

*Application of computer-based
methods to identify and investigate
ligands for proteins with high flexible
binding sites*

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der Naturwissenschaftlichen Fakultät I - Biowissenschaften
der Martin-Luther-Universität Halle-Wittenberg

von

Chiara Luise

geboren am 07. März 1988

Gutachter:

1. Prof. Dr. Wolfgang Sippl
2. Prof. Dr. Maria Letizia Barreca
3. Prof. Dr. Mike Schutkowski

Datum der Disputation:

28.09.2022

Life never ends

to my beloved nephew Jacopo, the joy of my life
to my dearest aunt Cinzia, who protects me from heaven

Page left intentionally blank.

ABSTRACT

Computer-aided drug design (CADD) methodologies are extensively used in pharmaceutical industries and academia, and nowadays play a pivotal role in the drug discovery pipeline. The current work illustrates the most widely used *in silico* techniques in drug discovery, emphasizing their applications, potential, limitations and alternative solutions to overcome those limitations –when possible. The results show how these methodologies have been applied in diverse projects to accelerate the discovery of new small molecule inhibitors for flexible protein binding pockets. Specifically, the studies presented herein cover different drug discovery aspects and try to address specific questions and challenges that are project/target related.

In the case of Spindlin1, the herein described work started when no inhibitors were reported, and only crystal structures in apo form and holo forms in complex with histone peptides were available. Nonetheless, a successful iterative virtual screening campaign was conducted, which was coupled with *in vitro* testing and lead optimization studies. This led to the identification of the first set of novel Spindlin1 inhibitors active in the micromolar range. Subsequently, with the availability of crystal structures in complex with inhibitors, we tried to tackle the pocket flexibility challenge by computational means. Our studies highlighted how pocket flexibility could play a role in pose prediction and how different *in silico* approaches perform in sampling pocket conformations. Subsequently, a validated protocol combining induced fit docking and short molecular dynamics simulation was proposed to predict the binding mode of small molecule ligands into a flexible binding pocket. Finally, a congeneric series of Spindlin1 inhibitors was designed and explored. In this context, CADD methodologies were helpful in explaining the lack of activity of some derivatives and defining the key protein-ligand interactions and their contributions to the binding. Moreover, using MM-GBSA as rescoring step, we developed a protocol that discriminates active from inactive compounds that can be applied to guide further optimization steps to prioritize compounds for synthesis and biological characterization.

Within the AChE and BChE work, computational methods such as docking and MM-GBSA were constructively used to predict the binding mode and explain the activity profile of a series of 8-hydroxy-2,7-naphthyridin-2-ium derivatives. In this study, we reported potent dual AChE and BChE inhibitors and a selective AChE inhibitor.

Moreover, we could identify the structural features that affect the biological activity of our series.

Finally, computer-based methods were used in the context of peptides to explore and rationalize from a structural point of view the cleavage behavior of some metalloproteinases, namely NEP, MMP-7, MMP-9, MMP-12 and MMP-14. Additionally, a detailed analysis and comparison of the MMPs pockets under investigation was performed.

Keywords: drug discovery, computer-aided drug design, molecular docking, molecular dynamics simulation, binding free energy calculation, MM-GBSA, virtual screening, pharmacophore modelling, small molecule inhibitors, peptides, epigenetics, Spindlin1, cholinesterases, AChE, BChE, metalloproteinases, NEP, MMP-14.

KURZFASSUNG

Computergestütztes Wirkstoffdesign (CADD) wird in der pharmazeutischen Industrie und in dem akademischen Bereich ausgiebig eingesetzt und spielt heute eine zentrale Rolle in der Wirkstoffentwicklung. Die vorliegende Arbeit befasst sich mit den am häufigsten verwendeten *in silico* Methoden, die in der Arzneimittelentwicklung eingesetzt werden, und hebt ihre Anwendungen, Potenzial, und Beschränkungen hervor. Die Ergebnisse der vorliegenden Arbeit zeigen, wie diese Methoden in verschiedenen Projekten angewandt wurden, um unter anderem die Entdeckung und Entwicklung neuer niedermolekularen Inhibitoren zu beschleunigen, die Flexibilität von Proteinen zu untersuchen und den Bindungsmodus von sowohl niedermolekularen als auch peptidischen Liganden zu erforschen. Die hier vorgestellten Studien befassen sich mit verschiedenen Aspekten der Arzneimittelentdeckung und versuchen, spezifische Fragen und Herausforderungen zu lösen, die mit dem Projekt/Target in Zusammenhang stehen.

Unsere Studien an Spindlin1 als Target begannen als noch keine Inhibitoren bekannt waren und nur Kristallstrukturen in Apo- und Holo-Form im Komplex mit Histonpeptiden vorlagen. Nichtsdestotrotz wurde eine erfolgreiche iterative Virtuelles-Screening-Kampagne durchgeführt, die mit *in vitro* Tests und Studien zur Leitstrukturoptimierung gekoppelt war und zur Identifizierung der ersten neuen Spindlin1-Inhibitoren, die im mikromolaren Bereich aktiv sind, führten. Mit der Verfügbarkeit von Spindlin1-Kristallstrukturen im Komplex mit Inhibitoren versuchten wir anschließend, die Herausforderung der Taschenflexibilität mit rechnerischen Mitteln anzugehen. Wir haben gezeigt, wie die Taschenflexibilität bei der Bindungsmodusvorhersage eine Rolle spielen könnte und wie verschiedene *in silico* Ansätze bei der Erfassung von Taschenkonformationen abschneiden. Daraufhin wurde ein validiertes Protokoll etabliert, das "induced fit docking" und kurze Molekulardynamiksimulationen kombiniert, um den Bindungsmodus von niedermolekularen Liganden in einer flexiblen Bindungstasche vorherzusagen. Schließlich wurde eine Serie von Spindlin1-Inhibitoren designt und synthetisiert, und deren Spindlin1 inhibitorische Aktivität mittels einem etablierten Assays bestimmt. In diesem Zusammenhang waren CADD-Methoden hilfreich, um die mangelnde Aktivität einiger Derivate zu erklären und die wichtigsten Protein-Ligand-Wechselwirkungen, die zur Bindung beitragen, zu bestimmen. Darüber hinaus haben wir unter Verwendung von MM-GBSA als Rescoring-Schritt ein Protokoll entwickelt, das aktive von inaktiven

Verbindungen unterscheidet und das als Leitfaden für weitere Strukturoptimierungsstudien dienen kann.

Im Rahmen der AChE- und BChE-Studien wurden computerdestützte Methoden wie Docking und MM-GBSA konstruktiv eingesetzt, um die Bindungsmodi einer Reihe von 8-Hydroxy-2,7-naphthyridin-2-ium-Derivaten vorherzusagen und deren Aktivitäten rational zu begründen. In dieser Studie wurden potente duale AChE- und BChE-Inhibitoren und einen selektiven AChE-Inhibitor entwickelt. Darüber hinaus konnten die strukturellen Merkmale, die die biologische Aktivität dieser Serie beeinflussen, identifiziert.

Schließlich wurden computergestützte Methoden bei peptidischen Liganden eingesetzt, um das Spaltungsverhalten der Metalloproteinasen NEP, MMP-7, MMP-9, MMP-12 und MMP-14 aus struktureller Sicht zu untersuchen. Zusätzlich wurde eine detaillierte Analyse und einen Vergleich der untersuchten MMP-Taschen durchgeführt.

Schlagwörter: Wirkstoffforschung, computergestütztes Wirkstoffdesign, molekulares Docking, Moleküldynamiksimulation, freie Bindungsenergie, MM-GBSA, virtuelles Screening, Pharmakophor-Modellierung, niedermolekulare Inhibitoren, Peptiden, Epigenetik, Spindlin1, Cholinesterasen, AChE, BChE, Metalloproteinasen, NEP, MMP-14.

ACKNOWLEDGEMENTS

The PhD time and thesis writing have been an incredible journey, made of challenging experiences and unforgettable memories. During this period, I gained professionally and personally at several levels, and I would like to acknowledge those who have contributed and to whom I am extremely grateful.

First and foremost, I would like to thank my supervisor Prof. Dr. Wolfgang Sippl, for his precious support, assistance and guidance. He has been a great mentor, patient and kind, from the very beginning. I am truly thankful for all the opportunities he gave me, for supporting my professional growth through interesting projects, valuable advice and scientific discussions. My deepest gratitude goes to him also for assigning me the position as laboratory teaching assistant, for letting me undertake the internship at Inte:Ligand and participate in the Europin project – those experiences, among many others, have shaped me immensely.

I would like to recognize the invaluable assistance of Dr. Dina Robaa during these years and thank her for her helpful advice and constructive comments. She has been a fantastic colleague and a friend as well – wise, encouraging, compassionate, always direct and honest.

I would like to gratefully acknowledge our collaborators, especially Prof. Dr. Manfred Jung and Dr. Andrea Heinz and their research groups, for their contributions to our projects.

My sincere thanks also go to Prof. Sharon Bryant and Prof. Dr. Thierry Langer for providing me with the opportunity to join their team as an intern at Inte:Ligand, for sharing their expertise, and for giving me access to their R&D laboratory and research facilities.

A big thank goes to all the colleagues I met along the way and with whom I shared my journey: the dear colleagues in our group, the researchers we had the privilege to host in Halle, and the colleagues I met in Vienna. A special thanks to Conrad, Lucas, Ece, Nehal and Riccardo – not just colleagues but friends. Each of you has given me something and been there for me every time I needed.

My sincere gratitude goes towards all my family members for their unconditional love and endless support over the years, to my parents who raised me and gave me the freedom and the courage to choose my own path in life and career. I am immensely grateful to my sister, no matter how physically far away we are from each other, she is always by my side, and I know that she will always be. I want to thank her for bringing

Jacopo to this world; he is simply the joy of my life, a true motivation for being a better person. I would like to extend my sincere thanks to Federico, my brother-in-law, for being my personal coach and encouraging me. I thank my friends scattered around Europe for sharing good and hard times with me, especially Serena and Valentina, because years have passed, but I can always count on them. I would also like to thank Brigitte Schmeja and the African community in Halle, I started joining the meetings thinking about doing voluntary work, and I found a second family.

Finally, words are not enough to thank Antonio. My success would not have been possible without the support, love and nurturing of my partner. Thank you, Antonio, for being understanding and encouraging, for sharing the path, the joy and sorrow.

“Wherever you go, go with all your heart.”

Confucius

Page left intentionally blank.

CONTENTS

| | |
|--|-----------|
| 1. INTRODUCTION | 1 |
| 1.1. <i>In Silico</i> techniques in medicinal chemistry: application, potential and limitations | 1 |
| 1.1.1. Molecular docking | 2 |
| 1.1.2. Molecular Dynamics simulations..... | 8 |
| 1.1.3. Binding Free Energy calculations | 11 |
| 1.1.4. Virtual Screening | 16 |
| 1.2. Relevance of the investigated protein targets | 23 |
| 1.2.1. Spindlin1 | 23 |
| 1.2.2. Acetylcholinesterase and Butyrylcholinesterase..... | 27 |
| 1.2.3. Metalloproteinases: Neprilysin and Matrix metalloproteinase-14 | 29 |
| 2. AIM OF THE WORK..... | 33 |
| 3. RESULTS..... | 35 |
| 3.1. Identification and Structure–Activity Relationship Studies of Small-Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1 | 35 |
| 3.2. Application of Virtual Screening Approaches for the Identification of Small Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1 | 36 |
| 3.3. Exploring aromatic cage flexibility of the histone methyllysine reader protein Spindlin1 and its impact on binding mode prediction: an in silico study..... | 37 |
| 3.4. Structure-Based Design, Docking and Binding Free Energy Calculations of A366 Derivatives as Spindlin1 Inhibitor | 38 |
| 3.5. Synthesis and biological evaluation of 8-hydroxy-2,7-naphthyridin-2-ium salts as novel inhibitors of acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) | 39 |
| 3.6. Degradation of tropoelastin and skin elastin by neprilysin..... | 40 |
| 3.7. MMP-14 degrades tropoelastin and elastin | 41 |
| 4. SUMMARY AND CONCLUSIONS | 43 |
| 4.1. Application of computer-based methods to search and investigate Spindlin1 inhibitors | 43 |
| 4.1.1. Identification of novel Spindlin1 inhibitors through an iterative VS campaign followed by optimization studies | 44 |
| 4.1.2. <i>In silico</i> studies to explore the aromatic cage flexibility and its impact on binding mode prediction | 47 |

| | |
|--|-----------|
| 4.1.3. Molecular Modeling Studies of A366 derivatives | 51 |
| 4.2. Design and binding mode prediction of AChE and BChE inhibitors..... | 56 |
| 4.3. Exploring the cleavage behavior of NEP and MMP-14 by computational means .. | |
| | 58 |
| 4.3.1. NEP | 58 |
| 4.3.2. MMP-14..... | 60 |
| 4.3.3. Common challenges and limitations | 61 |
| 4.4. General conclusions..... | 62 |
| 5. GLOSSARY OF TERMS USED..... | 65 |
| 6. REFERENCES..... | 69 |
| 7. APPENDICES..... | 83 |
| 7.1. Full-text publications and Supporting Information | 83 |
| 7.2. Book chapter: Lysine Reader Proteins | 83 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1.1. Schematic representation of different molecular docking approaches. | 5 |
| Figure 1.2. Examples of typical applications of Molecular Dynamics simulations in drug design projects. | 10 |
| Figure 1.3. Virtual Screening workflow. | 17 |
| Figure 1.4. Examples of pharmacophore modeling. | 20 |
| Figure 1.5. Crystal structure of Spindlin1 in complex with H3K4me3. | 25 |
| Figure 1.6. Chemical structures of the inhibitors mentioned in the text. | 26 |
| Figure 1.7. Crystal structures of human cholinesterases. | 29 |
| Figure 1.8. Crystal structure of NEP in complex with sacubitrilat. | 31 |
| Figure 1.9. Crystal structure of MMP-14 in complex with TIMP-1. | 32 |
| Figure 4.1. Workflow of the Spindlin1 VS campaign. | 45 |
| Figure 4.2. 2D chemical structures and predicted binding modes of compounds 1k and 4q in the Spindlin1 second domain binding site. | 46 |
| Figure 4.3. Analysis of the Spindlin1 aromatic cage. | 48 |
| Figure 4.4. Binding modes of A366 generated through rigid-body docking and IFD in the second domain of Spindlin1 and relative binding mode observed during the MD simulations. | 50 |
| Figure 4.5. Docking poses of A366 , 1c and 1e in the second domain of Spindlin1 and MD interaction heat map. | 53 |
| Figure 4.6. Box-plot representing the ΔG_{bind} distribution obtained for active and inactive Spindlin1 compounds using Prime MM-GBSA. Scaffold of A366 series. 2D structure and predicted binding modes of compound 1s | 54 |
| Figure 4.7. 2D chemical structures of the scaffold under analysis and two reported AChE and BChE inhibitors. | 56 |
| Figure 4.8. Predicted binding modes of compound 3e in AChE and BChE active sites. | 57 |
| Figure 4.9. 2D chemical structures and activity data of compounds 3e and 3f | 57 |
| Figure 4.10. Predicted docking pose of substrate peptide GVPGAIIPG at the active site of NEP. | 59 |
| Figure 4.11. Predicted docking pose of substrate peptide PGAGLGAL at the active site of MMP-14. | 61 |

LIST OF ABBREVIATIONS

| | |
|-----------------|---|
| 2D | Two-Dimensional |
| 3D | Three-Dimensional |
| ACE | Angiotensin-converting enzyme |
| AChE | Acetylcholinesterase |
| AD | Alzheimer's disease |
| AI | Artificial Intelligence |
| A β | Amyloid β |
| BAR | Bennett Acceptance Ratio |
| BChE | Butyrylcholinesterase |
| BFE | Binding Free Energy |
| CADD | Computer-Aided Drug Design |
| CAS | Catalytic Aactive Site |
| DNA | Deoxyribonucleic acid |
| DS | Docking Score |
| ECFPs | Extended-Connectivity Fingerprints |
| EMBL-EBI | European Bioinformatics Institute |
| EMBL | European Molecular Biology Laboratory |
| FDA | Food and Drug Administration |
| FEP | Free Energy Perturbation |
| GLP | G9a-like-protein |
| H3K4me3 | Histone H3 trimethyllysine 4 |
| H3K4me3K9me2 | Histone H3 trimethyllysine 4 demethyllysine 9 |
| H3K4me3K9me3 | Histone H3 trimethyllysine 4 trimethyllysine 9 |
| H3K4me3R8me2a | Histone4 H3 trimethyllysine 4 asymmetric dimethylarginine 8 |
| H4K20me3 | Histone H4 trimethyllysine 20 |
| H4K20me3R23me2a | Histone4 H3 trimethyllysine 20 asymmetric dimethylarginine 23 |
| HTS | High-Throughput Screening |
| IUPAC | International Union of Pure and Applied Chemistry |
| LB | Ligand-Based |
| LBDD | Ligand-Based Drug Design |
| LIE | Linear Interaction Energy |
| MACCS | Molecular ACCess System |
| MD | Molecular Dynamics |

| | |
|---------|--|
| MM-GBSA | Molecular Mechanics/Generalized Born Surface Area |
| MM-PBSA | Molecular Mechanics/Poisson-Boltzmann Surface Area |
| MME | Membrane Metallo-Endopeptidase |
| MMPs | Matrix Metalloproteinases |
| NEP | Neutral Endopeptidase |
| NMDA | N-methyl D-aspartate |
| NMR | Nuclear Magnetic Resonance |
| PAINS | Pan-Assay INterference compounds |
| PAS | Peripheral anionic site |
| PDB | Protein Data Bank |
| PTMs | Post-Translational Modifications |
| REOS | Rapid Elimination Of Swill |
| RMSD | Root-Mean-Square-Deviation |
| RMSF | Root-Mean-Square-Fluctuations |
| SAR | Structure–Activity Relationships |
| SB | Structure-Based |
| SBDD | Structure-Based Drug Design |
| TI | Thermodynamic Integration |
| TIMPs | Tissue Inhibitors of Metalloproteinases |
| VS | Virtual Screening |

Page left intentionally blank.

1. INTRODUCTION

This chapter is divided into two main parts, and it aims at introducing the computational approaches and the targets investigated in the thesis.

1.1. *In Silico* techniques in medicinal chemistry: application, potential and limitations

The use of *in silico* techniques in the field of computational chemistry and drug design has widely increased over the last four decades. The word “*in silico*” was initially coined in 1987 to describe experiments that were performed via computer simulations and were related to biological experiments. Several new approaches have emerged intending to assist scientists in studying and analyzing molecular systems, cellular behavior, gene expression, compounds properties, toxicity and pharmacokinetic properties of new drug candidates [1-4]. In the medicinal chemistry field, the employment of computers to investigate drug candidates has already started in the early 1980s. It can be dated back to 1980 and 1981 when the first article describing the use of computers at Merck was published which was followed by a cover article: i) a scientific overview in Science entitled “Three-dimensional molecular modeling and drug design” [5]; ii) a cover article in Fortune magazine entitled “The Next Industrial Revolution: designing drugs by computer at Merck” [6, 7].

Since then, computer-aided drug design (CADD) drew the interest of both the pharmaceutical industries and academia and nowadays plays a pivotal role in the drug discovery pipeline [7-14]. Indeed, CADD techniques have considerably influenced the development of several approved drugs such as captopril [10], nelfinavir and amprenavir [10], and tirofiban [15].

In the following sections, an overview of the most widely used *in silico* techniques in drug discovery is given, emphasizing their applications, potential, limitations and alternative solutions to overcome those limitations –when possible.

1.1.1. Molecular docking

Molecular docking is a widely used computational method that was pioneered in the early 1980s [16] and is, nowadays, not only employed in small molecule drug design but also in the study of protein-protein interactions and behavior of nanomaterials [17]. In a CADD scenario, docking studies are routinely applied in the hit-identification and lead-optimization phases. The primary objective of docking techniques is to generate a binding mode hypothesis of a molecule in a macromolecular target, typically proteins or nucleic acids, and to estimate its relative binding affinity [17, 18]. The structures of the target of interest are generally retrieved from the Protein Data Bank (PDB) [19], which to this day holds over 180,685 [20] macromolecular 3D structures solved by X-ray crystallography, and to a lesser extent NMR and electron microscopy. If an experimental structure of the protein target is not available, it can, under certain circumstances, be predicted by homology modeling using structural details of family-related members [21-23]. Very recently, a breakthrough came from the field of Artificial Intelligence (AI), as DeepMind, along with EMBL-EBI, has made available through AlphaFold2 highly accurate 3D structure predictions of over 100,000 unique proteins [24].

Both small molecules and peptide can be docked into a target structure, as well as protein to protein. Docking programs are used to “place” ligands into the binding site of the target of interest. In cases where the binding pocket is unknown, so-called “blind docking” can be performed to identify the part of the macromolecule that is more likely to be targeted and act as the binding site, or to search for possible allosteric pockets [25-28]. Nevertheless, in most cases the pocket is already known. Herein, the focus is set on molecular docking of small molecules and peptides to already known binding sites since these computational approaches were applied in all the studies and publications reported in the Ph.D. thesis work. In this context, the term ligand denotes both small molecules and peptides.

The key concepts at the basis of the molecular docking methods are search algorithms and scoring functions, which respectively sample conformations and orientations of the ligand into the pocket, and estimate a score for each of the predicted poses [29-32]. While it is well known that many robust and accurate docking algorithms are available, the reliability of the scoring functions continues to be the major limiting factor of docking

programs [17, 30, 33]. Through docking algorithms, docking programs generate different binding hypotheses for each docked ligand, diversifying its position, orientation and conformation. The binding modes can be similar or distinct based on various factors, e.g., size and flexibility of the ligand, shape of the binding pocket –buried or more solvent exposed–, number of poses outputted, presence of constraints applied during docking procedures, etc. The binding affinity of the obtained poses is then predicted by scoring functions, which estimate if a given molecule is potentially a suitable binder and which binding hypotheses are more favorable [30, 31]. The energy values associated with the poses are referred to as docking scores. Numerous scoring functions have been developed, and they can be classified as force field-based [30, 34], empirical [30, 35, 36], knowledge-based [30, 37, 38] and entropy-based [39-41]. Moreover, different scoring functions can be contemplated at the same time, leading to a consensus scoring approach [30, 42-44].

Empirical scoring functions are based on the assumption that the binding energies can be approximated by a sum of individual uncorrelated terms, a concept first proposed by Böhm, and then further extended [30, 36, 45]. Each empirically based scoring approach has specific terms, as example in **Equation 1** is reported the ChemScore function of Eldridge et al., which is the starting point for Glide scoring [30, 36, 46].

$$\begin{aligned} \Delta G_{bind} = \Delta G_0 + \Delta G_{H-bond} \sum_{il} g_1(\Delta r)g_2(\Delta\alpha) \\ + \Delta G_{metal} \sum_{aM} f(r_{aM}) \\ + \Delta G_{lipo} \sum_{IL} f(r_{IL}) + \Delta G_{rot}H_{rot} \end{aligned} \quad (1)$$

The free energy of binding (ΔG_{bind}) is approximated as a sum of contributing free energy terms of hydrogen bonding (H-bond), metal (metal), lipophilic (lipo) and ligand rotational entropy (rotor) components. ΔG_{H-bond} , ΔG_{metal} , ΔG_{lipo} and ΔG_{rot} are regression coefficients for each corresponding free energy term, whereas ΔG_0 is a regression constant. The summation $\sum_{il} g_1(\Delta r)g_2(\Delta\alpha)$ and $\sum_{IL} f(r_{IL})$ are calculated for all complementary possibilities of hydrogen bonds between ligand atoms (i) and receptor atoms (I), and for all lipophilic ligand atoms (I) and all lipophilic receptor atoms (L), respectively. The metal term, $\sum_{aM} f(r_{aM})$, is estimated for all acceptor and acceptor/donor atoms (a) in the ligand and any metal atoms (M) in the receptor. The free

energy terms are computed with functions (f , g_1 and g_2) which can depend on an angular ($\Delta\alpha$) and/or a distance (Δr) term.

Although the scoring functions vary for the way they approximate the energy of the binding event, they must strike a balance between speed and accuracy, which is ultimately shifted in favor of the speed.

1.1.1.1. Rigid and Flexible docking

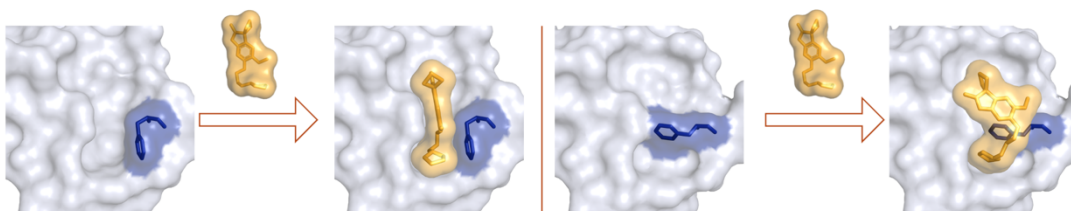
Initially, docking programs treated both ligand and protein as rigid entities, and only the three translational and three rotational degrees of freedom were included [16]. This approach evoked the “lock and key” concept introduced by Emil Fischer in 1894 to describe the enzyme-substrate interactions [47]. With the resolution of macromolecular 3D structures in apo and holo form (bound to a ligand), it became clear that the binding event is much more challenging. In fact, upon the binding of a ligand, the protein target can undergo conformational changes ranging from modest loop motions to hinge bending [48]. At the same time, the ligand can adopt different conformations and orientations in the binding pocket, of which one is more energetically favorable and defined as bioactive conformation. As mentioned above, search algorithms are applied to sample different ligand orientations/conformations and, thus, to treat ligand flexibility [16, 30, 49]. They can be divided into three categories [30]: systematic methods (conformational search, databases, incremental construction); simulation methods (molecular dynamics, energy minimization); and random or stochastic methods (genetic algorithms, Monte Carlo, tabu search).

Currently, there are many docking programs [50], which are based on diverse algorithms and scoring functions, and all of them treat only the ligand as flexible unless otherwise specified [17, 51]. This will be referred to as **rigid-body docking**, or merely docking, when only the ligand is treated as flexible while the protein (body) is kept rigid in its original state (**Figure 1.1**). Depending on the input protein conformation, the ligand adapts itself in the pocket in a different way giving binding modes that reflect the shape of the active site. Hence, the orientation of side chains or loops that impact the pocket will directly influence the poses generated by the docking program. **Figure 1.1.a** presents an example of rigid-body docking using two different protein conformations; in blue is highlighted a specific amino acid that shows alternative side chain orientations.

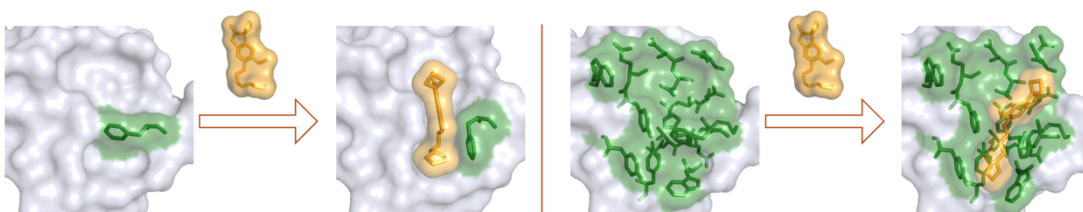
Various docking approaches, which allow the possibility to consider protein flexibility to some extent, have been developed to account for the protein plasticity,

including rotamer libraries [52, 53] and induced fit docking [54, 55]. These approaches will be referred to as **flexible docking**. Programs that employ rotamer libraries allow the user to introduce alternative conformations for specific side chains choosing from a database of rotamers. On the other hand, induced fit docking gives the possibility to treat either specific residues in the binding pocket or an area around the ligand as flexible, as well as allows specification of whole loops to move. This approach relies on the assumption that ligands are often known to induce conformational changes in the pocket upon binding. In **Figure 1.1.b** are illustrated two cases of flexible docking, and possible residues that can be treated as flexible are colored in green.

a. Rigid-body docking



b. Flexible docking



c. Ensemble docking (rigid-body docking)

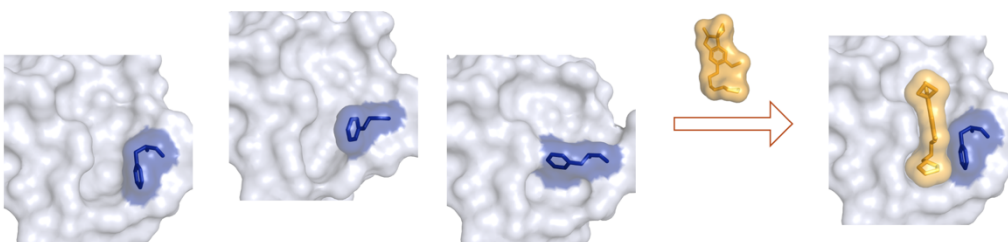


Figure 1.1. Schematic representation of different molecular docking approaches. The binding of a ligand (orange) to a protein target (white) is illustrated; in blue are highlighted different side chains orientations of the same amino acid (**a** and **c**), while in green possible residues that can be treated as flexible during flexible docking (**b**). The latter approach allows the flexibility of either specific residues (left) or an area around the ligand (right). When rigid-body docking is used, different binding hypotheses are generated based on the input protein conformation; in (**a**) two different protein conformations are given as separate examples while in (**e**) three different protein conformations as an example of ensemble docking.

