Universidad CEU San Pablo
CEINDO – CEU Escuela Internacional de Doctorado

PROGRAMA en CIENCIA Y TECNOLOGÍA DE LA SALUD

Computer-aided Drug Design based on Phosphatases (PTPRZ, PTPRGamma, PTP1B), Kinases (CK2) and Histone Deacetylases (HDAC1, HDAC6) as drug targets.

TESIS DOCTORAL

Presentada por:

Bruno Di Geronimo Quintero

Dirigida por:

Dra. Beatriz de Pascual-Teresa Fernández

Dra. Claire Coderch Boué

Madrid, 2019

Computer-aided Drug Design based on Phosphatases (PTPRZ, PTPRGamma, PTP1B), Kinases (CK2) and Histone Deacetylases (HDAC1, HDAC6) as drug targets.

Bruno Di Geronimo Quintero

Dra. Beatriz de Pascual-Teresa Fernández y Dra. Claire Coderch Boué, ambas pertenecientes a la Facultad de Farmacia de la Universidad San Pablo CEU,

CERTIFICAN:

Que la presente Memoria, titulada:

Computer-aided Drug Design based on Phosphatases (PTPRZ, PTPRGamma, PTP1B), Kinases (CK2) and Histone Deacetylases (HDAC1, HDAC6) as drug targets.

ha sido realizada bajo su dirección en el Departamento de Química y Bioquímica de la Universidad San Pablo CEU, por Bruno Di Geronimo Quintero y autorizan su presentación para ser calificada como Tesis Doctoral.

Montepríncipe, 28 de noviembre de 2019

Fdo. Dra. Beatriz de Pascual-Teresa Fernández

Fdo Dr.Claire Coderch Boué

VºBº
Directora del Departamento
Dña. Ana Ma Ramos González

List of original publications

The following publications have resulted to date from this thesis research work:

1. **Development of inhibitors of receptor protein tyrosine phosphatase β/ζ (PTPRZ1) as candidates for CNS disorders.** Pastor, M.; Fernández-Calle, R.; Di Geronimo, B.; Vicente Rodríguez, M.; Zapico, JM.; Gramage, E.; Coderch, C.; Pérez-García, C.; Lasek, AW.; Puchades-Carrasco, L.; Pineda-Lucena, A.; de Pascual-Teresa, B.; Herradón, G.; Ramos, A. *European Journal of Medicinal Chemistry* **2018**, *144*, 318-329.

2. **Pharmacological inhibition of Receptor Protein Tyrosine Phosphatase β/ζ (PTPRZ1) modulates behavioral responses to ethanol.** Fernández-Calle, R.; Vicente-Rodríguez, M.; Pastor, M.; Gramage, E.; Di Geronimo, B.; Zapico, JM.; Coderch, C.; Pérez-García, C.; Lasek, AW.; de Pascual-Teresa, B.; Ramos, A.; Herradón, G. *Neuropharmacology* **2018**, *137*, 86-95.

3. **Molecular Imaging Probes Based on Matrix Metalloproteinase Inhibitors (MMPIs).** Loganathan Rangasamy, Bruno Di Geronimo, Irene Ortín, Claire Coderch, José María Zapico, Ana Ramos and Beatriz de Pascual-Teresa. *Molecules* **2019**, *24*, 2982-3015.

This publication derives from the collaboration initiated with the CNIO through short stays during august.

4. **Disulfide Engineered Lipase to Enhance the Catalytic Activity: A Structure-Based Approach on BTL2.** César A. Godoy, Javier Klett, Bruno Di Geronimo, Juan A. Hermoso, José M. Guisán and César Carrasco-López. *International Journal of Molecular Sciences* **2019**, *20*, 5245-5261.

Acknowledgments

Firstly, I would like to acknowledge and express my sincerest appreciation to the University San Pablo CEU and CEINDO for giving me within the program the opportunity to execute and accomplish a Ph.D. study. Additionally, I want to demonstrate my gratitude to University San Pablo CEU for the financial support and aid of the FPI grant that provided a year's worth of doctoral study. Furthermore, I would like to recognize the Ministerio de Educación y Ciencia for the FPU scholarship which supported three years of educational research study.

Within the Doctorate program, I was provided with the invigorating opportunity to study four months abroad, in the Czech Republic, with the aid of a generous scholarship as well. Receiving scholarships such as these are very important for the development of doctoral dissertations not only within the educational realm of Spain, but also for global education. Moreover, these types of programs allow connection between both national and international research study groups, in my point of view, an important factor for higher education.

It is now my pleasure to convey my genuine appreciation to my advisors Professor Beatriz de Pascual-Teresa and Dr. Claire Coderch for their continuous support of my Ph.D. work and related research, for their motivation, persistence, patience, and profound, in-depth knowledge of this field. Their guidance facilitated my study and helped me in the writing of this dissertation. It would not have been possible without their constant encouragement and contribution. Their insightful comments and proposal of hard questions invigorated me to broaden and explore from various perspectives.

During these last four years, I want to thank my advisors for their patience and their collaboration in order to form a cohesive working environment. They were constantly by my side, revising scientific posters, helping me apply for scholarships and grants, as well as encouraging me to present and attend science symposiums.

Without their support, it would have not been possible to conduct and complete this research and I am indebted to them for their corroboration.

To continue, I would also like to thank my colleague Myriam Pastor for her continuous comradery and sharing our scientific doubts with one another. She taught me the most important lesson in science: “Read, read, read and read some more. Only then, can you begin to even start to think about it.”

I would like to give thanks to the Professor Ana Ramos for her support during our teaching lessons and the warmhearted breakfasts we shared together. I would also like to mention my appreciation for the helpfulness of the rest of my research group: Dr. José María Zapico, Dr. Irene Ortín, Dr. Regina Martín, and Dr. Loganathan Rangasamy.

Finally, I want to offer my gratefulness to all of the students with whom I have spent time within the laboratories. They will forever be memorable moments that have made these past four years treasured. Of course, it goes without saying, I must thank my parents, my brother and sister, my brother-in law, my niece, my future niece, my roommates, and my friends. It would have been unimaginable to complete this dissertation without their kindness, encouragement, and support. And, thank you to JRK for her providing me with English classes.

A mis hermanos.

“All models are approximations.
Essentially, all models are wrong, but some are useful.
However, the approximate nature of the model always be borne in mind.”

George E. P. Box

RESUMEN

Esta memoria está dividida en cinco capítulos.

En el capítulo I, se describen las principales herramientas computacionales utilizadas en este trabajo, su aplicación y su relevancia en el proceso de descubrimiento de fármacos, cuya importancia tanto en la industria farmacéutica como a nivel académico es innegable. Además, este capítulo describe brevemente las bases y principales aplicaciones de cada una de las técnicas utilizadas en diferentes etapas del desarrollo de esta tesis.

El capítulo II en primer lugar describe la relevancia de las fosfatasas en el descubrimiento de nuevos fármacos y su importancia como posibles dianas farmacológicas en diferentes enfermedades. En concreto, este capítulo está centrado en la estrategia del diseño de inhibidores del receptor de tirosina de membrana tipo Z o PTPRZ1 para el tratamiento del uso abusivo del alcohol y sus efectos reforzadores. El objetivo final de este trabajo ha sido el desarrollo de una nueva serie de pequeñas moléculas con capacidad de atravesar la barrera hematoencefálica, que presentan gran actividad y selectividad frente a PTPRZ1, para lo que el uso de las herramientas computacionales ha sido fundamental. La actividad de estas moléculas se basa en la inhibición directa de la actividad catalítica de PTPRZ1, con el fin de mimetizar la actividad de la pleiotrofina (PTN), una hormona cerebral implicada en enfermedades neurodegenerativas que es el ligando natural de esta fosfatasa. Como resultado, dos pequeñas moléculas **10a** y **12b** han sido seleccionadas por su alta actividad inhibitoria frente a PTPRZ1, con valores por debajo de micromolar, y por su capacidad de atravesar la membrana hematoencefálica en ensayos in vitro e in vivo. Además, estas pequeñas moléculas han sido ensayadas en otras fosfatasas con el objetivo de evaluar su de selectividad. Los prometedores resultados de dichas moléculas han permitido el ensayo directo en ratones, que ha corroborado tanto la implicación de PTPRZ1 en el mecanismo de adicción a drogas, como su potencial como diana farmacológica para este tipo de enfermedades. No sólo se ha visto una disminución en el consumo

de alcohol mediante la administración de estas pequeñas moléculas, sino que además se ha confirmado su efecto inhibidor en los mecanismos reforzadores del alcohol en ratones. Durante la realización de este trabajo, los métodos computacionales han sido utilizados para establecer modelos de unión ligando-receptor, los cuales no sólo han servido para explicar y racionalizar la actividad y selectividad de los compuestos desarrollados, sino que además, han servido de apoyo en el diseño de nuevas moléculas. Las técnicas de modelado por homología, docking, dinámica molecular, técnicas ab initio y de cálculo de energía libre han sido utilizadas en este capítulo para el estudio, desde un punto de vista estructural, del receptor PTPRZ1; así como del posible mecanismo de acción de las moléculas descritas por nuestro grupo de investigación.

En el capítulo III, se ha utilizado un nuevo enfoque en farmacología, conocido como “multidiana” La también denominada polifarmacología es un cambio en el paradigma clásico de la farmacoterapia, basado en la modulación de dos o más dianas por un mismo agente o molécula. En este capítulo, se describe el desarrollado de compuestos conocidos como inhibidores duales, ya que son capaces de inhibir dos enzimas implicadas en procesos tumorales: La Caseína Kinasa tipo 2 (CK2) y las histonas deacetilasas (HDAC) son enzimas que se encuentran relacionas entre si en una misma vía de señalización/activación. Es por ello que al actuar sobre ambas se espera una mayor respuesta farmacológica o sinergísmo. En este caso, el diseño racional de compuestos duales se ha basado en la unión de dos farmacóforos, uno selectivo de CK2 y otro de HDACs, a través de una cadena conectora en una misma molécula. El diseño, síntesis y modelado de estas moléculas ha permitido obtener compuestos con actividad en ambas enzimas del orden micromolar.

El capítulo IV resume la investigación llevada a cabo durante una estancia predoctoral en el Instituto de Química Orgánica y Bioquímica de Praga, el IOCB, bajo la supervisión del Dr. Lubomír Rulíšek. El trabajo realizado durante los cuatro meses de estancia, gracias a la concesión de una ayuda de movilidad de la CEINDO, está basado en la utilización de métodos de

química cuántica, en concreto el uso de la Teoría del Funcional de Densidad, más comúnmente conocida por sus siglas DFT, y su aplicación en el cálculo de perfiles energéticos en reacciones químicas y el modelado de pequeñas moléculas.

Se ha estudiado como ejemplo el cálculo del perfil energético teórico para la reacción de ciclopropanación de iluro de sulfonio. Posteriormente, esta misma metodología ha sido utilizada para el estudio de reacciones regioselectivas de alquilación en el TBB, uno de los inhibidores de CK2 descritos en el capítulo III y que forma parte de los fragmentos utilizados en la síntesis de inhibidores duales CK2/HDAC. Finalmente, la DFT se ha utilizado para la parametrización de los denominados “sigma-holes” en átomos de azufre y su implicación en interacciones intramoleculares en algunos de los inhibidores de PTPRZ1 descritos en el capítulo II de esta tesis.

Por último, el capítulo V constituye el inicio de una nueva línea de investigación dentro del grupo en el que se afronta otro enfoque novedoso dentro de la farmacoterapia, la degradación selectiva de las dianas farmacológicas. Son las denominadas moléculas PROTACs, cuyo nombre proviene del inglés “PROteolysis TArgeting Chimeras”. Estas moléculas bifuncionales se anclan selectivamente tanto a la diana que se desea degradar como a una ligasa selectiva que será la encargada de poliubiquitinar dicha diana. El objetivo final es marcar la diana farmacológica para que sea reconocida y pueda ser degradada por el sistema proteosómico. En este caso, se ha realizado un estudio del modo de unión de los PROTACs frente a CK2 recientemente descritos en la bibliografía. Nuestros estudios computacionales conducen a la propuesta de un modo de unión al sitio alostérico de la quinasa y nos han permitido proponer nuevas estructuras cuya síntesis se está llevando a cabo en la actualidad.

ABSTRACT

This thesis has been divided into five chapters.

Chapter I describes the main computational tools used in this work, their applications and their relevance in the drug discovery process, whose importance both in the pharmaceutical industry and academically is undeniable. In addition, this chapter briefly describes the bases and main applications of each technique used in the different stages of the development of this thesis.

Chapter II first describes the relevance of phosphatases in the discovery of new drugs and their importance as possible pharmacological targets in different diseases. Specifically, this chapter focuses on the strategy of the design of Receptor-type tyrosine-protein phosphatase zeta or PTPRZ1 for the treatment of alcohol use disorder and its rewarding effects. The final objective of this work has been the design of a new series of small molecules with the ability to cross the blood brain barrier, which present a great activity and selectivity against PTPRZ1. For this process, the use of computational tools has been fundamental. The mechanism of action of these molecules is based on the direct inhibition of the PTPRZ1 catalytic activity, in order to mimic the activity of pleiotrophin (PTN), a brain hormone involved in neurodegenerative diseases that is the natural ligand of this phosphatase. As a result, two small molecules **10a** and **12b** have been selected for their high activity against PTPRZ1, with values below micromolar, and for their ability to cross the blood brain membrane in in vitro and in vivo assays. In addition, these small molecules have been tested in other phosphatases in order to assess their selectivity profile. The promising results obtained with **10a** and **12b** have allowed testing these compounds directly in mice, which has corroborated the implication of PTPRZ1 in the mechanism of drug addiction, and its potential as a pharmacological target for this type of diseases. With this pharmacological assay we have not only observed a decrease in the alcohol consumption, additionally the inhibitory effect on alcohol-reinforcing mechanisms in mice has also been confirmed.

During this work, the computational methods have been used to establish ligand-receptor binding models, which have not only served to explain and rationalize the activity and selectivity of the developed compounds but have also served as support in the design of new molecules. Homology models, docking, molecular dynamics, ab initio and free energy calculation techniques have been used in this chapter to study, from a structural point of view, the PTPRZ1 receptor; as well as the possible mechanism of action of the molecules described by our research group.

In Chapter III, a new approach in pharmacology known as “multitarget” has been applied. This approach, also known as polypharmacology, constitutes a substantial change in the classical paradigm of the pharmacotherapy, based on the modulation of two or more targets by the same agent or molecule. In this chapter, the development of compounds known as dual inhibitors is described, since they are capable of inhibiting two types of enzymes involved in tumor processes: Casein Kinase type 2 (CK2) and histone deacetylase (HDAC). Those are two enzymes that have been related each other in the same signaling / activation pathway. That is why, when acting on both, a greater pharmacological response or synergism is expected. In this case, the rational design of dual compounds has been based on the linking of two scaffolds in the same molecule, one selective for CK2 and another for HDACs, through an appropriate connecting chain. The design, synthesis and modeling of these molecules has allowed to obtain compounds with activity in both enzymes at the micromolar range.

Chapter IV constitutes the initiation of a new line of research within the group in which another novel pharmacological approach is faced, the selective degradation of druggable targets. They are the so called PROTAC molecules, whose name stands for “PROteolysis TArgeting Chimeras”. These bifunctional molecules are selectively anchored to the target that is desired to be degraded and at the same time to a selective ligase that will be responsible for the polyubiquitinating of the target. The ultimate goal is to mark the pharmacological target that will be recognized and can be degraded by the proteasome system. In this case, a study of the binding

mode of a recently described in the literature PROTACs against CK2 has been carried out. Our computational studies lead to the proposal of a novel binding mode inside the allosteric site of the kinase and it has allowed us to propose new structures whose synthesis is currently being carried out.

Finally, Chapter VI summarizes the research carried out as part of a predoctoral visit at the Institute of Organic Chemistry and Biochemistry of Prague, the IOCB, under the supervision of Dr. Lubomír Rulíšek. The work done during the four months of stay, thanks to the CEINDO support and scholarship, was based on the use of quantum chemistry methods, specifically the application of the Density Functional Theory, more commonly known as DFT, the calculation of energy profiles in chemical reactions, and the modelling of small molecules.

The Sulfonium Ylide Cyclopropanation reaction was selected as a test example and the theoretical energy profiles for this reaction was carried out. Subsequently, this same methodology has been applied to the study of the regioselective alkylation in TBB, one of the CK2 inhibitors described in Chapter III, as one of the fragments used in the synthesis of dual CK2 /HDAC inhibitors. Finally, DFT has been used for the parameterization of the so-called “sigma-holes” in sulfur atoms and their involvement in the intramolecular interactions in some of the PTPRZ1 inhibitors described in Chapter II of this thesis.

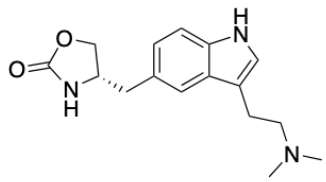
INDEX

CHAPTER I

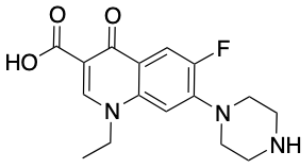
1. Introduction to Computer Aided Drug Design (CADD)

Drug design is an expensive and time-consuming process. It has been estimated that nowadays the price to develop and approve a new drug will be a stunning $2.6 billion (csdd.tufts.edu). The risk of drug discovery and development pipeline failure is high, and 90% of medicines entering clinical trials fail to obtain FDA approval and reach the consumer market. Approximately 75% of the price is due to drug discovery and design pipeline failures [1]. Given this huge number, many pharmaceutical companies are now looking for ways that can avoid these expenses through the application and development of novel CADD tools amongst others. The idea is simple (or not), trying to reduce costs and time by using computational tools. During the 1980s, the increasing availability of X-ray crystal structures, together with the improvement of computational resources created a special “cocktail” where pharmaceutical industries were able to apply CADD into the discovery of great number of compounds. The number, complexity and reliability in computational tools applied in drug discovery has increased since this period, and nowadays no one refuses the utility and place that CADD tools have in the development of new medicines. There are several examples of approved drugs that have been developed with the help of CADD; nevertheless, nowadays all pharmaceutical companies and the majority of the research groups have their own computational group that supports the drug development process at any stage. These well know examples are Zolmitriptan, Norfloxacin and Losartan.

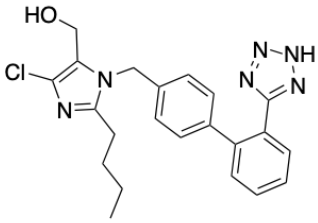
Zolmitriptan was obtained by Glaxo Wellcome R & D Ltd. using methysergide (5-HT receptor antagonist) as a lead compound and was further successfully developed as a migraine and headache treatment drug (Zomig®) after a series of active compounds were found (such as sumatriptan).



Zolmitriptan

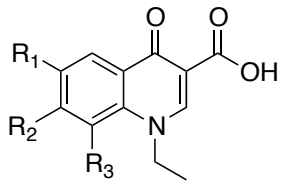


Norfloxacin



Losartan.

Norfloxacin, belongs to the class of fluoroquinolone antibiotics. It is used to treat urinary tract infections; it was discovered at Kyorin laboratories in Japan and subsequently licensed to Merck, patented in 1977 and approved for medical use in 1983. *Koga et al* outlined the tale of his discovery and the role of CADD in it [2]. A QSAR model was developed by the scientist of Kyorin laboratories for 6-, 7- or 8-monosubstituted compounds relating the antibacterial activity to steric parameters for R1 (Taft's Es parameter) and R3 (Verloop's B4 parameter). No connection was found at the R2 place for substituents, but it was noted that piperazine was a promising substituent. More specifically, the 6-fluoro-7-(1-piperazinyl) derivative was anticipated to be 10-fold more powerful than the de-fluoro analog. When this compound was synthesized and tested, the forecast was confirmed, as it was 16 times more powerful. On the basis of additional preclinical data, the compound was selected for clinical trials and subsequently became norfloxacin.



Losartan, is a drug mainly used to treat high blood pressure, belongs to the oral angiotensin II receptor inhibitors, and it was the first non-peptide to reach the market in 1995 marketed by Merck & Co and patented in 1986. Losartan was considered a selling blockbuster in 2006, reaching the 9th most prescribed medication in the United States, with more than 49 million prescriptions with sales of ~ US$3 billion in United States. This drug discovery process has been extensively described by *Duncia et al.*[3], where CADD have been the key point in the development of the biphenyltetrazole motif characteristic of the pharmacological group "sartans". For that purpose, the structure of the angiotensin II peptide was

solved and a ligand-structure alignment with the prototype molecule was carried out.

The four main fields developed in CADD are protein structure prediction, ligand-binding mode, ligand binding affinity and ligand ADME properties. All these methods have evolved substantially, giving rise to an independent research area and increasing the number of methods, algorithms and applications. Nevertheless, CADD works better when it is based on reliable experimental data and together with other fields such as medicinal chemistry, structural biology, and biochemistry. In this sense, a medicinal chemistry project needs the convergence of four main fields, organic synthesis, chemical biology, structural biology and computational chemistry (**Fig. 1.1**). Those must be in continuous contact and development to finally reach the desirable hit or lead molecule, which could modulate the proposed pharmacological target. Due to the extensive number of materials, methods and in silico tools available, here we present a summary and explanation of some of those methods applied in this thesis.

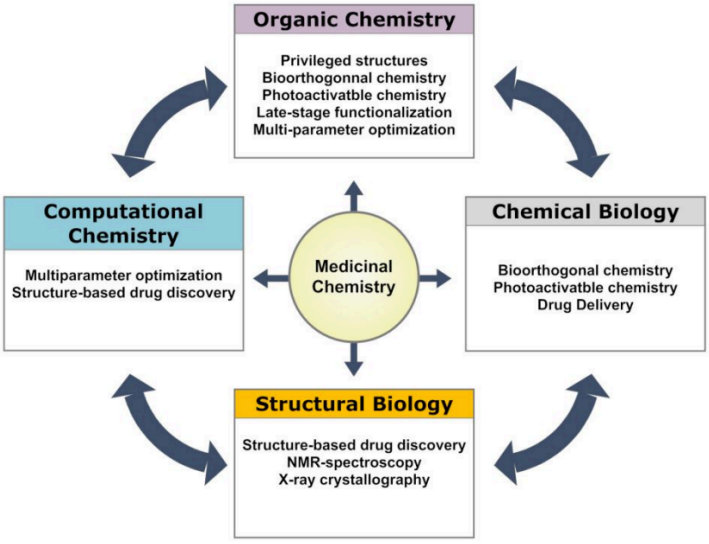


Figure 1.1. *Summarized scheme of a Medicinal Chemistry project in which the activities of the main involved areas are described.*

1.1. Materials.

1.1.1. PDB.

The Protein Data Bank (PDB) (www.rcsb.org) is a three-dimensional structural database of proteins and nucleic acids. The available data is usually obtained by X-ray crystallography (88.6%) or nuclear magnetic resonance (9.4%). Since 1971, of the 155830 structures that have been included, 29% are human and 6% come from Escherichia coli (**Fig. 1.2**). The principal asset is that all data collected in the PDB is under the public domain and can be used freely [4,5]. This has led to a huge number of references to this database (more than 100.000) and the opening of new frontiers and a research fields just based on the study of the structures, conformations, and chemical data. As an example of the relevance of structural biology, the number of Nobel Prizes related with three-dimensional structural determination or structural biology adds up to forty, including the 2017 Nobel Prize in Chemistry awarded to Jacques Dubochet, Joachim Frank and Richard Henderson for the development of cryo-electron microscopy. Note must be taken that not all deposited PDB structures are accurately refined and, therefore, they need to be thoroughly examined. In this line, the article *"Structures under Scrutiny"* published in Nature Chemical Biology, concludes that *"Structural biology advances have democratized access to biomolecular snapshots, but high standards must be maintained to ensure their utility"* [6]. Currently, in order to face this lack of accuracy on some deposited PDB structures, researchers have a large number of computational tools available, such as the open access PDB_Redo web server (pdb-redo.eu) or the molecular-graphics software Coot, that should be considered when analyzing any PDB structure [7–10].

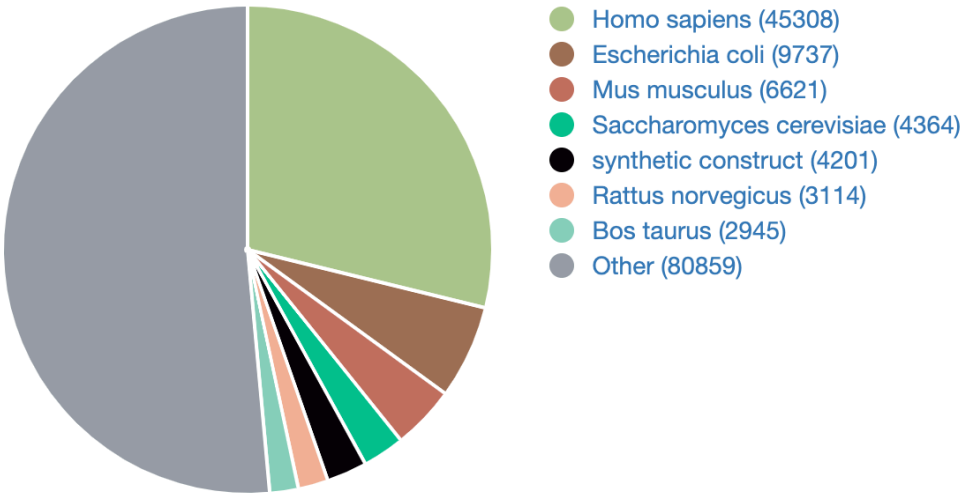


Figure 1.2. *Number of 3D structures available in the PDB database organized according to the organism they belong to.*

Particularly, on biomedical research, the significant interest of the protein-ligand complexes demands high-quality atomistic models, so the lack of accuracy is not an option [11]. To achieve this maximum standard, three main concepts must be taken into account: 1) the electron density pattern, 2) the specific stereochemistry of the small molecule and 3) the binding mode of the complex [12].

1.1.2. Chemical libraries

A chemical library is a repository that contains a wide variety of chemical compounds that represent the output (physical or virtual) of the drug development effort of certain companies or research groups. Generally, a chemical library contains a list of structures synthesized and physically available with a high degree of chemical diversity. However, there are many libraries with the molecules just virtually available [13]. The development of these chemical libraries has come along with the development of Virtual Screening campaigns that can be carried out, by means of docking techniques. The chemical libraries can be classified into two main groups: diverse, and targeted libraries. The first are based on random diversity while the second are based on one or various scaffolds with a variety of

modifications in order to explore possible derivatives [14,15]. Targeted libraries have increased in size due to the growth of chemical fragment libraries (Synthons) which can be attached to the backbone of the target molecule. One of the most remarkable free and public virtual libraries is ZINC15 (zinc15.docking.org) that contains more than 1.000 million compounds, divided into different groups according to their chemical availability, target, 3D structure or ADME properties among others [16].

Another example of free public libraries is The Open NCI Database (cactus.nci.nih.gov) that contains more than 260.000 structures issued from the NCI Development Therapeutics Program (DTP). Since its inception in 1955, the DTP has supported the development of more than 40 US-licensed anti-cancer agents through extensive collaborations with academic, pharmaceutical and biotechnological industries [17]. The interest of this Chemical Library is the free access to the virtual library as well as the benefit of physically ordering up to 30 compounds free of charge.

In addition, below are summarized other chemical libraries of interest. A full list of links can be obtained from www.click2drug.org.

- The MyriaScreen Diversity Collection from MERCK, with more than 10.000 compounds available virtually and physically (www.myriascreen.com).
- The Mu.Ta.Lig Virtual Chemotheca, which contains virtual compounds kindly provided by Mu.Ta.Lig participants who are the intellectual owners (chemotheca.unicz.it).
- The traditional Chinese medicine (TCM) database that is one of the biggest data set with natural products with more than 60.000 structures (tcm.cmu.edu.tw).

1.2. Methods

1.2.1. Homology Modelling.

In a drug discovery project, the ideal starting point for any researcher is to have the experimentally obtained 3D structure of the target protein (even

better if more than one is available), but it can be hampered by the lack of available 3D-structures. This situation has come about during the present work and the solution lies on homology modeling techniques. The term homology modeling refers to the modeling of a 3D structure using as template a known experimental structure of a homologous protein bearing approximately 30% of sequence identity. The methodology behind the homology model is quite simple: a known protein sequence has to be compared with the sequence of proteins with an experimentally available 3D-structure. This step is referred to as “template identification” and the result will be better when the amino acid sequence similarity between both the target and the template is higher. Generally, for a small sequence of 25 amino acids a 60% of identity is necessary, but for sequences of around 250 with just 20% identity would be enough [18]. This guideline depends on the target, and the target identification depends on such factors as the quality of the reported 3D structure, the presence of small molecules and the length of the sequence. Once the template has been selected, the next step is the sequence alignment with an alignment correction, that is, adding the missing or changed amino acids not present on the template. The secondary structure is generated from the backbone of the template bringing on the allowed dihedral ψ and φ angles to build the α-helixes and the β-sheets while the loops connecting them are then modelled. The side chains are latter modeled from the backbone and optimized in order to minimize the bad contacts or interactions.

One of the main challenges in protein structure prediction or modelling is the building of the loop regions [19]. A protein loop is a turn in the secondary structure where the polypeptide chain reverses its overall direction. For that reason, loops are one of the most flexible and variable positions of the protein. Moreover, loops present a highly variable amino acid sequence, which can often correspond to unaligned regions during the sequence alignment. In addition, loops usually correspond to solvent exposed zones of the protein, making the modeling of these areas extremely complex. For

this reason, standard modelling techniques are often not enough to reach this aim. Loops with less than 8 amino acids are commonly created by template alignment (Template Based Techniques) with similar known crystal structures, or even by simple geometry determination in the case of loops containing less than three amino acids. We must take in mind that by increasing the number of amino acids a loop is made of, the extent of the inaccuracy increases too [20]. In the case of more than 10 unresolved amino acids, de novo or ab initio techniques are applied (Non-Template Based Techniques). These methods are based on the statistical model to fill gaps generating the torsion angle pairs by revising the structure in order to maintain a realistic shape. This revision is based in the Ramachandran space of each amino acid and the available statistical data [21]. Few examples of this de novo techniques are softwares such as MODELLER, Loopy, and RAPPER, and the free web server Robetta (www.robetta.org) based on the software Rosseta and the Modeller-based Modloop (modbase.compbio.ucsf.edu/modloop) [22,23].

Once the homology model is built, the final step is the model validation. Albeit there is no reliable validation method, a good way to empirically validate models is to visually compare them with the template crystal structure. In addition, the common statistical parameter checked for assessing model quality is the Ramachandran plot, which shows the backbone distribution of the ψ and φ angles. This gives a graphical representation of the distribution of the conformational angles that shows how reasonable the folding and conformation of the protein is. However, depending on the applied methodology for this process, different quality parameters will be used for the model assessment.

One of the most relevant online servers used for homology modeling is SWISS-MODEL (https://swissmodel.expasy.org/). This server may build the model by automatically selecting the most accurate template based on sequence alignment with the target but can also build a model with a certain template provided by the user. In SWISS-MODEL, the model evaluation comes from two different estimators, GMQE and QMEAN. The GMQE

(Global Model Quality Estimation) is a quality estimation that combines properties from the target–template alignment and the template search method. The resulting number ranges from 0 to 1 where higher numbers indicate higher reliability of the model.

The QMEAN value is a composite estimator based on different geometrical properties that provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model [24]. The QMEAN values are in the range of -6 to 1 with zero being good. As a standard, values of -4.0 or below show low quality models but nevertheless the webpage itself indicates it by a “thumbs-up" or "thumbs-down" symbol next to the QMEAN score value whether we are faced or not with a good model [25].

A quite complete list of homology modeling softwares and web servers can be obtained from the web site: https://www.click2drug.org/

1.2.2. Docking Studies

Molecular docking is a computational method, that allows the prediction of the molecular recognition process more rapidly and at lower costs than other experimental methods such as X-Ray crystallography (but at the expense of accuracy), and it is probably the most heavily used tool in CADD [26]. Docking programs try to simulate the complementary relationship between ligand and protein in terms of volume and physical interactions. This process has been applied also to the study of protein-protein interactions (PPI), ligand-DNA/RNA interactions and even in materials.

Molecular docking addresses three questions: 1) finding the native binding pose of individual ligands in protein binding sites (the search algorithm); 2) estimate the relative binding affinity of each individual ligand (the scoring function); and, 3) discriminate ligands from non-binders (decoys). Despite all this, the main disadvantage of docking programs is related to accuracy and scoring, as in some cases interactions are overestimated.

Molecular Docking has been generally divided into three models:

- *Lock and Key Docking Model*, where receptor and ligand remain rigid during the calculations. It was the first and most simple developed method, but still useful in some cases, for example, to validate docking software (Re-docking) [27].
- *Ensemble Docking Model* is one of the most popular methods due to the equilibrium between accuracy and expensiveness of the calculation. It is based on the generation of an ensemble of ligand or receptor conformations usually by means of molecular dynamics algorithms. The most common are the generation of all the possible conformations of the flexible ligand within the rigid protein in order to obtain every possible receptor-ligand interaction [28–32].
- *Induced-Fit Docking Model*, is the most computationally expensive method because receptor and ligand are flexible during the calculations, thus creating a huge quantity of binding possibilities. This method is a hot-topic on software development due to the high standard results that are obtained, but still remains too computationally demanding or either does not provide enough conformation sampling [33–36]. By adding flexibility and degree of freedom to either the torsional angles of the amino acids side chains or the backbone of the receptor, we will approach the induced fit docking [37,38].

Another recent docking technique is the Dynamic Undocking (DUck), a computational procedure developed by *Xavier Barril* from the University of Barcelona [39]. This method calculates the work needed to break a key native contact obtaining a quasi-bound state (WQB), giving a tight relationship between WQB and the binding affinity (ΔG_{bind}). This methodology has shown its efficacy and usefulness in parallel with virtual screening techniques in the decoy discrimination (false positives) from a pull of docking hits. This is due to the fact that key hydrogen bonds are labile interactions for most of the docking decoys [40].

1.2.5.1. The search algorithm.

The search algorithm carries out the positioning of the molecule by modifying the location, conformation, and orientation in the selected volume of the protein. In the simplest model, where both (ligand and protein) are rigid, six degrees of translation and rotational freedom are considered. Then, the search algorithm checks all possible ligand positions and, together with the scoring algorithm, the best positions from the whole pull will be selected [41].

Monte Carlo is the most commonly used algorithm and is based on a random number generator that in this case chooses a position or orientation randomly. The process is simple: a pull of cartesian coordinates are chosen randomly, in which different orientations of the docked molecule are checked with those initial positions. The process restarts by choosing other positions, which will create other ligand orientations totally different from the first attempt. This process would be endless, but the volume and number of iterative attempts is previously defined in the input parameters. The main goal will be the exploration with the ligand of the maximum defined space of the protein.

Additionally, the original Monte Carlo algorithm can be modified by being coupled with other search methods and algorithms, such as the taboo search, simulated annealing, and the genetic algorithms [42].

- *The taboo algorithm* is based on the fundamental premise of not repeating a random position obtained by Monte Carlo by keeping track of what positions have been tested before.
- *The Monte Carlo simulated annealing* is based on random changes made in the ligand's current position, orientation, and conformation; then, the energy of the new position of the ligand is compared to its predecessor. If the new energy is lower than the previous one (which implies a better calculated affinity), this new position is immediately accepted, thus continuing the calculation at this new point.

Nonetheless, if the new position presents higher energy values than the previous one it will be rejected.

- *The genetic algorithm* is based on the cornerstone of modern biology, the natural selection. That is, from a pull of a diverse population, just the individuals with specific features will be selected to generate the next generation that will be selected again by the same specific feature and so on. In a genetic algorithm-based docking, a pull of conformations/rotations will be obtained randomly and those who fit the restriction (high score, for example) are selected to generate another pull.

1.2.5.2. The scoring functions

This function produces a scored quantitative energetic value calculated by the docking program depending on the different pose and conformation obtained by the search algorithm for the ligand in the defined volume of the target protein. As for the search algorithm, there are many types of scoring functions that range from the simplest, which calculates how well the ligand fits, to the more complex that include the entropy and enthalpy factors. The general equation to obtain the binding free energy of the ligand can be made up of various additive factors:

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib}$$

- ΔG_{bind} : free binding energy, total scoring.
- $\Delta G_{solvent}$: energy contribution due to the solvent effect that is calculated as the balance of the receptor-ligand interactions with the solvent and the respective receptor-solvent and ligand-solvent interactions.
- ΔG_{conf} : energy contribution of the conformational changes in the ligand and receptor obtained during the docking.
- ΔG_{int} : energy contribution obtained directly from the ligand-receptor interactions. This factor is obtained by summarizing the ΔG_{vdW} (the van der Walls energy) and the ΔG_{elect} (the electrostatic energy) of the interactions of the complex.

- ΔG_{rot}: energy derived from the entopic loss brought about by the loss of movement of the ligand upon binding.
- $\Delta G_{t/r}$: the energy derived from the loss of translational and rotational angles upon binding.
- ΔG_{vib}: the energy obtained by the changes in the vibrational mode, commonly avoided due to its complexity of calculation.

There is a relationship between de docking score (calculated binding energy) and the inhibition rate constant (K_i) which should be proportional. The simple Arrhenius equation (**Fig. 1.3**) could correlate both factors (K_i and E_a = scoring), so depending on the applied scoring function we can obtain a linear correlation or just a qualitative scoring of the molecules.

It always depends on the system and the portfolio of molecules. However, in any case, the docking score cannot be related to any cell assay or *in vivo* test.

$$K = Ae^{-E_a/RT} \implies \ln k = -\frac{E_a}{RT} + \ln A \implies \ln k = -\frac{E_a}{R}\left(\frac{1}{t}\right) + \ln A$$

K = Rate constant ($L \cdot mol^{-1} \cdot s^{-1}$)
A = Frequency Factor ($L \cdot mol^{-1} \cdot s^{-1}$)
E_a = Energy Activation ($Kcal \cdot mol^{-1}$)
R = Gas Constant (1.987×10^{-3} $Kcal \cdot K^{-1} \cdot mol^{-1}$)
T = Temperature (K)

Straight line plot ($y = mx + b$) for ln k versus $1/t$, where the slope is $-E_a/R$.

Figure 1.3. *Arrhenius equation*

1.2.5.3. The Grid

The grid is a selected and finite volume of the receptor where the molecule to be docked will be positioned and the docking score will be calculated. Docking software use pre-calculated interaction energy grid maps between the macromolecule such as a protein, DNA or RNA; and various atom probes of different atom types, such as aliphatic carbons, aromatic carbons, hydrogen bonding oxygens, and so on. The grid-based docking method first centers a cubic box in the area of the target in which the calculation is to be carried out to define the space. Then, this area is further divided by a 3D

grid that gives rise to the “nodes” or “grid points” that are separated by a distance called “grid spacing” (**Fig. 1.4**). The different atom probes are subsequently positioned on each grid point, and the potential interaction energy (both electrostatic and van der Waals) of the interaction of the probe with the protein in that grid point is calculated and stored in the corresponding grid file. This pre-calculation of the interaction energies allows faster calculations as the non-bonded lists do not need to be updated during the calculation. The total interaction energy of the ligand with the macromolecule will be the sum of all the precalculated energetic contributions.

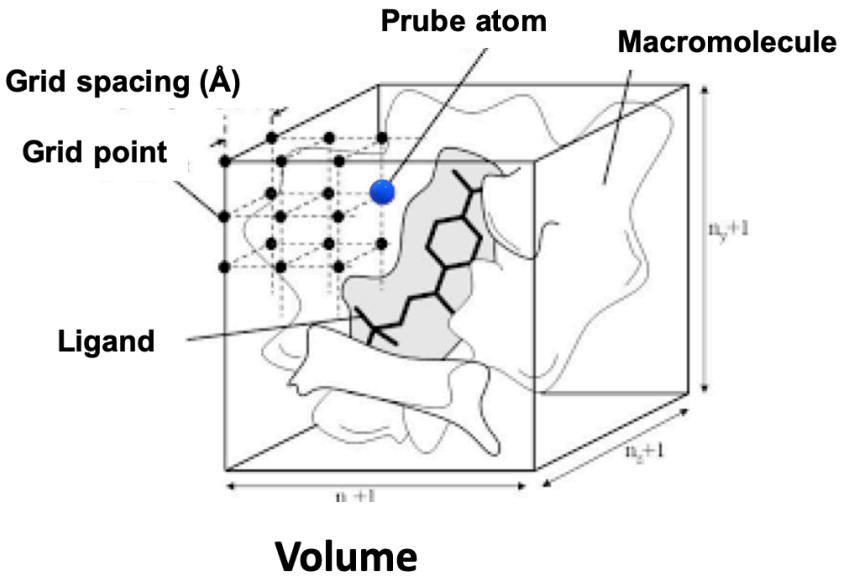


Figure 1.4 *Schematic representation of a GRID in docking. Image adapted from www.csb.yale.edu*

1.2.3. Hight Throughput Virtual Screening

The main goal of a Hight Throughput Virtual Screening (HTVS), is to differentiate from a wide set of small molecules or chemical library, those which can bind and fit better (hits) into the selected pocket or cavity of the receptor or receptors, from those that do not bind (decoys). The idea is quite simple but could become challenging when the set of compounds groups more than thousands of molecules, and the results have to be reliable and

obtained in a short period of time. This method has come as a theoretical alternative to the very expensive and time-consuming experimental High Throughput Screening (HTS) method. Nowadays, with the development of CPUs and computer code, which enables the user to parallelize calculations, is possible to carry out a HTVS within months, depending on the number of receptors as well as number of ligands to dock. Balancing between the accuracy of the results and the computational time is one of the key points in HTVS development. A HTVS protocol is based on sequential filters that, by increasing their computational complexity, eliminate possible decoys until obtaining a suitable number of potential candidates that are then forwarded to analyze their activity in biological assays. The schematic representation of a HTVS looks like a funnel, the big mouth represents the entire chemical library that is reduced in number by the consecutive filters ending with the smallmouth which are the selected final candidates. (**Fig. 1.4**).

HTVS has become an essential tool in drug discovery, especially in pioneering projects when there are no reported hits. This is the reason why HTVS has become a crucial technique in the pharma-industry with an outstanding increase [43]. There are two main types of HTVS:

- Ligand-based Virtual Screening (LBVS). This scenario comes when a group of molecules (set) with a known structure, present (or not) activity against a specific target. By using the collective information contained in the set of molecules it is possible to create a 3D pharmacophore model within a candidate set of molecules that can be compared in order to predict its activity. In order to use this methodology, it is not mandatory to know the 3D structure of the target, but to provide the pharmacophore model with enough structural diversity and activity data. The computational cost of this method is reduced, once the pharmacophore model has been built. This approach is based on similarity (chemical and structural) [44,45].

- Structure-Based Virtual Screening (SBVS). In this case, the 3D structure of the protein target is well-known, and the chemical library is screened by a set of sequential docking filters. It is the most computationally expensive method due to the high number of docking processes to be performed that can be queried by the parallel cluster.

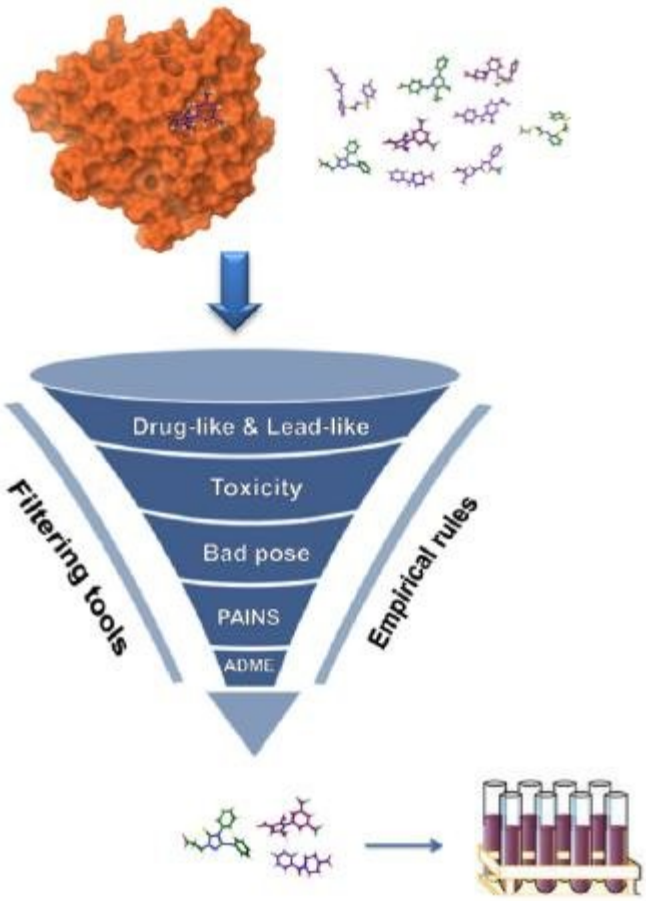


Figure 1.4. *Schematic representation of VS showing the different filters applied. Image taken from J.D.D.T. 2018, 8, 504-509.*

1.2.4. Molecular Dynamics

A crystal structure is an electronic density model. Molecular Dynamics (MD) is a computational tool to transform this rigid crystal structure into a mobile system by applying classical physics rules. During an MD simulation, the atoms are allowed to move and interact with each other for a short period of time, giving a view of the dynamic processes of the system. In the canonical version, trajectories of the atoms are determined by numerically

solving Newton's equations, where forces between the particles and their potential energies are calculated using interatomic potentials and Molecular Mechanics (MM) force fields. Although this is not as simple idea, in 1975 the first MD simulation of a simplified biological folding process was published in Nature [46]. A year later, a biological process was published where the way for understanding protein motion was studied as essential in function [47]. MD simulations have also been applied for the prediction of the molecular basis of the Gaucher disease due to the effect of the most common mutation N370S; or, for probing the mechanism of Tobacco Mosaic virus assembling where the required presence of RNA was confirmed, amongst other achievements [48,49]. MD is not only applied to biological systems, it has also been able to describe material properties, chemical reactions and physical processes. As part of this success, in 2013 Martin Karplus, Michael Levitt and Arieh Warshel were awarded the Novel price in Chemistry “for the development of multiscale models for complex chemical systems” (www.nobelprize.org). In the 1970s these three researchers stablished the basis of the powerful programs that are used nowadays to understand and predict chemical reactions and biological processes. Computer models mirroring real life have become crucial for most advances made in chemistry today. MD is based on the principles of classical mechanics and is one of the most used methodologies for the simulation of biological macromolecules. This is mainly because of its reliability and reduced computational cost compared to other computational methods such as DFT or QM/MM which will be discussed on the next section (**Fig. 1.5**).

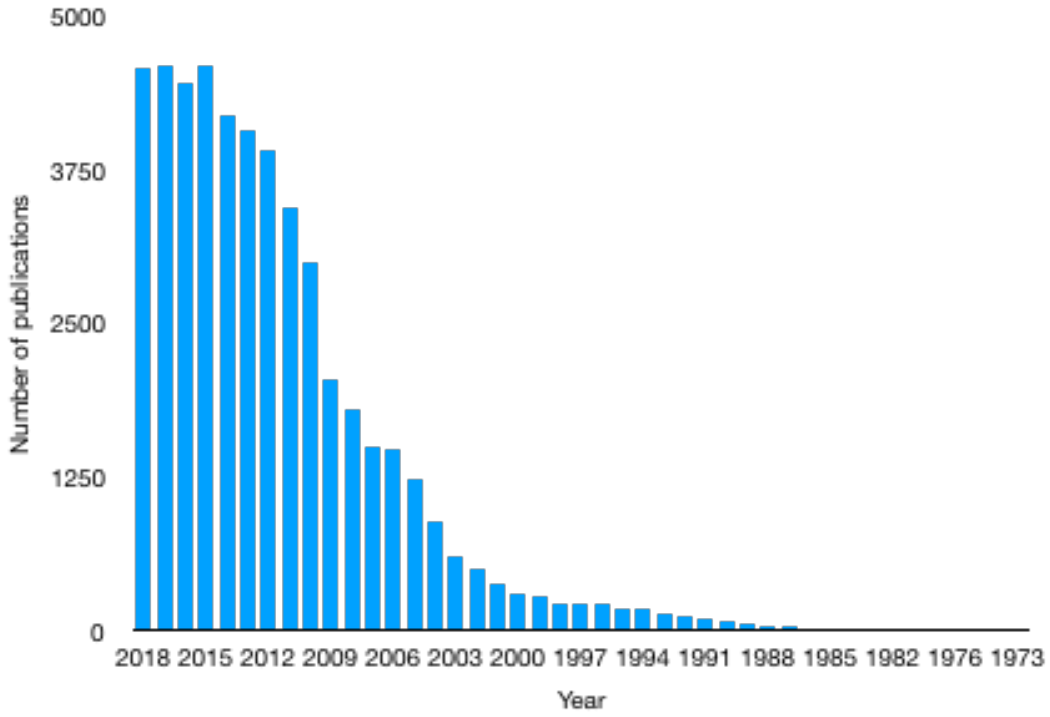


Figure 1.5 *Number of publication pear year from 1973 to 2018 which includes the key word "Molecular Dynamics". The data was obtained from PubMed database.*

MM is based on the Born–Oppenheimer approximation (BO)[50]. This approximation allows the wavefunction (Ψx,t) of a molecule to be simplified into its electronic and nuclear (vibrational, rotational) components. In MM, as the nucleus of the atom is much bigger than electrons, the total energy of the molecule is only considered by the function coming from the nucleus. This simplification allows the system to be described by the potential energy surface that is calculated with classical laws of physics. This final function is called force field (ff) and is based in these assumptions:

1) The atom (including electrons and nucleus) is considered to be a perfect sphere.

