and includes a sugar moiety (the 2-acetamido-2-deoxy-alpha-D-galactopyranose) linked to Thr5. The antibody subunits were removed and the peptide was prepared for docking simulations as described in Materials and methods section.

Firstly, we probed the capability of MUC1 peptide to simultaneously engages two SH3 domains of CIN85 similarly to Cbl-b. To this aim, the atomic coordinates of CIN85 SH3 domains co-crystallised with Cbl-b peptide (PDB code 2BZ8), were used as target protein for the peptide docking, performed by using Glide software [240, 241].

Among the obtained docking solutions, the conformation of the best scored pose resembles quite well that of the co-crystallised peptide Cbl-b, while in the other solutions the peptide assumed highly folded conformations. Therefore, the docking pose with the lowest score (gscore= -9.802 kcal/mol) was chosen for further analysis (Figure41).

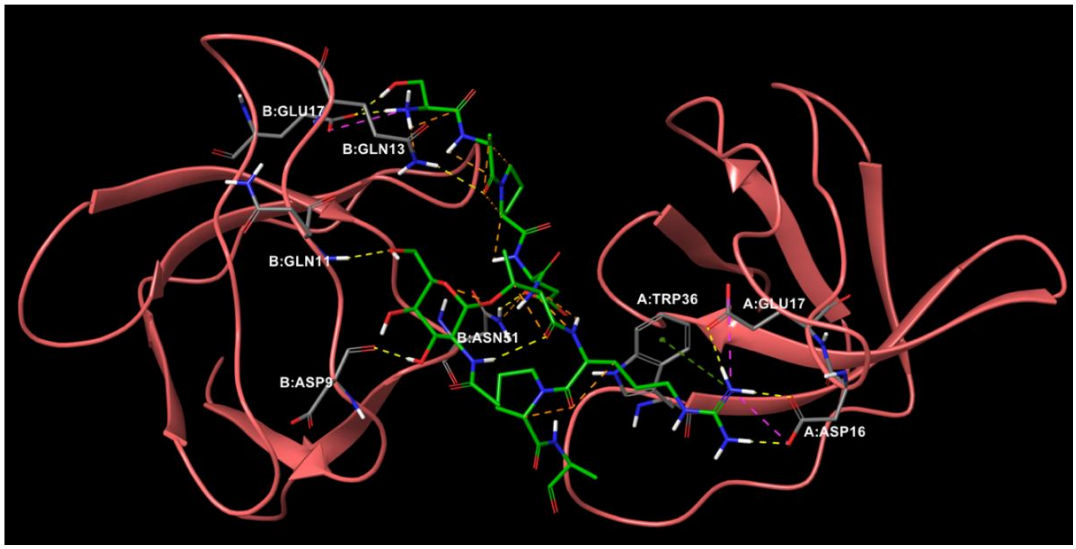


Figure41. Plausible binding mode of MUC1 peptide to CIN85 SH3A domain in its dimeric form. MUC1 peptide is represented as green sticks. The residues of CIN85 involved in the interactions are represented as grey sticks. The different types of interactions are highlighted as colour coded dotted lines: H-bonds are represented in yellow, salt bridge in violet and π-cation in green.

As shown in Figure42, MUC1 peptide could bind the SH3A domains of CIN85 establishing a network of interactions similar to that observed for Cbl-b. In details, one SH3 domain might engage MUC1 peptide by forming five H-bonds between i) the side chain of Asp4 of MUC1 and Asn51, ii) the side chain of Ser1 and Glu17, iii) the carbonyl backbone of Ala2 and Gln13, iv) the hydroxyl group attached to C6 of the sugar moiety and Gln11 and v) the hydroxyl group attached to C3 of the sugar moiety and Asp9. Moreover, hydrophobic contacts might be established between Phe52 and Pro3 of MUC1 peptide. Instead, the interactions observed with the other SH3 domain are mediated by the side chain of Arg6 of MUC1 that can establish π-cation interaction with Trp36 and H-bond and salt bridge both with Asp16 and Glu17.

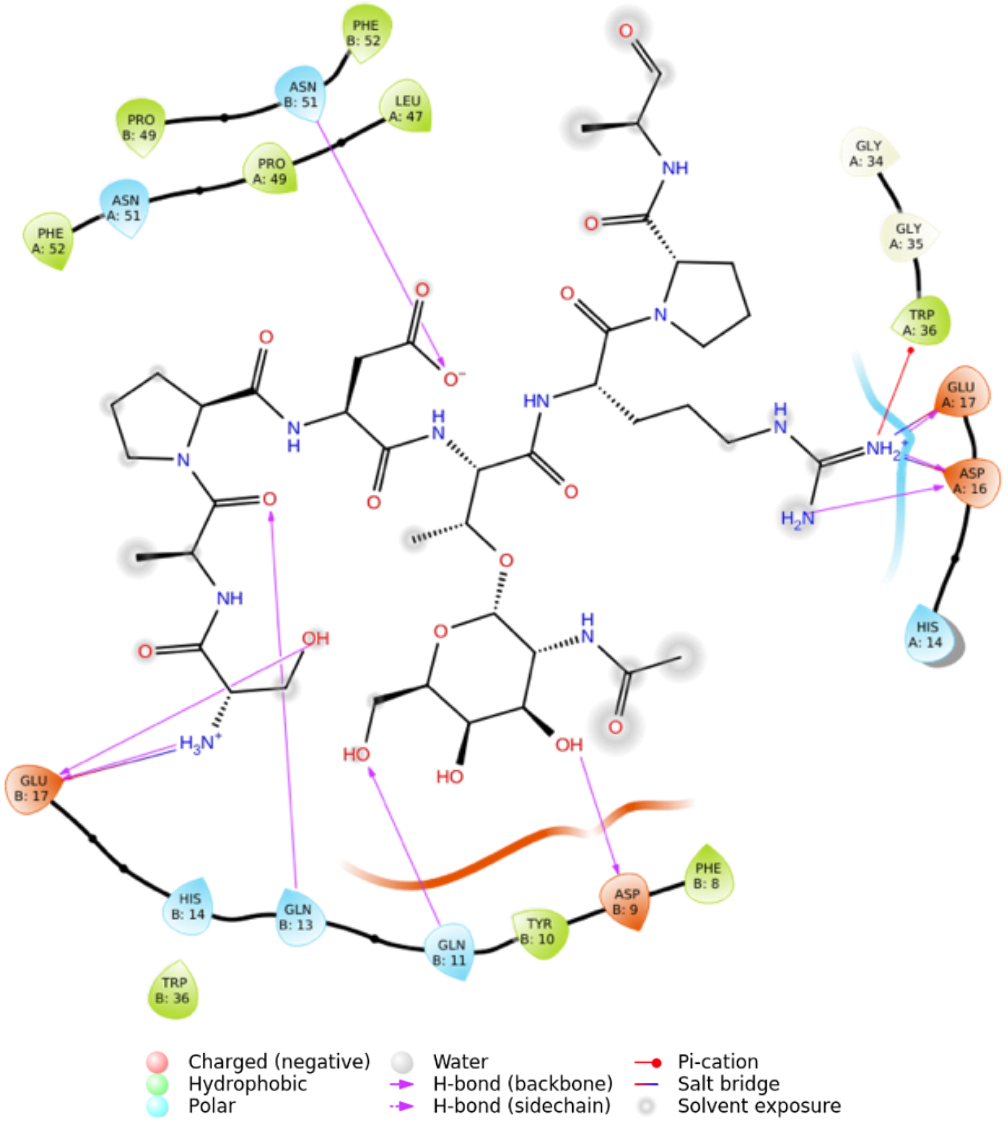


Figure42. 2D depiction of the interactions occurring in MUC1-CIN85 SH3A heterotrimeric complex obtained from the peptide docking studies.

To evaluate the stability of the interactions retrieved form docking experiments, the obtained complex was subjected to two independent 50 ns MD simulation by means of Desmond software [242]. The RMSD was calculated both for the protein and the peptide as shown in Figures S5 and S6 of Supporting Materials. From the analysis of the MD trajectories emerged that Asp16, Glu17 and Trp36 of CIN85 mediated the most stable interactions with MUC1, particularly with Arg6 (Figures 43 and 44). Differently from CIN85 SH3-Cbl-b complex, the interactions involving Asn51 are not stable especially in the second trajectory (Figure44). In Supporting Materials, the 2D representation of the protein-peptide contacts detected during the two trajectories is provided (Figures S7 and S8)

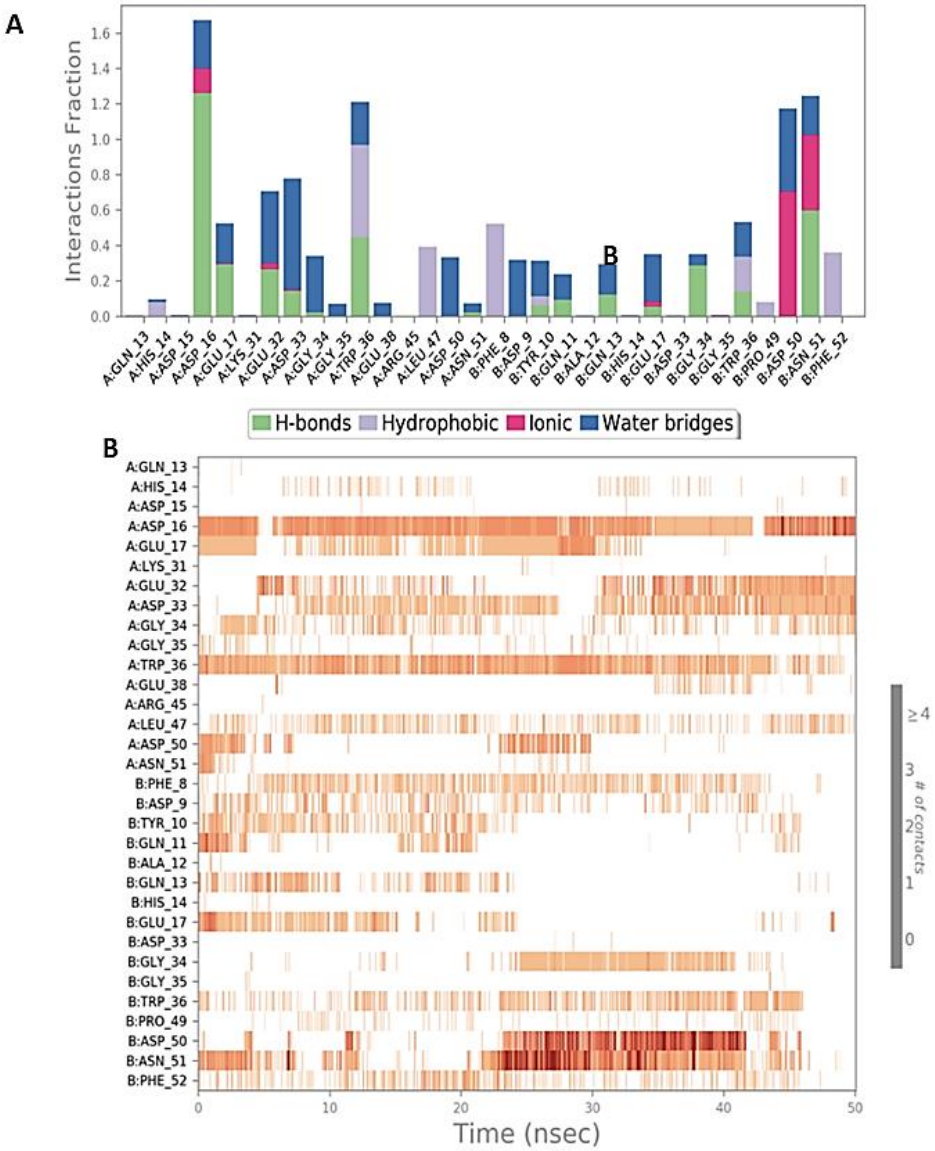


Figure43. Schematic representations of the interactions occurring in the first MD simulation of MUC1-CIN85 SH3A heterotrimeric complex obtained from the docking studies. A) Histogram plot of the protein-peptide contacts. B) Timeline representation of the interactions during the MD simulation.

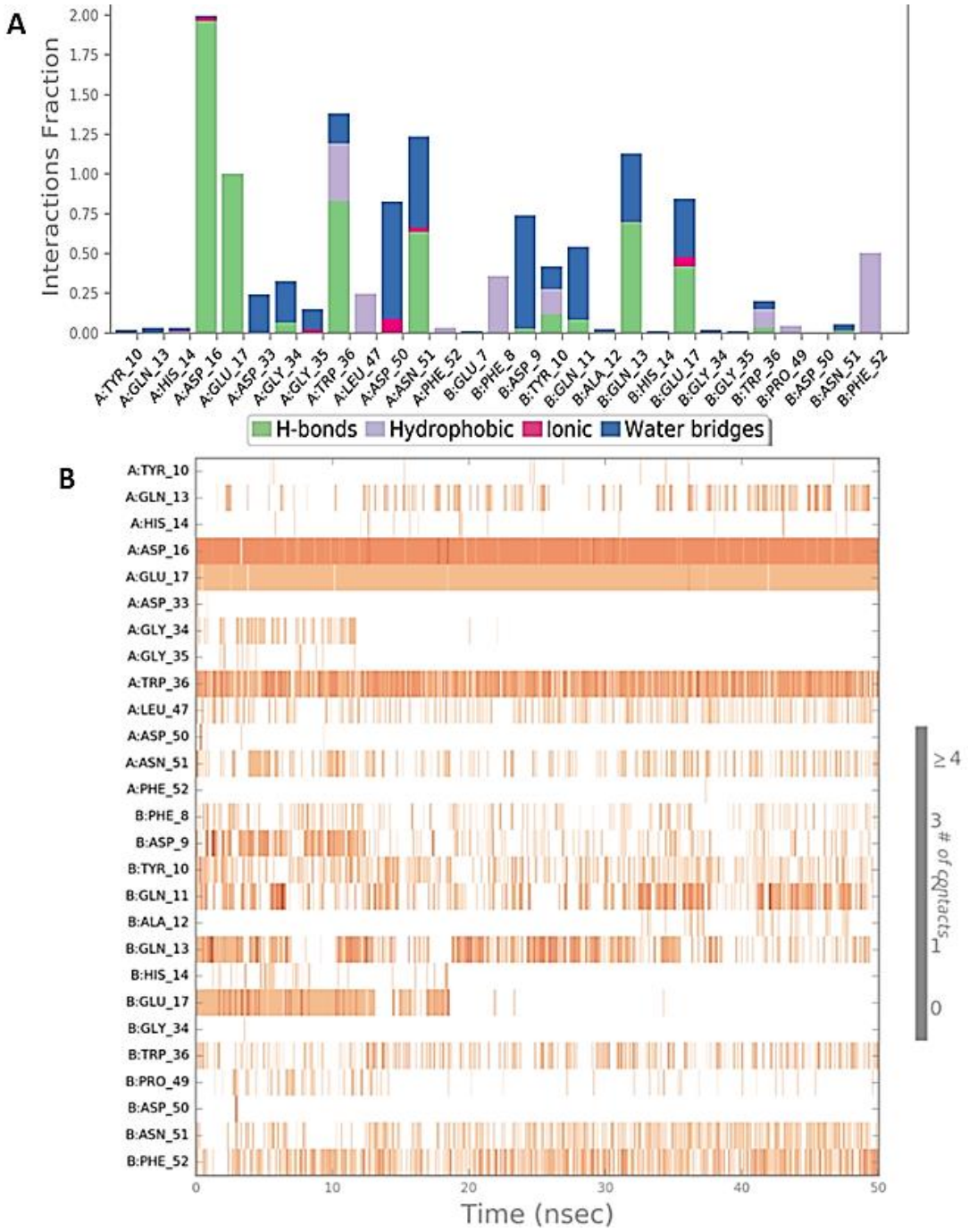


Figure44. Schematic representations of the interactions occurring in the second MD simulation of MUC1-CIN85 SH3A heterotrimeric complex obtained from the docking studies. A) Histogram plot of the protein-peptide contacts. B) Timeline representation of the interactions during the MD simulation.

The data reported above were collected and compared with those obtained from docking and MD simulations of MUC1 in complex with the monomeric state of CIN85 SH3 domain described in the following paragraph.

8.2.3 Peptide docking and MD simulation of CIN85 monomer and MUC1 peptide

The capability of SH3A domain of CIN85 to bind MUC1 forming a heterodimeric complex was probed through docking studies performed by Glide software [240,241]. To this aim, to allow an easier and reliable comparison with the former results, the 3D coordinates of a single SH3A domain of CIN85 were extracted from the complex (PDB code 2BZ8) used previously and exploited as target protein for the docking runs. Also in this case MUC1 VNTR derived peptide retrieved from the PDB structure 5OWP was employed for the docking studies. The best predicted docking pose (gscore= -3.698 kcal/mol), depicted in Figure45, was chosen and the putative protein-peptide interactions were analysed.