However, flexible docking is time-consuming and may not be feasible for big databases. Another way to account for pocket flexibility, trying to balance the speed of

the calculations, is to use rigid-body docking with an ensemble of protein conformations, which is generally defined as **ensemble docking (Figure 1.1.c)** [56-58]. Here different binding modes are generated and then the best binding hypothesis is selected for each ligand based on the docking score – regardless of the input protein conformation structure. Ensemble docking is performed to evaluate the impact of the protein flexibility on the ligand binding when a pool of multiple crystal structures of the same target is available. Alternatively, molecular dynamics simulations can be performed to generate such a set of different protein conformations [59-61]; the possibility to use molecular dynamics snapshots is discussed in section 1.1.2.

1.1.1.2. Advantages, Limitations, and Perspectives

Molecular docking is the most widely used technique to investigate the ligand binding mode and provide a rough estimation of the ligand binding affinity. Its major advantage is the speed of the calculation that allows the evaluation of millions of compounds (high-throughput docking) in short time and with reasonable computational cost [33, 62]. Moreover, the binding modes predicted by docking offer structural information on the nature of the interactions between the ligand and the protein, which in turn helps not only to explore the key interactions important for the activity but also to guide structure-based optimization studies.

In order to obtain reliable binding modes, the protocol that will be applied in docking studies is tested beforehand. Diverse settings are generally probed, and various strategies can be undertaken based on the data available for the target under investigation. In case there are co-crystal structures deposited in the PDB [19], the ability of the protocol to successfully reproduce the experimentally determined ligand binding modes is assessed (re-docking and cross-docking). When known inhibitors are reported, retrospective studies can be conducted to verify whether the setup being used can discriminate between active and inactive compounds and how the actives are ranked or enriched in comparison to a random selection. Alternatively, decoys can be used in the absence of known inactive ligands [63-65]. However, the use of experimentally confirmed inactive compounds should be privileged since decoys are assumed to be inactive but may be positives in reality [66]. Additionally, the accuracy of the docking protocol is highly related to protein and ligands preparation. As a consequence, some aspects should be carefully addressed. Particular attention has to be given to the protonation states of the amino acids in the binding site as well as to possible isomers and protonation states of the ligands. More

details on critical issues to consider during protein and ligand preparation are given in **chapter 3.2** [67].

Another challenging aspect to evaluate when performing docking is the correct placement of water molecules within the binding site [30, 68]. Indeed, structural water molecules should be incorporated only after careful analysis. In contrast, water molecules that are not directly involved in binding interactions or are not essential to the protein structure can be removed during docking [68]. Nevertheless, there might be cases in which it is difficult to determine which water molecules are conserved and which not. Additional considerations that are needed are: i) ligands can displace conserved water molecules; ii) even if water molecules are not conserved, they can still mediate interactions with ligands (bridging water molecules) [69, 70]. Thus, how to deal with water molecules remains a problematic issue. If the programs allow the possibility, specific water molecules in the binding site can be kept while using the function to toggle or to spin them. For example, the program GOLD offers such options; the toggle option leaves to the program the choice of whether the water molecule should be present or displaced by the ligand during the docking [71].

The main limitations of docking programs are the accuracy of the scoring functions and the possibility to treat protein flexibility [30, 33]. The scoring functions make various assumptions and simplifications in the evaluation of the binding affinity of protein-ligand complexes and, hence, rarely give an accurate estimation of the free energy of binding. Important physical phenomena like the thermodynamics of the free energy of binding and solvation/desolvation of the ligand are not considered into the energy values of most of the docking scoring functions [17]. A strategy to overcome these approximations is to post-process the top-ranked molecules using better but slower re-scoring methods, which are described in the binding free energy calculations section (**1.1.3**).

Proteins are dynamic systems and using only one static frame of the protein structure may lead to missing important aspects. At the same time, introducing flexibility into the docking process would exponentially increase the time for performing the calculations. Therefore, methods that try to balance the speed while considering the binding pocket flexibility have been developed and they have already been discussed above. In summary, alternatives are:

- Flexible docking (induced fit docking), selecting the residues to treat as flexible, e.g., residues that are known to be flexible or that should be explored.

- Ensemble docking, using diverse protein structures that show binding site plasticity.

Nonetheless, both flexible docking and ensemble docking are more computationally time demanding methodologies, and their applicability depends mainly on the objective of the project. For instance, flexible docking can be feasible for a small set of compounds up to a few hundred molecules but not in virtual screening (VS) campaigns (section 1.1.4).

The aspects considered so far are relevant for small molecules as well as for peptide docking, but additional considerations are needed for peptides. Because the accuracy of the scoring function is already a challenge for small molecules, when it comes to peptides, which are bigger in size and more flexible, putting too much trust in docking scores can be misleading. Consequently, peptide docking approaches are mainly used to obtain binding hypothesis, identify key interactions and to support experimental work, e.g., for the interpretation of ambiguous experimental data or to corroborate an assumption [51]. Additionally, the conformational sampling problem for very flexible ligands like peptides should be considered when performing docking. As peptides are more flexible, an exhaustive sampling is needed. To avoid that not enough conformations are sampled, it is advisable to extend the number of poses to generate during the calculations.

Overall, taking into account limitations and advantages and the fact that docking has been widely applied with success, it is considered a cornerstone of structure-based drug design [72].

1.1.2. Molecular Dynamics simulations

Molecular dynamics (MD) is a computational approach used to simulate the dynamic behavior of molecular systems as a function of time, which treats ligand, protein, ions and water molecules as flexible entities by applying Newton's equations of motion [73, 74]. When MD was first applied to a biomolecule in the late 1970s [75], there was great excitement about the potential of displaying the conformational flexibility inherent in proteins, unlike the static pictures retrieved from crystallography [9]. Since then, MD simulations have become powerful computational tools that can help drug design projects at several levels, going from the analysis of protein dynamics to highlighting the key protein-ligand interactions that are stable and likely responsible for the binding and activity [76, 77].

For instance, when it is known that a protein can exist in different conformations or that some side chains in the binding site are highly flexible, MD simulations can be employed to study the protein flexibility and generate an ensemble of multiple protein

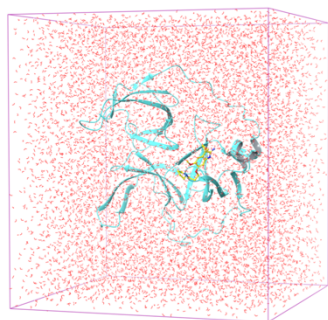
conformations [59-61, 77, 78], as exemplified in **Figure 1.2**. Hence, the obtained MD snapshots can provide an alternative way to create a structurally diverse subset of a protein structure, which can be used not only to evaluate the degree of flexibility but also to perform ensemble docking [79].

Another typical application of MD simulations is to assess the stability of the protein-ligand complex and confirm the binding mode hypothesis obtained by docking [76, 77, 80]. The root-mean-square-deviation (RMSD) and root-mean-square-fluctuations (RMSF) are analyzed to evaluate the stability of the protein and ligand during the simulation. High deviations and fluctuations are generally observed for protein motion, flexible amino acids, unstable binding mode, or ligand atoms that are more flexible. MD simulations can also be employed during optimization studies; here ligands that show unstable binding modes are discarded [77]. Besides, MD simulations provide helpful information to identify the interactions that are crucial for the binding of the inhibitors in the pocket [77]. The analysis of the interactions and their occupancy can shed light on the activity of the compounds and suggest promising chemical modifications that might improve the activity. As an example, in **Figure 1.2** is displayed the binding mode analysis of two ligands, one active and one inactive. A stable binding mode can be observed for the active ligand (Lig-1), which is also reflected in the low RMSD values. Meanwhile, diverse binding modes and high RMSD values can be seen for the inactive ligand (Lig-2). Further clarifications are given when the binding interactions and their occupancy are evaluated, indeed some key interactions are lost during the MD simulation of Lig-2.

Apart from the above-mentioned purposes, classical MD simulations can also be applied to (a) estimate the binding free energy of bound and unbound ligand-protein complexes [78]; (b) identify transient or cryptic sub-pockets [78, 81, 82]; (c) provide relevant information on water molecules and their contribution in modulating ligand binding [83-86].

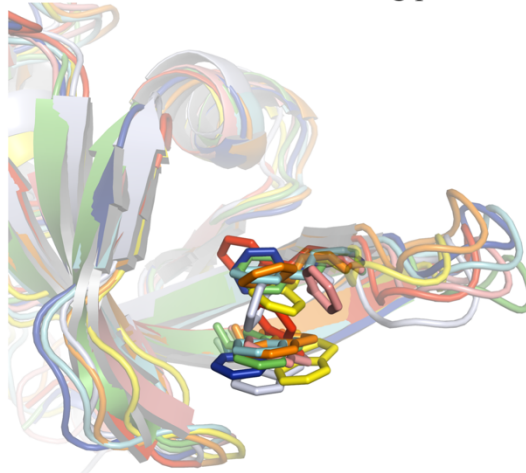
MD model system

Protein in explicit solvent



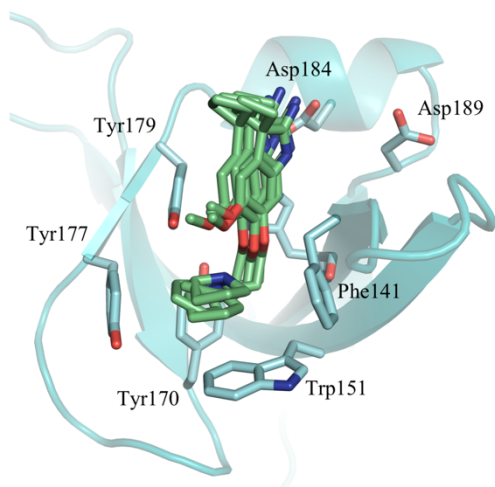
Investigation of protein flexibility

Example of clustering of specific amino acids in the binding pocket

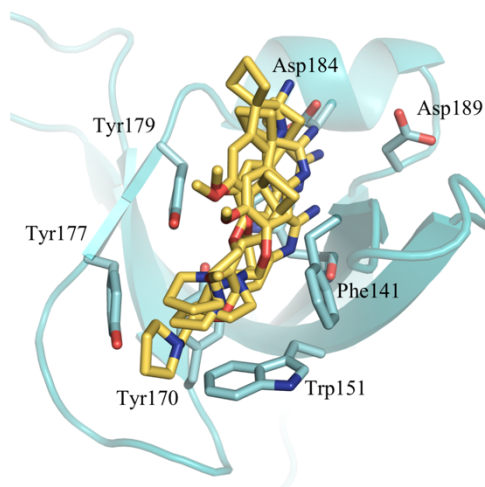


Investigation of binding mode stability and protein-ligand interactions through RMSD plot and heat map

Example of stable binding mode (active ligand, Lig-1)

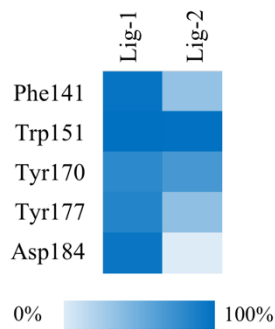
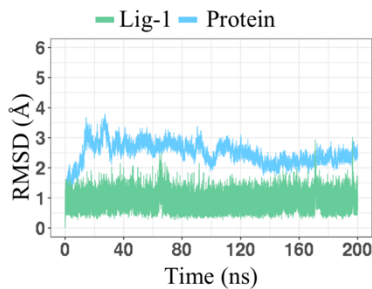


Example of unstable binding mode (inactive ligand, Lig-2)



Interaction heat map

RMSD plot



RMSD plot

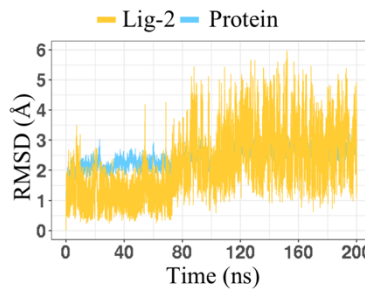


Figure 1.2. Examples of typical applications of Molecular Dynamics (MD) simulations in drug design projects.

The major limitations of MD simulations include the approximations of molecular force fields and the high computational costs [78]. As this approach is still a simulation, approximations are undoubtedly required to render it applicable in large-scale. In spite of that, progress has been made which allows to accurately predict important molecular motions by MD methods [78]. Thanks to the advance of supercomputer and Graphics Processing Units (GPUs), the computational costs associated with MD have been reduced and the speed of the simulation time is constantly increasing [87, 88]. Nevertheless, performing MD simulations is computationally more expensive than docking. Consequently, MD calculations are mostly implemented in the lead optimization phase and only for a limited number of molecules. For instance, simulations in the range of 1-10 nanoseconds can be performed for hundreds of molecules, whereas longer simulations (100-200 nanoseconds) are generally carried out only for a very small dataset of compounds.

The timescale reached during MD simulations can be extended; however, there will still be a gap compared to the timescale observed in experiments. Thus, the limitation of the simulation time leads to a sampling problem. With classical MD simulation, slow processes like folding and binding/unbinding processes cannot be observed. Other techniques that use enhanced sampling methods have been developed to address such problems in accessible timescale by introducing a bias force/potential to the MD simulations to increase the likelihood of detecting slow processes. Examples of enhanced sampling methods are replica exchange [89], metadynamics [90] and steered MD [91].

1.1.3. Binding Free Energy calculations

The ultimate goal of drug design strategies is to find new pharmaceutical compound(s) that bind to a macromolecular target with high affinity, which leads to a desired therapeutic effect. The binding event can be described with the chemical reaction (2):



where L represents the ligand and R the receptor (which typically is a protein); the binding strength is determined by the binding free energy, ΔG_{bind} . Many efforts have been made in developing computer-based methods that could decode the ligand binding interactions and estimate the binding affinity. Indeed, the ability to correctly estimate the

experimentally determined binding affinity and then to predict the activity of new compounds has been a major focus of CADD efforts [33].

As already described in the molecular docking section (1.1.1), docking programs also try to estimate the binding affinity; however, they are not particularly accurate. A strategy to overcome this limitation is to rescore the docked molecules using more accurate but slower approaches [92-97]. Consequently, binding free energy (BFE) calculations have received more attention in the last decades with the aim of improving the ligand binding affinity estimation [33].

BFE calculation methods can be classified in (a) end-state methods, including MM-GB(PB)SA [98] and Linear Interaction Energy (LIE) [99, 100]; and (b) alchemical perturbation methods, such as Thermodynamic Integration (TI) [101], Free Energy Perturbation (FEP) [102] and Bennett Acceptance Ratio (BAR) [103]. These methods either estimate the absolute protein-ligand BFE, which is the free energy difference between the bound and unbound state computed for the complex, or the relative BFE, which is the difference of the binding free energies between two or more ligands [33]. No method outperforms another, they all have pros and cons and the choice depends on the objective of the research and the system under investigation [33]. Even among the same approach, different software programs perform differently based on the target/data and on the protocol used. It is a good practice to evaluate the program and different settings beforehand. Albeit the same accuracy can be obtained by using different methods, the time required for performing the calculations vary significantly. Alchemical perturbation methods are among the more computationally intensive and expensive approaches as they require extensive sampling of the free ligand in solution and of the complex, as well as of unphysical intermediate states. Meanwhile, the end-state methods are based on the sampling of the end-point, which render them faster but still more demanding in comparison to docking scoring functions [98]. Due to the high computational costs, initially, BFE simulations were feasible only for a limited number of compounds [30]. Nowadays, the advances in modern computer hardware have significantly improved the speed of the calculations, making it possible to use these approaches for a bigger number of molecules.

BFE calculations are becoming more popular, and they are routinely applied as a rescoring step after docking [93-95]. If enough data (ligands and relative activity data) is available for a specific target, a protocol can be developed and subsequently used to guide the optimization steps of a series of compounds. Based on the dataset, binary models or

linear regression models can be generated to evaluate the discrimination power and the correlation between the computed BFE values and the experimental affinities. In order to create a reliable linear regression model, the range and distribution of the biological activity data should be as wide as possible covering at least 3 log units [104, 105]. Moreover, biological assays derived from different laboratories should never be considered together in discrimination/correlation studies as this can easily lead to errors and bias in the model. It is important to specify that BFE calculations are generally applied to evaluate congeneric series of compounds (ligands with similar structures) [30]; their accuracy is related to the dataset used and the results depend critically on the tested protein-ligand system [98]. Successful application in the evaluation of congeneric series by BFE calculations have been reported [92, 97].

FEP and TI methods are more restricted methodologies as they can handle only small variation in the ligand scaffold, whereas, MM-GB(PB)SA allows for larger variation [98]. Different studies have evaluated the ability to correlate BFE values retrieved from MM-GB(PB)SA approaches with experimental biological activity in case of more structurally dissimilar ligands, and depending on the target good correlation could be found [93].

1.1.3.1. MM-GB(PB)SA

Among BFE approaches, MM-GB(PB)SA has drawn interest due to its good balance between speed and accuracy [93-97, 106, 107]. The acronym stands for Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) and Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA), where molecular mechanics and continuum solvent models are combined to predict the protein-ligand BFE. The method was established by Kollman and Case in the late 1990s [108-110] with parallel work by other groups [111], and it has been further developed and modified since then [112-114].

The BFE calculated by MM-GB(PB)SA approaches is evaluated according to the Equations (3) and (4) [110].

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) \quad (3)$$

The total free energy of each term in the Equation (3) can be estimated as the sum of the gas-phase energy (E_{MM}), the solvation free energy (ΔG_{sol}) and the entropy contributions ($T\Delta S$), Equation (4).

$$G = E_{MM} + \Delta G_{sol} - T\Delta S \quad (4)$$

Where E_{MM} and ΔG_{sol} are computed as following:

$$E_{MM} = E_{int} + E_{ele} + E_{vdw} \quad (5)$$

$$\Delta G_{sol} = G_{PB/GB} + G_{SA} \quad (6)$$

E_{MM} represents the molecular mechanics free energy, which includes internal energy (E_{int}), electrostatic energy (E_{ele}) and van der Waals energy (E_{vdw}). Meanwhile ΔG_{sol} is the solvation free energy that consists of the polar contributions of the electrostatic solvation energy ($G_{PB/GB}$) and non-polar contributions of the non-electrostatic solvation energy (G_{SA}).

Calculating the entropy term, $T\Delta S$, is the most time-consuming step, thus, diverse attempts have been made to assign its contribution, and it has been proposed that it can be neglected if the BFE of similar ligands is analyzed [113]. In the original MM-PBSA method, the entropic contribution was obtained by a normal-mode analysis, which calculates the vibrational frequencies using a rigid-rotor harmonic-oscillator ideal-gas approximation [98]. Other methods for adding the entropic contribution have been proposed, such as assigning an entropic penalty to the rotatable bonds in the ligand and the protein [115], estimating the entropy by a quasi-harmonic analysis of the MD trajectories [116], or calculating it only for a subset of the snapshots taken from MD simulations [109, 110]. It has also been suggested that computing the entropy of the ligand can be enough, while other studies state that it can be omitted as it does not improve the results in large tests [93, 98, 117]. Indeed, there are many articles in the literature that omit the entropy term and show good accuracy of the BFE calculations [112, 113, 118], and many programs neglect it by default. In the studies presented in this thesis work, the entropic contribution was not evaluated when BFE calculations were performed.

1.1.3.2. Advantages and Limitations

Although BFE calculations has shown good results in estimating binding affinity there are limitations, many of which are applicable to BFE in general while others are methods related. Some limitations have already been discussed above and will be summarized here.

A typical drawback of BFE methods is that they are rigorous but computationally more demanding calculations; thus, they can only be applied for tens up to hundreds of compounds [33]. More often such approaches are used in the lead optimization step because at this stage the dataset being investigated is relatively small and encompasses a congeneric series of compounds. Some BFE methods are faster than others; nonetheless, they still remain too slow and not feasible for bigger databases usually used in VS campaigns. Moreover, their caveat in evaluating the binding affinity of dissimilar scaffolds should be carefully considered when applying these methodologies in a VS scenario (details about VS are given in section 1.1.4). On the other hand, BFE calculations correlate well with experimental biological activity in case of congeneric ligands [33, 92, 97, 119]. Because alchemical perturbation methods are based on energy perturbation, their accuracy is not only restricted to the use of congeneric series but also related to how many heavy atoms are perturbed [33, 120]; and they are more sensitive to changes in the net charge [98].

BFE methods require prior knowledge of the ligand binding mode, and their affinity estimation depends on the starting binding hypothesis. Thus, it is essential to have a high degree of confidence in the binding modes being used as input for BFE [94, 121]. Usually, starting ligand conformations and protein-ligand complexes are taken from docking. If co-crystallized complexes of similar ligands are known, this information can be used during docking studies to reproduce the important protein-ligand interactions and binding mode observed in the resolved 3D structures.

The accuracy of the BFE estimation can vary among the methods/protocol, and not all of them are able to capture small differences in affinities. Despite that, promising results are reported in the literature [92-95, 97, 112, 118, 120, 122, 123]. For example, it has been reported that FEP+ (a newer implementation of FEP) was able to estimate the relative binding affinity with a statistical accuracy in the order of 1 kcal/mol, lower than the accepted force field accuracy (1.5/2 kcal/mol) [33, 120]. In another study, an average unsigned error of about 1.0 kcal/mol was obtained between MM-PBSA values and experimental data [123]. However, the results are strictly associated with the dataset and protein-ligand system under investigation. Additionally, even if a good correlation is found, it is not entirely guaranteed that the method/protocol can reliably predict the activity of new compounds [33, 98].

Other potential challenges of BFE methods are the treatment of explicit water molecules and changes in water structure, and the approximation of the entropic contribution [124, 125].

Weaknesses aside, BFE calculations offer insights into the ligand binding affinity and play important roles in drug discovery [8, 126].

1.1.4. Virtual Screening

In the 1990s, a new approach became a cornerstone technology of pharmaceutical research [9]; the idea behind it was to generate billions of molecules and screen them all to find the right ligand that evoked the desired biological response. High-throughput screening (HTS) was, and still is, applied during the hit-identification stage for the search of new drug candidates. It requires an established assay for the target of interest along with a big library of synthesized compounds. After a naïve phase where HTS were widely applied, the use of this approach was reduced due to the combination of its high costs, extensive preparation and low hits rates [8, 9]. Moreover, the possibility of the identified hit to fail in other drug discovery and development stages highlighted the need to promote alternative strategies that could enhance the success rates at reduced costs and time [8, 9]. This led to the advent and rapid spread of virtual screening (VS) techniques, which have the same goal of HTS but with the potential of transforming random screening into more focused and integrated efforts. Indeed, knowledge about the target of interest is applied during VS campaigns in order to increase the hit rate by reducing the number of compounds to be experimentally tested. VS represents a rational, fast, and cost-effective pre-filtering tool widely implemented in the CADD scenario for the identification of novel and diverse hit structures [62, 127]. Biological evaluation of the selected virtual hits is carried out to confirm the activity of the compounds.

VS approaches can be classified into two branches based on the source of information used: ligand-based (LB) and structure-based (SB), **Figure 1.3**. The LB approaches rely on the availability of known active ligands. Meanwhile, the SB approaches require a 3D protein structure of the target obtained through methods like X-ray crystallography and NMR spectroscopy. In case an experimental structure is not available, homology models can be generated based on known 3D structures of related proteins [21-23]. The two VS approaches can be combined.

In both SB and LB strategies, usually large libraries of compounds are screened. Databases and related aspects are described in subsection **1.1.4.4**.

Virtual Screening workflow

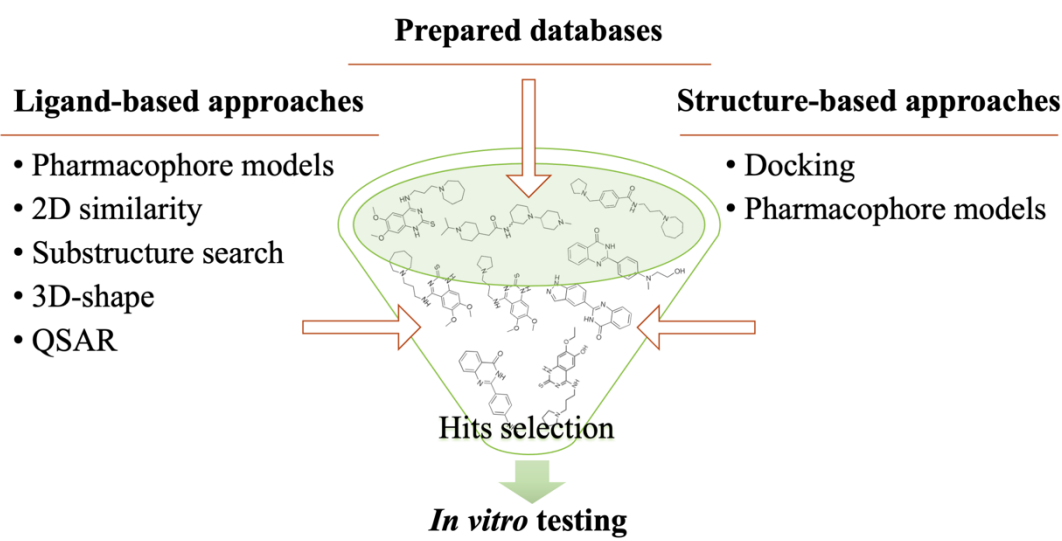


Figure 1.3. Virtual Screening workflow.

1.1.4.1. Ligand-based (LB) approaches

LB methods have been shown to yield promising hit compounds during both hit-identification and lead-optimization phases, either alone or in combination with SB approaches [128-141]. Different methods can be classified among the LB drug design strategies. In a VS scenario, techniques like substructures search, 2D similarity, 3D-shape and LB pharmacophore VS are commonly applied [127, 142-149]. These approaches rely on the idea that similar compounds generally show similar biological effects; a concept that is only partially true as even small substitutions can dramatically impact the activity as underlined by the “similarity paradox” described in the limitation subsection (1.1.4.5) [150, 151].

In the 2D similarity and substructures search VS, a known inhibitor(s) is used as query for screening large databases in a very short time [144-146, 152-154]. During a substructure search VS, the scaffold of an active molecule, which is known to be essential for the activity, is used as a query while searching for possible substituents that can enhance the potency. This strategy is applied primarily in the hit-expansion phase of a VS campaign to develop a SAR around an active molecule. Otherwise, if the VS objective is to search for new chemotypes using known inhibitor(s), 2D similarity search can be carried out. In this case, first molecular fingerprints are generated and then similarity indices, such as the Tanimoto coefficient of similarity, are computed to compare the query molecule to the compounds in the screening library [145, 146]. There are several types of molecular fingerprints, which differ in the method used to encode the structures of the

molecules as sequence of bits or boolean arrays. Some widely used approaches are i) substructure keys-based fingerprints, e.g., Molecular ACCess System (MACCS); ii) circular fingerprints, e.g., Extended-Connectivity Fingerprints (ECFPs). The latter method is more suitable for full structure similarity searching, while the first approach is also used for substructure queries [155].

If a 3D protein of the target under investigation is available, the virtual hits obtained from LB screenings can be subsequently docked to attain ideas about the binding modes and interactions established in the binding pocket before the final selection (SB strategy, the reader is referred to structure-based (1.1.4.2, below) and molecular docking (1.1.1)).

1.1.4.2. Structure-based (SB) approaches

SB methods can be carried out when a 3D protein structure of the target is available; they include docking and SB-pharmacophore VS. Such approaches are performed during the hit-identification phase of drug discovery projects, and they have been proven successful in identifying micromolar binders – which can be further optimized into more potent compounds [128, 139, 149, 156-170]. Additionally, focused libraries are generally screened in the course of hit-optimization studies [171].

Molecular docking is the most commonly used SB method, and it is usually employed in a VS campaign [30, 171, 172]. Docking studies can be used to screen full databases, prefiltered databases, or to better analyze and prioritize hits retrieved from previously run VS (SB-pharmacophore and LB). Using high-throughput docking procedures, databases of up to 10^6 small molecules can be screened in a reasonable time [173]. Often only the top-ranked compounds are visually inspected. Based on their estimated binding affinity and interactions in the binding site of the target, virtual hits are selected for further *in vitro* biological evaluation using an established assay [171, 174].

1.1.4.3. Pharmacophore modeling

According to the IUPAC, a pharmacophore is defined as “the ensemble of steric and electronic features that is necessary to ensure the optimal molecular interactions with a specific biological target structure and to trigger (or to block) its biological response” [175]. Based on the research areas, other definitions have been given. For instance, medicinal chemists have historically described the pharmacophore as the essential functional groups or structural elements of a molecule that are responsible for the biological activity [176, 177]. However, if the classical definition is considered, it can be

noticed that it does not describe a real molecule or association of functional groups, neither specific functional groups [176]. The concept of pharmacophore is indeed more abstract and accounts for common molecular interactions of a group of molecules for a given target protein [176].

Pharmacophore models can be generated using a LB (ligand-based) or a SB (structure-based) approach [149, 178, 179], examples can be seen in **Figure 1.4**. Classical pharmacophoric features are H-bond donors and acceptors (HBD, HBA), positive and negative ionizable groups (PI, NI), hydrophobic moieties (H) and aromatic rings (AR).

LB pharmacophore models are constructed based on the 3D alignment of active molecules: first, the known active ligands are transformed into 3D structures, then conformers are generated for each ligand with the aim of providing a good approximation of the bioactive conformation. Subsequently, the conformers are aligned based on shared chemical features, and hence diverse pharmacophore models are generated which represent the spatial orientation of the common features. Different models can be chosen and tested through validation studies, in case enough information is available for carrying out the validation procedure [180]. Finally, the selected model(s) are used for VS.