2) Bonds between atoms are calculated by harmonic potentials (springs).

3) Potential functions, force constant and equilibrium values come from experimental data, parameters calculated by quantum chemistry, or both.

4) Total potential energy comes from the summary of electronic,

vibrational, rotational and nuclear values (bonding and non-bonding values).

Molecular ffs allow the calculation of the potential energy of a system by dividing it into the bonded and non-bonded terms [51]. Bonded terms are defined by the interactions between linked atoms, and non-bonded or noncovalent interactions are described by the long range electrostatic and van der Walls forces (**Fig. 1.6**). In summary, ffs define a set of parameters for different types of atoms, usually obtained empirically. These parameters include values for atomic mass, van der Waals radius, and partial charge for any individual atoms, plus the bonded and non-bonded terms.

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$$

$$E_{\text{nonbonded}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}$$

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$$

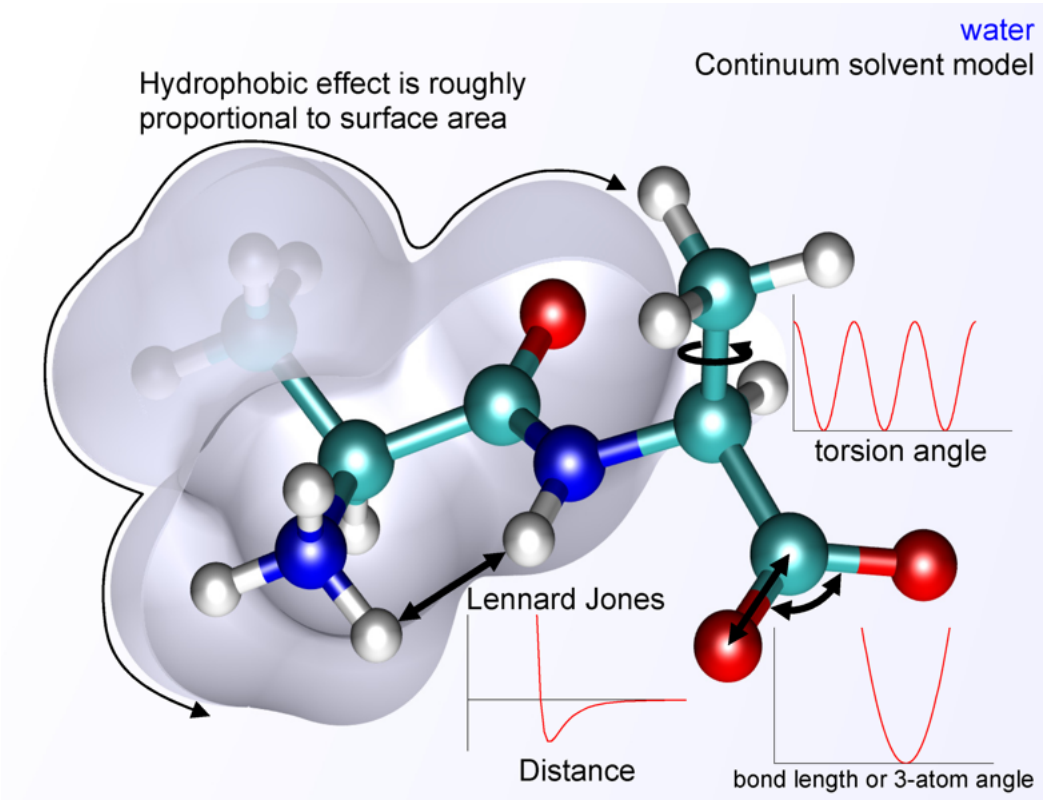
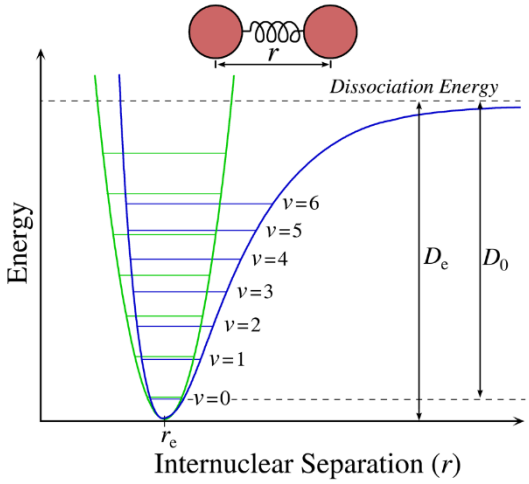


Figure 1.6. *Schematic representation of the most relevant parameters in a ff of a double alanine peptide represented in an implicit solvent model.*

1.2.5.4. The bonded term

- **The bond-stretching term.** All bonds have an optimal length that will be considered as the reference value, that outside the event of a chemical reaction, it will not deviate from the standard value. The bond-stretching energy is defined by the Morse potential obtained by the simplistic approach of Hooke's law formula, where K is the force constant, l is the bond length and l_0 is the equilibrium bond length adopted in a minimum energy structure.

$$Energy_{bond} = \frac{K}{2}(l - l_0)^2$$



- **The angle bending.** As for bonds, angles formed by three atoms in a plane also have an optimal value that corresponds to the minimum energy structure. The same formula from Hooke's law in which lengths are substituted by angles can be used to define the angle-bending energy.

$$Energy_{angle} = \frac{K}{2}(\emptyset - \emptyset_0)^2$$

- **The torsional term.** Unlike bonds and angles that rarely vary from their equilibrium values, the main conformational changes in a molecule come from the rotation about bonds. This term defines the energy variation brought about by the reorientation of two atoms (a and d) that are part of a dihedral angle (a-b-c-d). The energetic profile is expressed as a Fourier

series, where V_n is the maximum value of the torsion angle between *b-c; τ* is the current angle; and, δ the angle phase:

$$Energy_{torsional} = \frac{V_n}{2}(1 + \cos(\tau + \delta))^2$$

1.2.5.5. The non-bonded term

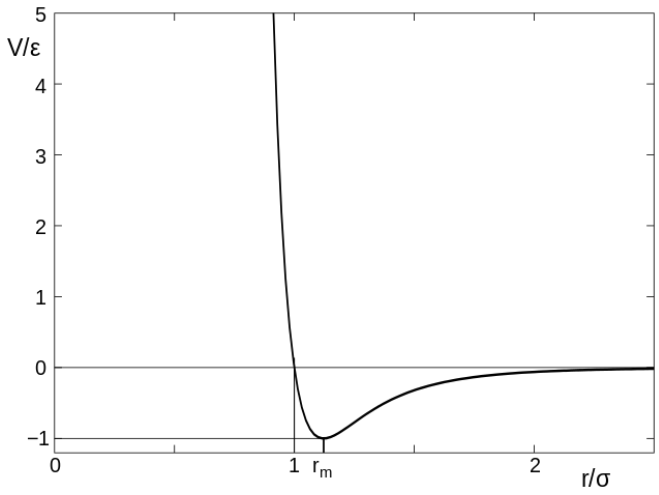
- **The van der Waals term**. This term is usually computed with the Lennard-Jones (LJ) potential which balances between the distance-dependent attractive and repulsive forces between two atoms or surfaces. The main features of the van der Walls forces are:

1. Additive and unlimited.
2. Non directional.
3. Weaker than normal covalent and electrostatic forces.
4. Distance dependent in a sense that the closer the molecules are the stronger the interactions.
5. Independent of temperature.

The simplified Lennard-Jones potential formula used by some simulation software is:

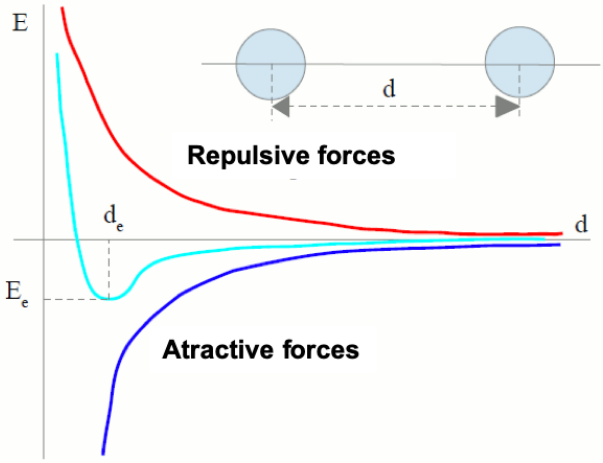
$$van\ der\ Waals_{Lennard-Jones} = \frac{a_{AB}}{r_{AB}^6} - \frac{b_{AB}}{r_{AB}^{12}}$$

Where *a* and *b* are specific constants (*a* describes the attractive long range and *b* Pauli repulsions) of the computed atoms *AB*, and r the distance between each other. The general plot of any (LJ) potential shows how the energy between two atoms tends to be cero (even negligible at long distances) and it decreases until the minimum r_m where both atoms present the higher attractive force. From this point if the distance decreases, the energy increases exponentially because of the repulsion due to the overlapping of the van der walls radii.



The electrostatic term. The electrostatic interactions between two molecules come from the added effect of the different atomic partial charges and the energy can be calculated by Coulomb's law. In the formula, $qA \cdot qB$ represent the partial charges of atoms *A* and *B*; ε_0 the corresponding dielectric constant; and, *r* the corresponding distance.

$$E_{electrostatic} = \frac{qA\ qB}{4\pi\epsilon_0\ r_{AB}}$$



The general plot of Coulomb's law is dependent of the distance and charge of the molecules. If both molecules are close enough and have the same partial charges only repulsive interactions will occur. On the contrary, if partial charges are of opposed sign attractive energies will arise. For any residue or molecule, the atomic partial charges are calculated by assigning a charge (positive or negative) to the set of atoms that compose the entire molecule, thus representing the global electrostatic property.

1.2.5.6. Parametrization

In addition to the mentioned term, the ff includes a set of parameters for different atoms types, chemical bonds, dihedral angles and van der Walls radius. These data come from different sources such as structural data, electronic and energetic values, and from ab-initio calculated values. For any ff, the obtaining and correct analysis of all this data demands a huge effort. The main goal of the ff is to simulate a model (protein, DNA, polymers, lipids, etc.) closest to the empirical values. To this aim they all have a period of validation with the empirical data available. Nowadays there are many developed ffs, each one with their advantages and disadvantages, usually focused on the simulation of a specific system. The first ff appeared in the 1960's, with the development of the MM method. It was developed to predict molecular structures, vibrational spectra and enthalpies of isolated molecules of small organic molecules [52,53]. The robustness and efficiency of this MM potential allowed it to survive and to be further developed, as an example the MM potentials developed by Allinger's group MM2, MM3 and MM4 [54–56]. Continuous software development is required to study new and larger macromolecules and lead to the creation of more widely applicable ffs. One of the classical examples is the Universal Force Field (UFF)[57] that contains parameters for all the atoms in the periodic table. The most applied ffs are the developed by CHARMM[58], AMBER[59], GROMOS[60], OPLS[61], and COMPASS[62], which have been continuously developing and updating the parameters. During the development of this thesis, AMBER has been extensively used for MM and MD simulations, for this reason we are going to focus on this ff. The evolution started with ff94, followed by ff99 and subsequently refined by *Duan et al*. [63] who introduced a more extensive modification of ff94/ff99 (called ff03), in which a fundamentally different concepts for the derivation of partial atomic charges were used. To improve the existing ff94 and ff99 force fields, reparameterization of the backbone dihedrals, which improved conformational preferences for typical secondary structures lead to the

modified force field ff99SB [64]. Later on, in 2015 the Amber ff14SB force field improved secondary structure content in small peptides and reproduction of NMR protein interactions. The protein secondary structure balance and dynamics was taken from the previous ff99, but weaknesses inside chain rotamer and backbone secondary structure was recalculated by multiple small adjustments of φ and ψ angle parameters that were compared with the experimental NMR data of short peptides. As an improvement of the ff14SB [65] the average errors in relative energies of pairs of conformations were under 1.0 kcal/mol as compared to QM, reduced 35% from ff99SB, for this reason this ff14SB is chosen as standard making it AMBER's primary protein model. There are also new developed ffs that are "seaworthy" when combined with standard and older water models. The ff15ipq force field [66] has been developed with a self-consistent physical model and rebuilds nearly all charge, angle, torsion, and some van-der Waals parameters for proteins using the Implicitly Polarized Q (IPolQ) [66] scheme for deriving implicitly polarized atomic charges in the presence of explicit solvent. It is primarily tailored for SPC/Eb water [67]. The ff15FB "Force Balance" force field is another re-derivation of the bonded parameters based on multiple sources of information [68]. This protein ff combines the well-established model of intermolecular interactions from AMBER94 with a systematic and thorough optimization of the intramolecular terms. In summary the ff15FB, the predictions of equilibrium thermodynamic properties were equivalent in accuracy to published models, and the predictions of temperature dependence were significantly improved with the updated TIP3P-FB [69]. According to *Lee-Ping Wang et al* this “model combination AMBER-FB15/TIP3P-FB will yield accurate predictions in simulations of proteins, particularly when fluctuations away from equilibrium, conformational changes, and/or temperature dependence are expected to play important roles. Another relevant field in ff development has been the halogen bonding parametrization, which is crucial in Drug Design (DD) [70]. A halogen bond is a noncovalent bond between a halogen atom (X) and a Lewis base (Y) and many efforts have been carried out to

parametrize this particular interaction [71]. The developed parameters try to simulate the attractive electrostatic interaction between the σ-hole and the Lewis base, using the data obtained by Quantum Mechanical (QM) calculations that are then translated to the non-bonded parameters. The most recent methods are the extra-point (EP) of charge developed by *Mahmoud A. A. Ibrahim* [72,73], where the σ-hole is imitated by a positive dummy atom and the polarizable ff based in the classical Drude oscillator model [74], which is able to reproduce QM molecular dipole moments and polarizabilities by modifying the LJ parameters.

1.2.5.7. Minimization process

The potential energy surface of a system is defined by the mode in which the energy of the molecules varies according to their coordinates. In molecular modeling, it is necessary to study the minimum points on the surface of potential energy that correspond to the stable states of the system [75]. The minimization process removes the excess energy from the system to prevent the simulation from collapsing energetically. This excess energy primarily results from bad interactions, such as two atoms being too close and have very high van der Waals repulsion energies. These normally arise as an artefact from either experimental data or computational modelling. The potential energy surface of a protein is very abrupt, and is made of many peaks (local maxima) and valleys (local minima), and the point with the lowest energetic minimum is known as the overall minimum or global minimum (**Fig. 1.7**), the highest energy point on the road between two minimum points is known as saddle points [76]. The algorithm that allows the identification of the geometries of the system that correspond to the points of minimum potential energy is called minimization algorithm. The most used methods for energy minimization in molecular modeling are those based on the derivatives of the potential function: the steepest descent and the conjugate gradient [77]. Both modify the coordinates chosen to move the system to the point of minimum energy. The first one

makes a search following the maximum slope and is commonly used when the structure to study is very far from the minimum. When approaching this minimum point, the steepest descent method becomes too oscillating and therefore can be replaced with the conjugate gradient method that avoids these oscillations.

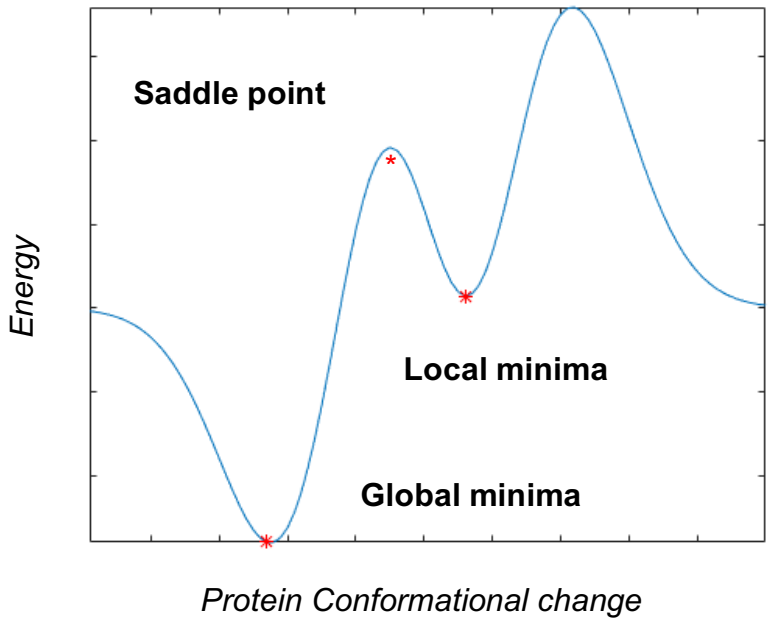


Figure 1.7. *Possible protein conformational changes and energies obtained during the minimization process.*

1.2.5. Molecular Dynamic Simulations

The minimization methods generate individual configurations of minimum energy that, in some cases, may be sufficient to predict certain properties of a system. However, these methods can hardly be applied when we want to study the structural and thermodynamic properties of macromolecular systems containing numerous minimum energy points [78]. Computational MD simulations methods are able to obtain a representative sample of these macroscopic systems at a given temperature during a certain time. There are two large simulation techniques: The Monte Carlo (MC) method, and the MD simulations [79]. While MC simulations are based on analyzing the energy of different randomly generated coordinates, the MD simulations allow the generation of a trajectory of points that evolve over time following

Newton's second law. It is therefore a deterministic method, that is, the state of a point in the trajectory allows to predict the state of the next one. The difference between both methods consist in that the MC simulations do not provide information about time evolution in contrast to the MD simulation, they provide an ensemble of representative configurations; and, consequently, conformations from which probabilities and relevant thermodynamic variables such as the free energy, may be calculated. MC simulations are not only important on their own, but they also play a fundamental role when designing complex and hybrid MD algorithms.

MD simulations are based on the Born-Oppenheimer approximation [50], so the atomic nuclei are heavy enough to be considered as classical particles whose dynamics can be studied by Newton's second equation, $F = m \cdot a$, that can also be described as the next equation:

$$\frac{d^2yi}{dt^2} = \frac{F_{xi}}{m_i}$$

(F_i) being the force applied to the particle (i) in the dimension (x) and (t) time. The total energy of a system is the sum of both the potential and kinetic energies that can be determined from the initial coordinates (x_0) and velocities (v_0).

$$E_{total} = E_{potential} + E_{kinetic}$$

The temporal evolution of the system can be monitored by applying numerical integration methods. In this way, small successive stages separated in time are obtained by a fixed interval δt (integration time). The force F acting on each particle at a time instant (t) is determined by the derivative of the potential energy with respect of the coordinates:

$$F = -\frac{\delta E_{potential}}{\delta x}$$

The value of the force F is assumed constant during each integration step. By summing all interaction forces in the system, the total force is obtained.

From these total forces we can determine the accelerations of the particles which can be combined with the positions *(x_0)* and speeds (v_0) at the initial time (*ti),* to calculate the positions and speeds in the next step (t_{i2}).

$$\frac{d^2yi}{dt^2} = \frac{F_{xi}}{m_i}$$

The integration of the equation of movement is calculated by applying Tailor's series. Using this approximation, the position of a particle at time *t* + δt is expressed according to its position, speed and acceleration:

$$x_i(t = \delta t) = x_i\,(t) + \frac{\delta x_i}{\delta t} + \frac{\delta^2 x_i}{2\delta t^2}$$

The integration time is selected taking into account that a too long δt will lead to instabilities due to overlapping of high energies; but a too short δt will study only a very limited region of the conformational space. In the case of biomolecules such as proteins and DNA, the quickest movements come from the vibrations of the bonds between heavy atoms and hydrogen (X-H) that take place on the femtosecond scale. However, this kind of movements do not contribute to the global behavior of the macromolecular system and would require more computational resources. In order to avoid this kind of unnecessary calculations, the SHAKE algorithm [80] is used which keeps the lengths of the X-H bonds constant in their equilibrium values and allows a longer integration time to be used, thus reducing the computational cost.

1.2.5.8. Periodic Boundary Conditions

Macromolecules are embedded in a solvent system. Usually biological systems, such as proteins and DNA are surrounded by water molecules. In order to simulate this environment during MD simulations the periodic bound conditions (PBC) have been developed. During a MD simulation, the biological system is centered in a water/solvent box [81–83]. To avoid anomalies, and ensure a complete immersion of the solute in the solvent, boundary conditions are used periodically so that the system (primary image) is considered surrounded by equal replicas (periodic images) in all directions, and the calculation of the simulation develops as if the system

were infinite in space (**Fig. 1.8**). The shape and size of the boxes is a key point during the calculation of the MD simulation, and it depends on the geometry of the solvated system, octahedral or cubic boxes can be selected. The bigger the size of the solvent box, the more time-consuming the calculation it will be.

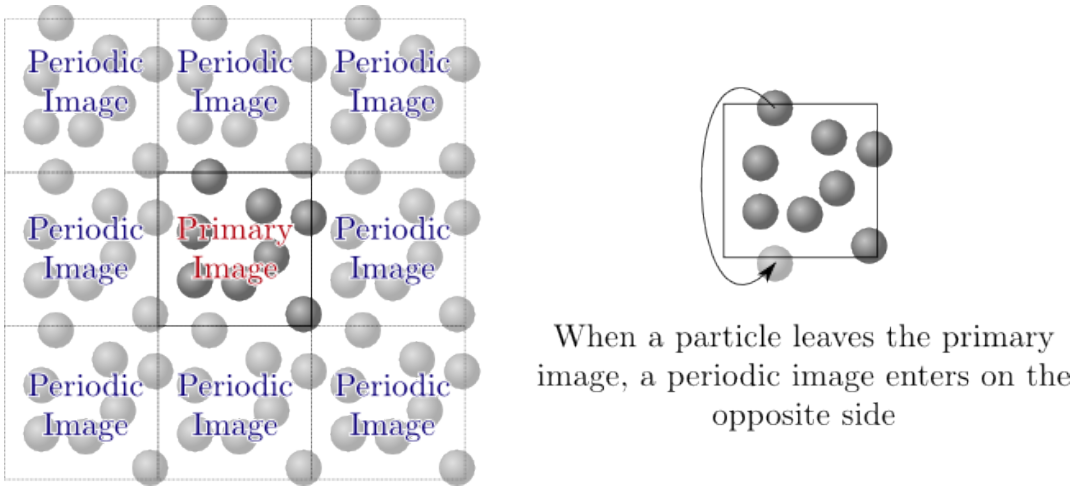


Figure 1.8. *Representative model of periodic boundary conditions where the periodic image and the behavior of the particles in the primary image can be observed.*

1.2.5.9. The Cut-off values

The calculation of the non-binding terms in a MD simulation is one of the most computationally expensive processes. A commonly used approach to accelerate it is to cancel the interactions between pairs of atoms that are separated by a greater distance than the cut-off value, which is a distance limit commonly between 10-12 Å. It is important to keep in mind that the distance limit must be less than the distance between any point of the box to any of its neighboring copies, otherwise it would be calculating interactions with same atom but in a different periodic image.

1.2.5.10. Amber MD protocol

There are so many protocols and conditions to carry out an MD simulation and almost every research group have their own way and perspective. Of course, there are some basic rules in conducting MD simulations, such as

the protein preparation process, minimization, heating and equilibration that have to be performed previous to the MD production phase.
Here, we present a standard protocol stablished in our laboratory to carry out MD simulations of protein-small molecule complexes (**Fig. 1.9**):

1) System preparation. The first step is getting everything ready to carry out the AMBER parametrization of the PDB file. For this purpose, the complex obtained from the docking or directly from any database has to be correctly protonated and every atom named according to the AMBER format. Residue protonation states of crystal structures can be calculated using the H++ web server (biophysics.cs.vt.edu/) that relies on the AMBER ff. H++ computes pKa values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. However, complexes issued from docking calculations are prepared and protonated in advance, in our case with the Protein Preparation Wizard included in the Schrödinger suit [84]. It is recommended before starting the AMBER parametrization to check the structure with the handy software pdb4amber that can check the correctness of the amino acids. As for the ligands, those issued from a docking calculation are already protonated; and in case we do not have this option, many webpages are available for this purpose, such as chemicalize, drugbank and corina. Additionally, free software such as babel and molden is it possible to create and prepare ligands. Once the 3D structures of the small molecules are ready, we must obtain the charges and parameters (bonds, angles, dihedrals and improper angles). For this purpose, Quantum Mechanics are applied using Gaussian [85] and the geometry optimization and RESP charges are calculated Hartree-Fock method and 6-31G** basis sets to match the parametrization used for the AMBER ff.

2) AMBER parametrization. As mentioned before the general force field ff14SB is used for the protein parametrization and the GAFF ff for the parametrization of the small molecules. The initial coordinate and topology

files of the system needed for the simulation are obtained using the tleap tool implemented in AMBER. It is important to note, that by software such as parmed or xparmed we can analyze any parameter created in the topology file and also modify it if it were necessary.

3) Vacuum minimization. The first step before running the MD simulation is the minimization of the starting structure in order to reduce and eliminate possible bad interactions and contacts. This process is strategically carried out in steps: first all hydrogens are minimized, after that the amino acid side chains and finally the whole complex.

4) Solvent embedding. The minimized structure is embedded in a TIP3P water box or octahedron using tleap. Different solvent boxes and sizes can be used depending on the system. Usually we adopt the octahedral system due to the more spherical conformation that adapts better to globular proteins, and also because it reduces the number of water molecules in comparison to the classical box, thus reducing the computational cost of the simulation. The cut-off value between the surface of the protein and the limit of the box varies between 10 to 15 Å but, usually 13 Å is the standard value applied. During the embedding, the whole system is also neutralized by adding Na+ or Cl- atoms in a random position inside the solvent box to obtain charge neutrality of the system.

5) Solvated minimization, heating and equilibration process. Before launching the MD simulation, the solvated system has to be heated and equilibrated. First of all, water molecules and corresponding ions are minimized, in order to delete any steric hindrance. After that, the system is heated in periodic boundary conditions and constant pressure (NPT) from 100K until 300K during 25ps. The heated system is later on equilibrated by progressively allowing the solute to move while the solvent molecules are free to move.

6) Molecular Dynamic simulation. Once the system is ready with stable density and temperature, MD simulations can be carried out without any restraints. Snapshots are collected each 20 ps for further analysis. The collected trajectories can be used to obtain the corresponding RMSD values, fluctuations or binding energies that will be discussed in following sections. The MD simulations obtained during the development of this thesis can vary from 10 ns to 100 ns, depending of the purpose of the simulation. Usually, ligand stabilities can be analyzed with just 10 to 20 ns of simulation time, but for instance loops conformational changes or organic solvent exploration of the protein surface can only be achieved with larger simulations times.

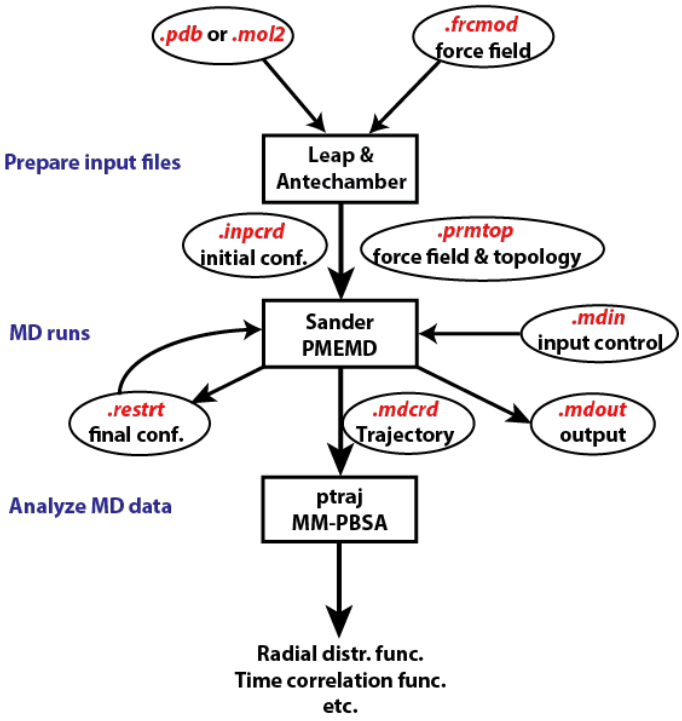


Figure 1.9. *General Proposed Amber protocol flowchart.*

1.2.5.11. MD analysis

The trajectories of the MD simulations can be analyzed by different softwares, depending on the information to be extracted that can either be density, volumes, distances, energies, diffusion, stability, or energies among others. Visual inspection is always mandatory to check the stability

of the simulated system and it is used as a first filter. For this purpose, the trajectory is transformed into PDB files that can be read by the molecular graphics program PyMOL [86].

One of the most used measurements to analyze the stability of a complex is the Root-Mean-Square Deviation of atomic positions (or simply Root-Mean-Square Deviation, RMSD) which is the measure of the average distance between a group of defined atoms in a specific 3D coordinate compared to the same atom in a reference position [87]. This can be applied to the backbone atoms of the protein, and also to the ligand to check the stability of the complex. Another related measure is the Root-Mean Square (RMS), which only measures the difference distance between two atomic positions, commonly applied when two related structures are aligned.

As MD simulations produce long trajectories of atomic positions as a function of time, to provide several representative structures of the simulation constitutes an advantage [88]. This approximation is called clustering and can be applied to the whole simulation or to a fragment of it; and can be applied to a selected number of atoms or the whole system.

The electrostatic and van der Waals components of the ff will generate attractive or repulsive forces that cover the binding of a ligand to its target. These forces can be measured along the MD simulation and translated into binding energies. These energies will allow the analysis of ligand-receptor and protein-protein interactions along the simulation time that will give an idea of the quality of the studied complex. The most popular approaches to estimate the free energy of the binding of small ligands to biological macromolecules are MM/PBSA and MM/GBSA, which combine molecular mechanics (MM) energies with the Poisson–Boltzmann (PBSA) or generalized Born and surface area continuum solvation (GBSA) [89]. From the MM term that both methods apply, energy is calculated by the sum of the bonded (bond, angle and dihedral) and non-bonded (electrostatic and van der Waals interactions) terms. The MM/PBSA method was developed by *Kollman et al.*[90] in the late 90s whereas the MM/GBSA was developed

by *David A. Case* in the early 2000s [91]. In both methods, the binding free energy (ΔG_{bind}) between a ligand (L) and a receptor (R) to form a complex RL is calculated as:

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw}$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$$

In these equations ΔE_{MM}, ΔG_{sol} and $-T\Delta S$ are the changes of MM energy in the gas phase, the solvation free energy, and the conformational entropy upon binding, respectively. ΔE_{MM} includes $\Delta E_{internal}$ (bond, angle and dihedral energies), electrostatic $\Delta E_{electrostatic}$ and van der Waals energies ΔE_{vdw}. ΔG_{solv} is the sum of $\Delta G_{PB/GB}$, which is the electrostatic solvation energy (polar contribution), and ΔG_{SA} as the non-electrostatic solvation component (non-polar contribution). The polar contribution is calculated using either GB or PB model, while the non-polar energy is estimated by solvent accessible surface area (SASA). $\Delta G_{PB/GB}$ is typically obtained by solving the Poisson-Boltzmann (PB) equation giving the MM/PBSA estimation whereas by using the generalized Born (GB) model the MM/GBSA is obtained. Through the solution of this equation an estimated binding energy can be obtained. Nevertheless both methods include some disadvantages such as: the computational cost of the calculations; overestimation of the differences between sets of ligands and questionable entropy estimation of both methods; the disregard of the conformational contribution; and, the overlook of the effect of water molecules lodged in the binding-site, amongst others [89].

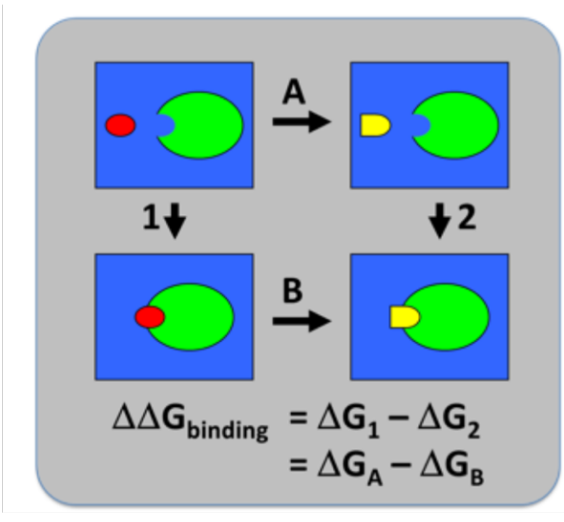
Another recently developed method to calculate the binding energy of a complex is the ultrafast MM-ISMSA developed by *Klett et al*. [92]. In this case, the terms PB or GB have been modified by the Implicit Solvent Model

(ISM) method that comes from the Deby-Sack's theory of polar liquid. The MM-ISMSA function includes: i) a molecular mechanics (MM) term based in Lennard-Jones potential 12-6; ii) an electrostatic component based on the implicit solvent model (ISM) that includes the calculation of the individual desolvation of each of the elements that form the complex, plus a term for hydrogen bonds; and, iii) the contribution for loss of surface area (SA) to take into account the decrease in the number of contacts with water molecules when the complex is formed. The ISM approach reduces significantly the complexity of the calculation carried out by PB or GB methods giving a reliable result with lower computational resources.

1.2.5.12. Free Energy Perturbation (FEP)

FEP is a method based on statistical mechanics that is used in computational chemistry for computing free energy differences from MD or MC simulations. This method is used to calculate the free energy difference of going from state A to state B by applying the Zwanzig equation [93].

$$\Delta F(\mathbf{A} \rightarrow \mathbf{B}) = F_{\mathbf{B}} - F_{\mathbf{A}} = -k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{E_{\mathbf{B}} - E_{\mathbf{A}}}{k_{\mathrm{B}}T}\right)\right\rangle_{\mathbf{A}}$$



$\Delta G_A - \Delta G_B = AG_1 - \Delta G_2 = \Delta\Delta G_{binding}$

Where T is the temperature, k_B is Boltzmann's constant, and $E_B - E_A$ is the free energy map (also known as Potential Mean Force, PMF) along both states. FEP calculations can be achieved when the difference between the two states is small enough (and not too big) in a way that it is possible to divide a perturbation into a series of smaller "windows", which can be computed independently. The application of FEP is huge and actually it

could be said that is one of the most promising CADD applications [94,95]. For example, this technique is usually applied to calculate the energy difference in binding energy between a series of ligands. Using a reference active compound (A), we can artificially obtain a series of compounds (B) that we know present better calculated binding energy. The advantages of this method come from the fact that when predicting relative binding energies, both enthalpic and entropic term are taken into account by including the receptor and the ligand. Moreover, the calculations can be carried out in explicit solvent. The study carried out by *Wang et al.* [96] where, through the FEP+ software [97] of the Schrödinger suit, a pull of known inhibitors and different receptors are considered, shows the high reliability of the results obtained by this technique (**Fig. 1.10**).

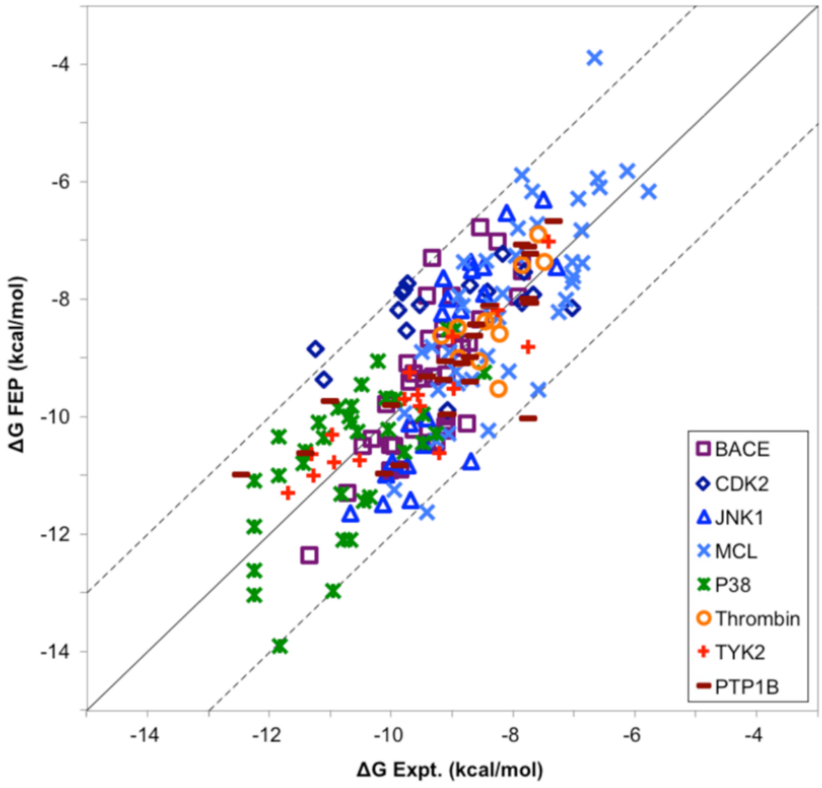


Figure 1.10. *Correlation between FEP-predicted binding free energies and experimental data for all eight systems studied. FEP-predicted binding free energies for most of the ligands are within 1.0 kcal/mol of their experimental values, and only nine of 199 studied ligands deviate from their experimental free energies by more than 2 kcal/mol. Figure taken from J. Am. Chem. Soc.* ***2015****, 137, 2695−2703.*

Additionally, FEP can be applied to pKa predictions, analysis of the solvent effects on reactions, and enzymatic reactions among other types of in silico experiments. The increase of FEP techniques has promoted its application by using Graphic Processor Units (GPU) instead of the Core Processor Units (CPUs), which represent higher computational power, thus reducing the calculations almost 20 times. Several software companies have been developing FEP methods such as Schrödinger (schrodinger.com/fep), AMBER [98] or SeeSar-BioSolveIT (biosolveit.de/SeeSAR/).

1.2.5.13. MD simulations in Nonaqueous Solvents

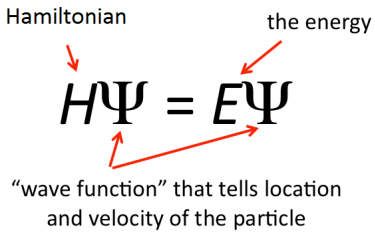
Pocket druggability is a term used to identify suitable binding pockets for small molecules, which could be considered as one of the main issues at the beginning of the process in CADD [99]. Predicting protein druggability is a key point and hotspot in computational modelling due to its applicability and number of publications [100]. Structural biology has been extensively used to find novel and allosteric pockets by crystallizing small ligands and fragments by soaking techniques; but despite its reliability, this methodology presents the main disadvantage of its expensive cost and the limitation in the number of molecules used for pocket exploration [101,102]. The high cost of this technique lays in the fact that many crystals must be diffracted using as many probe molecules as possible in order to obtain satisfactory data of novel pockets, increasing the cost or this type of experiments. Taking this is mind, MD simulations with water-mixed organic solvents have been a novel application to further overtake this challenge. Instead of carrying out conventional MD simulations with water molecules, organic solvent molecules are added to the solvent box in order analyze if those are able to reach any other unknown pocket or explore in depth the well-known pockets [103]. One of the main disadvantages of this methodology is the possibility to obtain exacerbated conformational changes in the protein that may reflect denaturation of the protein due to high concentrations of organic solvents [103,104]. Moreover, not all probe molecules have shown to be

effective for all systems. For this reason, it is needed to carry out many different attempts with different solvent mixtures at different concentrations to find the best way to explore the druggability of the target protein.

Another issue to take into account is the “false-positive” pockets that can be determined through this type of simulations that need to be addressed and evaluated to be considered as real druggable pockets. A wide variety of probe molecules have been used to search for new cryptic pockets such as benzene, methanol, hexane, ethanol, isopropanol, purine, chloroform, toluene, chlorobenzene, pyridine, imidazole, indole, pyrimidine, tetrahydropyran, acetic acid and ammonia amongst others [104–108]. All these probes present their own particular physicochemical properties that makes them suitable or not to interact with the different domains of the protein. In summary, all published methodologies start from the “apo” structure with a bound ligand and are simulated until obtaining the “apo-bound” structure, which presents an organic solvent molecule from the mixture inside a new pocket that was not present in the apo structure [107]. Regarding the nomenclature, some authors call the “apo-bound” structures “holo structures”, and this term can be erroneously correlated with the holoprotein definition in enzymology, which means a complete protein with its prosthetic group [109]. For this reason, in this thesis we consider a protein with a bound organic solvent molecule the apo-bound structure. The “apo-bound” conformation is generated by the displacement of some amino acids (generally hydrophobic) brought about by the organic solvent molecules. The interactions with the pocket established by the probe molecules can be later on evaluated, and the long term and stable pockets can be suitable for further studies to evaluate its druggability. Numerous softwares have been developed to identify, find and evaluate the cryptic pockets such as pyMDMix (mdmix.sourceforge.net), Fpocket(fpocket), CAVER (caver.cz), and MDpocket(omictools.com/mdpocket-tool) [110–112].

1.3. Quantum Mechanics and Thermochemistry

Quantum chemistry, also called molecular Quantum Mechanics (QM) is a branch of theoretical chemistry where quantum mechanics and quantum field theory are applied [113]. Mathematically, QM describes the fundamental behavior of matter on a molecular scale. QM methods study the ground state (the lowest energy state) of individual atoms and molecules, and the excited states, and transition states that occur during chemical reactions. QM is applied nowadays to study molecular and physicochemical properties of molecules and materials, light emission/absorption, spectra, chemical reactions, chemical interactions, electronic research, amongst others. QM begins with the Schrödinger equation that describes the wave function of a defined system [114]. The exact solution for the Schrödinger equation can only be obtained for the hydrogen atom [115]. All other atomic, or molecular systems, involve the motions of three or more "particles", their Schrödinger equations cannot be solved exactly, for this reason and approximation must be applied. Here is where QM is applied. The different approximations to solve the Schrödinger equation have led to the different QM methodologies developed such as Hartree-Fock (HF), Density Functional Theory (DFT) or Heitler–London–Slater–Pauling (HLSP) method. The HF methods converts the multielectronic Schrödinger equation into several simpler monoelectronics, where the interelectronic repulsion is treated on average. Each monoelectronic equation is solved by providing a monoelectronic wave function (orbital) and an energy that is the energy of the orbital. The solution of the HF equation provides a set of orthonormal orbitals [116]. The post-Hartree–Fock methods are the set of methods developed to improve on the original HF. They add electron correlation, which is a more accurate way of including the repulsions between electrons than in the Hartree–Fock



method where repulsions are only averaged. This post-Hartree–Fock methods includes the Multi-configuration time-dependent Hartree (MCTDH) and Møller–Plesset perturbation theory (MP2, MP3, MP4). The DFT method has become very popular in recent years, the main advantage of this method is that the energy of a molecule is determined form the electron density instead of a wave function used by the previous methods [117]. The electron density is expressed as a linear combination of basic functions that is used to obtain the energy. In DFT there are more than 200 functionals developed, all of them with advantages and disadvantages [118]. The DFT method tends to be classified as ab initio methods and its main advantage is that the Coulombic repulsion integrals are calculated over the 3D electron density instead of the whole functional. For this reason, DFT is faster than HF for systems with more than 10-15 non-hydrogen atoms and this fact, together with its accuracy, has made DFT one the most popular methods in thermochemistry. Both DFT and HF methods must use a basis set. A basis set is a set of functions (called basis functions) that are used to represent the electronic wave function in order to turn the equation suitable for efficient implementation in a computer [119]. As we mentioned, DFT presents more than 200 methods, only the mention of each one would exceed this introduction [118]. For this reason, B3LYP, the most cited method for organic molecules calculations will be addressed [120,121].

B3LYP is a hybrid functional developed in the late 80s based in exchange correlation approximation [122–125]. Decoding B3LYP in terms: B3 is Becke's 3 parameter exchange correlation functional which uses 3 parameters to mix in the exact Hartree-Fock exchange correlation; and, LYP is the Lee Yang and Parr correlation functional that recovers dynamic electron correlation. B3LYP is one of the first DFT methods that was a significant improvement over Hartree-Fock. B3LYP is generally faster than most Post Hartree-Fock techniques and usually yields comparable results. Another advantage is that it is not as heavily parameterized as other hybrid functionals, having only 3 where other could have up to 26. Actually, empirical dispersion (d3) is also included in any B3LYP calculation to

incorporate a correction for long-range dispersion interactions [126].

1.4. ADME-Tox

Among the improvements in hit identification and lead optimization in drug discovery and development, significant progress has also been made in the prediction of absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) properties by computational tools [127]. The two main reasons for clinical trial failure are the lack of efficacy (the expected pharmacological effect is not observed) and the toxicity (the molecule presents excessive side effects) [128]. For this reason, ADME-tox prediction is a growing field in CADD, and increasingly reliable and predictive models are being built. These models can be applied virtually at the early stages of the discovery process with the aim to minimize the non-optimal compounds and prevent the failure of those before the in vivo assays.

In order to create any model a data set is needed, in this case, for the ADME-Tox calculation, this data could come from the open-source collections like the US Food and Drug Administration (FDA) [129,130], but mainly the majority of the data comes from the pharmaceutical industry [131]. In silico models try to correlate the chemical structure or chemical fingerprints of the molecule with the ADME-tox parameter; for this reason, the more data and chemical diversity used to build the model, the more complex and reliable the model will be. The models implemented are based in genetic algorithms, machine-learning, vector machine algorithms, neuronal networks, and ultimately, deep learning. ADME models can be classified by the prediction field, for instance: i) physicochemical properties (pKa, permeability, Topological Polar Surface Area (TPSA), solubility, Lipinski rules, PAINS); ii) ADME properties (gastro intestinal absorption, bioavailability, distribution, protein binding, Blood Brain Barrier (BBB) cross, partition coefficients, metabolism and clearance); iii) drug toxicity (hepatotoxicity and carcinogenicity, cardiotoxicity, CYP inhibition). Actually,

there is software and several free databases (crdd.osdd.net/admet) to calculate and predict ADME-tox properties. During the development of this thesis the following have been used:

- Molinspiration (molinspiration.com) with more than 4500 citations is based on the prediction physicochemical properties.
- The online chemical database Ochem (ochem.eu) predicts the physicochemical properties such as solubility, LopP and ADME-tox properties like carcinogenicity and CYP inhibition.
- SwissADME (swissadme.ch) computes physicochemical descriptors as well as to predict ADME parameters and pharmacokinetic properties.
- QikProp (schrodinger.com/qikprop) distributed by Schrödinger provides rapid ADME predictions for drug candidates.

1.5. Ligand-receptor interaction energies and constants

A straightforward thermodynamic equation can be used to describe binding mechanisms. When levels stay unchanged the reversible binding mechanisms of a ligand (L) to a receptor (R) are in balance. The relationship is dynamic and depends on the association rate (K_{on}) and the dissociation rate (K_{off}). This equilibrium is defined by the K_d, that relates both terms K_{on} & K_{off}. Another ligand-binding constant is the K_i or inhibition constant, which is in relation to the displacement of a competitor.

$$R + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} RL \qquad K_d = \frac{k_{on}}{k_{off}}$$

The spontaneous process of binding only happens when the total Gibbs free energy increases. The Gibbs equation depends on the ideal gas constant R; the absolute temperature T; the standard reference concentration $C^{\ominus}$ (1 mol·L^{-1}); the enthalpic change ΔH; and, the entropic term ΔS:

$$\Delta G = -RT \ln\left(\frac{K_d}{C^\theta}\right) = \Delta H - T\Delta S$$

One particular binding process may be either enthalpy-driven or entropy-driven, depending on the sign and comparative weight of the *ΔH* and *ΔS* terms [132].