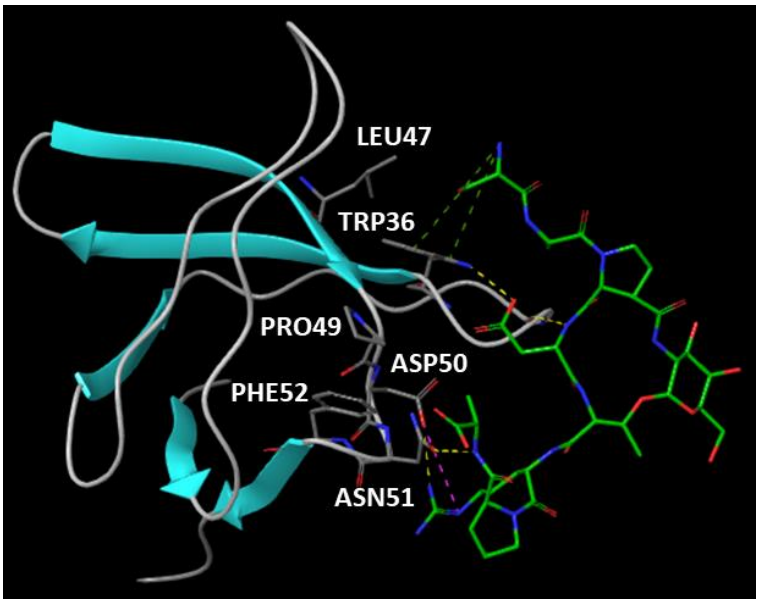


Figure45. A) Plausible binding mode of MUC1 peptide with CIN85 SH3A domain. MUC1 VNTR derived peptide is depicted as green sticks while the amino acids residues of the binding surface of SH3A domain are represented as grey sticks. The different types of interactions are shown as colour coded dotted lines: H-bonds are represented in yellow, salt bridge in violet and π-cation in green.

In details, MUC1 peptide might bind the SH3 domain of CIN85 by forming three H-bonds between: i) the backbone NH of Asp 4 of MUC1 and Gly34 of CIN85, ii) the NH of the backbone of Ala8 and the side chain of Asn51 and iii) the side chain of Arg6 and carboxylic group of Asp50. Additionally, a salt bridge was detected between Arg6 and Asp50 (Figure46).

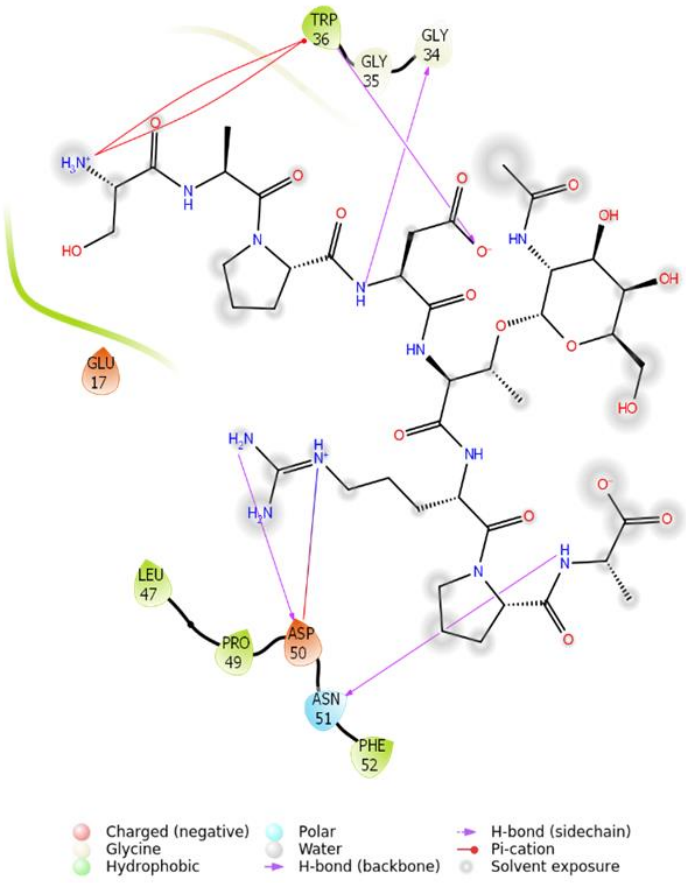


Figure46. 2D illustration of the protein-peptide interactions detected in MUC1-CIN85 SH3A heterodimeric complex obtained from docking.

The above-described complex was subjected to two short independent 50 ns MD simulations by means of Desmond software [242] to further explore the interactions between MUC1 peptide and the SH3A domain of CIN85. The RMSD was computed for both trajectories and plotted as displayed in Figures S9 and S10. Then, the protein-peptide interactions were analysed to deeply investigate their stability during the two simulations. The results showed that in both trajectories the contacts mediated by Asp50 and Asn51 of CIN85 disappeared while the H-bond between Trp36 and the Asp4 of MUC1 binding motif remained stable throughout the simulation time (Figure47 and 48). Moreover, additional interactions to those detected in the docking complex were observed in both MD outputs. Concerning the first simulation, a further H-bond between the NH of the backbone of Asp4 of MUC1 and Asp33 of CIN85 get formed at around 8 ns and stays stable for the rest of the trajectory (Figure47B). Instead, in the second MD output, two new H-bonds were found between Asp16 and Glu17 of CIN85 and the N-terminal Ser1 of MUC1

peptide which appear at the beginning of simulation and remain stable throughout the trajectory (Figure48B). Furthermore, the 2D depiction of the interactions occurring in the two trajectories are displayed in Supporting Materials (Figures S11 and S12).

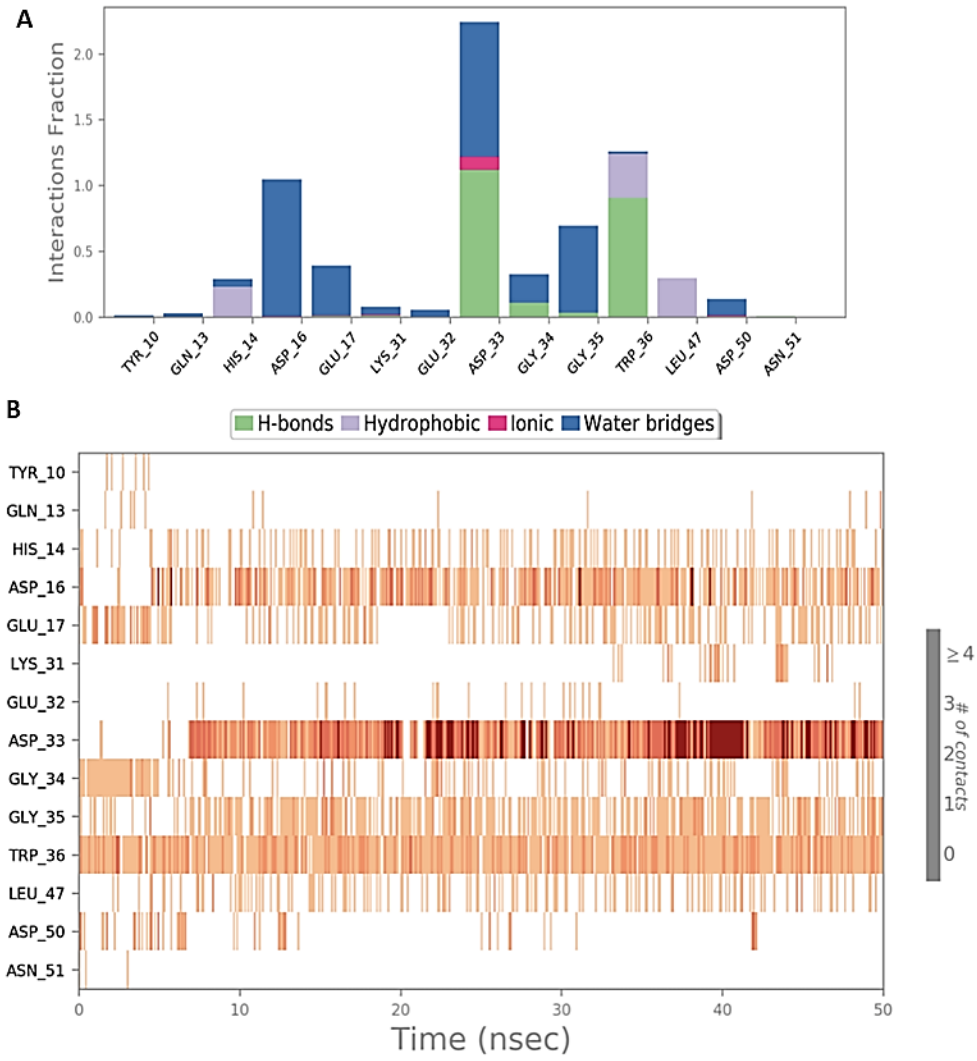


Figure47. Schematic representations of the interactions occurring in the first MD simulation of obtained from the docking studies. A) Histogram plot of the protein-peptide contacts. B) Timeline representation of the interactions during the MD simulation.

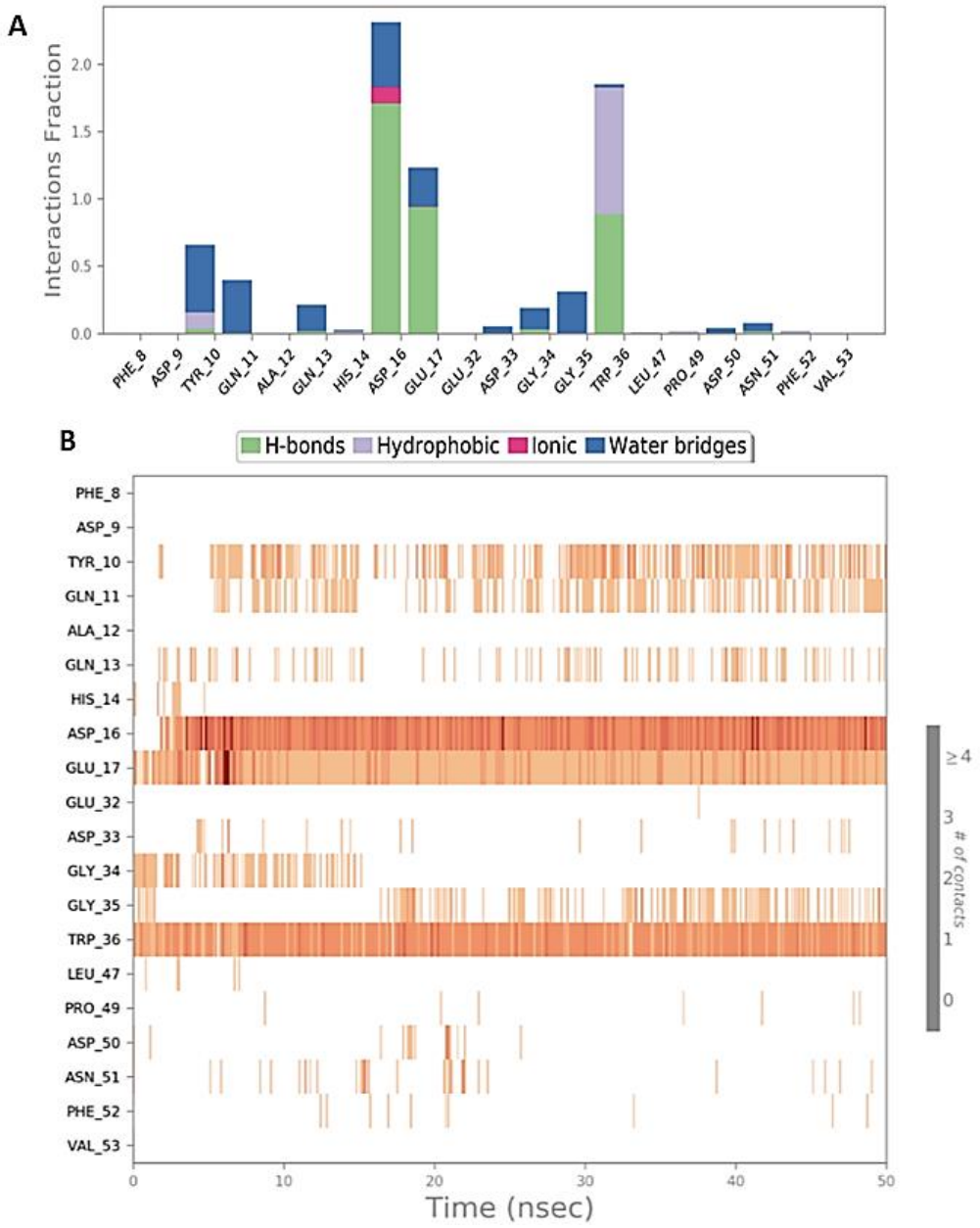


Figure48. Schematic representations of the interactions occurring in the second MD simulation of MUC1-CIN85 SH3A heterodimeric complex obtained from the docking studies. A) Histogram plot of the protein-peptide contacts. B) Timeline representation of the interactions during the MD simulation.

Based on the above reported outcomes, Trp36 of CIN85 is involved in the most stable interactions both in the heterodimeric and heterotrimeric complex, in analogy also to Cbl-b peptide for which mutational studies have highlighted the pivotal role of Trp36 residue in the binding to Cbl-b. In particular, we observed that Asp4 of MUC1 mostly contributes to the formation of the heterodimeric complex through its interaction with Trp36, while Arg6 plays a central role in the generation of the heterotrimeric complex through the engagement of H-bonds with Asp16 and

Glu17 and π-cation interactions with Trp36. Additionally, in both MD trajectories involving CIN85 SH3A domain in its monomeric state, the residues of MUC1 included between Thr5 and Ala8 establish almost no contacts with CIN85 during the entire simulation time.

8.2.4 MM-GBSA calculation of MD simulation

The average binding free energy of the MD snapshots was estimated by MM-GBSA methods for each MD simulation, included the one related to the X-ray complex CIN85-Cbl-b. As reported in Table 3, the experimentally solved complex displayed the lowest $\Delta G_{binding}$ average values equal to -141.449 kcal/mol and -136.904 kcal/mol for the first and the second simulation, respectively. Concerning MUC1-CIN85 complexes, in the heterotrimeric form it reported $\Delta G_{binding}$ average values (-54.624 and -62.681 kcal/mol) approximately two-fold lower than those retrieved for the heterodimeric complex (-36.009 and -26.516 kcal/mol).

Table 3. $\Delta G_{binding}$ values calculated for each MD simulation.

		MD Simulations of the complexes		
		CIN85 dimer – Cbl-b peptide	**CIN85 dimer – MUC1 peptide**	**CIN85 monomer – MUC1 peptide**
First MD	**Average $\Delta G_{binding}$ (kcal/mol)**	-141.449	-54.624	-36.009
	$\Delta G_{binding}$ range (kcal/mol)	-164.158 to -112.656	-101.514 to -18.227	-55.595 to -10.065
Second MD	**Average $\Delta G_{binding}$ (kcal/mol)**	-136.904	-62.681	-26.516
	$\Delta G_{binding}$ range (kcal/mol)	-163.5629 to -116.8524	-109.318 to -34.366	-47.012 to -6.81

These results pointed out that, in terms of Gibbs free energy, MUC1 peptide should preferentially binds two SH3 domains of CIN85 simultaneously, as also suggested by the above-mentioned experimental studies according to which in the presence of a MUC1 peptide able to promote dimerization, the interaction between the two proteins was two-fold higher than in its absence. The data gained from MM-GBSA calculations are also in accordance with the MD output; indeed, while in the heterotrimeric complex the peptide remains appreciably bound to both SH3 domains of CIN85, in heterodimeric one some of the contacts found in the docking

result disappeared within the first nanoseconds of the simulation time letting the interested amino acid residues freely moving in the aqueous environment.

8.3 Materials and Methods

8.3.1 Protein-peptide docking

Protein-peptide docking was performed by employing the crystal structure of N-terminal SH3 domain of CIN85 bound to Cbl-b peptide (PDB code 2BZ8) [213] as protein target. The Cbl-peptide was removed; both chain A and B were used to simulate the dimeric state of CIN85, while only chain A was maintained to generate its monomeric state. MUC1 peptide containing the binding motif PDTRP present in the VNTR was extracted from the crystal structure of glycopeptide "SAfPDT*RPAP" in complex with scFv-SM3 (PDB code 5OWP) [239]. This MUC peptide structure contains a fluoroproline residue that was converted into natural Pro. The two PDB structures were prepared by using the “Protein preparation wizard” [243] implemented in Schrodinger Suite (release 2018-4). Two grids were generated for the docking calculation; for the first one that includes CIN85 in its dimeric form, the binding site was defined selecting Cbl-b peptide, while the second grid, concerning the monomeric state, was set by selecting the conserved amino acids present in the binding surface of CIN85 SH3A domain Phe8, Tyr10, Asp16, Glu17, Trp36, Pro49 and Phe52. Furthermore, for both systems the option to create a grid suitable for peptide docking was checked. 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 OPLS3e [244]. Then, the docking computations were performed by using the “ligand docking” tool of Schrödinger suite [240,241]. The selected protocol was “SP-Peptide” and the flexible ligand sampling method was chosen. 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.