SB pharmacophore models can be generated only when a 3D structure of the target protein is available [179]. They mainly rely on the use of a protein-ligand complex as the source of the interaction patterns to assign the pharmacophoric features. Such models reflect the information about the bioactive conformation of the active ligand, the binding interactions between the ligand and the protein, and the shape restrictions of the binding pocket since exclusion volumes can be implemented [149, 177]. Additionally, SB pharmacophore models can be constructed using an apo form protein structure [181-183]. Nevertheless, because this approach misses the concept behind the pharmacophore (molecular interactions information retrieved from the ligand), it is rarely employed. Recently, another strategy has emerged for the generation of SB pharmacophore models which rely on the use of MD simulations of a protein-ligand complex [182-188]. The MD trajectory is decomposed and each inputted snapshot is treated as separate entities for the generation of pharmacophore models. Hence, the evolution of the features over time can be analyzed, as well as their occupancy. Similar to LB pharmacophore modeling, validation studies are performed to evaluate the performance of the generated SB pharmacophore models before using them in VS campaigns [180].

Pharmacophore modeling is commonly applied in VS campaigns and is considered a valuable tool for drug discovery projects [10, 149, 189, 190].

Pharmacophore modeling

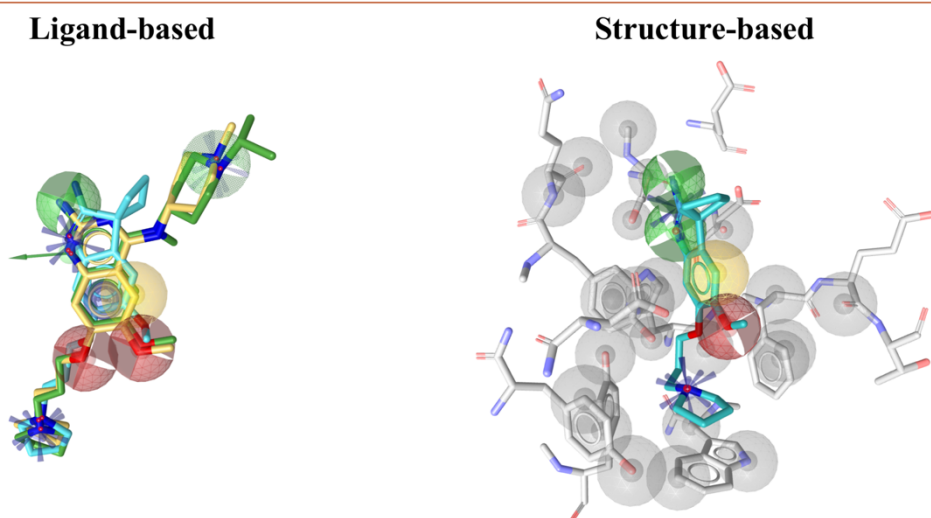


Figure 1.4. Examples of pharmacophore modeling. On the left is depicted a pharmacophore model generated on a set of known inhibitors, while on the right by using a protein-ligand complex. The yellow spheres represent hydrophobic features, red spheres hydrogen-bond acceptor, green arrows/spheres hydrogen-bond donor and the blue spheres positive ionizable features. The excluded volumes are indicated as grey spheres.

1.1.4.4. Databases

Virtual chemical databases are essential for all VS approaches as they represent the source of potential hit compounds. There is a big variety of public and private libraries, such as ChemBridge [191], ChemDiv [192], Enamine [193], IBScreen [194], Sigma-Aldrich [195], Princeton [196], Cayman [197], Maybridge [198], Prestwick [199], Otava [200], etc. The databases can be generic libraries covering large sets of compounds or contain focused and targeted compounds (designed for a specific target and/or containing certain features), as well as already filtered molecules based on drug-like and lead-like properties or favorable physicochemical properties. Moreover, libraries of approved drugs are also available for a drug repositioning strategy [203], e.g., the DrugBank database which contains FDA-approved drugs [204]. It should be emphasized that the libraries to use in a VS campaign must fit the goal of the project. For instance, to identify novel scaffolds for new or relatively unexplored targets, chemically diverse databases are particularly attractive. If the purpose of the VS is lead optimization, then chemical libraries with high intermolecular similarity are a more suitable source [66].

Already in the era 1992–1995, there were debates over how large the total space of molecules is and how the libraries could properly sample that space [9]. To cope with the demands of bigger databases, different vendors provide virtual libraries of billions of compounds [205]. As an example, Enamine offers the REAL Space database that

comprises 14.1 billion make-on-demand molecules and is currently the most extensive library of commercially available compounds [206]. Very recently, a new trend has emerged which favors the use of ultra-large databases in VS with the idea that by covering a bigger chemical space, the chance of finding novel inhibitors increases [207-210]. On the other hand, a viewpoint published in the "New Trends in Virtual Screening" special issue of the Journal of Chemical Information and Modeling tried to counterbalance this tendency by highlighting small-scale VS campaigns that led to the discovery of clinical candidates [211].

Filters, which serve to reduce the number of compounds to be screened, are often applied either during the database preparation or in the refinement of the database prior to the VS [212]. Based on the target, substructure filters can be applied to generate focused libraries containing only molecules with desired chemical features or fragments which are known to be essential for the activity. Typical filters are based on the physicochemical properties of the compounds like the Lipinski's rule of five for drug-likeness [213], Oprea's criteria for lead-likeness [214], or on the reactivity of the molecules like the REOS (rapid elimination of swill) filter [215]. Another frequently applied filter is the PAINS (pan-assay interference compounds) filter, which identifies functional groups and scaffolds of biologically promiscuous compounds known to be non-specific and active against numerous proteins [216-218]. However, PAINS were found among 6–7% of approved drugs [219, 220]; hence some concerns have been raised about the use of PAINS filters as they can lead to discarding high-quality chemical probes and drug candidates [221]. The same happens for some approved drugs, like antibiotics, which fall outside the rule of five for drug-likeness [66, 222]. Therefore, another strategy is to not discard but to flag those compounds that do not pass these filters. Unsuitable compounds could instead be discarded while analyzing the results of the VS bearing in mind that screening hits in general are only starting points which will further undergo optimization steps. Thus, some functional groups can be replaced or removed, and physicochemical properties can be optimized in subsequent steps.

Although numerous databases are available online, there are cases in which custom-made libraries are needed, for example to explore potential chemical modifications that can enhance the activity and physicochemical properties of a validated hit. In this context, virtual focused libraries can be designed in-house using a confirmed scaffold and programs that can suggest novel derivatives for the synthesis according to synthetic and parametric considerations [171, 223].

1.1.4.5. Limitations and Perspectives

Even though VS might be considered a probabilistic game or challenging like finding needles in haystacks, it is largely applied and with success. To enhance its odds, it is important to plan the steps carefully and to consider current limitations and pitfalls [66, 172].

The “similarity paradox” encompasses the notion that minor chemical modifications can render the molecules either active or inactive. This was for example observed for the monoamine oxidase inhibitor pargyline, where only an additional methyl group could lead to a complete loss in inhibitory activity [150, 151]. Therefore, selecting only one or a few representative ligands from a series of similar compounds might turn a successful VS into a failure based on the selected molecules. To avoid such a potential problem, it is advisable to select more representative ligands for the biological evaluation and/or analyze the interactions established at the binding pocket (binding modes predicted by docking) to better evaluate the virtual hits and understand which modification might be beneficial for the activity. Inevitably VS yields active and inactive compounds, but the rationalization of the results can help to explain its successes and guide the optimization studies [30]. Besides, considering that the primary objective of VS is to increase the number of hits compared to a random selection and that it is applied to screen large chemical databases, errors are commonly expected and tolerated [66].

An aspect that might be seen as a drawback of VS is that the hits retrieved are generally active only in the μM range [66, 172], and optimization studies are needed to improve the activity and other properties [224]. This should be further contextualized, as not all weak hits are the same. Indeed, the discovery of other weakly active ligands might be questionable in the case of a well-known target, for which numerous potent inhibitors active have already been reported. By contrast, in the case of novel targets for which only little information is available, weakly or borderline active VS hits can be considered valuable resources [172].

The quality of the derived hits is an additional issue concerning VS results as numerous studies report promiscuous binders and assay-interfering molecules, or often experimental evaluation is not consistent [66, 172]. This can be in part easily resolved by checking if the hits fall into the PAINS filters and then, based on the target, decide whether to test the compounds or not. Again, in case there are no reported inhibitors, it might be worthwhile to have low-quality starting points. Regarding the experimental evaluation, validation of the virtual hits through biological tests is essential and should

always be performed. This might seem obvious; nevertheless, publications that lack biological assays are present in literature [143, 172].

Further pitfalls are method validation and benchmark studies [66, 143, 172]. It can be difficult to validate the protocols, either because it is an intrinsic challenge of the method itself (e.g., for similarity and substructure search) or because the information needed for validation studies is not yet available. Contrariwise, in the case of pharmacophore and docking VS, if there are known active and inactive compounds, it is highly recommended to evaluate and validate the ability of the protocol to discriminate between active and inactive ligands before running the VS campaigns.

Lastly, the structural novelty of the identified hits is another limitation that has been raised since the “rediscovery” of previously known or very similar active compounds by VS is reported [129, 172]. To avoid such cases, resources like CAS SciFinder [225] and PubChem [226] can be applied to evaluate if the identified VS hits are outside the current scientific and patent space for the target being screened.

Despite the limitations mentioned above, VS techniques are valuable resources for finding new hit molecules and exploring the chemical space around the hits.

1.2. Relevance of the investigated protein targets

In the following sections, the protein targets investigated in this work are described. For each target, the physiological roles, the implications in pathological conditions, as well as the available inhibitors, are introduced.

1.2.1. Spindlin1

Epigenetics is the study of heritable changes in gene activity and expression that occur without alteration in the DNA sequence. These non-genetic alterations are tightly regulated by two major epigenetic modifications: chemical modifications of cytosine residues located in the DNA (DNA methylation) and histone modifications, referred to as post-translational modifications (PTMs) of histone tails [227, 228]. Three groups of proteins play a role in the PTMs and are classified as writers, erasers and readers, which respectively introduce, remove and recognize specific modifications. Typical modifications that alter the chromatin structure by adding chemical groups on the histone tails are acetylation, methylation, phosphorylation, sumoylation and ubiquitylation [229]. Additionally, reader proteins are recruited at the histone tail to read the marker and guide further downstream signaling cascades [230-233].

A growing interest in epigenetic mechanisms has led to the recognition that PTMs mechanisms contribute to the genesis and development of several human diseases, most importantly cancer, diabetes and neurodegenerative diseases [234]. Up to date, only histone deacetylase inhibitors have been approved by the FDA [235-237]. Meanwhile, other fields of epigenetics have not yet been sufficiently investigated, although they comprise some highly attractive biological targets for the field of drug discovery.

Spindlin1 is a relatively newly identified chromatin reader protein, which is composed of three Tudor domains (**Figure 1.5**). It recognizes the H3K4me3 (H3 trimethylated at lysine 4) and H4K20me3 (H4 trimethylated at lysine 20) marks. While the recognition of H3K4me3 by Spindlin1 has been more examined [238-240], its interaction with H4K20me3 has been recently described and further studies are needed to investigate the biological relevance and functions correlated to it [241, 242]. Moreover, the binding of the H4K20me3 histone mark displays a weaker affinity than that of H3K4me3; thus, it has been postulated that the H4K20me3 mark may act as a secondary substrate for Spindlin1 [242].

Up to date, 15 crystal structures of Spindlin1 have been deposited in the PDB [19, 243], which comprise apo form crystal structure [244], as well as holo forms in complex with histone peptides [239, 240, 242, 245, 246] and inhibitors [247-249]. The latter structures have been resolved in the last three years. In **Figure 1.5** is depicted the crystal structure of Spindlin1 in complex with H3K4me3 (PDB ID: 4H75 [239]). The analysis of the crystal structures, together with mutagenesis studies of some residues of the first and second Tudor domains, clearly showed that the second domain is responsible for the recognition of the trimethylated lysine [239, 240, 242, 250]. In particular, a so-called aromatic cage formed by four aromatic residues (Phe141, Trp151, Tyr170, Tyr177) binds the trimethylated lysine through cation-pi interactions and van der Waals contacts (**Figure 1.5**). Meanwhile, the first domain can bind to asymmetrically dimethylated arginine residues (Rme2a) and to positive nitrogen moieties of bivalent inhibitors, which simultaneously bind to the first and second domains. Of note, the interaction with Rme2a displays different effects on the histone peptide affinity: it increases the affinity for H3K4me3 (H3K4me3R8me2a), whereas, decreases the binding affinity of H4K20me3 (H4K20me3R23me2a) [240, 242]. Very recently, another bivalent methylation pattern on H3 has been described, namely H3K4me3K9me3/2. The study reported that by binding to the first domain, the trimethylation and demethylation of lysine 9 enhances the histone affinity of H3K4me3 [251].

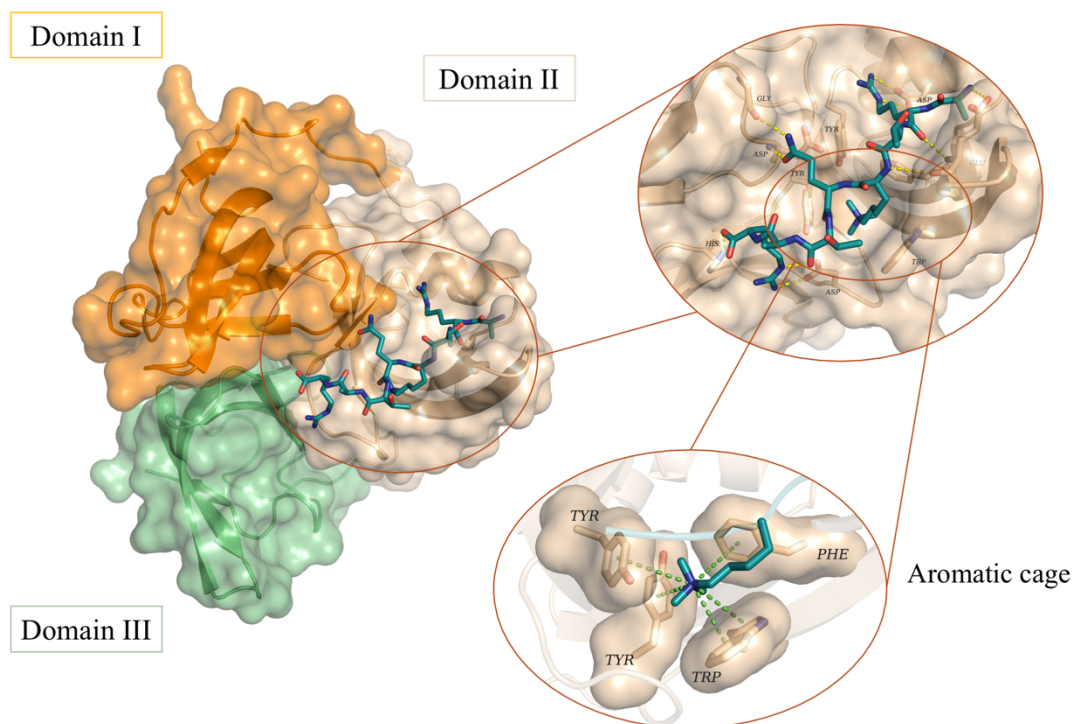


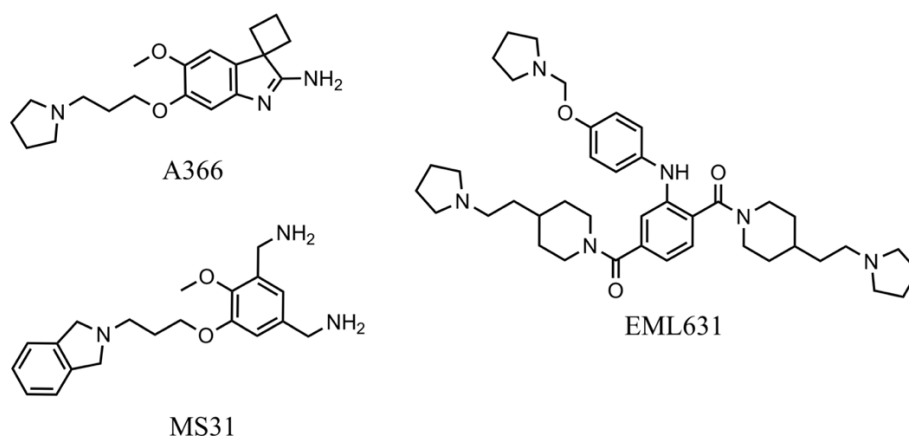
Figure 1.5. Crystal structure of Spindlin1 in complex with H3K4me3, PDB ID: 4H75. Only the side chains of the surrounding residues are shown for clarity, and they are displayed as sticks. Hydrogen bonds are represented as dashed yellow lines and cation-pi interactions as dashed green lines.

Spindlin1 has been related to several types of malignant tumors, including ovarian cancer, non-small-cell lung cancers, breast cancer and triple-negative breast cancer, liposarcoma and very recently to liver cancer [250, 252-256]. Likewise, a study established that Spindlin1, which is negatively regulated by the miR-148/152 family, enhances adriamycin-resistance, thus linking Spindlin1 to drug-resistant breast cancer [257]. Additionally, it has been reported that *spindlin1* may play a role in tumorigenesis, and was proposed to be a proto-oncogene [258-262]. Diverse signaling pathways have been associated with some of the Spindlin1 functions mentioned above, such as Wnt [240, 263], PI3K/Akt [253], RET [250], uL18-MDM2-p53 [261] pathways and SREBP1c-triggered FASN signaling [256]. Besides, Spindlin1 was found to control skeletal muscle development in mice and to be involved in the first meiotic division of mammalian oocytes. Accordingly, Spindlin1 is also potentially related to human skeletal muscle diseases and human infertility [264, 265].

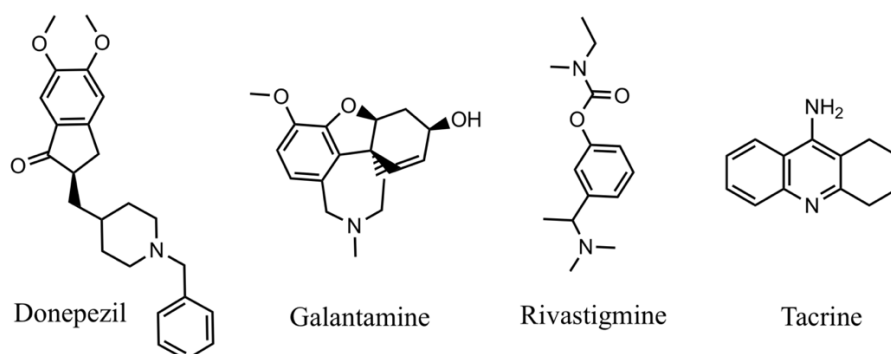
Owing to the associations of Spindlin1 to diverse pathological conditions, it can be considered as an interesting target for therapeutic purposes, which, however, needs to be further investigated. Consequently, a growing interest in this reader protein has arisen, which has led to the identification of several inhibitors in recent years. In 2016, **A366** –a

previously described G9a inhibitor (IC₅₀: 3.3 nM (AlphaLISA), [266])– was reported as the first Spindlin1 inhibitor discovered using a screening platform (IC₅₀: 186.3 nM (AlphaLISA), [267]). Applying the same approach in the screening of an epigenetic compound library, followed by optimization studies, other inhibitors have been reported in 2019, such as the fragment-like inhibitor **MS31** (IC₅₀ of 77 nM (AlphaLISA) and 243 nM (fluorescence polarization), [249]). Bae et al. presented the first study describing the development of bivalent inhibitors through protein microarrays [247]; as an example, compound **EML631** (K_d of 3 μM) is shown in **Figure 1.6** alongside with **A366** and **MS31**. Lately, starting from **A366** other bivalent inhibitors have been developed [248].

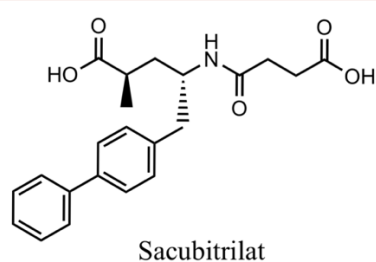
Spindlin1 inhibitors



AChE and BChE inhibitors



NEP inhibitor



MMPs inhibitor

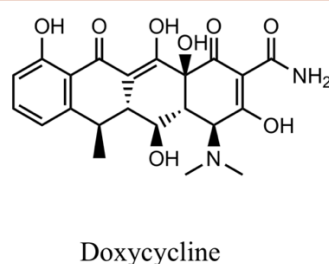


Figure 1.6. Chemical structures of the inhibitors mentioned in the text.

Our efforts in finding novel Spindlin1 inhibitors, in investigating a congeneric series of **A366** analogs, as well as in studying the binding pocket plasticity resulted from this Ph.D. thesis are reported in **chapter 3**.

1.2.2. Acetylcholinesterase and Butyrylcholinesterase

Acetylcholinesterase (AChE) and butyrylcholinesterase (BChE) are two esterase enzymes that belong to the cholinesterase family, and they are responsible for terminating the cholinergic signal transmission through the hydrolysis of choline esters [268]. Although the two enzymes have similar catalytic properties, they show different cellular and extracellular distributions [269]. AChE is highly selective for the hydrolysis of acetylcholine, and it is mainly found in chemical synapses of the central and peripheral nervous systems and in the membranes of erythrocytes [269, 270]. This renders AChE a key regulator of cholinergic neurotransmission. On the other hand, BChE is a nonspecific cholinesterase enzyme that hydrolyzes many different choline-based esters, as well as neuroactive peptides, and it is present in the blood plasma at high concentrations [269, 271]. Because BChE does not have an endogenous substrate, its physiological role is not totally clear unless for its well-recognized function as metabolizer of bioactive esters in the diet and drugs [272]. Recently, BChE has been found to have an important role in regulating the peptide hormone ghrelin in mice; a hormone also called “hunger hormone”, which affects not only appetite and weight gain but also emotional states, fear, anxiety and aggression [273]. It has also been proposed that BChE may act as a backup enzyme for AChE since it is a co-regulator of acetylcholine levels in the human brain [271, 274].

The correlation of AChE and BChE to human diseases is well known, as the dysregulation of cholinergic signaling has been linked to several pathological conditions, including myasthenia gravis, glaucoma and Alzheimer’s disease (AD) [275-277]. In the context of AD, two main events play a role in its onset and progression, which are the formation of amyloid β ($A\beta$) plaques and a decrease of cholinergic signaling, due to degeneration of cholinergic neurons, associated with low levels of acetylcholine [271, 278]. The cholinergic signaling is important in cognition, and consequently AD patients show cognitive, functional and behavioral symptoms. Therefore, the first therapeutic line for AD is to treat its symptoms by increasing the levels of acetylcholine through AChE inhibitors [271]. Numerous AChE inhibitors have been developed; donepezil, galantamine and rivastigmine are three FDA-approved drugs for the treatment of AD (**Figure 1.6**) – galantamine and rivastigmine are approved for mild-to-moderate stages, whereas donepezil for all stages of AD [271, 279, 280]. Donepezil and galantamine

compete with acetylcholine for the AChE binding, and they act as reversible inhibitors. Meanwhile, rivastigmine forms a covalent adduct in the AChE catalytic pocket, and it is classified as a “pseudo-irreversible” AChE inhibitor [279]. In addition to the cholinergic activity, various studies demonstrated that AChE inhibitors are also able to prevent A β oligomerization to some extent [281]. However, these inhibitors, as well as the other FDA-approved drugs for the treatment of AD (memantine, an N-methyl D-aspartate (NMDA) receptor antagonist prescribed to improve memory and attention; and a combination of donepezil and memantine), can only reduce and slow down the symptoms, but they cannot delay or stop its progression [271, 280].

Interestingly, donepezil and galantamine selectively inhibit AChE, while rivastigmine is a dual cholinesterase inhibitor that targets both AChE and BChE [282]. Several studies have supported the use of dual inhibitors in the treatment of neurological disorders [269, 282]; consequently, efforts have been made in the development of reversible dual AChE-BChE inhibitors [283-285]. Considering also that the drugs currently approved for AD have limited therapeutic efficacy and several side effects, the discovery of novel cholinesterase inhibitors is still an attractive area of research. Our contribution in this field is reported in **chapter 3.5**.

In **Figure 1.7** are illustrated the crystal structures of human AChE in complex with donepezil (PDB ID: 6O4W [286]) and human BChE with tacrine (PDB ID: 4BDS [287]). The latter inhibitor is a previously FDA-approved drug for the treatment of AD, which was withdrawn from use in 2013 because of concerns over safety [288]. The 2D structure of tacrine is reported in **Figure 1.6**. AChE and BChE share 65% sequence identity, and they show a similar tertiary structure [269, 271]. The binding pockets of both enzymes present a deep groove (gorge), composed of aromatic residues, that gives access to the active site and shape the pocket. At the bottom of the gorge is located the catalytic active site (CAS), which encompasses the conserved catalytic triad (His, Ser and Glu) and other residues that differ based on the cholinesterase and the species. At the top of the gorge, the peripheral anionic binding site (PAS) acts as the entrance and guides the substrate down [271]. BChE shows a larger PAS since it contains smaller amino acids than in AChE, e.g., Ala (A277 in hBChE) substitutes Trp (W286 hAChE and W297 in TcAChE).

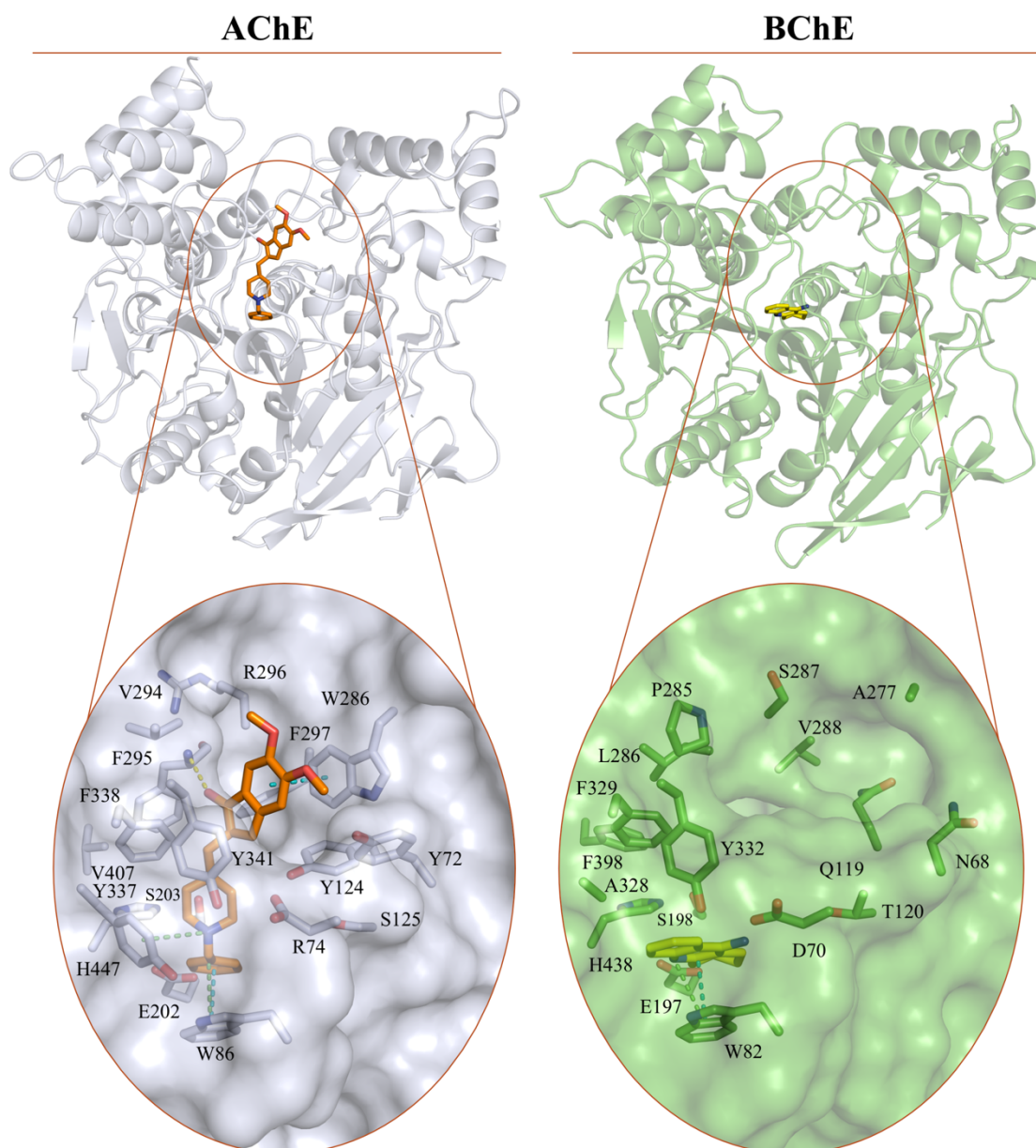


Figure 1.7. Crystal structures of human cholinesterases. AChE (white) and BChE (green) in complex with donepezil (orange sticks) and tacrine (yellow sticks), respectively. PDB IDs: 6O4W (AChE) and 4BDS (BChE). Only the side chains of the surrounding residues are shown for clarity and they are displayed as sticks; for residue F295 (AChE) the main chain is shown. Binding interactions are represented with dashed lines colored in green (cation- π), cyan (π - π stacking) and yellow (hydrogen bond).

1.2.3. Metalloproteinases: Neprilysin and Matrix metalloproteinase-14

Other targets that have been investigated in this thesis work are members of the metalloproteinase family, a big family of enzymes that uses metal ions for the catalytic mechanism and to which all protease enzymes belong. The metalloproteinases are further divided into two subgroups, exopeptidases (metalloexopeptidases) and endopeptidases

(metalloendopeptidases) [289]. Neprilysin and Matrix metalloproteinase-14 are both members of the endopeptidase subfamily.