ΔH can be approximated to the process in which new chemical bonds are formed as well other are cleaved, and the difference between the initial and final state will be ΔH. This term, in ligand binding includes the intramolecular and intermolecular weak interactions, as well as the costs of breaking the hydrogen bonds when solvating and desolvating the protein and the ligand.

ΔS relies in the distributed molecular energy from the system. This term can be approximated to the loss of ligand and protein flexibility, and the rearrangements in the solvent network.

According to the previous equation, ΔG can be obtained through the K_d, by performing biochemical binding experiments. ΔH can be characterized by using other experimental techniques such as isothermal titration calorimetry (ITC), where the heat that is either released or absorbed during a biomolecular binding event is directly measured. The ΔS can only be calculated by solving the previous equation with known values for the ΔG and ΔH terms. It is interesting to mention that studies carried out by Charles H. Reynolds & M. Katharine Holloway show that, out of a pull of experimental binding thermodynamics extracted form 100 protein−ligand complexes, ΔG are not generally correlated to either ΔH or ΔS. Moreover, they highlighted that when considering analogous series of compounds, the increase of ΔH is related to the decrease of ΔS, and vice versa. That means, ligands with favorable enthalpies have negative entropies and ligands with favorable entropies have unfavorable enthalpies [133].

From the computational point of view, this is a great issue since the majority

of the software and scoring functions tend to provide the enthalpy part by measuring the ligand-receptor interactions energies but not taking into account the entropy changes. Nevertheless, new techniques such as MD, FEP and other free energy techniques are currently facing this issue as they can obtain information on the change of conformational entropy of both the ligand and receptor upon binding [134,135].

1.5.1. Ligand-receptor interactions

Drug specificity and selectivity is brought about by the sum of all the non-covalent ligand-receptor interactions. Nowadays the so-called atomic interactions established between ligands and proteins are receiving special attention. An extensive analysis of this type of interactions where the most frequent were extracted was carried out using the PDB and reported by Renato Ferreira de Freitasa and Matthieu Schapira (**Fig. 1.11**) [136].

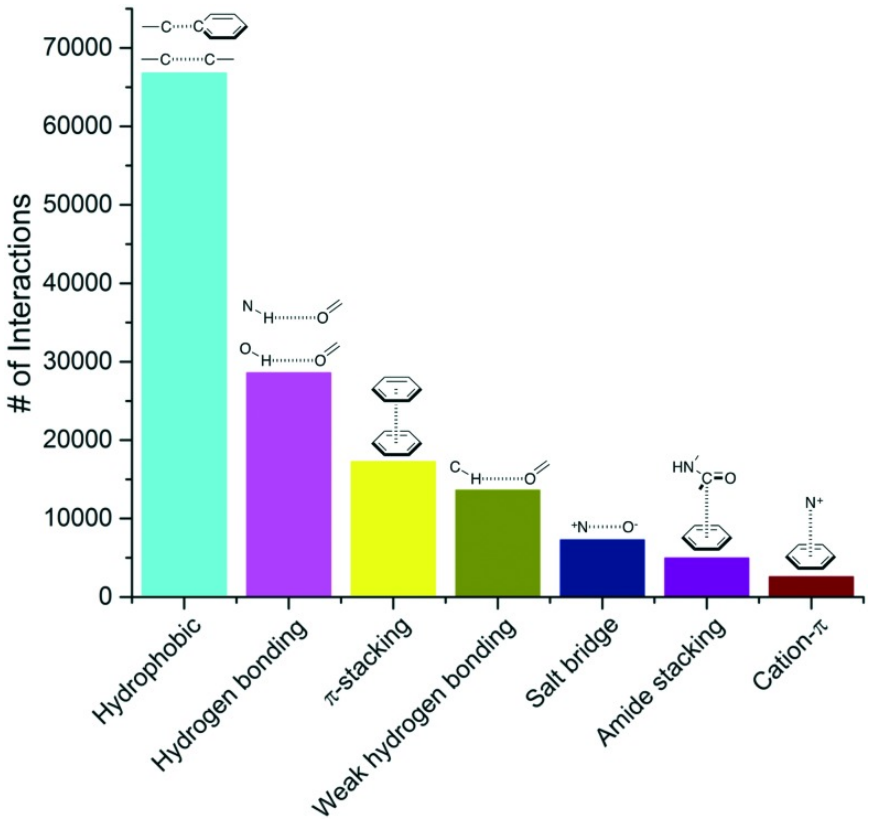


Figure 1.11 *Frequency distribution of the most common non-covalent interactions observed in protein–ligand complexes extracted from the PDB. Image taken from Med. Chem. Comm. 2017, 8, 1970-1981.*

The main types of interactions are columbic (formed by two charged groups), dipole-dipole (when the partially negative portion of one of a dipole approaches the partially positive portion), halogen bonds, and other sigma holes interactions, and, the van der Waals interaction including the Debye, London dispersions and Keesom forces. The strength of the interactions varies, but the order is usually as following: ionic > hydrogen-bond > dipole-dipole > van der Waals.

1.5.2. Salt bridges

Salt bridges, also known as ionic interactions, are the strongest amongst all known non-covalent molecular interactions, with energy values ranging from −15 to −21 kcal·mol^{-1} [137]. The salt bridge is a non-covalent interaction between two ionized molecules that mixe two contributions: a hydrogen bond, and an electrostatic one. This interaction is mainly seen between a protonated amine and a deprotonated carboxylic acid. The proton involved in this interaction can migrate from the charged amine to the side chain of the carboxylic acid group and vice versa. Nevertheless, pure salt bridges, that is non-hydrogen bridged, can also be found between fully alkylated quaternary ammonium groups and charged carboxylic acids [138]. Geometrically, a salt bridge angle between two moieties may be oriented in the range 147−176°, with a distance that can vary from 2.70 Å to 3.30 Å (**Fig 1.12**).

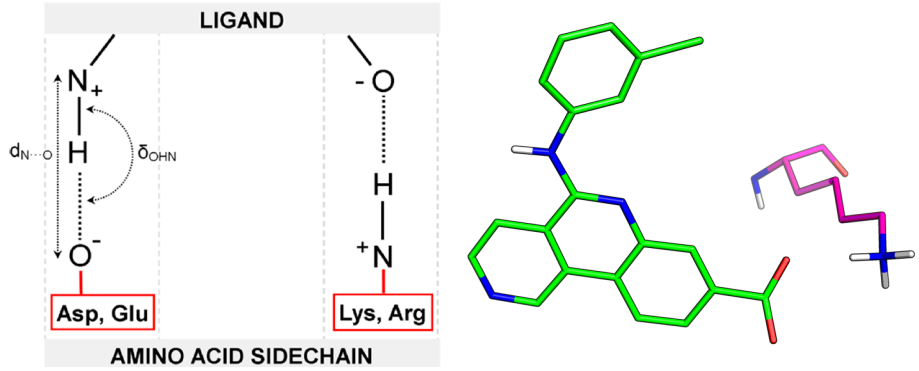


Figure 1.12 *Scheme of ionic interactions. Salt bridge between the charged carboxylate group from CX4945 and Lys68 from the CK2 receptor (PDB code **3PE1**).*

1.5.3. Hydrogen bonds

Hydrogen bonds are the most significant interactions in the biological recognition process. A hydrogen bond is an interaction between a proton donor (D) (electronegative atom, particularly nitrogen or oxygen) and a proton acceptor or lone pair carrier (A) [139]. The strength of the hydrogen bonds can range, depending on the distance, between 1-12kcal·mol-1. The typical distance (D-A) is between 2.8 and 3.2Å, and shorter distances will lead to lower energy values. The preferred angle between the donor and acceptor (D—H—A) is 150°, which corresponds to the lone pair direction of the donor, but it can range also from 170 to 180° depending on the implicated groups (**Fig. 1.13)** [139,140]. Sometimes, aromatic rings can be used also as a weak hydrogen acceptor, as the aromaticity generates a structure rich in electrons above and below the benzene ring that exhibits a partial negative charge [141].

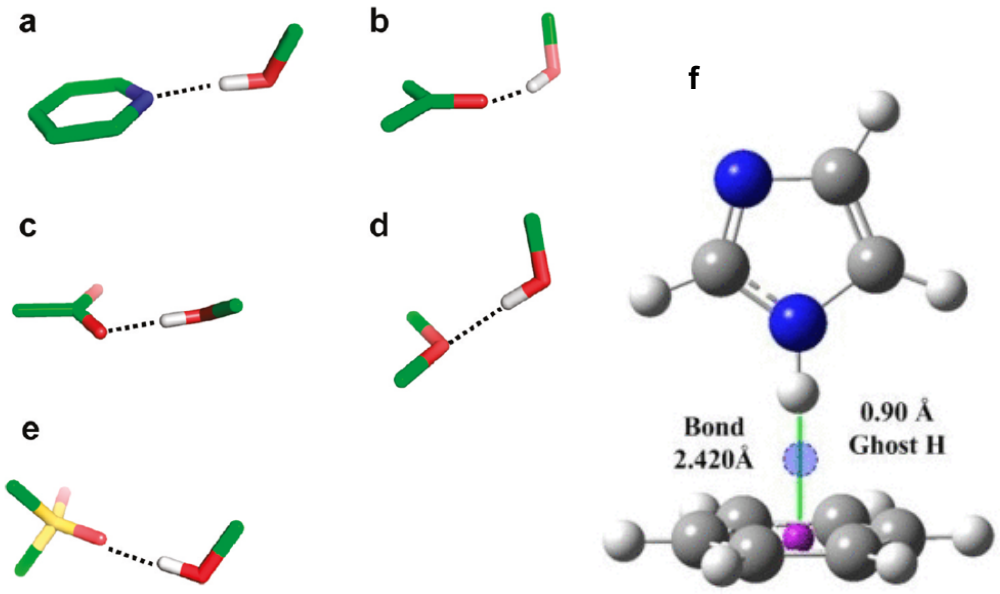


Figure 1.13. *Schematic depiction of the most preferred geometries of hydrogen bond interactions with various types of acceptors: (a) pyridine nitrogen, (b) carbonyl oxygen, (c) carboxylic acid, (d) ether oxygen, (e) sulfonyl group and (f) hydrogen bond inside aromatic ring, histidine and benzene molecules. Figure taken from Journal of Medicinal Chemistry,* ***2010****, Vol. 53, No. 14.*

The strength of the hydrogen bond depends in the geometries

(directionality) and interatomic distances between the implicated atoms. This strength can be calculated by the ionic character ΔpK_a determinant and when this value approaches zero the strength gets closer to the maximum value [142]. Additionally, another important measurement is the pK_{BHX} or hydrogen bond basicity, which is defined as the capacity or strength of a functional group to be a hydrogen acceptor [143]. Interestingly, this measurement is not directly linked with basicity, because it is involved with the desolvation process and interaction geometries [144]. Also, the pK_{BHX} can be correlated with ligand affinity giving important information for the SAR development [145].

1.5.4. Halogen and sulfur bonds

Halogen bonds were discovered in 1986 by *N. Ramasubbu et al.* through a computer analysis of structures with C-X bonds (were X= Cl, Br, I) extracted from the Cambridge Crystallographic Data Base (ccdc.cam.ac.uk) [146]. In this study, halogen bonds were discussed for the first time, leading to the conclusion that the halogen (X) from the C-X bond is capable of interacting with electrophiles, nucleophiles, and other halogens (**Fig. 1.14-a**). Halogen bonds occur when a σ-hole is formed in the electronic cloud of Cl, Br and I atoms (**Fig. 1.14-b)**. This σ-hole presents a positive changed density and acts as a Lewis base that is able to accept a lone pair of electrons [147]. Therefore, the σ-hole can lead to attractive interactions with classical H-bond acceptors such as carbonyl groups (**Fig. 1.15**), but also the electronic density of the π-clouds of aromatic groups [71]. The relevance of this type of bonding as provided up to 1,000-fold increase of some inhibitors in specificity and affinity towards their targets, giving the halogen substitution a great potential tool in medicinal chemistry [148].

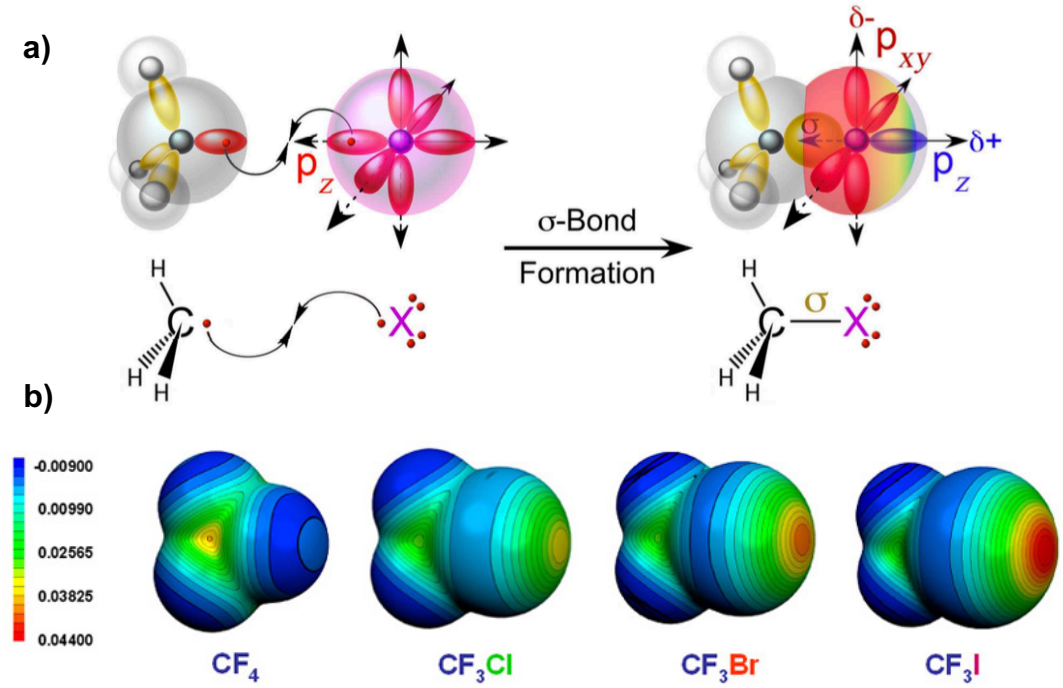


Figure 1.14. a) *The σ-hole formation of a covalent carbon-halogen bond (a C—X σ-bond), as a result of the bonding the p_z orbital of the halogen opposite to the carbon atom becomes depopulated leading to the σ-hole.* ***b)*** *Electrostatic potential calculated at the 0.001 au contour level for perfluoromethane, chlortrifluoromethane, bromotrifluoromethane and iodofluoromethane. Red and blue denote regions of negative and positive potential, respectively. Figure taken from crystalengineering.blogspot.com.*

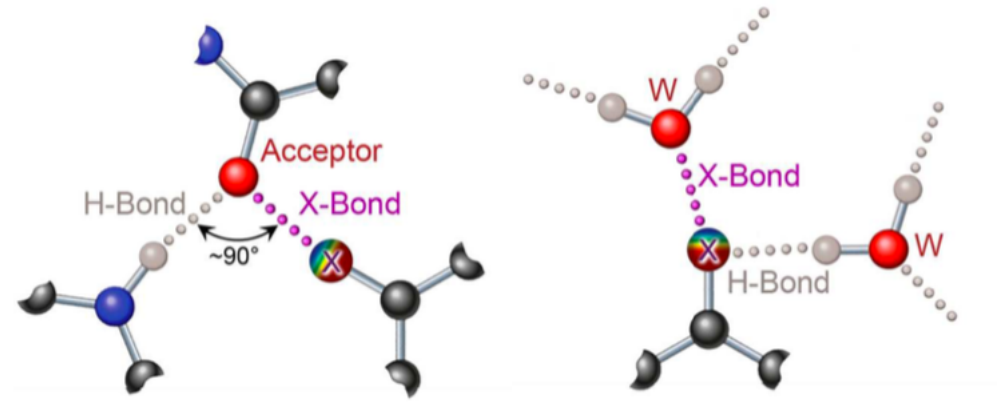


Figure 1.15. *Different type of halogen bonds with donor and acceptor molecules. Figure taken from crystalengineering.blogspot.com.*

Sulfur atoms also present σ-holes and this fact is not as commonly known and as exploited as for halogens [149]. The C−S bond presents a low lying

σ* orbital on the S atoms that could be referred also as a σ-hole, leading to interactions within electron donating atoms, such as N and O [150]. In contrast to halogen atoms, the positive region or σ-hole-like potential in sulfur is present near the atom, although slightly offset from the C−S bond axis (**Fig. 1.16**) [150].

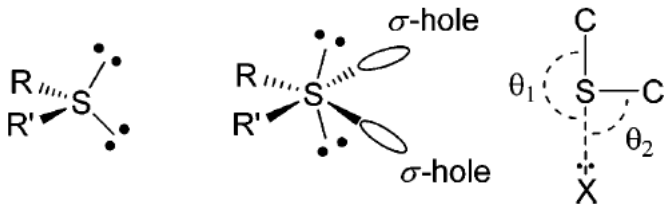


Figure 1.16. *Classical representation with the two lone pairs of electrons of a sulfur atom with the σ* orbitals, and lone pairs of electrons; and, the attractive Ẍ-S interaction. Figure taken from J. Med. Chem.* ***2015****, 58, 4383−4438.*

In medicinal chemistry, the O-S and, particularly, N-S interactions in intermolecular context, have been used to further stabilize molecular conformations and reduced the molecular flexibility with the aim to increase the activity of the designed inhibitors (**Fig. 1.17**) [151,152].

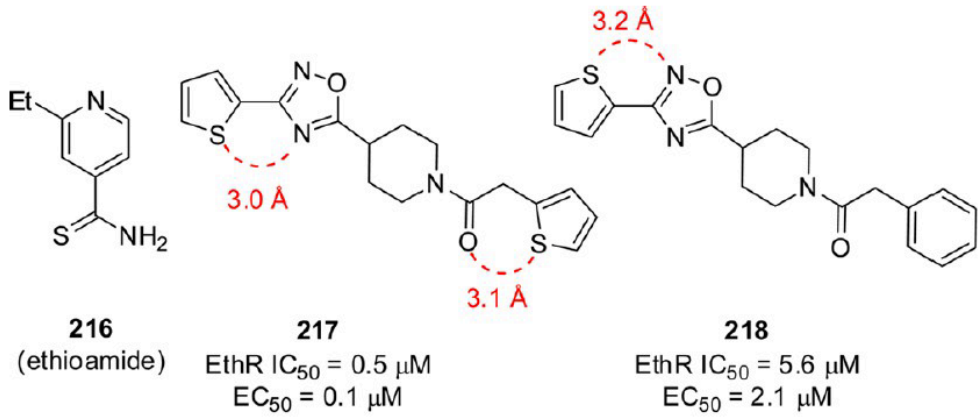


Figure 1.17. *Antituberculosis drug ethionamide and derivatives* ***216****,* ***217*** *and* ***218;*** *showing the EthR (flavin adenine dinucleotide-containing monooxygenase) inhibition. Image taken from J. Med. Chem.* ***2015****, 58, 4383−4438.*

Moreover, sulfur atoms can interact with aromatic residues giving a S-π interaction, where the van der Waals contacts between the sulfur and the

carbon present a maximum distance of 5.3 Å (**Fig. 1.18**). These types of contacts are common in intraprotein contacts established between the sulfhydryl groups of cysteines and other aromatic rings [153].

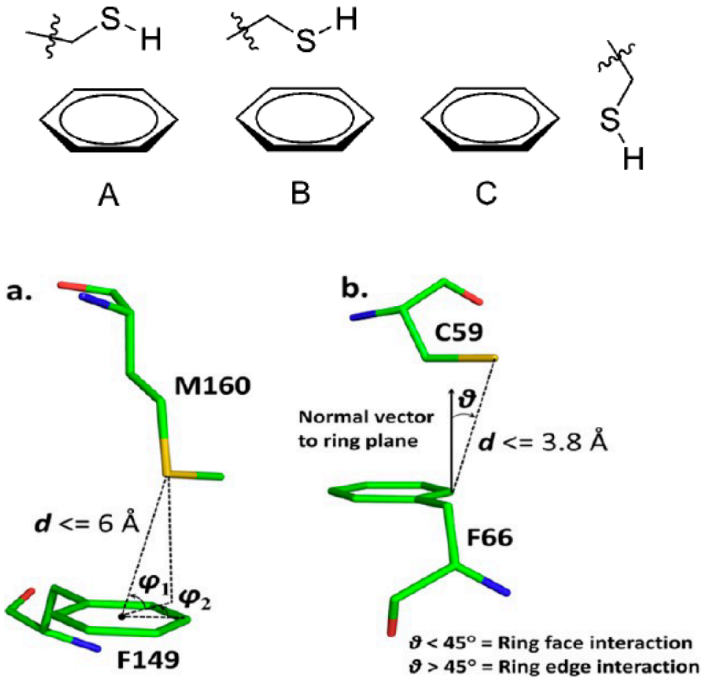


Figure 1.18. A, B, C *Different structures and types of the S-π interaction.* ***a.*** *The Met160/Phe149 S-π interaction in the PDB code* ***3TBD****.* ***b.*** *The S···π interaction between Cys59 and Phe66 in the PDB code* ***3TBD****. Image taken from J. Med. Chem. 2015, 58, 4383−4438.*

1.5.5. Hydrophobic interactions

The hydrophobic effect consists in the tendency to self-association of non-polar molecules in the presence of aqueous solution. This short-range attractive energy presents both enthalpic and entropic effects [154]. It describes the energetic preference of nonpolar molecular surfaces to interact with other nonpolar molecular surfaces, and thereby displace water molecules from the interacting surfaces. The association of two nonpolar molecules in water reduces the total nonpolar surface area exposed to the solvent, increasing the entropy of association and reducing the amount of structured water between both molecules. In detail, desolvation and cooperative effects govern free energy gains through hydrophobic

interactions [155]. As for van der Waals forces, hydrophobic interactions are individually weak (0.1 to 0.2 kJ/mol for every square angstrom of solvent-accessible hydrocarbon surface) but all together generate non-negligible interaction energy [156].

1.5.5.1. π interactions

Aromatic moieties are common drug fragments, so interactions involving π systems are critical to protein-ligand recognition since the side chains of aryl containing amino acid such as Trp, Phe, Tyr, and His are often exposed to binding sites. These interactions include π-π stacking, alkyl-aryl, and cation/anion-π [157]. These types of interactions are formed due to the charge distribution and the aromatic nature of the interacting groups. The aromatic π systems generate a negative charge density above and below the planes of the ring, while the periphery of the ring exhibits a positive charge density. The most commonly known is the π-π stacking interaction established by a benzene dimer (**Fig. 1.19**) where π system of a benzene interacts with another π system in a parallel orientation [158,159]. Additionally, there are the parallel displaced and T-shaped modes that are stable and favored interactions, but in this case the parallel-displaced arrangement is 0.5-0.75 kcal/mol more stable than T-shaped, at least in protein structures [160]. Interestingly, when the aromatic ring is polysubstituted the polarity of the π-system may be reversed allowing the charge transfer from an electron-rich π-system to the polysubstituted one.

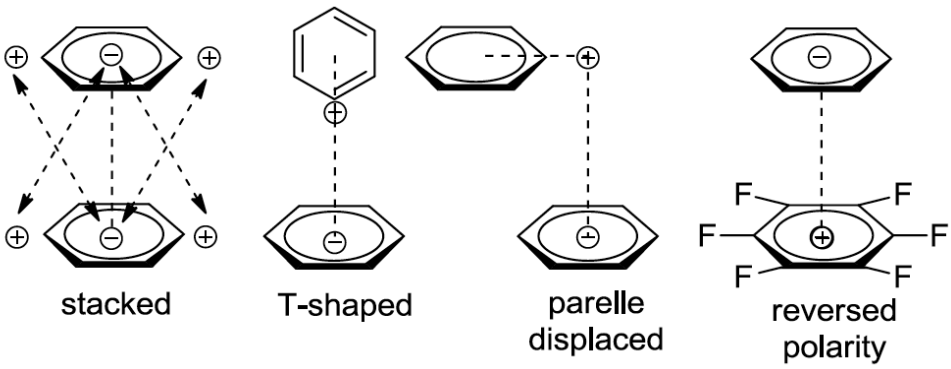


Figure 1.19. *Typical π–π aromatic interactions. Adapted from Nature,* ***2017.*** *543, 637-646.*

1.5.5.2. Cation-π interactions

These are non-covalent intermolecular interactions established between the face of an electron-rich π system and cation and energetically comparable to a typical hydrogen bond [161]. In the case of a ligand-receptor interaction, the π system usually comes from the receptor in the form of the amino acids containing aryl groups, while cations are supplied by the protonated amines (RNH3+), quaternary ammonium ions, sulfonium ions and carbocations (**Fig. 1.20**). On the other side, charged amino acids (Lysine and arginine) that could interact through cation- π interactions with ligands. Many examples can be highlighted where cation-π interactions occur showing the relevance of this type of interaction in drug discovery. For example, the ligand interactions with the GABA receptor, acetyl cholinesterase or the D2 dopamine receptor [161–164].

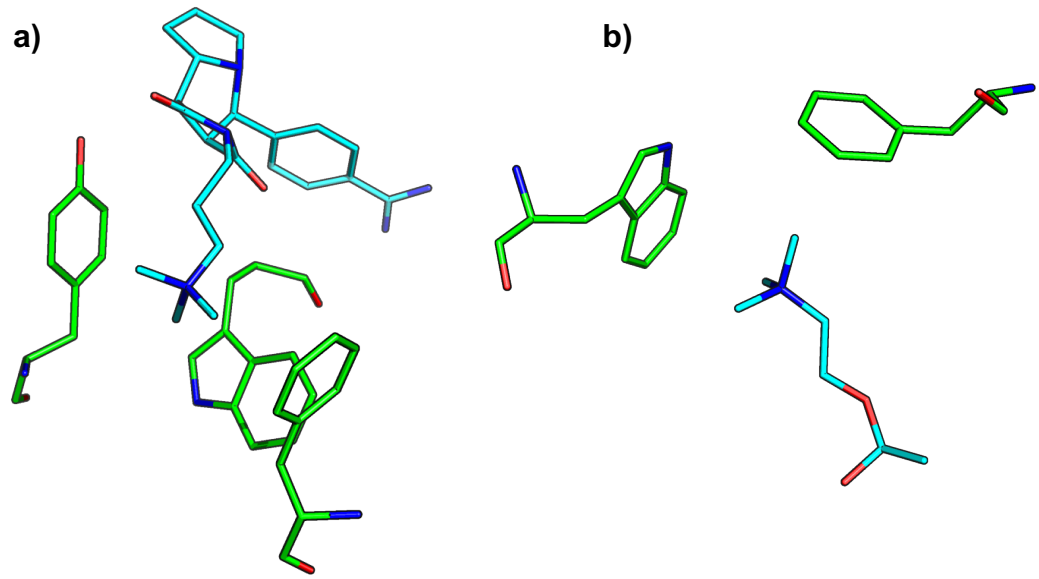


Figure 1.20. *Cation-π interactions.* ***a)*** *Ligand 78A interacting with the coagulation receptor factor X, the trimethylammonium moiety interacts with Tyr99, Phe174 and Trp215 (PDB code* ***2BOK****).* ***b)*** *Acetylcholine interacting with the acetylcholinesterase, the trimethylammonium moiety interacts with Trp84 and Phe330 (PDB code* ***2ACE****).*

1.6. References

1. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33.

2. Koga, H.; Itoh, A.; Murayama, S.; Suzue, S.; Irikura, T. Structure-activity relationships of antibacterial 6,7- and 7,8-disubstituted 1-alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic acids. *J. Med. Chem.* **1980**, *23*, 1358–1363.

3. Duncia, J. V.; Chiu, A. T.; Carini, D. J.; Gregory, G. B.; Johnson, A. L.; Price, W. A.; Wells, G. J.; Wong, P. C.; Calabrese, J. C.; Timmermans, P. B. M. W. M. The discovery of potent nonpeptide angiotensin II receptor antagonists: a new class of potent antihypertensives. *J. Med. Chem.* **1990**, *33*, 1312–1329.

4. Minor, W.; Dauter, Z.; Jaskolski, M. The young person's guide to the PDB. *Postepy Biochem.* **2016**, *62*, 242–249.

5. Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. **2017**, *1607*, 627–641.

6. Structures under scrutiny. *Nat. Chem. Biol.* **2015**, *11*, 301–301.

7. Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **2014**, *1*, 213–220.

8. Joosten, R. P.; Joosten, K.; Cohen, S. X.; Vriend, G.; Perrakis, A. Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. *Bioinformatics* **2011**, *27*, 3392–3398.

9. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.

10. Emsley, P.; Cowtan, K. Coot : model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.

11. Deller, M. C.; Rupp, B. Models of protein–ligand crystal structures: trust, but verify. *J. Comput. Aided. Mol. Des.* **2015**, *29*, 817–836.

12. Pozharski, E.; Deller, M. C.; Rupp, B. Validation of Protein–Ligand Crystal Structure Models: Small Molecule and Peptide Ligands. **2017**, *1607*, 611–625.

13. Walters, W. P. Virtual Chemical Libraries. *J. Med. Chem.* **2019**, *62*, 1116–1124.

14. John Harris, C.; D. Hill, R.; W. Sheppard, D.; J. Slater, M.; F.W. Stouten, P. The Design and Application of Target-Focused Compound Libraries. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 521–531.

15. Sheppard, D.; Lipkin, M.; Harris, C.; Catana, C.; Stouten, P. Strategies for Small Molecule Library Design. *Curr. Pharm. Des.* **2014**, *20*, 3314–3322.

16. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.

17. Monga, M.; Sausville, E. Developmental Therapeutics Program at the NCI: molecular target and drug discovery process. *Leukemia* **2002**, *16*, 520–526.

18. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng. Des. Sel.* **1999**, *12*, 85–94.

19. Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.

20. Lee, J.; Lee, D.; Park, H.; Coutsias, E. A.; Seok, C. Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 3428–3436.

21. Adhikari, A. N.; Peng, J.; Wilde, M.; Xu, J.; Freed, K. F.; Sosnick, T. R. Modeling large regions in proteins: Applications to loops, termini, and folding. *Protein Sci.* **2012**, *21*, 107–121.

22. Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363–382.

23. Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500–2501.

24. Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011**, *27*, 343–350.

25. Benkert, P.; Tosatto, S. C. E.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins Struct. Funct. Bioinforma.* **2008**, *71*, 261–277.

26. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput.*

Aided. Drug Des. **2011**, *7*, 146–57.

27. Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402–406.

28. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

29. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748.

30. Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.

31. Tripathi A, B. V. A. Molecular Docking: From Lock and Key to Combination Lock. *J. Mol. Med. Clin. Appl.* **2018**, *2*, 1–9.

32. Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114*, 2271–2278.

33. Zhong, H.; Tran, L. M.; Stang, J. L. Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *J. Mol. Graph. Model.* **2009**, *28*, 336–346.

34. Xu, M.; Lill, M. A. Induced fit docking, and the use of QM/MM methods in docking. *Drug Discov. Today Technol.* **2013**, *10*, 411–418.

35. Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160–3173.

36. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534–553.

37. Nabuurs, S. B.; Wagener, M.; de Vlieg, J. A Flexible Approach to

Induced Fit Docking. *J. Med. Chem.* **2007**, *50*, 6507–6518.

38. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins Struct. Funct. Genet.* **2002**, *47*, 409–443.

39. Ruiz-Carmona, S.; Schmidtke, P.; Luque, F. J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **2017**, *9*, 201–206.

40. Ruiz-Carmona, S.; Barril, X. Docking-undocking combination applied to the D3R Grand Challenge 2015. *J. Comput. Aided. Mol. Des.* **2016**, *30*, 805–815.

41. Novič, M.; Tibaut, T.; Anderluh, M.; Borišek, J.; Tomašič, T. The Comparison of Docking Search Algorithms and Scoring Functions. *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery*; 2016; p. 99–127.

42. Dias, R.; de Azevedo Jr., W. Molecular Docking Algorithms. *Curr. Drug Targets* **2008**, *9*, 1040–1047.

43. Lavecchia, A.; Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20*, 2839–2860.

44. Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. *J. Chem. Inf. Model.* **2008**, *48*, 2108–2117.

45. Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.* **2010**, *53*, 539–558.

46. Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694–698.

47. Warshel, A. Bicycle-pedal model for the first step in the vision process.

Nature **1976**, *260*, 679–683.

48. Offman, M. N.; Krol, M.; Silman, I.; Sussman, J. L.; Futerman, A. H. Molecular Basis of Reduced Glucosylceramidase Activity in the Most Common Gaucher Disease Mutant, N370S. *J. Biol. Chem.* **2010**, *285*, 42105–42114.

49. Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* **2006**, *14*, 437–449.

50. Köuppel, H.; Domcke, W.; Cederbaum, L. S. Multimode Molecular Dynamics Beyond the Born-Oppenheimer Approximation. *Adv. Chem. Phys.* **2007**, *52*, 59–246.

51. Monticelli, L.; Tieleman, D. P. Force Fields for Classical Molecular Dynamics.; 2013; p. 197–213.

52. Gavezzotti, A. *Molecular Aggregation*; Oxford University Press, 2006.

53. Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.

54. Schnur, D. M.; Dalton, D. R. MM2 force field parameters for oximes. *J. Org. Chem.* **1988**, *53*, 3313–3316.

55. Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

56. Allinger, N. L.; Chen, K.; Lii, J.-H. An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **1996**, *17*, 642–668.

57. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

58. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al.

CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

59. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.

60. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

61. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

62. Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338–7364.

63. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

64. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.

65. Maier, J. A.; Simmerling, C.; Wickstrom, L.; Hauser, K. E.; Martinez, C.; Kasavajhala, K. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

66. Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model

and Implicit Polarization. *J. Phys. Chem. B* **2013**, *117*, 2328–2338.

67. Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *J. Phys. Chem. B* **2012**, *116*, 6279–6287.

68. Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039.

69. Sajadi, F.; Rowley, C. N. Simulations of lipid bilayers using the CHARMM36 force field with the TIP3P-FB and TIP4P-FB water models. *PeerJ* **2018**, *6*, 5472–5698.

70. Hernandes, M.; Cavalcanti, S. M.; Moreira, D. R.; de Azevedo Junior, W.; Leite, A. C. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **2010**, *11*, 303–314.

71. Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564–2574.

72. Ibrahim, M. A. A. Molecular mechanical perspective on halogen bonding. *J. Mol. Model.* **2012**, *18*, 4625–4638.

73. Ibrahim, M. A. A. AMBER Empirical Potential Describes the Geometry and Energy of Noncovalent Halogen Interactions Better than Advanced Semiempirical Quantum Mechanical Method PM6-DH2X. *J. Phys. Chem. B* **2012**, *116*, 3659–3669.

74. Lin, F.-Y.; MacKerell, A. D. Polarizable Empirical Force Field for Halogen-Containing Compounds Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2018**, *14*, 1083–1098.

75. Bakken, V.; Helgaker, T. The efficient optimization of molecular geometries using redundant internal coordinates. *J. Chem. Phys.* **2002**, *117*, 9160–9174.

76. Henkelman, G.; Jónsson, H. A dimer method for finding saddle points

on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

77. Machado-Charry, E.; Béland, L. K.; Caliste, D.; Genovese, L.; Deutsch, T.; Mousseau, N.; Pochet, P. Optimized energy landscape exploration using the ab initio based activation-relaxation technique. *J. Chem. Phys.* **2011**, *135*, 34102–34111.

78. Mackay, D. H. J.; Cross, A. J.; Hagler, A. T. The Role of Energy Minimization in Simulation Strategies of Biomolecular Systems. *Prediction of Protein Structure and the Principles of Protein Conformation*; Springer US: Boston, MA, 1989; p. 317–358.

79. Paquet, E.; Viktor, H. L. Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review. *Biomed Res. Int.* **2015**, *2015*, 1–18.

80. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

81. de Leeuw, S. W.; Perram, J. W.; Smith, E. R. Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1980**, *373*, 27–56.

82. Makov, G.; Payne, M. C. Periodic boundary conditions in ab initio calculations. *Phys. Rev. B* **1995**, *51*, 4014–4022.

83. Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.

84. Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2019.

85. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson,

H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, and D. J. F. (Gaussian, Inc., Wallingford CT, 2009). *Gaussian 09*.

86. L DeLano, W. Pymol: An open-source molecular graphics tool. *{CCP4} Newsl. Protein Crystallogr.* **2002**, *40*.

87. Coutsias, E. A.; Wester, M. J. RMSD and Symmetry. *J. Comput. Chem.* **2019**, *40*, 1496–1508.

88. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.

89. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.

90. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

91. Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.

92. Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, Á.; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: An Ultrafast and Accurate Scoring Function for Protein–Protein Docking. *J. Chem. Theory Comput.* **2012**, *8*, 3395–3408.

93. Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.

94. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; 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*, 2695–2703.

95. Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632.

96. Wang, L.; Chambers, J.; Abel, R. Protein–Ligand Binding Free Energy Calculations with FEP+.; 2019; p. 201–232.

97. Schrödinger, L. Schrödinger Release 2019-3: FEP+ 2019.

98. Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043–2050.

99. Owens, J. Determining druggability. *Nat. Rev. Drug Discov.* **2007**, *6*, 187–187.

100. Chilingaryan, Z.; Yin, Z.; Oakley, A. J. Fragment-Based Screening by Protein Crystallography: Successes and Pitfalls. *Int. J. Mol. Sci.* **2012**, *13*, 12857–12879.

101. Badger, J. Crystallographic Fragment Screening. **2012**, *841*, 161–177.

102. Drwal, M. N.; Jacquemard, C.; Perez, C.; Desaphy, J.; Kellenberger, E. Do Fragments and Crystallization Additives Bind Similarly to Drug-like Ligands? *J. Chem. Inf. Model.* **2017**, *57*, 1197–1209.

103. Cuchillo, R.; Pinto-Gil, K.; Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *J. Chem. Theory Comput.* **2015**, *11*, 1292–1307.

104. Wedberg, R.; Abildskov, J.; Peters, G. H. Protein Dynamics in Organic Media at Varying Water Activity Studied by Molecular Dynamics Simulation.

J. Phys. Chem. B **2012**, *116*, 2575–2585.

105. Lama, D.; Brown, C. J.; Lane, D. P.; Verma, C. S. Gating by Tryptophan 73 Exposes a Cryptic Pocket at the Protein-Binding Interface of the Oncogenic eIF4E Protein. *Biochemistry* **2015**, *54*, 6535–6544.

106. Tan, Y. S.; Spring, D. R.; Abell, C.; Verma, C. The Use of Chlorobenzene as a Probe Molecule in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 1821–1827.

107. Uehara, S.; Tanaka, S. Cosolvent-Based Molecular Dynamics for Ensemble Docking: Practical Method for Generating Druggable Protein Conformations. *J. Chem. Inf. Model.* **2017**, *57*, 742–756.

108. Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 14257–14263.

109. *Saunders Comprehensive Veterinary Dictionary, 3 ed.*; Elsevier, I., Sud.; 2007.

110. Alvarez-Garcia, D.; Barril, X. Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. *J. Med. Chem.* **2014**, *57*, 8530–8539.

111. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276–3285.

112. Jurcik, A.; Bednar, D.; Byska, J.; Marques, S. M.; Furmanova, K.; Daniel, L.; Kokkonen, P.; Brezovsky, J.; Strnad, O.; Stourac, J.; et al. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **2018**, *34*, 3586–3588.

113. *Reviews in Computational Chemistry, Volume 31*; Parrill, A. L.; Lipkowitz, K. B., Sud.; Reviews in Computational Chemistry; Wiley, 2018.

114. Rechenberg, H. Erwin Schrödinger and the creation of wave mechanics. *Acta Phys. Pol. B* **1987**, *19*, 683–695.

115. McEvoy, J. P.; Zarate, O. *Introducing Quantum Theory*; Totem Book.; 2000.

116. Echenique, P.; Alonso, J. L. A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry. *Mol. Phys.* **2007**, *105*, 3057–3098.

117. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **2015**, *87*, 897–923.

118. Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315–2372.

119. Jensen, F. Atomic orbital basis sets. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 273–295.

120. Riley, K. E.; Op't Holt, B. T.; Merz, K. M. Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties. *J. Chem. Theory Comput.* **2007**, *3*, 407–433.

121. Peverati, R.; Truhlar, D. G. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372*, 1–52.

122. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

123. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

124. Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical

analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

125. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

126. Reckien, W.; Janetzko, F.; Peintinger, M. F.; Bredow, T. Implementation of empirical dispersion corrections to density functional theory for periodic systems. *J. Comput. Chem.* **2012**, *33*, 2023–2031.

127. Piñero, J.; Furlong, L. I.; Sanz, F. In silico models in drug development: where we are. *Curr. Opin. Pharmacol.* **2018**, *42*, 111–121.

128. Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* **2018**, *11*, 156–164.

129. Yang, C.; Valerio, L. G.; Arvidson, K. B. Computational Toxicology Approaches at the US Food and Drug Administration. *Altern. to Lab. Anim.* **2009**, *37*, 523–531.

130. Lin, A.; Giuliano, C. J.; Palladino, A.; John, K. M.; Abramowicz, C.; Yuan, M. Lou; Sausville, E. L.; Lukow, D. A.; Liu, L.; Chait, A. R.; et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **2019**, *11*, 1–18.

131. Valerio Jr., L. G. In silico toxicology for the pharmaceutical sciences. *Toxicol. Appl. Pharmacol.* **2009**, *241*, 356–370.

132. Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M. The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *Annu. Rev. Biophys.* **2018**, *47*, 223–250.

133. Reynolds, C. H.; Holloway, M. K. Thermodynamics of Ligand Binding and Efficiency. *ACS Med. Chem. Lett.* **2011**, *2*, 433–437.

134. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17*, 144–178.

135. Deganutti, G.; Moro, S. Estimation of kinetic and thermodynamic ligand-binding parameters using computational strategies. *Future Med. Chem.* **2017**, *9*, 507–523.

136. Ferreira de Freitas, R.; Schapira, M. A systematic analysis of atomic protein–ligand interactions in the PDB. *Medchemcomm* **2017**, *8*, 1970–1981.

137. Kurczab, R.; Śliwa, P.; Rataj, K.; Kafel, R.; Bojarski, A. J. Salt Bridge in Ligand–Protein Complexes—Systematic Theoretical and Statistical Investigations. *J. Chem. Inf. Model.* **2018**, *58*, 2224–2238.

138. Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, *3*, 604–617.

139. Zhou, S.; Wang, L. Unraveling the structural and chemical features of biological short hydrogen bonds. *Chem. Sci.* **2019**, *10*, 7734–7745.

140. Du, Q.-S.; Wang, Q.-Y.; Du, L.-Q.; Chen, D.; Huang, R.-B. Theoretical study on the polar hydrogen-π (Hp-π) interactions between protein side chains. *Chem. Cent. J.* **2013**, *7*, 92–100.

141. Sarkhel, S.; Desiraju, G. R. N-H...O, O-H...O, and C-H...O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins Struct. Funct. Bioinforma.* **2003**, *54*, 247–259.

142. Gilli, P.; Pretto, L.; Bertolasi, V.; Gilli, G. Predicting Hydrogen-Bond Strengths from Acid−Base Molecular Properties. The p K a Slide Rule: Toward the Solution of a Long-Lasting Problem. *Acc. Chem. Res.* **2009**, *42*, 33–44.

143. Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J.-Y.; Renault, E. The p K BHX Database: Toward a Better Understanding of Hydrogen-Bond Basicity for Medicinal Chemists. *J. Med. Chem.* **2009**, *52*, 4073–4086.

144. Kenny, P. W.; Montanari, C. A.; Prokopczyk, I. M.; Ribeiro, J. F. R.;

Sartori, G. R. Hydrogen Bond Basicity Prediction for Medicinal Chemistry Design. *J. Med. Chem.* **2016**, *59*, 4278–4288.

145. Hao, M.-H. Theoretical Calculation of Hydrogen-Bonding Strength for Drug Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 863–872.

146. Ramasubbu, N.; Parthasarathy, R.; Murray-Rust, P. Angular preferences of intermolecular forces around halogen centers: preferred directions of approach of electrophiles and nucleophiles around carbon-halogen bond. *J. Am. Chem. Soc.* **1986**, *108*, 4308–4314.

147. Ford, M. C.; Ho, P. S. Computational Tools To Model Halogen Bonds in Medicinal Chemistry. *J. Med. Chem.* **2016**, *59*, 1655–1670.

148. Xu, Z.; Yang, Z.; Liu, Y.; Lu, Y.; Chen, K.; Zhu, W. Halogen Bond: Its Role beyond Drug–Target Binding Affinity for Drug Discovery and Development. *J. Chem. Inf. Model.* **2014**, *54*, 69–78.

149. Beno, B. R.; Yeung, K.-S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A. A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. *J. Med. Chem.* **2015**, *58*, 4383–4438.

150. Murray, J. S.; Lane, P.; Politzer, P. Simultaneous σ-hole and hydrogen bonding by sulfur- and selenium-containing heterocycles. *Int. J. Quantum Chem.* **2008**, *108*, 2770–2781.

151. Markham, G. D.; Bock, C. W. Intramolecular sulfur-oxygen interactions: An ab initio molecular orbital and density functional theory investigation. *J. Mol. Struct. THEOCHEM* **1997**, *418*, 139–154.

152. Lin, S.; Wrobleski, S. T.; Hynes, J.; Pitt, S.; Zhang, R.; Fan, Y.; Doweyko, A. M.; Kish, K. F.; Sack, J. S.; Malley, M. F.; et al. Utilization of a nitrogen–sulfur nonbonding interaction in the design of new 2-aminothiazol-5-yl-pyrimidines as p38α MAP kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5864–5868.

153. Pal, D.; Chakrabarti, P. Different Types of Interactions Involving Cysteine Sulfhydryl Group in Proteins. *J. Biomol. Struct. Dyn.* **1998**, *15*, 1059–1072.

154. Schaeffer, L. The Role of Functional Groups in Drug–Receptor Interactions. *Pract. Med. Chem.* **2008**, 359–378.

155. Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.

156. Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chemie Int. Ed.* **2004**, *43*, 6596–6616.

157. Martinez, C. R.; Iverson, B. L. Rethinking the term "pi-stacking". *Chem. Sci.* **2012**, *3*, 2191–2201.

158. Zhao, R.; Zhang, R.-Q. A new insight into π–π stacking involving remarkable orbital interactions. *Phys. Chem. Chem. Phys.* **2016**, *18*, 25452–25457.

159. Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting non-covalent π interactions for catalyst design. *Nature* **2017**, *543*, 637–646.

160. McGaughey, G. B.; Gagné, M.; Rappé, A. K. π-Stacking Interactions. *J. Biol. Chem.* **1998**, *273*, 15458–15463.

161. Dougherty, D. A. The Cation−π Interaction. *Acc. Chem. Res.* **2013**, *46*, 885–893.

162. Dvir, H.; Silman, I.; Harel, M.; Rosenberry, T. L.; Sussman, J. L. Acetylcholinesterase: From 3D structure to function. *Chem. Biol. Interact.* **2010**, *187*, 10–22.

163. Lummis, S. C. R.; L. Beene, D.; Harrison, N. J.; Lester, H. A.; Dougherty, D. A. A Cation-π Binding Interaction with a Tyrosine in the Binding Site of the GABAC Receptor. *Chem. Biol.* **2005**, *12*, 993–997.

164. Dougherty, D. A. Cation-π Interactions Involving Aromatic Amino Acids. *J. Nutr.* **2007**, *137*, 1504–1508.

CHAPTER II

2. Protein Tyrosine Phosphatase

2.1. Introduction

Protein phosphorylation and the opposite process dephosphorylation, are the main tools of post-translational modifications (PTMs). Protein Kinases (PK) are in charge of the phosphorylation processes in which usually the adenosine triphosphate (ATP) acts as a phosphate donor, whereas Protein Phosphatases (PP) are in charge of the dephosphorylations. Therefore, the balance between the activity of both enzymes determine whether or not a substrate is phosphorylated and are critical in cell signaling (**Fig. 2.1**) [1].