8.3.2 MD simulations and MM-GBSA calculations

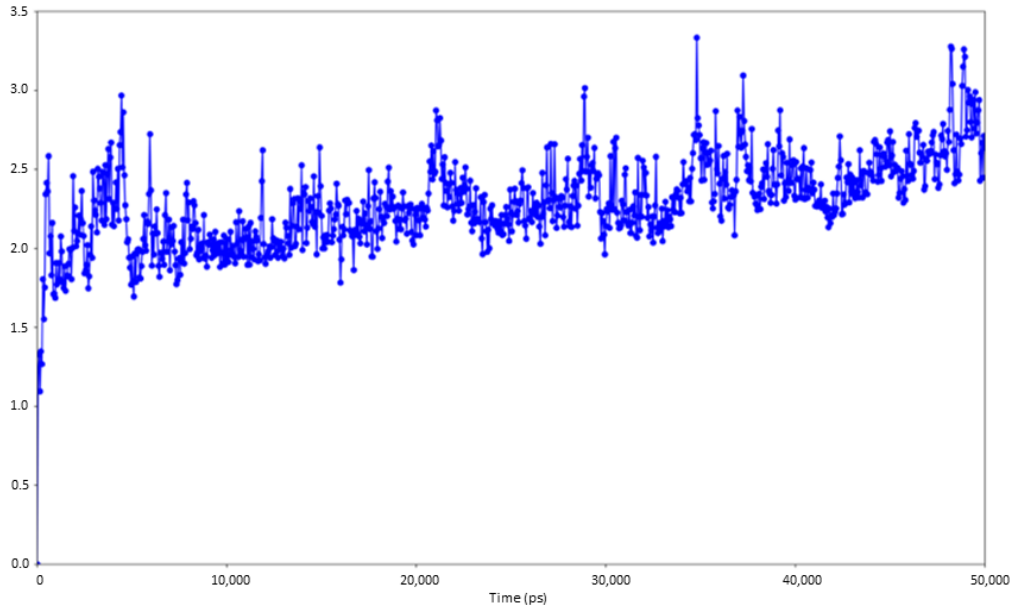
Two independent 50 ns MD simulation were run for each complex by using Desmond software [242]. Each system was solvated in an orthorhombic box using TIP3P water model and neutralized by adding Na+ ions through the “System builder” tool implemented in Schrödinger suite. The simulation was set up through

the “Molecular dynamics” tool. The system was relaxed before the simulation by using the protocol implemented in Desmond and then it was simulated for 50 ns by NTP ensemble at 300 K and 1 atm by using respectively and Nose-Hoover chain thermostat and Martyna-Tobias-Klein barostat. Random seeds were employed as starting point of MD simulation and the force field was set as OPLS3e[244]. Finally, the obtained trajectories were used to compute MM-GBSA calculations through Schrödinger suite by exploiting the Python script “thermal_mmgbsa.py”.

8.4 Supporting Materials

Figure S1. RMSD plots related to the first MD simulation of X-ray structure Cbl-b-CIN85.

A) RMSD Plot of protein heavy atoms



B) RMSD Plot of ligand

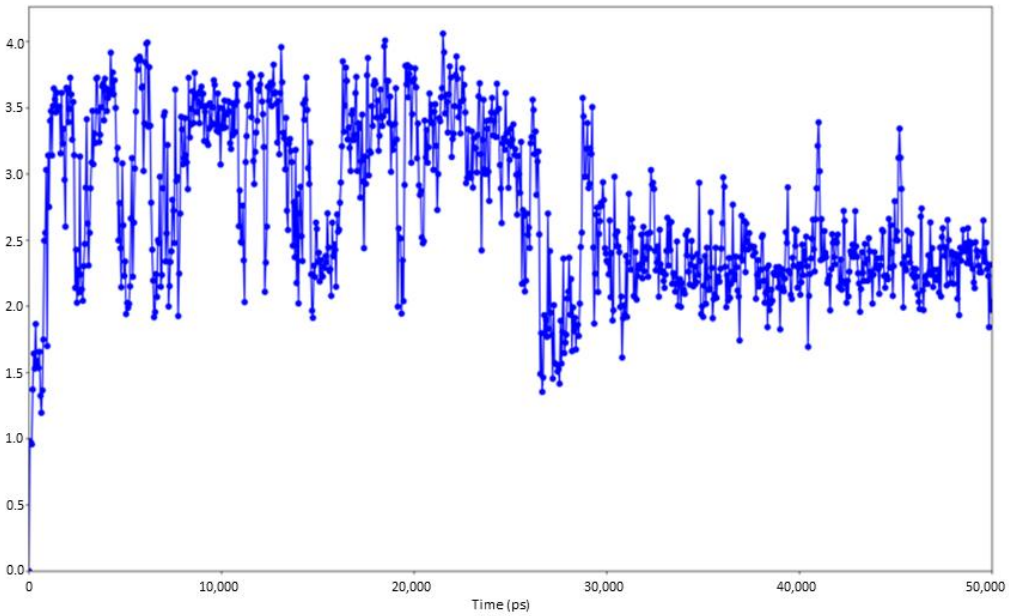
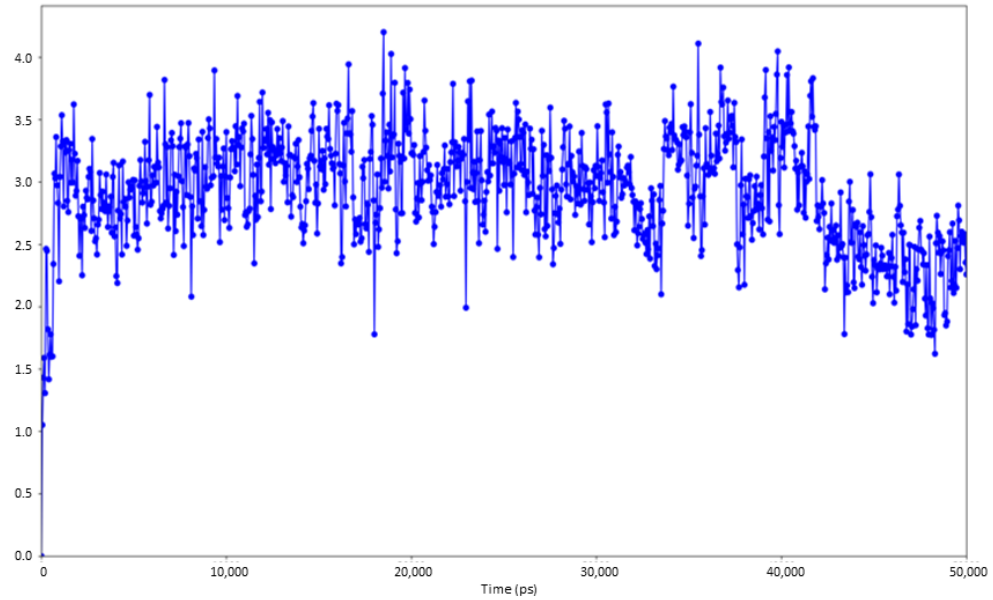
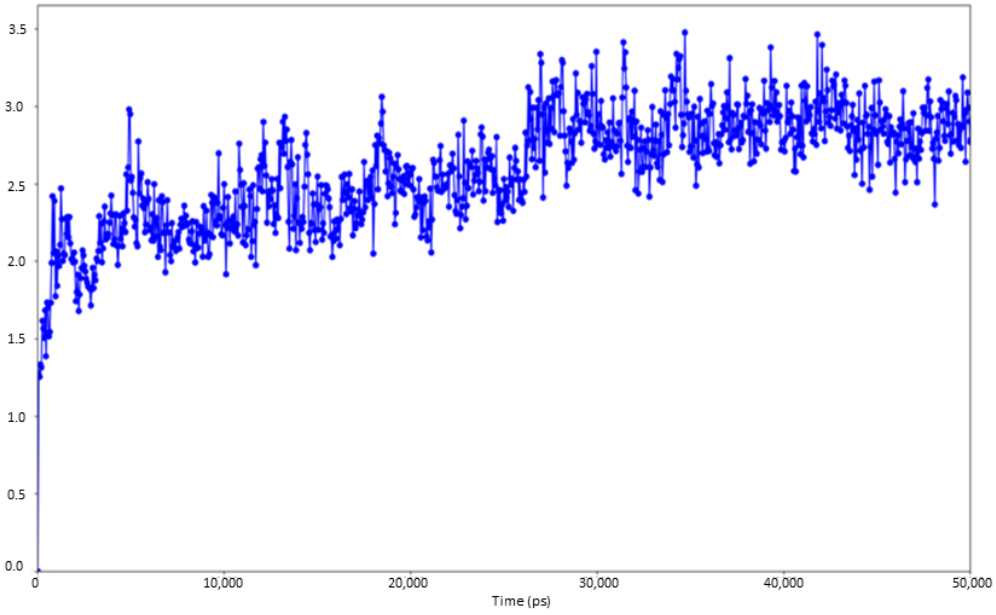


Figure S2. RMSD plots related to the second MD simulation of X-ray structure Cbl-b-CIN85.

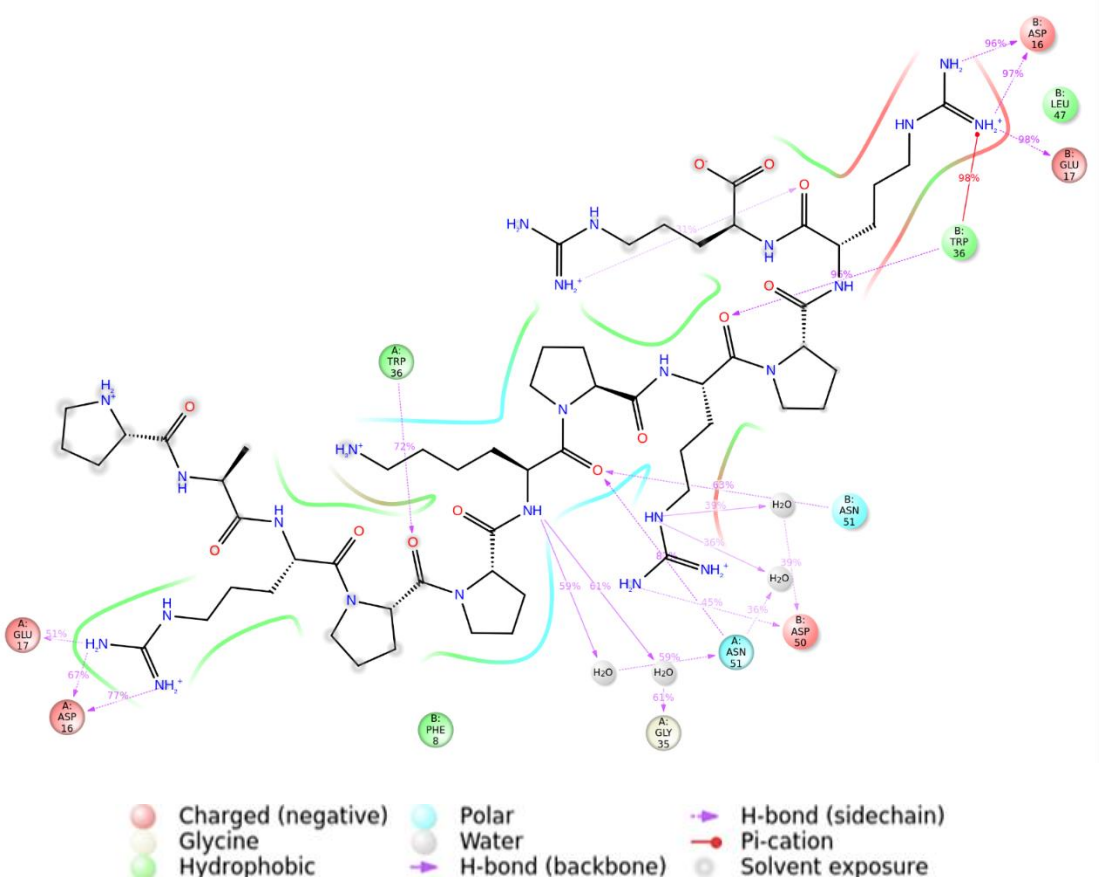
A) RMSD Plot of protein heavy atoms



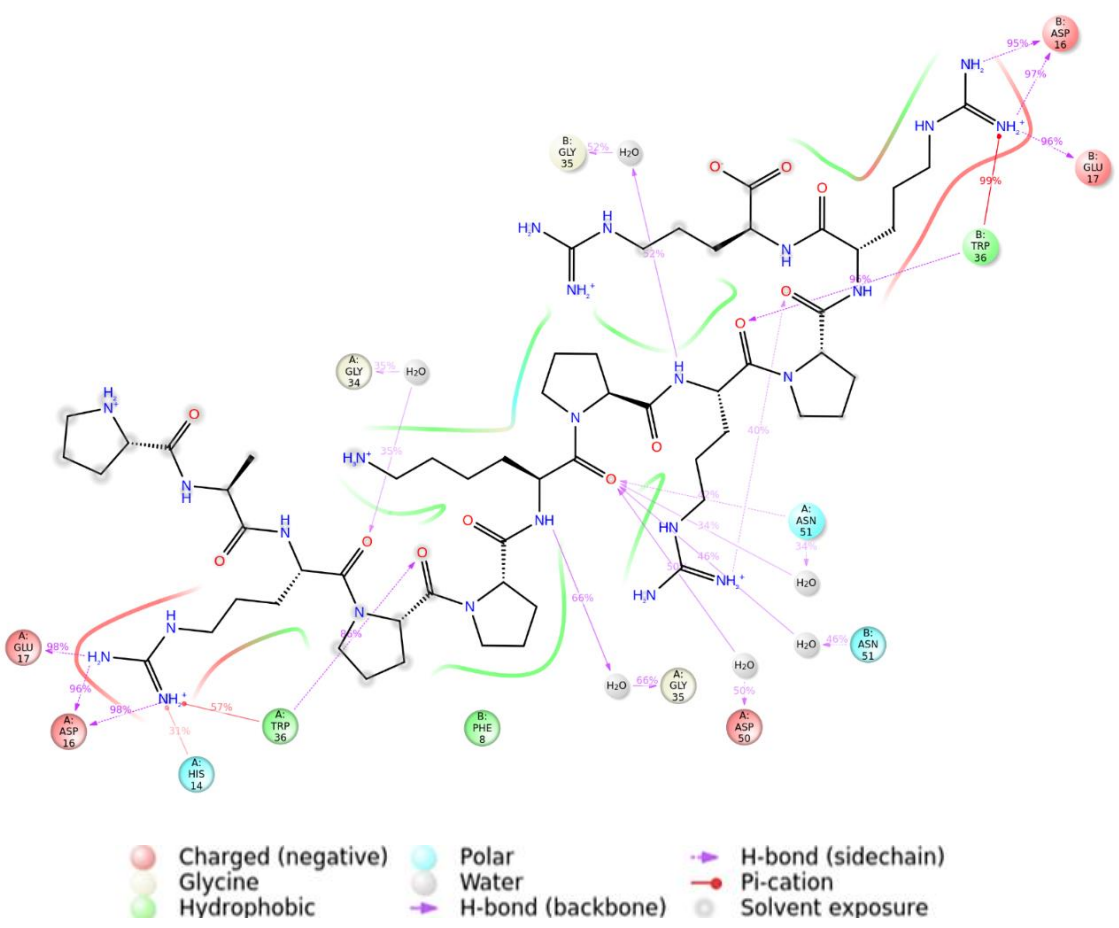
B) RMSD Plot of ligand



FigureS3. 2D depiction of the interactions occurring in the first MD simulation of the X-ray structure Cbl-b-CIN85.

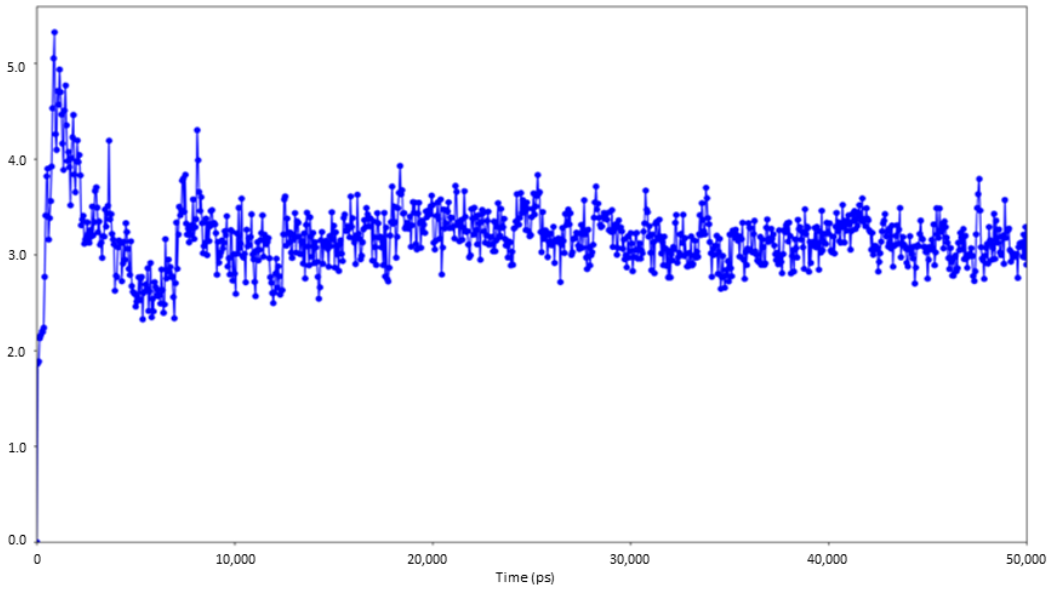


FigureS4. 2D depiction of the interactions occurring in the second MD simulation of the X-ray structure Cbl-b-CIN85.



FigureS5. RMSD plots related to the first MD simulation of MUC1-CIN85 SH3A heterotrimeric complex.