1.2.3.1. Neprilysin

Neprilysin is also known as membrane metallo-endopeptidase (MME) or neutral endopeptidase (NEP). It is a type-II integral membrane zinc-containing endopeptidase, which is involved in the extracellular catabolism of numerous bioactive peptides, including enkephalins, substance P, endothelin, bradykinin and atrial natriuretic factor peptide [290, 291]. Due to its physiological implications, NEP has been a focus of interest in the field of analgesics, anti-hypertensive and heart failure drugs [291-293]. First attempts in the discovery of NEP inhibitors resulted in the development of diverse candidates for the treatment of hypertension and heart failure, the use of which in the long term showed that the initial reduction in blood pressure was not sustained [293]. This effect has been linked to another physiological role of NEP. Indeed, it has been found that NEP also breaks down the peptide angiotensin II. Thus, the inhibition of NEP alone also leads to increasing the levels of angiotensin II, which in turn counterbalances the actions of the natriuretic peptide [293]. Consequently, to overcome this limitation dual inhibitors have been proposed that simultaneously bind to NEP and angiotensin-converting enzyme. Another strategy, which has been successful and resulted in an FDA-approved drug (sacubitril/valsartan) for the treatment of heart failure, is the combination of NEP and ACE inhibitors [293]. In **Figure 1.6** is reported the NEP inhibitor sacubitrilat, the active form of the prodrug sacubitril, while the 3D structure of NEP in complex with sacubitrilat is represented in **Figure 1.9** (PDB ID: 5JMY, [294]).

Apart from the above-mentioned implications, NEP has also been found to degrade components of the elastic fiber system [295, 296] and it has been postulated that NEP may play a crucial role in wrinkle formation through the degradation of dermal elastic fibers [296, 297]. Moreover, the up-regulation and overexpression of NEP in mice exposed to UV and under skin aging has been associated with impairment of the elastic fiber network and subsequent loss of skin elasticity [297]. Our contribution in understanding the cleavage behavior of NEP on human skin elastin is reported in **chapter 3.6**.

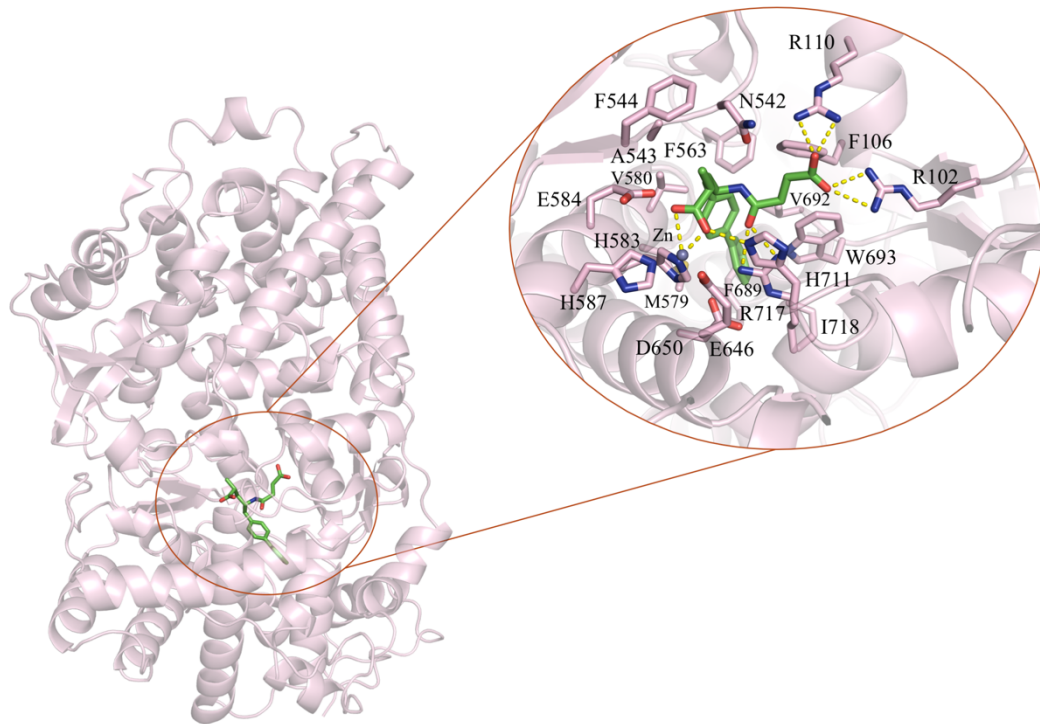


Figure 1.8. Crystal structure of NEP in complex with sacubitrilat (PDB ID: 5JMY). Only the side chains of the surrounding residues are shown for clarity and they are displayed as light pink sticks, whereas the inhibitor is colored in green. The zinc ion is shown as a grey sphere, and the hydrogen bonds are represented as dashed yellow lines.

1.2.3.2. Matrix metalloproteinase-14

Matrix metalloproteinases (MMPs) are also known as matrix metalloproteinases or matrixins, and they are a large group of zinc- and calcium-dependent endopeptidases that regulate several connective tissue remodeling processes [298]. For instance, owing to MMPs' ability to degrade extracellular matrix proteins, including collagens, proteoglycans, fibronectin, laminin and elastin, MMPs are known to be involved in embryonic development, pregnancy, growth and wound healing [299]. Furthermore, MMPs have been associated with processes like vascular remodeling and angiogenesis [300].

In physiological conditions, the activity of MMPs is further controlled by endogenous specific tissue inhibitors, namely TIMPs (tissue inhibitors of metalloproteinases) [301]. As an example, the crystal structure of MMP-14 in complex with TIMP-1 is depicted in **Figure 1.9** (PDB ID: 3MA2, [302]). The dysregulation of the MMPs-TIMPs balance has been linked to different pathological conditions such as pulmonary emphysema, rheumatoid arthritis, atherosclerosis, myocardial infarction, as well as tumor growth and metastasis [301, 303]. Additionally, the overexpression of MMPs has also been related to

diseases like periodontitis, hepatitis, autoimmune disorders of the skin and dermal photoaging [304, 305]. Due to their implication in pathological disorders, MMPs are considered as potential drug targets [305-307]. Thus, efforts have been made in the discovery of MMPs inhibitors, which, unfortunately, have not led to encouraging outcomes as most of the clinical trials highlighted severe side effects of MMPs inhibitors related to their poor selectivity for a specific MMP [308, 309]. Nevertheless, one inhibitor has been approved by the FDA and it is on the market for the treatment of periodontal disease [308]; namely the known antibiotic doxycycline (**Figure 1.6**).

Among the 28 MMPs known so far, our interest was set on MMP-14 in order to characterize its cleavage behavior on human skin elastin and to compare it with other MMPs that are known to cleave elastin, like MMP-7, MMP-9 and MMP-12. Indeed, even though it has been suggested that MMP-14 plays a central role in macrophage-mediated elastin degradation [310], its association with human elastin was not yet been described, unlike for other MMPs [311-313]. A detailed comparison of the pockets of the MMPs under our analysis, as well as our findings, is reported in **chapter 3.7**.

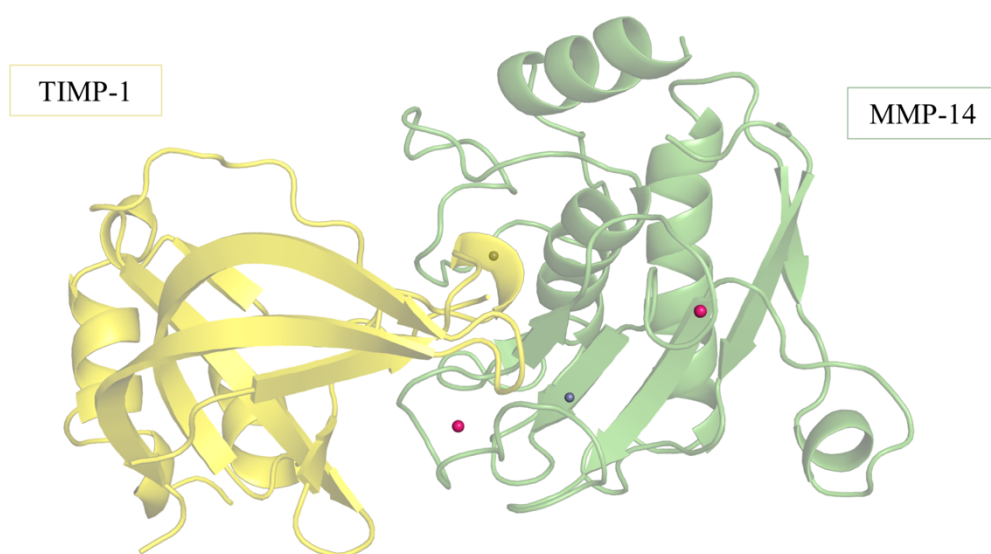


Figure 1.9. Crystal structure of MMP-14 (green) in complex with TIMP-1 (yellow). The metal ions are shown as spheres, respectively in grey (zinc) and hot pink (calcium).

2. AIM OF THE WORK

The studies described in this thesis aim to apply computational approaches to drug discovery and design projects to identify new inhibitors, rationalize their activity, get insight into protein flexibility, and investigate the binding mode of small molecules and peptides. A special focus is given to proteins with flexible binding sites for which traditional computational approaches often fail. Different computational techniques are evaluated based on the project and related challenges, including structure-based pharmacophore modeling, VS, docking (classical rigid-body docking and flexible docking), MD simulations and MM-GBSA calculations.

To search for new Spindlin1 inhibitors, an iterative VS campaign encompassing structure- and ligand-based approaches is conducted, followed by lead optimization studies (**chapter 3.1** and **chapter 3.2**). Knowing the potential flexibility of the aromatic cage, the focus is set to study the conformational changes of the pocket by structural analysis and MD simulations. Another aim is to establish a relatively fast protocol able to rightly predict the binding mode of small molecule ligands in flexible binding pockets (**chapter 3.3**). Lastly, in an effort to improve the activity and explore the SAR of the Spindlin1 inhibitor **A366**, a series of 21 derivatives is investigated by means of molecular modeling tools like docking, MD simulations and MM-GBSA calculations (**chapter 3.4**).

In **chapter 3.5**, computational methods are implemented to support the discovery of dual inhibitors of AChE and BChE as well as inhibitors showing a preferential inhibition of AChE over BChE. In this project, docking and MM-GBSA are applied to predict the binding mode, rationalize the biological activity and identify the structural features that affect the activity of our series.

Finally, computer-based approaches are applied in **chapter 3.6** and **chapter 3.7** to explore and rationalize from a structural point of view the cleavage behavior of NEP, MMP-14 and other MMPs on specific human elastin fibers. The binding mode of the peptides is investigated, and an analysis and comparison of the MMPs pockets under investigation is performed.

Overall, the global aim is to integrate CADD methods in the academic pipeline; hence, successful examples of collaboration among different research fields are shown.

Page left intentionally blank.

3. RESULTS

In this chapter are presented the abstracts of seven manuscripts relevant to this thesis work. The full texts with the relative supporting information, and a book chapter that covers lysine reader proteins, are given in the **Appendices**.

3.1. Identification and Structure–Activity Relationship Studies of Small-Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1

D. Robaa, T. Wagner, **C. Luise**, L. Carlino, J. McMillan, R. Flaig, R. Schüle,
M. Jung, W. Sippl

ChemMedChem, 11, 2327-2338, 2016

DOI: 10.1002/cmdc.201600362

The methyllysine reader protein Spindlin1 has been implicated in the tumorigenesis of several types of cancer and may be an attractive novel therapeutic target. Small-molecule inhibitors of Spindlin1 should be valuable as chemical probes as well as potential new therapeutics. We applied an iterative virtual screening campaign, encompassing structure- and ligand-based approaches, to identify potential Spindlin1 inhibitors from data-bases of commercially available compounds. Our *in silico* studies coupled with *in vitro* testing were successful in identifying

novel Spindlin1 inhibitors. Several 4-aminoquinazoline and quinazolinethione derivatives were among the active hit compounds, which indicated that these scaffolds represent promising lead structures for the development of Spindlin1 inhibitors. Subsequent lead optimization studies were hence carried out, and numerous derivatives of both lead scaffolds were synthesized. This resulted in the discovery of novel inhibitors of Spindlin1 and helped explore the structure–activity relationships

3.2. Application of Virtual Screening Approaches for the Identification of Small Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1

C. Luise and D. Robaa

Rational Drug Design. Methods in Molecular Biology, 1824, 347-370, 2018

DOI: 10.1007/978-1-4939-8630-9_21

Computer-based approaches represent a powerful tool which helps to identify and optimize lead structures in the process of drug discovery. Computer-aided drug design techniques (CADD) encompass a large variety of methods which are subdivided into structure-based (SBDD) and ligand-based drug design (LBDD) methods. Several approaches have been successfully used over the last three decades in different fields. Indeed also in the field of epigenetics, virtual screening (VS) studies and structure-based approaches have been applied to identify novel chemical modulators of epigenetic targets as well as to predict the binding mode of active ligands and to study the protein dynamics. In this chapter, an iterative VS approach using both SBDD and LBDD methods, which was successful in identifying Spindlin1 inhibitors, will be described. All protocol steps, starting from structure-based pharmacophore modeling, protein and database preparation along with docking and similarity search, will be explained in details.

3.3. Exploring aromatic cage flexibility of the histone methyllysine reader protein Spindlin1 and its impact on binding mode prediction: an in silico study

C. Luise, D. Robaa, W. Sippl

Journal of Computer-Aided Molecular Design, 35, 695-706, 2021

DOI: 10.1007/s10822-021-00391-9

Some of the main challenges faced in drug discovery are pocket flexibility and binding mode prediction. In this work, we explored the aromatic cage flexibility of the histone methyllysine reader protein Spindlin1 and its impact on binding mode prediction by means of in silico approaches. We first investigated the Spindlin1 aromatic cage plasticity by analyzing the available crystal structures and through molecular dynamic simulations. Then we assessed the ability of rigid docking and flexible docking to rightly reproduce the binding mode of a known ligand into Spindlin1, as an example of a reader protein displaying flexibility in the binding pocket. The ability of induced fit docking was further probed to test if the right ligand binding mode could be obtained through flexible docking regardless of the initial protein conformation. Finally, the stability of generated docking poses was verified by molecular dynamic simulations. Accurate binding mode prediction was obtained showing that the herein reported approach is a highly promising combination of in silico methods able to rightly predict the binding mode of small molecule ligands in flexible binding pockets, such as those observed in some reader proteins.

3.4. Structure-Based Design, Docking and Binding Free Energy Calculations of A366 Derivatives as Spindlin1 Inhibitor

C. Luise, D. Robaa, P. Regenass, D. Maurer, D. Ostrovskiy, L. Seifert, J. Bacher, T. Burgahn, T. Wagner, J. Seitz, H. Greschik, K.S. Park, Y. Xiong, J. Jin, R. Schüle, B. Breit, M. Jung, W. Sippl
International Journal of Molecular Sciences, 22, 5910, 2021
DOI: 10.3390/ijms22115910

The chromatin reader protein Spindlin1 plays an important role in epigenetic regulation, through which it has been linked to several types of malignant tumors. In the current work, we report on the development of novel analogs of the previously published lead inhibitor **A366**. In an effort to improve the activity and explore the structure–activity relationship (SAR), a series of 21 derivatives was synthesized, tested in vitro, and investigated by means of molecular modeling tools. Docking studies and molecular dynamics (MD) simulations were performed to analyze and rationalize the structural differences responsible for the Spindlin1 activity. The analysis of MD simulations shed light on the important interactions. Our study highlighted the main structural features that are required for Spindlin1 inhibitory activity, which include a positively charged pyrrolidine moiety embedded into the aromatic cage connected via a propyloxy linker to the 2-aminoindole core. Of the latter, the amidine group anchor the compounds into the pocket through salt bridge interactions with Asp184. Different protocols were tested to identify a fast in silico method that could help to discriminate between active and inactive compounds within the **A366** series. Rescoring the docking poses with MM-GBSA calculations was successful in this regard. Because **A366** is known to be a G9a inhibitor, the most active developed Spindlin1 inhibitors were also tested over G9a and GLP to verify the selectivity profile of the **A366** analogs. This resulted in the discovery of diverse selective compounds, among which **1s** and **1t** showed Spindlin1 activity in the nanomolar range and selectivity over G9a and GLP. Finally, future design hypotheses were suggested based on our findings.

3.5. Synthesis and biological evaluation of 8-hydroxy- 2,7-naphthyridin-2-ium salts as novel inhibitors of acetylcholinesterase (AChE) and butyrylcholinesterase (BChE)

M. Schiedel, A. Fallarero, C. Luise, W. Sippl, P. Vuorelab and M. Jung

Med. Chem. Commun., 8, 465-470, 2017

DOI: 10.1039/c6md00647g

By analogy with the natural product chelerythrine, which has been identified as an inhibitor of both acetylcholinesterase (AChE) and butyrylcholinesterase (BChE), we prepared a small series of 8-hydroxy-2,7-naphthyridin-2-ium salts. Spectroscopic analyses allowed us to elucidate the zwitterionic nature of 2,7-naphthyridin-1(1H)-ones, the neutral state of 8-hydroxy-2,7-naphthyridin-2-ium salts. Among the tested compounds, we identified dual inhibitors of AChE and BChE as well as an inhibitor showing a preferential inhibition of AChE over BChE. By in vitro characterization in combination with docking studies, we were able to identify structural features that influence the biological activity of 8-hydroxy-2,7-naphthyridin-2-ium salts.

3.6. Degradation of tropoelastin and skin elastin by neprilysin

A.C.M. Huertas, C.E.H. Schmelzer, **C. Luise**, W. Sippl, M. Pietzsch,

W. Hoehenwarter, A. Heinz

Biochimie, 146, 73-78, 2018

DOI: 10.1016/j.biochi.2017.11.018

Neprilysin is also known as skin fibroblast-derived elastase, and its up-regulation during aging is associated with impairments of the elastic fiber network, loss of skin elasticity and wrinkle formation. However, information on its elastase activity is still limited. The aim of this study was to investigate the degradation of fibrillar skin elastin by neprilysin and the influence of the donor's age on the degradation process using mass spectrometry and bioinformatics approaches. The results showed that cleavage by neprilysin is dependent on previous damage of elastin. While neprilysin does not cleave young and intact skin elastin well, it degrades elastin fibers from older donors, which may further promote aging processes. With regards to the cleavage behavior of neprilysin, a strong preference for Gly at P1 was found, while Gly, Ala and Val were well accepted at P10 upon cleavage of tropoelastin and skin elastin. The results of the study indicate that the progressive release of bioactive elastin peptides by neprilysin upon skin aging may enhance local tissue damage and accelerate extracellular matrix aging processes.

3.7. MMP-14 degrades tropoelastin and elastin

N. Miekus, **C. Luise**, W. Sippl, T. Baczek, C.E.H. Schmelzer, A. Heinz

Biochimie, 165, 32-39, 2019

DOI: 1016/j.biochi.2019.07.001

Matrix metalloproteinases are a class of enzymes, which degrade extracellular matrix components such as collagens, elastin, laminin or fibronectin. So far, four matrix metalloproteinases have been shown to degrade elastin and its precursor tropoelastin, namely matrix metalloproteinase-2, -7, -9 and -12. This study focuses on investigating the elastinolytic capability of membrane-type 1 matrix metalloproteinase, also known as matrix metalloproteinase-14. We digested recombinant human tropoelastin and human skin elastin with matrix metalloproteinase-14 and analyzed the peptide mixtures using complementary mass spectrometric techniques and bioinformatics tools. The results and additional molecular docking studies show that matrix metalloproteinase-14 cleaves tropoelastin as well as elastin. While tropoelastin was well degraded, fewer cleavages occurred in the highly cross-linked mature elastin. The study also provides insights into the cleavage preferences of the enzyme. Similar to cleavage preferences of matrix metalloproteinases-2, -7, -9 and -12, matrix metalloproteinase-14 prefers small and medium-sized hydrophobic residues including Gly, Ala, Leu and Val at cleavage site P1'. Pro, Gly and Ala were preferably found at P1-P4 and P20-P40 in both tropoelastin and elastin. Cleavage of mature skin elastin by matrix metalloproteinase-14 released a variety of bioactive elastin peptides, which indicates that the enzyme may play a role in the development and progression of cardiovascular diseases that go along with elastin breakdown.

Page left intentionally blank.

4. SUMMARY AND CONCLUSIONS

In chapter 3, the results of our efforts to apply computer-based methods in drug discovery projects are given in details. Several approaches were employed to identify and study small-molecule inhibitors and peptides. In this chapter, a summary of the results with an emphasis on the techniques used, the main achievements as well as the related challenges and limitations is given. Furthermore, general conclusions and perspectives are provided.

4.1. Application of computer-based methods to search and investigate Spindlin1 inhibitors

To search for new Spindlin1 inhibitors, computer-based methods such as structure-based pharmacophore modelling, VS, rigid-body and flexible docking, MD simulations and MM-GBSA calculations were applied.

Firstly, an iterative VS campaign, encompassing structure- and ligand-based approaches, was conducted based on the peptide binding mode and then on the first validated hits. Subsequently, lead optimization studies were carried out. This led to the synthesis of numerous derivatives in order to obtain compounds with improved inhibitory activities and to deduce the SAR of these new series of Spindlin1 inhibitors (**chapter 3.1** and **chapter 3.2**). Furthermore, knowing that the aromatic cage of the reader proteins can undergo conformational changes, special attention was given to the structural flexibility. Hence, MD simulations and induced fit docking were performed to gain insight into the binding site flexibility and binding mode prediction (**chapter 3.3**). Finally, with the availability of a co-crystal structure of Spindlin1 in complex with **A366**, the binding mode of an in-house congeneric series of **A366** derivatives was investigated via MD simulations, docking and MM-GBSA calculations (**chapter 3.4**).

4.1.1. Identification of novel Spindlin1 inhibitors through an iterative VS campaign followed by optimization studies

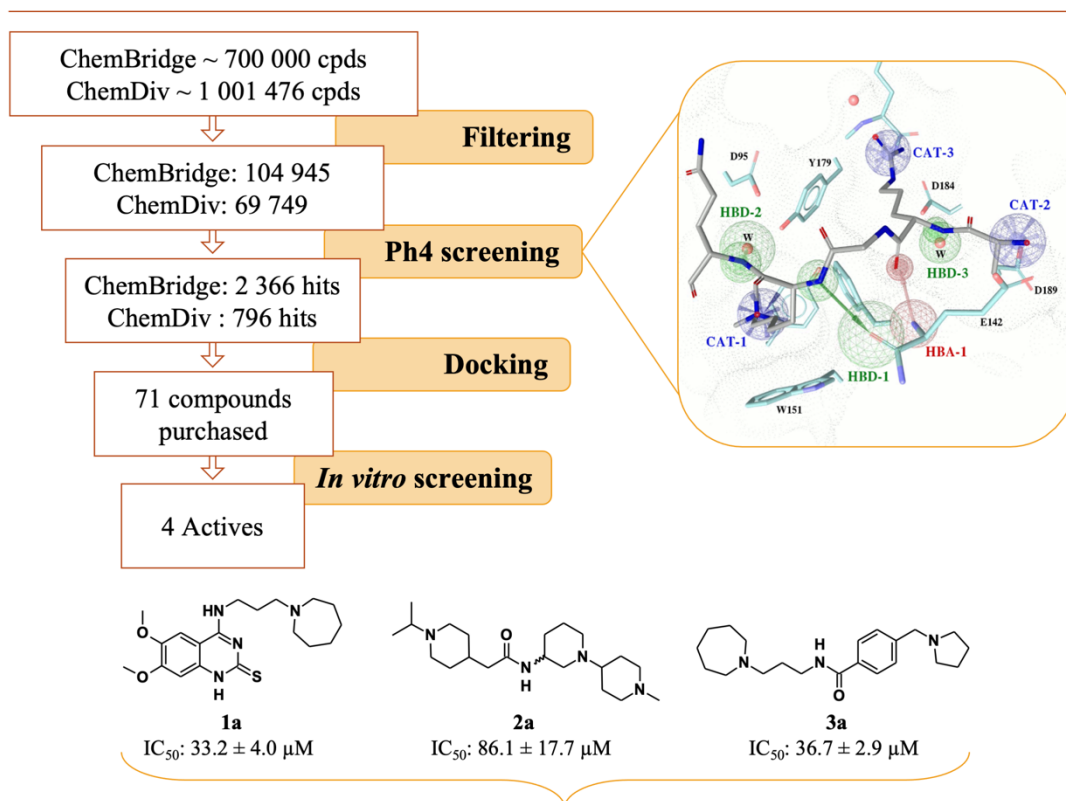
The main objective of this part of the work was to discover novel Spindlin1 inhibitors from databases of commercially available compounds through *in silico* techniques (**chapter 3.1**) and to provide insights and all protocol steps of a VS campaign (**chapter 3.2**).

At the beginning of the study, no Spindlin1 inhibitors were reported, and only crystal structures in apo form and holo forms in complex with histone peptides were deposited in the PDB. The main challenge of this project was to identify small molecule inhibitors for a relatively new target having little information about the important features responsible for the binding and activity. To overcome this, an iterative VS campaign, including structure- and ligand-based approaches coupled with *in vitro* testing, was performed (**Figure 4.1**).

First, a pharmacophore screening based on the experimentally determined peptide binding mode was carried out. The crystal structure of Spindlin1 in complex with H3K4me3 peptide (PDB ID: 4H75) was retrieved from the PDB and pharmacophoric features describing the protein-peptide interactions were assigned. Six different pharmacophore models were generated and used to screen databases of commercially available compounds, namely ChemBridge and ChemDiv. One feature (CAT-1, **Figure 4.1**), representing the cation- π interactions of the trimethylated lysine moiety with the aromatic residues of the cage (**Figure 1.5**), was considered essential and kept in all pharmacophore models. The other features were, instead, alternatively combined. To decrease the number of compounds to be screened, the databases were previously filtered by accounting for the molecular weight and the presence of structural features. Thus, compounds with a molecular weight greater than 500 Da and compounds containing no positively charged nitrogen atoms (N^+) or H-bond donors were removed.

The virtual hits obtained from the pharmacophore screening were subsequently docked to the crystal structure. The docking results were analyzed and only the top-scored compounds showing a protonated nitrogen rightly positioned in the aromatic cage were considered. The most promising compounds were purchased and subjected to *in vitro* screening. With this approach, we were able to identify four compounds that showed *in vitro* inhibition of Spindlin1 using an AlphaLISA assay established by our collaborators (**Figure 4.1**).

Structure-based VS



Ligand-based VS

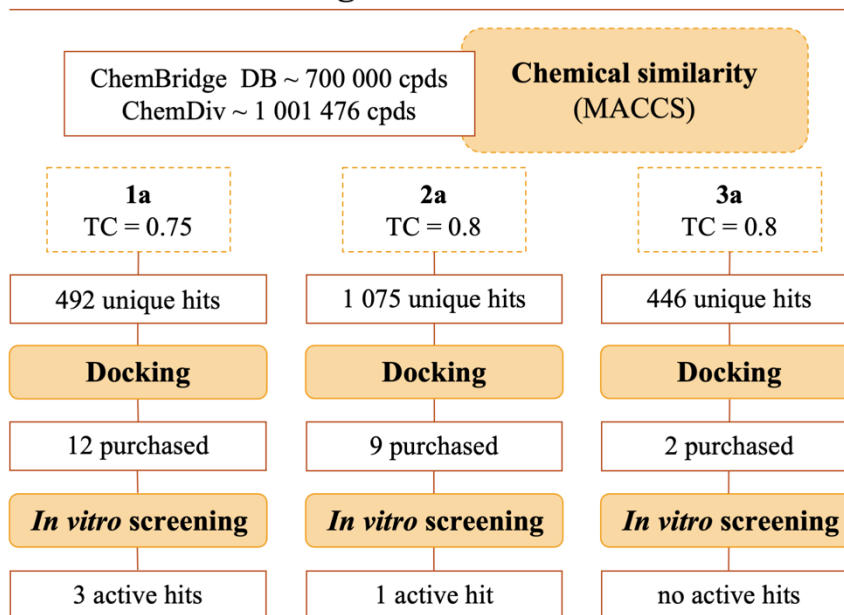


Figure 4.1. Workflow of the VS campaign. Spindlin1 crystal structure in complex with H3K4Me3 (PDB ID: 4H75) was used for pharmacophore models generation and docking. Top-right figure: all pharmacophoric features; blue spheres: cationic feature (CAT), green spheres and green arrow: H-bond donor feature (HBD), red spheres and red arrow: H-bond acceptor feature (HBA). The N-terminal residues of H3K4me3 are displayed as grey sticks, the binding pocket residues as cyan sticks, and water molecules as red spheres. TC: Tanimoto Coefficient. MACCS: Molecular ACCESS System.

Secondly, to obtain further derivatives and/or structural analogs of the active hits, a 2D-chemical similarity search using the identified hits as queries was executed. Again, the retrieved hits were consequently docked into the Spindlin1 crystal structure and a total of 23 compounds were purchased and tested. This resulted in uncovering four additional active hits (**Figure 4.1**).