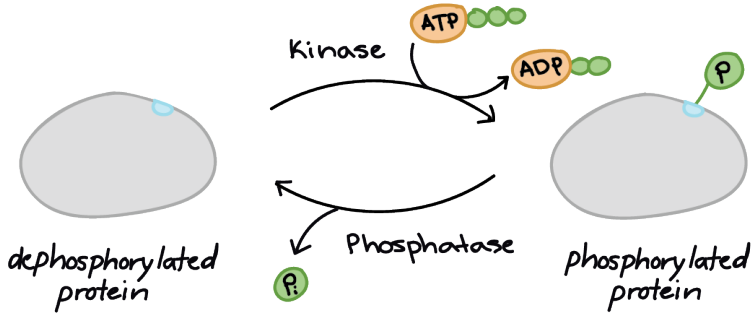


Figure 2.2 *General scheme for phosphorylation-dephosphorylation process in the post translational modifications.*

This reversible mechanism consists on the addition of a phosphate group (PO_4^{2-}) through a phosphoester bond to amino acids with a free hydroxyl group such as tyrosine, serine and threonine **(Fig. 2.2)**.

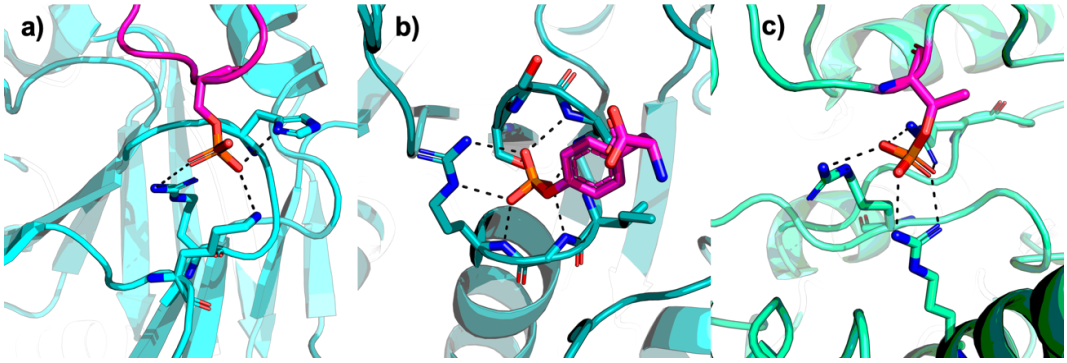


Figure 2.3. *General interaction between phosphorylated residues and basic amino acids.* ***a)*** *Phosphoserine interacting with arginine, lysine and histidine (PDB code* ***5JEJ****).* ***b)***

*Phosphotyrosine interacting with arginine and nitrogens from the backbone (PDB code **1PTY**). **c)** Phosphothreonine interacting with arginines (PDB code **2ERK**).*

Less common are the phosphoramidate bonds established between the nitrogens of histidine, lysine and arginine, and with the carboxylic groups of aspartic acid and glutamic acid through mixed anhydride linkages. Thus, these phosphorylated residues are termed non-canonical amino acids, which are currently a hot topic in molecular biology due to their peculiarity and the increased complexity of their analysis and isolation [2,3].

The phosphorylation process of the protein yields an increase of the polarity on a specific area of the protein, which can lead to changes on the conformation and subsequently on the interactions with other proteins. This biochemical process transforms notably the properties of the amino acid by the addition of three oxygens, rendering a total negative charge of two that increases the polarity and hydrogen bonding pattern. Not only amino acids are susceptible to phosphorylation and dephosphorylation; nucleotides, sugars and fatty acids can also undergo these processes. The human genome holds approximately 20.000 coding genes, which can produce a range of about 80.000 to almost 40.000 proteins due to alternative splicing [4,5]. It has been calculated that more than two-thirds of those proteins can be phosphorylated [6,7]. Moreover, there are more than one phosphorylation site for each protein, therefore the final number is quite challenging.

Some open-access bioinformatics projects such as PhosphoSitePlus (www.phosphosite.org) and Kinexus PhosphoNET (www.phosphonet.ca) provide comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, amongst others. These two websites list over 200,000 and 760,000 additional sites, respectively, that are likely to be phosphorylated. These numbers are huge, so a comprehensive understanding of the phosphorylation process remains a major challenge [8].

Nowadays, the human genome includes approximately 568 protein kinases

and 189 protein phosphatases that regulate phosphorylation events, although this number could vary depending on the source. These enzymes regulate many important cellular processes such as protein synthesis, cell division, signal transduction, cell growth and development, therefore any variation on their signal/regulation could alter the normal cell state thus converging in disease [9,10].

It is notable that the number of kinases is almost four times higher than the number of phosphatases; this difference is by no means negligible and acquires special relevance in drug discovery. Focusing on the common residues that are usually phosphorylated, it has been stipulated that serine residues are the most commonly modified with approximately 86% from the total, threonine 12% and only 2% of tyrosine residues are phosphorylated [10–12].

2.1.1. The Phosphatome and different PP classification

PPs are not only the counter parts of kinases; they are in essence critical signalling regulators by their own right, playing a crucial role under normal and pathological situations, and providing a novel approach to new therapeutic strategies never approached before. In contrast to the well-known kinome that was mentioned for the first time in 2002 by Gerard Manning and colleagues [13], the phosphatome has been recently stablished in 2017 by same authors (phosphatome.net) [14]. In contrast to kinases, that present a well stablished catalytic domain and reaction mechanism; phosphatases present different catalytic cores and reaction mechanisms, making their study intrinsically more complex. The study undergone by Gerard Manning, was carried out on nine different species, and shed light on the fact that phosphatases have evolved from different enzymatic ancestors, giving as a result 10 different protein folds or superfamilies, 21 families, and 178 subfamilies (**Table 2.1**) [14]. Previous to this phosphatome classification, PPs have been classified by other authors, such as Nicholas K. Tonks who stablished that PPs can be subdivided into three main families, depending on the residues that they dephosphorylate:

Ser/Thr phosphatases (PPP and PPM), Tyr phosphatases (PTPs) including dual-specificity phosphatases (DSPs) and the halo acid dehalogenases (HADs) [15].

Following this classification, the PTP family represents more than 60% of the total of general phosphatases, and nowadays has been one of the most studied and well know family. (**Fig. 2.3**) They have in common the presence of a catalytic cysteine defined by the active-site signature motif HCX_5R, in which the cysteine residue functions as a nucleophile and is essential for catalysis. PTPs can be categorized as receptor-like (RL) or non-transmembrane (NT) proteins myotubularins (MTMs) and dual specific phosphatases (DSPs) with a total number around 105, when their Protein Tyrosine Kinases (PTK) counterparts only reach up to 90 family members in humans [16]. PTP transmembrane receptor-like proteins (PTPR) include a transmembrane domain that translates a signal through ligand-controlled protein tyrosine dephosphorylation. From this group, a total of 8 subfamilies (R1-R8) have been described, some of which include an extracellular domain which can be involved in processes of cell–cell and cell–matrix contact and recognition. Subfamily R5 includes Protein Tyrosine Phosphatase Receptor type Z1 (PTPRZ1) and Protein Tyrosine Phosphatase Receptor type Gamma (PTPRG). The general motif in this family, in addition to the transmembrane segment, is the phosphatase domain D1 (PD1), which has the intrinsically active phosphatase activity. Some of them, also include a pseudophosphatase domain (PD2), structurally related to PD1 but without enzymatic activity, but with extreme relevance in the PTPR activity regulation and inhibition through the dimerization process (oligodimerization) [17–19].

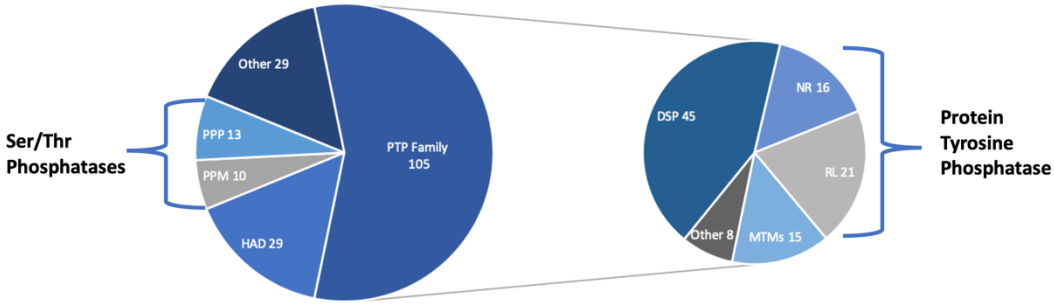


Figure 2.4. *General scheme of the phosphatase family by simplified classification.*

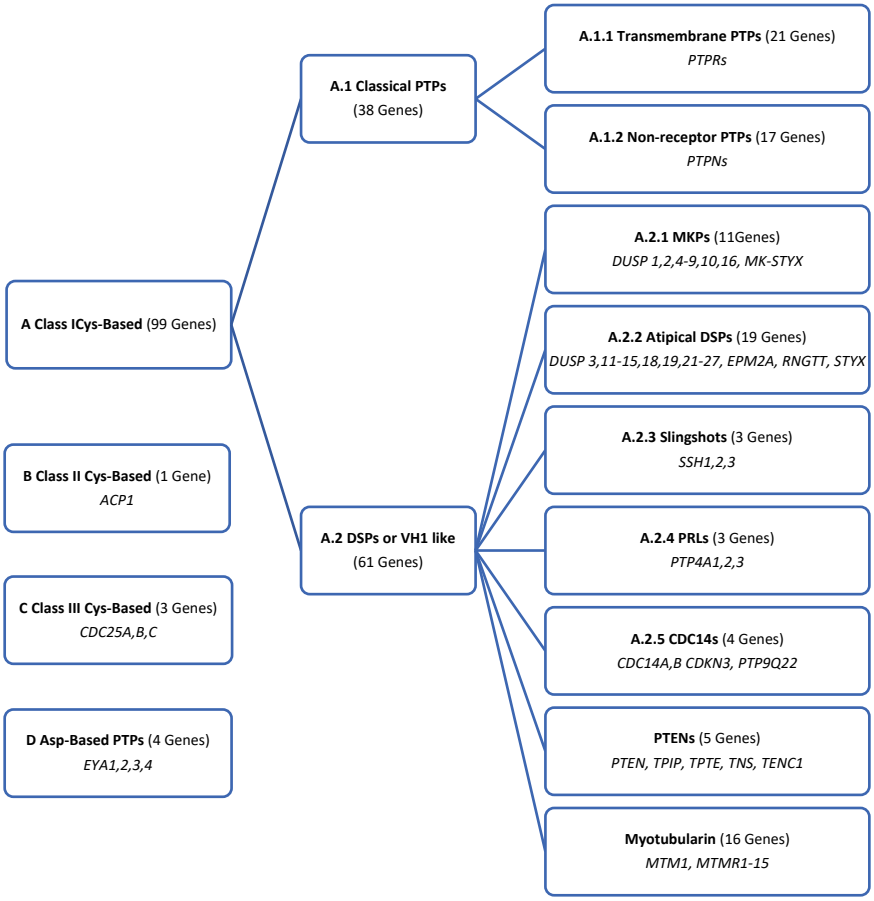
Table 2.1. *Classification of the phosphatome by superfamily and family with the corresponding substrate types. The presence or not in the human genome is also marked with the corresponding number.*

Superfamily	**Family**	**Substrate**	**Present in human / Gene number**
CC1 (Cysteine-based Class I)	PTP	pTyr	Yes/37
	DSP	pTyr, pSer/pThr, small molecule	Yes/40
	PTEN	lipid	Yes/8
	Myotubularin	lipid	Yes/15
	SAC	lipid	Yes/5
	Paladin	inactive	Yes/1
	OCA	?	No

CC2 (Cysteine-based Class II)	LMWPTP	pTyr	Yes/1
	SSU72	pSer	Yes/1
	ArsC	Arsenate	No
CC3 (Cysteine-based Class III)	CDC25	pTyr, pThr	Yes/3
HAD (HaloAcid Dehalogenase)	EYA	pTyr	Yes/4
	FCP	pSer	Yes/8
	NagD	pTyr, pSer	Yes/5
PPPL (PPP-Like)	PPP	pSer/pThr	Yes/13
	PAP	?	Yes/2
PPM (PPM)	PPM	pSer/pThr	Yes/20
A P (Alkaline phosphatase)	AP	pTyr, perhaps pSer/pThr	Yes/4
HP (Histidine phosphatase)	HP1	pTyr, pSer/pThr	Yes/12
	HP2	pTyr, small molecules	Yes/8
PHP (Histidine specific	PHP	pHis	Yes/1

phosphatase)			
RTR1 (CTD Phosphatase)	RTR1	pSer of CTD repeats	Yes1

The non-transmembrane PTPs, comprise approximately 280 residues including the catalytic domain. They are made up of several defined short sequence motifs as well as a regulatory sequence which can modulate and control the phosphatase activity subcellular distribution. This group is divided into 8 subfamilies and includes the well-known Protein Tyrosine Phosphatase type B (PTP1B) that will be discussed later. The DSP family of enzymes are the less conserved and the most structurally diverse group among all PTPs as they possess little sequence similarity in the catalytic domains if compared to the classical PTPs. This group includes about 43 encoding genes. They include the VH1-like related enzymes, which are well known to be involved in the immune response. The name of this classification comes from the feature that these enzymes can remove phosphate groups from serine (P-Ser) or threonine (P-Thr) residues as well as pTyr residues [20]. In addition to the classification carried out by *Nicholas K. Tonks*, the excellent review of *Alonso et al.* [21] in which a total of 107 human PPs have been classified, is recommended here for a detailed classification that has been outlined below (**Scheme 2.1**).



Scheme 2.1. *The Set of PTP Genes in the Human Genome classified by Alonso et al. [21].*

2.1.2. PTPs, the general structure

In 1988 the first PTP, in detail PTP1B, was characterized and biochemically reported by *K. Tonks et al.* [22]. It wasn't until 1994, when the first crystal structure of PTP was solved and deposited in the PDB database (entry ID **1PTU**) [23]. This was the structure of the 37-kDa PTP1B or PTPN1, member of the PTP family, which had been isolated from human placenta. Ever since, more than 264 3D structures of PTP1B have been deposited. In fact, PTP1B has been one of the top targets in drug discovery during more than 20 years due to its pharmacological profile involved in obesity and diabetes. This has not only provided structural knowledge thanks to the many crystal structures obtained with different substrates and ligands; but has also, allowed a deeper insight into the mechanism of action of PTPs brought about by the mutation studies carried out on this enzyme [24]. The general

catalytic domain of PTPs (cysteine-based Class I or Class I PTPs) have in common various folding motifs; all of them contain a central core with the highly conserved active site, consisting of a twisted β-sheet surrounded by several α-helices (**Fig. 2.4**). The active site motif is made up by the (H/V)CX_5R(S/T) sequence, being the X_5-segment the one that varies largely between PTPs subclasses. This motif also called the P-loop comes from the main α-helix (4) located on the central core of the globular domain. This highly conserved loop shares a common folding pattern among all cysteine based PTPs. It is notable that the catalytic cysteine is surrounded by the backbone nitrogen atoms of the α-helix; thus, providing a high basic environment that allows the deprotonation of the cysteine. Furthermore, the P-loop also includes a conserved arginine, which is involved in the substrate recognition process apart from its implication on the catalytic mechanism. This P-loop is surrounded by several loops (pTyr-loop, WPD-loop, E-loop and Q-loop) involved in the mechanism of action. The pTyr-loop is in charge of substrate recognition (pTyr) and reaction mechanism presenting a well-conserved tyrosine, which ensures the substrate binding through π-π interactions towards the active site.

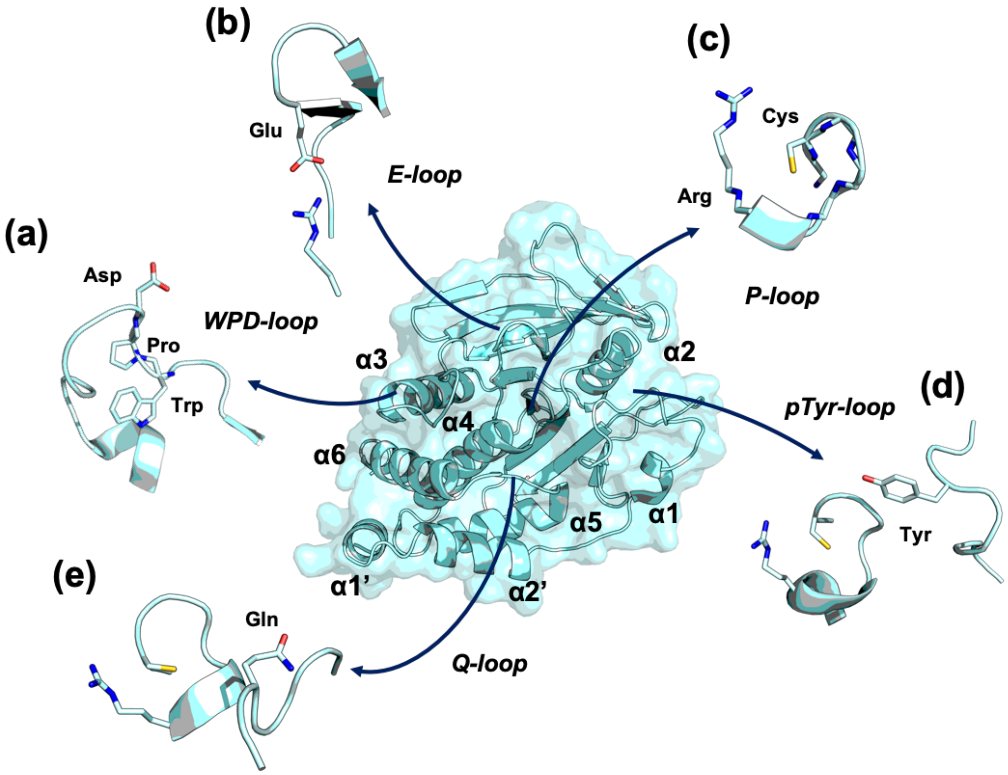


Figure 2.4. *General structure of the PTP class I catalytic domain.* ***(a)*** *WPD-loop;* ***(b)*** *E-loop;* ***(c)*** *Catalytic site or P-loop;* ***(d)*** *pTyr-loop;* ***(e)*** *Q-loop. Residues that are part of the corresponding loop are labeled.*

The WPD-loop (Trp-Pro-Asp) plays a special role in the different PTP conformations, a fact that will be discussed later, and is implicated in substrate recognition and catalytic cycle, also the WPD-loop includes a catalytic aspartic acid involved in the PTP recycling mechanism. The E-loop located adjacent to the active site, has been described to play a coordinate behavior by the interaction between the side chains of the Lys residue present in this loop and the Asp present in the WPD-loop. Finally, the Q-loop located next to the active site, plays a remarkable action by stabilizing and activating a water molecule implicated in the catalytic cycle through the side chain of a Gln [25–27]. On the other hand, the catalytic domains of Class II and III PTP differ from Class I. In Class II the P-loop is located at the N-terminal end of the phosphatase, whereas in Class I is located in the central core. As an example, the catalytically active cysteine of the bovine liver LMW-PTP (Class II) is located at position 12, in contrast this amino

acid occupies the position 215 on the soluble PTP1B type (Class I). Moreover, the P-loop in LMW-PTPs is surrounded by two loops: i) the V-loop that corresponds to the Q-loop in Class I and differs in length and sequence among the other PTP types; and, ii) the DPYY-loop that corresponds to the WPD-loop in Class I and is as close to the active site as the WPD-loop, but only presents a common glutamic acid (**Fig. 2.5**) [28].

In the Class III or CDC25 family, the α-helix following C-terminal after the P-loop matches well with the configuration found in Classes I and II. However, the catalytic domain of CDC25 contains a smaller number of β-strands and α-helices compared to Class I. Moreover, the loop regions surrounding the substrate binding pocket in Classes I and II is absent in CDC25, such us the pTyr-loop, WPD-loop and Q-loop (**Fig. 2.5**) [28]. All this results in that Class III is by far the less conserved PTP compared to the rest of the family.

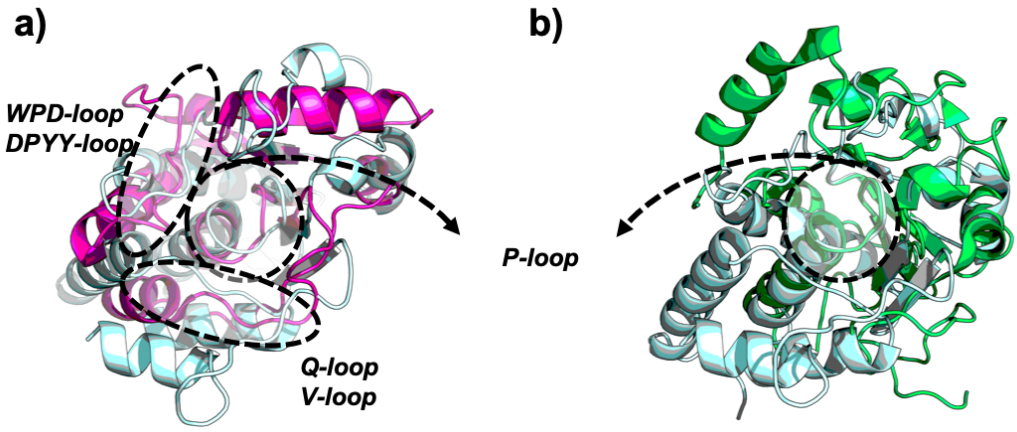


Figure 2.5. *Structural alignment of the catalytic domain of PTPs from Class I, II & III.* ***a)*** *The catalytic domain of the human Class I PTP VHR (PDB code* ***1VHR****) pale blue and Class II (PDB code* ***1DG9****) in pink. Dashed circles show the different positions occupied by the WPD-loop/DPYY loop and Q-loop/V-loop.* ***b)*** *Class I PTP pale blue and the Class III PTP, CDC25b (PDB code* ***1QB0****) in green superimposed where only the active site (P-loop) fits.*

2.1.3. The mechanism of action on PTPs

Catalysis by PTPs occurs through a two-step mechanism, also called ping-pong, involving a phosphocysteine intermediate [27]. The first step is the nucleophilic attack from the non-protonated cysteine (thiolate group) to the

phosphate ester moiety from the phosphotyrosine residue (**Fig. 2.6**). This first step releases the peptidyl tyrosine leaving the catalytic cysteine phosphorylated. The second step is the recycling of the catalytic amino acid through the hydrolysis of the phosphocysteine by an activated water molecule by as aspartic acid, yielding an inorganic phosphate group and the regenerated enzyme. During all this mechanism of action, the presence of an arginine is mandatory to further stabilize the different complexes.

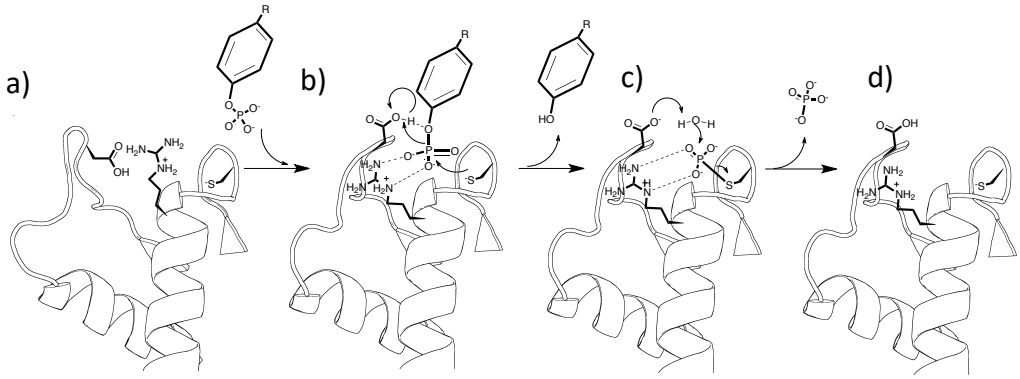


Figure 2.6. *General PTP dephosphorylation mechanism.* ***a)*** *Apo structure in an open conformation.* ***b)*** *Substrate binding leading to the closed conformation and the nucleophilic attack from the activated cysteine.* ***c)*** *Hydrolysis of the phosphocysteine by the catalytic water activated by the glutamic acid residue, and release of the inorganic phosphate.* ***d)*** *Recycling of the enzyme in the closed conformation ready to come back to the open conformation.*

Prior to the catalytic process, the substrate needs to be recognized and well positioned into the P-loop. PTPs recognize the substrate primarily through the pTyr-loop, which presents a tyrosine able to stablish a π-π interaction with the substrate. Not only the pTyr-loop is in charge of the substrate recognition, the E-loop and WPD-loop play an important role on this process as well as in the rearrangement of the conformation of the protein. The WPD-loop is usually in a so-called open conformation when no substrate or competitive ligand is attached. After substrate recognition, the WPD-loop closes by interacting with the P-loop leading to the so-called closed conformation. The high basic environment of the P-loop brought about by the amide protons and the conserved arginine residue induces the decrease of the pKa of the catalytic cysteine. This prompts the loss of the proton

yielding the nucleophilic thiolate group. As mentioned above, once the substrate is recognized, the pTyr-substrate locates the phospho group inside the P-loop establishing various hydrogen bonds and an ionic interaction between the negatively charged phosphate group and the positively charged arginine. The reaction begins with the nucleophilic attack of the thiolate group of the catalytic cysteine to the phosphate group of the substrate, cleaving the phosphoesther bond (**Fig. 2.6-b**). In this step, the protonated aspartic acid from the WPD-loop plays an acid role by donating the proton to the tyrosinate oxygen and releases the tyrosine. After that, the glutamate behaves as a base and takes a proton from the catalytic water, which is held tight in place by the glutamine side chain present in the Q-loop. The resulting hydroxyl group attacks the phosphoenzyme cleaving the phosphotioester bond and releasing the catalytic cysteine and an inorganic phosphate [25,29]. The complete reaction mechanism has been elucidated and corroborated in PTP1B by obtaining the corresponding reaction intermediates and Transition States Analogues (TSA) by means of X-ray crystallography (**Fig. 2.7**). The strategy followed by the different research groups to complete the structures of the reaction mechanism was to use inactive mutants and vanadate complexes for mimicking the theoretical TS. Vanadate has been described as a potent PTPs inhibitor as complexes based on vanadate have the advantage of displaying the same overall charge as phosphate and can easily derivatize and interact with ligands at their axial positions thus mimicking the substrate. The first complex between a PTP and the phosphotyrosine substrate was deposited in 1997 under the PDB code **1PTY** [30]**,** in the C215S PTP1B mutant form (**Fig. 2.7-b**). Serine decreases notably the catalytic activity from the P-loop giving the possibility to isolate this complex with its general substrate phosphotyrosine. Later on, in 2010 two different analogues of the theoretical transition states were obtained through the complexes with substrates that involve vanadate (**Fig. 2.7-c, e)**. The first one (TSA1) mimics the TS of the nucleophile attack by the presence of metavanadate (VO_3^-) and the peptide sequence DADEYL (**3I7Z**) which is a PTP substrate. The trigonal bipyramid geometry obtained

by the VO_3^- and the oxygen from the tyrosine substrate can reflect and mimic the non-stable phosphoderivative transition states (**Fig. 2.7-c**). The second TS analogue (TSA2) (**3I80**) mimics the hydrolysis of the phosphotioester bond by the presence of orthovanadate ion $(VO_4)^{3-}$. In water solution vanadate ions present a tetrahedral structure, while in complex with PTPs they form a stable pentavalent structure due to the presence of the extra thiolate cysteine (**Fig. 2.7-e**). This whole pathway of the reaction catalyzed by PTP1B was described by *Brandão et al.* [30] using the mentioned crystal structures and other also available in the PDB to fully understand the PTP mechanism of action (**Fig. 2.7**). For clarity, the PTP mechanism of action has also been drawn in 2D (**Fig. 2.8**). Not only vanadium has shown PTP inhibitory activity, nitrate has also been crystallized inside the active site (**4BJO**) showing its potential as inhibitor by mimicking the equatorial portion of the trigonal bipyramidal transition state [31].

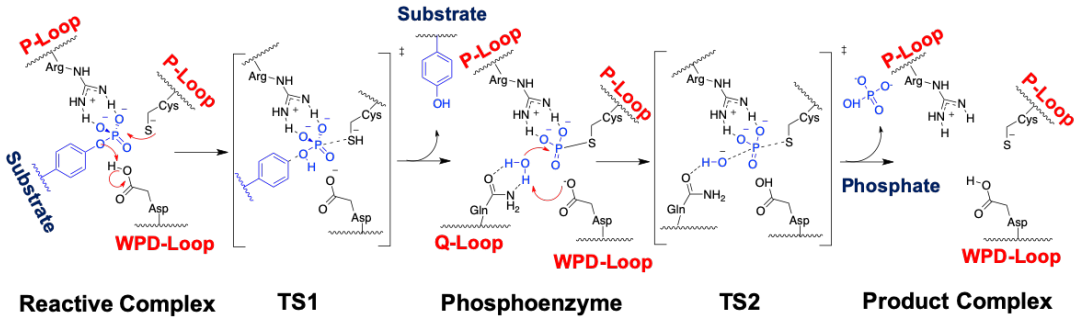


Figure 2.8. *General scheme of the reaction catalyzed by PTPs.*

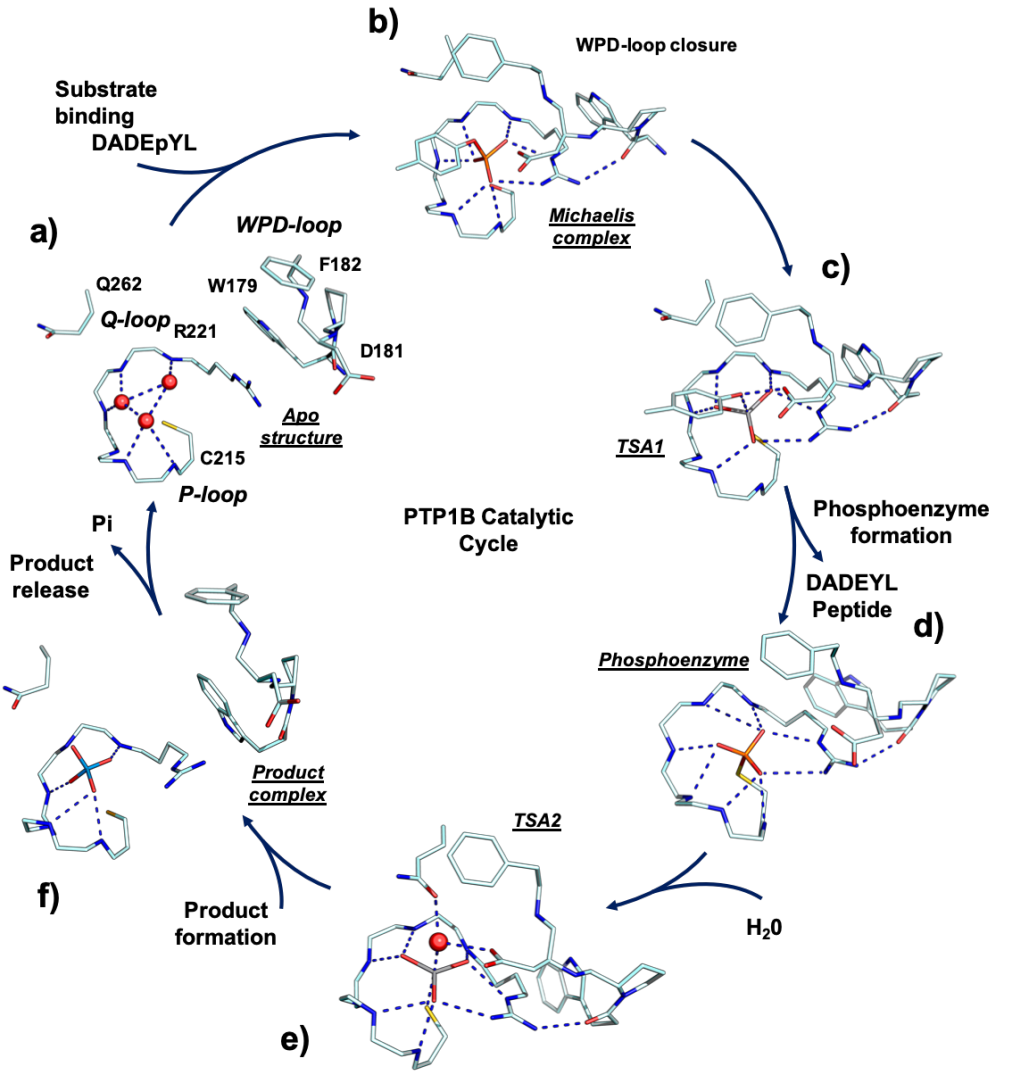


Figure 2.7. *The different crystal structures solved to date that corroborate the proposed mechanism for the phosphatase reaction in PTP1B. For clarity, only the backbone and some selected side chains are shown.* ***a)*** *PTP1B apoenzyme (**2CM2**).* ***b)*** *Michaelis complex between PTP1B C215S and the peptide DADEpYL (**1PTU**).* ***c)*** *First transition state (TSA1) complex between native PTP1B, metavanadate, and the Tyr in the peptide DADEYL (**3I7Z**);* ***d)*** *Mutant PTP1B Q262A cysteinyl-phosphate intermediate enzyme (**1A5Y**).* ***e)*** *Second transition state (TSA2) complex, between native PTP1B and orthovanadate (**3I80**).* ***f)*** *PTP1B-tungstate product analog complex (**2HNQ**).*

2.2. PTPs in human diseases

During the last 2 decades, PTPs have been showing their potential as druggable targets, mainly due to their implication in several diseases and pathological processes. Nevertheless, most recent research highlights the

relevance of phosphatases that contribute to human diseases and disorders. Yet they remain not well established druggable targets due to their structure similarity and lack of deeper studies when compared to kinases [32]. The variety of diseases were PTPs are implicated is huge, and varies from mental disorders to cancer, metabolic disorders, Parkinson, Alzheimer and other rare diseases. In general, the pathological process could come from alteration of the normal PTP activity, which could be linked either with its overexpression, depletion or faulty regulation. Furthermore, PTPs are implicated in numerous cell processes such as cell renewal, differentiation, embryonic development, osteogenesis, and angiogenesis, among others, making them easy culprits when the normal activity is altered. Here we present some of the most common pathological processes linked to PTPs, but there is likely still a plethora of unknown diseases that are yet to be studied.

2.2.1. Stem cells and differentiation processes

Stem cells are undifferentiated cells that can turn into specific cells, as the body needs them. Differentiation includes the transformation to any specific cell line such as neurons, muscle cells or epithelia. During this process, the SHP2, a cytoplasmic PTP encoded by the PTPN11 gene is required in the initial stages of development for cell differentiation. SHP2 has become a well-known target in drug discovery due to its implication in several types of cancers. The upregulated activity of this enzyme can be linked to increased cell proliferation, survival, and self-renewal [33,34]. On the contrary, SHP2-knockdown cells or cells with decreased SHP2 activity, have been linked to reduced proliferation and differentiation into hematopoietic cell lineages as well as with aberrant proliferation and differentiation [35]. The deficient expression of SHP2 was observed for the first time in the Noonan syndrome, an autosomal dominant disorder associated with several craniofacial abnormalities, short stature, cardiac defects and learning disabilities [36,37]. Other PTPs, such as PTPRσ, SHP1, PTP1B and PTPRζ, are also implicated in proliferation processes. PTPRσ or

PTPRSigma, negatively regulate the renewal of hematopoietic stem cells and knock out models have shown and increased hematopoietic growth [38]. SHP1 has been associated with adipogenic processes. Knockout mice models have increased adipogenic and decreased osteogenic markers that lead to an increase in adipose tissue and a reduction in bone mass [39,40]. The actions of PTP1B and PTPRζ, as well as PTPRγ will be discussed later.

2.2.2. Angiogenesis

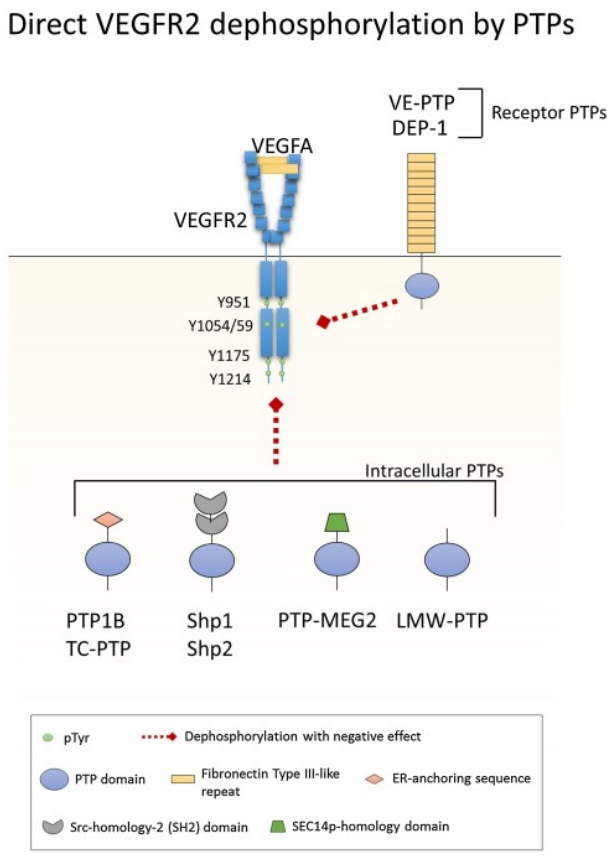


Figure 2.9. *Direct VEGFR2 dephosphorylation by PTPs. Figure taken from Pharmacological Research. 2017, 115, 107–123.*

Angiogenesis is a process that has been related to the action of the VEGF tyrosine kinase receptor family. The blood vessel formation is regulated by VEGF mediated signaling which is the natural ligand of the VEGF receptor (VEGFR) [41]. That induces VEGFR tyrosine phosphorylation, hence some PTPs play a critical role in its regulation. Angiogenesis is the process to form new blood vessels from pre-existing blood vessels by the sprouting of endothelial cells activated by VEGF. The endothelial cells possess many VEGFR, especially type II, which produces cell proliferation and migration towards the VEGF gradient, giving as a result the elongation of the blood vessel [42]. Needless to say, that this type of receptor is on the spotlight in drug discovery, especially in cancer research. One of the most important PTPs involved in phospho-VE-Cadherin levels is PTP1B (**Fig. 2.9**) [43]. The overexpression of PTP1B

inhibits VEGF mediated phosphorylation of VEGFR2 and reduces the phospho-VE-Cadherin levels. This reduces the cell-cell contacts between endothelial cells and disrupts the angiogenesis process [44]. Moreover, animal models with PTP1B deficient or PTP1B(-/-) knockout, have been useful to demonstrate enhanced angiogenesis concluding that the inhibition of PTP1B reduces endothelial dysfunction in various associated cardiovascular diseases [45,46]. Another PTP related to endothelial vascular activity is RPTPβ also known as vascular endothelial PTP (VE-PTP), implicated in blood vessel maintenance [47,48].

2.2.3. Autoimmune diseases

Some of the autoimmune pathologies including type 1 diabetes, rheumatoid arthritis and systemic lupus erythematosus, are related to the alteration of PTPN22 [49]. Recent studies have linked different polymorphisms (G788A and R263Q) with susceptibility to autoimmune diseases; however, the mechanism of action remains unclear [50]. TPN22 regulates the activity and effector functions of multiple important immune cell types, including lymphocytes, granulocytes and myeloid cells; and genetic alteration could increase the anti-tumour T-cell responses leading to autoimmune disorders [51].

2.2.4. Cancer

PTPs play an important role as novel pharmacological targets. Some reported examples to date constitute just the tip of the iceberg. Again, SHP2 has been classified as an oncogene with various gain-of-function mutants found in several types of leukemia as well as in prostate, lung and breast cancers [34,52]. Another PTP related with cancer is CDC25, which is overexpressed and altered in a large number of tumor types [53]. The "CDC" in its name refers to "Cell Division Cycle", therefore this enzyme regulates the transitions between cell-cycle phases (from G1 to S and from G2 to M phases) by dephosphorylating different types of CDKs (Cycline Dependent Kinases) [54,55]. Recent studies highlight the significance of

CDC25 overexpression in a subset of cancers with remarkably poor prognosis such as triple-negative breast cancer or endometrial cancer, among others [56,57]. Another common PTP involved in cancer is PTPN22, also known as the lymphoid tyrosine phosphatase (LYP) whose overexpression and lack of regulation could eventually lead to tumorigenesis processes, especially myeloid leukemia [58,59]. Other examples are phosphatases of regenerating liver-3 (PRL-3 or PTP4A3) that have been found to be overexpressed in the metastasis phase of colorectal cancer when compared to primary stages of the tumor.[60] In various human cancers the Dual specificity protein phosphatase 1 (DUSP1) has been observed to be abnormally expressed and has also been associated in patient prognosis. The molecular mechanisms of DUSP1 includes extracellular signal-regulation of different protein kinases (ERKs) and downregulation of p38 MAPKs and JNKs signaling by direct dephosphorylation [61]. Furthermore, DUSP1 has been found to play a role in tumor chemotherapy, immunotherapy, and biotherapy by inducing resistance to the pharmacological treatment [61–63].

2.3. PTPs in drug discovery

The number of publications that include sentences such as “protein tyrosine phosphatase inhibitors” have been increasing exponentially since 1990. A period that corresponds to two years after the first solved crystal structure of a PTP was deposited (***Fig. 2.10***). Since then, more than 300 articles have been published along more than 20 years, and the trend line seems to be continuing. However, despite the fact that some PTPs represent classical targets that are subjected to thorough investigation by scientists, there is still a large number of PTPs to be studied in more depth. Here, we discuss the most common targets, such as SHP2, CDC25, PRL3, and VEGFR among others, as recent advances and their druggable approach that have inspired the development of this thesis. Since PTPRZ1, PTPRG and PTP1B present a special situation, an exclusive section has been created to

address their peculiarities.

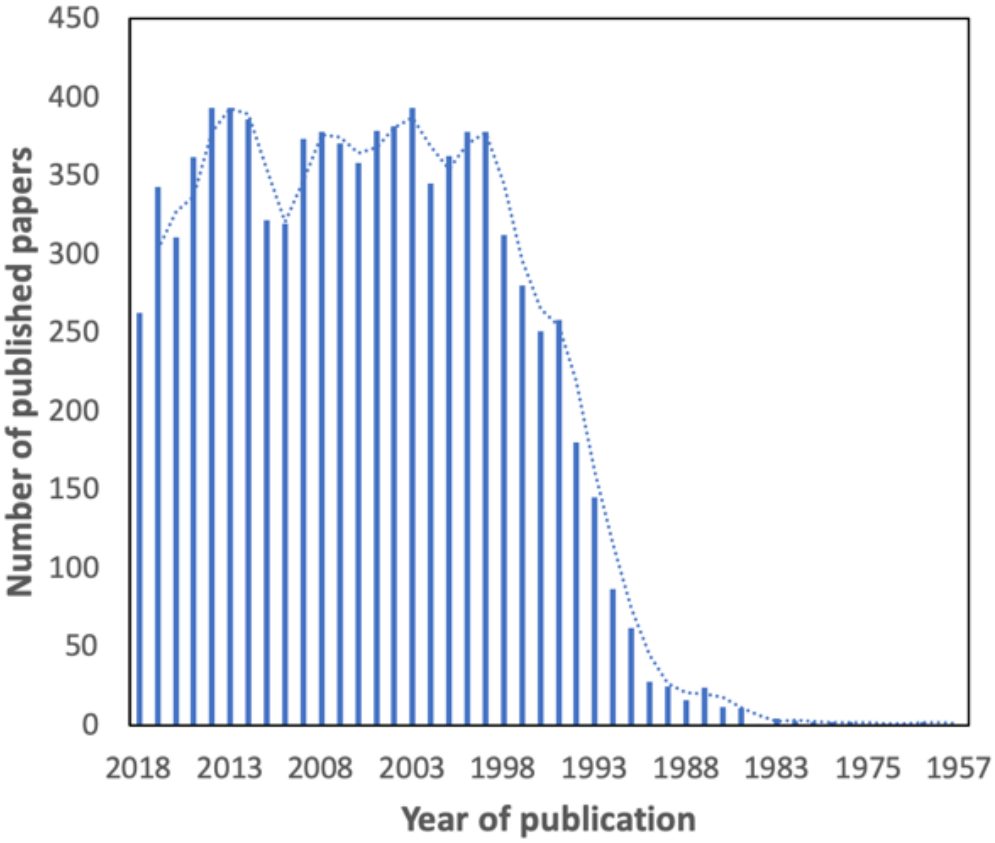


Figure 2.10. *Increase of publications related to PTPs in drug discovery. Search has been performed in Pubmed using words "protein tyrosine phosphatase inhibitors" (https://www.ncbi.nlm.nih.gov).*

2.3.1. SHP2 (PTPN11)

This target has been the object of numerous projects in drug discovery because it has been observed to be hyperactivated and/or dysregulated in multiple types of cancers and tumorigenic processes [52]. First attempts to develop a small molecule inhibitor targeting SHP2 resulted in **Estramustine phosphate** (***Fig. 2.11.***). This drug has been approved by the FDA for the treatment of prostate cancer, although it presents modest activity (IC_{50}=17 μM) and poor selectivity against other PTPs [64]. Later on, another screening campaign found a nonconserved pocket proximal to the catalytic domain of SHP2 (***Fig. 2.12***), which led to the development of a potent

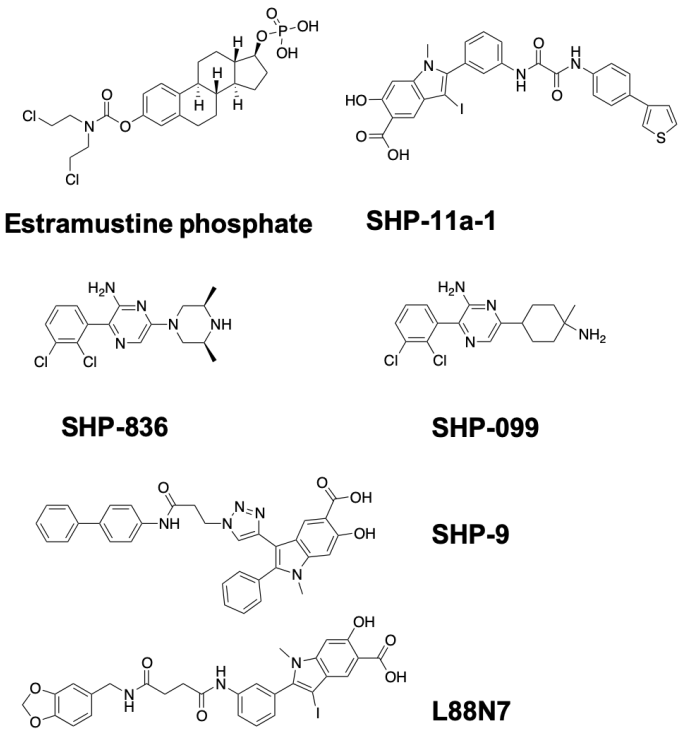


Figure 2.11. Developed small molecule SHP2inhibitors.

bidentate inhibitor with quite good selectivity profile against other PTPs. **SHP-11a-1** developed by *Zeng et al.* [65] presented inhibitory activity (IC_{50} = 200 nM) and more than 5-fold in vitro selectivity against 20 other PTPs. This compound has demonstrated direct activity in cells, proving the antiproliferative cancer action in lung cancer cell lines, and, due to its suitable ADME properties, has also been tested in melanoma-bearing mice models [33].