A) RMSD Plot of protein heavy atoms



B) RMSD Plot of ligand

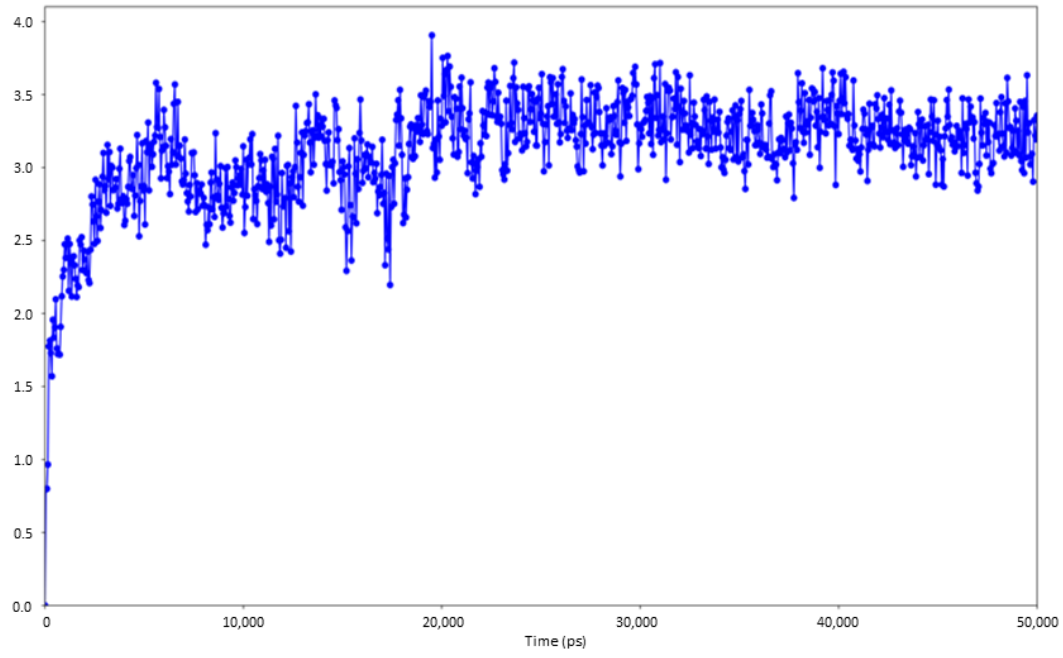
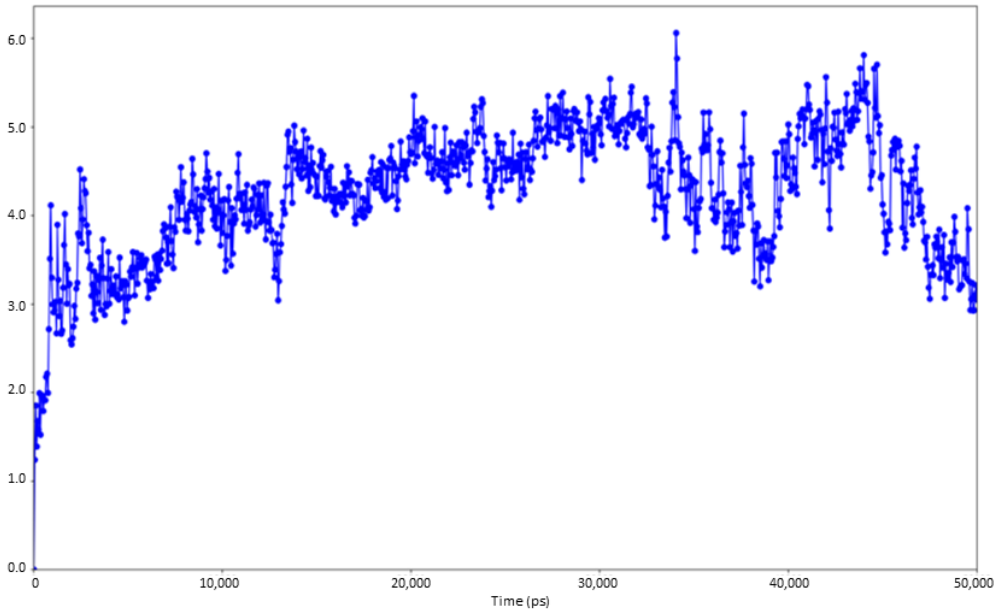
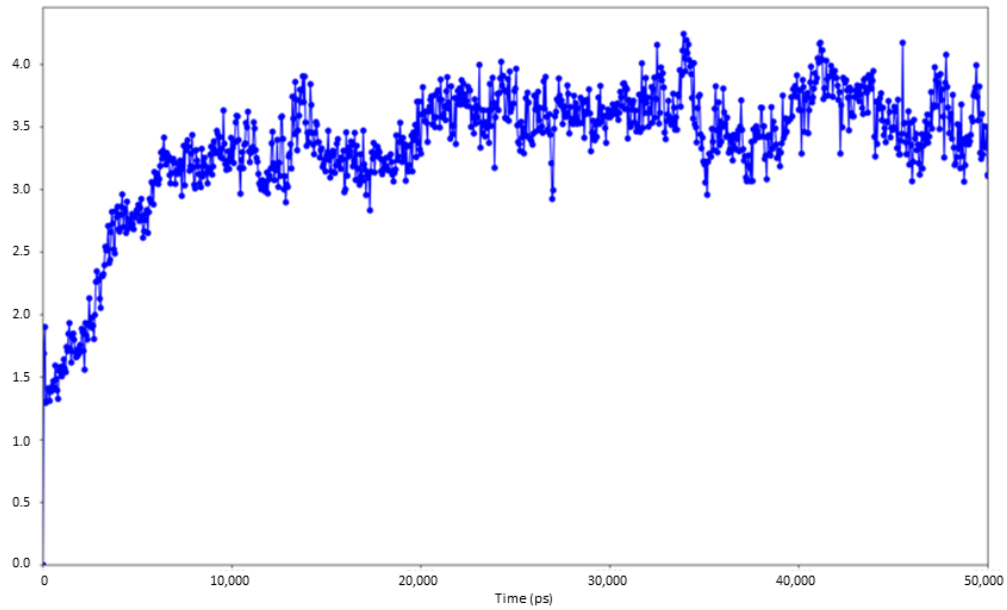


Figure S6. RMSD plots related to the second MD simulation of MUC1-CIN85 SH3A heterotrimeric complex.

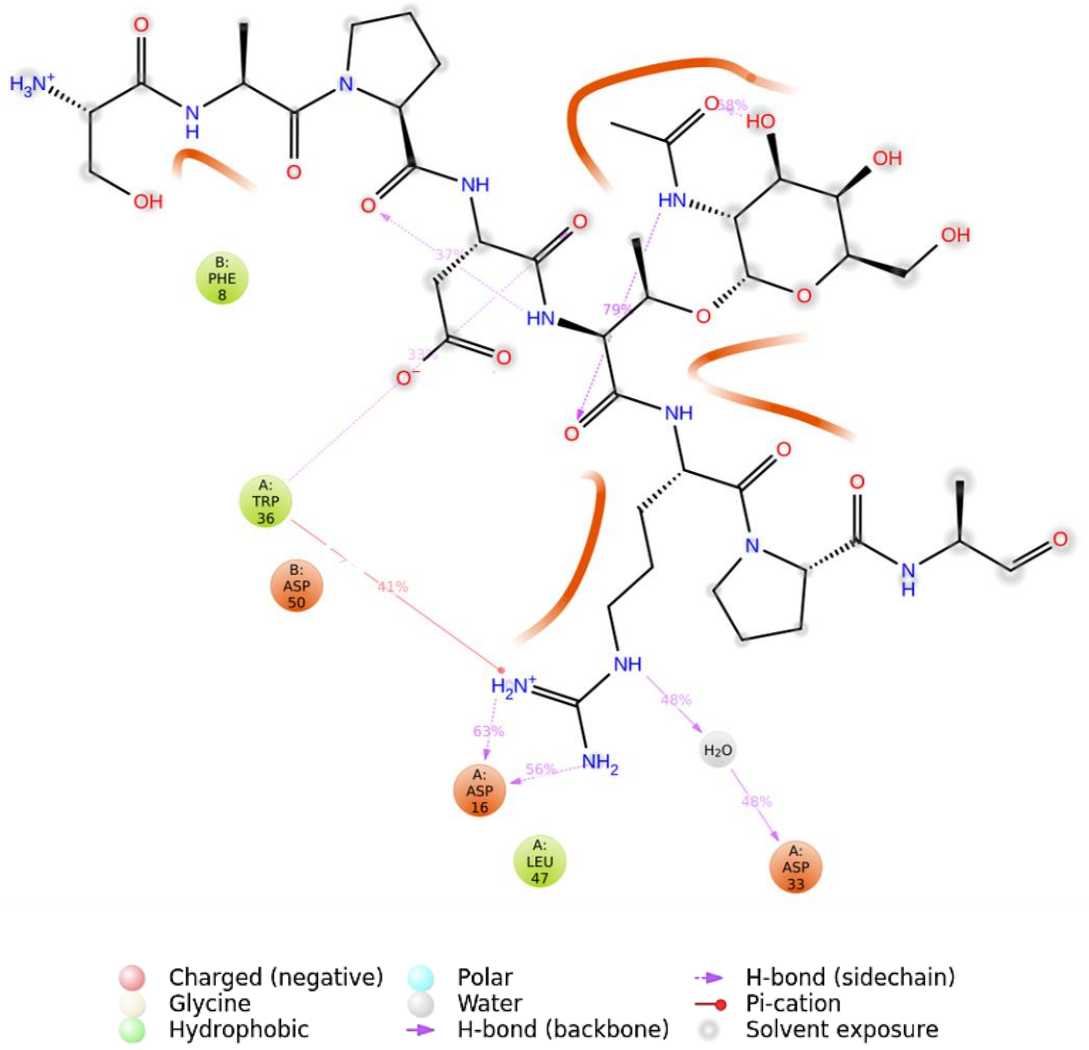
A) RMSD Plot of protein heavy atoms



B) RMSD Plot of ligand



FigureS7. 2D depiction of the interactions occurring in the first MD simulation of MUC1-CIN85 SH3A heterotrimeric complex.



FigureS8. 2D depiction of the interactions occurring in the second MD simulation of MUC1-CIN85 SH3A heterotrimeric complex.

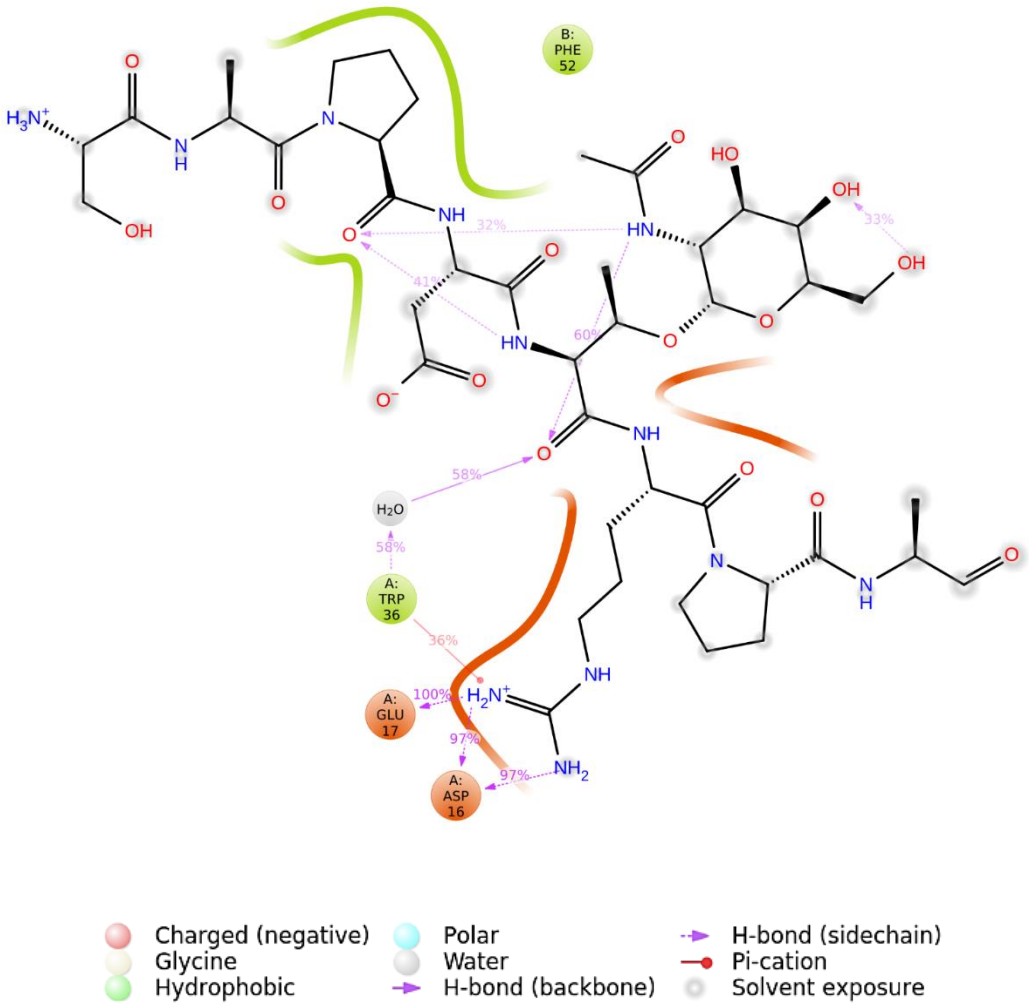
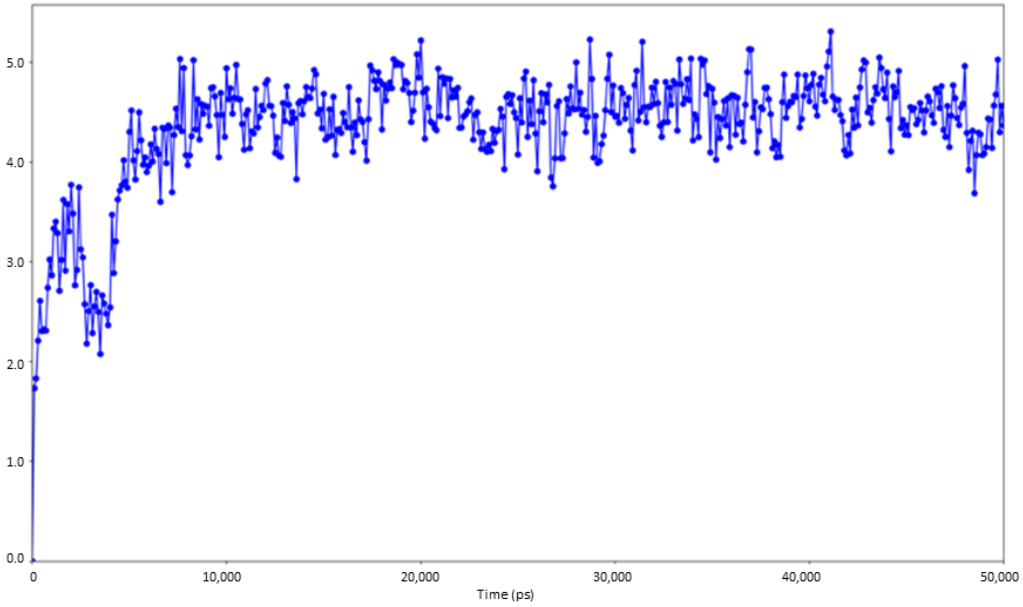


Figure S9. RMSD plots related to the first MD simulation of MUC1-CIN85 SH3A heterodimeric complex.