Finally, lead optimization studies of two scaffolds (4-aminoquinazolinethiones and thioquinazolinones) were carried out, and numerous derivatives were synthesized, which allowed to expand the SAR of the series and attain compounds with improved inhibitory activities. Altogether, the obtained data highlighted that the aliphatic basic group acts as a lysine-mimetic, and its nature has a significant effect on the activity. Thus, the choice of an appropriate lysine-mimetic group is crucial for achieving Spindlin1 inhibition. Nonetheless, additional modifications are needed to further improve the activity, as shown by the increase in potency of the compounds bearing a benzyloxy substituent at position 7 of the 4-aminoquinazolinethione and thioquinazolinone scaffolds. As an example, the predicted binding mode of two of the most active compounds (**1k** and **4q**, IC_{50} values of 3.5 and 4.5 μ M, respectively) are illustrated in **Figure 4.2**.

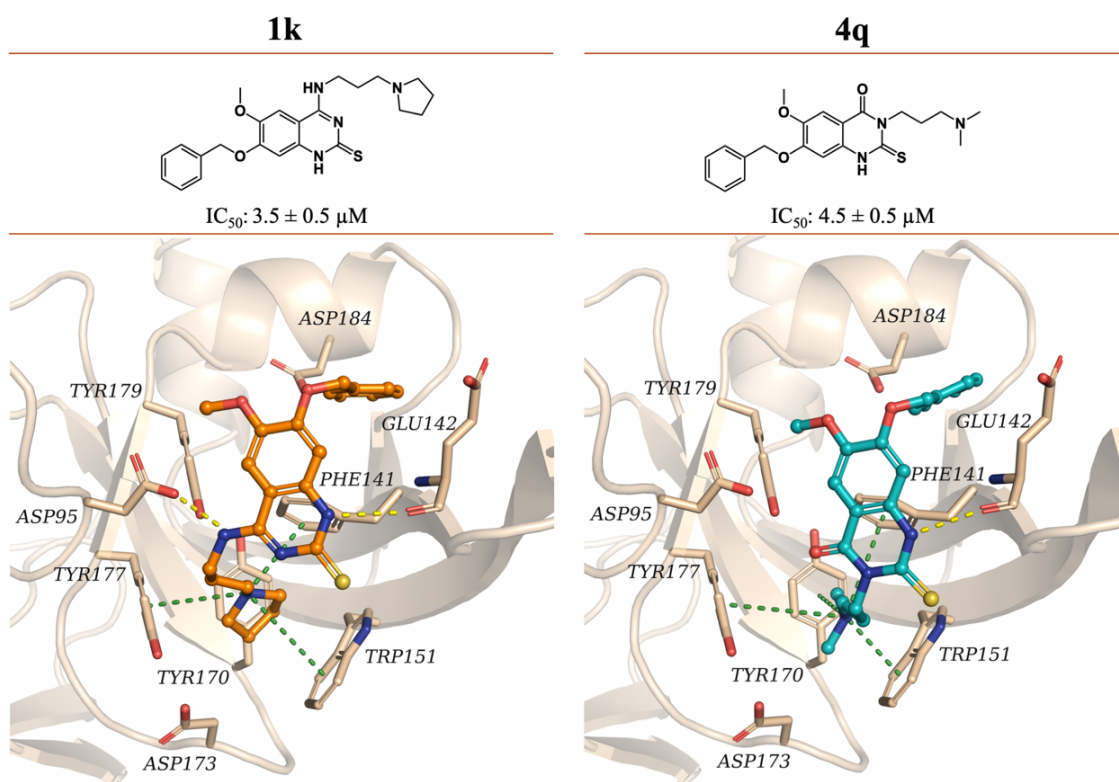


Figure 4.2. 2D chemical structures and predicted binding modes of compounds **1k** and **4q** in the Spindlin1 second domain binding site (PDB ID: 4H75). Docking poses are shown as orange (**1k**) and cyan (**4q**) balls and sticks, while the side chains of the surrounding residues are depicted as beige sticks. Hydrogen bond interactions are represented with yellow dashed lines while cation-pi interactions as dark green dashed lines.

In conclusion, the *in silico* studies coupled with *in vitro* testing applied in this work were successful in identifying the first set of novel Spindlin1 inhibitors active in the micromolar range. Meanwhile, structural modifications of the VS hits allowed four- to six-fold improvement in the inhibitory activity yielding six Spindlin1 inhibitors active in the low micromolar range.

4.1.1.1. Limitations and conclusions

As discussed in detail in the introduction (subparagraph 1.1.4.5), VS campaigns have some limitations. Generally, only hits active in the micromolar range are retrieved, and optimization studies are needed to improve the activity. This is partially due to the type of approaches used at this stage, mainly fast approaches like pharmacophore screening, similarity search and docking that do not take into account important contributions for the binding. Moreover, the nature of the compounds present in the database also plays a key role. For well-known targets, tailored and enriched databases can be used, which will facilitate the finding of more active compounds. However, for new targets with unknown inhibitors, the primary goal of VS and HTS is to find new hits/lead structures, which will be further optimized.

In this project, we encountered additional challenges related to the little knowledge available for the target at the beginning of the project. Indeed, no Spindlin1 inhibitors and crystal structures bound to inhibitors were reported, and the potential flexibility of the pocket upon the binding of ligands was still unknown. Nevertheless, micromolar hits could be found which were optimized into more active compounds.

Parallel to this work, our colleagues performed a focused screening of an in-house database containing known epigenetic probes and discovered that **A366**, a potent inhibitor of the histone methyltransferase G9a, is also a Spindlin1 nanomolar inhibitor. Hence, our subsequent efforts focused on **A366** derivatives (section 4.1.3).

4.1.2. *In silico* studies to explore the aromatic cage flexibility and its impact on binding mode prediction

This part of the work focused on investigating the pocket flexibility of Spindlin1 by molecular modeling methodologies (chapter 3.3). After the release of diverse Spindlin1 crystal structures in complex with inhibitors, it became clear that the aromatic cage responsible for the binding of the trimethylated lysine and mimetic moieties can undergo conformational changes. Because subtle changes in the side chain conformation can impact the binding mode prediction and related *in silico* studies, we set to explore the

pocket flexibility and to assess the ability of various computational methods to rightly reproduce the binding mode of a known Spindlin1 inhibitor.

To investigate the conformational plasticity of the binding pocket, we first analyzed the available Spindlin1 crystal structures deposited in the Protein Data Bank focusing on the aromatic cage residues (Phe141, Trp151, Tyr170, Tyr177). The analysis underlined that, among those residues, Phe141 and Trp151 display higher temperature factor values, and hence a higher degree of flexibility. Of note is the side chain of Phe141, as it can adopt two different orientations that lead to two diverse shapes of the aromatic cage, which we defined as closed cage and open cage (**Figure 4.3**). The open conformation is present in all the X-ray ligand-bound forms but one, indicating that the flip of the Phe141 side chain is induced by the ligands.

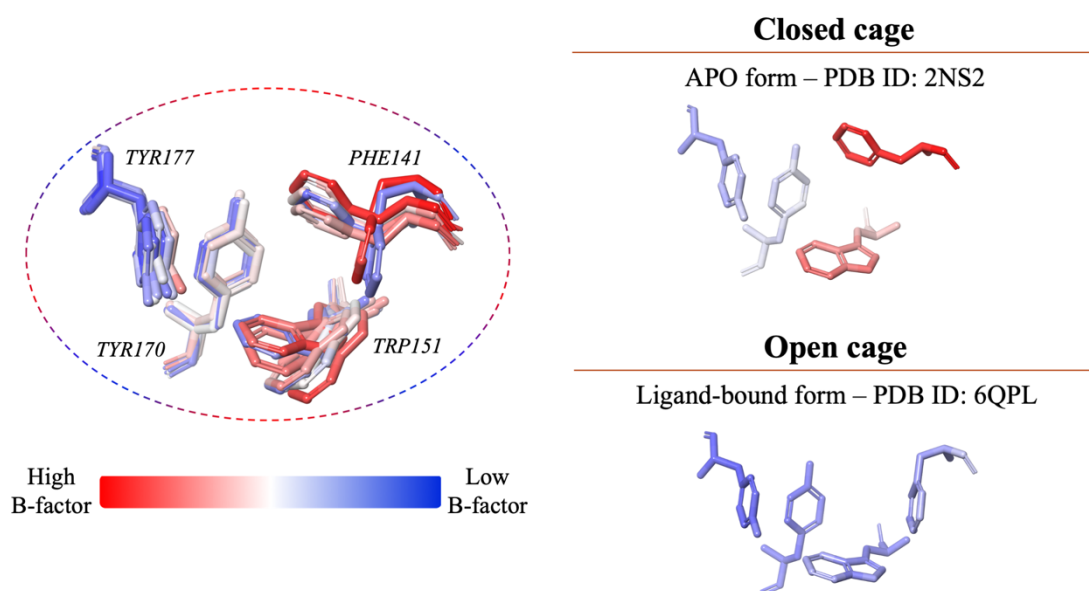


Figure 4.3. Analysis of the Spindlin1 aromatic cage. Left side: superimposition of the investigated crystal structures (PDB IDs: 2NS2, 4H75, 4MZF, 5Y5W, 5JSG, 5JSJ, 6QPL, 6I8Y). Right side: focus on the cage conformation. The protein residues are colored according to the PDB B-factor values and are shown as sticks.

Subsequently, we performed MD simulations on the apo protein to probe whether it is possible to obtain the open cage starting from the closed cage conformation and to further assess the flexibility of the second domain binding site. The results of the MD simulations were analyzed by several means: i) RMSD plots of the backbone atoms; ii) RMSF values of the heavy atoms of the amino acids that constitute the pocket; iii) occupancy of the hydrogen bond networking established by the latter residues during the simulation time; iv) clustering of the trajectories based on the RMSD of Ph141. From our study we could conclude that some residues are relatively steady (His139, Tyr170,

Tyr177, Tyr179), whereas others show flexibility during the simulation (e.g., Phe141, Trp151). Hydrogen bonds among the steady residues, which are maintained during simulation, play a role in stabilizing their conformations. Contrariwise, Phe141 lacks the possibility to form hydrogen bonds with its side chain and displays plasticity and diverse conformations. On the other hand, Phe141 interacts via pi-pi stacking with Trp151, leading the side chains of both residues to be closer and move together during the simulations. Consequently, the MD simulations could not yield an open cage conformation suitable for the ligand binding. Indeed, only closed cage or disorganized cage conformations could be obtained. We attributed the failure of the MD simulations to produce the open cage to the hydrophobic nature of the pocket: in the absence of any ligand, Phe141 is driven by pi-pi stacking interactions towards Trp151, which results in the aromatic cage being mainly in the closed conformation.

In order to test not only if flexible docking could generate the open cage but also how the methodology performs using different protein conformations as starting points, we then investigated the ability of induced fit docking (IFD) to correctly reproduce the experimentally determined X-ray binding mode of **A366** (PDB ID: 6I8Y). At the same time, we also carried out **rigid-body docking** (which refers to classical docking – protein kept rigid in its original conformation, ligand treated as flexible; subparagraph 1.1.1.1) to emphasize that, in cases where the pocket can exhibit flexibility upon ligand binding, treating the protein as a rigid entity can be a limiting factor if the right protein conformation is not available. Thus, open and closed conformations were used for the docking studies. Additionally, diverse IFD protocols were examined by changing the number of the residues treated as flexible with the aim to establish a relatively fast and efficient protocol. The selection of the residues was guided by the crystal structures analysis and MD simulations results.

As expected, the experimentally determined binding mode of **A366** could not be reproduced through rigid-body docking into the closed cage conformation (**Figure 4.4**). Whereas IFD could nicely reproduce the X-ray binding mode of **A366** starting both from open and closed pockets, with RMSD values in the range of 0.48-1.45 Å. In fact, when using the closed cage as protein input, **A366** induced the flip of Phe141, and IFD methodology proved to be able to generate an open cage with the correct ligand binding mode (**Figure 4.4**).

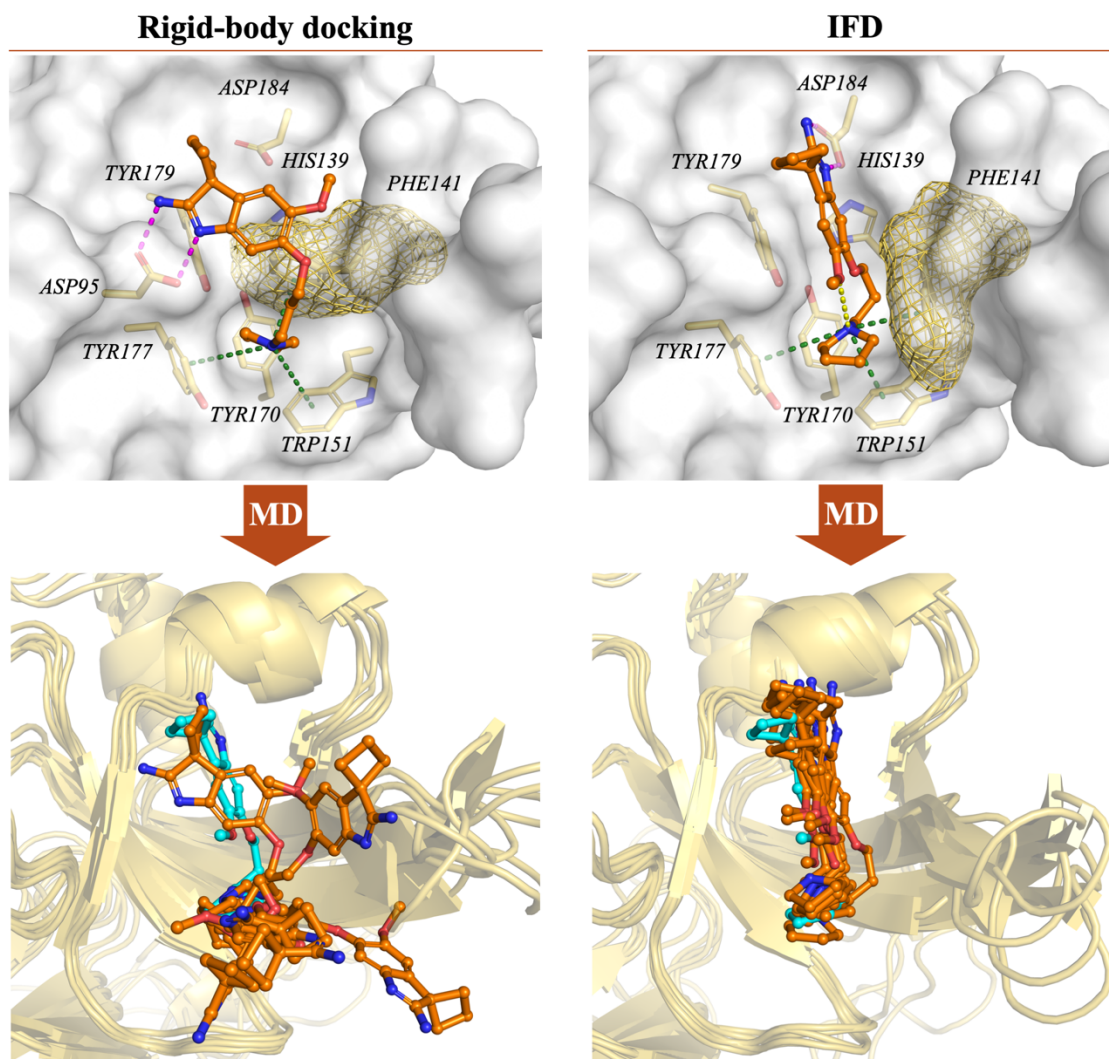


Figure 4.4. The upper panel shows the predicted binding modes of **A366** generated through rigid-body docking and IFD in the second domain of Spindlin1 (PDB ID: 2NS2 – closed cage). The IFD pose reported refers to the protocol with three residues were treated as flexible (Phe141, Trp151, Asp184). The docking poses are displayed as orange balls and sticks, binding pocket residues as yellow sticks, and the protein as white surface. Phe141 is also illustrated as yellow mesh surface. Binding interactions are represented as dashed lines colored in magenta (salt bridge), yellow (hydrogen bonds) and dark green (cation-pi interactions). The lower panel shows the relative binding mode of **A366** (orange balls and sticks) observed during the MD simulations at 0 ns, 10 ns, 20 ns, 30 ns, 40 ns and 50 ns superimposed with the experimentally determined **A366** X-ray ligand conformation (cyan ball and stick, PDB ID: 6I8Y). The protein is displayed as yellow cartoon.

To evaluate the stability of the attained binding mode, we subsequently subjected the obtained docking poses-complexes to short MD simulations (50 ns). The study highlighted that the binding mode obtained into the closed aromatic cage through rigid-body docking is highly unstable and generates diverse binding modes (**Figure 4.4**). Since Phe141 does not flip during the simulation, the experimentally determined pose of **A366** is not reproduced. Assuming that the **A366** X-ray conformation would not have been available for comparison, this binding mode hypothesis could have been discarded based

on the MD results. Meanwhile the binding mode predicted into the closed cage by IFD is highly stable during the MD simulation, with the initial pose being maintained throughout the simulation time – a pose that perfectly reproduces the experimentally observed **A366** X-ray conformation (**Figure 4.4**). These results are good examples of how MD simulations can be beneficial to verify the predicted binding modes by analysing their stability.

To sum up, we outlined how pocket flexibility can play a role in binding mode prediction and how diverse *in silico* methodologies perform in sampling pocket conformations and pose prediction. We highlighted that previous pocket analysis could shed light on the residues to treat as flexible during IFD, allowing faster and accurate results. The combination of IFD followed by short MD simulation reported in this work is a highly promising approach to rightly predict the binding mode of small molecule ligands in flexible binding pockets, such as observed in Spindlin1.

4.1.2.1. Limitations

One limitation encountered in this study is related to the MD simulation of the apo protein. Our work highlighted that for such a protein where stable pi-pi interactions between side chains are formed, it is challenging to obtain clusters that sample suitable open cage conformations. Even if the open cage is present during the simulation time, the event is so negligible that it is not perceived in the analysis. On the other hand, the MD simulations of the apo protein gave confidence about the residues to treat as flexible during IFD. Likewise, the MD simulations of the docked poses were helpful to evaluate the stability and reliability of the predicted binding mode hypotheses.

IFD proved to be a viable and relatively fast approach to observe the movement of side chains and the consequent modification of the shape of the pocket induced by the ligand. However, as fast as IFD might be, it can be performed on a small number of compounds, but it is not feasible on a large scale – e.g., databases used in VS. In fact, VS methods are meant to be fast and using IFD would be time-consuming.

4.1.3. Molecular Modeling Studies of **A366** derivatives

After investigating the pocket flexibility and how to obtain the X-ray binding mode of **A366** starting from a closed conformation by computational means, we focused our attention on developing novel **A366** analogs (**chapter 3.4**). In an attempt to improve the activity and explore the SAR of the lead inhibitor, a congeneric series of **A366** derivatives

was designed, synthesized, tested *in vitro*, and investigated through molecular modeling approaches. The dataset encompassed 21 analogs, of which 17 displayed IC₅₀ values in the range of 0.15–11.8 μ M, two were found to be weakly active, and three were considered inactive as they showed a considerably decreased activity. Since **A366** is originally known as a potent G9a inhibitor, the most active Spindlin1 inhibitors were also tested *in vitro* against G9a and GLP (G9a like protein) to verify their selectivity profile. The crystal structure of Spindlin1 in complex with **A366** (PDB ID: 6I8Y – open cage) was used within this project to carry out docking and subsequent *in silico* studies.

Knowing that the difference between the active and inactive analogs lies in the linker chain length, we first employed docking studies and MD simulations to analyze and rationalize this information by computational means. Unfortunately, the docking results could not explain the compounds' inactivity, as similar binding modes and docking scores were detected for **A366** and the inactive analogs. This prompted us to further explore the predicted binding modes of both inactive compounds (**1c**, **1e**) and **A366** through long MD simulations (200 ns) (**Figure 4.5**). The stability of the poses and the interactions established during the simulations were examined. Interestingly, the analysis of the MD simulations shed light on the important interactions and their contributions. Specifically, the study highlighted the crucial role of the spacer length for the presence of the intramolecular hydrogen bond and how this influences the formation of the cation– π interactions with the aromatic cage residues and the interactions of the amidine moiety with Asp184. The inactive compounds, which bear either an ethyl or a butyl chain spacer, show a negligible occupancy rate of the intramolecular hydrogen bond and dramatically decreased values of the cation– π interactions compared to **A366** (**Figure 4.5**). In summary, the key structural features and interactions associated with the nanomolar activity of **A366** underlined from the MD analysis are:

- A positively charged pyrrolidine moiety embedded into the aromatic cage, which undergoes stable cation– π interactions with all the four surrounding amino acids.
- A propyl spacer, which is entropically favored as it stabilizes the bioactive conformation through the formation of an intramolecular hydrogen bond between the methoxy group and the positively charged pyrrolidine-NH.
- The amidine group of the 2-aminoindole core which anchors the compound into the pocket through salt bridge interactions with Asp184.

The absence of one of these interactions results in inactive **A366** analogs, as shown by **1c** and **1e**, which both lack a stable intramolecular hydrogen bond. Moreover, **1c** exhibits salt bridge interactions with Asp184 but the occupancy rates of the cation–pi interactions are generally low, while **1e** is able to establish cation–pi interactions to a certain extent but fails to preserve stable interactions with Asp184. The interaction heat map obtained from the 200 ns MD simulations is displayed in **Figure 4.5**.

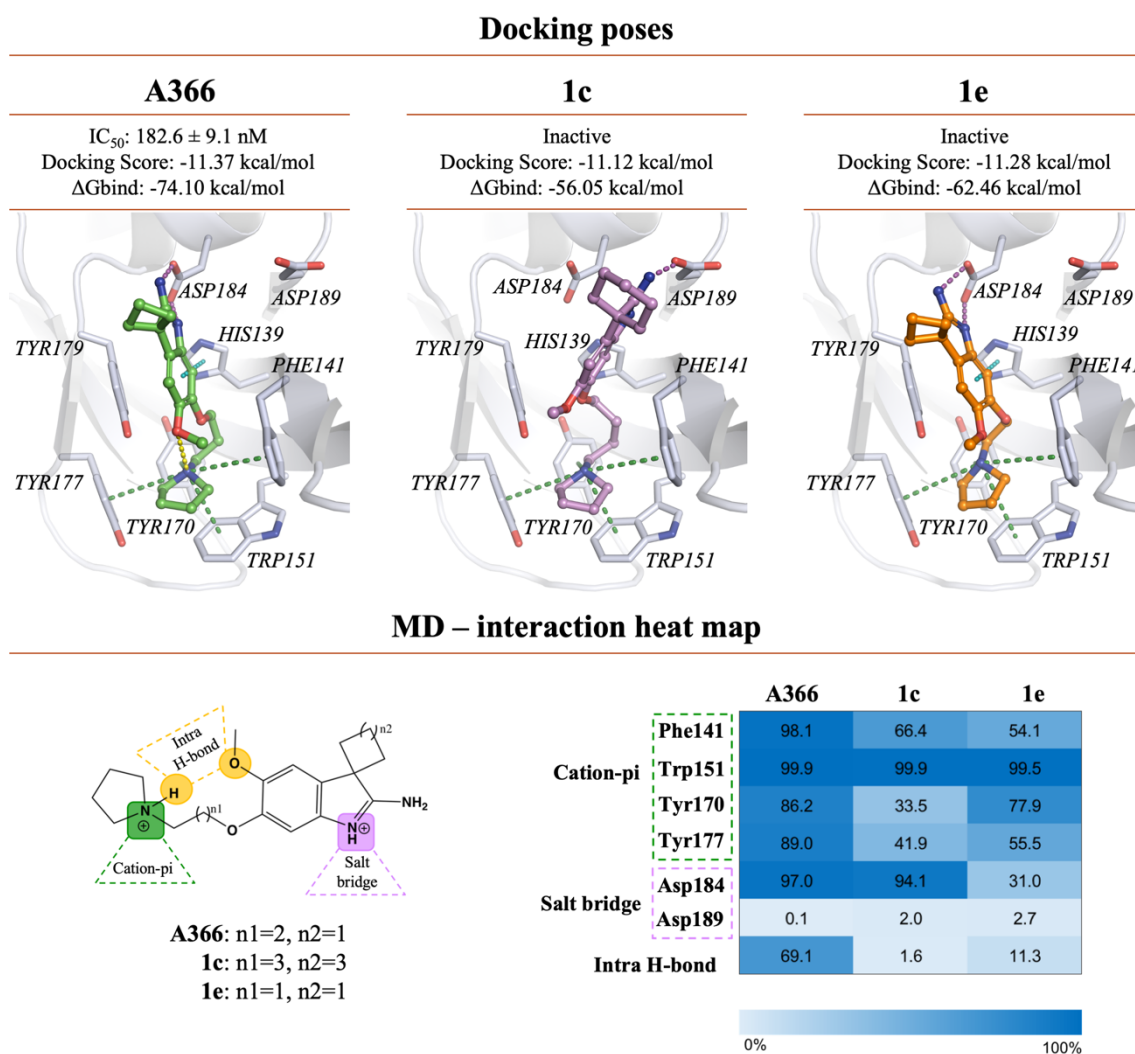


Figure 4.5. The upper panel shows the docking poses of **A366**, **1c** and **1e** in the second domain of Spindlin1 (PDB ID: 6I8Y – open cage). The ligands are displayed as balls and sticks, while the side chains of the surrounding residues as white sticks. Binding interactions are represented as dashed lines colored in magenta (salt bridge), yellow (hydrogen bonds), dark green (cation–pi interactions) and cyan (pi–pi stacking). The lower panel shows on the left side the 2D structure of the compounds highlighting the atoms involved in the interactions, and on the right side the heat map with the occupancy rates of the binding interactions identified during the 200 ns MD simulations of **A366**, **1c** and **1e** docking poses.

Although MD simulations gave insight into the compounds' activity, the approach remains too time-consuming for more extensive datasets. With the purpose to identify a

faster method capable of discriminating between active and inactive compounds within the **A366** series, we performed binding free energies calculations as rescoring steps. Interestingly, Prime MM-GBSA resulted in being up to the challenge providing good discrimination ranges between active and inactive compounds (MM-GBSA ΔG_{bind} mean values of -76.3 kcal/mol (actives) and -60.35 kcal/mol (inactives)). In **Figure 4.6** is reported the box-plot of the computed binding affinity values.

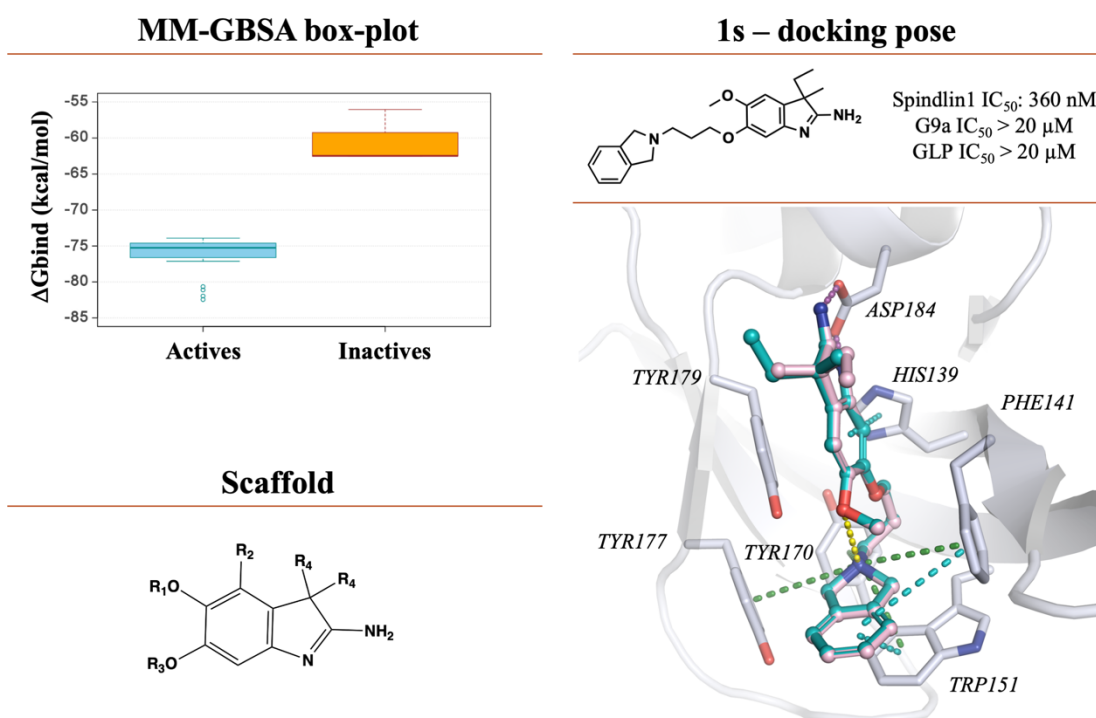


Figure 4.6. The left upper panel shows the box-plot representing the ΔG_{bind} distribution obtained for active and inactive compounds using Prime MM-GBSA (protocol when only ligands are relaxed). In the left lower panel is depicted the scaffold of our series. The right panel shows the 2D structure of compound **1s** as well as the predicted binding modes (S, cyan; R, light pink). The ligands are displayed as balls and sticks, while the side chains of the surrounding residues as white sticks. Binding interactions are represented as dashed lines colored in magenta (salt bridge), yellow (hydrogen bonds), dark green (cation-pi interactions) and cyan (pi-pi stacking).