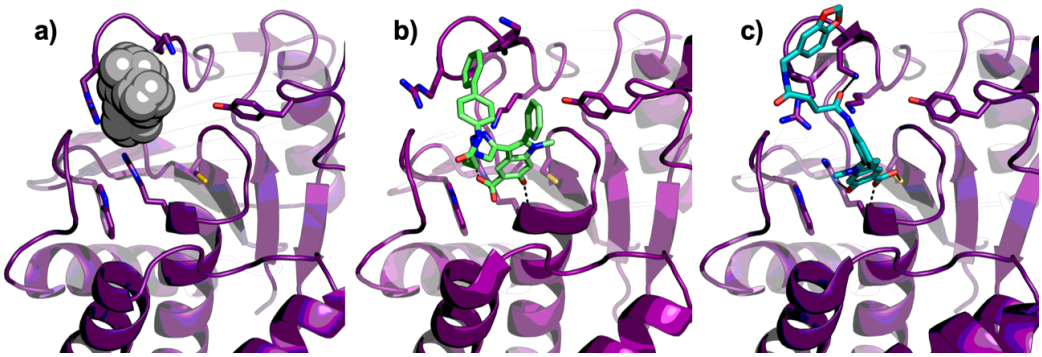


***Figure 2.12. a)** Apo SHP2 PDB Code **3B70**, spheres represent the non-conserved pocket beside the active site. **b)** Complex between SHP2 and **SHP-9** synthesized by Zeng et al, PDB Code **3O5X**. **c)** Complex between SHP2 and Compound **L88N79** synthesized by Zeng et al, PDB Code **4PVG**. **SHP-9** and **L88N79** interact with SHP2 in a similar binding mode than **SHP-11a-1** with no crystal structure reported.*

Most recently, inhibitor **SHP099** (IC_{50} = 0.071 μM in **5HER**), and its precursor **SHP836** (IC_{50} = 5 μM), have been depicted out of 900 primary

hits, by a high throughput screen (HTS) carried out on an enzyme allosterically activated by a phosphorylated peptide [66]. Both compounds bind to a specific allosteric site, a hydrophobic pocket formed by the interface of the so-called N-SH2 and C-SH2 domains (other domains noted apart from the catalytic domain which complete the whole SHP2 structure), and PTP domains (**Fig. 2.13-a**). The interactions between these compounds and those SHP2 domains inhibit the catalytic activity through stabilization of the inactive conformation. These molecules have opened the door to a novel approach in the development of SHP2 inhibitors, as these are a good example of potent, selective and orally bioavailable allosteric specific SHP2 inhibitors that present efficacy in patient-derived tumor xenograft models [66]. The mechanism of action of these allosteric compounds is based on the intrinsic SHP2 ability to exchange between closed (inactive) and open (active) conformations by stabilizing the SHP2 in the inactive conformation with the N-terminal SH2 domain blocking the active site (**Fig. 2.13-b**).

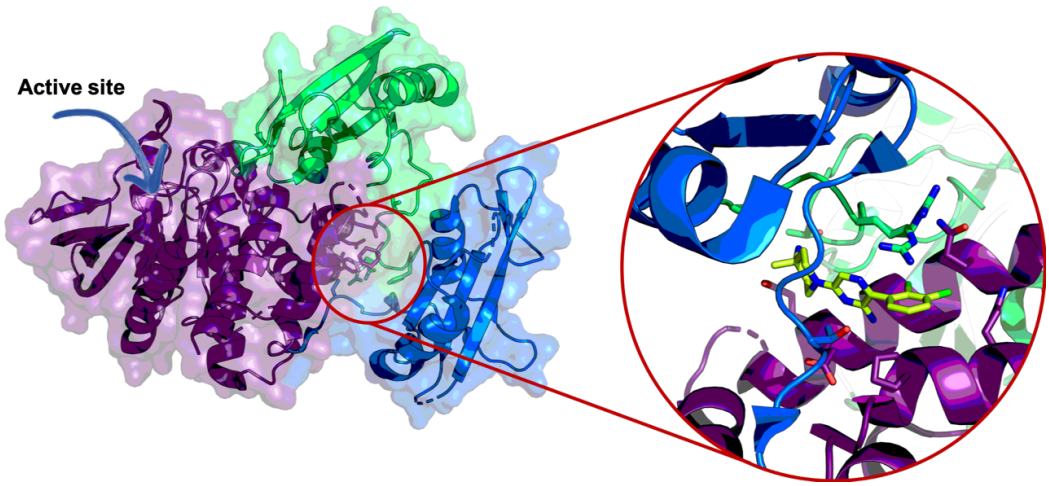


Figure 2.13-a. *Action mechanism of allosteric compounds in complex with SHP2. In blue C-SH2 Domain, green C-SH2 and purple catalytic domain. Whole system is* ***2SHP****, and the interaction with* ***SHP099*** *that corresponds to* ***5EHR****.*

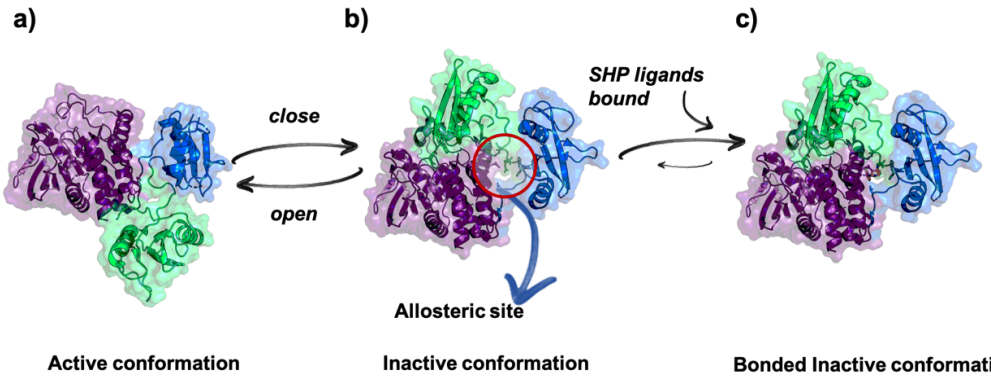


Figure 2.13-b. *The mechanism of action of SHP derivatives. The SHP2 alternates between opened* ***a)*** *and closed* ***b)*** *conformation,* ***6CRF*** *and* ***4DGP****, respectively.* ***c)*** *SHP molecules stabilize the inactive conformation* ***6CMS****.*

2.3.2. SHP1 (PTPN6)

This is one of the unique cases in PTPs, in which the process of drug discovery started by the development of an agonist instead of the classical inhibitor. **Sorafenib** and **Regorafenib** increase the phosphatase activity of SHP1 by direct interaction, depleting the autoinhibition of SHP1 by interfering with the inhibitory N-SH2 domain (**Fig. 2.14**) [67,68]. N-SH2 intrinsically blocks the SHP1 active site in the inactive conformation, when the N-SH2 is shifted away the SHP1 recovers its activity. **Sorafenib** and **Regorafenib** stabilize the shifted conformation of N-SH2. From this fact, analogues have been developed inspired by those two molecules, like compounds **SC-40** and **SC-43** that present better activity both in vitro and in vivo. Moreover, combination of these two molecules with other antitumorigenic agents has shown substantial cell apoptosis in hepatocellular carcinomas and breast cancer, making SHP1 a promising anticancer lead for further therapeutic development [69]. Studies published in 2018 by the same group, reported that **SC-43** and **SC-78** present an interesting activity against the Colorectal Cancer (CRC) impairing that of **regorafenib** [70]**.** The clear advantages of these compounds over **regorafenib** is that **SC** derivatives do not present the amide-pyridine ring, which can bind in the ATP-binding domain in kinases, avoiding the binding to multiple kinases. However, **regorafenib** is nowadays an approved

molecule in the treatment of CRC [71].

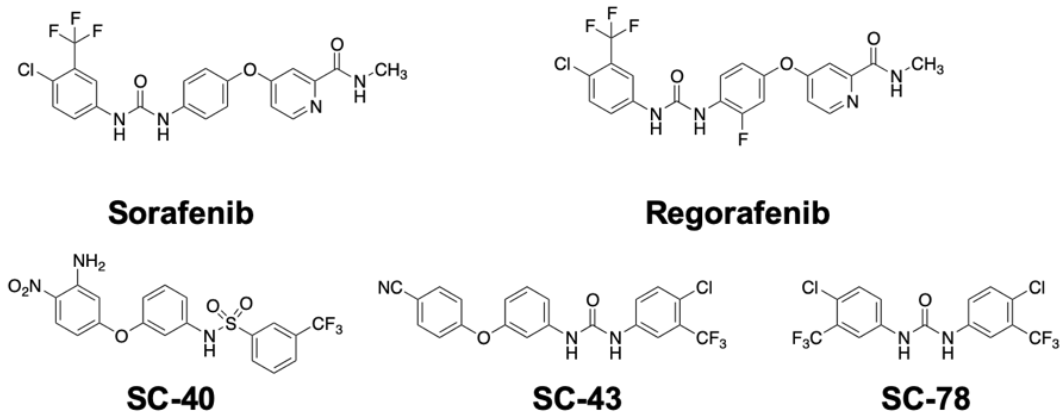


Figure 2.14. *Developed small molecules against SHP1.*

2.3.3. CDC25

As mentioned before, CDC25 plays a critical role in regulating cell cycle progression. It has been observed to be involved in many human tumors, which has made it to be a well-defined target in drug discovery. One of the first discovered molecules was the covalent inhibitor **DA-3003-1** (**NSC 663284**) (**Fig. 2.15**) [72]. **DA-3003-1** presents in vitro Ki values of 29, 95, and 89 nM for CDC25 subtypes A, B and C, respectively, and also some in vivo activity at 5 mg/kg in colon cancer cell lines HT 29 injected in mice [73]. Another CDC25 irreversible inhibitor is **IRC-083864** (**Fig. 2.15**) that presents a redox mechanism of action [74]. This bisquinone inhibitor possesses an IC_{50} value against CDC25A, B and C of ~20–50 nM and some selectivity against others PTPs. The oral administration of **IRC-083864** was also found to be effective in some types of cancers on xenograft models. Another bisquinone, **CDC25-13** presents a Ki of ~2 μM (IC_{50} 3.18 μM) against CDC25B with a reversible competitive inhibition mechanism, has proven to mediate as a bypass of the DNA damage-induced mitotic checkpoint blockade and proliferation of cancer cells [75]. The most recent developed inhibitors are compounds **CDC25-1** and **CDC25-7** (PDB Codes **4WH7** for **CDC25-1** and **4WH9** for **CDC25-7)** with an IC_{50} of 5μM and 1 μM, respectively (**Fig. 2.15**). These molecules were obtained by a fragment-

based screening approach that interestingly bind in an allosteric site positioned at ~15 Å away from the active site (**Fig. 2.16**) [76].

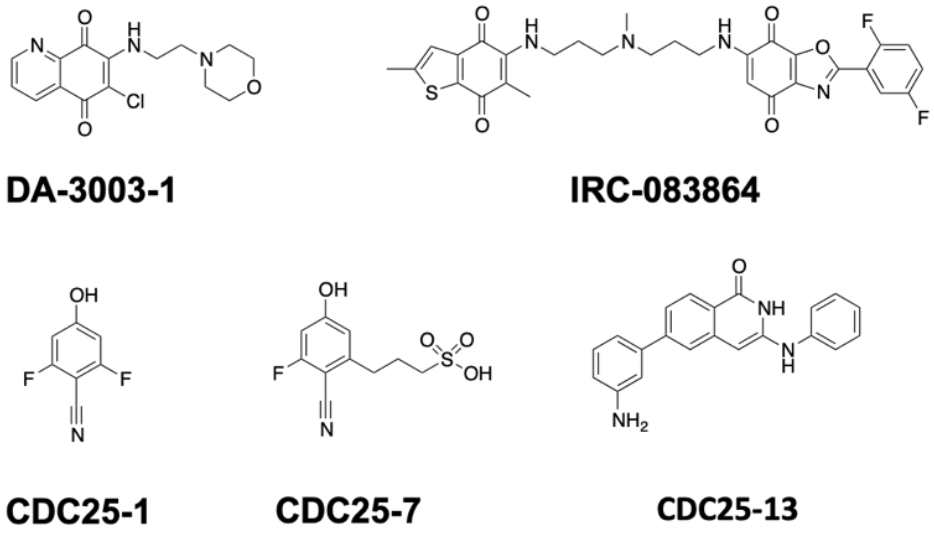


Figure 2.15. *Developed small molecules against CDC25.*

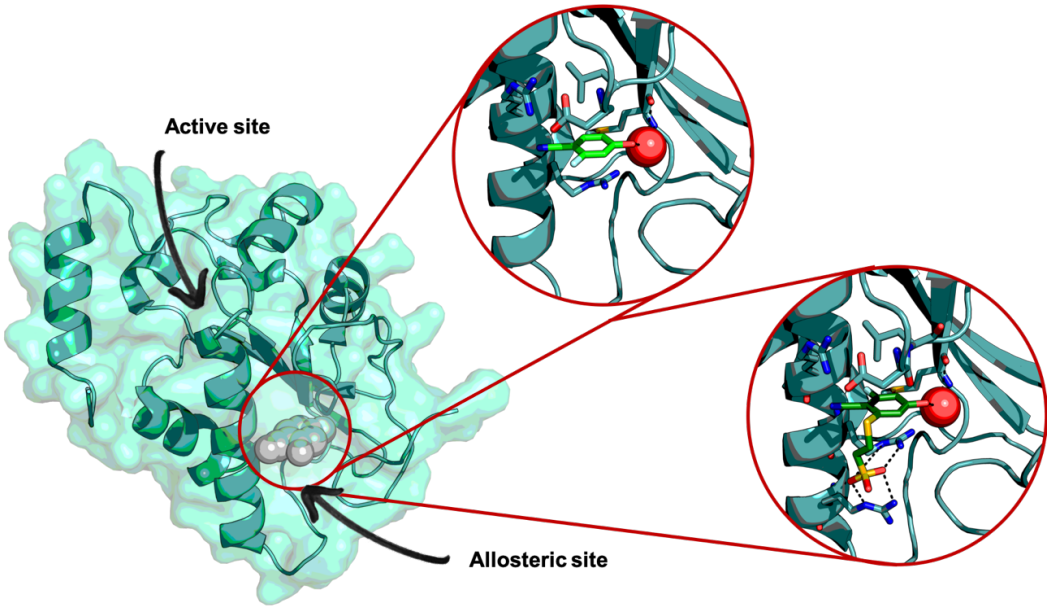


Figure 2.16. *Action mechanism of the allosteric inhibitors* ***CDC25-1*** *(****4WH7*** *compound colored in green) and* ***CDC25-7*** *(****4WH9*** *compound colored in pale green) in complex with CDC25. The red spheres represent a bound water molecule in the complex, and the grey spheres represent the allosteric pocket on the surface in contact with CDK2.*

The proposed action mechanism of this type of compounds is the disruption of the protein-protein interaction between CDC25B and CDK2/cyclin A, thereby inhibiting the dephosphorylation of CDK2. Even though **CDC25-7** is

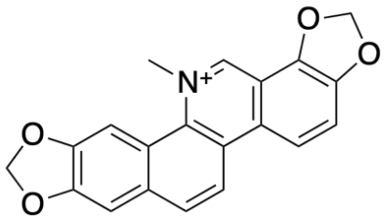
a very weak inhibitor, it constitutes an important starting point for follow-up SAR development of more potent inhibitors using this ppi (protein-protein interaction) disrupting approach in CDC25.

2.3.4. DUSP1

HTS assay has been also the starting point of DUSP1 hits [77]. The screened molecules came from the chemical library of the National Institutes of Health Molecular Libraries Screening Center Network (MLSCN), also now known as The Molecular Libraries and Imaging (MLI) Program (https://commonfund.nih.gov/molecularlibraries/index). After the screening of the 65,000 compounds of this library, 46 compounds were identified with IC_{50} values <50 μM, as well as four with IC_{50} values < 1.0 μM. Unfortunately, these proposed compounds presented chemical reactivity, poor solubility, and lack of cellular potency. This example shows us how difficult the hit search in PTPs could be due to its intrinsical activity and structure. Another described molecule against DUSP1 is **sanguinarine**, a natural product obtained from the screening of The MicroSource Natural Products Library (Discovery Systems, Inc., Gaylordsville, CT, http://www.msdiscovery.com/natprod.html), which is a collection of more than 720- pure natural products and their derivatives (nowadays this chemical library contains 800 pure compounds) [78]. From this assay, **sanguinarine** was selected as hit with an IC_{50} = 17,3 μM against DUSP1. (**Fig. 2.17**). This alkaloid comes from the toxic plant *Chelidonium majus* (Greater Celandine), the extracts of which have long been used in traditional herbal medicine. More documented properties of sanguinarine include antiviral, antimicrobial, and even antitumor activity [79]. Unfortunately, the lack of a potent and selective DUSP1 inhibitor, plus the lack of available crystal structures with bound small molecules, have made of DUSP1 a neglected target showing that upcoming research will be necessary to

assess the feasibility of this target.

Chelidonium majus



Sanguinarine

Figure 2.17. *The herbaceous perennial plant Chelidonium majus and one of its corresponding alkaloids sanguinarine.*

2.3.5. PRL3

Also known as PTP4A-1, -2 &-3, has been studied as target in cancer therapy due to its role in cell migration, invasion, proliferation, and tumor growth. The first discovered small molecule against PRL3 was the approved drug pentamidine in 2002, but its potency and selectivity were quite low among the multiple side effects of this controversial molecule [80]. In 2006 a potent inhibitor **BR-1** was described with an in vitro IC_{50} of 0.9 μM (***Fig. 2.18***) [81]. In 2008, a more selective and potent inhibitor, the **thienopyridone**, with an IC_{50} of 0.457 μM, was discovered by screening the Roche chemical library against PRL3, and showing also promising in vivo results [82]. Moreover, the thienopyridine photooxygenation yielded a more potent PTP4A3 inhibitor (**JMS-053**) with an IC_{50} of 18 nM. This inhibitor proved to be the most potent molecule described at this time [83]. A recent publication has shown that this increased activity of the oxidized thienopyridine comes mainly from its capacity to oxidize the catalytic cysteine from PRL3 depleting the nucleophilic feature of the catalytic cysteine of this phosphatase [84]. The proposed binding modes seem to be allosteric, near the WPD-loop of the enzyme stabilizing the closed conformation due to the non-competitive inhibition kinetics. Nevertheless, more studies of this binding mode need to be performed to confirm such hypothesis. Most recently, a first-in-class humanized antibody (PRL3-zumab) has been developed and has reached phase 1 clinical trials

(NCT03191682), constituting one of the first antibodies against PTPs in study (clinicaltrials.gov) [85].

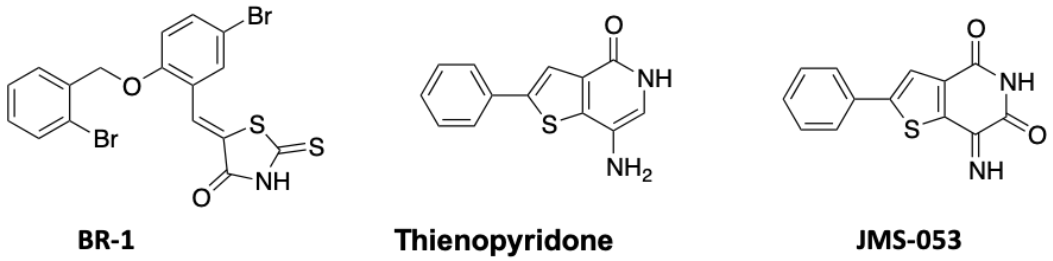


Figure 2.18. *Developed small molecules against PRL3.*

2.4. PTPRZ1

The Receptor-type tyrosine-protein phosphatase zeta (PTPRZ1 or PTPRζ UniProtKB - P23471) is a transmembrane phosphatase member of the CC1, Cysteine-based Class I superfamily (**Table 2.2**). PTPRZ1 has three different isoforms P23471-1, -2, -3 being the first one the canonical. This receptor is made up of an extracellular domain (aa 25–1636), a transmembrane domain (aa 1637–1662) and an intracellular domain (aa 1663–2315). The extracellular domain is formed by an Alpha-carbonic anhydrase (CA aa 36-300) and a Fibronectin type-III (FN-III aa 314-413) domain. The intracellular domain is formed by a Tyrosine-protein phosphatase domain 1 (PD1 aa 1717-1992) and a Tyrosine-protein phosphatase domain 2 (PD2 aa 2023-2282) being the first one the only one active. The transmembrane domain is formed by a helix (aa 1637 – 1662) (**Fig. 2.19**). The catalytic cysteine is positioned at 1933 and the substrate-binding site is located between positions 1901-1977. There are 24 sites for glycosylation, 2 disulfide bonds (aa 56 ↔ 240 and 133 ↔ 264) and 5 phosphorylation sites (aa 637, 639, 1684, 1687 and 2055) [87,88]. PTPRZ1 is a tissue specific phosphatase, mainly expressed in the central nervous system (CNS), specifically in astrocytes and subsets of neurons, where it is localized in the Purkinje cell layer of the cerebellum, the dentate gyrus, and the subependymal layer of the anterior horn of the lateral ventricle [89].

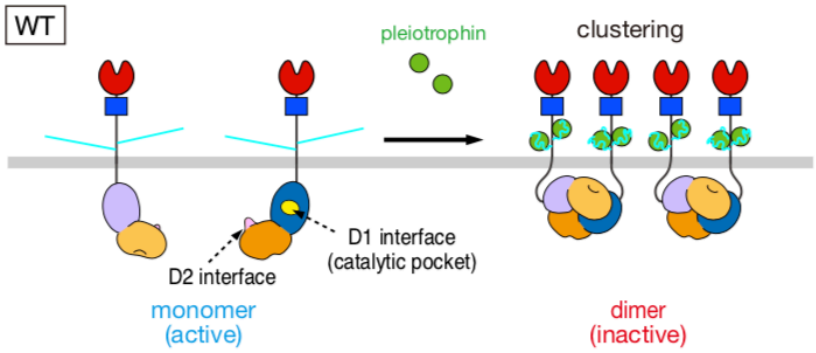


Figure 2.19. *General structure of the PTPRZ1, showing the oligodimerization and the head to toe mechanism upon PTN binding. Figure taken from* [86].

On the CNS the carbonic-anhydrase like domain could interact with contactin (Neural Cell Adhesion Protein CNTN1 UniProtKB - Q12860), midkine (MK UniProtKB - P21741) and pleiotrophin (PTN UniProtKB - P63089), those ligand-receptor interactions deplete the intrinsic phosphatase activity of the PD1 increasing the number of phosphorylated residues on the inside the neuronal tissue [87,90,91].

2.4.1. The PTPRZ1 signaling pathway

Has mentioned before, PTPRZ1 is intrinsically active. Only when the natural ligand attaches to the extracellular domain, the intracellular domain oligo-dimerizes blocking its phosphatase activity. At the molecular level, the proposed ligand-receptor mechanism was for the first time reported in 2009, by *Barr AJ et al,* a novel "head to toe dimer" structure for the PTPRG cytoplasmic domains was solved where the PD1 interacted with PD2 and vice versa (**2NLK**). This novel crystal structure allowed the understanding of how the PD2 was able to block the phosphatase activity of PD1, both coming from different receptors (**Fig. 1.19**) [92]. Ten years later, this same conformation has been solved for PTPRZ1 by *A. Fujikawa et al* (PDB code **6J6U**) [86]. The well-known PTPRZ1 phosphorylation substrates are β-catenin, Fyn, ALK, β-adducin, GIT1/Cat-1 and p190/RhoGAP (**Fig. 2.20**). In normal conditions, β-catenin stabilizes the cytoskeleton through its interaction with cell adhesion molecules (CAM), cadherin and intracellular

cytoskeletal protein actin, while its phosphorylation deteriorates this binding contributing to the contact loss breaking down of the cell adhesion and remodeling of the cytoskeleton [93]. Fyn is an intrinsically inactive Tyrosine-kinase, but after the phosphorylation process (induced by the PTN effect) it regulates cytoskeletal remodeling process by phosphorylating several microtubule-associated proteins such as MAP2 and MAPT. It also plays a role in neural processes by phosphorylating other substrates such as DPYSL2 (Dihydropyrimidinase-related protein 2 or DRP-2), a multifunctional adapter protein within the central nervous system, ARHGAP32 (Rho GTPase-activating protein 32), a regulator for Rho family GTPases implicated in various neural functions, and SNCA, a small pre-synaptic protein [94,95]. β-adducin is a membrane-cytoskeleton-associated protein that promotes the assembly of the spectrin-actin network, thus maintaining the stability of the cell membrane and cytoskeleton. Furthermore, β-adducin can regulate certain transcription factors due to its ability to module chromatin stability [96,97].

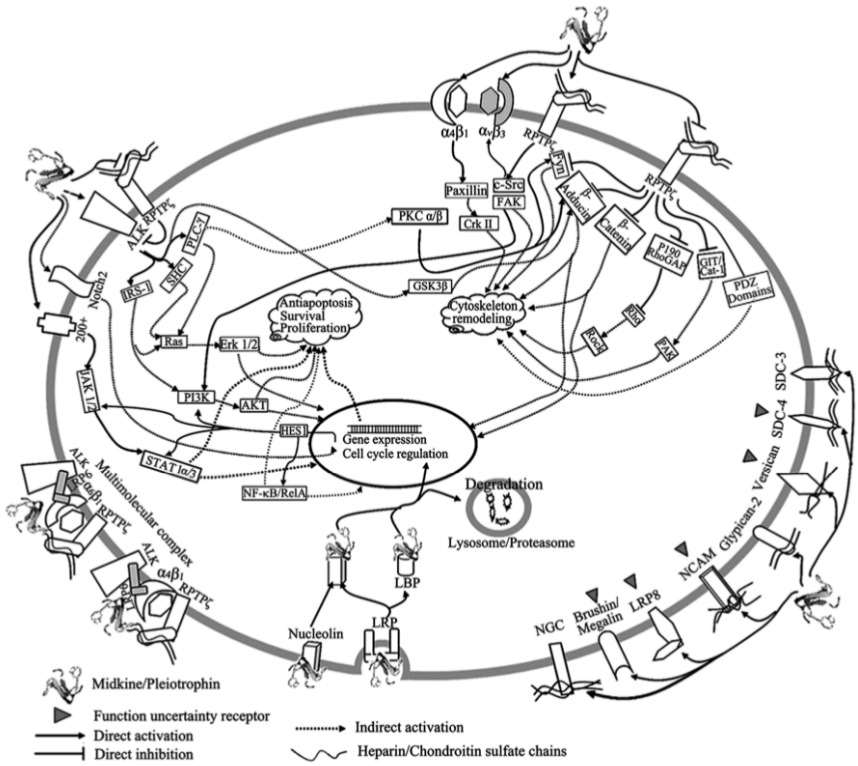


Figure 2.20. *Schematic representation of the PTPRZ1 signalling pathway showing the different dephosphorylated substrates and its mechanism of action. Figure taken from Biol. Pharm. Bull.* ***2014****,37(4), 511–520.*

The GIT1/Cat-1 is another of the PTPRZ1 substrates. When this enzyme is activated by phosphorylation it may further activate p21-activated serine threonine kinase (PAK) which plays an important role in the cytoskeleton turnover and remodeling, and finally on neuron migration [98]. The neuronal receptor tyrosine kinase (ALK) auto-activates itself through its own kinase activity, being PTPRZ1 its negative regulator. ALK plays an important role in the genesis and differentiation of the nervous system through the phosphorylation of other substrates such MAPK1/ERK2 and MAPK3/ERK1, CBL, FRS2, IRS1 and SHC1 [99]. In general, the PTN-PTPRZ1 pathway promotes on the CNS the regulation of cell growth and survival, cell adhesion, integrin-mediated signaling, cytoskeletal remodeling, cell motility, immune response and axon guidance. In addition, it may be involved in the differentiation of neuronal cells during the formation of neurite extensions and activity-dependent actin reorganization in dendritic spines.

2.4.2. Crystal structures and Inhibitors of PTPRZ1

The complete 3D structure of PTPRZ1 has not been fully elucidated yet. The first fragment crystallized and solved was the CA domain, deposited in the PDB under the code **3JXF** in 2009 (**Fig. 2.21**) [87]. A complex between contactin-1 and CA bond to CNTN1 was later solved (PDB code **3S97**) [100]. It wasn't until 2016, when the first intracellular catalytic domain was crystallized under PDB code **5AWX,** without any bound inhibitor and thus with the WPD-loop in the open conformation [101]. This structure was initially attempted to obtain in complex with inhibitor **SCB4380**, however, despite the IC_{50} of 0.4 μM, the bound conformation was not solved and in order to model the binding mode, docking techniques were applied (**Fig. 2.22**).

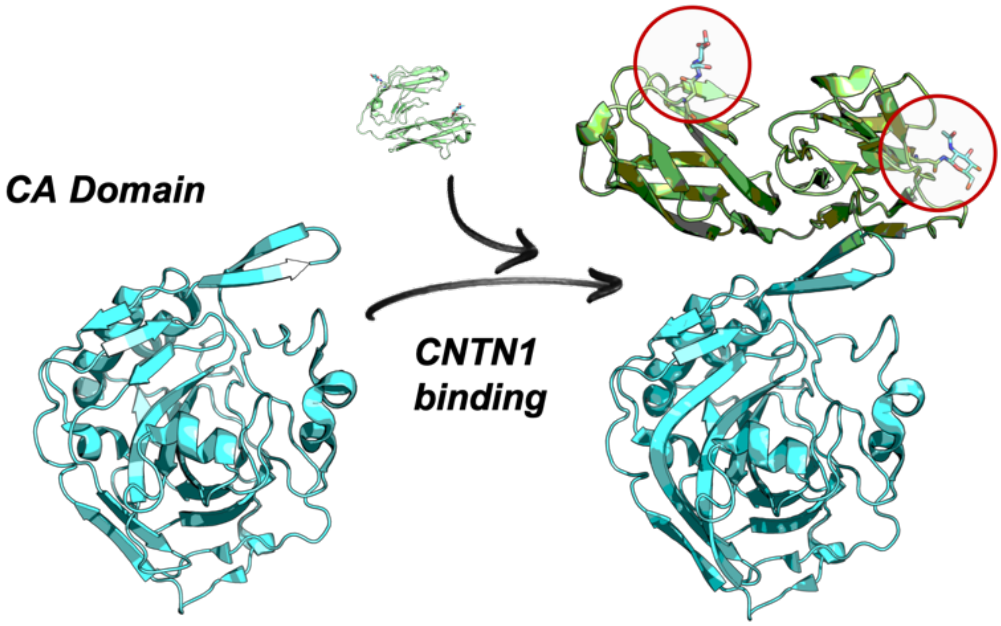


Figure 2.21. *Crystal structure of PTPRZ1 CA Domain (PDB code **3JXF**) and its complex with the CNTN1 domain (PDB code **3S97**). Notice the presence of two glucosamine molecules in CNTN1 highlighted in the figure.*

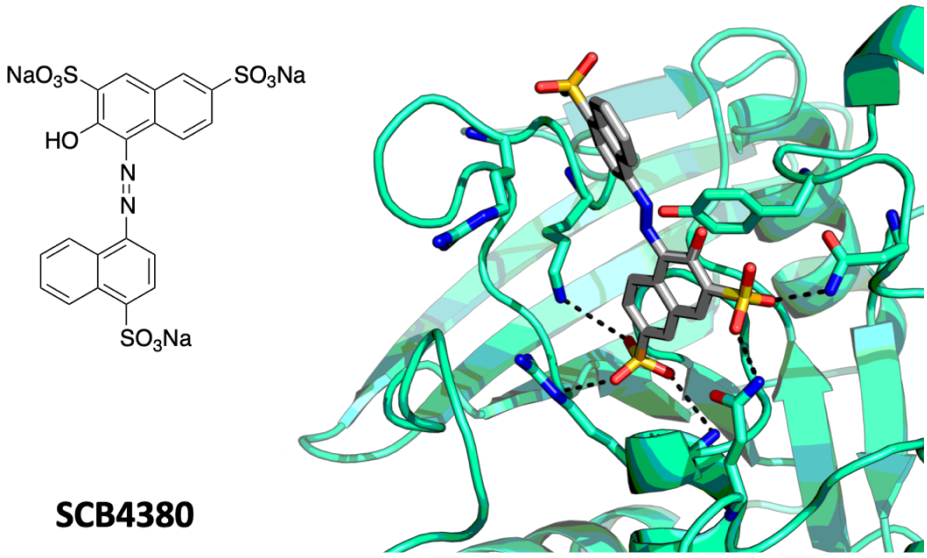


Figure 2.22. *Chemical structure of* ***SCB4380*** *(left) and bound complex obtained by modelling using the* ***5AWX*** *crystal structure.*

Nevertheless, a year later the same research group published the first crystal structure of PTPRZ1 in complex with a small molecule, **NAZ2329**, under the PDB code **5H08** [102]**.** In this structure the WPD-loop was displaced out of the active site, giving the so called super-open conformation. As mentioned before, **SCB4380** was published as a potent and selective PTPRZ1 inhibitor. As it failed to crystallize with PTPRZ1, the binding mode of this compound was modelled by docking techniques and stablished to bind into the active site through the sulfonate moiety that mimicks of the substrate's phosphate group. Furthermore, this small molecule has been the first PTPRZ1 inhibitor to be tested in an in vivo mice model against rat C6 glioblastoma by using liposomes as carrier due to its low ADME properties. Compound **NAZ2329** was designed with the aim to solve this ADME issue as a cell-permeable small molecule that allosterically inhibits both PTPRZ1 and PTPRG (IC_{50} of 7.5 and 4.8 μM respectively). As deduced from the crystal structure, **NAZ2339** penetrates between the α_4-helix and the WPD-loop, displacing Trp1899, and thus creating the superopen conformation (**Fig. 2.23**). The binding mode of this molecule is based mainly on the electrostatic interaction between the **NAZ2339** sulfonamide group and Arg1939 plus the van der Waals interaction of the

trifluoro-methyl, methoxy-benzyl moiety inside the hydrophobic pocket. This molecule has shown allosteric properties because it does not touch the PTPRZ1 active site, it displaces the WPD-loop where the crucial Asp1901 is located. This leads to the loss of phosphatase activity because Asp1901 will be too far away from the catalytic site and will not be able to recycle the phosphorylated catalytic cysteine. A complete summary of PTPRZ1 crystal structures deposited is in **Table 2.2**.

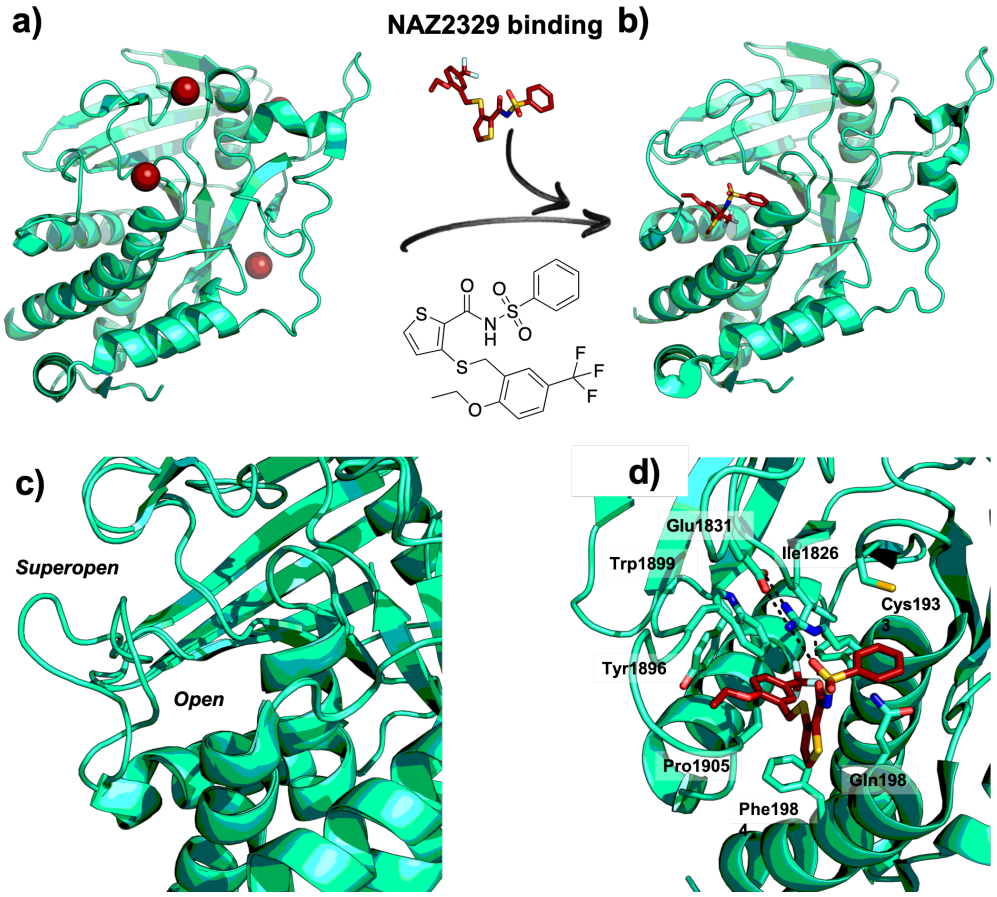
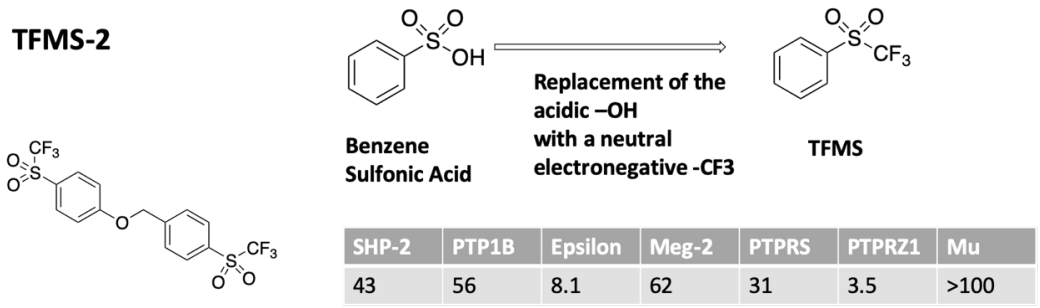


Figure 2.23. a) *General structure of PTPRZ1 PD1 (PDB code **5AWX**), red spheres represent bromine atoms.* ***b)*** *PTPRZ1 in superopen conformation after **NAZ2329** binding (PDB code **5H08**).* ***c)*** *Structural differences between the superopen conformation and the open conformation of the two superimposed PTPRZ1 PD1.* ***d)*** *Binding mode of **NAZ2329** inside PTPRZ1 where the main interactions are highlighted.*

During the efforts to identify new chemical templates for PTP1B inhibitors, specifically phosphate bioisosters such as sulfonic acid and trifluoro-methylsulfonyl (TFMS) derivatives, the first PTPRZ1 inhibitor was reported

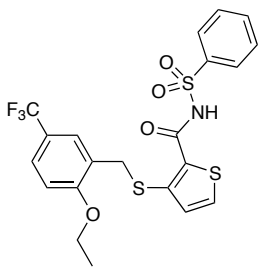
(**TFMS-2)** [103]. This inhibitor presents an interesting activity and selectivity profile towards PTPRZ1 despite being designed as a PTP1B inhibitor (**Fig. 2.24**) [103]. The general structure of the series of inhibitors presented two TFMS moieties aiming to occupy the two phosphate-binding sites of PTP1B (see section **2.6.1**).



SHP-2	PTP1B	Epsilon	Meg-2	PTPRS	PTPRZ1	Mu
43	56	8.1	62	31	3.5	>100

Figure 2.24. *Chemical structure of* ***TFMS-2*** *and design rationale. Table shows the* IC_{50} *values in different PTPs.*

Table 2.2. *PDB codes related to UniProtKB entry P23471 (PTPRZ_HUMAN).*

PDB Code	Domain	Conformation	Ligand	IC_{50} µM
3JXF	CA	-	-	-
3S97	CA-CNTN1	-	-	-
5AWX	PD1	Open	-	-
5H08	PD1	Superopen	7WL – NAZ2329 	7.5
6J6U	PD1 +PD2	Open	-	-

2.4.3. PTPRZ1 related diseases

The clinical relevance of PTPRZ1 has been highlighted through its implication in several types of diseases and physiopathological processes [104]. The main role of this receptor is connected to neuronal survival, remyelination processes, and promotion of neuronal inflammation. Therefore, the disruption of the normal behavior could be of interest in diseases such as tumor processes, neurological disorders, or drug addiction [102,105–107]. The first time that PTPRZ1 was linked to a human disease was in 1994, when in a pulmonary adenocarcinoma cell line A427, PCR experiments described PTPRZ1 as a tumor suppressor factor due to its low level's expression [108]. Later on, other types of tumors, such as colorectal cancers, were also linked with this low PTPRZ1 expression [109]. However, PTPRZ1 cannot always be linked to tumor suppression, as it has been found to be activated or overexpressed in several types of cancerous processes such as: primary human glioblastomas [110,111], glioma cell migration [112], human cutaneous melanomas [113,114], gastric cancer [115], anaplastic lymphoma [99], breast cancer [116,117], cervical carcinoma and adenocarcinoma [118], small cell lung cancer [119], and renal cell carcinoma (RCC) [120].

Additionally, PTPRZ1 plays a negative role in oligodendrocyte and remyelination in demyelinating CNS diseases through the dephosphorylation of p190RhoGAP [121]. In an induced CNS injury, it has been observed that an induction of PTPRZ1 with high mRNA expression levels in sprouting axonal areas may contribute to injury recovery and axonal regrowth. These results suggest that PTPRZ1 is involved in recovery and survival from demyelinating diseases [122]. Furthermore, PTPRZ1 has been proposed as a necessary factor for maintaining the integrity in the CNS during demyelinating processes, becoming an important target of axonal protection in CNS inflammatory diseases such as multiple sclerosis, periventricular leukomalacia, or Parkinson's disease [123]. Furthermore, some studies have shown that after neuronal injury, PTN, the endogenous

PTPRZ1 ligand, is released. This inactivates PTPRZ1 and promotes oligodendrocyte differentiation for the remyelination process [124]. For all of these reasons, PTPRZ1 has been proposed as novel target for the treatment of devastating diseases such as Alzheimer or multiple sclerosis, based on delayed recovery from demyelinating injuries [125]. In Parkinson's disease increased levels of PTN in the substantia nigra have been observed in patients, suggesting that PTN could exhibit protective effects. Moreover, mice models deficient in PTPRZ1 have proven to increase the susceptibility to experimental autoimmune encephalomyelitis (EAE), which further supports previous hypothesis [122,126].

Related to mental diseases, it is interesting to mention that PTPRZ1 has been also linked to schizophrenia. A meta study carried out in the United Kingdom with more than 1400 patients shows a genetic association between the PTPRZ1 expression and the incidence in schizophrenia. The role of PTPRZ1 in the modulation of neuregulin (NRG1)/ERBB4 signaling as possible action mechanism [127]. This hypothesis was also deduced by mutated mice with PTPRZ1 overexpression (PTPRZ1-tag mice), which enhanced the pathogenesis of schizophrenia [128]. But nevertheless, another meta study carried out in Japan with 576 schizophrenic patients and 768 controls, rendered no clear evidence of the relationship between the development of schizophrenia and PTPRZ1, at least in the Japanese population [129].

Lastly, PTPRZ1 has been related to *Helicobacter pylori* colonization through the toxin VacA, which possesses the capability to bind to the cell surface by the direct interaction with this receptor [130]. In BHK-21 cells, the overexpression of PTPRZ1 conferred higher VacA sensitivity than normal which led to higher *H. pylori* colonization [131]. Another study showed how the fetal calf serum (FCS) inhibited the *H. pylori* vacuolization through the inhibition of the ppi between PTPRZ1 and VacA [132]. The essential extra cellular domain from PTPRZ1 which interacts with VacA was detected by immunoprecipitation methods, sequence QTTQP at positions 747–751, confirmed by genetic designed mutants T748A and T749A that had

diminished vacuolation [133]. All these studies point to the inhibition of the direct ppi between PTPRZ1 and VacA as a novel therapeutic strategy in the development of new antimicrobial agents.

PTPRZ1 and one of its main ligands, pleiotrophin (PTN), are overexpressed in multiple diseases making them interesting and potential novel demonstrated targets for the development of new immunotherapeutic treatments [134]. Some of the small molecules developed by *P. Huang*, *A. Fujikawa*, have their potential therapeutic effect by inhibiting PTPRZ1. Not only small molecules have been developed, also monoclonal antibodies targeting PTPRZ1 were studied showing its efficiency in delaying tumor growth in a glioblastoma model [135]. Also, glioblastomas have also been challenged by interference RNA with promising results in both in vitro and in vivo models [136].

2.5. PTPRG

The Receptor-type tyrosine-protein phosphatase gamma (PTPRG or PTPRγ UniProtKB - P23470) is also a transmembrane phosphatase member of the CC1 superfamily, Cysteine-based Class I (**Table 2.3.**). PTPRG presents the same extracellular (aa 20-736), transmembrane (aa 737-762) and intracellular domains (aa 763-1445) as PTPRZ1. In fact, the total sequence similarity is around 73%, so the same action mechanism and structure is shared with PTPRZ1. PTPRG interacts mainly with CNTN3, CNTN4, CNTN5 and CNTN6, whereas PTPRZ1 binds only to CNTN1 [87]. In this line, PTN is also a putative ligand for PTPRG as it shares common structural domains with PTPRZ1 [137]. In contrast to PTPRZ1, not so many phosphorylated substrates have been reported for PTPRG apart from Fibroblast Growth Factor Receptor 1 (FGFR) [138]. But it seems that both receptors could share more common phosphorylated substrates accordingly to their structures.

PTPRG is widely expressed in the CNS (in neocortex, hippocampus, striatum and cerebellum), especially during the grain development and in

adulthood, but in contrast to PTPRZ1, it is also expressed in many other tissues [139]. The ppi between PTPRG CA domain and different CNTNs has been stablished through crystallography (**5E5R**), which has allowed to deduce that this interaction does not induce dimerization of either PTPRZ or PTPRG with CNTN1 and CNTN4, respectively [87]. It has been proposed that the interaction with CNTNs specifies the location of dephosphorylation reactions instead of blocking the phosphatase activity directly. A plausible explanation is that the role for CNTNs would be to guide both PTPRG and PTPRZ1 toward specific cellular regions, controlling the dephosphorylation process in specific locations. PTPRG only presents two known isoforms, and the phosphorylated substrates are not well stablished above.

2.5.1. Crystal structures and Inhibitors of PTPRG

Despite the apparently lower interest of PTPRG as a druggable target, when compared to PTPRZ1, many more structures have been crystallized to date, and especially with a bond ligand. The CA domain has been crystallized twice and deposited under the PDB codes **3JXH** (only the CA domain) and **5E5R** (CA domain plus the CNTN4) (**Fig. 2.25**) [87]. The intracellular PD1 domain has been crystallized in three WPD-loop conformations (closed, open and superopen) and also together with the PD2 (**Fig. 2.26**). All the structures deposited are summarized in **Table 2.3**. The structures that contain the PD1 domain in the open conformation are PDB codes **2H4V**, **3QCB** (in complex with sulfate), **2HY3** and **3QCC-chain A** (in complex with vanadate). The closed conformation is deposited under PDB codes **3QCC-chain B** and **3QCD**, all in complex with vanadate. PDB code **2NLK** contains both domains PD1 and PD2, which has allowed to structurally solve the head to toe mechanism, as it can be deduced form the study of the neighboring proteins in the asymmetric unit (**Fig. 2.26)**. The head to toe blocking mechanism can be explained by the interaction between the Asp1305-Asp1306-Tyr1307 sequence present in PD2 that interacts directly with the amino acids Lys958 and Arg960 from the P-loop, and Tyr880 from the pTyr-loop (**Fig. 2.27**). These interactions are so relevant that a double

mutation in Asp1305Lys and Asp1306Lys produces the loss of auto inhibitory activity in the receptor. In a similar manner, variations in the E-loop such as Lys958Asp and Arg960Asp mutants also lead to a loss of auto inhibitory activity because this type of acid-basic changes provoke electrostatic repulsive interactions that disrupts the ppi (**Fig. 2.27**) [92].