A) RMSD Plot of protein heavy atoms



B) RMSD Plot of ligand

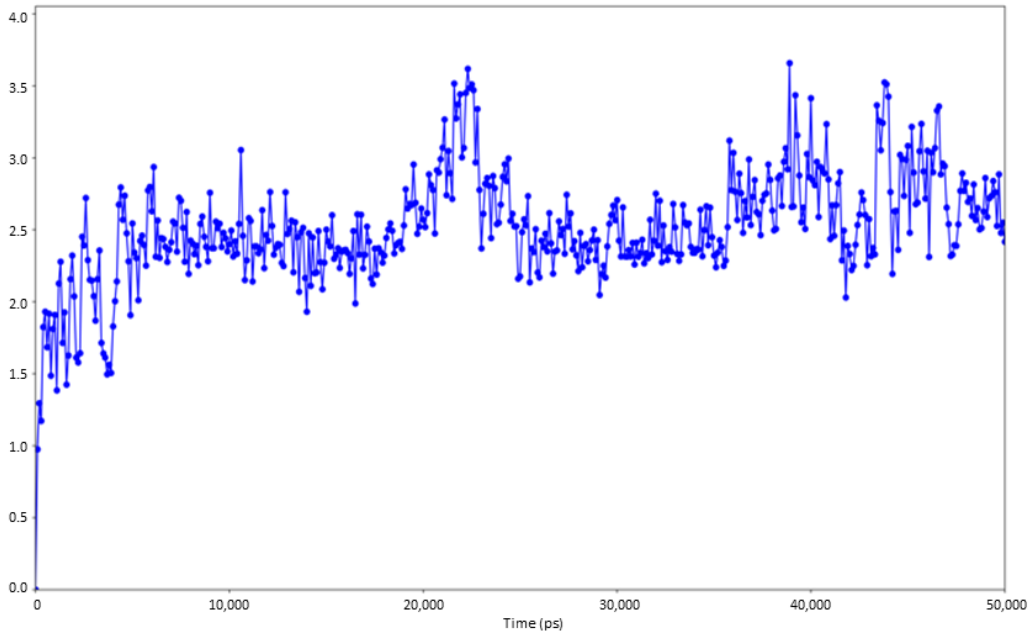
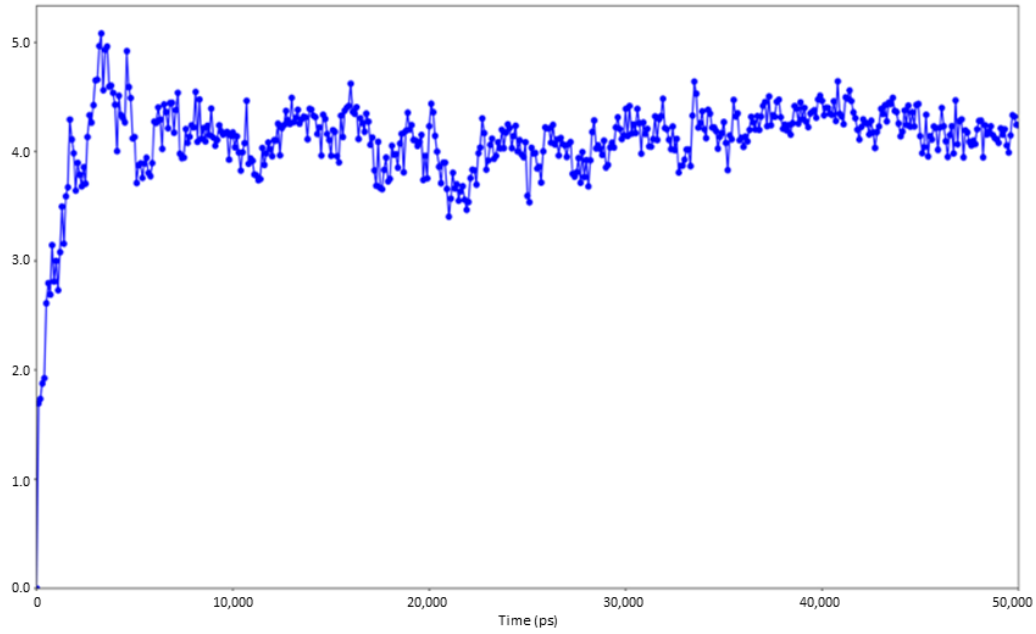
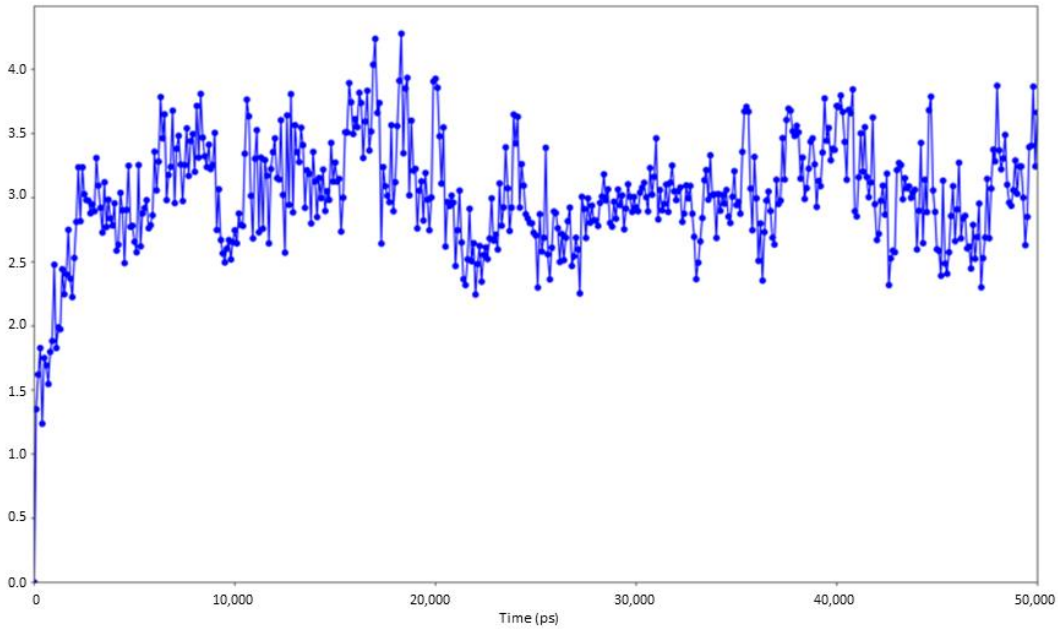


Figure S10. RMSD plots related to the second MD simulation of MUC1-CIN85 SH3A heterodimeric complex.

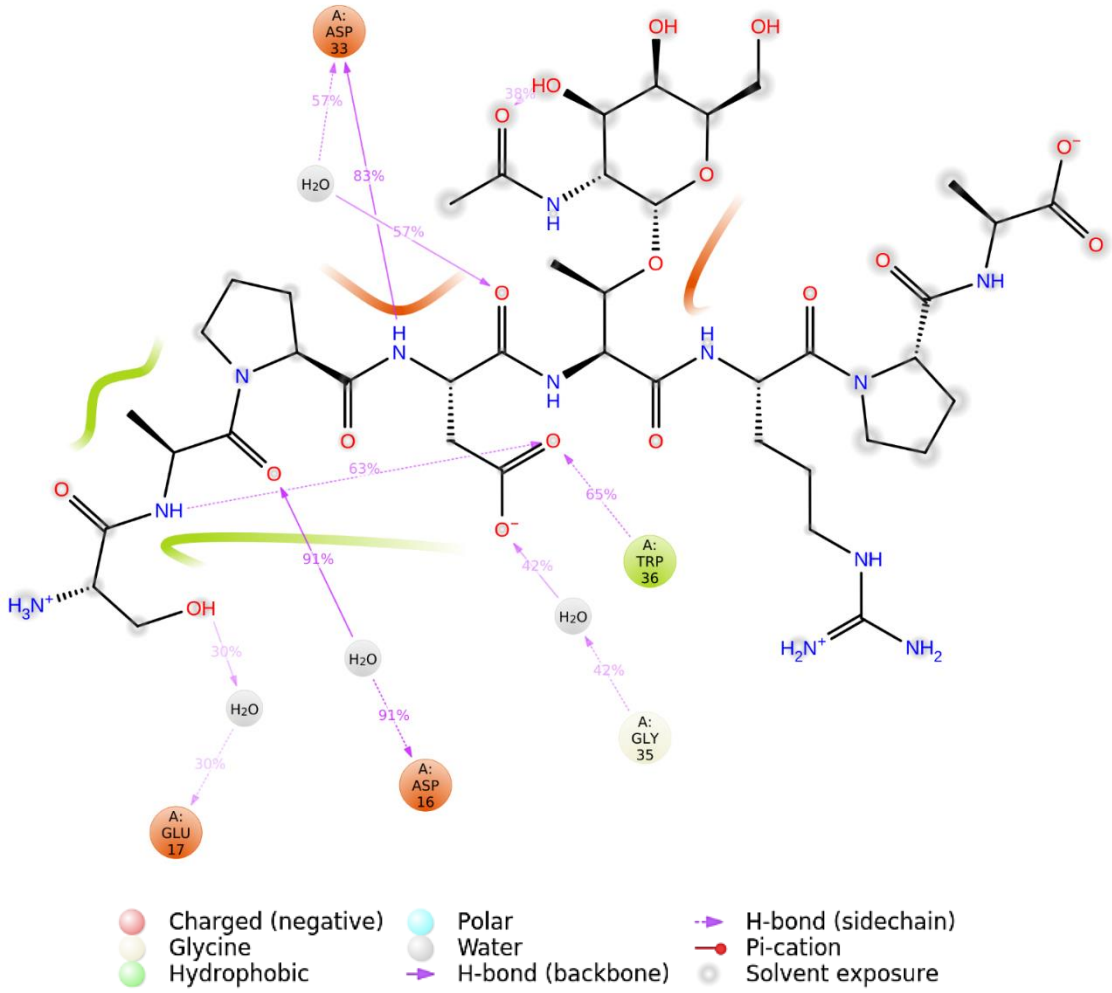
A) RMSD Plot of protein heavy atoms



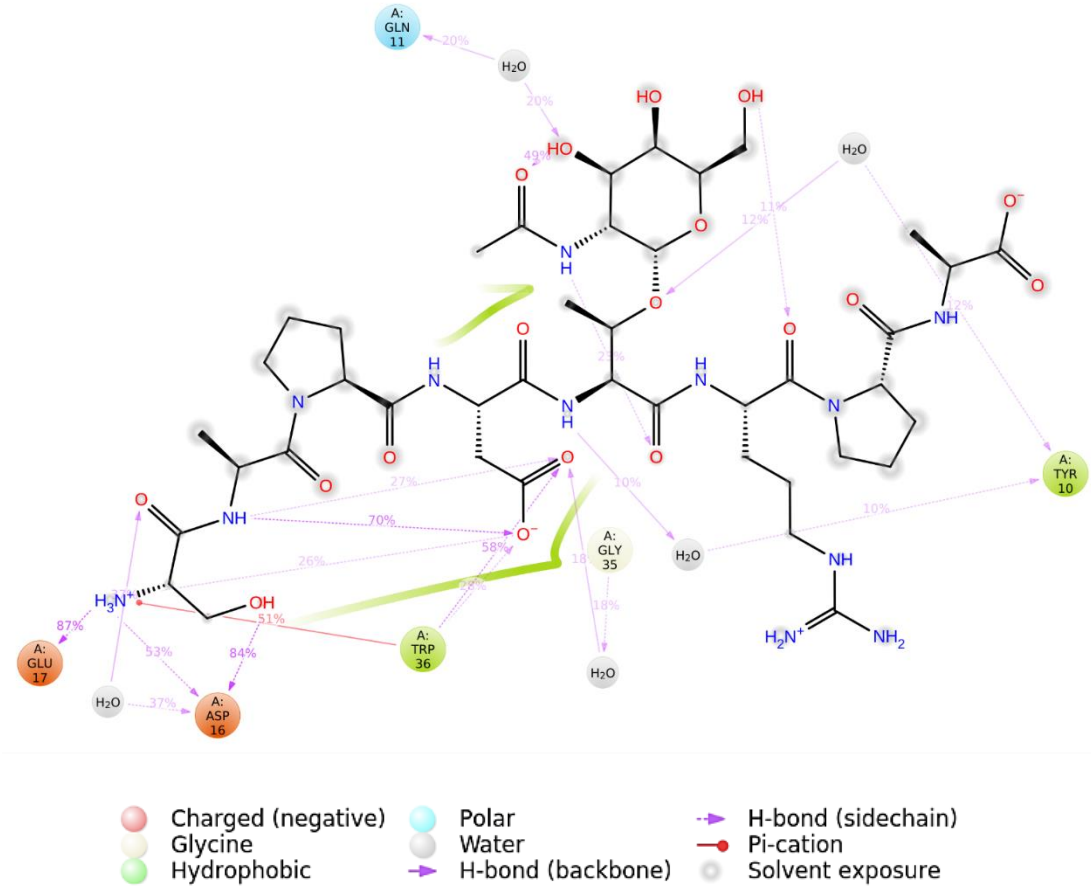
B) RMSD Plot of ligand



FigureS11. 2D depiction of the interactions occurring in the first MD simulation of MUC1-CIN85 SH3A heterodimeric complex.



FigureS12. 2D depiction of the interactions occurring in the second MD simulation of MUC1-CIN85 SH3A heterodimeric complex.



Conclusions

Conclusions

MUC1-CIN85 PPI represents a viable target for the development of new anticancer drugs. However, the absence of the experimental structure of the complex hampers the design of small molecules inhibitors. In this context, computational approaches proved to be useful means to study PPIs allowing to obtain useful structural information.

In details, during my abroad experience at University of Vienna, *in silico* methods were applied to investigate the presence of putative ligand-binding sites on SH3A domain of CIN85. Considering that the opening of binding sites requires structural rearrangements, MD simulation was performed in order to obtain representative protein conformations by clustering procedure. For each representative frame, the presence of druggable binding sites was probed by F-pocket and FTMap. The first one allowed to identify two druggable binding pockets, while the second one predicted seven druggable hotspot regions. Notably, both tools revealed the presence of a druggable binding site on SH3A binding interface. Compounds able to interact with this region might inhibit MUC1-CIN85 PPI through an orthosteric mechanism, while the other predicted binding sites might accommodate allosteric ligands. To further investigate our *in silico* hypothesis, the binding site detected in consensus by both approaches was used to perform docking studies by using the Asinex PPI library. The compounds selected from the docking procedure were purchased and their ability to disrupt MUC1-CIN85 PPI is currently under evaluation.

Instead, during my internship at Ri.MED Foundation, my research activity was focused on the investigation of the molecular contacts between MUC1 and CIN85 SH3 domain. For this purpose, protein-peptide docking was employed to generate the 3D model of the complex both in its heterodimeric and heterotrimeric form. Interestingly, when bound as dimer, CIN85 SH3 domains might interact with MUC1 by establishing a similar interaction network of that found in the x-ray structure of CIN85-Cbl-b complex. To probe the stability of the contacts retrieved in the docking studies, MD simulation was carried out on the obtained complexes. For comparative purpose MD simulation was performed also on the experimental structure of CIN85-Cbl-b. The MD outputs revealed that in the trimeric form the

most stable interactions are in accordance with the x-ray structure, whereas in the heterodimeric complex the molecular contacts found in the docking studies disappear at the beginning of the simulation. Finally, MM-GBSA calculations were executed for each MD trajectory showing that the formation of the heterotrimeric complex should be energetically favoured respect to the heterodimeric one, thus supporting experimental studies for which CIN85 would engage MUC1 in its dimeric form.

Overall, both the studies reported in this section provide useful hints that could be exploited for the structure-based design of MUC1-CIN85 PPI modulators as potential anti-metastatic agents.

References

References

1. Mabonga, L.; Kappo, A. P., Protein-protein interaction modulators: advances, successes and remaining challenges. *Biophys Rev* **2019,** *11* (4), 559-581.

2. Stevers, L. M.; Sijbesma, E.; Botta, M.; MacKintosh, C.; Obsil, T.; Landrieu, I.; Cau, Y.; Wilson, A. J.; Karawajczyk, A.; Eickhoff, J.; Davis, J.; Hann, M.; O'Mahony, G.; Doveston, R. G.; Brunsveld, L.; Ottmann, C., Modulators of 14-3-3 Protein-Protein Interactions. *J Med Chem* **2018,** *61* (9), 3755-3778.

3. Ryan, D. P.; Matthews, J. M., Protein-protein interactions in human disease. *Curr Opin Struct Biol* **2005,** *15* (4), 441-6.

4. De Las Rivas, J.; Fontanillo, C., Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* **2010,** *6* (6), e1000807.

5. Zinzalla, G.; Thurston, D. E., Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Med Chem* **2009,** *1* (1), 65-93.

6. Arkin, M. R.; Tang, Y.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **2014,** *21* (9), 1102-14.

7. Voet, A.; Berenger, F.; Zhang, K. Y., Electrostatic similarities between protein and small molecule ligands facilitate the design of protein-protein interaction inhibitors. *PLoS One* **2013,** *8* (10), e75762.

8. 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.

9. Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** *450* (7172), 1001-9.

10. Fry, D. C., Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers* **2006,** *84* (6), 535-52.

11. Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J.; Zhang, G.; et al., Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J Med Chem* **1992,** *35* (24), 4640-2.

12. Two i.v. antiplatelet agents marketed for coronary disease. *Am J Health Syst Pharm* **1998,** *55* (14), 1440, 1443.

13. Schiff, P. B.; Horwitz, S. B., Taxol stabilizes microtubules in mouse fibroblast cells. *Proc Natl Acad Sci U S A* **1980,** *77* (3), 1561-5.

14. Metz, A.; Ciglia, E.; Gohlke, H., Modulating protein-protein interactions: from structural determinants of binding to druggability prediction to application. *Curr Pharm Des* **2012,** *18* (30), 4630-47.

15. Chene, P., Drugs targeting protein-protein interactions. *ChemMedChem* **2006,** *1* (4), 400-11.

16. Sable, R.; Jois, S., Surfing the Protein-Protein Interaction Surface Using Docking Methods: Application to the Design of PPI Inhibitors. *Molecules* **2015,** *20* (6), 11569-603.

17. Wang, H.; Liu, C.; Deng, L., Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting. *Sci Rep* **2018,** *8* (1), 14285.

18. Keskin, O.; Ma, B.; Nussinov, R., Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* **2005,** *345* (5), 1281-94.

19. Burgoyne, N. J.; Jackson, R. M., Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **2006,** *22* (11), 1335-42.

20. Nooren, I. M.; Thornton, J. M., Diversity of protein-protein interactions. *EMBO J* **2003,** *22* (14), 3486-92.

21. Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R., Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem Rev* **2008,** *108* (4), 1225-44.

22. Villoutreix, B. O.; Kuenemann, M. A.; Poyet, J. L.; Bruzzoni-Giovanelli, H.; Labbe, C.; Lagorce, D.; Sperandio, O.; Miteva, M. A., Drug-Like Protein-Protein Interaction Modulators: Challenges and Opportunities for Drug Discovery and Chemical Biology. *Mol Inform* **2014,** *33* (6-7), 414-437.

23. Fischer, G.; Rossmann, M.; Hyvonen, M., Alternative modulation of protein-protein interactions by small molecules. *Curr Opin Biotechnol* **2015,** *35*, 78-85.

24. Jin, L.; Wang, W.; Fang, G., Targeting protein-protein interaction by small molecules. *Annu Rev Pharmacol Toxicol* **2014,** *54*, 435-56.

25. Garner, T. P.; Lopez, A.; Reyna, D. E.; Spitz, A. Z.; Gavathiotis, E., Progress in targeting the BCL-2 family of proteins. *Curr Opin Chem Biol* **2017,** *39*, 133-142.

26. Birkinshaw, R. W.; Gong, J. N.; Luo, C. S.; Lio, D.; White, C. A.; Anderson, M. A.; Blombery, P.; Lessene, G.; Majewski, I. J.; Thijssen, R.; Roberts, A. W.; Huang, D. C. S.; Colman, P. M.; Czabotar, P. E., Structures of BCL-2 in complex with venetoclax reveal the molecular basis of resistance mutations. *Nat Commun* **2019,** *10* (1), 2385.