Finally, through the SAR of our series, we could identify substitution patterns that can be used to design selective compounds. The biological activity of the **A366** derivatives highlighted the modifications that lead to a significant drop in the activity for G9a and GLP whilst maintaining the potent inhibitory activity against Spindlin1. Namely, small substitutions on the pyrrolidine moiety such as 3-methyl and 3-hydroxymethyl or replacing the pyrrolidine ring with an isoindoline moiety. As an example, **Figure 4.6** depicts the binding mode of **1s**, a good Spindlin1 binder and the most selective compound of our series (Spindlin1 IC_{50} : 360 ± 20 nM, G9a IC_{50} >20 μ M, GLP IC_{50} >20 μ M). Conversely, other modifications were well tolerated by both targets. For instance, the

most active derivative of the series (compound **1p**), which bears a dimethyl group instead of the spiro moiety on R4, showed no selectivity over G9a and GLP (Spindlin1 IC₅₀: 157 ± 12 nM, G9a IC₅₀: 759 nM, GLP IC₅₀: 470 nM).

In conclusion, the **A366** dataset yielded a variety of analogs with diverse potent and selective Spindlin1 inhibitors. Using MD simulations, we rationalized the lack of activity of some derivatives and elucidated the critical interactions and their contributions to the binding. We developed a rescoring protocol able to successfully discriminate active from inactive compounds that can be applied to guide further optimization steps to prioritize compounds for synthesis and biological characterization.

4.1.3.1. Limitations and Perspectives

The results of this work highlighted the limitations of docking scores, which, in some cases, fail to discriminate active compounds from inactive even within a congeneric series. Consequently, we rescored the docked molecules using MM-GBSA, a more accurate but slower approach. The dataset at our disposal enabled us to generate satisfactory binary models but not linear regression models. In fact, the range of the data is too narrow; moreover, the compounds are mainly spread in the nanomolar range, thus limiting the generation of linear regression models. Further work in this line would include probing new predictive models once a larger sample containing ligands with a broader range of activity will be available.

Taking into account the SAR of our series and the structural features of the binding site of Spindlin1 and G9a, perspectives for further modifications can be drawn. Knowing the good selectivity profile of analogs **1s** (R3: 3-(isoindolin-2-yl)propyl) and **1t** (R3: 3-((*R*)-3-methyl-pyrrolidin-1-yl)propyl), other bicyclic basic moieties on R3 –scaffold is shown in **Figure 4.6**– as well as bisubstitutions on the pyrrolidine moiety will be explored to attain new potent and selective compounds. Furthermore, position R2 will be investigated with longer and hydrophilic substitutions. Such modifications are hypothesized to pick up new interactions with the surrounding Spindlin1 amino acids (e.g., Asp95 and Glu142), gaining potency, and at the same time clash with the smaller G9a pocket and thus ensure selectivity.

4.2. Design and binding mode prediction of AChE and BChE inhibitors

In this project, the determined biological activities of a series of 8-hydroxy-2,7-naphthyridin-2-ium derivatives was rationalized by means of molecular docking studies (**chapter 3.5**). With the aim to identify dual inhibitors of AChE and BChE as well as inhibitors showing a preferential inhibition of AChE over BChE, we developed a set of derivatives starting from the knowledge that the natural products of the 8-methoxyisoquinoline group, e.g., chelerythrine (**Figure 4.7**), have been identified as AChE and BChE inhibitors.

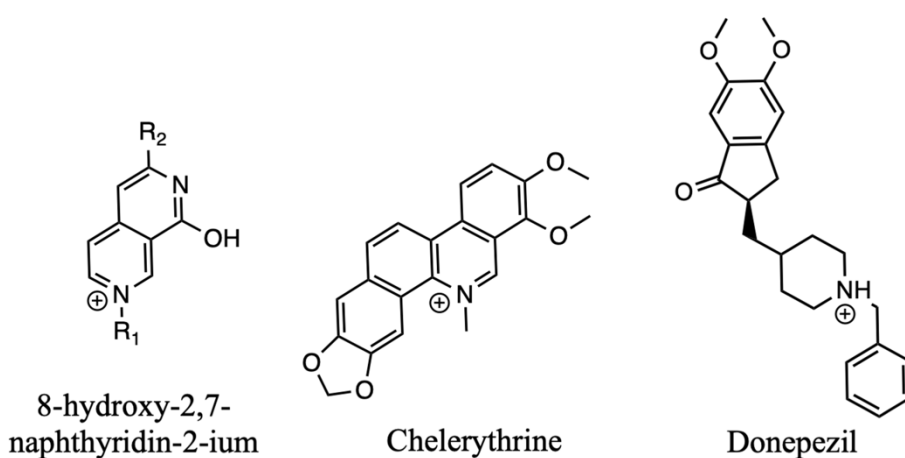


Figure 4.7. 2D chemical structures of our scaffold and two reported AChE and BChE inhibitors.

The 8-hydroxy-2,7-naphthyridin-2-ium derivatives were designed focusing on three main aspects: i) generate structurally different analogs to gain insights into the structural motifs that influence the biological activity; ii) synthetical accessibility of the atoms at positions 2 (R2) and 6 (R1) of the investigated scaffold (**Figure 4.7**); iii) analogy with the *N*-benzylated moiety of donepezil, a selective AChE inhibitor (**Figure 4.7**). The R1 position of the scaffold was thus explored by comparing *N*-benzylated to *N*-methylated and *N*-ethylated moieties. Subsequently, the synthesized ligands were docked into AChE and BChE pockets, minimized and rescored with MM-GBSA to identify plausible binding modes and rationalize their activity.

The *in vitro* inhibition data revealed that the *N*-benzylated derivatives are more active than the relative *N*-methylated and *N*-ethylated analogs allowing us to identify structural features that influence the activity of our series. From our docking studies, we could identify the role of the benzyl group in the AChE and BChE active sites: the moiety helps the compounds be anchored in the pocket, adopt the correct orientation, and establish

some of the key interactions (cation–pi and pi–pi interactions). As an example, the predicted binding modes of the most potent dual inhibitor **3e** in the AChE and BChE pockets are shown in **Figure 4.8**, whereas the 2D structure is reported in **Figure 4.9**. Moreover, the study highlighted that quaternary nitrogen of the active compound **3e** adopts the same position as observed for the protonated amino group of donepezil in the X-ray of AChE (**Figure 1.7**).

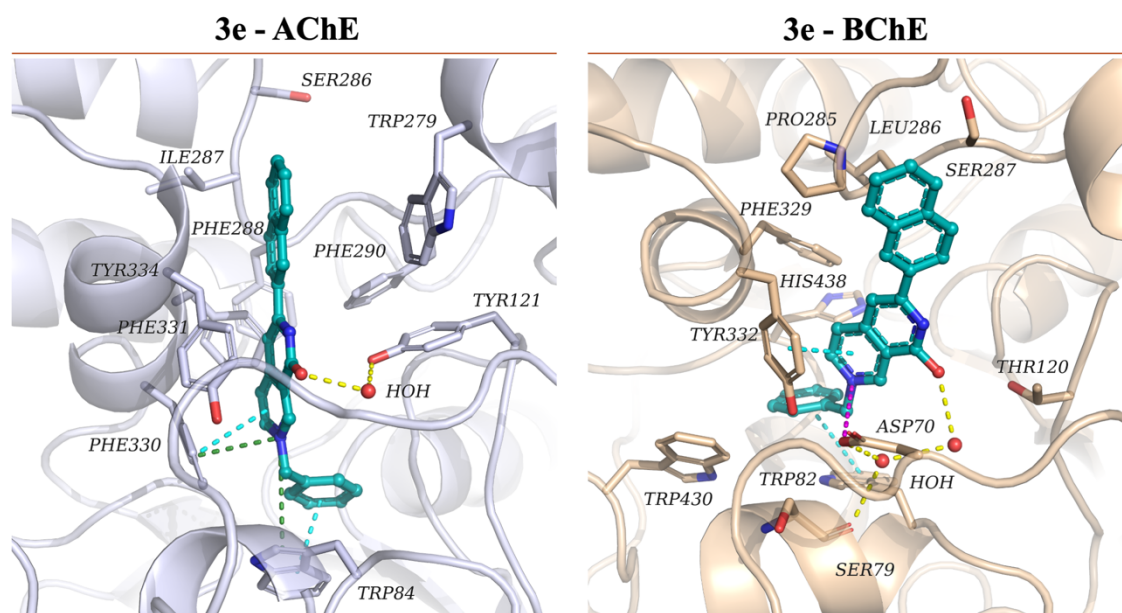


Figure 4.8. Predicted binding modes of compound **3e** in AChE and BChE active sites. The docking poses are depicted as balls and sticks colored in dark cyan, while the side chains of the surrounding residues are depicted as white (AChE) and beige (BChE) sticks. Only the side chains of the surrounding amino acid residues are shown for clarity; for residue Ser79 the main chain is shown. Binding interactions are represented as dashed lines colored in magenta (salt bridge), yellow (hydrogen bonds), dark green (cation-pi interactions) and cyan (pi-pi stacking).

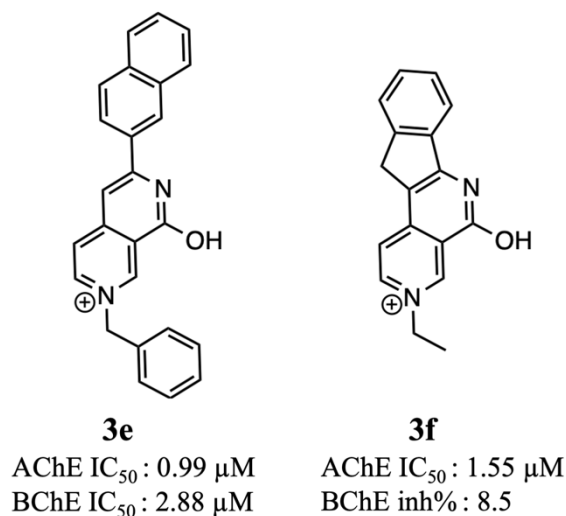


Figure 4.9. 2D chemical structures and activity data of compounds **3e** and **3f**.

Notably, compound **3f** showed preferential inhibition of AChE over BChE (**Figure 4.9**). The selectivity of **3f** is consistent with the binding mode hypotheses which suggest that when the benzyl group is missing, the molecule is not able to adopt a favorable orientation at the larger BChE binding pocket, and the critical interactions are not formed (salt bridge with Asp70, cation–pi interaction with Tyr332, pi–pi interaction with Trp82). On the other hand, the absence of the benzyl group at the smaller AChE active site results in less active compounds. Considering the activity data of the 8-hydroxy-2,7-naphthyridin-2-ium derivatives, compound **3f** can be viewed as a promising starting point for further selectivity studies.

The main challenge of this work was to understand the binding mode of our series and the structural features that influence the AChE and BChE activity; we addressed this by synthesis, *in vitro* characterization and *in silico* studies. In conclusion, we reported potent dual AChE and BChE inhibitors and a selective AChE inhibitor, and we could identify the structural features that affect the biological activity.

4.3. Exploring the cleavage behavior of NEP and MMP-14 by computational means

Our contribution in these projects focused on exploring and rationalizing from a structural point of view the cleavage behavior of NEP (**chapter 3.6**), MMP-14 and other MMP proteins (**chapter 3.7**) on specific human elastin fibers. Additionally, we performed a detailed analysis and comparison of the MMPs pockets under investigation.

To help the readers to understand the terminology following is a summary of the nomenclature used: the peptide sequence is defined as P4-P3-P2-P1-P1'-P2'-P3'-P4', where residues P4-P3-P2-P1 indicate directionality from N-terminal to the cleavage site, while P1'-P2'-P3'-P4' from the cleavage site to C-terminal. The corresponding binding sites on the enzymes are defined as S4-S3-S2-S1 and S1'-S2'-S3'-S4', respectively [314].

4.3.1. NEP

Using mass spectrometry, our collaborators identified cleavage sites and preferences for specific residues on the peptide sequences of fibrillar skin elastin; among these residues, some have already been reported, whereas others were not described before. Specifically, a preference for Pro at P2 and P3', Gly at P1 and bulky residues at P1' and P2' was observed. With the purpose of investigating these preferences by computational means, we performed docking studies for the peptides under analysis. From our study,

we could elucidate the preference for specific residues and suggest favorable interactions for others.

- Residue P1 lies in subpocket S1, which shows limited size, and thus, Gly seems to be favorable at this position.
- Subpocket S1', formed by Phe563 and Val692, is buried and hydrophobic in nature, and it is able to accommodate and interact with bulky residues (such as Phe, Ile, Leu and Lys) situated at P1' and P2' on the peptides.
- Pro at position P2 and P3' interacts with the aromatic ring of His587 and His711, respectively, and therefore appears to be a suitable residue in these positions.

Hydrogen bonds that stabilize the substrate peptides in the pocket and help place the scissile bond close to the zinc ion were detected in the docking results. These include a hydrogen bond with a conserved water bound to the zinc ion and several hydrogen bonds to NEP side chains (Arg102, Arg110, Asn542, His711, Arg717) and backbone residues (Ala543, Tyr545, Ser547) – slight variations are observed among diverse peptides. As an example, the predicted docking pose of GVPGAIPG at the active site of NEP is displayed in **Figure 4.10**.

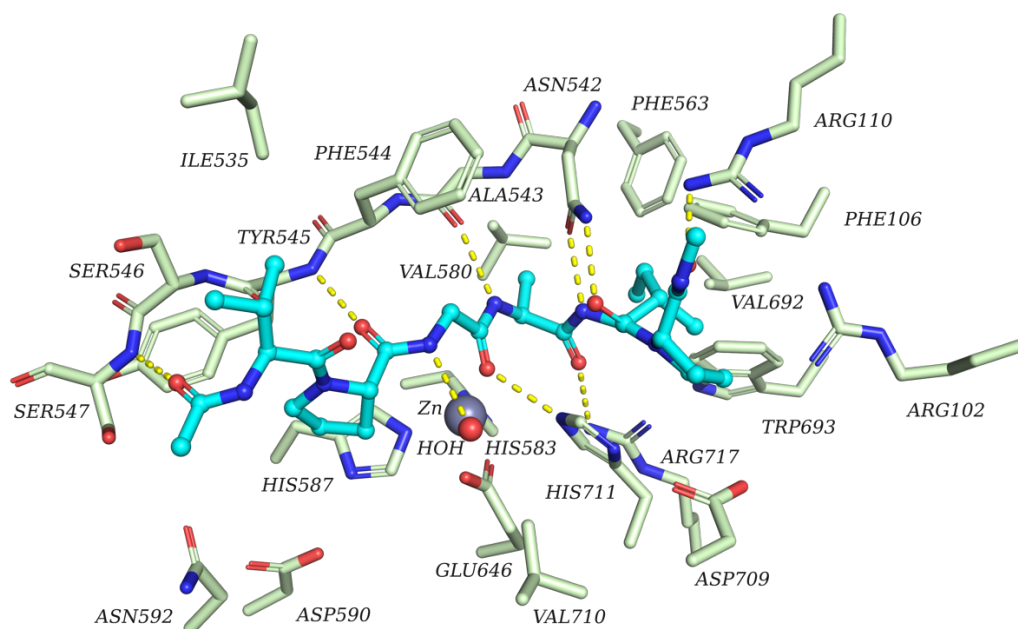


Figure 4.10. Predicted docking pose of substrate peptide GVPGAIPG (cyan ball and sticks) at the active site of NEP (pale green sticks). Hydrogen bonds are represented as yellow dashed lines. The water molecule is displayed as red sphere while the catalytic zinc ion as grey sphere. For clarity, only the central six residues of the peptide and the side chains of the surrounding amino acid residues are shown; for residues Asn542, Ala543, Phe544, Tyr545, Ser546 and Ser547 main chains are displayed.

4.3.2. MMP-14

Since MMP-14 was not yet described as an elastase, this work aimed to investigate the elastolytic capability of this protein and to compare it with other MMPs that were already known to cleave elastin, like MMP-7, MMP-9 and MMP-12. To this end, we used diverse techniques such as complementary mass spectrometric, bioinformatics tools and molecular docking studies to systematically characterize and compare the cleavage behavior and the cleavage site specificities of the MMPs under our analysis. Specifically, our collaborators identified by mass spectrometric analysis that tropoelastin and elastin peptides are digested by MMP-14 and other MMPs. Furthermore, cleavage site specificities and preferences covering the cleavage positions P4-P4' were determined. To gain insight into the binding mode of the digested peptides, we performed docking studies for the four MMPs. Furthermore, we carried out a comprehensive structure analysis and comparison of these MMPs with an emphasis on S1' subpocket, which highlighted similarities and main differences among the pockets. The readers are referred to **chapter 3.7** for details about pocket analysis. Herein are summarized the findings of the docking studies at MMP-14, MMP-12, MMP-9 and MMP-7 active sites for six octapeptides derived from cleavage sites found in the MMP-14 digests of elastin.

- The docking poses of all analyzed octapeptides displayed similar binding modes at the active sites, suggesting that MMP-14, MMP-12, MMP-9 and MMP-7 are able to accommodate and cleave the same peptides.
- Small and medium-sized hydrophobic residues including Gly, Ala, Leu and Val at cleavage site P1' are preferred by MMP-14, MMP-12, MMP-9 and MMP-7 as shown by mass spectrometric analysis. Nonetheless, docking studies highlighted that bigger residues such as Lys, Phe and Tyr, are likewise well accepted and fitted in the MMP-14, MMP-12, MMP-9 and MMP-7 active sites.

The docking studies could identify hydrogen bonds that stabilize the substrate peptides in the MMPs pockets and favor the placement of the scissile bond close to the zinc ion. Specifically, residue P1' establishes hydrogen bonds with the backbones of Leu and Ala (conserved residues among these four MMPs). Moreover, the other peptide residues are involved in a conserved hydrogen bond with the zinc ion and several hydrogen bonds to MMPs backbone residues that further stabilize the peptides into the pocket. Based on the nature of the octapeptides, pi-pi stacking and cation-pi interactions

with MMPs side chain residues are also observed. The predicted binding mode of PGAGLGAL in the active site of MMP-14 is shown in **Figure 4.11** as an example.

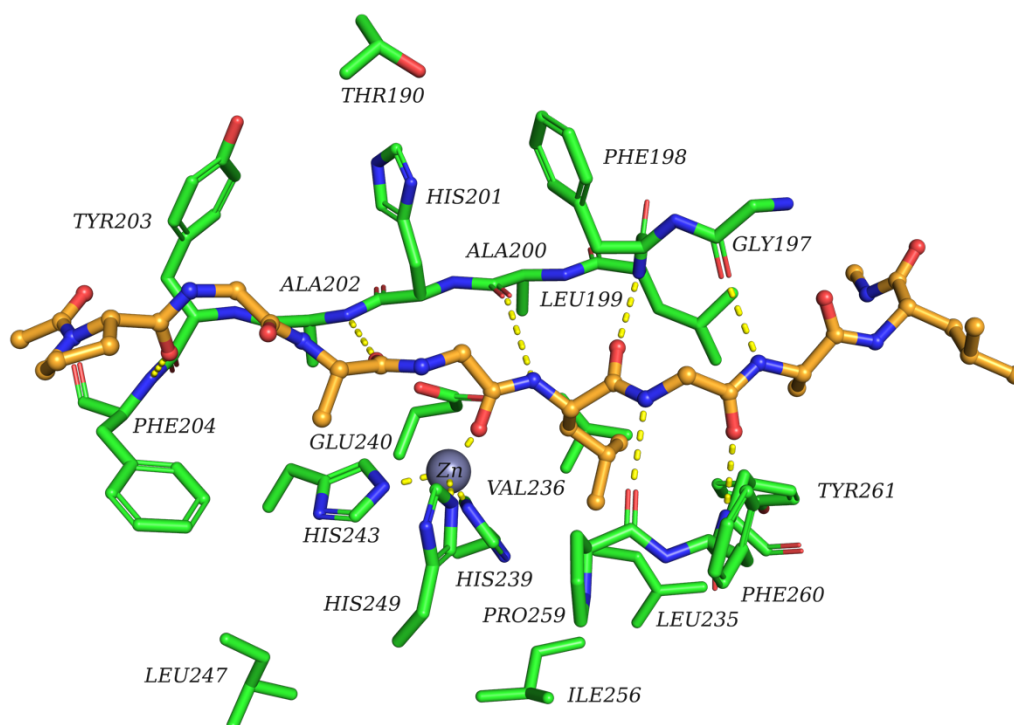


Figure 4.11. Predicted docking pose of substrate peptide PGAGLGAL (orange ball and sticks) at the active site of MMP-14 (green sticks). Hydrogen bonds are represented as yellow dashed lines, while the catalytic zinc ion as grey sphere. Only the side chains of the surrounding amino acid residues are shown for clarity; for residues Gly197, Phe198, Leu199, Ala200, His201, Ala202, Tyr203, Phe204, Pro259, Phe260 and Tyr261 main chains are displayed.

4.3.3. Common challenges and limitations

These two projects shared the same challenges and limitations as both tried to address the same questions. Within these works, we carried out docking studies to suggest binding mode hypothesis and investigate the cleavage behavior of specific metalloproteinases. As discussed in detail in the introduction (subparagraph 1.1.1.2), molecular docking is the most widely used technique, although it has limitations. One above all is the low accuracy of the scoring functions, which is even more evident for peptide docking due to the bigger size and increased flexibility of the peptides. In fact, it is difficult to differentiate which peptide would serve as a better substrate based only on the obtained docking scores. Thus, the analysis of the binding mode and the interactions established at the active site give better insight than relying solely on the docking scores. Furthermore, another aspect to consider when performing docking is the conformational sampling problem for very flexible ligands with many rotatable bonds like peptides. General procedures used to avoid this drawback are to extend the number of poses generated and use constraints to

ensure that the correct poses are sampled; both strategies were applied in these studies to reduce the sampling problem.

4.4. General conclusions

CADD methodologies are widely used both in pharmaceutical industries and academia and have become an essential part of the drug discovery pipeline. The general aim of this thesis was to apply computational approaches to drug discovery and design projects to identify new inhibitors, rationalize their activity, get insight into protein flexibility and investigate the binding mode of small molecules and peptides. We used diverse approaches, including structure-based pharmacophore modeling, VS, docking (classical rigid-body docking and flexible docking), MD simulations and MM-GBSA calculations.

Through the studies presented herein, we have shown how computer -based methods can be successfully used to address specific questions and challenges that are project/target related. For instance, by applying different VS approaches, coupled with *in vitro* testing, we could identify novel hits even for a new target that lacked known inhibitors. Not surprisingly, the hits were active in the micromolar range; nonetheless, the study pointed out how structural optimization could lead to higher inhibitory activity. Concerning molecular docking, this approach was constructively applied in all the projects to predict the ligands binding mode and, for AChE and BChE, also to explain the selectivity between the two targets. However, as we have seen in the case of Spindlin1, rigid-body docking fails to retrieve the proper binding mode when dealing with flexible pockets. In such a scenario, more thorough computational methods like flexible docking combined with short MD simulations can be used to achieve a better binding mode prediction. Moreover, when possible, we have overcome the limitations of docking scores by rescoring the ligand poses with MM-GBSA. For Spindlin1, we have proved that satisfactory MM-GBSA models that discriminate between active from inactive compounds can be generated.

To conclude, *in silico* techniques are undoubtedly valuable tools to search and investigate potential drug candidates with reduced cost and time. Nevertheless, some limitations still need to be addressed, such as more accurate scoring functions, incorporating solvent effects and faster evaluation of pocket flexibility in the docking procedure, and increasing computational efficiency.

Finally, this work has also emphasized that collaborations between different fields are essential to achieve progress in drug discovery.

Page left intentionally blank.

5. GLOSSARY OF TERMS USED

Follow a brief definition of the technical terms used in the thesis. The glossary was readapted from <http://www.drugdesign.com/web/teaching/glossary> [315].

Affinity - The affinity of a ligand is its ability to bind to its biological target.

Analog - A molecule structurally similar to another, generally based on the same scaffold.

Antagonist - A molecule that blocks the activation of a receptor.

Apo - 3D structure of a macromolecule without a ligand.

Benchmark - A dataset by which the validity of a method can be measured and judged.

Binding mode - Orientation and geometry adopted by a chemical substance when it is bound to a specific protein.

Bioactive conformation - The geometry adopted by a ligand when it binds to its biological target.

False positive - A hit that is erroneously recognized as good (positive).

FDA - Food and Drug Administration. Agency in the USA responsible for safety regulations.

Force field - A set of mathematical equations and parameters used for assessing the energy of a given chemical system.

Genetic algorithm - A computerized search technique inspired by evolutionary biology and used in computer science to find approximate solutions to optimization and search problems.

Harmonic - A function whose frequency is an integral multiple of the frequency of a reference function.

Holo - 3D structure of a macromolecule with a bound ligand.

Homology Modeling - A method for predicting the 3D structure of a protein, based on its amino acid sequence and the 3D-structure of analog protein(s).

In silico - Refers to the use of silicon-based computer technologies to perform simulations, modeling and experiments.

In vivo - Refers to the technique of performing a given procedure in a controlled environment outside of a living organism. The studies are performed with microorganisms, cells, or biological molecules outside their normal biological context.

Inhibitor - Chemical substance that blocks or suppress the activity of a given protein.

Isomers - Isomers are compounds with the same chemical formula but different 2D-structures.

Ligand - In Biochemistry: substance that binds to a biological target. In Chemistry: an atom or group of atoms.

Minimization - Minimization treatments consist of successive alterations of the geometry of the molecule until a minimum is found on the conformational potential surface.

Moieties - Fragments, functional groups or portions of chemical compounds.

Molecular descriptor - Molecular descriptors are numerical values that capture the structure and properties of molecules.

Molecular dynamics - Molecular dynamics (MD) is a computer simulation technique which follows the time evolution of a molecular system in 3D. Successive integration of Newtons equations of motion over time enables to obtain information about time-dependent properties of the system.

Molecular mechanics - Molecular mechanics is a computer simulation technique for modeling the molecular geometry and energy of a system based on the energy minimization of its potential energy function. The set of potential functions used to calculate the energy is known as the force-field.

Nuclear Magnetic Resonance, NMR - An analytical method that allows the spectroscopic detection of structural information of molecules. It requires the application of a strong magnetic field.

Pharmacophore - Specific 3D arrangement of chemical groups common to active molecules and essential to their biological activities.

Promiscuous - Some molecules emerge repeatedly as hits in diverse unrelated target systems. These so-called promiscuous hits act non-competitively, show little structure–activity relationships and have poor selectivity.

Protein Data Bank - The "Protein Data Bank". A worldwide repository source of 3D structures of proteins obtained by X-ray crystallography or NMR studies.

Protein folding - A spontaneous process directed by the physical and chemical properties of the amino acid sequence that dictate the folding of a protein into a particular conformation.

QSAR, Quantitative Structure Activity Relationships - Quantitative Structure Activity Relationships (QSAR) - Mathematical equation linking chemical structure and biological activity for a series of compounds.

RMS - Root Mean Square: the square root of the arithmetic average of the square's set of values.

RMSD - The Root Mean Square Deviation is a measure of the differences between the values predicted by a model and the values experimentally observed.

Stochastic - Stochastic means "random" in opposition to "deterministic" (where random phenomena are not involved). Stochastic models are based on random trials that are guided by computerized stochastic algorithms.

Stochastic algorithm - Process with random variables.

Structure–activity relationships - Structure–Activity Relationships (SAR) is the analysis of the relationships between chemical structure and biological activity.

Trajectory - Trajectory = positions + velocities.

Virtual library - A library which exists solely in electronic form and used in the design and evaluation of possible real libraries.

Page left intentionally blank.

6. REFERENCES

- Murray, D., et al., *In silico gene expression analysis--an overview*. Mol Cancer, 2007. **6**: p. 50.
- Katiyar, R.S. and P.K. Jha, *Molecular simulations in drug delivery: Opportunities and challenges*. WIREs Computational Molecular Science, 2018. **8**(4): p. e1358.
- Gleeson, M.P., *Generation of a Set of Simple, Interpretable ADMET Rules of Thumb*. Journal of Medicinal Chemistry, 2008. **51**(4): p. 817-834.
- Goller, A.H., et al., Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. Drug Discov Today, 2020.
- Gund, P., et al., *Three-dimensional molecular modeling and drug design*. Science, 1980. **208**(4451): p. 1425-1431.
- G, B., *A new industrial revolution is on the way*. Fortune, 1981. **104**: p. 106-114.
- Brown, F.K., et al., *The evolution of drug design at Merck Research Laboratories*. J Comput Aided Mol Des, 2017. **31**(3): p. 255-266.
- Jorgensen, W.L., *The many roles of computation in drug discovery*. Science, 2004. **303**(5665): p. 1813-8.
- Van Drie, J.H., *Computer-aided drug design: the next 20 years*. J Comput Aided Mol Des, 2007. **21**(10-11): p. 591-601.
- Kubinyi, H., Success Stories of Computer-Aided Design, in Computer Applications in Pharmaceutical Research and Development, B.W.a.S. Ekins, Editor. 2006. p. 377-424.
- Muegge, I., A. Bergner, and J.M. Kriegl, *Computer-aided drug design at Boehringer Ingelheim*. J Comput Aided Mol Des, 2017. **31**(3): p. 275-285.
- Warr, W.A., *A CADD-alog of strategies in pharma*. J Comput Aided Mol Des, 2017. **31**(3): p. 245-247.
- Stahl, M., W. Guba, and M. Kansy, Integrating molecular design resources within modern drug discovery research: the Roche experience. Drug Discov Today, 2006. **11**(7-8): p. 326-33.
- Hillisch, A., N. Heinrich, and H. Wild, Computational Chemistry in the Pharmaceutical Industry: From Childhood to Adolescence. ChemMedChem, 2015. **10**(12): p. 1958-62.
- Hartman, G.D., et al., Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. J Med Chem, 1992. **35**(24): p. 4640-2.
- Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions*. J Mol Biol, 1982. **161**(2): p. 269-88.
- Chen, Y.C., *Beware of docking!* Trends Pharmacol Sci, 2015. **36**(2): p. 78-95.
- Yan, Y., et al., HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res, 2017. **45**(W1): p. W365-W373.
- Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Research, 2000. **28**(1): p. 235-242.
- <https://www.rcsb.org>. (Accessed August 3, 2021).
- Biasini, M., et al., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 2014. **42**(W1): p. W252-W258.
- Baker, D. and A. Sali, *Protein structure prediction and structural genomics*. Science, 2001. **294**(5540): p. 93-6.
- Šali, A. and T.L. Blundell, *Comparative Protein Modelling by Satisfaction of Spatial Restraints*. Journal of Molecular Biology, 1993. **234**(3): p. 779-815.
- Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
- Hetenyi, C. and D. van der Spoel, Blind docking of drug-sized compounds to proteins with up to a thousand residues. FEBS Lett, 2006. **580**(5): p. 1447-50.