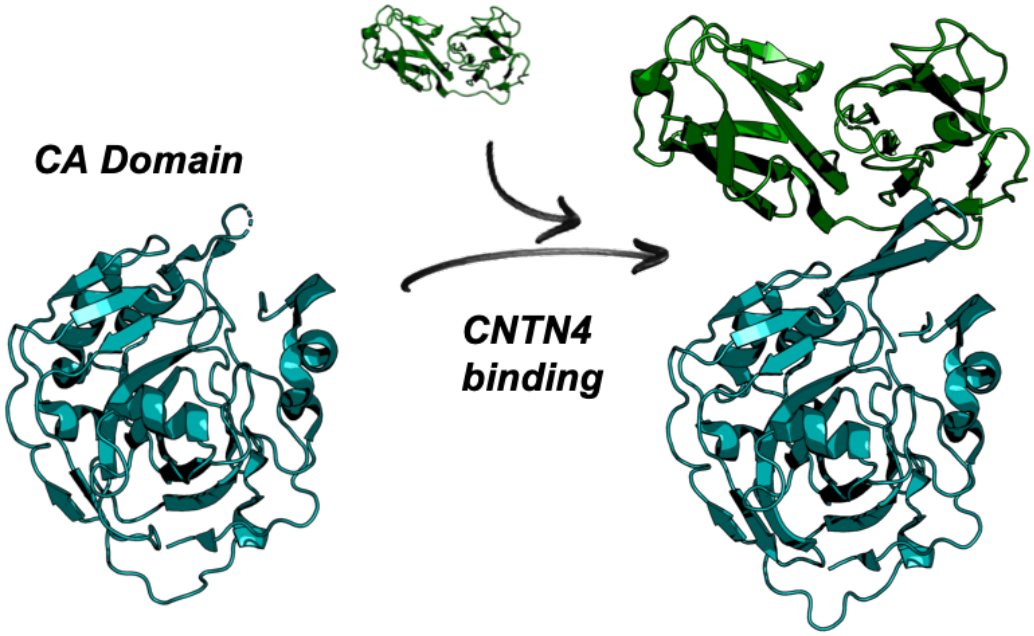


Figure 2.25. *Crystal structure of PTPRG CA Domain (PDB code **3JXH**) and its complex with the CNTN1 domain (PDB code **5E5R**).*

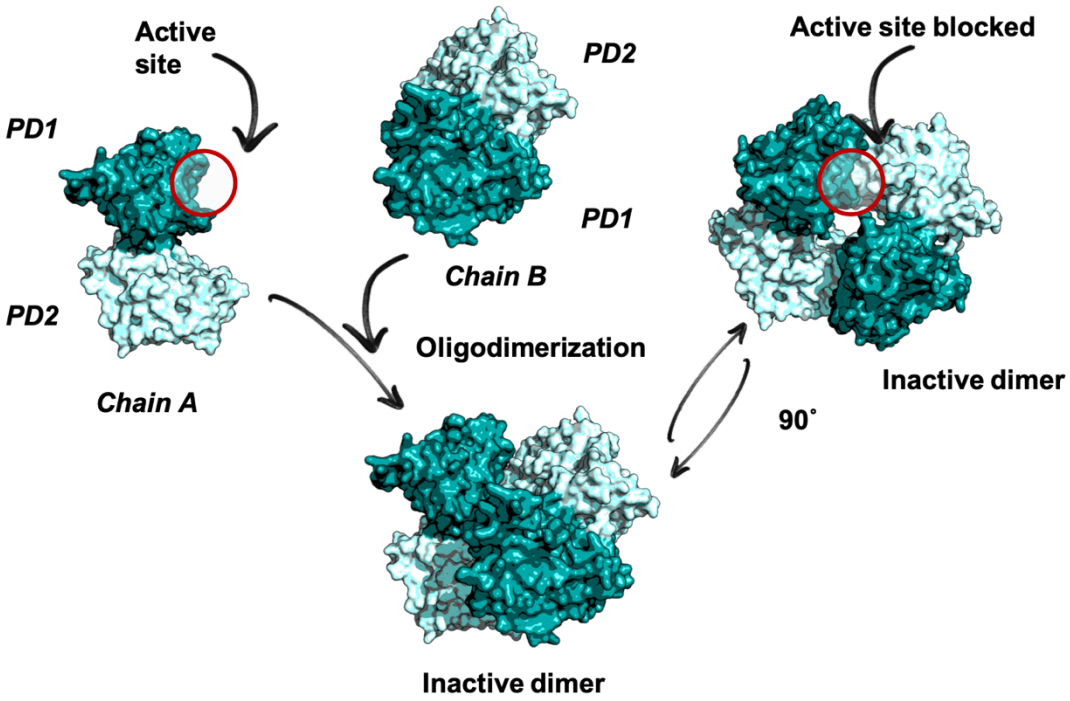


Figure 2.26. *The head to toe mechanism of dimerization in PTPRG intracellular domains.*

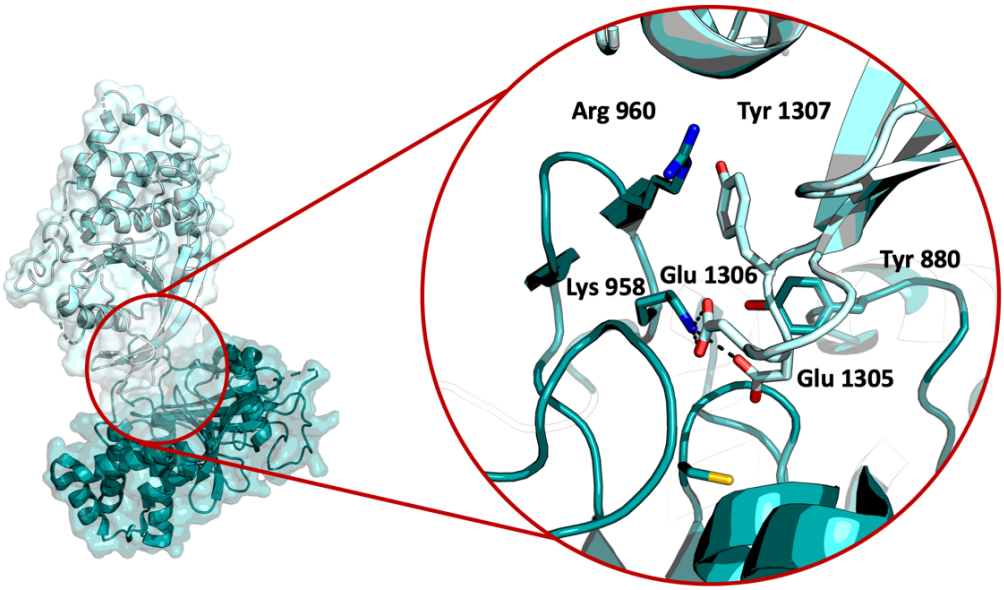


Figure 2.27. *Molecular mechanism of the oligodimerization chain A of PD1 and chain B of PD2. The specific amino acids that interact in the ppi with the active site are highlighted.*

Additionally, there are 13 PTPRG structures complexed with small molecules in total that have allowed identifying for the first time the WPD-loop in the "superopen-conformation" (**Table 2.3**). These inhibitors were

published in 2011 by *Sheriff et al*. [140]. All of these ligands present a benzyl-halide moiety (**NXY** and derivatives) that induces this superopen conformation by displacing the side chain of Trp1026 of the WPD-loop, thus forcing the opening of a new cryptic pocket. Compound **NXY** was obtained in 2011 through a sophisticated HTS of 26235 compounds. **NXY** was the only one that presented a moderate PTPRG activity (IC_{50} = 5 μM) and selectivity against other common phosphatases (PTP1B IC_{50} = 70 μM and CD45 IC_{50} = > 120 μM) [140]. From this first PTPRG hit, *Sheriff's group* carried out an intensive SAR in order to increase the potency, obtaining **NX4** with an IC_{50} = 0.5 μM, PDB code **3QCJ** and **3QCE**. (**Fig. 2.28**). The binding mode of this series is similar to one previously mentioned for compound **NAZ2339**, except that the atoms that go deepest inside the cryptic pocket are halogens in all cases. Derivatives with no halogen atoms decrease 20 times the activity as shown in the **Table 2.3.** These molecules also present a thiophen and thioether group, which link the benzyl-halide moiety to the rest of the molecule. Interestingly, changing the thioether linker by an ether, amine or sulfonate decreases considerably the activity (see section **Chapter V**). The first series with a terminal carboxylic acid was generated by adding a sulfonamide moiety and methyl or benzyl moieties to the thiophene, leading to **NX4** that presents the highest activity. This molecule is able to interact with the active site by a hydrogen bridge between the charged carboxylic acid and the catalytic water, as expected the esterification of the carboxylic acid reduces the inhibitory activity. This moiety also interacts through a hydrogen bond established with the methoxy substituent with Gln1104 present in the Q-loop. Additionally, compounds **NXV** and **NXW** were designed on the opposite direction, instead of reaching the active site of the PTPG, they were grown towards the α3 and α6 helixes with different alkyne derivatives (**Fig. 2.29**).

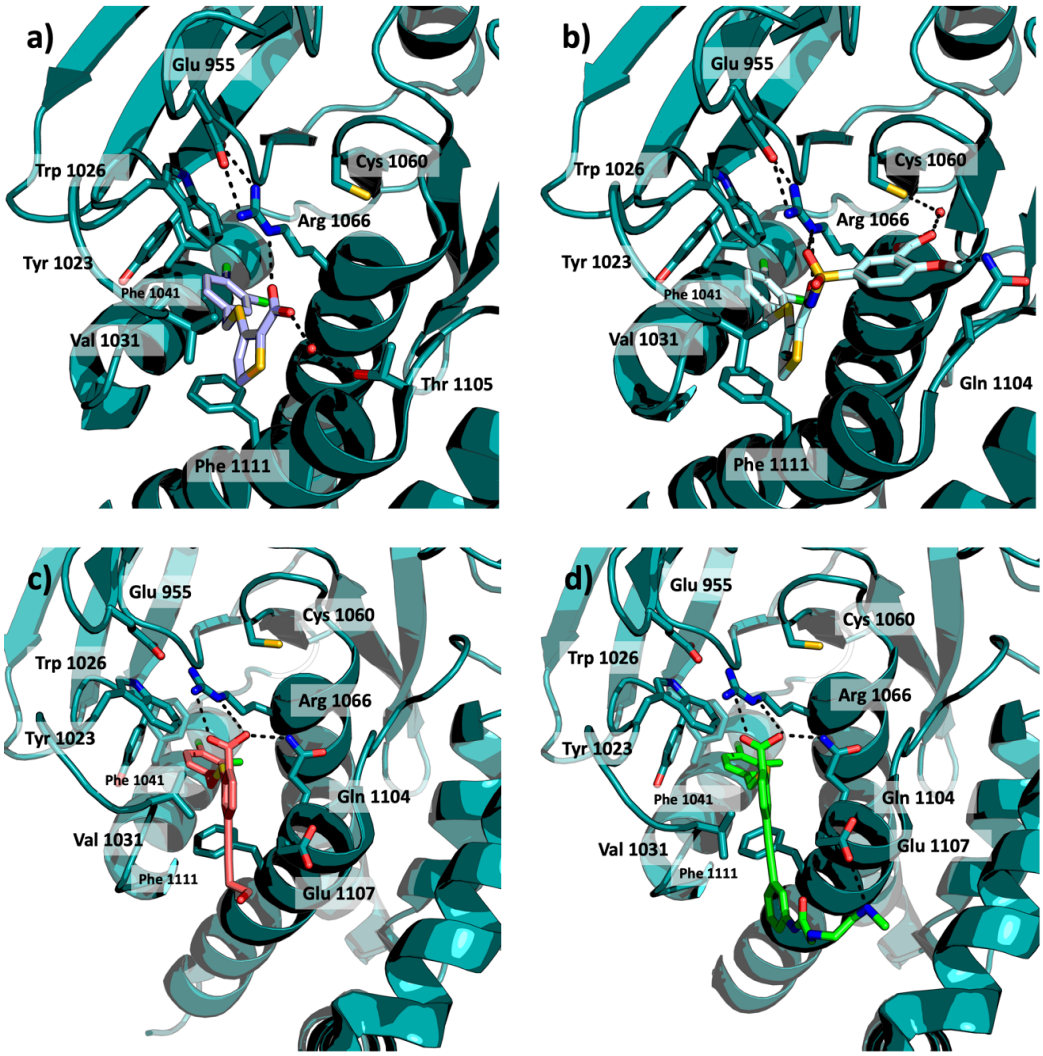
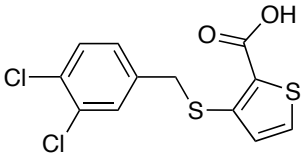
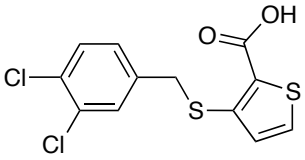
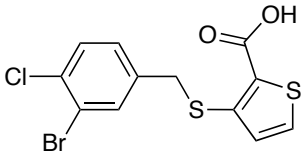
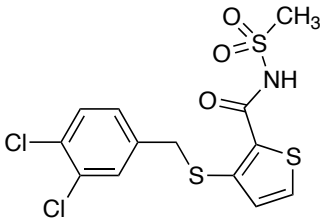
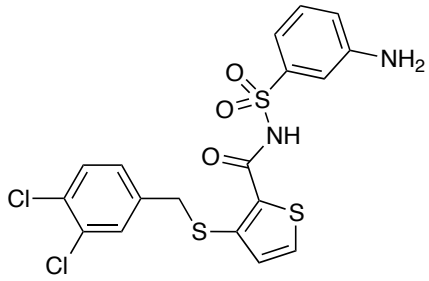
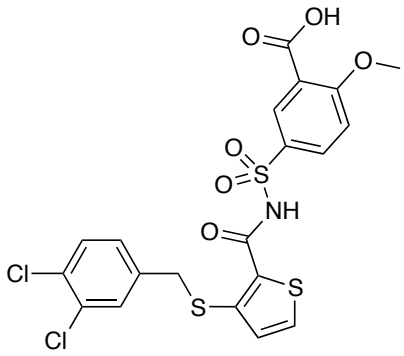
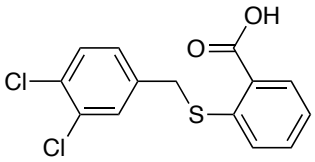
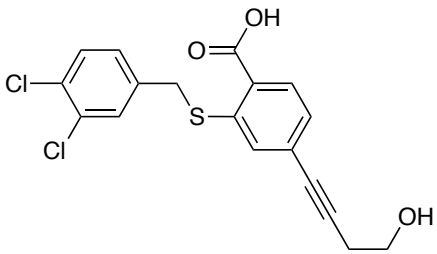


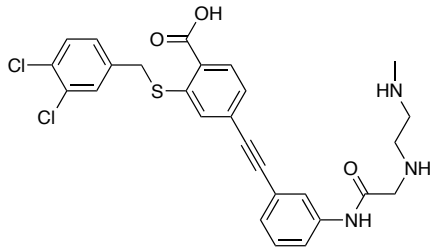
Figure 2.29. *PTPRG complexes structures with **NXY a)**, **NX4 b)**, **NXV c)**, and **NXW d)** with PTPRG (**3QCE**, **3QCJ**, **3QCL** and **3QCM**). Main interactions are highlighted.*

Table 2.3. *PDB codes related to UniProtKB entry P23470 (PTPRG_HUMAN).*

PDB Code	Domain	Conformation	Ligand	IC_{50} (μM)
2H4V	PD1	Open	-	-
2HY3	PD1	Open	Vanadate	-
2NLK	PD1 + PD2	Open	-	-
2PBN	PD1	Open	Sulfate	-

3JXH	CA-CNTN4	-	-	-
3QCB	PD1	Superopen	Sulfate	-
3QCC	PD1	Open/Closed	Vanadate	-
3QCD	PD1	Closed	Vanadate	-
3QCE	PD1	Superopen/Open	 NXY	5
3QCF	PD1	Superopen/Open	 NXY	5
3QCG	PD1	Superopen/Open	 NX1	-
3QCH	PD1	Superopen	NX2 	3.3
3QCI	PD1	Superopen	NX3	3.8

				
3QCJ	PD1	Superopen	NX4 	0.5
3QCK	PD1	Superopen	 NX5	5.8
3QCL	PD1	Superopen	NXV 	4.0
3QCM	PD1	Superopen	NXW	-

				
3QCN	PD1	Open	-	-
5E5R	CA - CNTN4	-	-	-

2.5.2. PTPRG related diseases

PTPRG has been implicated in different types of diseases, especially in specific mutated cancer types, and has also been identified as a tumor suppressor. However, in contrast to PTPRZ1, the majority of the diseases related to PTPRG come mainly from the lack of activity or depletion of this phosphatase. The first time that PTPRG gene was characterized as a tumor suppressor candidate, was in 1991 when the PTPRG allele was discovered to be lost in 3 of 5 renal carcinoma cell lines and 5 of 10 lung carcinoma tumor samples by mRNA analysis [141]. Since then, PTPRG has been considered to be a tumour suppressor in multiple cancer cell lines, in vivo models and samples, using different types of techniques such as PCR, direct protein quantification or DNA methylation. Here, we only summarize some of the PTPRG activities where it has identified as a tumor suppressor. Hypermethylation analysis in primary cutaneous T-cell lymphoma (CTCL), show how PTPRG and other tumor suppressor genes (BCL7a and thrombospondin) were significantly hypermethylated resulting in its under expression [142]. In myeloid leukemia patients, PTPRG is also downregulated, especially in chronic myeloid leukemia (CML) patients [143]. Also, RAS mutations promote the PTPRG methylation gene in a complementary event in oncogenesis inducing acute lymphoblastic leukemia [144]. PTPRG methylation has been proposed as a potential

biomarker for early detection of colon cancer [145]. Following hypermethylation genes, in CTCL this process inactivates several types of tumors suppressing routes those involved in DNA repair, cell cycle, and apoptosis signaling pathways, all of which include PTPRG, highlighting the relevance of this enzyme as tumor suppressor [142].

In human breast cancer, PTPRG expression promotes tumorigenesis by increasing proliferation, stimulating cell migration and reducing apoptosis [146,147]. The proposed mechanism is that PTPRG may up-regulate p21(cip) and p27(kip) proteins through the ERK1/2 pathway, and consequently exhibit the characteristics of a breast tumor suppressor, at least in MCF-7 cells lines [148]. In this sense, PTPRG has also been detected to be under expressed or dramatically depleted in other types of tumors such as renal cell carcinoma, nasopharyngeal carcinoma and lung cancer [141,149,150].

2.6. PTP1B

PTP1B or PTPN1 is a non-receptor PTP expressed in several tissues and one of the most relevant phosphatases has it has been validated as a target for diabetes and obesity. Consequently, it was the first PTP to be isolated, purified and studied. These facts make PTP1B a subject of interest for both the pharmaceutical industry and the academic institutions as it is an attractive and desirable drug target for therapeutic intervention.

PTP1B belongs to the non-receptor class 1 PTP subfamily (UniProtKB - P18031). Its catalytic domain presents 275 amino acids (aa 3-277) and it is considered as a structural standard because, as mentioned before, it was the first PTP to be crystallized [23]. Therefore, the general structure described before can also be applied in this enzyme with the notorious exception that PTP1B presents an extra C-terminal α7-helix (aa 285-295), which has been related with autoinhibition processes and allosteric inhibition. Moreover, two proline-rich motifs and an Endoplasmic Reticulum (ER)-targeting domain, both not crystallized yet, play an important

regulatory role. The proline-rich motifs (aa 300–308 and 309–314) mediate the binding of SH3 domain-containing proteins present in p130, Grb2, Crk and Src. These proteins establish a complex with PTP1B that leads to different type of signals such as cell migration [151]. The ER-targeting domain locates PTP1B on the cytoplasmic face of the ER membrane. The C-terminal domain constituted by 35 aa anchor the PTP to the ER so that it only dephosphorylates intracellular substrates [152]. However, the cysteine protease calpain is able to cleave this motif thus releasing PTP1B from this cellular niche into the whole cytoplasmic space with enhanced catalytic activity [153]. Typical post-translational modifications are phosphorylation at Ser50 responsible for its enzymatic activation, and also Ser242 and Ser243. This process is carried out by the CLK family (Dual specificity protein kinase, CLK1 and CLK2) [154], and the putative Akt (Protein Kinase B), involved in the signal transduction PI3K/AKT/mTOR pathway [155]. Also, a reduction–oxidation reaction in the catalytic Cys215 of PTP1B inhibits substrate binding and the catalytic activity, which can be restored by reduction. Henceforth, this has been proposed as a regulatory mechanism of the intrinsic activity of PTP1B. This regulatory mechanism has been studied by both experimental and theoretical techniques, through the reversible oxidization of the catalytic cysteine to sulphenic acid (Cys-SOH). This latter could react with the alpha-amino group from the following Ser216 forming a sulfonamide cross-link, triggering a conformational change in the P-loop and blocking the activity. (**Fig. 2.30**) [156,157].

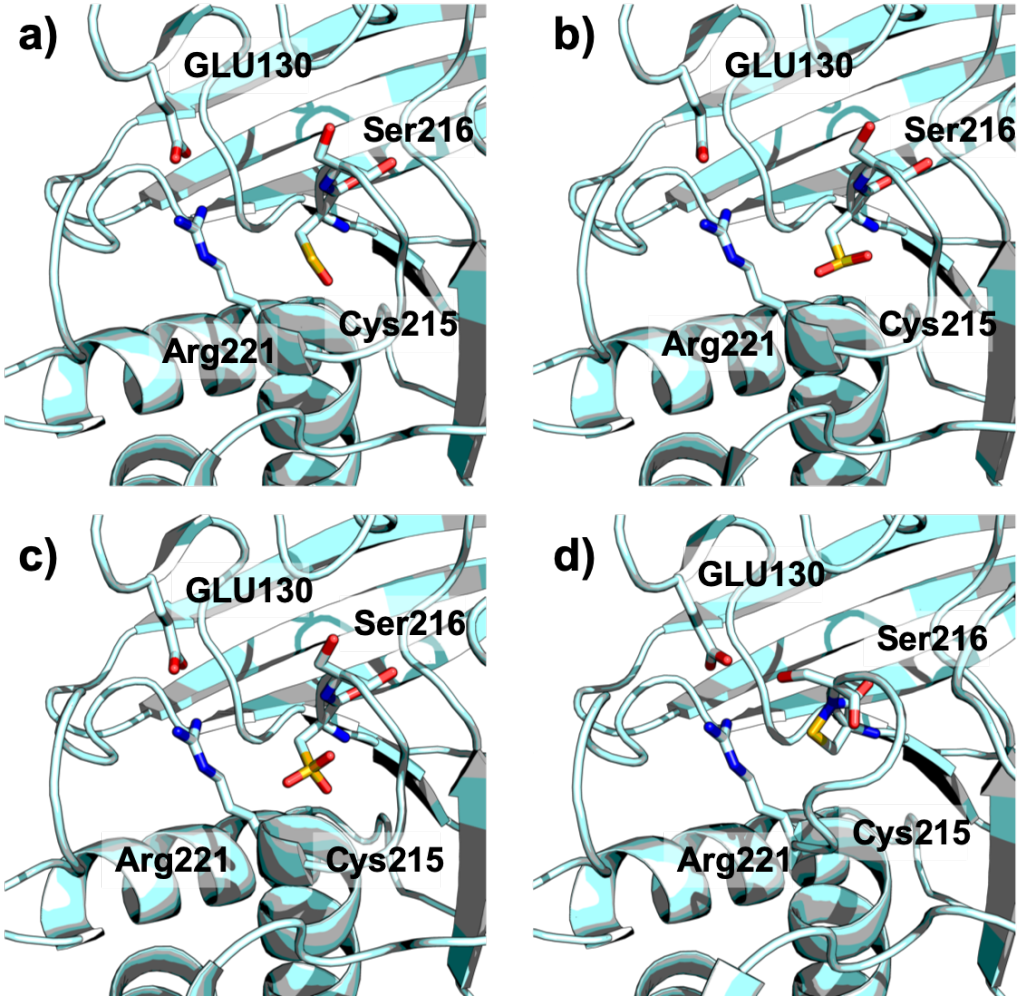


Figure 2.30. *Different PTP1B oxidized states* ***a)*** *sulfenic acid* ***1OET****,* ***b)*** *sulfinic acid* ***1OEU****,* ***c)*** *sulfonic acid* ***1OEV****,* ***d)*** *cyclic sulfonamide* ***1OES****. Notice the overall disrupting of the P-loop in this last structure.*

This enzyme plays a wide diversity of cellular regulatory activities with more than 30 known substrates and 13 interaction partners, but nevertheless more substrates remain yet to be discovered [158]. At first, PTP1B was well known to be involved in the Insulin Receptor (IR) mechanism. Indeed, this is one of the hot spots related to the development of PTP1B inhibitors. The first time that PTP1B was identified to be related with glucose intake was in 1994 where desensitization of insulin receptor function by a high glucose condition was associated with the activation of PTP1B [159]. Since then, several articles have been published with the aim of shedding light on this mechanism and regulatory processes. IR is a kinase transmembrane receptor that dimerizes after insulin binding, and self phosphorylates. This

promotes the recognition by the IR Substrate (IRS) proteins that leads to activation of phosphatidylinositol (PI) 3-kinase and AKT. The PI3K activation into the membrane increases the inositol-3-phosphate levels, which leads to the activation of different cellular pathways such as PKB/Akt, mTOR and PKC-alpha, that finally increase the number of glucose receptors on the cell membrane (**Fig. 2.31**) [160–162].

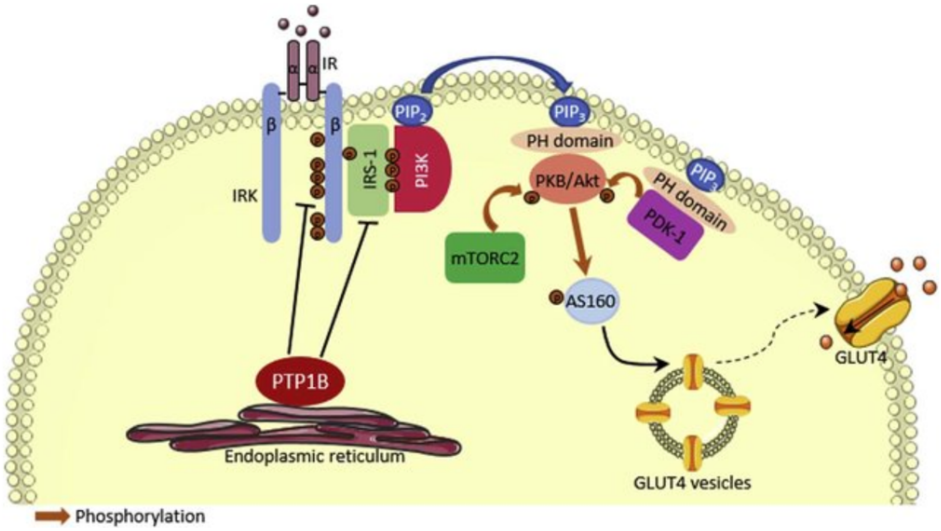


Figure 2.31. *Signaling pathway by PTP1B promotes the negative regulation of insulin by dephosphorylation of both Insulin Receptor Substrates 1 (IRS-1) and Insulin Receptor Kinase (IRK). Insulin binds to the extracellular α-subunit of its receptor and induces the Insulin Receptor (IR) activation by autophosphorylation. PI3K also activates increasing the levels of PIP3 inducing the activation of other enzymes such as PKB/Akt and PDK-1, mTORC2 and PKB which finally facilitates the translocation of GLUT4 to the membrane. Figure taken from Food Chem Toxicol.* ***2018****,111, 474-481.*

PTP1B has been shown to antagonize this signaling pathway by dephosphorylating different substrates and thus increasing the insulin resistance. Moreover, PTP1B decreases the effect of leptin by dephosphorylating the downstream effect caused by the Janus kinase 2 (JAK2) [163]. As well as insulin, leptin is a key hormone that plays an important role related with energy balance and appetite/hunger [164]. It has been proven that PTP1B$^{-/-}$ knock-out mice have decreased leptin signaling and body fat ratios, are less responsive to leptin in hypothalamic neurons, and are able to overcome resistance to obesity [165].

2.6.1. Crystal structures and inhibitors of PTP1B

As mentioned before, the number of PTP1B crystal structures and molecules designed to modify the activity of PTP1B is huge. Here we only remark some of the historically relevant and recent advances on PTP1B inhibitors, and recently published crystal structures. One of the common ways by which a competitive inhibitor of a protein target is designed is by constructing molecules that are either substrate or transition state analogues, or product mimics. Despite this being the typical initial drug design approach, for PTPs this is not the most appropriate technique due to the active site similarity amongst all PTPs. Therefore, novel strategies for PTP1B drug design have been developed in the last 10 years. A phosphorylated Tyrosine (P-Tyr) is the natural substrate of PTPs, so the initial approach was to design substrate mimics and explore surrounding areas in order to find adjacent pockets which could be useful to increase the affinity. In 1997, using a PTP1B inactive mutant (C215S) and soaking methods, a second aryl phosphate binding site was discovered using bis-(para-phosphophenyl) methane (BPPM) (**1AAX**) and pTyr (**1PTY**) (**Fig. 2.32**) [30]. This novel aryl phosphate-binding site was named “site-B” and the most important residues implicated in the binding mode are Arg24 and Arg254, which interact with the phosphate. Additionally, favorable interactions were determined, such as the water-mediated hydrogen bonds with Met258 and Gln262 from the Q-loop and van der Waals contacts with the side chains of Ile219, Asp48, and Val49.

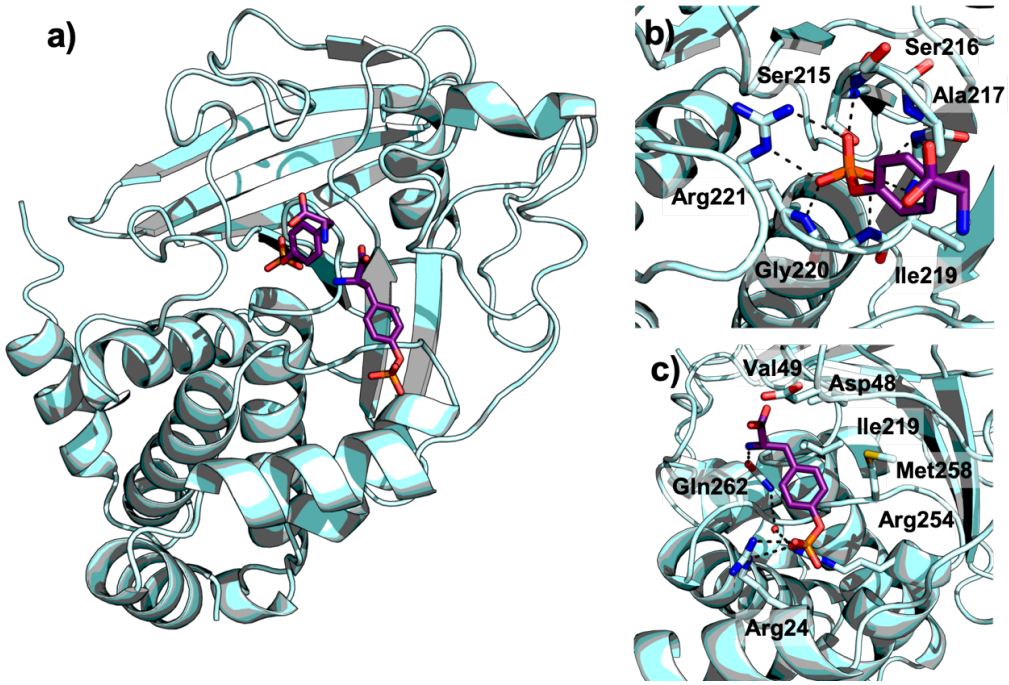


Figure 2.32. a) *General PTP1B structure with P-Tyr residues occupying catalytic site and site B (**1PTY**) where the catalytic cysteine is mutated to serine.* ***b)*** *Detailed interactions in the active site.* ***c)*** *Detailed interactions in the site B.*

One of the first phospho tyrosine mimicks was phosphonomethylphenyl alanine (Pmp) and phosphonodifuromethylphenyl alanine (F2Pmp), which provided an excellent non-hydrolysable phosphate derivative and displayed enhanced binding to the PTP1B active site. These developed molecules presented an IC_{50} close to 0.1 μM. (**Fig. 2.33**) [166].

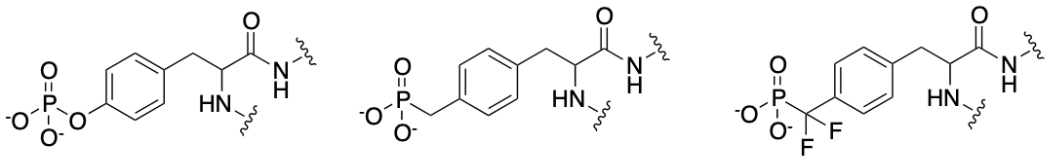
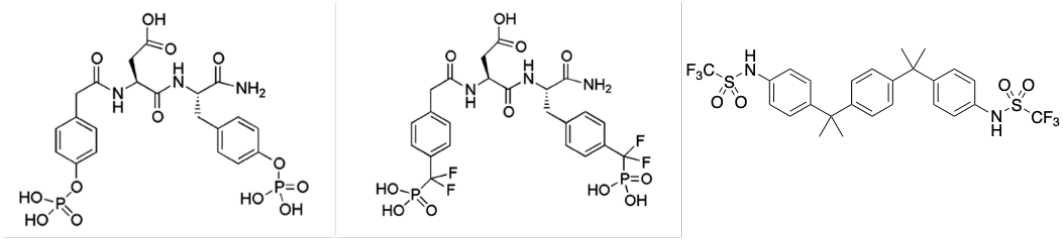


Figure 2.33. *Initially developed PTP1B inhibitors. The rest of the peptide has been removed for clarity.*

In order to increase the activity, both aryl phosphate-binding sites were explored with bivalent ligands using different phosphomimics, obtaining different series of PTP1B inhibitors. *Zhang et al.* reported the first phosphate derivative (**Zhang 1**) that reached both sites; and then, by adding a non-hydrolyzable group they obtained **Zhang 2** with a K_i 2.5 nM and higher

specificity (**Fig. 2.34**). Other groups developed **Zhang 7** by using trifluoromethyl sulfonyl (TFMS) benzene derivatives. Although it presented an IC_{50} of 2.5 μM, **Zhang 7** showed low selectivity against other phosphatases [103].



Zhang 1 Zhang 2 Zhang 7

Figure 2.34. *Developed Bidentate ligands for PTP1B reported by Zhang et al.* [103].

In 2003, Abbott laboratories published a novel inhibitor based on the site-directed oxalylarylaminobenzoic acid-based pharmacophore that led to a series of potent PTP1B inhibitors with 6-fold selectivity over the highly homologous T-cell PTPase (TCPTP), and high selectivity over other phosphatases. The complex with the analogue **ligand 418** was crystallized in complex with PTP1B under PDB code **1PH0** (**Fig. 2.35**). In 2005, another approach was applied by *Black et al.* using 1,2,5-thiadiazolidin-3-one1,1-dioxide (**Black 1**) as phospho mimicks, thus obtaining **Black 10** with a IC_{50} = 2.47 μM that was crystallized in complex with PTP1B under PDB code **2BGD** (**Fig. 2.36**) [167].

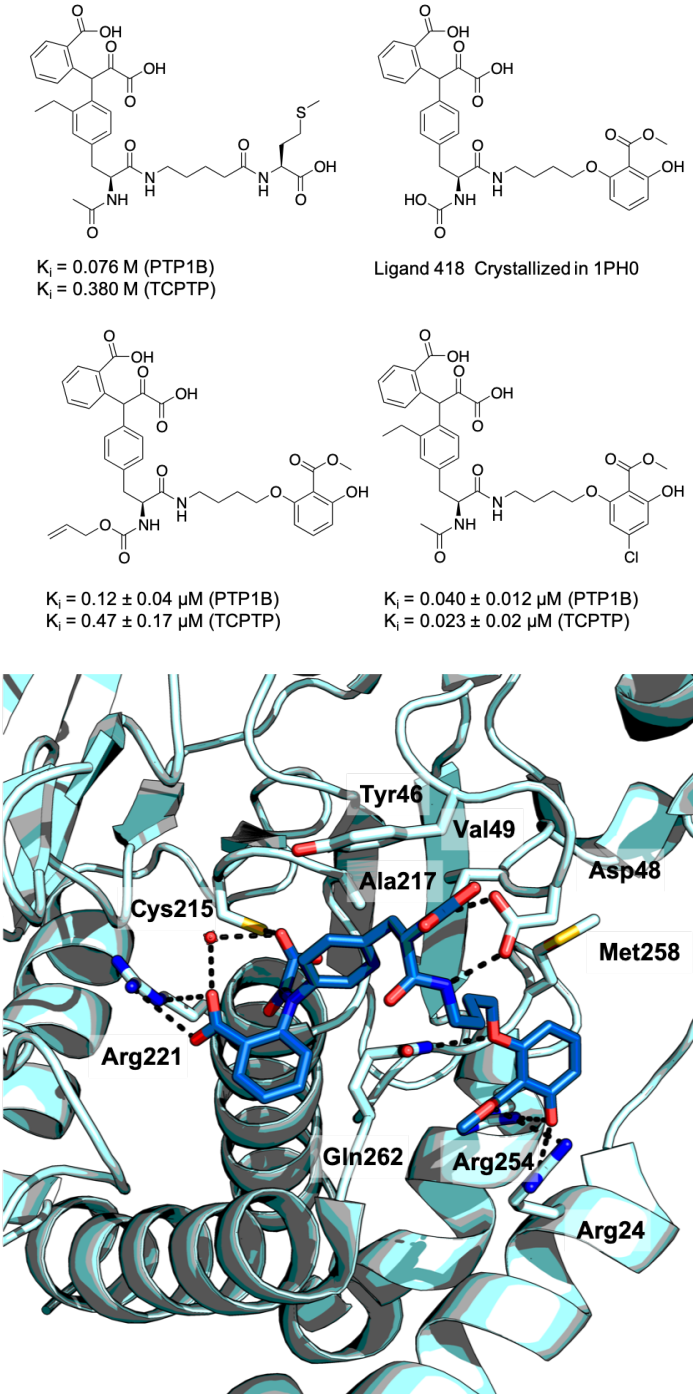


Figure 2.35. *Top, chemical structure of the most relevant oxalylarylaminobenzoic acids and their corresponding K_i. Bottom, PDB code* ***1PH0*** *with* ***ligand 418*** *inside PTP1B. The benzoic acid moieties reach the catalytic site and the site B.*

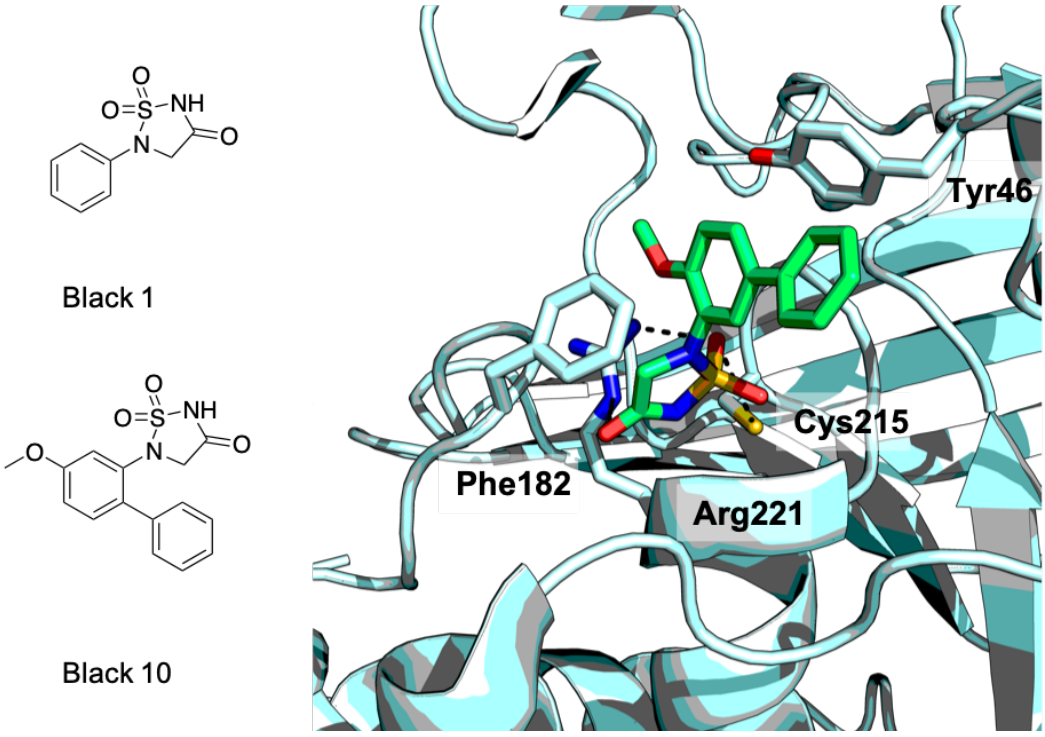


Figure 2.36. *1,2,5-thiadiazolidin-3-one-1,1-dioxide derivatives.* ***Black 1*** *was the first compound that after optimization led to* ***Black 10****, the most potent of this series. Black 10 has been crystallized inside PTP1B (****2BGD****).*

2.6.2. Allosteric inhibitors

A breakthrough in PTP1B inhibitors field came in 2004, when the research group of *K Tonks* discovered a novel allosteric pocket that was selectively filled by small molecules.[168] The previously described benzbromarone derivative named **BB3** [169] was described as a moderate inhibitor (IC_{50} = 350 μM) and crystallized in complex with PTP1B (**1T48**) (**Fig. 2.37**). That crystal structure showed the binding site at more than 20 Å away from the active site occupying the site of Trp291, thus blocking the interaction between helix α_7 (residues 282–298) and helixes α_3 and α_6. Interestingly, the first authors predicted the complex of this benzbromarone inside the PTP1B active site by docking techniques.

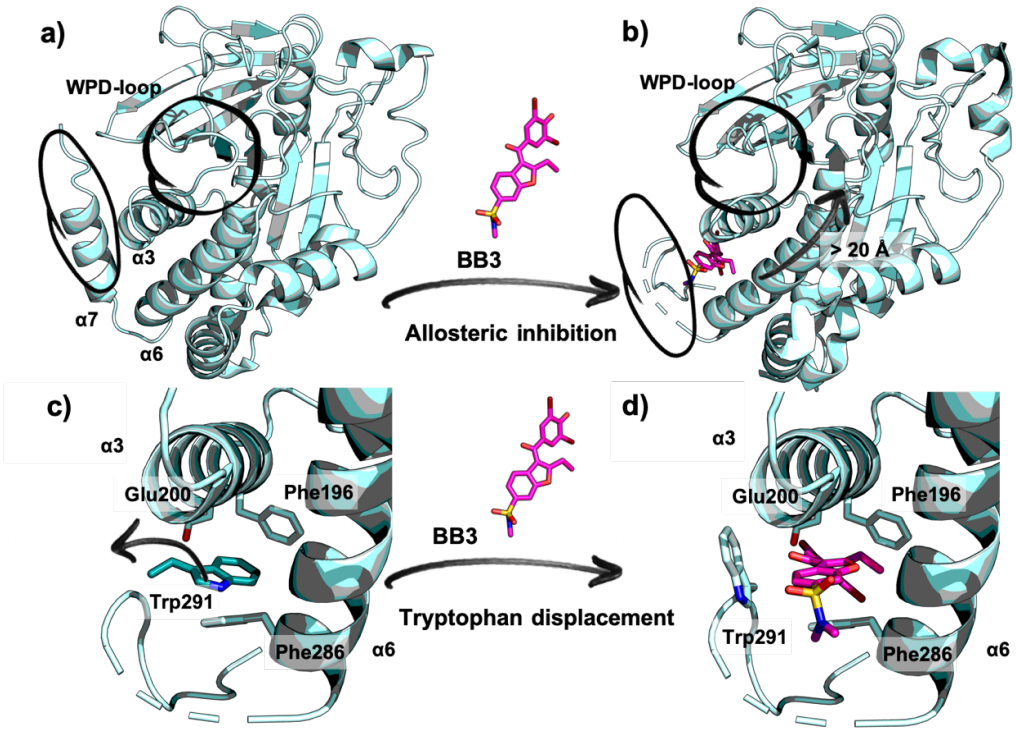


***Figure 2.37**. **a)** PTP1B in close conformation (PDB code **1PTY**). **b)** General binding mode of **BB3** inside the PTP1B allosteric site (PDB code **1T48**). **c)** The general structure of allosteric site (in pale cyan) without the ligand and the position the Trp 291 in the close conformation (in deep teal). **d)** Trp 291 displacement after the **BB3** binding. Notice the overall α-7 disruption after the **BB3** binding between a) & b) and how the WPD-loop changes from closed to open.*

After **BB3**, other compounds were subsequently developed by growing the sulfonamide group with hydrophobic moieties trying to surround the Phe208 and increase the π-π interactions. As a result, compounds **892** and **FRJ** were obtained with IC_{50} values of 22 and 8, respectively; and crystallized under PDB codes **1T49** and **1T4J**, respectively (**Fig. 2.38**). Visual inspection of these allosteric-complexes showed the WPD-loop in the open conformation, leading to the hypothesis that somehow these molecules prevented the closure of the WPD-loop during the PTP1B catalytic cycle. Computational tools became necessary to understand the precise mechanism of action of this type of compounds. Wei Cui et al shed light on this mechanism by studying the free energy variations based on Umbrella

Sampling (US) [170]. They confirmed that these small molecules blocked the normal movement of the WPD-loop increasing by 7.95 kcal/mol the movement of the Cα of Trp179, which is involved on the close-open movement of the WPD-loop. The result is the increased rigidity of the WPD-loop and the α_3 and α_6 helixes upon ligand binding. Shuai Li et al using principal component analysis (PCA) and dynamic cross-correlation matrices (DCCM) also confirmed these results, wherein they showed that the specific molecular mechanism is made up by the rotation of Trp179 which causes the downward movement of the WPD loop [171].

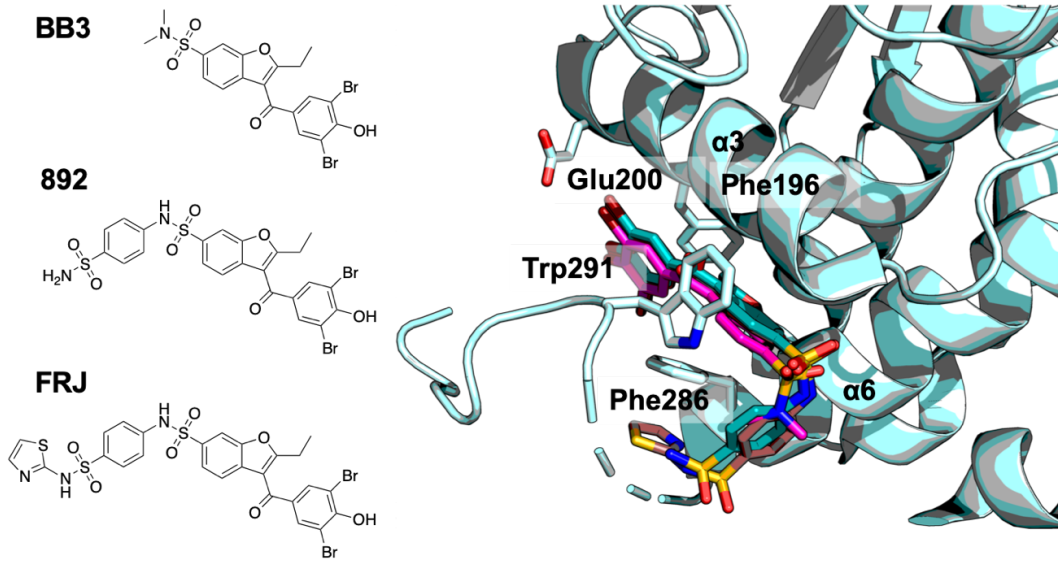


Figure 2.38. *Superimposition of* ***BB3*** *(magenta),* ***892*** *(deep teal) and* ***FRJ*** *(brown) inside the allosteric site of PTP1B. Notice how by growing the molecule it surrounds the side chain of Phe286 increasing its interactions.*

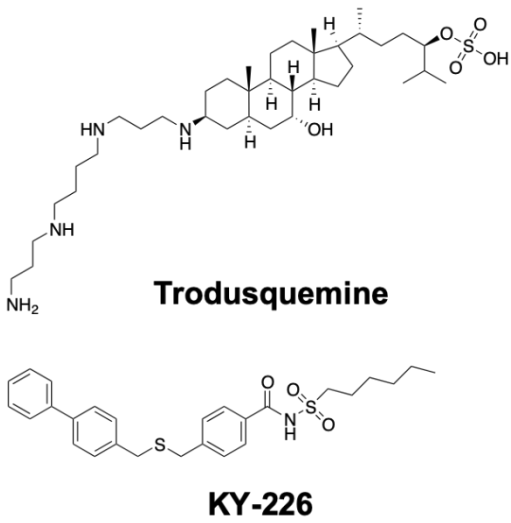


Figure 2.39. *Latest developed small molecules against PTP1B.*

A similar mechanism of action was proposed for the sterol trodusquemine, (also known as **MSI-1436**) that was under clinical trials (NCT02524951), was proposed to present the same type of action mechanism (**Fig. 2.39**) [172]. The experimental technique transverse relaxation optimized spectroscopy (TROSY) was applied for the spectra analysis of PTP1B and upon MDI-1436 complex, showing that α_7-helix and residues Leu299, His310, Ile311, Val334 and Ser393 were most affected with chemical shift perturbations confirming the allosteric mechanism [172]. One of the latest small molecules developed against PTP1B with promising results was **KY-226**, with an IC_{50} = 0.28 μM, that exhibits no agonist activity against peroxisome proliferator-activated receptor γ (PPARγ) a validated anti-target [173]. In vivo experiments in mice have provided promising results for **KY-226** as a potent and efficient anti diabetic agent. Moreover **KY-226** appears to exert anti-obesity effects via hypothalamus leptin signal in obese mice, in contrast to other antidiabetic molecules such as pioglitazone [174].