27. Ni, D.; Lu, S.; Zhang, J., Emerging roles of allosteric modulators in the regulation of protein-protein interactions (PPIs): A new paradigm for PPI drug discovery. *Med Res Rev* **2019,** *39* (6), 2314-2342.

28. Gonzalez-Ruiz, D.; Gohlke, H., Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem* **2006,** *13* (22), 2607-25.

29. Silvian, L. F.; Friedman, J. E.; Strauch, K.; Cachero, T. G.; Day, E. S.; Qian, F.; Cunningham, B.; Fung, A.; Sun, L.; Shipps, G. W.; Su, L.; Zheng, Z.; Kumaravel, G.; Whitty, A., Small molecule inhibition of the TNF family cytokine CD40 ligand through a subunit fracture mechanism. *ACS Chem Biol* **2011,** *6* (6), 636-47.

30. Bakail, M., Ochsenbein, F., Targeting protein-protein interactions, a wide open field for drug design. *C. R. Chimie* **2016,** *19*, 19-27.

31. Lau, J. L.; Dunn, M. K., Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem* **2018,** *26* (10), 2700-2707.

32. Hill, T. A.; Shepherd, N. E.; Diness, F.; Fairlie, D. P., Constraining cyclic peptides to mimic protein structure motifs. *Angew Chem Int Ed Engl* **2014,** *53* (48), 13020-41.

33. Cussol, L.; Mauran-Ambrosino, L.; Buratto, J.; Belorusova, A. Y.; Neuville, M.; Osz, J.; Fribourg, S.; Fremaux, J.; Dolain, C.; Goudreau, S. R.; Rochel, N.; Guichard, G., Structural Basis for alpha-Helix Mimicry and Inhibition of Protein-Protein Interactions with Oligourea Foldamers. *Angew Chem Int Ed Engl* **2021,** *60* (5), 2296-2303.

34. Kannan, S.; Aronica, P. G. A.; Ng, S.; Gek Lian, D. T.; Frosi, Y.; Chee, S.; Shimin, J.; Yuen, T. Y.; Sadruddin, A.; Kaan, H. Y. K.; Chandramohan, A.; Wong, J. H.; Tan, Y. S.; Chang, Z. W.; Ferrer-Gago, F. J.; Arumugam, P.; Han, Y.; Chen, S.; Renia, L.; Brown, C. J.; Johannes, C. W.; Henry, B.; Lane, D. P.; Sawyer, T. K.; Verma, C. S.; Partridge, A. W., Macrocyclization of an all-d linear alpha-helical peptide imparts cellular permeability. *Chem Sci* **2020,** *11* (21), 5577-5591.

35. Higueruelo, A. P.; Jubb, H.; Blundell, T. L., Protein-protein interactions as druggable targets: recent technological advances. *Curr Opin Pharmacol* **2013,** *13* (5), 791-6.

36. Walensky, L. D.; Bird, G. H., Hydrocarbon-stapled peptides: principles, practice, and progress. *J Med Chem* **2014,** *57* (15), 6275-88.

37. Tisato, V.; Voltan, R.; Gonelli, A.; Secchiero, P.; Zauli, G., MDM2/X inhibitors under clinical evaluation: perspectives for the management of hematological malignancies and pediatric cancer. *J Hematol Oncol* **2017,** *10* (1), 133.

38. Eyrisch, S.; Helms, V., Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem* **2007,** *50* (15), 3457-64.

39. Li, B.; Rong, D.; Wang, Y., Targeting Protein-Protein Interaction with Covalent Small-Molecule Inhibitors. *Curr Top Med Chem* **2019,** *19* (21), 1872-1876.

40. Shin, W. H.; Christoffer, C. W.; Kihara, D., In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods* **2017,** *131*, 22-32.

41. Milroy, L. G.; Grossmann, T. N.; Hennig, S.; Brunsveld, L.; Ottmann, C., Modulators of protein-protein interactions. *Chem Rev* **2014,** *114* (9), 4695-748.

42. Qiu, Y.; Li, X.; He, X.; Pu, J.; Zhang, J.; Lu, S., Computational methods-guided design of modulators targeting protein-protein interactions (PPIs). *Eur J Med Chem* **2020,** *207*, 112764.

43. Cossar, P. J.; Lewis, P. J.; McCluskey, A., Protein-protein interactions as antibiotic targets: A medicinal chemistry perspective. *Med Res Rev* **2020,** *40* (2), 469-494.

44. Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z., Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* **2014,** *14* (16), 1923-38.

45. Wu, K. J.; Lei, P. M.; Liu, H.; Wu, C.; Leung, C. H.; Ma, D. L., Mimicking Strategy for Protein-Protein Interaction Inhibitor Discovery by Virtual Screening. *Molecules* **2019,** *24* (24).

46. Aminpour, M.; Montemagno, C.; Tuszynski, J. A., An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules* **2019,** *24* (9).

47. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol* **2011,** *9*, 71.

48. Zhong, S.; Macias, A. T.; MacKerell, A. D., Jr., Computational identification of inhibitors of protein-protein interactions. *Curr Top Med Chem* **2007,** *7* (1), 63-82.

49. Feng, T.; Barakat, K., Molecular Dynamics Simulation and Prediction of Druggable Binding Sites. *Methods Mol Biol* **2018,** *1762*, 87-103.

50. Macari, G.; Toti, D.; Polticelli, F., Computational methods and tools for binding site recognition between proteins and small molecules: from classical geometrical approaches to modern machine learning strategies. *J Comput Aided Mol Des* **2019,** *33* (10), 887-903.

51. Roche, D. B.; Brackenridge, D. A.; McGuffin, L. J., Proteins and Their Interacting Partners: An Introduction to Protein-Ligand Binding Site Prediction Methods. *Int J Mol Sci* **2015,** *16* (12), 29829-42.

52. Dukka, B. K., Structure-based Methods for Computational Protein Functional Site Prediction. *Comput Struct Biotechnol J* **2013,** *8*, e201308005.

53. Batool, M.; Ahmad, B.; Choi, S., A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019,** *20* (11).

54. Lin, X.; Li, X.; Lin, X., A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules* **2020,** *25* (6).

55. Salmaso, V.; Moro, S., Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front Pharmacol* **2018,** *9*, 923.

56. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J., Protein-ligand docking: current status and future challenges. *Proteins* **2006,** *65* (1), 15-26.

57. Macalino, S. J. Y.; Basith, S.; Clavio, N. A. B.; Chang, H.; Kang, S.; Choi, S., Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery. *Molecules* **2018,** *23* (8).

58. Keskin, O.; Tuncbag, N.; Gursoy, A., Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem Rev* **2016,** *116* (8), 4884-909.

59. Soni, N.; Madhusudhan, M. S., Computational modeling of protein assemblies. *Curr Opin Struct Biol* **2017,** *44*, 179-189.

60. Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J.; Critical Assessment of, P. I., CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **2003,** *52* (1), 2-9.

61. Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G.; Strategies for 3D pharmacophore-based virtual screening. *Drug Discov Today Technol* **2010,** *7* (4), e203-70.

62. Wermuth, C. G.; Ganellin, C.R.; Lindberg, P.; Mitscher, L.A.; Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl Chem* **1998,** *70*, 1129-1143.

63. Seidel, T.; Wieder, O.; Garon, A.; Langer, T., Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. *Mol Inform* **2020,** *39* (11), e2000059.

64. Schaller, D., Next generation 3D pharmacophore modeling. *WIREs Comput Mol Sci* **2020**, 10:e1468.

65. Kaserer, T.; Beck, K. R.; Akram, M.; Odermatt, A.; Schuster, D., Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules* **2015,** *20* (12), 22799-832.

66. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* **2012,** *55* (14), 6582-94.

67. Vittorio, S.; Seidel, T.; Germano, M. P.; Gitto, R.; Ielo, L.; Garon, A.; Rapisarda, A.; Pace, V.; Langer, T.; De Luca, L., A Combination of Pharmacophore and Docking-based Virtual Screening to Discover new Tyrosinase Inhibitors. *Mol Inform* **2020,** *39* (3), e1900054.

68. Wieder, M.; Garon, A.; Perricone, U.; Boresch, S.; Seidel, T.; Almerico, A. M.; Langer, T., Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *J Chem Inf Model* **2017,** *57* (2), 365-385.

69. Perricone, U.; Wieder, M.; Seidel, T.; Langer, T.; Padova, A.; Almerico, A. M.; Tutone, M., A Molecular Dynamics-Shared Pharmacophore Approach to Boost Early-Enrichment Virtual Screening: A Case Study on Peroxisome Proliferator-Activated Receptor alpha. *ChemMedChem* **2017,** *12* (16), 1399-1407.

70. Bock, A.; Bermudez, M.; Krebs, F.; Matera, C.; Chirinda, B.; Sydow, D.; Dallanoce, C.; Holzgrabe, U.; De Amici, M.; Lohse, M. J.; Wolber, G.; Mohr, K., Ligand Binding Ensembles Determine Graded Agonist Efficacies at a G Protein-coupled Receptor. *J Biol Chem* **2016,** *291* (31), 16375-89.

71. Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C.; Glick, M.; Davies, J. W., How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *J Chem Inf Model* **2009,** *49* (1), 108-19.

72. Stumpfe, D., Bajorath, J., Similarity searching. *WIREs Comput Mol Sci* **2011,** *1*, 260-282.

73. Kumar, A.; Zhang, K. Y. J., Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Front Chem* **2018,** *6*, 315.

74. Kristensen, T. G.; Nielsen, J.; Pedersen, C. N., Methods for Similarity-based Virtual Screening. *Comput Struct Biotechnol J* **2013,** *5*, e201302009.

75. Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J., Molecular similarity in medicinal chemistry. *J Med Chem* **2014,** *57* (8), 3186-204.

76. Shin, W. H.; Zhu, X.; Bures, M. G.; Kihara, D., Three-dimensional compound comparison methods and their application in drug discovery. *Molecules* **2015,** *20* (7), 12841-62.

77. Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N., Virtual screening in drug discovery -- a computational perspective. *Curr Protein Pept Sci* **2007,** *8* (4), 329-51.

78. Gimeno, A.; Ojeda-Montes, M. J.; Tomas-Hernandez, S.; Cereto-Massague, A.; Beltran-Debon, R.; Mulero, M.; Pujadas, G.; Garcia-Vallve, S., The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int J Mol Sci* **2019,** *20* (6).

79. Finn, P. W., Morris, G.M., Shape-based similarity searching in chemical databases. *WIREs Comput Mol Sci* **2013,** *3*, 226-241.

80. Bali, A., Molecular-field-based three-dimensional similarity studies on quinoline-based CNS active agents. *ISRN Pharm* **2011,** *2011*, 186943.

81. 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* (6), 920-936.

82. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A., Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem* **2016,** *59* (9), 4035-61.

83. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Adv Protein Chem* **2003,** *66*, 27-85.

84. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **2020,** *16* (1), 528-552.

85. Jorgensen, W. L., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996,** *118*, 11225-11236.

86. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* **2016,** *12* (1), 281-96.

87. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** *102* (18), 3586-616.

88. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017,** *14* (1), 71-73.

89. Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* **2004,** *25* (13), 1656-76.

90. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **2011,** *40* (7), 843-56.

91. Meller, J.; Molecular Dynamics. In *eLS* **2001**.

92. Alonso, H.; Bliznyuk, A. A.; Gready, J. E., Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* **2006,** *26* (5), 531-68.

93. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.

94. Lazim, R.; Suh, D.; Choi, S., Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int J Mol Sci* **2020,** *21* (17).

95. Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-level characterization of protein-protein association. *Proc Natl Acad Sci U S A* **2019,** *116* (10), 4244-4249.

96. Paul, A., Yadav, K.S., Parkinson's disease: Current drug therapy and unraveling the prospects of nanoparticles. *J Drug Deliv Sci Tec* **2020,** *58*, 101790.

97. Balestrino, R.; Schapira, A. H. V., Parkinson disease. *Eur J Neurol* **2020,** *27* (1), 27-42.

98. Kalia, L. V.; Lang, A. E., Parkinson's disease. *Lancet* **2015,** *386* (9996), 896-912.
99. Dickson, D. W., Parkinson's disease and parkinsonism: neuropathology. *Cold Spring Harb Perspect Med* **2012,** *2* (8).

100. Jankovic, J.; Tan, E. K., Parkinson's disease: etiopathogenesis and treatment. *J Neurol Neurosurg Psychiatry* **2020,** *91* (8), 795-808.

101. Chakraborty, A.; Brauer, S.; Diwan, A., Possible therapies of Parkinson's disease: A review. *J Clin Neurosci* **2020,** *75*, 1-4.

102. Carrera, I.; Cacabelos, R., Current Drugs and Potential Future Neuroprotective Compounds for Parkinson's Disease. *Curr Neuropharmacol* **2019,** *17* (3), 295-306.

103. Borovac, J. A., Side effects of a dopamine agonist therapy for Parkinson's disease: a mini-review of clinical pharmacology. *Yale J Biol Med* **2016,** *89* (1), 37-47.

104. Shihabuddin, L. S.; Brundin, P.; Greenamyre, J. T.; Stephenson, D.; Sardi, S. P., New Frontiers in Parkinson's Disease: From Genetics to the Clinic. *J Neurosci* **2018,** *38* (44), 9375-9382.

105. Frey, B.; AlOkda, A.; Jackson, M. P.; Riguet, N.; Duce, J. A.; Lashuel, H. A., Monitoring alpha-synuclein oligomerization and aggregation using bimolecular fluorescence complementation assays: What you see is not always what you get. *J Neurochem* **2020**.

106. Oliveri, V., Toward the discovery and development of effective modulators of alpha-synuclein amyloid aggregation. *Eur J Med Chem* **2019,** *167*, 10-36.

107. Rezaeian, N.; Shirvanizadeh, N.; Mohammadi, S.; Nikkhah, M.; Arab, S. S., The inhibitory effects of biomimetically designed peptides on alpha-synuclein aggregation. *Arch Biochem Biophys* **2017,** *634*, 96-106.

108. Recchia, A.; Debetto, P.; Negro, A.; Guidolin, D.; Skaper, S. D.; Giusti, P., Alpha-synuclein and Parkinson's disease. *FASEB J* **2004,** *18* (6), 617-26.

109. Fouka, M.; Mavroeidi, P.; Tsaka, G.; Xilouri, M., In Search of Effective Treatments Targeting alpha-Synuclein Toxicity in Synucleinopathies: Pros and Cons. *Front Cell Dev Biol* **2020,** *8*, 559791.

110. Sulzer, D.; Edwards, R. H., The physiological role of alpha-synuclein and its relationship to Parkinson's Disease. *J Neurochem* **2019,** *150* (5), 475-486.

111. Serpell, L. C.; Berriman, J.; Jakes, R.; Goedert, M.; Crowther, R. A., Fiber diffraction of synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation. *Proc Natl Acad Sci U S A* **2000,** *97* (9), 4897-902.

112. Pena-Diaz, S.; Pujols, J.; Ventura, S., Small molecules to prevent the neurodegeneration caused by alpha-synuclein aggregation. *Neural Regen Res* **2020,** *15* (12), 2260-2261.

113. Sangwan, S.; Sahay, S.; Murray, K. A.; Morgan, S.; Guenther, E. L.; Jiang, L.; Williams, C. K.; Vinters, H. V.; Goedert, M.; Eisenberg, D. S., Inhibition of synucleinopathic seeding by rationally designed inhibitors. *Elife* **2020,** *9*.

114. Luk, K. C.; Kehm, V. M.; Zhang, B.; O'Brien, P.; Trojanowski, J. Q.; Lee, V. M., Intracerebral inoculation of pathological alpha-synuclein initiates a rapidly progressive neurodegenerative alpha-synucleinopathy in mice. *J Exp Med* **2012,** *209* (5), 975-86.

115. Luk, K. C.; Kehm, V.; Carroll, J.; Zhang, B.; O'Brien, P.; Trojanowski, J. Q.; Lee, V. M., Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **2012,** *338* (6109), 949-53.

116. Masuda-Suzukake, M.; Nonaka, T.; Hosokawa, M.; Oikawa, T.; Arai, T.; Akiyama, H.; Mann, D. M.; Hasegawa, M., Prion-like spreading of pathological alpha-synuclein in brain. *Brain* **2013,** *136* (Pt 4), 1128-38.

117. Longhena, F.; Faustini, G.; Brembati, V.; Pizzi, M.; Bellucci, A., The good and bad of therapeutic strategies that directly target alpha-synuclein. *IUBMB Life* **2020,** *72* (4), 590-600.

118. Brundin, P.; Dave, K. D.; Kordower, J. H., Therapeutic approaches to target alpha-synuclein pathology. *Exp Neurol* **2017,** *298* (Pt B), 225-235.

119. Wong, Y. C.; Krainc, D., alpha-synuclein toxicity in neurodegeneration: mechanism and therapeutic strategies. *Nat Med* **2017,** *23* (2), 1-13.

120. Mittal, S.; Bjornevik, K.; Im, D. S.; Flierl, A.; Dong, X.; Locascio, J. J.; Abo, K. M.; Long, E.; Jin, M.; Xu, B.; Xiang, Y. K.; Rochet, J. C.; Engeland, A.; Rizzu, P.; Heutink, P.; Bartels, T.; Selkoe, D. J.; Caldarone, B. J.; Glicksman, M. A.; Khurana, V.; Schule, B.; Park, D. S.; Riise, T.; Scherzer, C. R., beta2-Adrenoreceptor is a regulator of the alpha-synuclein gene driving risk of Parkinson's disease. *Science* **2017,** *357* (6354), 891-898.