26. Davis, I.W., et al., Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci*, 2009. **18**(9): p. 1998-2002.
27. Hassan, N.M., et al., Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. *Sci Rep*, 2017. **7**(1): p. 15451.
28. Hetenyi, C. and D. van der Spoel, Toward prediction of functional protein pockets using blind docking and pocket search algorithms. *Protein Sci*, 2011. **20**(5): p. 880-93.
29. Abagyan, R. and M. Totrov, *High-throughput docking for lead generation*. *Current Opinion in Chemical Biology*, 2001. **5**(4): p. 375-382.
30. Kitchen, D.B., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 2004. **3**(11): p. 935-49.
31. Kroemer, R.T., *Structure-based drug design: docking and scoring*. *Curr Protein Pept Sci*, 2007. **8**(4): p. 312-28.
32. Huang, S.Y. and X. Zou, *Advances and challenges in protein-ligand docking*. *Int J Mol Sci*, 2010. **11**(8): p. 3016-34.
33. Limongelli, V., *Ligand binding free energy and kinetics calculation in 2020*. *WIREs Computational Molecular Science*, 2020. **10**(4): p. e1455.
34. Yin, S., et al., MedusaScore: an accurate force field-based scoring function for virtual drug screening. *J Chem Inf Model*, 2008. **48**(8): p. 1656-62.
35. Bohm, H.J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des*, 1994. **8**(3): p. 243-56.
36. Eldridge, M.D., et al., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*, 1997. **11**(5): p. 425-45.
37. Muegge, I. and Y.C. Martin, A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem*, 1999. **42**(5): p. 791-804.
38. Gohlke, H., M. Hendlich, and G. Klebe, *Knowledge-based scoring function to predict protein-ligand interactions*. *J Mol Biol*, 2000. **295**(2): p. 337-56.
39. Ruvinsky, A.M., Calculations of protein-ligand binding entropy of relative and overall molecular motions. *J Comput Aided Mol Des*, 2007. **21**(7): p. 361-70.
40. Chang, M.W., et al., Empirical entropic contributions in computational docking: evaluation in APS reductase complexes. *J Comput Chem*, 2008. **29**(11): p. 1753-61.
41. Lee, J. and C. Seok, A statistical rescoring scheme for protein-ligand docking: Consideration of entropic effect. *Proteins*, 2008. **70**(3): p. 1074-83.
42. Charifson, P.S., et al., Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem*, 1999. **42**(25): p. 5100-9.
43. Feher, M., *Consensus scoring for protein-ligand interactions*. *Drug Discov Today*, 2006. **11**(9-10): p. 421-8.
44. Teramoto, R. and H. Fukunishi, Consensus scoring with feature selection for structure-based virtual screening. *J Chem Inf Model*, 2008. **48**(2): p. 288-95.
45. Bohm, H.J., LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des*, 1992. **6**(6): p. 593-606.
46. Friesner, R.A., et al., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 2004. **47**(7): p. 1739-1749.
47. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme*. *Berichte der deutschen chemischen Gesellschaft*, 1894. **27**(3): p. 2985-2993.
48. Jorgensen, W.L., Rusting of the lock and key model for protein-ligand binding. *Science*, 1991. **254**(5034): p. 954-5.
49. Brooijmans, N. and I.D. Kuntz, *Molecular recognition and docking algorithms*. *Annu Rev Biophys Biomol Struct*, 2003. **32**: p. 335-73.
50. A comprehensive list of docking software can be found at <http://www.Click2Drug.org> - SIB, Swiss Institute of Bioinformatics. (Accessed June 25, 2020).
51. Ciemny, M., et al., *Protein-peptide docking: opportunities and challenges*. *Drug Discov Today*, 2018. **23**(8): p. 1530-1537.

52. Leach, A.R., *Ligand docking to proteins with discrete side-chain flexibility*. Journal of Molecular Biology, 1994. **235**(1): p. 345-356.
53. Jones, G., et al., Development and validation of a genetic algorithm for flexible docking. J Mol Biol, 1997. **267**(3): p. 727-48.
54. Sherman, W., et al., Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem, 2006. **49**(2): p. 534-53.
55. Bottegoni, G., et al., A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). J Comput Aided Mol Des, 2008. **22**(5): p. 311-25.
56. Knegtel, R.M., I.D. Kuntz, and C.M. Oshiro, *Molecular docking to ensembles of protein structures*. J Mol Biol, 1997. **266**(2): p. 424-40.
57. Huang, S.Y. and X. Zou, Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins, 2007. **66**(2): p. 399-421.
58. Totrov, M. and R. Abagyan, Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol, 2008. **18**(2): p. 178-84.
59. Lin, J.H., et al., Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc, 2002. **124**(20): p. 5632-3.
60. Amaro, R.E., R. Baron, and J.A. McCammon, *An improved relaxed complex scheme for receptor flexibility in computer-aided drug design*. J Comput Aided Mol Des, 2008. **22**(9): p. 693-705.
61. Amaro, R.E., et al., *Ensemble Docking in Drug Discovery*. Biophys J, 2018. **114**(10): p. 2271-2278.
62. Shoichet, B.K., *Virtual screening of chemical libraries*. Nature, 2004. **432**(7019): p. 862-5.
63. Graves, A.P., R. Brenk, and B.K. Shoichet, *Decoys for docking*. J Med Chem, 2005. **48**(11): p. 3714-28.
64. Wallach, I. and R. Lilien, *Virtual decoy sets for molecular docking benchmarks*. J Chem Inf Model, 2011. **51**(2): p. 196-202.
65. Mysinger, M.M., et al., Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem, 2012. **55**(14): p. 6582-94.
66. Scior, T., et al., *Recognizing pitfalls in virtual screening: a critical review*. J Chem Inf Model, 2012. **52**(4): p. 867-81.
67. Luise, C. and D. Robaa, Application of Virtual Screening Approaches for the Identification of Small Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1. Methods Mol Biol, 2018. **1824**: p. 347-370.
68. Elokely, K.M. and R.J. Doerksen, *Docking challenge: protein sampling and molecular docking performance*. J Chem Inf Model, 2013. **53**(8): p. 1934-45.
69. Roberts, B.C. and R.L. Mancera, *Ligand-protein docking with water molecules*. J Chem Inf Model, 2008. **48**(2): p. 397-408.
70. Ruvinsky, A.M., et al., The Role of Bridging Water and Hydrogen Bonding as Key Determinants of Noncovalent Protein-Carbohydrate Recognition. ChemMedChem, 2018. **13**(24): p. 2684-2693.
71. Verdonk, M.L., et al., *Modeling water molecules in protein-ligand docking using GOLD*. J Med Chem, 2005. **48**(20): p. 6504-15.
72. Daina, A., et al., Drug Design Workshop: A Web-Based Educational Tool To Introduce Computer-Aided Drug Design to the General Public. Journal of Chemical Education, 2017. **94**(3): p. 335-344.
73. Allen, M.P., *Introduction to Molecular Dynamics Simulation*, in *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Note*, K.B. Norbert Attig, Helmut Grubmüller, Kurt Kremer, Editor. 2004, John von Neumann Institute for Computing - NIC-Directors: Jülich. p. 1-28.
74. Frenkel, D. and B. Smit, *Chapter 4 - Molecular Dynamics Simulations*, in *Understanding Molecular Simulation (Second Edition)*, D. Frenkel and B. Smit, Editors. 2002, Academic Press: San Diego. p. 63-107.
75. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**(5612): p. 585-90.

76. Zhao, H. and A. Caflisch, *Molecular dynamics in drug design*. Eur J Med Chem, 2015. **91**: p. 4-14.
77. De Vivo, M., et al., Role of Molecular Dynamics and Related Methods in Drug Discovery. J Med Chem, 2016. **59**(9): p. 4035-61.
78. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery*. BMC Biol, 2011. **9**: p. 71.
79. Wong, C.F., et al., Molecular docking of balanol to dynamics snapshots of protein kinase A. Proteins, 2005. **61**(4): p. 850-8.
80. Sakano, T., et al., *Molecular dynamics analysis to evaluate docking pose prediction*. Biophys Physicobiol, 2016. **13**: p. 181-194.
81. Schames, J.R., et al., *Discovery of a novel binding trench in HIV integrase*. J Med Chem, 2004. **47**(8): p. 1879-81.
82. Stank, A., et al., TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. Nucleic Acids Research, 2017. **45**(W1): p. W325-W330.
83. Mustata, G. and J.M. Briggs, *Cluster analysis of water molecules in alanine racemase and their putative structural role*. Protein Engineering, Design and Selection, 2004. **17**(3): p. 223-234.
84. Rudling, A., A. Orro, and J. Carlsson, Prediction of Ordered Water Molecules in Protein Binding Sites from Molecular Dynamics Simulations: The Impact of Ligand Binding on Hydration Networks. J Chem Inf Model, 2018. **58**(2): p. 350-361.
85. Jukic, M., et al., ProBiS H2O MD Approach for Identification of Conserved Water Sites in Protein Structures for Drug Design. ACS Med Chem Lett, 2020. **11**(5): p. 877-882.
86. Schaller, D., S. Pach, and G. Wolber, *PyRod: Tracing Water Molecules in Molecular Dynamics Simulations*. J Chem Inf Model, 2019. **59**(6): p. 2818-2829.
87. van Meel, J.A., et al., *Harvesting graphics power for MD simulations*. Molecular Simulation, 2008. **34**(3): p. 259-266.
88. Nobile, M.S., et al., Graphics processing units in bioinformatics, computational biology and systems biology. Brief Bioinform, 2017. **18**(5): p. 870-885.
89. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1): p. 141-151.
90. Laio, A. and M. Parrinello, *Escaping free-energy minima*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12562-6.
91. Isralewitz, B., M. Gao, and K. Schulten, *Steered molecular dynamics and mechanical functions of proteins*. Current Opinion in Structural Biology, 2001. **11**(2): p. 224-230.
92. Rapp, C., et al., A molecular mechanics approach to modeling protein-ligand interactions: relative binding affinities in congeneric series. J Chem Inf Model, 2011. **51**(9): p. 2082-9.
93. Guimaraes, C.R. and M. Cardozo, *MM-GB/SA rescoring of docking poses in structure-based lead optimization*. J Chem Inf Model, 2008. **48**(5): p. 958-70.
94. Lyne, P.D., M.L. Lamb, and J.C. Saeh, Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. J Med Chem, 2006. **49**(16): p. 4805-8.
95. Lindstrom, A., et al., Postprocessing of docked protein-ligand complexes using implicit solvation models. J Chem Inf Model, 2011. **51**(2): p. 267-82.
96. Mulakala, C. and V.N. Viswanadhan, Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies? J Mol Graph Model, 2013. **46**: p. 41-51.
97. Pu, C., et al., Assessing the performance of docking scoring function, FEP, MM-GBSA, and QM/MM-GBSA approaches on a series of PLK1 inhibitors. Medchemcomm, 2017. **8**(7): p. 1452-1458.
98. Genheden, S. and U. Ryde, *The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities*. Expert Opin Drug Discov, 2015. **10**(5): p. 449-61.
99. Aqvist, J., C. Medina, and J.E. Samuelsson, *A new method for predicting binding affinity in computer-aided drug design*. Protein Eng, 1994. **7**(3): p. 385-91.

100. Gutierrez-de-Teran, H. and J. Åqvist, *Linear interaction energy: method and applications in drug design*. Methods Mol Biol, 2012. **819**: p. 305-23.
101. Straatsma, T.P. and J.A. McCammon, *Computational Alchemy*. Annual Review of Physical Chemistry, 1992. **43**(1): p. 407-435.
102. Kollman, P., Free energy calculations: Applications to chemical and biochemical phenomena. Chemical Reviews, 1993. **93**(7): p. 2395-2417.
103. Bennett, C.H., *Efficient estimation of free energy differences from Monte Carlo data*. Journal of Computational Physics, 1976. **22**(2): p. 245-268.
104. Hevener, K.E., et al., Quantitative structure-activity relationship studies on nitrofuranyl anti-tubercular agents. Bioorg Med Chem, 2008. **16**(17): p. 8042-53.
105. Avram, S., et al., Evaluation of the pharmacological descriptors related to the induction of antimicrobial activity by using QSAR and computational mutagenesis, in Recent Trends on QSAR in the Pharmaceutical Perceptions. 2012. p. 63-98.
106. Rastelli, G., et al., Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem, 2010. **31**(4): p. 797-810.
107. Li, Y., Z. Liu, and R. Wang, *Test MM-PB/SA on true conformational ensembles of protein-ligand complexes*. J Chem Inf Model, 2010. **50**(9): p. 1682-92.
108. Cheatham, T.E., 3rd, et al., Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. J Biomol Struct Dyn, 1998. **16**(2): p. 265-80.
109. Srinivasan, J., et al., *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
110. Kollman, P.A., et al., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res, 2000. **33**(12): p. 889-97.
111. Vorobjev, Y.N. and J. Hermans, ES/IS: Estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophysical Chemistry, 1999. **78**(1): p. 195-205.
112. Foloppe, N. and R. Hubbard, Towards predictive ligand design with free-energy based computational methods? Curr Med Chem, 2006. **13**(29): p. 3583-608.
113. Homeyer, N. and H. Gohlke, Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. Mol Inform, 2012. **31**(2): p. 114-22.
114. Sirin, S., D.A. Pearlman, and W. Sherman, *Physics-based enzyme design: predicting binding affinity and catalytic activity*. Proteins, 2014. **82**(12): p. 3397-409.
115. Raha, K. and K.M. Merz, Jr., Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. J Med Chem, 2005. **48**(14): p. 4558-75.
116. Gao, C., M.S. Park, and H.A. Stern, Accounting for ligand conformational restriction in calculations of protein-ligand binding affinities. Biophys J, 2010. **98**(5): p. 901-10.
117. Yang, T., et al., *Virtual screening using molecular simulations*. Proteins, 2011. **79**(6): p. 1940-51.
118. Homeyer, N., et al., Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. J Chem Theory Comput, 2014. **10**(8): p. 3331-44.
119. Gutiérrez-de-Terán, H. and J. Åqvist, *Linear Interaction Energy: Method and Applications in Drug Design*, in *Computational Drug Discovery and Design*, R. Baron, Editor. 2012, Springer New York: New York, NY. p. 305-323.
120. Wang, L., et al., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J Am Chem Soc, 2015. **137**(7): p. 2695-703.
121. Cappel, D., et al., Impact of Different Automated Binding Pose Generation Approaches on Relative Binding Free Energy Simulations. J Chem Inf Model, 2020. **60**(3): p. 1432-1444.
122. Karaman, B. and W. Sippl, *Docking and binding free energy calculations of sirtuin inhibitors*. Eur J Med Chem, 2015. **93**: p. 584-98.

123. Wang, J., et al., Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J Am Chem Soc*, 2001. **123**(22): p. 5221-30.
124. Mobley, D.L. and K.A. Dill, Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure*, 2009. **17**(4): p. 489-98.
125. Cournia, Z., B. Allen, and W. Sherman, Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J Chem Inf Model*, 2017. **57**(12): p. 2911-2937.
126. Sliwoski, G., et al., *Computational methods in drug discovery*. *Pharmacol Rev*, 2014. **66**(1): p. 334-95.
127. Lill, M., *Virtual screening in drug design*. *Methods Mol Biol*, 2013. **993**: p. 1-12.
128. Robaa, D., et al., Identification and Structure-Activity Relationship Studies of Small-Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1. *ChemMedChem*, 2016. **11**(20): p. 2327-2338.
129. Zhu, T., et al., Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *J Med Chem*, 2013. **56**(17): p. 6560-72.
130. Schlegel, B., et al., Generation of a homology model of the human histamine H(3) receptor for ligand docking and pharmacophore-based screening. *J Comput Aided Mol Des*, 2007. **21**(8): p. 437-53.
131. Singh, J., et al., Successful Shape-Based Virtual Screening: The Discovery of a Potent Inhibitor of the Type I TGF β Receptor Kinase (T β RI). *Bioorganic & Medicinal Chemistry Letters*, 2003. **13**(24): p. 4355-4359.
132. Barreca, M.L., et al., *Efficient 3D database screening for novel HIV-1 IN inhibitors*. *J Chem Inf Comput Sci*, 2004. **44**(4): p. 1450-5.
133. De Luca, L., et al., A refined pharmacophore model for HIV-1 integrase inhibitors: Optimization of potency in the 1H-benzylindole series. *Bioorg Med Chem Lett*, 2008. **18**(9): p. 2891-5.
134. Barreca, M.L., et al., Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors. *J Med Chem*, 2005. **48**(22): p. 7084-8.
135. Sabatini, S., et al., Re-evolution of the 2-phenylquinolines: ligand-based design, synthesis, and biological evaluation of a potent new class of Staphylococcus aureus NorA efflux pump inhibitors to combat antimicrobial resistance. *J Med Chem*, 2013. **56**(12): p. 4975-89.
136. Cai, H., et al., *Combined Virtual Screening and Substructure Search for Discovery of Novel FABP4 Inhibitors*. *Journal of Chemical Information and Modeling*, 2017. **57**(9): p. 2329-2335.
137. Lesnik, S., et al., LiSiCA: A Software for Ligand-Based Virtual Screening and Its Application for the Discovery of Butyrylcholinesterase Inhibitors. *J Chem Inf Model*, 2015. **55**(8): p. 1521-8.
138. Divsalar, D.N., et al., Novel Histone Deacetylase Inhibitors and HIV-1 Latency-Reversing Agents Identified by Large-Scale Virtual Screening. *Front Pharmacol*, 2020. **11**: p. 905.
139. De Luca, L., et al., Pharmacophore-based discovery of small-molecule inhibitors of protein-protein interactions between HIV-1 integrase and cellular cofactor LEDGF/p75. *ChemMedChem*, 2009. **4**(8): p. 1311-6.
140. De Vita, D., et al., Discovery of in vitro antitubercular agents through in silico ligand-based approaches. *Eur J Med Chem*, 2016. **121**: p. 169-180.
141. Goldmann, D., et al., Novel scaffolds for modulation of TRPV1 identified with pharmacophore modeling and virtual screening. *Future Med Chem*, 2015. **7**(3): p. 243-56.
142. Koeppen, H., Kriegl, J., Lessel, U., Tautermann, C.S. and Wellenzohn, B., *Ligand-Based Virtual Screening*, in *Virtual Screening*, H.K. R. Mannhold, G. Folkers and C. Sotriffer, Editor. 2011, John Wiley & Sons, Ltd. p. 61-85.
143. Ripphausen, P., B. Nisius, and J. Bajorath, *State-of-the-art in ligand-based virtual screening*. *Drug Discovery Today*, 2011. **16**(9): p. 372-376.

144. Eckert, H. and J. Bajorath, Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discovery Today*, 2007. **12**(5): p. 225-233.
145. Willett, P., *Similarity-based virtual screening using 2D fingerprints*. *Drug Discov Today*, 2006. **11**(23-24): p. 1046-53.
146. Muegge, I. and P. Mukherjee, *An overview of molecular fingerprint similarity search in virtual screening*. *Expert Opin Drug Discov*, 2016. **11**(2): p. 137-48.
147. Banegas-Luna, A.-J., J.P. Cerón-Carrasco, and H. Pérez-Sánchez, A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future Medicinal Chemistry*, 2018. **10**(22): p. 2641-2658.
148. Dilip, A., et al., Ligand-based virtual screening interface between PyMOL and LiSiCA. *J Cheminform*, 2016. **8**(1): p. 46.
149. Vuorinen, A. and D. Schuster, Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*, 2015. **71**: p. 113-134.
150. Bajorath, J., *Integration of virtual and high-throughput screening*. *Nat Rev Drug Discov*, 2002. **1**(11): p. 882-94.
151. Martin, Y.C., J.L. Kofron, and L.M. Traphagen, *Do structurally similar molecules have similar biological activity?* *J Med Chem*, 2002. **45**(19): p. 4350-8.
152. Breault, G.A., et al., Exploring 8-benzyl pteridine-6,7-diones as inhibitors of glutamate racemase (MurI) in gram-positive bacteria. *Bioorg Med Chem Lett*, 2008. **18**(23): p. 6100-3.
153. Kristensen, T.G., J. Nielsen, and C.N. Pedersen, *Methods for Similarity-based Virtual Screening*. *Comput Struct Biotechnol J*, 2013. **5**: p. e201302009.
154. Karatzas, E., et al., ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing. *Bioinformatics*, 2020. **36**(8): p. 2602-2604.
155. Cereto-Massague, A., et al., *Molecular fingerprint similarity search in virtual screening*. *Methods*, 2015. **71**: p. 58-63.
156. Heinke, R., et al., Virtual screening and biological characterization of novel histone arginine methyltransferase PRMT1 inhibitors. *ChemMedChem*, 2009. **4**(1): p. 69-77.
157. Simoben, C.V., et al., A Novel Class of Schistosoma mansoni Histone Deacetylase 8 (HDAC8) Inhibitors Identified by Structure-Based Virtual Screening and In Vitro Testing. *Molecules*, 2018. **23**(3).
158. Gruneberg, S., M.T. Stubbs, and G. Klebe, Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem*, 2002. **45**(17): p. 3588-602.
159. Enyedy, I.J., et al., Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem*, 2001. **44**(25): p. 4313-24.
160. Trosset, J.Y., et al., Inhibition of protein-protein interactions: the discovery of druglike beta-catenin inhibitors by combining virtual and biophysical screening. *Proteins*, 2006. **64**(1): p. 60-7.
161. Doman, T.N., et al., Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem*, 2002. **45**(11): p. 2213-21.
162. Sancineto, L., et al., Computer-aided design, synthesis and validation of 2-phenylquinazolinone fragments as CDK9 inhibitors with anti-HIV-1 Tat-mediated transcription activity. *ChemMedChem*, 2013. **8**(12): p. 1941-53.
163. Kannan, S., et al., Discovery of inhibitors of Schistosoma mansoni HDAC8 by combining homology modeling, virtual screening, and in vitro validation. *J Chem Inf Model*, 2014. **54**(10): p. 3005-19.
164. Marchand, J.R., et al., Discovery of Inhibitors of Four Bromodomains by Fragment-Anchored Ligand Docking. *J Chem Inf Model*, 2017. **57**(10): p. 2584-2597.
165. Gamba, E., et al., Identification of novel 2-benzoxazolinone derivatives with specific inhibitory activity against the HIV-1 nucleocapsid protein. *Eur J Med Chem*, 2018. **145**: p. 154-164.

166. Tintori, C., et al., High-throughput docking for the identification of new influenza A virus polymerase inhibitors targeting the PA-PB1 protein-protein interaction. *Bioorg Med Chem Lett*, 2014. **24**(1): p. 280-2.
167. Lyne, P.D., et al., Identification of Compounds with Nanomolar Binding Affinity for Checkpoint Kinase-1 Using Knowledge-Based Virtual Screening. *Journal of Medicinal Chemistry*, 2004. **47**(8): p. 1962-1968.
168. Chua, M.J., et al., Activity of bromodomain protein inhibitors/binders against asexual-stage *Plasmodium falciparum* parasites. *Int J Parasitol Drugs Drug Resist*, 2018. **8**(2): p. 189-193.
169. Roatsch, M., et al., Substituted 2-(2-aminopyrimidin-4-yl)pyridine-4-carboxylates as potent inhibitors of JumonjiC domain-containing histone demethylases. *Future Med Chem*, 2016. **8**(13): p. 1553-71.
170. Grabowski, M., et al., Identification of a pyrogallol derivative as a potent and selective human TLR2 antagonist by structure-based virtual screening. *Biochem Pharmacol*, 2018. **154**: p. 148-160.
171. Lionta, E., et al., Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem*, 2014. **14**(16): p. 1923-38.
172. Ripphausen, P., et al., Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J Med Chem*, 2010. **53**(24): p. 8461-7.
173. Sledz, P. and A. Caflisch, *Protein structure-based drug design: from docking to molecular dynamics*. *Curr Opin Struct Biol*, 2018. **48**: p. 93-102.
174. Verdonk, M.L., et al., Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci*, 2004. **44**(3): p. 793-806.
175. Wermuth, C.G., et al., *Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)*. *Pure and Applied Chemistry*, 1998. **70**(5): p. 1129.
176. Wermuth, C.G., Pharmacophores: Historical Perspective and Viewpoint from a Medicinal Chemist, in *Pharmacophores and Pharmacophore Searches*, H.K. R. Mannhold, G. Folkers, T. Langer, R.D. Hoffmann, Editor. 2006. p. 1-13.
177. Seidel, T., et al., The Pharmacophore Concept and Its Applications in Computer-Aided Drug Design. *Prog Chem Org Nat Prod*, 2019. **110**: p. 99-141.
178. Langer, T. and G. Wolber, *Pharmacophore definition and 3D searches*. *Drug Discov Today Technol*, 2004. **1**(3): p. 203-7.
179. Wolber, G. and T. Langer, LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*, 2005. **45**(1): p. 160-9.
180. Triballeau, N., H.O. Bertrand, and F. Acher, *Are You Sure You Have a Good Model?*, in *Pharmacophores and Pharmacophore Searches*, H.K. R. Mannhold, G. Folkers, T. Langer, R.D. Hoffmann, Editor. 2006. p. 325-364.
181. Yang, S.Y., Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today*, 2010. **15**(11-12): p. 444-50.
182. Schuetz, D.A., et al., *GRAIL: GRids of phArmacophore Interaction fieLds*. *J Chem Theory Comput*, 2018. **14**(9): p. 4958-4970.
183. Schaller, D., et al., *Next generation 3D pharmacophore modeling*. *WIREs Computational Molecular Science*, 2020. **10**(4): p. e1468.
184. Wieder, M., et al., Evaluating the stability of pharmacophore features using molecular dynamics simulations. *Biochem Biophys Res Commun*, 2016. **470**(3): p. 685-689.
185. Wieder, M., et al., Comparing pharmacophore models derived from crystal structures and from molecular dynamics simulations. *Monatsh Chem*, 2016. **147**: p. 553-563.
186. Wieder, M., et al., Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *J Chem Inf Model*, 2017. **57**(2): p. 365-385.
187. Muchtaridi, M., et al., Cytotoxicity Of Chalcone Of *Eugenia aquea* Burm F. Leaves Against T47D Breast Cancer Cell Lines And Its Prediction As An Estrogen Receptor Antagonist Based On Pharmacophore-Molecular Dynamics Simulation. *Adv Appl Bioinform Chem*, 2019. **12**: p. 33-43.
188. Schaller, D. and G. Wolber, PyRod Enables Rational Homology Model-based Virtual Screening Against MCHR1. *Mol Inform*, 2020. **39**(6): p. e2000020.

189. Langer, T. and R.D. Hoffmann, *Virtual screening: an effective tool for lead structure discovery?* *Curr Pharm Des*, 2001. **7**(7): p. 509-27.
190. Kaserer, T., et al., Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules*, 2015. **20**(12): p. 22799-832.
191. Groom, C.R., et al., *The Cambridge Structural Database*. *Acta Crystallogr B Struct Sci Cryst Eng Mater*, 2016. **72**(Pt 2): p. 171-9.
192. <https://www.chemdiv.com/complete-list-of-compounds-libraries>. (Accessed July 04, 2020).
193. <https://enamine.net>. (Accessed July 04, 2020).
194. <https://www.ibscreen.com>. (Accessed July 04, 2020).
195. <https://www.sigmaaldrich.com>. (Accessed July 04, 2020).
196. <http://www.princetonbio.com>. (Accessed July 04, 2020).
197. <https://www.caymanchem.com>. (Accessed July 04, 2020).
198. <http://www.maybridge.com>. (Accessed July 04, 2020).
199. <http://www.prestwickchemical.com>. (Accessed July 04, 2020).
200. <http://www.otavachemicals.com>. (Accessed July 04, 2020).
201. Irwin, J.J., et al., *ZINC: a free tool to discover chemistry for biology*. *J Chem Inf Model*, 2012. **52**(7): p. 1757-68.
202. Sterling, T. and J.J. Irwin, *ZINC 15--Ligand Discovery for Everyone*. *J Chem Inf Model*, 2015. **55**(11): p. 2324-37.
203. Hodos, R.A., et al., *In silico methods for drug repurposing and pharmacology*. *WIREs Systems Biology and Medicine*, 2016. **8**(3): p. 186-210.
204. Wishart, D.S., et al., *DrugBank: a comprehensive resource for in silico drug discovery and exploration*. *Nucleic Acids Research*, 2006. **34**(suppl_1): p. D668-D672.
205. Hoffmann, T. and M. Gastreich, The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discov Today*, 2019. **24**(5): p. 1148-1156.
206. <https://enamine.net/library-synthesis/real-compounds/real-space-navigator>. (Accessed July 04, 2020).
207. Lyu, J., et al., *Ultra-large library docking for discovering new chemotypes*. *Nature*, 2019. **566**(7743): p. 224-229.
208. Grebner, C., et al., Virtual Screening in the Cloud: How Big Is Big Enough? *J Chem Inf Model*, 2019.
209. Gloriam, D.E., *Bigger is better in virtual drug screens*. *Nature*, 2019. **566**(7743): p. 193-194.
210. Mullard, A., Supersized virtual screening offers potent leads. *Nat Rev Drug Discov*, 2019.
211. Clark, D.E., Virtual Screening: Is Bigger Always Better? Or Can Small Be Beautiful? *J Chem Inf Model*, 2020.
212. Bologna, C.G., O. Ursu, and T.I. Oprea, *How to Prepare a Compound Collection Prior to Virtual Screening*. *Methods Mol Biol*, 2019. **1939**: p. 119-138.
213. Lipinski, C.A., et al., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 1997. **23**(1): p. 3-25.
214. Oprea, T.I., *Property distribution of drug-related chemical databases*. *J Comput Aided Mol Des*, 2000. **14**(3): p. 251-64.
215. Walters, W.P., M.T. Stahl, and M.A. Murcko, *Virtual screening—an overview*. *Drug Discovery Today*, 1998. **3**(4): p. 160-178.
216. Baell, J.B. and G.A. Holloway, New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem*, 2010. **53**(7): p. 2719-40.
217. Baell, J. and M.A. Walters, *Chemistry: Chemical con artists foil drug discovery*. *Nature*, 2014. **513**(7519): p. 481-3.
218. Dahlin, J.L., J. Inglese, and M.A. Walters, *Mitigating risk in academic preclinical drug discovery*. *Nat Rev Drug Discov*, 2015. **14**(4): p. 279-94.
219. Senger, M.R., et al., Filtering promiscuous compounds in early drug discovery: is it a good idea? *Drug Discov Today*, 2016. **21**(6): p. 868-72.