2.6.3. Natural products

Due to their chemical diversity, natural products have been usually used in drug discovery in order to find robust and viable lead candidates [175]. One of the most extensive review of PTP1B inhibitors from natural origins, cite more than 300 compounds and metabolites with inhibitory activities against PTP1B, including a chemical diversity [176]. Flavonoids are one of the first groups of molecules to be identified with notable activity against PTP1B. Five natural flavonoids from *Broussonetia papyrifera* showed inhibitory activity against PTP1B with IC_{50} values ranging from 4.3 to 36.8 μM [177].

This was one of the first flavonoids series to be published, which was followed by more than 100 structures including chalcone, dibezylmethane, flavanone, dihydroflavonol, isoflavanone, isoflavone, flavone, biflavonoid, and flavonoid glycoside derivatives. Despite this huge number of molecules, the inhibitory activity only reaches the micromolar range and further selectivity studies are needed [178].

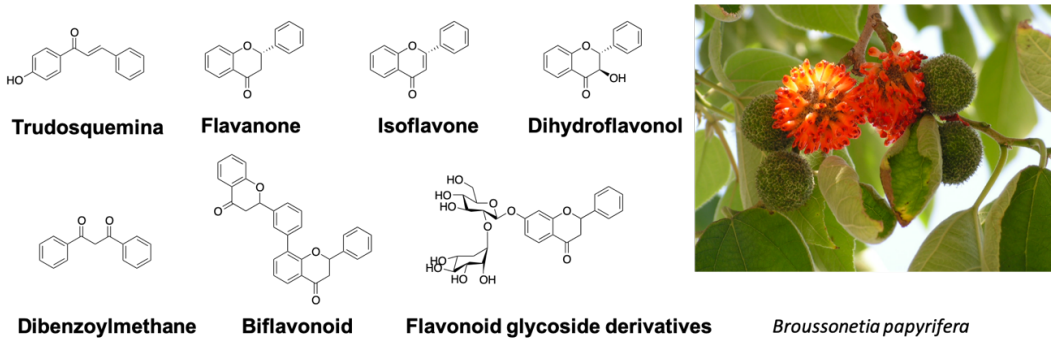


Figure 2.40. *Some chemical structures of the natural products isolated from Broussonetia papyrifera.*

Bromophenols are also an interesting group of molecules that show PTP1B inhibitory activity. This group of molecules comes mainly from marine algae extracts that have been widely used in traditional Chinese medicine due to their anti-hyperglycemic effect. In fact, the ethanol extract from *Rhodomela confervoides* shows this mentioned effect in rats. *Shi Dayong et al.* [179] measured the inhibitory activity of the isolated bromophenol derivatives (**Fig. 2.41**) against PTP1B. These compounds present IC_{50} values ranging from 0.84 to 2.4, these results suggest that the traditionally application of this plant as anti-hyperglucemic may be due to PTP1B inhibition as a possible mechanism of action.

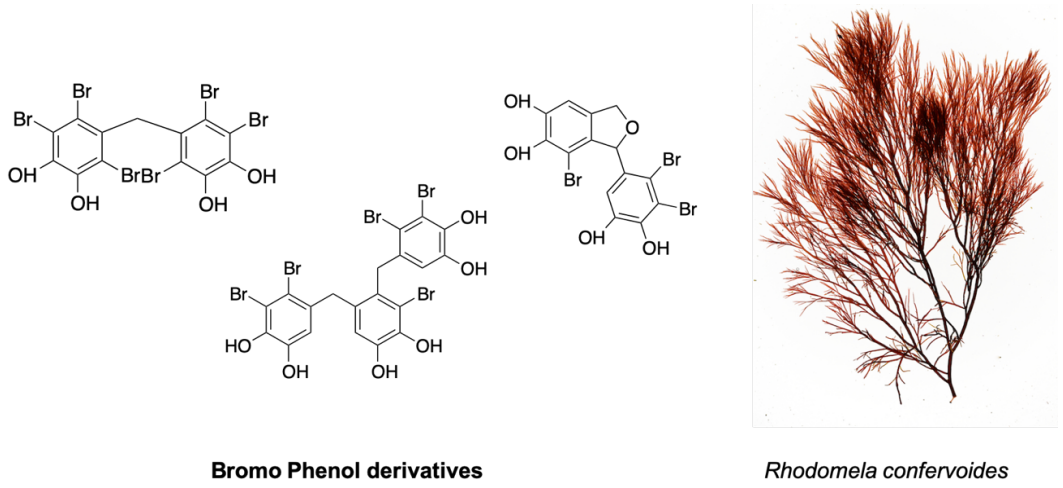


Figure 2.41. *Chemical structure of some of the bromophenol derivatives extracted from Rhodomela confervoides.*

Pentacyclic triterpenoids are also identified as PTPB inhibitors. Lithocholic acid (LCA) is a natural product that inhibits PTP1B with moderate activity (IC_{50} = 12.74 μM) (**Fig. 2.42**). A SAR study carried out *Bing et al.* [180] provided compound **BING-14B**, a potent PTP1B inhibitor with IC_{50} =1.62 μM and 14-fold selectivity over the homologous enzyme TCPTP. The binding mode of these triterpene derivatives was studied by docking techniques and the authors suggested that the carboxylic acid mimicked the phosphate group, thus fitting inside the active site.

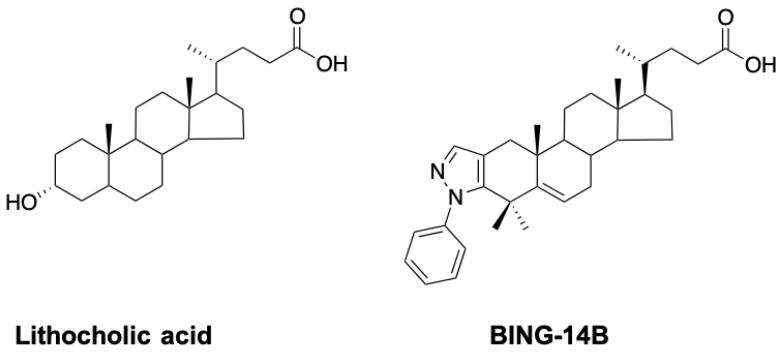


Figure 2.42. *Chemical structure of Lithocholic acid and **BING-14B**.*

2.6.4. Vanadium derivatives

As mentioned above (**Section 2.1.3**), vanadium derivatives are a potent

PTP inhibitors. The pharmacological relevant oxidation states of vanadium are oxidovanadium or metavanadate (IV) ($VO(OH)_2$) and vanadate (V) (anions $H_2VO_4^-$). Recalling the action mechanism, vanadate could complex the PTP active site by mimicking the substrate phosphate group and also the transition state analogues (TSA) (**Fig. 2.7**). In fact, direct treatment with vanadate derivatives have shown in vivo enhances glucose uptake and fat metabolism, which has been explained by its PTP1B inhibitory activity [181]. Another fact is that vanadyl (IV) (VO^{2-}) acts on the cell plasma membranes facilitating the glucose cell permeation and impairing lipolysis [182]. Vanadate not only interacts with PTPs; due to its similarity with phosphate in geometry, charge and van der Walls radius, it could also bind to other enzymes such as kinases [183,184]. The ADME properties of vanadate and chelating derivatives have been widely studied, showing notably increased activities when those are coordinated because of the high cell permeability [185,186]. Although it is noteworthy that vanadium is a metal present in nature and human body tissues in small concentrations (around 60 nM), the lethal doses were measured as LD_{50} 0.15 mM per kg of body weight for sodium metavanadate. The oral doses of vanadyl sulfate that present pharmacological responses in patients were under of 1mM per day, giving us an idea of the wide therapeutic window [187,188]. Despite these promising results, vanadate and vanadyl are not the best molecules for oral administration; however, vanadium complexes with maltolate such as Bis(ethylmaltolato)oxidovanadium(IV) (BEOV) and BMOV-Bis(maltolato)oxidovanadium(IV) are the best way to increase (up to 3 times) absorption and bioavailability [189]. Taking into account these results, it seems contradictory that at the moment no clinical trials or further studies on vanadate as a clinical candidate are being carried out. A critical review by *Scior* et al.[190] highlighted the relevance of this therapeutic option for diabetes treatment and the neglect of the pharmaceutical industry to develop future pharmacotherapies. In the conclusions of this review it is written:

"Yet, in a more critical view, developing new organic ligand variations of

vanadium-containing antidiabetic drugs could merely have an impact on pharmacokinetics - and since all variations would end up in bioliganded oxidovanadium (IV) or oxidovanadium (V) species no means would exist to profile vanadium drugs to bind to PTP1B as their specific target without binding to other PTPs. Reaching that point - apparently very soon – the "big" transnational pharmaceutical industries had to conclude that vanadium compounds can never replace insulin therapy, and therefore vanadium can be dismissed as a drug-like option for R&D."

2.6.5. PTP1B related diseases

Apart from the above-mentioned diabetes and obesity, PTP1B has been related to other diseases such as cancer, inflammation and heart diseases. In cancer, this phosphatase acts both as tumor-suppressing and tumor-promoting agent, however this always depends in the context of the tumour type and expression [191,192].

2.6.5.1. Cancer

In breast cancer, PTP1B has been linked clearly with an increased overexpression in 75% of the cases, including different stages of the disease [193]. This is because PTP1B has been associated with the HER2 (hormone epidermal receptor type 2) with high levels of coexpression, which may be translated to PTP1B and HER2 cooperating in the pathogenesis of this subtype of breast cancers [194]. Nevertheless, another study published in 2011 showed that no correlation existed between the expression of PTP1B and HER2. A meta-analysis of more than 1402 tissue samples in different ages, tumor sizes, T stages, histologic grades, lymph node statuses, or histological subtypes with 49% PTP1B overexpression shows no significant association between PTP1B expression and HER2 expression, making PTP1B expression an independent factor [194]. A recent publication shed light on this issue stablishing that PTP1B promotes breast cancer metastasis by regulating phosphatase and tensin homolog (PTEN) but not epithelial-mesenchymal transition (EMT), being PTP1B an

independent factor from HER2 positive cancers but increasing its poor prognosis [195]. This fact was also suggested by *Sheng-rong Sun et al*. [196] who presented a study with 67 breast cancer samples HER2 positive with 70% of PTP1B overexpression. A deep analysis showed how the action of PTP1B was linked to increased levels of phosphorylated signal transduction and activator of transcription 3 (STAT3), and chemokine ligand 5 (CCL5), which can play a crucial role in the development of breast cancer. Another type of tumor linked with PTP1B is prostate cancer in which PTP1B has been well established to play a regulatory role in metabolic signal induction by androgen stimulation, increasing its activity. It has been found that in prostate cancer PTP1B and androgen receptor (AR) genes were amplified and associated with a subset of high-risk primary tumors [197]. Moreover, in vitro studies show how PTP1B depletion or inhibition decreases migration and invasion of androgen-dependent human prostate tumor cells [198]. These findings make PTP1B a suitable candidate for preclinical testing in human prostate cancer [199].

2.6.5.2. Inflammation

Another important role played by PTP1B is the modulation of immune signaling. This comes from its relationship with the JAK-STAT pathway, where PTP1B specifically targets JAK2. Moreover, recent evidences may consider obesity as an inflammatory condition, linking the fields of immunology and metabolism, where PTP1B presents a key role in both. PTP1B is able to negatively regulate cytokines and TLR signaling by decreasing TN alpha, Il-6 and IFN-beta production coming from macrophages stimulated by LPS. This increased concentration of circulating cytokines has been observed in PTP1B $^{-/-}$ knock out mice, which present higher number of immature IgM, IgD and B cell proliferation [200]. Furthermore, *Zabolotny et al.*[201] showed that PTP1B overexpression is linked with higher adipose tissue macrophage markers with CD68 and TNFa; of which, TNFa induces the expression of PTP1B through transcriptional activation by NF-kB. With this data, the authors suggested

that PTP1B overexpression in multiple tissues in obesity is regulated by inflammation and that PTP1B may be a target for anti-inflammatory therapies. PTP1B is likely to contribute to the downregulation of the inflammatory process, especially considering the interplay between obesity and chronic systemic inflammation. The finding that PTP1B expression was controlled by inflammation in mice gave increased support to the above-mentioned hypothesis.

2.6.5.3. Heart diseases

Gomez et al. [202] showed for the first time the implication of PTP1B in heart diseases, focusing their research in chronic heart failure and endothelial cells. They evaluated the impact of PTP1B in mice with chronic heart failure, using two approaches: either a 2-month pharmacological PTP1B inhibition; or, a gene deletion of PTP1B $^{-/-}$. Their results show an improved insulin sensitivity added to a restoration of the endothelial function, which finally leads to reducing cardiac fibrosis, myocyte hypertrophy and cardiac expression of atrial natriuretic peptide (ANP). On the contrary, elevated activity of PTP1B in rats show to be linked with induced pressure overload causing cardiac insulin resistance that proceeds and accompanies mitochondrial and systolic dysfunction, which can be associated with heart failure [203]. An extensive review of PTPs in cardiac physiology and pathophysiology has been carried out by *Poizat et al.* [204] highlighting the role of PTP1B. Recent findings show that either the lack of the inhibition of PTP1B produces resistance against heart failure via cardiac improvement of VEGF and angiogenesis signaling [205,206]. Moreover, these authors also underline that a series of compounds inhibit both PTP1B and also LMWPTP, a low molecular weight protein tyrosine phosphatase that has been described as potential target for heart failure by increasing the phosphorylated levels of the insulin receptor, Akt and PI3-K [207]. Recent in-vivo zebrafish studies have shown that treatment with **MSI-1436** stimulated the regeneration ratio of heart muscle and thin tissue after induced cardiac ischemia. In addition, improved heart function in mice has

been observed as well as increased survival in the infarcted zone with the administration of **MSI-1436** 24 h after the injury [208].

2.6.5.4. Alzheimer disease (AD)

Recent advances in AD have linked its origin to an autoinflammatory disease and with diabetes. In fact, some authors name AD as diabetes type III [209–211]. As it was mentioned before, PTP1B has been linked with both types of diseases, diabetes and autoinflammatory disease, so there could be a direct connection between AD and PTP1B. Defective neuronal insulin and leptin signaling have been linked with AD, in fact there is no leptin signaling within degenerating neurons and the insulin resistance in brain is highly increased in individuals with AD [212,213]. In this sense, it has been proven that PTP1B down-regulation restores hypothalamic insulin and leptin signaling [214]. Moreover, PTP1B was recently described as a positive regulator of neuroinflammation processes and is found increased in the microglia and neuronal endoplasmic reticulum (ER) during stress [215]. Due to this fact, some authors have proposed PTP1B as a promising target in AD and other neurological diseases [216–218].

2.6.6. PTP1B Clinical trials

During all this introduction on PTPs the relevance of bioavailability and selectivity as the major challenge in the discovery of PTP inhibitors has been highlighted. This has led to some molecules reaching clinical trials with more or less success.

One of the recent failures is the application of MSI-1436C in metastatic breast cancer (NCT02524951) which started its phase 1 in 2015 and was terminated due to the lack of interest by the sponsor in 2018.

PTP1B inhibitor (ISIS-PTP1BRx) from Ionis Pharmaceuticals, Inc. is an antisense RNA that blocks the translation of the PTP1B encoding mRNA. This molecule has reached phase 2 studies where it proved good results for safety, tolerability, and efficacy in type 2 diabetes (NCT01918865) [219].

Despite this success, currently there is no more clinical trials going on for this molecule.

Nowadays the only active clinical trial (under recruitment) is NCT03189355 that aims to study the expression of PTP1B and modification on the body composition in patients with septic shock. This study is based in the association between PTP1B and loss of muscle mass in patients in sepsis and in resuscitation. This is because PTP1B could play a role in the protein synthesis abnormalities observed during sepsis leading clinically to impaired body composition including the loss of muscle body mass.

2.7. Pleiotrophin and Midkine

Pleiotrophin (PTN UniProtKB - P21246) and Midkine (MK UniProtKB - P21741) are neurotrophic factors that function as cytokines and growth factors by regulating many processes like cell growth, adhesion, migration, differentiation, proliferation and also inflammatory responses and tissue regeneration [220–222]. These two natural ligands share over 50% of identity in amino acid sequences (**Fig. 2.43**) and have been proven to bind PTPRZ1, but they also interact with other receptors. For example PTN binds to PTPRS (PTPR Sigma) [223], and MK with LDL receptor-related protein (LRP), Neural cell adhesion molecule (NCAM also called CD56), and anaplastic lymphoma kinase (ALK) [221,224]. The 3D structure of these neurotrophic factors has been solved by NMR methods (**Table 2.4**). Although the detailed binding mode of both ligands to PTPs remains unsolved, the interactions with the PTRZ1 chondroitin sulfate chains is quite well experimentally supported [91,225].

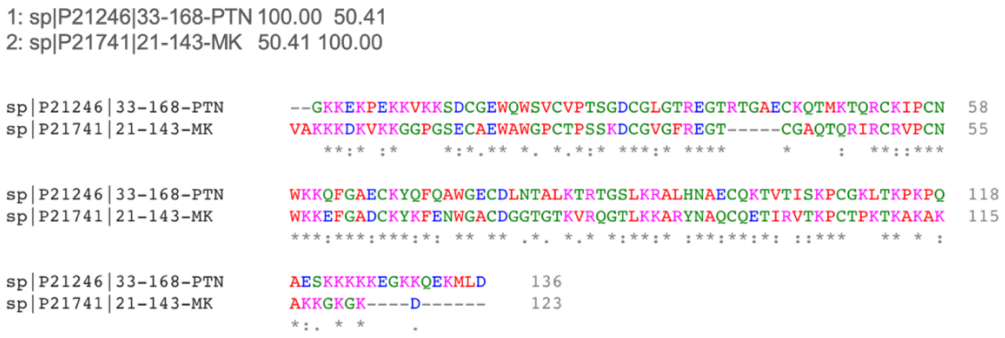


Figure 2.43. *PTN and MK sequence alignment. Data retrieved from Clustal Omega web server. Sequences were obtained from the corresponding entry number in Uniprot.*

Table 2.4. *Available 3D dimensional structures solved by NMR methods for PTN and MK in the PDB database.*

Protein	***PDB Code***	***Sequence***
PTN	2N6F	33-168
MK	1MKC	84-126
MK	1MKN	23-81

MK and PTN are widely expressed during embryogenesis [226–228]. However, this expression becomes restricted to a few cell types in adults, especially in the CNS. This fact has been linked to their roles in cell development and repair [229]. Both types of ligands induce oligodimerization of PTPRZ1, which inactivates the intrinsic PTP activity (**see section 2.4.1**). The expression of PTN and MK is upregulated generally as a response to tissue damage such as ischemic injury in the brain, inflammation, cell migration, repair or insolating the damage in CNS [230–232]. Also the overexpression of MK and PTN is related to malignant transformations that have been linked in many tumors and chronic inflammation processes [233–235]. Interestingly, these two factors have been found to be increased in different CNS diseases such as Parkinson, Alzheimer disease (AD) and also in drug addiction [236–238]. For this

reason, the therapeutic potential of these factors has been widely explored using mice models and pharmacological treatments in an attempt to mimick the effect of both factors. Recent experiments suggest that drug abuse, including alcohol, regulates the expression of PTN and MK in the brain, which could be involved in the regulation of drug effects, such as addictions [239]. For instance, PTN (PTN-/-) and MK (MK-/-) knockout mice have shown increased drug addiction to amphetamines and increased vulnerability to amphetamine-induced neurotoxic effects [240,241]. Also, this model has shown to be more sensitive to the conditioning effects of alcohol and has an increased amphetamine-seeking behavior [106,107,242]. On the contrary, mice with overexpressed PTN (PTN-Tg) have shown to present neuroprotective effects after the injection of oxidopamine, also known as 6-hydroxydopamine or “6-OHDA” and similar to MPTP, a potent neurotoxic compound that triggers Parkinson's disease (PD) [243,244]. In addition, this PTN overexpressed model did not exhibit the addictive exacerbated behavior present in knock-out mice. Moreover, the rewarding effect of alcohol was also studied applying the conditioned place preference (CPP) experiment in mice [239,245]. Results show that after alcohol administration, the PTN-/- mice are clearly influenced by the ethanol rewarding effect; but on the contrary, PTN-Tg mice were not showing any sign of ethanol influence. For this reason, PTN is thought to modulate behavioral responses to ethanol, making it a suitable approach for drug discovery. Mimicking the effects of PTN through small molecules has been one of the aims of this thesis and the results will be explained and discussed later (section **2.10**).

2.8. Alcoholism

Alcoholism, also called alcohol use disorder (AUD) is a psychological disorder in which a person has a desire or physical need to consume alcohol, even though it has a negative impact on his life. The National Institute on Alcohol Abuse and Alcoholism (NIAAA www.niaaa.nih.gov) describes alcohol use disorder as a "drinking problem that becomes

severe". This definition is directly linked to the meaning of drug addiction. The National Institutes of Health (NIH) define it as “uncontrollable, compulsive drug seeking and use, even in the face of negative health and social consequences.”[246]. Just to briefly summarize, the impact of this disease, according to the World Health Organization (WHO www.who.int) is 3 million deaths every year resulting from the harmful use of alcohol, which is a 5.3% from the total number of deaths per year. Moreover, the harmful use of alcohol is related with more than 200 diseases and injuries and this is more exacerbated in young people (the age group ranging from 20 to 39 years old) which approximately causes the 13.5 % of the total deaths.

There are a variety of factors that affect AUD, including individual and social factors that can increase the vulnerability to this disease, all of them are implicated in the magnitude of alcohol-related problems. The WHO summarizes in **Figure 2.44** the causes and health problems in alcohol consumption. The impact of alcohol consumption depends mainly on two related dimensions: i) the drinking pattern, and ii) the amount of drinking. Following the NIAAA drinking levels, moderate alcohol consumption is considered up to one drink per day for women and up to two drinks per day for men. On the contrary, heavy alcohol consumption is considered with five or more binge drinking in the past month, being binge drinking five or more alcoholic drinks for males or four or more alcoholic drinks for females.

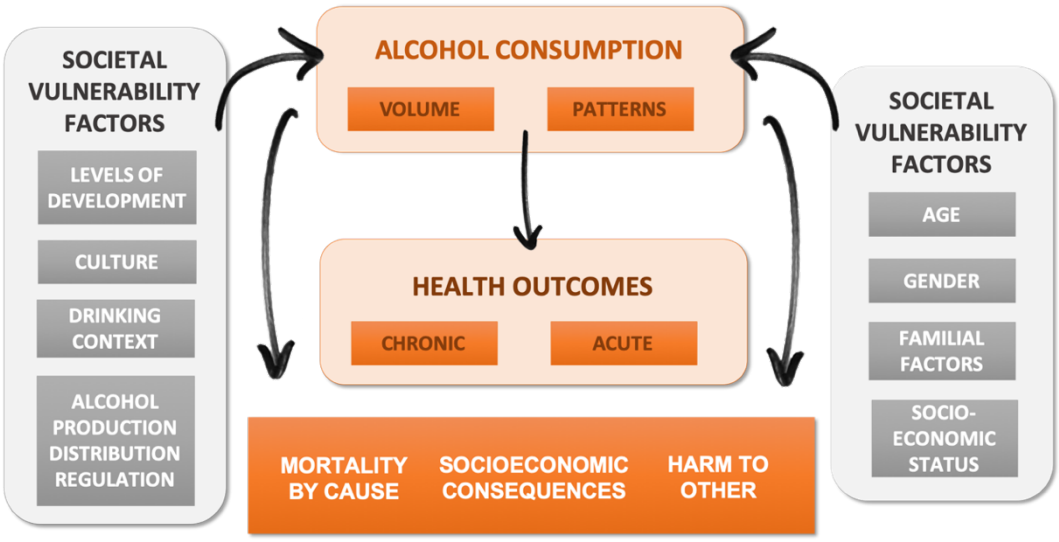


Figure 2.44. *WHO scheme describing influential factors, risks and effects of the AUD.*

2.8.1. Molecular basis of alcoholism

Chronic alcohol abuse induces a widespread of neuroadaptations in the CNS that can last a lifetime. This process involves the remodeling in synapses and neuronal connections that depend ultimately on gene expression [247]. Alcohol dysregulates neural circuits that mediate both motivated and goal-directed behaviors. The dopaminergic connection between the ventral tegmental area (VTA) and the nucleus accumbens are the most affected [248]. This area has been linked to the rewarding and behavioral effects of alcohol and other drug abuse substances. Ethanol is a small molecule in comparison with other abuse substances, and for this reason, its target or targets at the molecular level remain unclear. There are well-known targets of alcohol that come mainly from animal studies, but yet the mechanism of action remains unclear in many cases [249]. One of these targets is DNA, or more specifically gene expression. Chronic ethanol consumption produces neuroadaptation processes where the normal chromatin remains condensed, decreasing the gene expression, and finally producing anxiogenic effects (**Fig. 2.45**) [250]. At the molecular level it has been shown that chronic consumption of alcohol increases levels of Histone Deacetylase (HDAC) and decreases the levels of histones H3 and H4 [247]. Interestingly, the blockage of HDAC activity by small molecules such as

Trichostatin (see **Chapter III**) has shown to prevent the withdrawal-related anxiety and reinforcing effects of ethanol, at least in animal models [251–253]. This discovery has made HDACs, especially HDAC1, an interesting novel target in the AUD treatment, and there are currently some HDAC inhibitors approved for this purpose.

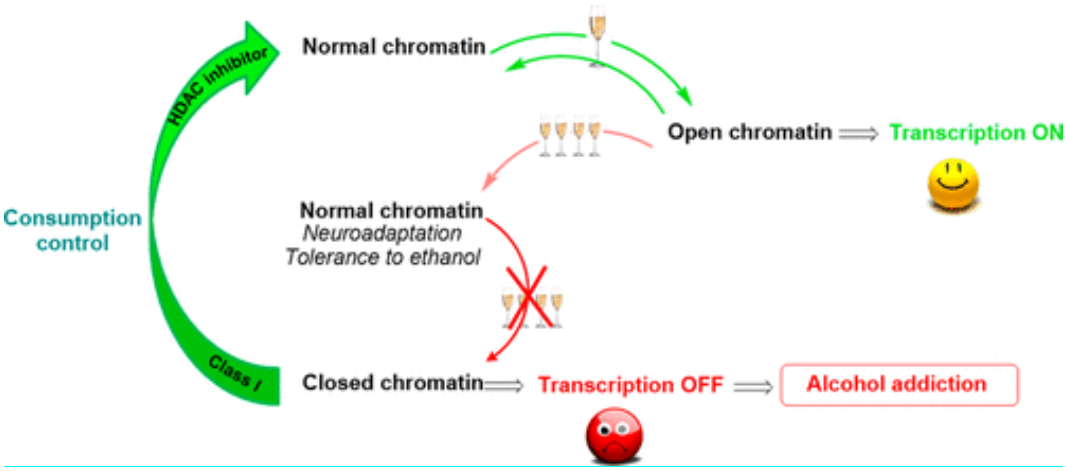


Figure 2.45. *Proposed mechanism of action of alcohol that alters gene transcription. Figure taken from J. Med. Chem.2018, 615, 1745-1766.*

Moreover, in dopaminergic (DA) neurons, alcohol consumption decreases intensively DNA transcription, thus decreasing its activity, which may play a crucial role during both addiction and the AUD process [247]. The dopaminergic system is highly affected during chronic alcohol consumption by the decreased levels of dopamine, metabolites, and protein tyrosine hydroxylase in the striatum, producing the associated symptoms of dysphoria, craving and relapse [254,255].

The gamma-aminobutyric acid (GABA) neurotransmission system is also affected and modified by chronic alcohol consumption [256]. The $GABA_A$ receptor can be modulated allosterically, and the subunit expression can be differentially altered by chronic ethanol consumption, which induces tolerance and decrease of its activation. $GABA_A$ receptors play an important role in the mesolimbic dopamine (DA) reward pathway and interacting in the stress circuitry. This produces vulnerability to alcoholism and anxiety [257]. GABA modulates emotion and response to stress, so when this homeostasis is altered subsequent addiction can occur.

Moreover, the serotonergic system is also affected by alcohol consumption and has also been related to AUD as it plays a key role in the alcohol intake regulation. It has been shown that among other alterations, alcohol consumption affects the functionality and expression of 5-HTT (5-hydroxi tryptamine transporter), decreasing its levels in the postsynaptic cleft, especially in the hypothalamus and amygdala, which leads to rewarding effects of alcohol [258,259]. Increased serotonin levels have been linked increased alcohol consumption, therefore, the selective blockage of the 5-HT receptor (5-HT3) has shown to decrease alcohol self-administration [260].

Potassium channels have also been related to prolonged alcohol drinking and alcohol dependence due to an increase of plasticity and intrinsic excitability [261]. AUD produces adaptations in potassium channels (KCa2, KV7, and GIRK) leading to impaired regulation of the after-hyperpolarization and aberrant cell firing. Interestingly, among the 79 potassium channels, the KV7 (Included in gene Kcnd2) has been found to be less expressed in high drinking conditions [262]. Retigabine, a FDA-approved KV7 channel positive modulator, has shown to be more effective at reducing alcohol intake in high versus low drinking rat populations [262,263]. As mentioned above, pleiotrophin (PTN) and midkine (MK) are two cytokines with important functions in the survival and differentiation of dopaminergic neurons, which are overexpressed in different brain areas after drug administration, including ethanol. In mice, it has been demonstrated that PTN-/- and PMK-/- mice models are more sensitive to the reinforcing effects of alcohol, meanwhile transgenic mice that overexpress PTN (PTN-tag) do not present this exacerbated rewarding effects produced by drinking ethanol [106,231,239].

2.8.2. Options for the treatment of AUD

- Behavioral treatment based on changing drinking behavior by counseling led by health professionals such as psychologists,

social workers and alcohol counselors. It is based on Cognitive–Behavioral and Motivational Enhancement therapy.

- Mutual-Support Groups: Alcoholics Anonymous (AA), provide peer support for people quitting or cutting back on their drinking and usually combined with treatment led by health professionals.
- Medications: There are three approved drugs for the treatment of alcoholism.

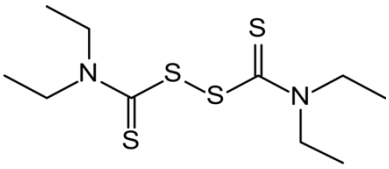
There is no fixed treatment for AUD, but the NIAAA reflects that ultimately choosing to get treatment may be more important than the approach used, as long as the approach avoids heavy confrontation and incorporates empathy, motivational support, and a focus on changing drinking behaviors.

2.8.3. Drugs in AUD

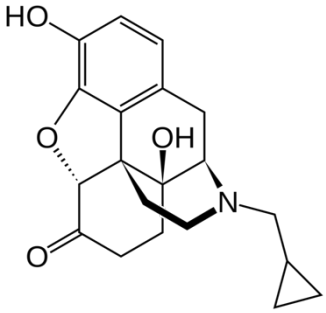
There are three drugs approved for the treatment of AUD.

2.8.3.1. Approved drugs

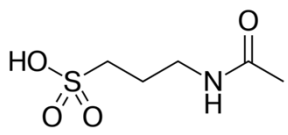
Disulfiram: works by inhibiting the acetaldehyde dehydrogenase enzyme, causing many of the non-desired effects of a hangover to be felt immediately following alcohol consumption. Disulfiram does not reduce alcohol cravings, so a major problem associated with this drug is the extremely poor compliance and finally failure to follow the treatment. This is the main reason why disulfiram has been substituted by the combination of the new drugs, naltrexone and acamprosate.



Naltrexone: naltrexone and its active metabolite 6β-naltrexol are competitive antagonists of the μ-opioid receptor (MOR), the κ-opioid receptor (KOR) to a lesser extent, and, to a far lesser extent, of the δ-opioid receptor (DOR). As an opioid antagonist, naltrexone blocks the positive-reinforcement effects of alcohol and allows the person to stop or reduce drinking. Naltrexone has been shown to decrease the amount and frequency of drinking, and the desire for drinking alcohol to a greater extent.

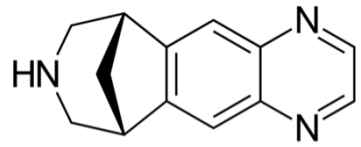


Acamprosate: works as an antagonist the NMDA receptor and as a positive allosteric modulator of the $GABA_A$ receptor. Acamprosate acts by reducing the glutamate surge induced by alcohol and the enhancement of GABA signaling at $GABA_A$ receptors via positive allosteric receptor modulation. This drug makes it easier to maintain abstinence. The main side effects are the increase of suicidal behavior, major depressive disorder, and kidney failure, for these reasons this drug needs to be monitored.

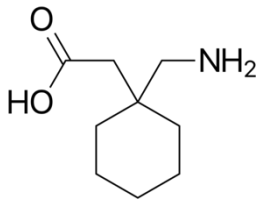


2.8.3.2. Drugs under study

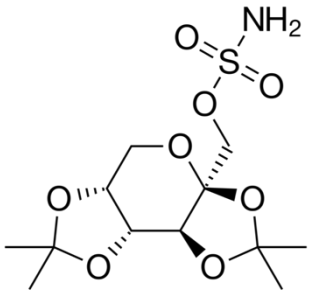
Varenicline: It is a high-affinity full agonist on α_7 nicotinic acetylcholine receptor and is a partial agonist on the $\alpha_4\beta_2$, $\alpha_3\beta_4$, that leads to the release of dopamine in the nucleus accumbens. The anti-smoking drug varenicline significantly reduced alcohol consumption and craving among people with alcoholism and specially in individuals who also smoked [264]. This fact was observed when smoking patients were treated with varenicline and showed a concomitant reduction in drinking [265]. Varenicline is currently in Phase 4 with more than 20 clinical studies completed.



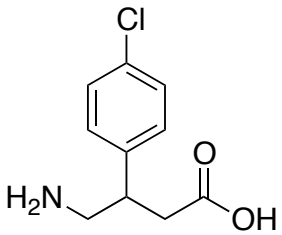
Gabapentin: acts as an inhibitor of $\alpha_2\delta$ subunit-containing voltage-dependent calcium channels (VDCCs). Gabapentin was shown to increase abstinence and reduce heavy drinking. Those individuals taking the medication also reported fewer alcohol cravings and improved mood and sleep with a favorable safety profile (NCT01141049) [266]. Gabapentin presents 11 studies completed to treat AUD.



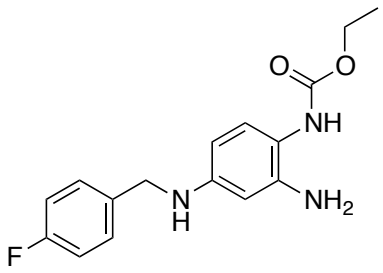
Topiramate: the mechanism of action of topiramate is not fully understood, but it has been proposed that it may alter the activity of its targets by modifying their phosphorylation state [267]. The targets of topiramate include sodium and calcium channels, GABA receptors, AMPA/kainate receptors, and the carbonic anhydrase isoenzymes [268]. Topiramate was shown to help people curb problem drinking, particularly among those with a certain genetic makeup that appears to be linked to treatment's effectiveness [269]. Topiramate presents twelve completed clinical assays for the treatment of AUD.



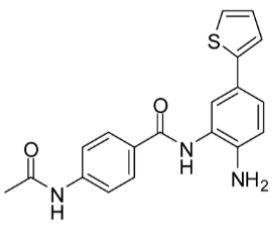
Baclofen: is a GABA agonist used as a skeletal muscle relaxant used for pain relief. Some research points to baclofen as an effective drug for the decrease or suppression of alcohol craving, and has been the subject of fourteen studies for AUD treatment [270]. Nevertheless, recent studies highlight the lack of efficacy compared to placebo or, the reduction in alcohol consumption but not the maintenance of abstinence [271,272].



Retigabine: or Ezogabine (D23129) is an approved antiepileptic agent, that selectively activates potassium ion channels Kv7.2-Kv7.5 but not the cardiac Kv 7.1 channel, thereby avoiding cardiac side effects. Its pharmacological activity in the reduction of alcohol intake has been confirmed in rats and rabbits [262,273,274]. Nevertheless, there is still no clinical trials to prove its efficacy in humans, although there are 18 completed trials for the treatment of epilepsy and seizures.

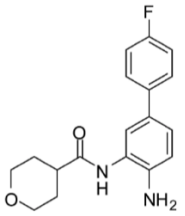


HDAC inhibitors (HDACIs): Only a few HDACIs have been tested in preclinical studies for AUD and other psychiatric diseases [275,276]. BBB permeability is mandatory to develop HDACIs for AUD. An extensive study with the aim to find selective HDAC-1 and HDAC-2 inhibitors has been carried out [277]. This study highlighted compounds **Cpd-60** and **BRD4884** as possible candidates for further studies in preclinical studies of AUD due to their nanomolar IC_{50} ranges [251]. In 2018, the same research group developed **SAH**, a selective compound against HDCA1 that demonstrated its effectivity in reducing ethanol intake and relapse when tested in rats [277].



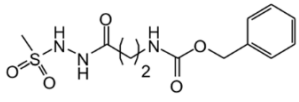
Cpd-60

IC_{50} (nM)
HDAC 1 = 7
HDAC 2 = 49
HDAC 3 = 10 000
HDAC 4 = > 10 000
HDAC 5 = > 10 000
HDAC 6 = > 10 000
HDAC 7 = > 10 000
HDAC 8 = > 10 000



BRD4884

IC_{50} (nM)
HDAC 1 = 29
HDAC 2 = 62
HDAC 3 = 1090
HDAC 4 = > 33 330
HDAC 5 = > 33 330
HDAC 6 = > 33 330
HDAC 7 = > 33 330
HDAC 8 = > 33 330
HDAC 9 = > 33 330



SAH

2.9. Computational approach in the inhibition of PTPRZ1 in the treatment of alcoholism

AUD is a global impact disease with no specific treatment and with frequent relapses. Therefore, novel therapeutic approximations are urgently needed. PTN and MK have shown promising experimental results in the prevention of neurotoxicity induced by drug abuse, and also in the depletion of the positive-reinforcement effects of drug and alcohol abuse. For these reasons, our aim is to mimic the effect of both cytokines. From the pharmacokinetic point of view, the direct administration of PTN or MK is not the optimal way to study their effects. Both proteins present poor druglike properties such as high molecular weight (MW), present labile peptide bonds and the inability to directly cross the blood-brain barrier (BBB)[278,279]. Even though, other similar peptides have been effectively administrated, expenses overcome the benefits. In the initial stages of any target validation project, drug discovery is needed in order to confirm the proposed effect through the action of a small molecule. The initial hypothesis and main goal of this project is to modulate the PTN/MK signaling pathway with a small molecule. This approach involves the computer-aided design of compounds with restrictive properties such as potency, selectivity, non-toxic effects and BBB permeability. As PTN and MK are the natural ligands of PTPRZ1, to reproduce their effects with a small molecule, the intrinsic phosphatase activity of PTPRZ1 has to be blocked, thus making this phosphatase the biological target of this project. Various pharmacological strategies could be used to block the activity of PTPRZ1 [280], such as: the direct activation of the extracellular domain; increasing the stability of the inactive PTPRZ1 dimer through ppi stabilizers; direct depletion of the receptor through proteolysis; inhibition of its peptide synthesis through RNAi; among others. However, in this project, we have selected the classical approach, that is, the direct inhibition of catalytic activity in PTPRZ1.

Hit identification becomes necessary in order to have a main chemical

scaffold to modify in order to increase the activity, potency and/or selectivity. The “hit to lead” development will involve the optimization of the pharmacodynamic and pharmacokinetic properties of the selected hits so as to select a suitable preclinical candidate for further in-vivo studies.

Molecular modelling techniques are a critical point during Drug Discovery. During the development of this thesis, the rational modelling of the PTPRZ1 in complex with the developed inhibitors becomes one of the mains targets. At the beginning of this project no crystal structure of the PTPRZ1 PD1 was available, for this reason, models were built using the structures of PTPRG (closed, open and superopen) due to the high amino acid similarity (**Scheme 2.2**). In order to support the plausibility of this models, specially the superopen conformation, MD simulations were carried out using PTPRG as reference. In addition, PTP1B was included in this section using it as contrast reference since the superopen conformation is not available. That conformation was modelled, and the rest, the closed and open conformations, were obtained from the corresponding crystal structures. This first computational approach gives us the opportunity to study the behavior of the WPD-loop in all of three PTPRZ1, PTPRG and PTP1B conformations. Given the high similarity of both PTPRZ1 and PTPRG we carried out an extra modelling study using MD simulations with organic mixed solvents in order to find slight differences in the WPD-loop behavior that could help us reach and inhibit PTPRZ1 instead of PTPRG. During the development of this last study, the first PTPRZ1 structure was released in the apo conformation with no ligand bound **5AWX**, and later on the structure **5H08** bound to ligand **7WL**. For this reason, both open conformations **5AWX** and **2PBN** were used as initial structures in the solvent modelling study. Moreover, a comparative study between the available superopen crystal structures and models of PTPRZ1 and PTPRG was carried out giving showing notable differences.

As part of this project, a portfolio of developed molecules was synthesized and tested (in vitro and in vivo) against our main target PTPRZ1 but also in PTPRG and PTP1B. The obtained biological results are summarized and

further discussed. Giving these results, molecular modelling techniques were applied in order to rationalize and understand the mechanism of action and selectivity profile of the most active and interesting molecules. For this aim, all previous mentioned structures coming from crystal structures and models were used as receptors for docking techniques and the obtained complexes for further carrying out MD simulations with the aim to check stability and total binding energy. Other precise techniques such as FEP has also been applied. Finally, given the in silico results we have discussed and proposed a plausible binding mode for our molecules that allows us to rationalize their selectivity profile.

2.10. Objectives

The previous mentioned objectives related to the development of the PTPRZ1 inhibitors have been summarized and sorted:

- Molecular modelling of PTPRZ1.
- Model validation and WPD-loop study.
- Exploration of the structural differences of PTPRZ1 vs PTPRG.
- Design and synthesis of the chemical portfolio.
- Biological in vitro and in vivo studies.
- Modelling of the complexes by docking, MD simulations and FEP techniques.
- Rationalization of the in vitro and in silico results.
- Proposed action mechanism and further discussion.

2.11. Materials and Methods

Protein preparation. All PDB structures were obtained from the PDB databank (https://www.rcsb.org/). The homology models were built through the SWISS-MODEL web server using the corresponding template [281]. Water molecules were removed from crystal structures unless needed to be taken into account for further analysis, given that the Glide scoring function

considers the solvent contribution. For the protein preparation which includes the building of incomplete side chains and loops, Protein Preparation Wizard module of the Schrödinger Suite (http://www.Schrodinger.com) was applied [282]. The protonation states of charged amino acids were optimized within the PROPKA module and interestingly, in all cases the catalytic cysteine was protonated. Finally, hydrogens were optimized within the Schrödinger forcefield (ff) OPLS-AA in order to eliminate bad contacts [283].

During the course of this thesis, ensemble docking has been used by means of the Schrödinger suite of programs, where flexible ligands were prepared with the LigPrep module and docked into rigid receptors using the default parameters supplied in the Glide docking module of the suite of programs.

The ligand preparation. The entire portfolio of molecules in this section was first obtained in the corresponding 2D structures in sdf format file. The possible ionization states of all the ligand candidates were generated by the LigPrep module with default parameters at pH 7 ± 2 [284]. The global minimum energy conformation of each structure was identified using the OPLS 2005 force field and tautomers were also generated [285].

Grid generation. The grid was built by the **Glide** grid generation module. The box center was fixed either in the ligand binding site, the center of the active site (if there is no ligand present for instance in homology models) with a 13-15 Å box size. Halogen atoms (excluding the fluorine atoms) were considered as hydrogens acceptors. Default parameters were applied for the rest [286].

Docking studies. Based on the Glide module, docking studies were performed using the above-mentioned ligands and grids by choosing the XP (Extra Precision) method. The standard methodology applied in our laboratory is based on the XP method with default parameters, the van der Waals scaling factor was 1.0, and partial charge cutoff of 0.25 [287,288]. We also selected between 15 and 25 output docking poses depending on the ligand flexibility.

Binding pose selection. For the final docking pose selection, we commonly based our criteria on the best docking score, nevertheless visual inspection was crucial in all cases.

VS procedure. The HTVS was carried out using the Virtual Screening Workflow (VSW) of the Maestro Schrödinger Suite [288]. The set of ligands was previously filtered by physicochemical properties and compliance of Lipinski's rule of five. A series of three-filter docking (HTS High-trough docking, SP simple precision and XP extra precision) calculations were carried out on the three models based on the hierarchical screening approach of the software [287].

MD simulations. Apo non-bound structures obtained from Protein Preparation Wizard module were checked by the pdb4amber module in advance. Then, structures were parametrized using the tleap module from AMBER, using the popular protein force field ff14SB. In addition, for those complexes with ligands, geometry optimization and charge distributions were calculated for the selected compounds by quantum mechanics (RHF/3-21G*//RHF/6-31G**) with Gaussian 03 [289]. For ligands extracted from crystallized structures, only charge calculations were carried out without previous optimization. The general amber force field (GAFF) was used to parametrize ligands for AMBER14 [290]. All systems were hydrated by using boxes containing explicit TIP3P water molecules [291] in a truncated octahedron box of approximately 10.000 and also adding counter ions to maintain electro neutrality. Solvent molecules and counter ions were relaxed by energy minimization and then allowed to redistribute around the positional restrained structures during a 50 ps run at constant temperature (300 K) and pressure (1 atm). These initial harmonic restraints were gradually reduced in a series of progressive energy minimizations steps until they were completely removed. The resulting systems were heated again from 100 to 300 K during 20 ps and allowed to equilibrate in the absence of any restraints for 1.0 ns. Periodic boundary conditions (PBC) and the Particle Mesh Ewald (PME) methods were used to treat long-range electrostatic effects [292]. The SHAKE algorithm was applied to all bonds

and an integration step of 2.0 fs was used throughout [293]. After that, systems were ready to carry out 10-100 ns MD simulations without any restraints, generating snapshots (500-5.000) each 20 ps for further analysis. The trajectories were collected and the cpptraj module of AMBER14 was used to calculate the root mean square deviation (RMSD) of the atomic positions of the ligands and the most populated conformers [294]. The MM-ISMSA program was used to calculate the total binding energies for each complex giving the global energy binding and per-residue energy decomposition [295].

Mixed solvent: Equilibrated mixed solvent ethanol, isopropanol and acetonitrile at 20% v/v were obtained from the versatile pyMDMIX free software [296,297]. The initial coordinates of both systems where obtained with the Packmol free software [298,299]. The organic molecules were prepared using the same protocol and parametrization as the one applied for ligands. Additionally, for those molecules containing halogen atoms, RESP Partial Charge Assignments were also calculated [300,301]. Once the system was parametrized, MD simulations were carried out, monitoring density until a stable parameter along the simulation was obtained. In order to obtain equilibrated mixed solvent benzene 1M and chlorobenzene 1M and avoid phase separation, the Lennard-Jones attractive force parameters were reduced in half. In the case of halogenated solvents, the Extra Point (EP) parametrization methodology developed by *M. A. A. Ibrahim* was used. In this methodology the positive region centered on the halogen atom is represented by an extra-point (EP) of positive charge with no mass or van der Waals radii (**Fig. 2.46**). The EP parameterization was obtained at MP2/aug-cc-pVDZ level with a polarization function for the halogen atoms [302,303]. These calculations have not been performed in our laboratory due to their complexity, but in the Mahmoud A. A. Ibrahim´s CompChem lab (http://compchem.net). Non-bonded and bonded EP parameters were also provided by the CompChem lab.