121. Cox, D.; Selig, E.; Griffin, M. D.; Carver, J. A.; Ecroyd, H., Small Heat-shock Proteins Prevent alpha-Synuclein Aggregation via Transient Interactions and Their Efficacy Is Affected by the Rate of Aggregation. *J Biol Chem* **2016,** *291* (43), 22618-22629.

122. El-Agnaf, O. M.; Paleologou, K. E.; Greer, B.; Abogrein, A. M.; King, J. E.; Salem, S. A.; Fullwood, N. J.; Benson, F. E.; Hewitt, R.; Ford, K. J.; Martin, F. L.; Harriott, P.; Cookson, M. R.; Allsop, D., A strategy for designing inhibitors of alpha-synuclein aggregation and toxicity as a novel treatment for Parkinson's disease and related disorders. *FASEB J* **2004,** *18* (11), 1315-7.

123. Kim, Y. S.; Lim, D.; Kim, J. Y.; Kang, S. J.; Kim, Y. H.; Im, H., beta-Sheet-breaking peptides inhibit the fibrillation of human alpha-synuclein. *Biochem Biophys Res Commun* **2009,** *387* (4), 682-7.

124. Chemerovski-Glikman, M.; Rozentur-Shkop, E.; Richman, M.; Grupi, A.; Getler, A.; Cohen, H. Y.; Shaked, H.; Wallin, C.; Warmlander, S. K.; Haas, E.; Graslund, A.; Chill, J. H.; Rahimipour, S., Self-Assembled Cyclic d,l-alpha-Peptides as Generic Conformational Inhibitors of the alpha-Synuclein Aggregation and Toxicity: In Vitro and Mechanistic Studies. *Chemistry* **2016,** *22* (40), 14236-46.

125. Huggins, K. N.; Bisaglia, M.; Bubacco, L.; Tatarek-Nossol, M.; Kapurniotu, A.; Andersen, N. H., Designed hairpin peptides interfere with amyloidogenesis pathways: fibril formation and cytotoxicity inhibition, interception of the preamyloid state. *Biochemistry* **2011,** *50* (38), 8202-12.

126. Hornedo-Ortega, R.; Alvarez-Fernandez, M. A.; Cerezo, A. B.; Richard, T.; Troncoso, A. M. A.; Garcia-Parrilla, M. A. C., Protocatechuic Acid: Inhibition of Fibril Formation, Destabilization of Preformed Fibrils of Amyloid-beta and alpha-Synuclein, and Neuroprotection. *J Agric Food Chem* **2016,** *64* (41), 7722-7732.

127. Ardah, M. T.; Paleologou, K. E.; Lv, G.; Abul Khair, S. B.; Kazim, A. S.; Minhas, S. T.; Al-Tel, T. H.; Al-Hayani, A. A.; Haque, M. E.; Eliezer, D.; El-Agnaf, O. M., Structure activity relationship of phenolic acid inhibitors of alpha-synuclein fibril formation and toxicity. *Front Aging Neurosci* **2014,** *6*, 197.

128. Di Giovanni, S.; Eleuteri, S.; Paleologou, K. E.; Yin, G.; Zweckstetter, M.; Carrupt, P. A.; Lashuel, H. A., Entacapone and tolcapone, two catechol O-methyltransferase inhibitors, block fibril formation of alpha-synuclein and beta-amyloid and protect against amyloid-induced toxicity. *J Biol Chem* **2010,** *285* (20), 14941-54.

129. Hornedo-Ortega, R.; Cerezo, A. B.; Troncoso, A. M.; Garcia-Parrilla, M. C., Protective effects of hydroxytyrosol against alpha-synuclein toxicity on PC12cells and fibril formation. *Food Chem Toxicol* **2018,** *120*, 41-49.

130. Meng, X.; Munishkina, L. A.; Fink, A. L.; Uversky, V. N., Effects of Various Flavonoids on the alpha-Synuclein Fibrillation Process. *Parkinsons Dis* **2010,** *2010*, 650794.

131. Caruana, M.; Hogen, T.; Levin, J.; Hillmer, A.; Giese, A.; Vassallo, N., Inhibition and disaggregation of alpha-synuclein oligomers by natural polyphenolic compounds. *FEBS Lett* **2011,** *585* (8), 1113-20.

132. Das, S.; Pukala, T. L.; Smid, S. D., Exploring the Structural Diversity in Inhibitors of alpha-Synuclein Amyloidogenic Folding, Aggregation, and Neurotoxicity. *Front Chem* **2018,** *6*, 181.

133. Hong, D. P.; Fink, A. L.; Uversky, V. N., Structural characteristics of alpha-synuclein oligomers stabilized by the flavonoid baicalein. *J Mol Biol* **2008,** *383* (1), 214-23.

134. Meng, X.; Munishkina, L. A.; Fink, A. L.; Uversky, V. N., Molecular mechanisms underlying the flavonoid-induced inhibition of alpha-synuclein fibrillation. *Biochemistry* **2009,** *48* (34), 8206-24.

135. Zhu, M.; Rajamani, S.; Kaylor, J.; Han, S.; Zhou, F.; Fink, A. L., The flavonoid baicalein inhibits fibrillation of alpha-synuclein and disaggregates existing fibrils. *J Biol Chem* **2004,** *279* (26), 26846-57.

136. Zhang, L.; Li, C.; Lin, G.; Krajcsi, P.; Zuo, Z., Hepatic metabolism and disposition of baicalein via the coupling of conjugation enzymes and transporters-in vitro and in vivo evidences. *AAPS J* **2011,** *13* (3), 378-89.

137. Zhang, L.; Lin, G.; Kovacs, B.; Jani, M.; Krajcsi, P.; Zuo, Z., Mechanistic study on the intestinal absorption and disposition of baicalein. *Eur J Pharm Sci* **2007,** *31* (3-4), 221-31.

138. Zhu, M.; Han, S.; Fink, A. L., Oxidized quercetin inhibits alpha-synuclein fibrillization. *Biochim Biophys Acta* **2013,** *1830* (4), 2872-81.

139. Faria, A.; Pestana, D.; Teixeira, D.; Azevedo, J.; De Freitas, V.; Mateus, N.; Calhau, C., Flavonoid transport across RBE4 cells: A blood-brain barrier model. *Cell Mol Biol Lett* **2010,** *15* (2), 234-41.

140. Zhao, J.; Liang, Q.; Sun, Q; Chen, C.; Xu, L.; Zhou, P.; (-) Epigallocatechin-3-gallate (EGCG) inhibits fibrillation, disaggregates amyloid fibrils of a-synuclein, and protects PC12 cells against a-synuclein-induced toxicity. *RSC Adv* **2017,** *7* (52), 32508-32517.

141. Konijnenberg, A.; Ranica, S.; Narkiewicz, J.; Legname, G.; Grandori, R.; Sobott, F.; Natalello, A., Opposite Structural Effects of Epigallocatechin-3-gallate and Dopamine Binding to alpha-Synuclein. *Anal Chem* **2016,** *88* (17), 8468-75.

142. Lorenzen, N.; Nielsen, S. B.; Yoshimura, Y.; Vad, B. S.; Andersen, C. B.; Betzer, C.; Kaspersen, J. D.; Christiansen, G.; Pedersen, J. S.; Jensen, P. H.; Mulder, F. A.; Otzen, D. E., How epigallocatechin gallate can inhibit alpha-synuclein oligomer toxicity in vitro. *J Biol Chem* **2014,** *289* (31), 21299-310.

143. Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E., EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity. *Proc Natl Acad Sci U S A* **2010,** *107* (17), 7710-5.

144. Muronetz, V. I.; Barinova, K.; Kudryavtseva, S.; Medvedeva, M.; Melnikova, A.; Sevostyanova, I.; Semenyuk, P.; Stroylova, Y.; Sova, M., Natural and Synthetic Derivatives of Hydroxycinnamic Acid Modulating the Pathological Transformation of Amyloidogenic Proteins. *Molecules* **2020,** *25* (20).

145. Medvedeva, M.; Barinova, K.; Melnikova, A.; Semenyuk, P.; Kolmogorov, V.; Gorelkin, P.; Erofeev, A.; Muronetz, V., Naturally occurring cinnamic acid derivatives prevent amyloid transformation of alpha-synuclein. *Biochimie* **2020,** *170*, 128-139.

146. Nadeem, M.; Imran, M.; Aslam Gondal, T.; Imran, A.; Shahbaz, M.; Muhammad Amir, R.; Wasim Sajid, M.; Batool Qaisrani, T.; Atif, M.; Hussain, G.; Salehi, B.; Adrian Ostrander, E.; Martorell, M.; Sharifi-Rad, J.; C. Cho, W.; Martins, N.; Therapeutic Potential of Rosmarinic Acid: A Comprehensive Review. *Appl Sci* **2019,** *9* (15), 3139.

147. Rao, J. N.; Dua, V.; Ulmer, T. S., Characterization of alpha-synuclein interactions with selected aggregation-inhibiting small molecules. *Biochemistry* **2008,** *47* (16), 4651-6.

148. Lam, H. T.; Graber, M. C.; Gentry, K. A.; Bieschke, J., Stabilization of alpha-Synuclein Fibril Clusters Prevents Fragmentation and Reduces Seeding Activity and Toxicity. *Biochemistry* **2016,** *55* (4), 675-85.

149. Lendel, C.; Bertoncini, C. W.; Cremades, N.; Waudby, C. A.; Vendruscolo, M.; Dobson, C. M.; Schenk, D.; Christodoulou, J.; Toth, G., On the mechanism of nonspecific inhibitors of protein aggregation: dissecting the interactions of alpha-synuclein with Congo red and lacmoid. *Biochemistry* **2009,** *48* (35), 8322-34.

150. Shaltiel-Karyo, R.; Frenkel-Pinter, M.; Rockenstein, E.; Patrick, C.; Levy-Sakin, M.; Schiller, A.; Egoz-Matia, N.; Masliah, E.; Segal, D.; Gazit, E., A blood-brain barrier (BBB) disrupter is also a potent alpha-synuclein (alpha-syn) aggregation inhibitor: a novel dual mechanism of mannitol for the treatment of Parkinson disease (PD). *J Biol Chem* **2013,** *288* (24), 17579-88.

151. Valiente-Gabioud, A. A.; Miotto, M. C.; Chesta, M. E.; Lombardo, V.; Binolfi, A.; Fernandez, C. O., Phthalocyanines as Molecular Scaffolds to Block Disease-Associated Protein Aggregation. *Acc Chem Res* **2016,** *49* (5), 801-8.

152. Lee, E. N.; Cho, H. J.; Lee, C. H.; Lee, D.; Chung, K. C.; Paik, S. R., Phthalocyanine tetrasulfonates affect the amyloid formation and cytotoxicity of alpha-synuclein. *Biochemistry* **2004,** *43* (12), 3704-15.

153. Lamberto, G. R.; Binolfi, A.; Orcellet, M. L.; Bertoncini, C. W.; Zweckstetter, M.; Griesinger, C.; Fernandez, C. O., Structural and mechanistic basis behind the inhibitory interaction of PcTS on alpha-synuclein amyloid fibril formation. *Proc Natl Acad Sci U S A* **2009,** *106* (50), 21057-62.

154. Fonseca-Ornelas, L.; Schmidt, C.; Camacho-Zarco, A. R.; Fernandez, C. O.; Becker, S.; Zweckstetter, M., Small-Molecule-Induced Soluble Oligomers of alpha-Synuclein with Helical Structure. *Chemistry* **2017,** *23* (53), 13010-13014.

155. Fonseca-Ornelas, L.; Eisbach, S. E.; Paulat, M.; Giller, K.; Fernandez, C. O.; Outeiro, T. F.; Becker, S.; Zweckstetter, M., Small molecule-mediated stabilization of vesicle-

associated helical alpha-synuclein inhibits pathogenic misfolding and aggregation. *Nat Commun* **2014,** *5*, 5857.

156. Lamberto, G. R.; Torres-Monserrat, V.; Bertoncini, C. W.; Salvatella, X.; Zweckstetter, M.; Griesinger, C.; Fernandez, C. O., Toward the discovery of effective polycyclic inhibitors of alpha-synuclein amyloid assembly. *J Biol Chem* **2011,** *286* (37), 32036-44.

157. Hayden, E. Y.; Kaur, P.; Williams, T. L.; Matsui, H.; Yeh, S. R.; Rousseau, D. L., Heme Stabilization of alpha-Synuclein Oligomers during Amyloid Fibril Formation. *Biochemistry* **2015,** *54* (30), 4599-610.

158. Ono, K.; Yamada, M., Vitamin A potently destabilizes preformed alpha-synuclein fibrils in vitro: implications for Lewy body diseases. *Neurobiol Dis* **2007,** *25* (2), 446-54.

159. Ardah, M. T.; Paleologou, K. E.; Lv, G.; Menon, S. A.; Abul Khair, S. B.; Lu, J. H.; Safieh-Garabedian, B.; Al-Hayani, A. A.; Eliezer, D.; Li, M.; El-Agnaf, O. M., Ginsenoside Rb1 inhibits fibrillation and toxicity of alpha-synuclein and disaggregates preformed fibrils. *Neurobiol Dis* **2015,** *74*, 89-101.

160. Hong, D. P.; Fink, A. L.; Uversky, V. N., Smoking and Parkinson's disease: does nicotine affect alpha-synuclein fibrillation? *Biochim Biophys Acta* **2009,** *1794* (2), 282-90.

161. Tavassoly, O.; Kakish, J.; Nokhrin, S.; Dmitriev, O.; Lee, J. S., The use of nanopore analysis for discovering drugs which bind to alpha-synuclein for treatment of Parkinson's disease. *Eur J Med Chem* **2014,** *88*, 42-54.

162. Kardani, J.; Roy, I., Understanding Caffeine's Role in Attenuating the Toxicity of alpha-Synuclein Aggregates: Implications for Risk of Parkinson's Disease. *ACS Chem Neurosci* **2015,** *6* (9), 1613-25.

163. Gonzalez-Lizarraga, F.; Ploper, D.; Avila, C. L.; Socias, S. B.; Dos-Santos-Pereira, M.; Machin, B.; Del-Bel, E.; Michel, P. P.; Pietrasanta, L. I.; Raisman-Vozari, R.; Chehin, R., CMT-3 targets different alpha-synuclein aggregates mitigating their toxic and inflammogenic effects. *Sci Rep* **2020,** *10* (1), 20258.

164. Wagner, J.; Ryazanov, S.; Leonov, A.; Levin, J.; Shi, S.; Schmidt, F.; Prix, C.; Pan-Montojo, F.; Bertsch, U.; Mitteregger-Kretzschmar, G.; Geissen, M.; Eiden, M.; Leidel, F.; Hirschberger, T.; Deeg, A. A.; Krauth, J. J.; Zinth, W.; Tavan, P.; Pilger, J.; Zweckstetter, M.; Frank, T.; Bahr, M.; Weishaupt, J. H.; Uhr, M.; Urlaub, H.; Teichmann, U.; Samwer, M.; Botzel, K.; Groschup, M.; Kretzschmar, H.; Griesinger, C.; Giese, A., Anle138b: a novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease. *Acta Neuropathol* **2013,** *125* (6), 795-813.

165. Levin, J.; Schmidt, F.; Boehm, C.; Prix, C.; Botzel, K.; Ryazanov, S.; Leonov, A.; Griesinger, C.; Giese, A., The oligomer modulator anle138b inhibits disease progression in a Parkinson mouse model even with treatment started after disease onset. *Acta Neuropathol* **2014,** *127* (5), 779-80.

166. Wrasidlo, W.; Tsigelny, I. F.; Price, D. L.; Dutta, G.; Rockenstein, E.; Schwarz, T. C.; Ledolter, K.; Bonhaus, D.; Paulino, A.; Eleuteri, S.; Skjevik, A. A.; Kouznetsova, V. L.; Spencer, B.; Desplats, P.; Gonzalez-Ruelas, T.; Trejo-Morales, M.; Overk, C. R.; Winter, S.; Zhu, C.; Chesselet, M. F.; Meier, D.; Moessler, H.; Konrat, R.; Masliah, E., A de novo

compound targeting alpha-synuclein improves deficits in models of Parkinson's disease. *Brain* **2016,** *139* (Pt 12), 3217-3236.

167. Price, D. L.; Koike, M. A.; Khan, A.; Wrasidlo, W.; Rockenstein, E.; Masliah, E.; Bonhaus, D., The small molecule alpha-synuclein misfolding inhibitor, NPT200-11, produces multiple benefits in an animal model of Parkinson's disease. *Sci Rep* **2018,** *8* (1), 16165.

168. Pujols, J.; Pena-Diaz, S.; Lazaro, D. F.; Peccati, F.; Pinheiro, F.; Gonzalez, D.; Carija, A.; Navarro, S.; Conde-Gimenez, M.; Garcia, J.; Guardiola, S.; Giralt, E.; Salvatella, X.; Sancho, J.; Sodupe, M.; Outeiro, T. F.; Dalfo, E.; Ventura, S., Small molecule inhibits alpha-synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proc Natl Acad Sci U S A* **2018,** *115* (41), 10481-10486.