220. Capuzzi, S.J., E.N. Muratov, and A. Tropsha, *Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay Interference Compounds*. J Chem Inf Model, 2017. **57**(3): p. 417-427.
221. Lagorce, D., et al., Pan-assay interference compounds (PAINS) that may not be too painful for chemical biology projects. Drug Discov Today, 2017. **22**(8): p. 1131-1133.
222. DeGoey, D.A., et al., Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. J Med Chem, 2018. **61**(7): p. 2636-2651.
223. Langer, T. and G. Wolber, Virtual combinatorial chemistry and in silico screening: Efficient tools for lead structure discovery? Pure and Applied Chemistry, 2004. **76**(5): p. 991.
224. Stumpfe, D. and J. Bajorath, Current Trends, Overlooked Issues, and Unmet Challenges in Virtual Screening. J Chem Inf Model, 2020.
225. <https://www.cas.org/products/scifinder>. (Accessed July 07, 2020).
226. <https://pubchem.ncbi.nlm.nih.gov>. (Accessed July 07, 2020).
227. Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape*. Cell, 2007. **128**(4): p. 635-8.
228. Eccleston, A., et al., *Epigenetics*. Nature, 2007. **447**: p. 395.
229. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
230. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone readers*. Nat Struct Mol Biol, 2012. **19**(12): p. 1218-27.
231. Yap, K.L. and M.M. Zhou, Keeping it in the family: diverse histone recognition by conserved structural folds. Crit Rev Biochem Mol Biol, 2010. **45**(6): p. 488-505.
232. Yun, M., et al., *Readers of histone modifications*. Cell Res, 2011. **21**(4): p. 564-78.
233. Taverna, S.D., et al., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol, 2007. **14**(11): p. 1025-1040.
234. Kelly, T.K., D.D. De Carvalho, and P.A. Jones, *Epigenetic modifications as therapeutic targets*. Nat Biotechnol, 2010. **28**(10): p. 1069-78.
235. Lundstrom, K., *What is the potential of epigenetics in drug development?* Future Med Chem, 2015. **7**(3): p. 239-42.
236. Yoon, S. and G.H. Eom, *HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases*. Chonnam Med J, 2016. **52**(1): p. 1-11.
237. Cattaneo, A. and M. Chirichella, *Targeting the Post-translational Proteome with Intrabodies*. Trends Biotechnol, 2019. **37**(6): p. 578-591.
238. Wang, W., et al., Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. EMBO Rep, 2011. **12**(11): p. 1160-6.
239. Yang, N., et al., Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17954-9.
240. Su, X., et al., Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. Genes Dev, 2014. **28**(6): p. 622-36.
241. Shanle, E.K., et al., Histone peptide microarray screen of chromo and Tudor domains defines new histone lysine methylation interactions. Epigenetics Chromatin, 2017. **10**: p. 12.
242. Wang, C., et al., Spindlin-1 recognizes methylations of K20 and R23 of histone H4 tail. FEBS Lett, 2018. **592**(24): p. 4098-4110.
243. <https://www.rcsb.org>. (Accessed September 9, 2021).
244. Zhao, Q., et al., Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation. J Biol Chem, 2007. **282**(1): p. 647-56.
245. Zhao, F., et al., Molecular basis for histone H3 "K4me3-K9me3/2" methylation pattern readout by Spindlin1. J Biol Chem, 2020. **295**(49): p. 16877-16887.
246. Du, Y., et al., Structural mechanism of bivalent histone H3K4me3K9me3 recognition by the Spindlin1/C11orf84 complex in rRNA transcription activation. Nat Commun, 2021. **12**(1): p. 949.

247. Bae, N., et al., Developing Spindlin1 small-molecule inhibitors by using protein microarrays. *Nat Chem Biol*, 2017. **13**(7): p. 750-756.
248. Fagan, V., et al., A Chemical Probe for Tudor Domain Protein Spindlin1 to Investigate Chromatin Function. *J Med Chem*, 2019. **62**(20): p. 9008-9025.
249. Xiong, Y., et al., Discovery of a Potent and Selective Fragment-like Inhibitor of Methyllysine Reader Protein Spindlin 1 (SPIN1). *J Med Chem*, 2019. **62**(20): p. 8996-9007.
250. Franz, H., et al., The histone code reader SPIN1 controls RET signaling in liposarcoma. *Oncotarget*, 2015. **6**(7): p. 4773-89.
251. Zhao, F., et al., Molecular basis for histone H3 "K4me3-K9me3/2" methylation pattern readout by Spindlin1. *J Biol Chem*, 2020.
252. Jiang, F., et al., Expression, purification, crystallization and preliminary X-ray analysis of human spindlin1, an ovarian cancer-related protein. *Protein Pept Lett*, 2006. **13**(2): p. 203-5.
253. Chen, X., et al., Suppression of SPIN1-mediated PI3K-Akt pathway by miR-489 increases chemosensitivity in breast cancer. *J Pathol*, 2016. **239**(4): p. 459-72.
254. Drago-Ferrante, R., et al., Suppressive role exerted by microRNA-29b-1-5p in triple negative breast cancer through SPIN1 regulation. *Oncotarget*, 2017. **8**(17): p. 28939-28958.
255. Song, Q., et al., miR-409 Inhibits Human Non-Small-Cell Lung Cancer Progression by Directly Targeting SPIN1. *Mol Ther Nucleic Acids*, 2018. **13**: p. 154-163.
256. Zhao, M., et al., SPIN1 triggers abnormal lipid metabolism and enhances tumor growth in liver cancer. *Cancer Lett*, 2020. **470**: p. 54-63.
257. Chen, X., Y.W. Wang, and P. Gao, SPIN1, negatively regulated by miR-148/152, enhances Adriamycin resistance via upregulating drug metabolizing enzymes and transporter in breast cancer. *J Exp Clin Cancer Res*, 2018. **37**(1): p. 100.
258. Gao, Y., et al., Spindlin1, a novel nuclear protein with a role in the transformation of NIH3T3 cells. *Biochem Biophys Res Commun*, 2005. **335**(2): p. 343-50.
259. Zhang, P., et al., Overexpression of spindlin1 induces metaphase arrest and chromosomal instability. *J Cell Physiol*, 2008. **217**(2): p. 400-8.
260. Yuan, H., et al., Overexpression of SPINDLIN1 induces cellular senescence, multinucleation and apoptosis. *Gene*, 2008. **410**(1): p. 67-74.
261. Fang, Z., et al., SPIN1 promotes tumorigenesis by blocking the uL18 (universal large ribosomal subunit protein 18)-MDM2-p53 pathway in human cancer. *Elife*, 2018. **7**.
262. Janecki, D.M., et al., SPIN1 is a proto-oncogene and SPIN3 is a tumor suppressor in human seminoma. *Oncotarget*, 2018. **9**(65): p. 32466-32477.
263. Wang, J.X., et al., SPINDLIN1 promotes cancer cell proliferation through activation of WNT/TCF-4 signaling. *Mol Cancer Res*, 2012. **10**(3): p. 326-35.
264. Greschik, H., et al., The histone code reader Spin1 controls skeletal muscle development. *Cell Death Dis*, 2017. **8**(11): p. e3173.
265. Choi, J.W., et al., Spindlin1 alters the metaphase to anaphase transition in meiosis I through regulation of BUB3 expression in porcine oocytes. *J Cell Physiol*, 2019. **234**(6): p. 8963-8974.
266. Sweis, R.F., et al., Discovery and development of potent and selective inhibitors of histone methyltransferase g9a. *ACS Med Chem Lett*, 2014. **5**(2): p. 205-9.
267. Wagner, T., et al., Identification of a small-molecule ligand of the epigenetic reader protein Spindlin1 via a versatile screening platform. *Nucleic Acids Res*, 2016. **44**(9): p. e88.
268. Silver, A., *The biology of cholinesterases*. 1974, Amsterdam; Oxford; North-Holland publ. comp.; New York: Elsevier.
269. Lane, R.M., S.G. Potkin, and A. Enz, *Targeting acetylcholinesterase and butyrylcholinesterase in dementia*. *Int J Neuropsychopharmacol*, 2006. **9**(1): p. 101-24.
270. Chatonnet, A. and O. Lockridge, *Comparison of butyrylcholinesterase and acetylcholinesterase*. *Biochem J*, 1989. **260**(3): p. 625-34.
271. Agatonovic-Kustrin, S., C. Kettle, and D.W. Morton, *A molecular approach in drug development for Alzheimer's disease*. *Biomed Pharmacother*, 2018. **106**: p. 553-565.

272. Lockridge, O., Review of human butyrylcholinesterase structure, function, genetic variants, history of use in the clinic, and potential therapeutic uses. *Pharmacol Ther*, 2015. **148**: p. 34-46.
273. Brimijoin, S., et al., *Physiological roles for butyrylcholinesterase: A BChE-ghrelin axis*. *Chem Biol Interact*, 2016. **259**(Pt B): p. 271-275.
274. Mesulam, M., et al., Widely spread butyrylcholinesterase can hydrolyze acetylcholine in the normal and Alzheimer brain. *Neurobiol Dis*, 2002. **9**(1): p. 88-93.
275. Fambrough, D.M., D.B. Drachman, and S. Satyamurti, *Neuromuscular junction in myasthenia gravis: decreased acetylcholine receptors*. *Science*, 1973. **182**(4109): p. 293-5.
276. Alward, W.L., *Medical management of glaucoma*. *N Engl J Med*, 1998. **339**(18): p. 1298-307.
277. Coyle, J.T., D.L. Price, and M.R. DeLong, *Alzheimer's disease: a disorder of cortical cholinergic innervation*. *Science*, 1983. **219**(4589): p. 1184-90.
278. Rafii, M.S. and P.S. Aisen, *Recent developments in Alzheimer's disease therapeutics*. *BMC Med*, 2009. **7**: p. 7.
279. Camps, P. and D. Munoz-Torrero, *Cholinergic drugs in pharmacotherapy of Alzheimer's disease*. *Mini Rev Med Chem*, 2002. **2**(1): p. 11-25.
280. <https://www.alz.org/alzheimers-dementia/treatments/medications-for-memory>. (Accessed July 27, 2020).
281. Munoz-Torrero, D., Acetylcholinesterase inhibitors as disease-modifying therapies for Alzheimer's disease. *Curr Med Chem*, 2008. **15**(24): p. 2433-55.
282. Kandiah, N., et al., Rivastigmine: the advantages of dual inhibition of acetylcholinesterase and butyrylcholinesterase and its role in subcortical vascular dementia and Parkinson's disease dementia. *Clin Interv Aging*, 2017. **12**: p. 697-707.
283. Tasso, B., et al., Quinolizidinyl derivatives of bi- and tricyclic systems as potent inhibitors of acetyl- and butyrylcholinesterase with potential in Alzheimer's disease. *Eur J Med Chem*, 2011. **46**(6): p. 2170-84.
284. Vezekov, L., et al., Galantamine-based hybrid molecules with acetylcholinesterase, butyrylcholinesterase and gamma-secretase inhibition activities. *Curr Alzheimer Res*, 2012. **9**(5): p. 600-5.
285. Tin, G., et al., Tricyclic phenothiazine and phenoselenazine derivatives as potential multi-targeting agents to treat Alzheimer's disease. *MedChemComm*, 2015. **6**(11): p. 1930-1941.
286. Gerlits, O., et al., A new crystal form of human acetylcholinesterase for exploratory room-temperature crystallography studies. *Chem Biol Interact*, 2019. **309**: p. 108698.
287. Nachon, F., et al., Crystal structures of human cholinesterases in complex with huprine W and tacrine: elements of specificity for anti-Alzheimer's drugs targeting acetyl- and butyryl-cholinesterase. *Biochem J*, 2013. **453**(3): p. 393-9.
288. <https://www.ncbi.nlm.nih.gov/books/NBK547868/>, LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; 2012-. Tacrine. . [Updated Jan 15, 2020] - (Accessed July 27, 2020).
289. Auld, D.S., *Metalloproteases*, in *Encyclopedia of Biological Chemistry (Second Edition)*, W.J. Lennarz and M.D. Lane, Editors. 2013, Academic Press: Waltham. p. 86-89.
290. Sexton, T., et al., Active site mutations change the cleavage specificity of neprilysin. *PLoS One*, 2012. **7**(2): p. e32343.
291. Oefner, C., et al., Structure of human neutral endopeptidase (Neprilysin) complexed with phosphoramidon. *J Mol Biol*, 2000. **296**(2): p. 341-9.
292. Bayes-Genis, A., J. Barallat, and A.M. Richards, *A Test in Context: Neprilysin: Function, Inhibition, and Biomarker*. *J Am Coll Cardiol*, 2016. **68**(6): p. 639-653.
293. Jhund, P.S. and J.J. McMurray, The neprilysin pathway in heart failure: a review and guide on the use of sacubitril/valsartan. *Heart*, 2016. **102**(17): p. 1342-7.
294. Schiering, N., et al., Structure of neprilysin in complex with the active metabolite of sacubitril. *Sci Rep*, 2016. **6**: p. 27909.

295. Szendroi, M., et al., On the presence of a metalloprotease in human skin fibroblasts that degrades the human skin elastic fiber system. *J Invest Dermatol*, 1984. **83**(3): p. 224-9.
296. Morisaki, N., et al., Nepriylsin is identical to skin fibroblast elastase: its role in skin aging and UV responses. *J Biol Chem*, 2010. **285**(51): p. 39819-27.
297. Nakajima, H., et al., Epithelial-mesenchymal interaction during UVB-induced up-regulation of neutral endopeptidase. *Biochem J*, 2012. **443**(1): p. 297-305.
298. Murphy, G. and H. Nagase, *Progress in matrix metalloproteinase research*. *Mol Aspects Med*, 2008. **29**(5): p. 290-308.
299. Chin, J.R. and Z. Werb, Matrix metalloproteinases regulate morphogenesis, migration and remodeling of epithelium, tongue skeletal muscle and cartilage in the mandibular arch. *Development*, 1997. **124**(8): p. 1519-30.
300. Benjamin, M.M. and R.A. Khalil, Matrix metalloproteinase inhibitors as investigative tools in the pathogenesis and management of vascular disease. *Exp Suppl*, 2012. **103**: p. 209-79.
301. Maskos, K. and W. Bode, Structural basis of matrix metalloproteinases and tissue inhibitors of metalloproteinases. *Molecular Biotechnology*, 2003. **25**(3): p. 241-266.
302. Grossman, M., et al., The intrinsic protein flexibility of endogenous protease inhibitor TIMP-1 controls its binding interface and affects its function. *Biochemistry*, 2010. **49**(29): p. 6184-92.
303. Kahari, V.M. and U. Saarialho-Kere, Matrix metalloproteinases and their inhibitors in tumour growth and invasion. *Ann Med*, 1999. **31**(1): p. 34-45.
304. Kahari, V.M. and U. Saarialho-Kere, *Matrix metalloproteinases in skin*. *Exp Dermatol*, 1997. **6**(5): p. 199-213.
305. Acharya, M.R., et al., Chemically modified tetracyclines as inhibitors of matrix metalloproteinases. *Drug Resist Updat*, 2004. **7**(3): p. 195-208.
306. Baker, A.H., D.R. Edwards, and G. Murphy, *Metalloproteinase inhibitors: biological actions and therapeutic opportunities*. *J Cell Sci*, 2002. **115**(Pt 19): p. 3719-27.
307. Raeeszadeh-Sarmazdeh, M., L.D. Do, and B.G. Hritz, Metalloproteinases and Their Inhibitors: Potential for the Development of New Therapeutics. *Cells*, 2020. **9**(5).
308. Verma, R.P., Hydroxamic acids as matrix metalloproteinase inhibitors, in *Matrix Metalloproteinase Inhibitors*. 2012, Springer. p. 137-176.
309. Benjamin, M.M. and R.A. Khalil, Matrix metalloproteinase inhibitors as investigative tools in the pathogenesis and management of vascular disease, in *Matrix metalloproteinase inhibitors*. 2012, Springer. p. 209-279.
310. Xiong, W., et al., Membrane-type 1 matrix metalloproteinase regulates macrophage-dependent elastolytic activity and aneurysm formation in vivo. *J Biol Chem*, 2009. **284**(3): p. 1765-71.
311. Sbardella, D., et al., Human matrix metalloproteinases: an ubiquitous class of enzymes involved in several pathological processes. *Mol Aspects Med*, 2012. **33**(2): p. 119-208.
312. Cui, N., M. Hu, and R.A. Khalil, *Biochemical and Biological Attributes of Matrix Metalloproteinases*. *Prog Mol Biol Transl Sci*, 2017. **147**: p. 1-73.
313. Heinz, A., et al., Degradation of tropoelastin by matrix metalloproteinases--cleavage site specificities and release of matrikines. *FEBS J*, 2010. **277**(8): p. 1939-56.
314. Schechter, I. and A. Berger, On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem Biophys Res Commun*, 1968. **32**(5): p. 898-902.
315. <http://www.drugdesign.com/web/teaching/glossary>. (Accessed August 3, 2021).

Page left intentionally blank.

7. APPENDICES

7.1. Full-text publications and Supporting Information

7.2. Book chapter: Lysine Reader Proteins

J. Bacher, D. Robaa, C. Luise, W. Sippl, M. Jung. In Epigenetic Drug Discovery, vol 74; 2019. Wiley-VCH. DOI:10.1002/9783527809257

IDENTIFICATION AND STRUCTURE–ACTIVITY
RELATIONSHIP STUDIES OF SMALL-MOLECULE INHIBITORS
OF THE METHYLLYSINE READER PROTEIN SPINDLIN1

<https://doi.org/10.1002/cmdc.201600362>

APPLICATION OF VIRTUAL SCREENING APPROACHES FOR
THE IDENTIFICATION OF SMALL MOLECULE INHIBITORS OF
THE METHYLLYSINE READER PROTEIN SPINDLIN1

https://doi.org/10.1007/978-1-4939-8630-9_21

EXPLORING AROMATIC CAGE FLEXIBILITY OF THE HISTONE
METHYLLYSINE READER PROTEIN SPINDLIN1 AND ITS
IMPACT ON BINDING MODE PREDICTION: AN IN SILICO
STUDY

<https://doi.org/10.1007/s10822-021-00391-9>

**STRUCTURE-BASED DESIGN, DOCKING AND BINDING FREE
ENERGY CALCULATIONS OF A366 DERIVATIVES AS
SPINDLIN1 INHIBITOR**

<https://doi.org/10.3390/ijms22115910>

SYNTHESIS AND BIOLOGICAL EVALUATION OF 8-HYDROXY-
2,7-NAPHTHYRIDIN-2-IUM SALTS AS NOVEL INHIBITORS OF
ACETYLCHOLINESTERASE (ACHE) AND
BUTYRYLCHOLINESTERASE (BCHÉ)

<https://doi.org/10.1039/c6md00647g>

DEGRADATION OF TROPOELASTIN AND SKIN ELASTIN BY
NEPRILYSIN

<https://doi.org/10.1016/j.biochi.2017.11.018>

MMP-14 DEGRADES TROPOELASTIN AND ELASTIN

<https://doi.org/10.1016/j.biochi.2019.07.001>

BOOK CHAPTER: LYSINE READER PROTEINS

<https://doi.org/10.1002/9783527809257.ch14>



CHIARA LUISE

◦ DETAILS ◦

luisechi@hotmail.it

◦ LANGUAGES ◦

Italian

English

German

◦ SKILLS ◦

- Schrödinger's Small Molecule Drug Discovery Suite
 - Live Design
 - Stardrop
 - MOE
 - Inte:Ligand
 - AMBER
- GOLD, AutoDock
- PyMOL, VMD
- KNIME, R, Pipeline Pilot
- Bash and Python scripting
- ChemDraw, ChemSketch, MarvinSketch
- OpenEye, VolSurf+, Moka, OpenBabel
- Operating Systems: Linux, MacOS, Windows
- Graphic design tools: Gimp, iMovie

◦ HOBBIES ◦

Cooking, Dog lover, Graphic design, Photography, Biking, Community Involvement

📁 EMPLOYMENT HISTORY

• **Computational Chemist at Drug Discovery Unit - School of Life Sciences, University of Dundee, UK**

September 2020 — Present

Research: apply computer-based methods to design and investigate novel inhibitors for the targets of interest within the DDU portfolios.

Other responsibilities: provide lectures and training, supervise Ph.D. students.

• **Research and Laboratory Teaching Assistant at Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, DE**

October 2015 — August 2020

Research: Structure- and ligand-based design and optimization of inhibitors of several targets.

Teaching Assistant: • "Protein Modeling and Simulation" course within the MSc program in Pharmaceutical and Industrial Biotechnology • "Quantitative Pharmaceutical Analysis" course within the MSc program in Pharmacy.

Other responsibilities: Supervise visiting researchers.

• **Laboratory Teaching Assistant at Department of Pharmaceutical Sciences, University of Perugia, IT**

October 2014 — December 2014

Teaching Assistant: "Molecular Modeling" and "Advanced Biomaterials" courses within the MSc program in Pharmaceutical Biotechnologies.

🎓 EDUCATION

• **Ph.D. study - Computational Chemistry, Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, DE**

October 2015 — August 2020

• **Master in Pharmacy (110/110 cum Laude), Department of Pharmaceutical Sciences, University of Perugia, IT**

October 2008 — February 2015

👥 INTERNSHIPS

• **Research Internship at Inte:Ligand - Software Entwicklungs und Consulting GmbH, Vienna, AT**

October 2017 — November 2017

Short-term scientific mission Grant: Cost Action CA15135.

• **Training in Molecular Modeling at Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, DE**

March 2015 — September 2015

Grant: Erasmus+ Mobility for Traineeship.

• **Experimental work in Computer-Aided Drug Design laboratory at Department of Pharmaceutical Sciences, University of Perugia, IT**

March 2013 — May 2014

Thesis: CDK9: working hypothesis for computer-aided inhibitors design.

★ **SCIENTIFIC CONTRIBUTION: INVITED TALK**

EUROPIN Summer School on Drug Design, University of Vienna, AT

September 2019

★ **SCIENTIFIC CONTRIBUTION: PUBLICATIONS**

- **C. Luise, D. Robaa, W. Sippl. Exploring aromatic cage flexibility of the histone methyllysine reader protein Spindlin1 and its impact on binding mode prediction: an in silico study.** *Journal of Computer-Aided Molecular Design*, 35, 695-706, 2021.
- **C. Luise, D. Robaa, P. Regenass, D. Maurer, D. Ostrovskiy, L. Seifert, J. Bacher, T. Burgahn, T. Wagner, J. Seitz, H. Greschik, K.S. Park, Y. Xiong, J. Jin, R. Schüle, B. Breit, M. Jung, W. Sippl. Structure-Based Design, Docking and Binding Free Energy Calculations of A366 Derivatives as Spindlin1 Inhibitor.** *International Journal of Molecular Sciences*, 22, 5910, 2021.
- **K.T. Shukur, T. Ercetin, C. Luise, W. Sippl, O. Sirkecioglu, M. Ulgen, G.P. Coskun, M. Yarim, M. Gazi, H.O. Gulcan. Design, synthesis, and biological evaluation of new urolithin amides as multitarget agents against Alzheimer's disease.** *Arch Pharm (Weinheim)*, 354(5), e2000467, 2021.
- **B. Noshadi, T. Ercetin, C. Luise, M.Y. Yuksel, W. Sippl, M.F. Sahin, M. Gazi, H.O. Gulcan. Synthesis, Characterization, Molecular Docking, and Biological Activities of Some Natural and Synthetic Urolithin Analogs.** *Chem Biodivers*, 17(8), e2000197, 2020.
- **D. Robaa, J. Melesina, C. Luise, W. Sippl. Structure-based design of epigenetic inhibitors.** In *Chemical Epigenetics - Topics in Medicinal Chemistry*; Springer International Publishing, 455-486, 2020.
- **N. Miekus, C. Luise, W. Sippl, T. Baczek, C.E.H. Schmelzer, A. Heinz. MMP-14 degrades tropoelastin and elastin.** *Biochimie*, 165, 32-39, 2019.
- **C. Luise, T. Heimburg, B. Karaman, D. Robaa, W. Sippl. Computer-based lead identification for epigenetic targets.** In *Epigenetic Drug Discovery*, vol 74; 2019. Wiley-VCH.
- **J. Bacher, D. Robaa, C. Luise, W. Sippl, M. Jung. Lysine Reader Proteins.** In *Epigenetic Drug Discovery*, vol 74; 2019. Wiley-VCH.
- **C. Luise, D. Robaa. Application of Virtual Screening approaches for the identification of small molecule inhibitors of the methyllysine reader protein Spindlin1.** In *Rational Drug Design. Methods in Molecular Biology*, vol 1824; 2018. Humana Press.
- **A.C.M. Huertas, C.E.H. Schmelzer, C. Luise, W. Sippl, M. Pietzsch, W. Hoehenwarter, A. Heinz. Degradation of tropoelastin and skin elastin by neprilysin.** *Biochimie*, 146, 73-78, 2018.
- **M. Schiedel, A. Fallarero, C. Luise, W. Sippl, P. Vuorela, M. Jung. Synthesis and biological evaluation of 8-hydroxy-2,7-naphthyridinium salts as potent inhibitors of acetylcholinesterase (AChE) and butyrylcholinesterase (BChE).** *MedChem-Comm.* 8, 465-470, 2017.
- **D. Robaa, T. Wagner, C. Luise, L. Carlino, J. McMillan, R. Flaig, R. Schüle, M. Jung, W. Sippl. Identification and SAR studies of Small-Molecule Inhibitors of the Methyllysine Reader Protein Spindlin1.** *ChemMedChem.* 11, 2327-2338, 2016.

★ SCIENTIFIC CONTRIBUTION: POSTERS

- C. Luise, D. Robaa, T. Wagner, M. Jung, W. Sippl. **Molecular Modeling Studies of Spindlin1 inhibitors: Investigation of Protein Flexibility and Pharmacophore Development.** 14th German Conference on Chemoinformatics, November 11-13, 2018, Mainz, DE.
- C. Luise, D. Robaa, T. Wagner, R. Schüle, M. Jung, W. Sippl. **Virtual Screening Based Identification of Novel Small Molecule Inhibitors of the Methyl-Lysine Reader Spindlin1.** EUROPIN Summer School on Drug Design, September 17 -22, 2017, University of Vienna, AT.
- C. Luise, D. Robaa, W. Sippl. **Docking and Binding Free Energy Calculations of Ligand Binding to L3MBTL3 Methyl-Lysine Reader Domain.** 11th European Workshop in Drug Design and Mu.Ta.Lig. Training School, May 21-26, 2017, Siena, Certosa di Pontignano, IT.
- C. Luise, D. Robaa, W. Sippl. **Binding Free Energy Prediction for Ligand Binding to L3MBTL3 Methyl-Lysine Reader Domain.** 3rd Freiburg Epigenetic Spring Meeting: Chemical Biology of Epigenetics, April 10-13, 2016, Freiburg, DE.

✿ AWARDS

- **Poster Prize - EUROPIN Summer School on Drug Design, University of Vienna, AT**
September 2017
- **Selected among the best graduates in the Department of Pharmaceutical Sciences - University of Perugia, IT**
October 2015

✿ EXTRA-CURRICULAR ACTIVITIES

- Member of the organizing committee of the "Pharma-Research Day" at Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, DE
- Member of EUROPIN (A European Pharmacoinformatics Initiative)
- Member of a DAAD bilateral project with the University of Ljubljana (DE/SI)

DECLARATION OF AUTHORSHIP

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift selbständig und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die aus ihnen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Die Arbeit wurde ausschließlich der Mathematisch-Naturwissenschaftlichen Fakultät der Martin-Luther- Universität Halle-Wittenberg vorgelegt und an keiner anderen Universität oder Hochschule weder im In- und Ausland zur Erlangung des Doktorgrades eingereicht.

28.11.2021



Chiara Luise