169. Pena-Diaz, S.; Pujols, J.; Conde-Gimenez, M.; Carija, A.; Dalfo, E.; Garcia, J.; Navarro, S.; Pinheiro, F.; Santos, J.; Salvatella, X.; Sancho, J.; Ventura, S., ZPD-2, a Small Compound That Inhibits alpha-Synuclein Amyloid Aggregation and Its Seeded Polymerization. *Front Mol Neurosci* **2019,** *12*, 306.

170. Pena-Diaz, S.; Pujols, J.; Pinheiro, F.; Santos, J.; Pallares, I.; Navarro, S.; Conde-Gimenez, M.; Garcia, J.; Salvatella, X.; Dalfo, E.; Sancho, J.; Ventura, S., Inhibition of alpha-Synuclein Aggregation and Mature Fibril Disassembling With a Minimalistic Compound, ZPDm. *Front Bioeng Biotechnol* **2020,** *8*, 588947.

171. Toth, G.; Neumann, T.; Berthet, A.; Masliah, E.; Spencer, B.; Tao, J.; Jobling, M. F.; Gardai, S. J.; Bertoncini, C. W.; Cremades, N.; Bova, M.; Ballaron, S.; Chen, X. H.; Mao, W.; Nguyen, P.; Tabios, M. C.; Tambe, M. A.; Rochet, J. C.; Junker, H. D.; Schwizer, D.; Sekul, R.; Ott, I.; Anderson, J. P.; Szoke, B.; Hoffman, W.; Christodoulou, J.; Yednock, T.; Agard, D. A.; Schenk, D.; McConlogue, L., Novel Small Molecules Targeting the Intrinsically Disordered Structural Ensemble of alpha-Synuclein Protect Against Diverse alpha-Synuclein Mediated Dysfunctions. *Sci Rep* **2019,** *9* (1), 16947.

172. Maqbool, M.; Gadhavi, J.; Hivare, P.; Gupta, S.; Hoda, N., Diphenyl triazine hybrids inhibit alpha-synuclein fibrillogenesis: Design, synthesis and in vitro efficacy studies. *Eur J Med Chem* **2020,** *207*, 112705.

173. Wolber, G.; Langer, T., LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model* **2005,** *45* (1), 160-9.

174. Gitto, R.; De Luca, L.; Ferro, S.; Russo, E.; De Sarro, G.; Chisari, M.; Ciranna, L.; Alvarez-Builla, J.; Alajarin, R.; Buemi, M. R.; Chimirri, A., Synthesis, modelling and biological characterization of 3-substituted-1H-indoles as ligands of GluN2B-containing N-methyl-d-aspartate receptors. *Bioorg Med Chem* **2014,** *22* (3), 1040-8.

175. Pujols, J.; Pena-Diaz, S.; Conde-Gimenez, M.; Pinheiro, F.; Navarro, S.; Sancho, J.; Ventura, S., High-Throughput Screening Methodology to Identify Alpha-Synuclein Aggregation Inhibitors. *Int J Mol Sci* **2017,** *18* (3).

176. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009,** *30* (16), 2785-91.

177. Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M.; George, J. M.; Rienstra, C. M., Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat Struct Mol Biol* **2016,** *23* (5), 409-15.

178. Pedretti, A.; Villa, L.; Vistoli, G., VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. *J Comput Aided Mol Des* **2004,** *18* (3), 167-73.

179. BIOVIA, D. S. *BIOVIA Discovery Studio Visualizer*, 20; San Diego, 2020.

180. Zoete, V.; Daina, A.; Bovigny, C.; Michielin, O., SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. *J Chem Inf Model* **2016,** *56* (8), 1399-404.

181. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 3-26.

182. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* **2002,** *45* (12), 2615-23.

183. Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G., Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **2008,** *3* (3), 435-44.

184. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997,** *267* (3), 727-48.

185. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *J Cheminform* **2011,** *3*, 33.

186. Gladysz, R.; Dos Santos, F. M.; Langenaeker, W.; Thijs, G.; Augustyns, K.; De Winter, H., Spectrophores as one-dimensional descriptors calculated from three-dimensional atomic properties: applications ranging from scaffold hopping to multi-target virtual screening. *J Cheminform* **2018,** *10* (1), 9.

187. Armstrong, M. S.; Finn, P. W.; Morris, G. M.; Richards, W. G., Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension. *J Comput Aided Mol Des* **2011,** *25* (8), 785-90.

188. Bose, M.; Mukherjee, P., Potential of Anti-MUC1 Antibodies as a Targeted Therapy for Gastrointestinal Cancers. *Vaccines (Basel)* **2020,** *8* (4).

189. Nath, S.; Mukherjee, P., MUC1: a multifaceted oncoprotein with a key role in cancer progression. *Trends Mol Med* **2014,** *20* (6), 332-42.

190. Brayman, M.; Thathiah, A.; Carson, D. D., MUC1: a multifunctional cell surface component of reproductive tissue epithelia. *Reprod Biol Endocrinol* **2004,** *2*, 4.

191. Dhar, P.; McAuley, J., The Role of the Cell Surface Mucin MUC1 as a Barrier to Infection and Regulator of Inflammation. *Front Cell Infect Microbiol* **2019,** *9*, 117.

192. Gao, T.; Cen, Q.; Lei, H., A review on development of MUC1-based cancer vaccine. *Biomed Pharmacother* **2020,** *132*, 110888.

193. Beckwith, D. M.; Cudic, M., Tumor-associated O-glycans of MUC1: Carriers of the glyco-code and targets for cancer vaccine design. *Semin Immunol* **2020,** *47*, 101389.

194. Suzuki, H.; Shoda, J.; Kawamoto, T.; Shinozaki, E.; Miyahara, N.; Hotta, S.; Iizuka, Y.; Nakahara, A.; Tanaka, N.; Yanaka, A.; Irimura, T., Expression of MUC1 recognized by monoclonal antibody MY.1E12 is a useful biomarker for tumor aggressiveness of advanced colon carcinoma. *Clin Exp Metastasis* **2004,** *21* (4), 321-9.

195. Hollingsworth, M. A.; Swanson, B. J., Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* **2004,** *4* (1), 45-60.

196. Pourjafar, M.; Samadi, P.; Saidijam, M., MUC1 antibody-based therapeutics: the promise of cancer immunotherapy. *Immunotherapy* **2020,** *12* (17), 1269-1286.

197. Ibrahim, N. K.; Yariz, K. O.; Bondarenko, I.; Manikhas, A.; Semiglazov, V.; Alyasova, A.; Komisarenko, V.; Shparyk, Y.; Murray, J. L.; Jones, D.; Senderovich, S.; Chau, A.; Erlandsson, F.; Acton, G.; Pegram, M., Randomized phase II trial of letrozole plus anti-MUC1 antibody AS1402 in hormone receptor-positive locally advanced or metastatic breast cancer. *Clin Cancer Res* **2011,** *17* (21), 6822-30.

198. Corraliza-Gorjon, I.; Somovilla-Crespo, B.; Santamaria, S.; Garcia-Sanz, J. A.; Kremer, L., New Strategies Using Antibody Combinations to Increase Cancer Treatment Effectiveness. *Front Immunol* **2017,** *8*, 1804.

199. Venepalli, N. K., Gandhi, C.C., Ozer, H.; Ho, D., Lu, Y., Xie, H., Berg, S.A., Chowdhery, R.A., Gargano, M.A.,; Braun, A. H., Phase Ib study of PGG beta glucan in combination with anti-MUC1 antibody (BTH1704) and gemcitabine for the treatment of advanced pancreatic cancer. *J Clin Oncol* **2015,** *33* (3), TPS493.

200. Curry, J. M.; Thompson, K. J.; Rao, S. G.; Besmer, D. M.; Murphy, A. M.; Grdzelishvili, V. Z.; Ahrens, W. A.; McKillop, I. H.; Sindram, D.; Iannitti, D. A.; Martinie, J. B.; Mukherjee, P., The use of a novel MUC1 antibody to identify cancer stem cells and circulating MUC1 in mice and patients with pancreatic cancer. *J Surg Oncol* **2013,** *107* (7), 713-22.

201. Moore, L. J.; Roy, L. D.; Zhou, R.; Grover, P.; Wu, S. T.; Curry, J. M.; Dillon, L. M.; Puri, P. M.; Yazdanifar, M.; Puri, R.; Mukherjee, P.; Dreau, D., Antibody-Guided In Vivo Imaging for Early Detection of Mammary Gland Tumors. *Transl Oncol* **2016,** *9* (4), 295-305.

202. Bose, M., Mukherjee, P. In *A novel antibody blocks anti-apoptotic activity of MUC1 in pancreatic cancer cell lines*, Proceedings of the American Association for Cancer Research Annual Meeting 2019, Atlanta, GA, Cancer Res: Atlanta, GA, 2019.

203. Yonezawa, S.; Kitajima, S.; Higashi, M.; Osako, M.; Horinouchi, M.; Yokoyama, S.; Kitamoto, S.; Yamada, N.; Tamura, Y.; Shimizu, T.; Tabata, M.; Goto, M., A novel anti-MUC1 antibody against the MUC1 cytoplasmic tail domain: use in sensitive identification of poorly differentiated cells in adenocarcinoma of the stomach. *Gastric Cancer* **2012,** *15* (4), 370-81.

204. Dian, D.; Janni, W.; Kuhn, C.; Mayr, D.; Karsten, U.; Mylonas, I.; Friese, K.; Jeschke, U., Evaluation of a novel anti-mucin 1 (MUC1) antibody (PankoMab) as a potential diagnostic tool in human ductal breast cancer; comparison with two established antibodies. *Onkologie* **2009,** *32* (5), 238-44.

205. Danielczyk, A.; Stahn, R.; Faulstich, D.; Loffler, A.; Marten, A.; Karsten, U.; Goletz, S., PankoMab: a potent new generation anti-tumour MUC1 antibody. *Cancer Immunol Immunother* **2006,** *55* (11), 1337-47.

206. Fiedler, W.; DeDosso, S.; Cresta, S.; Weidmann, J.; Tessari, A.; Salzberg, M.; Dietrich, B.; Baumeister, H.; Goletz, S.; Gianni, L.; Sessa, C., A phase I study of PankoMab-GEX, a humanised glyco-optimised monoclonal antibody to a novel tumour-specific MUC1 glycopeptide epitope in patients with advanced carcinomas. *Eur J Cancer* **2016,** *63*, 55-63.

207. Dikic, I., CIN85/CMS family of adaptor molecules. *FEBS Lett* **2002,** *529* (1), 110-5.

208. Kowanetz, K.; Husnjak, K.; Holler, D.; Kowanetz, M.; Soubeyran, P.; Hirsch, D.; Schmidt, M. H. H.; Pavelic, K.; De Camilli, P.; Randazzo, P. A.; Dikic, I., CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors. *Mol Biol Cell* **2004,** *15* (7), 3155-66.

209. Kurochkina, N.; Guha, U., SH3 domains: modules of protein-protein interactions. *Biophys Rev* **2013,** *5* (1), 29-39.

210. Saksela, K.; Permi, P., SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Lett* **2012,** *586* (17), 2609-14.

211. Oneyama, C.; Nakano, H.; Sharma, S. V., UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug. *Oncogene* **2002,** *21* (13), 2037-50.

212. Larson, S. M.; Davidson, A. R., The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. *Protein Sci* **2000,** *9* (11), 2170-80.

213. Jozic, D.; Cardenes, N.; Deribe, Y. L.; Moncalian, G.; Hoeller, D.; Groemping, Y.; Dikic, I.; Rittinger, K.; Bravo, J., Cbl promotes clustering of endocytic adaptor proteins. *Nat Struct Mol Biol* **2005,** *12* (11), 972-9.

214. Hashimoto, S.; Hirose, M.; Hashimoto, A.; Morishige, M.; Yamada, A.; Hosaka, H.; Akagi, K.; Ogawa, E.; Oneyama, C.; Agatsuma, T.; Okada, M.; Kobayashi, H.; Wada, H.; Nakano, H.; Ikegami, T.; Nakagawa, A.; Sabe, H., Targeting AMAP1 and cortactin binding bearing an atypical src homology 3/proline interface for prevention of breast cancer invasion and metastasis. *Proc Natl Acad Sci U S A* **2006,** *103* (18), 7036-41.

215. Desrochers, G.; Cappadocia, L.; Lussier-Price, M.; Ton, A. T.; Ayoubi, R.; Serohijos, A.; Omichinski, J. G.; Angers, A., Molecular basis of interactions between SH3 domain-containing proteins and the proline-rich region of the ubiquitin ligase Itch. *J Biol Chem* **2017,** *292* (15), 6325-6338.

216. Mayer, B. J., SH3 domains: complexity in moderation. *J Cell Sci* **2001,** *114* (Pt 7), 1253-63.

217. Ceregido, M. A.; Garcia-Pino, A.; Ortega-Roldan, J. L.; Casares, S.; Lopez Mayorga, O.; Bravo, J.; van Nuland, N. A.; Azuaga, A. I., Multimeric and differential binding of CIN85/CD2AP with two atypical proline-rich sequences from CD2 and Cbl-b*. *FEBS J* **2013,** *280* (14), 3399-415.

218. Kurakin, A. V.; Wu, S.; Bredesen, D. E., Atypical recognition consensus of CIN85/SETA/Ruk SH3 domains revealed by target-assisted iterative screening. *J Biol Chem* **2003,** *278* (36), 34102-9.

219. Buchse, T.; Horras, N.; Lenfert, E.; Krystal, G.; Korbel, S.; Schumann, M.; Krause, E.; Mikkat, S.; Tiedge, M., CIN85 interacting proteins in B cells-specific role for SHIP-1. *Mol Cell Proteomics* **2011,** *10* (10), M110 006239.

220. Gaidos, G.; Soni, S.; Oswald, D. J.; Toselli, P. A.; Kirsch, K. H., Structure and function analysis of the CMS/CIN85 protein family identifies actin-bundling properties and heterotypic-complex formation. *J Cell Sci* **2007,** *120* (Pt 14), 2366-77.

221. Li, Q.; Yang, W.; Wang, Y.; Liu, W., Biochemical and Structural Studies of the Interaction between ARAP1 and CIN85. *Biochemistry* **2018,** *57* (14), 2132-2139.

222. Hutchings, N. J.; Clarkson, N.; Chalkley, R.; Barclay, A. N.; Brown, M. H., Linking the T cell surface protein CD2 to the actin-capping protein CAPZ via CMS and CIN85. *J Biol Chem* **2003,** *278* (25), 22396-403.

223. Havrylov, S.; Redowicz, M. J.; Buchman, V. L., Emerging roles of Ruk/CIN85 in vesicle-mediated transport, adhesion, migration and malignancy. *Traffic* **2010,** *11* (6), 721-31.

224. Cascio, S.; Farkas, A. M.; Hughey, R. P.; Finn, O. J., Altered glycosylation of MUC1 influences its association with CIN85: the role of this novel complex in cancer cell invasion and migration. *Oncotarget* **2013,** *4* (10), 1686-97.

225. Ciborowski, P.; Finn, O. J., Non-glycosylated tandem repeats of MUC1 facilitate attachment of breast tumor cells to normal human lung tissue and immobilized extracellular matrix proteins (ECM) in vitro: potential role in metastasis. *Clin Exp Metastasis* **2002,** *19* (4), 339-45.

226. Dennis, J. W.; Granovsky, M.; Warren, C. E., Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **1999,** *1473* (1), 21-34.

227. Kohlgraf, K. G.; Gawron, A. J.; Higashi, M.; Meza, J. L.; Burdick, M. D.; Kitajima, S.; Kelly, D. L.; Caffrey, T. C.; Hollingsworth, M. A., Contribution of the MUC1 tandem repeat and cytoplasmic tail to invasive and metastatic properties of a pancreatic cancer cell line. *Cancer Res* **2003,** *63* (16), 5011-20.

228. Cascio, S., Sciurba, J., Hughey, R., Camacho, C., Finn, O. In *Muc1/Cin85 complex is a new molecular target for control of cancer invasion and metastasis.* , Proceedings of the 105th Annual Meeting of the American Association for Cancer Research, San Diego, CA, San Diego, CA, 2014.

229. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **2009,** *10*, 168.

230. Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **2015,** *10* (5), 733-55.

231. Landon, M. R.; Amaro, R. E.; Baron, R.; Ngan, C. H.; Ozonoff, D.; McCammon, J. A.; Vajda, S., Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem Biol Drug Des* **2008,** *71* (2), 106-16.

232. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** *30* (10), 1545-614.

233. Tubiana, T.; Carvaillo, J. C.; Boulard, Y.; Bressanelli, S., TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries. *J Chem Inf Model* **2018,** *58* (11), 2178-2182.

234. Schmidtke, P.; Barril, X., Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* **2010,** *53* (15), 5858-67.

235. Schuetz, D. A.; Seidel, T.; Garon, A.; Martini, R.; Korbel, M.; Ecker, G. F.; Langer, T., GRAIL: GRids of phArmacophore Interaction fieLds. *J Chem Theory Comput* **2018,** *14* (9), 4958-4970.

236. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** *31* (2), 455-61.

237. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008,** *29* (11), 1859-65.

238. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

239. Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martinez-Saez, N.; Castro-Lopez, J.; Garcia-Martin, F.; Companon, I.; Hinou, H.; Nishimura, S. I.; Jimenez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F., The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer. *J Am Chem Soc* **2017,** *139* (50), 18255-18261.

240. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **2004,** *47* (7), 1750-9.

241. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** *47* (7), 1739-49.

242. Bowers, K. J., Chow, E., Xu, H, Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y. and Shaw D.E. In *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, Tampa, Florida, 2006.

243. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **2013,** *27* (3), 221-34.

244. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D., OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J Chem Theory Comput* **2019,** *15* (3), 1863-1874.