Figure 2.46. *Screenshot of the EP-Dichlorobenzene molecule showing the corresponding atom type, notice the EP atoms present positive charge (last column) mimicking the halogen sigma hole. Molden stick representation of the EP-Dichlorobenzene, carbon atoms (brown), hydrogens (white), chlorines (green) and the EP dummy atoms (yellow).*

Pocket analysis. For the protein pocket cavity analysis, aligned PDBs were obtained from each trajectory and later analyzed by means of the Fpocket software based on the Voronoi tessellation algorithm [304]. The Mdpocket module was applied to obtain the corresponding volumes in each MD simulation step using as a pocket reference the first one obtained from the PDB code **3QCH** [305]. The volume of the pockets recorded along the simulation were represented as trend lines and violin plots using the RStudio software and the package ggplot2 (RStudio Team (2015) https://www.rstudio.com).

Free energy perturbation (FEP). The Amber16 CPU TI code implementation was used to calculate the free energies of solvation (ΔG_{solv}) for both ligand free and complex systems [306]. For proteins the Amber ff14SB force field, with theTIP3P model for water were used and ligands were also parametrized by quantum mechanics (RHF/3-21G*//RHF/6-31G**) with Gaussian 03 [289]**.** The solvated ligands and complexes were initially minimized for 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient. Solute atoms were restrained with a restraint weight of 5 kcal/mol. Heating until 300 K and pressuring up to 1.0 atm was maintained through 100ps. After that, MD simulations productions were run in the NPT ensemble with a Langevin thermostat set to 300 K with collision

frequency of 5.0 ps-1 and relaxation time of 2.0 ps. All calculations were carried out using the multiscale pmemd.MPI module from AMBER14 [307]. The Relative Free Energy of Binding (RFEB) calculations were run with 11 equal windows with a spaced lambda window ranging from 0 to 1. For the linear de– and re–charging transformations this will be ifsc=0. All integrations were carried out using a cubic spline over 11 lambda windows and the final ΔG of binding was calculated by: $\Delta G_{complex}$ - $\Delta G_{ligands}$.

2.12. Molecular Modelling of PTPRZ1 and WPD-loop study

As it was mentioned above, at the beginning of this project there were no available crystal structures of the intracellular phosphatase domain of PTPRZ1. For this reason, homology models were necessary in order to obtain a suitable 3D structure. The sequence of the PTPRZ1 phosphatase domain (PD1) was obtained (Uniprot - P23471) and analyzed using the web server BLAST (www.uniprot.org/blast) using only the human target database. Results summarized in **Figure 2.47** show that the nearest closed amino acid sequence is that of PTPRG, with a 69.3 % of identity. Following the ranges of amino acid sequence similarity with PTPRZ1, and taking the maximum result from the isoforms, we obtained also PTPRF (UniProtKB - P10586), PTPRD (UniProtKB - P23468) and PTPRA (UniProtKB - P18433) with 40.4%, 39.8% and 38.2% of sequence identity, respectively (**Fig. 2.47**). Further studies carried out using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo) show that after comparing the full sequences of both PTPRZ1 and PTPRG they presented 52.23% of identity, while comparing only the PD1 domain they show 73.61% of sequence identity. This fact highlights that the larger differences arise from the extracellular and transmembrane domains.

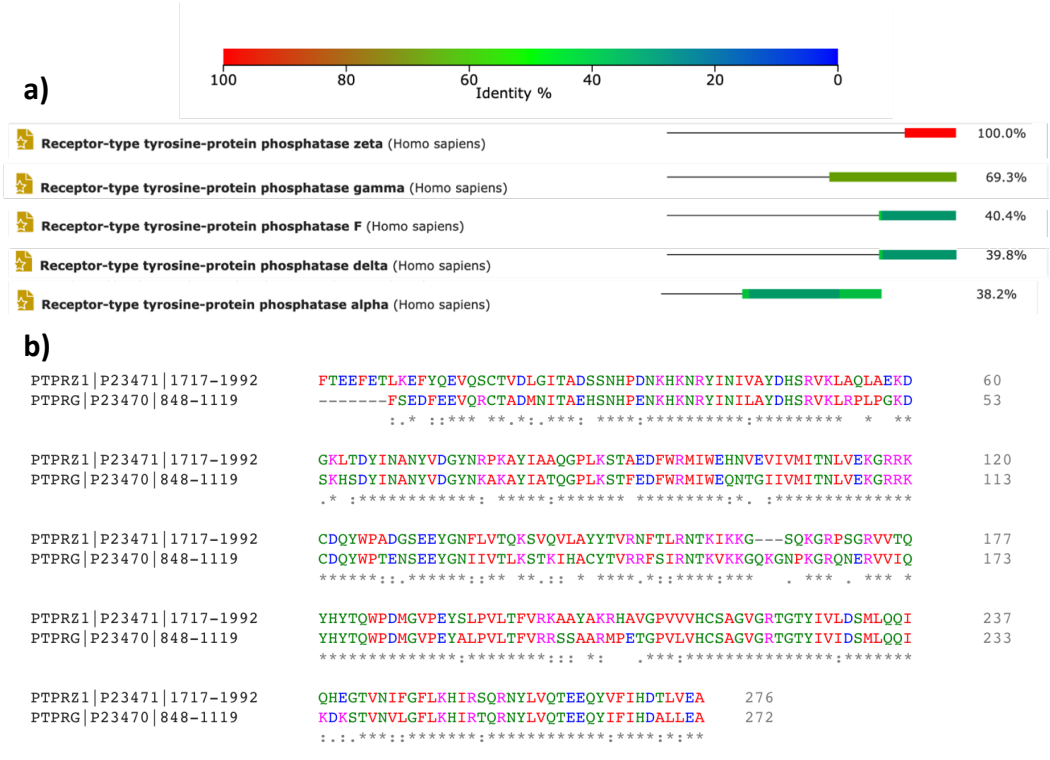


Figure 2.47. a) *Amino acid sequence of PTPRZ1 PD1 and the comparison with other PTPs carried out by BLAST web server.* ***b)*** *The identity matrix percentage are shown for each PTP. Above, the total amino acid comparison between PTPRZ1 vs PTPRG.*

Luckily, PTPRG (described in **section 2.5**) was the most similar to PTPRZ1 as it provided us with several high quality PD1 crystal structures in either the apo conformation or bound to different ligands. On the other hand, other phosphatases contained only the PD1-PD2 of PTPRF (PDB code **1LAR**). For this reason, as well as for the high sequence similarity with PTPRZ1, PTPRG was selected as template for the homology modeling procedures.

The PTPRZ1 homology models were built using the SWISS Model webserver given its reliability. Different crystal structures of PTPRG were used as templates. The results of these models are summarized in the next table as well as the RMS value of the models upon superimposition with their corresponding templates using PyMOL alignment (**Table 2.5**).

Table 2.5. *Obtained Models and crystal structures used as templates. Identity sequence, QMQE and QMEAN values obtained from Swissmodel. RMS value comes from the Cα alignment in Pymol.*

Model PTPRZ1	**Template**	**Seq. Identity (%)**	**GMQE**	**QMEAN**	**RMS (Å)**
PD1 Closed	3QCC_B	74.62	0.82	-1.55	0.064
PD1 Open	2PBN	75.0	0.82	-1.68	0.071
PD1 Superopen	3QCH	74.81	0.81	-1.29	0.157
PD1 Superopen	3QCJ	75.10	0.81	-2.69	0.174
PD1+PD2 Open	2NLK	74.04	0.76	-1.60	0.063
Model PTP1B	**Template**	**Seq.Identity (%)**	**GMQE**	**QMEAN**	**RMS (Å)**
PD1 Superopen	3QCH	37.11	0.73	-2.33	0.079

The GMQE (Global Model Quality Estimating) value is a quality estimate that mixes the target-template alignment characteristics with the template search technique. This value ranges from 0 to 1, being 1 when the modelled structure is the same as the template. The QMEAN value is a composite estimator based on distinct geometric characteristics and offers absolute quality estimates for both global (i.e. for the entire structure) and local (i.e. per residue) based on a single model. A QMEAN value under -4.0 is indicative of a low-quality model. The RMS compares only the alignment of

the positions of the Cα atoms, that is the backbone of the built models, so the nearest the value is to 0 the higher the similarity of the 3D structure. All the obtained data conclude that the built models are of high quality and reliable. The RMS values under 0.2Å imply the high similarity of the model and the corresponding template crystal structure. In order to validate the PTPRZ1 models in their three conformations we decided to compare their dynamic behavior with that of PTPRG. We expected the results to be similar for these two phosphatases due to their high similarity. However, we needed a "negative control", that is, a different phosphatase that was known to be unable to adopt the superopen conformation. For that purpose, we built a PTP1B model in the superopen conformation. This phosphatase cannot naturally adopt this WPD-loop superopen conformation due to the steric hindrance brought about by the presence of the bulky Phe191.

2.12.1. MD simulations of the closed, open and superopen conformations

This far, the built models have only been evaluated by statistical methods, but no experimental validation by methods as X-Ray or NMR was available. To assess the reliability of the models, MD simulations were carried out with the aim of comparing the dynamic behaviors of PTPRZ1 and PTPRG while using the dynamic behavior of PTP1B as a control. The structures issued from the MD simulations were compared with that of the templates.

To analyze the dynamic behaviors, we needed a reliable marker that could be used in all the MD simulations to compare models versus crystals. Previous studies performed by different research groups [170,308] on the WPD-loop flexibility of PTP1B established a methodology to assess the behavior of the loop by using structural descriptors that defined the conformational state of the WPD-loop.

As mentioned in **section 2.5.1**, the variety of PTPRG and PTPRZ1 structures can be classified by the WPD-loop conformation (**Fig. 2.48**):

i) The closed conformation, when WPD-loop closes the active site. Obtained when the phosphatase domain is bound to a substrate-like ligand or a competitive inhibitor in **3QCC** [309,310].

ii) The open conformation, when the protein is in the apo form **2NLK** and **5AWX** (PTPRG and PTPRZ1, respectively) [92,101,168].

iii) The superopen conformation, when the PD1 is bound to an inhibitor that alters the overall conformation of the WPD-loop **3QCJ** and relatives, and **5H08** (PTPRG and PTPRZ1, respectively) [102,309]. This altered superopen conformation was first described by *Sheriff et al* [309] and is induced by the binding of a hydrophobic benzyl-halide moiety of a ligand that displaces the side chain of Trp1026/Trp1899 thus forcing the opening of a new cryptic pocket.

Models of the closed, open and superopen conformations of PTPRZ1 and PTP1B, and their crystallographic counterparts of PTPRG (**Table 2.6**) were subjected to 100 ns MD simulations in order to assess the differences and similarities of the dynamic behaviors. The behavior of the WPD-loop was assessed in each system by comparing the three conformations using as reference the studies carried out by *Kamerlin et al.* [308]. The most important of these descriptors is the distance between the WPD-loop and the active site. This distance was measured by three points that have been indicated as the reaction path for the conformational change of the WPD-loop. In the following descriptors, amino acid numbers correspond to PTP1B sequence (**Fig. 2.49**) and the corresponding number for PTPRZ1 and PTPRG are also shown in **Table 2.7**:

i) Descriptor 1 (**D1**): the distance between the α carbons (CA) of Phe182 at the WPD-loop and Gly218 at the active site (Phe182-Gly218).

ii) Descriptor 2 (**D2**): the distance between the nitrogen epsilon (Nε) of the guanidinium of Arg221 at Helix7 and the carbon of the peptide bond (C) of Trp179 at the WPD-loop (Trp179-Arg221).

iii) Descriptor 3 (**D3**): the Cα-dihedral (CA) established between Thr178-Trp179-Pro180-Asp181 (PTP-1B-Cα) in the WPD-loop.

***Table 2.6**. Summary of the starting structures for the MD simulations. In parenthesis are the used templates.*

Conformation	PTPRZ1	PTPRG	PTP1B
Closed	Model (3QCC)	3QCC	1PTY
Open	Model (2PBN)	2PBN	2HNP
Superopen	Model (3QCH)	3QCH	Model (3QCH)

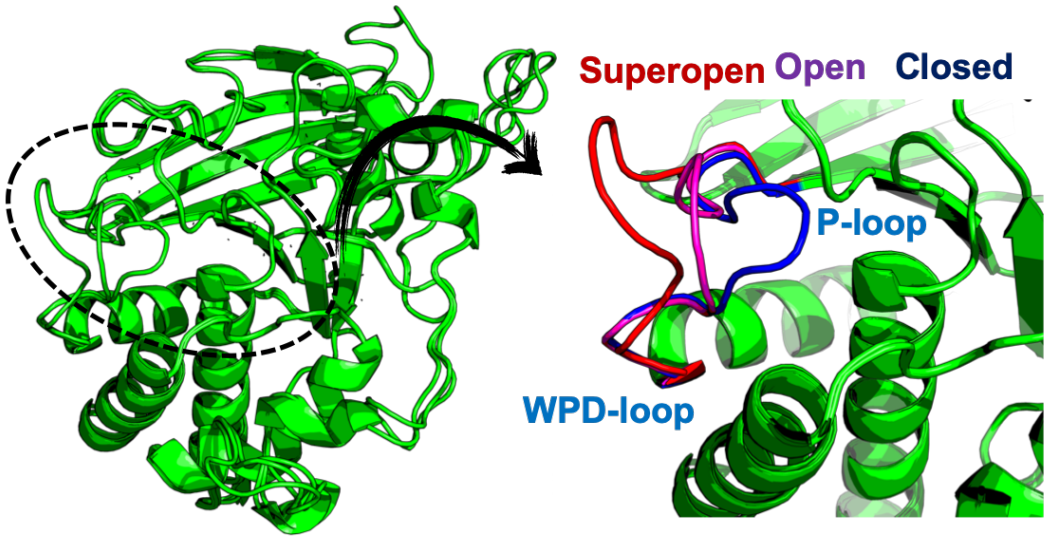


***Figure 2.48.** General structure of the PD1 of PTPRG and the different conformations of the WPD-loop (zoomed).*

Table 2.7. *Summary of the amino acids and corresponding atom types that represent the descriptor used.*

Descriptor	PTPRZ1	PTPRG	PTP1B
D1	M1902.CA - G1063.	M1029.CA - G1063.CA	F182.CA - G218.CA
D2	W1899.C - R1939.NE	W1026.C - R1066.NE	W179.C - R221.NE
D3	Q1898.CA W1899.CA P1900.CA D1901.CA	Q1025.CA W1026.CA P1027.CA D1026.CA	T178.CA W179.CA P180.CA D181.CA

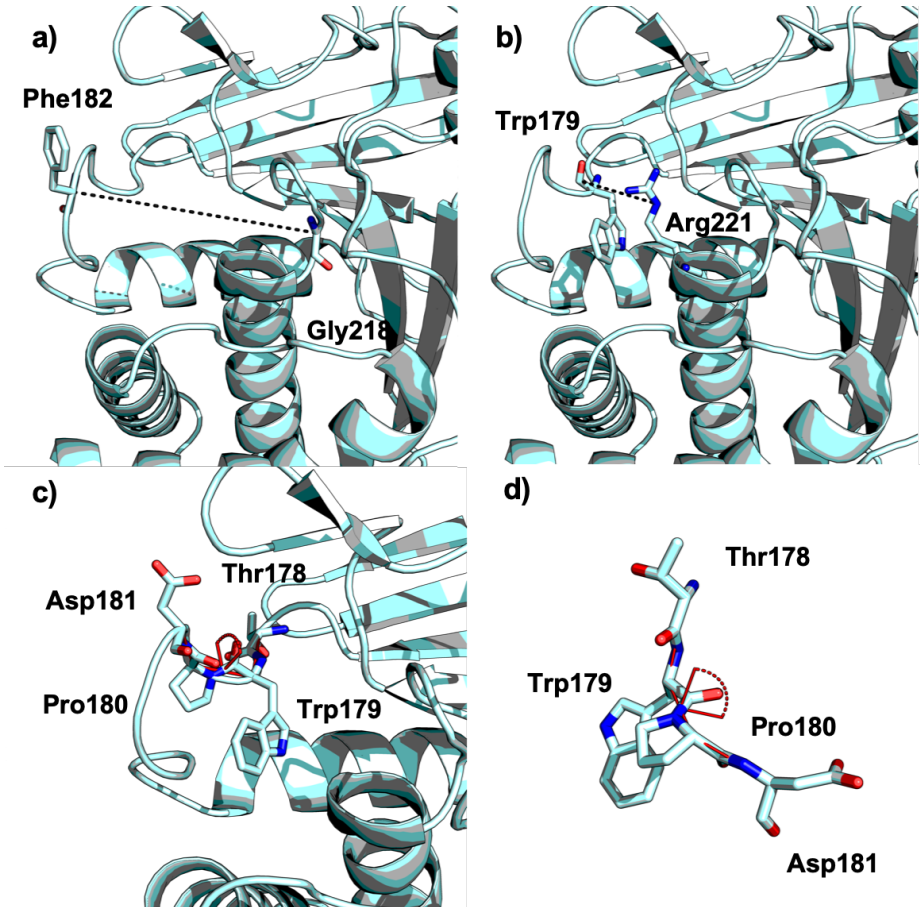


Figure 2.49. *PTP1B descriptors.* ***a)*** *D1 descriptor,* ***b)*** *D2 descriptor,* ***c)*** *and* ***d)*** *D3 descriptor. Distances are highlighted in black dashed lines and dihedral angles in red.*

The simulations starting from the closed conformations rendered results that were in agreement with previously reported data [308] in which the closed conformations tend to open (**Fig. 2.50**). Notably, there is a small opening of the WPD-loop, greater for PTP1B in which D1 varies 4 Å, whereas for PTPRZ1 and PTPRG this variation is less than 2 Å. This result suggests a same dynamic behavior for both related PTPRZ1 and PTPRG in contrast to the less similar PTP1B.

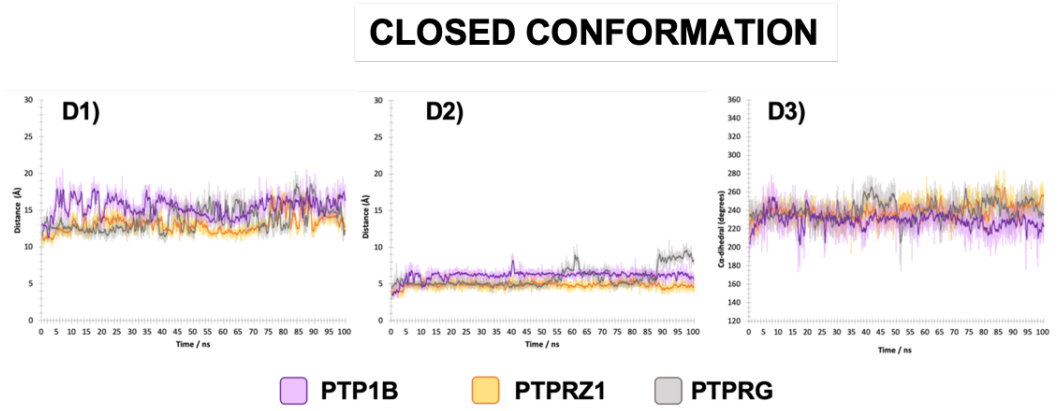


Figure 2.50. *Descriptors D1, D2 and D3 monitored during 100 ns MD simulations in PTP1B, PTPRZ1 and PTPRG, all three started from the closed conformation.*

The simulations starting from the open conformation showed no drastic movements when comparing the descriptors for each system, thus preserving the initial conformation throughout the simulation time with a behavior in all three enzymes showing a non-dire oscillating movement (**Fig. 2.51**). These results were also in agreement with published work [308].

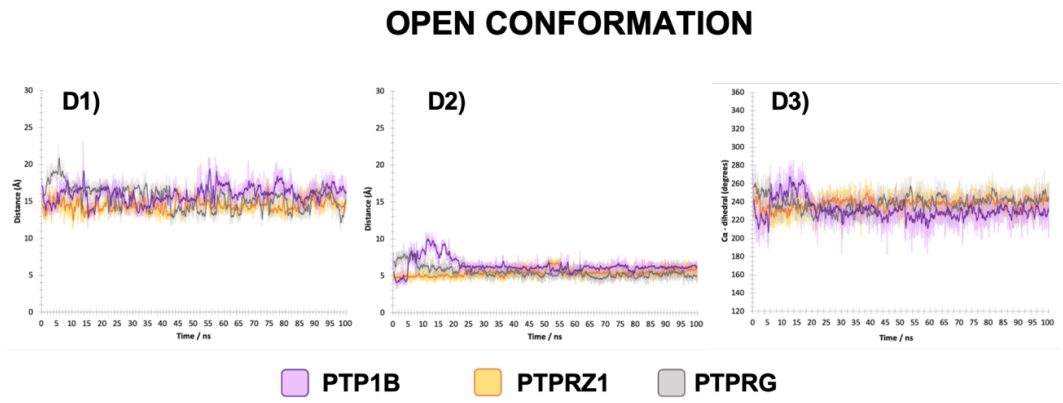


Figure 2.51. *Descriptors D1, D2 and D3 monitored during 100 ns MD simulations in PTP1B, PTPRZ1 and PTPRG, all three started from the open conformation.*

The simulations that started from the superopen conformation, however, showed mixed results, but an overall different WPD-loop behavior has been observed in this case when comparing PTP1B with PTPRZ1 and PTPRG. During the first 65 ns of the simulation time, a WPD-loop oscillation can be observed for each enzyme, with opening and closure movements over the active site in a similar way that the open conformation. After the initial 65 ns, the pocket closure of PTP1B started, changing the conformation of the WPD-loop more than 10 Å. In contrast, PTPRZ1 and PTPRG only changed 4 Å during all the simulation time (**Fig. 2.52**).

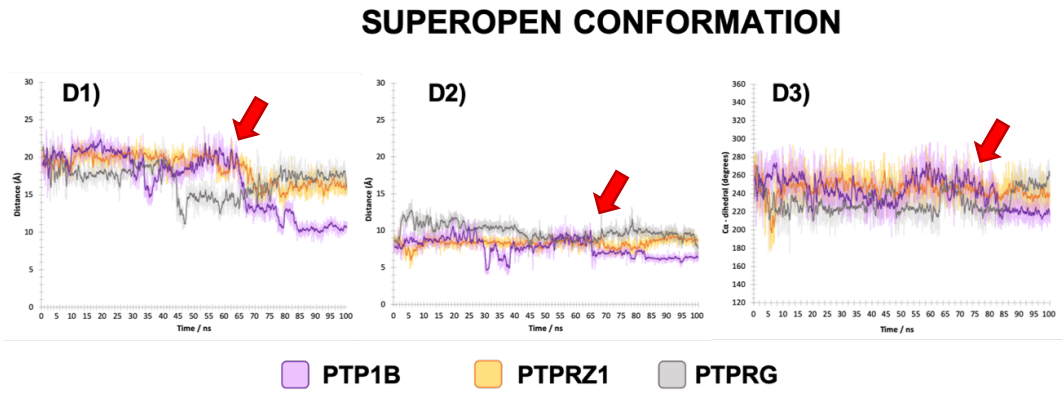


Figure 2.52. *Descriptors D1, D2 and D3 monitored during 100 ns MD simulations in PTP1B, PTPRZ1 and PTPRG, all three started from the superopen conformation. The red arrow shows the point where the pocket closure of PTP1B started.*

Visual inspection confirmed how PTPRZ1 and PTPRG remained in the

superopen conformation while PTP1B closed at the end of the simulation (**Fig. 2.53**). The monitoring of the D3 descriptor showed a similar behavior for PTPRZ1 and PTPRG, but both different from PTP1B. In the case of the latter, the value of the angle decreases from 260º to 215º, which is indicative of the closing of the WPD-loop over the active site. On the other hand, the PTPRZ1-Cα and PTPRG-Cα maintained values around 240º during the entire simulation (**Fig. 2.52**). The monitoring of the C_α dihedrals further confirm that the initial coordinates in the superopen conformation of PTPRZ1 and PTPRG remain stable during the simulation, whereas the conformation of the WPD-loop of PTP1B changed to a closed conformation at the end of the simulation. This conformation was the only one that presented notable differences in the three systems. To account for the change in pocket size, the solvent-accessible surface area (SASA) of the cryptic pocket of the proteins was also explored during the MD simulation [311]. The cryptic pocket was defined through the amino acids shown in **Table 2.8** and **Figure 2.54**.

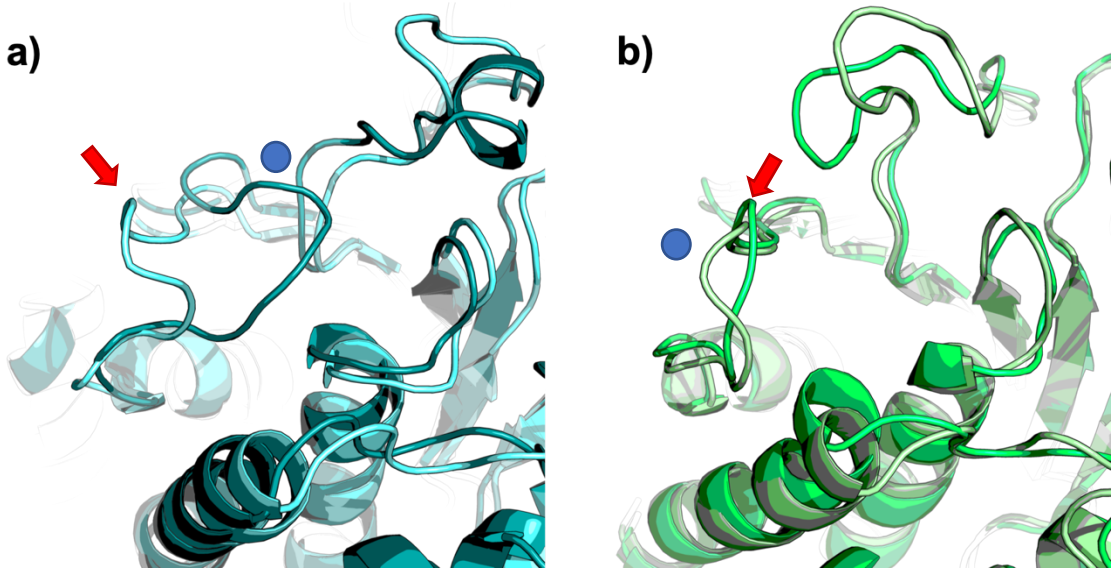


Figure 2.53. *The two most populated conformers (marked with a blue sphere) of* ***a)*** *PTP1B and* ***b)*** *PTPRZ1 obtained during the MD simulation superimposed to the initial conformation (marked with a red arrow). It can be appreciated that the closure of the WPD-loop towards the active site on PTP1B is complete compared to that of PTPRZ1 that remains stably opened.*

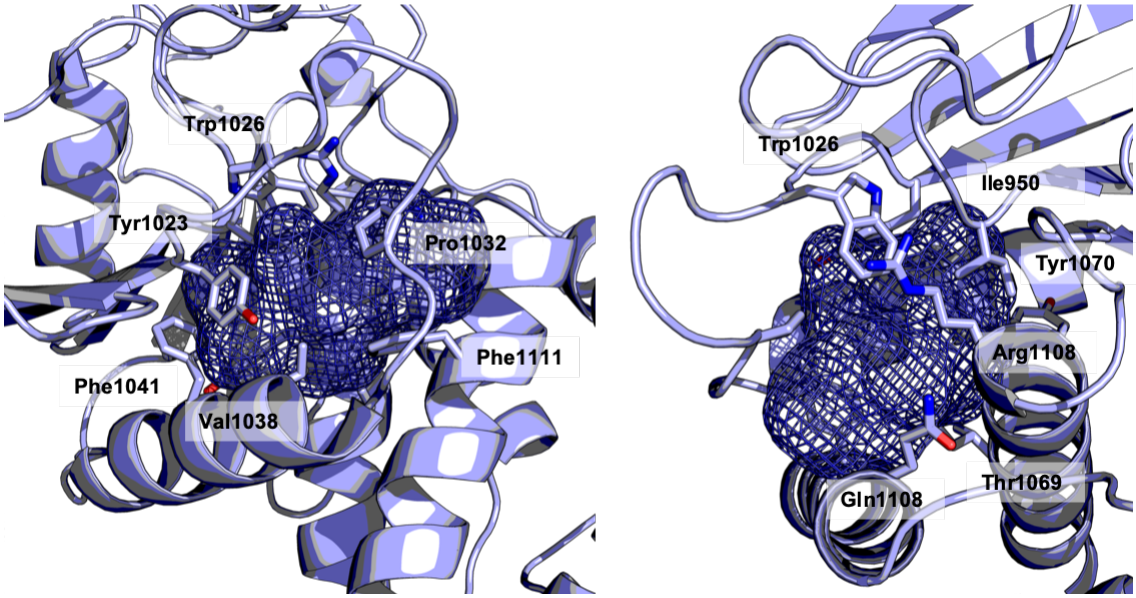


Figure 2.54. *Cryptic pocket definition for the SASA analysis of PTPRG front and side views (**3QCJ**).*

Table 2.8. *Cryptic pocket amino acids used for the SASA analysis*

System	Amino acids
PTPRZ1	Ile1826, Tyr1896, Trp1899, Pro1905, Val1911, Phe1914, Arg1939, Thr1942, Tyr1943, Gln1981 and Phe1984
PTPRG	Ile950, Tyr1023, Trp1026, Pro1032, Val1038, Phe1041, Arg1066, Thr1069, Tyr1070, Gln1108 and Phe1111
PTP1B	Leu110, Tyr176, Trp179, Pro185, Phe191, Phe194, Arg221, Thr224, Gln266 and Phe269

The results shown in **Figure 2.55** highlight how this pocket decreases in size in all of the proteins, but while PTPRZ1 and PTPRG preserve a cryptic pocket volume higher than 200 Å^3, in PTP1B the volume decreases notably after 70 ns of the simulation, resulting in a volume lower than 100 Å^3. These results indicate that although the cryptic pocket presents a tendency to close in the absence of an inhibitor, the superopen conformation in PTPRZ1 and PTPRG appears to be more stable than in PTP1B.

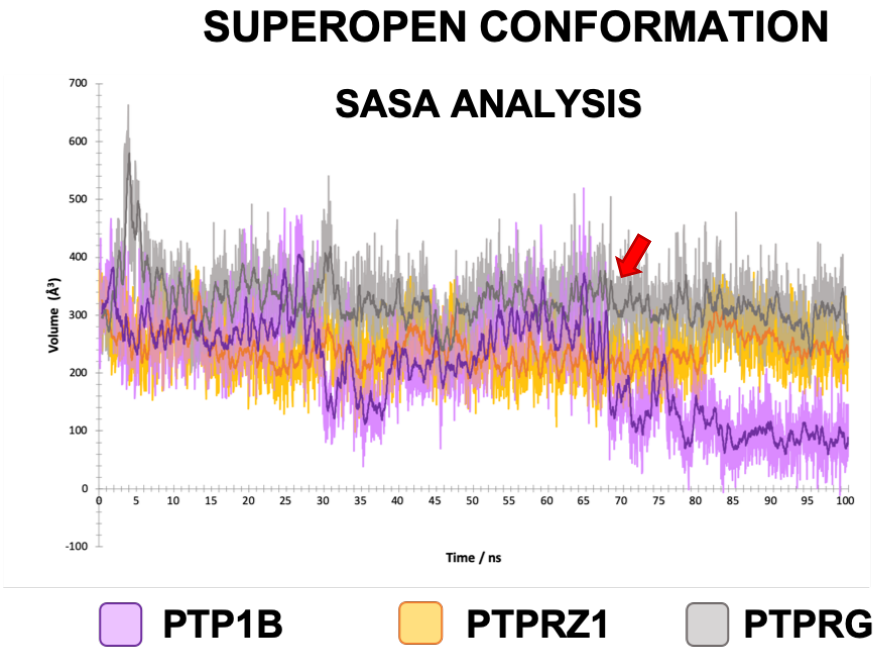


Figure 2.55. *SASA monitored during the 100 ns MD simulations of PTP1B, PTPRZ1 and PTPRG, all three started from the superopen conformation. The red arrow shows the point where PTP1B started to close the WPD-loop.*

These results strongly suggest that PTPRZ1 and PTPRG being members of the same subfamily, have the same dynamic behaviour, and can adopt the superopen conformation. Therefore, PTPRZ1 should be inhibited by structurally similar compounds as the ones described by Sheriff et al. [309]. This assumption was later demonstrated when the PTPRZ1 crystal structure **5H08** bound to **7WL** was published with the superopen conformation, the deep analysis of this structure and comparison with the other models will be discussed in **section 2.4.2**.

Moreover, the fact that PTP1B cannot stabilize the superopen conformation paves the way for selective PTPRZ1 vs PTP1B inhibition.

For further exploring the cryptic pocket in the superopen conformation and in an attempt to explore the slight differences between PTPRG and PTPRZ1, statistical analysis was carried out measuring the volume by means of the Fpocket software [304]. The module Mdpocket [305] was selected for this purpose due to its capacity to analyse simulated systems along the time. For this purpose, a reference pocket must be used to

compare this volume with the set of snapshots taken from the 100 ns MD simulation. We selected the pocket formed by compound **NX2** from the crystal structure **3QCH** (**Fig. 2.56**). The data obtained were analysed and plotted by means of RStudio software (RStudio Team-2015) and the package ggplot2 for the violin plots figures (**Fig. 2.57**). Results show the notable differences in the pocket volume between the three enzymes. For PTP1B the pocket in the closed and open conformations maintain almost a similar shape, signalling the opening of the closed conformation and the closure of the open, and both reaching a similar average volume around 120 $Å^3$. The same fact can be observed for both PTPRZ1 and PTPRG with the notorious difference that their average volume values are higher than for PTP1B, about 140 $Å^3$. In the superopen conformation, PTP1B also shows lower values (150 $Å^3$) compared to PTPRZ1 and PTPRG (200 $Å^3$). When comparing the superopen conformations of both PTPRZ1 and PTPRG, we noticed how PTPRG tends to present a bigger volume than PTPRZ1 with a higher average volume of 245 $Å^3$ in contrast to 210 $Å^3$. This fact will be analysed in more detail and discussed on the next section. These MD simulations and analysis allow us to confirm that PTPRZ1 and PTPRG present a similar behaviour in contrast to PTP1B.

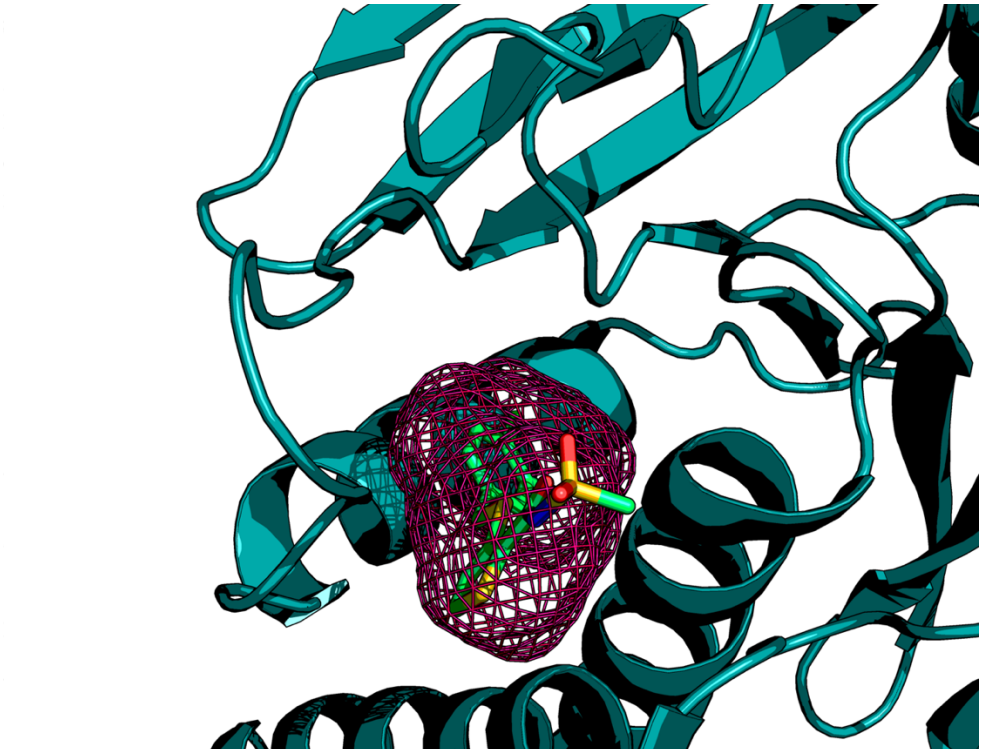


Figure 2.56. *Pocket representation in red lines used as reference for measuring the opening/closure of the WPD-loop in the PDB code* ***3QCH****. This pocket corresponds to the volume of the bound ligand* ***NX2****, with the exception of the methyl-sulfonamide moiety which is exposed to the solvent.*

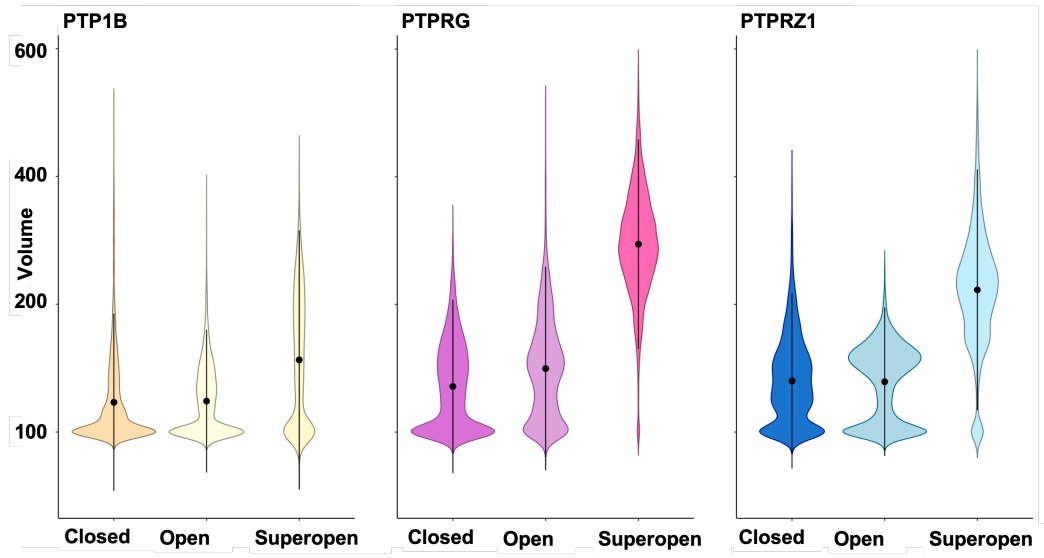


Figure 2.57. *Violin plots representing the volume in Å^3 for PTP1B, PTPRG and PTPRZ1 along the 100 ns MD simulation. Central dots represent the average value from all measures of the pocket while the trend lines are the deviation values.*

These differences in movement and stability, notably in the superopen conformations, are likely to be due to the different amino acid composition in the cryptic pocket of the three enzymes. While PTP1B presents phenylalanine at position 225, PTPRZ1 and PTPRG present a valine at the equivalent positions (positions 1911 and 1038, respectively). The side chain of Phe225 present in PTP1B interacts strongly with the sidechain of Trp179 located on the WPD-loop; the aromatic interactions established between these amino acids, further favour the closure of the loop towards the active site of the superopen model of PTP1B. The presence of Val1911 and Val1038 reduces the van der Waals component of the interaction with the side chain of the corresponding tryptophanes (Trp1899 and Trp1026) thus reducing the closing rate of the WPD-loop towards the active site of PTPRZ1 and PTPRG.

2.12.2. MD simulations in mixed organic solvents

Based on the above-mentioned results, we wanted to deepen on the origin of this slight but relevant differences in the pocket opening between PTPRZ1 and PTPRG. These slight differences can open up the possibility of achieving selectivity within these two phosphatases through the appropriate fine tuning of small molecule inhibitors.

As solvent mixtures in MD simulations have shown to be a predictive method to find cryptic pockets and also pocket exposure, we have applied this methodology in order to better understand the dynamics of the WPD-loop [312]. Many methods have been published using different organic solvents such as benzene, catechol, acetone, ethanol and acetonitrile, among others [297]. It is noteworthy that some protocols include the adjustment of the solvent-protein parameters in order to increase the hydrophobic interactions and facilitate the solvent access on cryptic pockets [313]. Nevertheless, in order to explore the WPD-loop and the superopen conformation in our systems, a standard protocol was selected for the mixed solvent MD simulations. In this case, a list of organic solvents was selected as probes based on their physicochemical properties and molecular volume

in order to provide a range of different dielectric constants, which are directly linked with the molecule polarity. The selected probe molecules were; benzene, chlorobenzene, dichlorobenzene, ethanol, acetonitrile and isopropanol, mixed in water at 1M concentration for benzene and chlorobenzene, and 20% v/v for the rest. For the halogenated solvents, we have used two different parametrization methods, one based only in the general amber force field; and, the other within the EP Parameterization (EP-chlorobenzene and EP-dichlorobenzene) developed by *M. A. A. Ibrahim.* [301]. A summary of the applied scheme in this study is depicted in the **Figure 2.58**.

Benzene 1M: benzene is on the top of common ring fragments in the FDA approved drugs [314], non-miscible in water and contains a non-polar structure, with the lowest dielectric constant (2.28).

Chlorobenzene 1M: chlorobenzene is a non-polar structure, non-miscible in water. The substitution with chlorine is common in drug design due its capability to create halogen bonds and change the dielectric constant of aromatic rings (5.69). Chlorobenzene has also been proved as a useful probe for the discovery of halogenated inhibitors [315].

Dichlorobenzene or orthodichlorobenzene (ODCB) 1M: This probe was selected in order to increase the number of halogen atoms with respect to chlorobenzene, and also because it is one of the main moieties which has been crystallized inside the majority of the superopen bound PTPRG crystal structures. Its dielectric constant is 9.93, which is slightly less polar than chlorobenzene.

Isopropanol 20% v/v: is miscible in water and presents a polar and non-polar substructure. In this case, the non-polar substructure presents two methyl groups and is therefore bigger than the one of acetonitrile and ethanol. Given the larger size of the non-polar substructure, the dielectric constant is correspondingly lower (18.3). Isopropanol has been recently proved to be a useful tool for the exploration of different holo conformations [316].

Ethanol 20% v/v: ethanol is a small molecule with 46.07 MW and dielectric constant of 24.6, miscible in water and has a polar and a non-polar structure able to establish hydrogen bonds and van der Waals interactions.

Acetonitrile 20% v/v: acetonitrile is miscible in water and is comparable to ethanol as it also has a polar and non-polar substructure and similar MW, although it has dielectric constant of 36.64, higher than ethanol. Acetonitrile has proven to be useful in the exploration of hydrophobic pockets due to its low molecular weight, high polarizability and its shovel-shaped methyl end, which makes it able to recognize hydrophobic pockets as well as to establish hydrogen bonds [317].

The apo structures of PTPRZ1 (PTPRZ1 model **2PBN**) and PTPRG (**2PBN**) were immersed in these solvent mixture boxes and 100ns MD simulations were carried out. The pocket opening was analyzed by means of Mdpocket and by visual inspection, following the protocol depicted below. As in the previous simulation of the apo structures, here the pocket obtained from the PDB code **3QCH** crystal structure was also used as reference for all simulated systems.

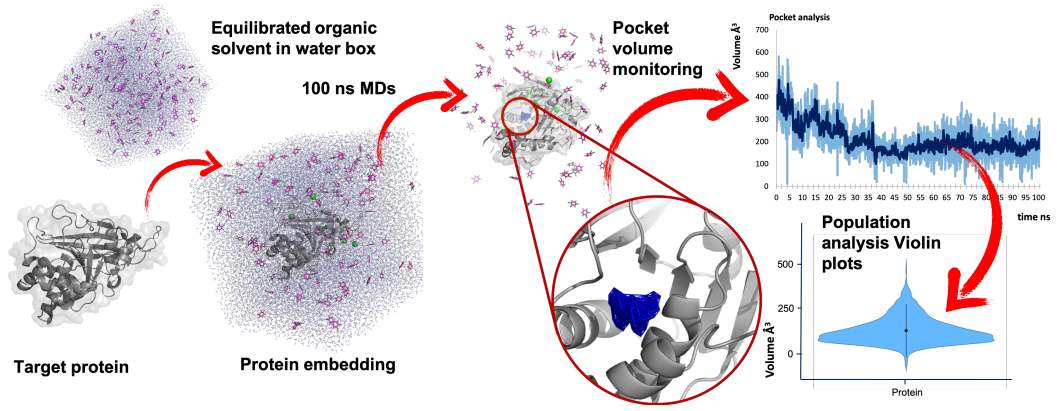


Figure 2.58. *Scheme of the protocol followed to explore the opening of the cryptic pocket by means of organic mixed solvent MD simulations.*

The Mdpocket results were plotted in violin plot format, to make comparison easier. The summarized results are represented below (**Fig. 2.59**). The majority of the plots present the typical violin shape, which means that major populations of volumes exist in the simulation; one of the populations

represents the closure of the pocket with values near to 0 Å^3; the other higher one comes from the pocket exposure in the open conformation. If a third higher population appears, it means either greater pocket exposure or opening.

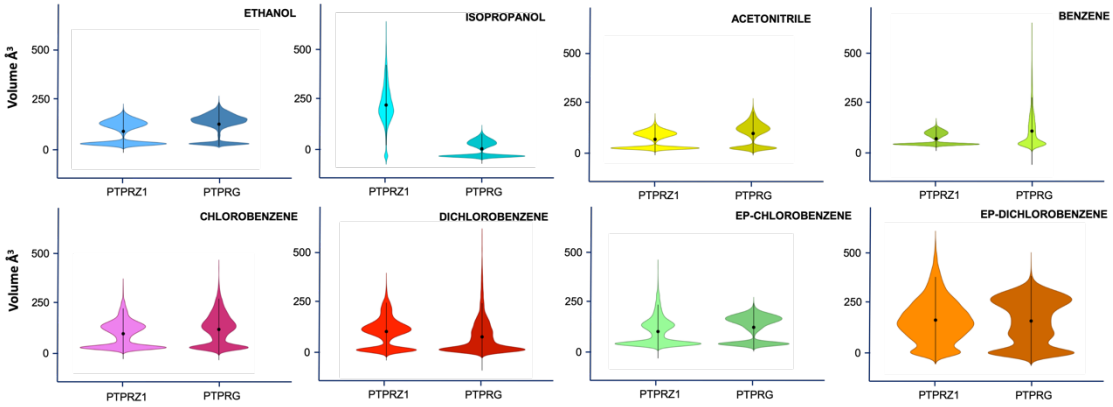


Figure 2.59. *Violin plots showing the volume obtained from each MD simulation using the corresponding solvent for PTPRZ1 and PTPRG.*

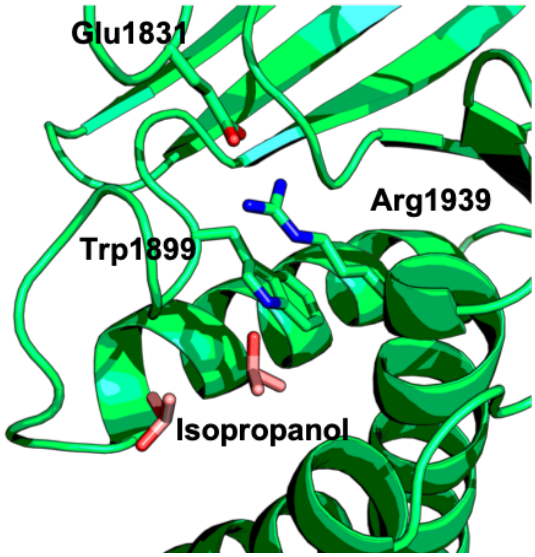


Figure 2.60. *Last snapshot from the MD simulation of PTPRZ1 with isopropanol molecules in pink color. Notice the entrance of the isopropanol molecules inside the pocket and the overall disruption of the WPD-loop.*

For the non-halogenated solvents, the pocket exposure was not observed, preserving the open conformation or either closing it. Remarkably, the MD simulation with isopropanol in PTPRZ1 presents a needle shaped violin plot, this fact is due to some of the isopropanol molecules interacting with the WPD-loop thus blocking its normal movement and giving as a result an average volume around 200 Å^3. Analyzing in deep this effect, we noticed that in any of the represented snap-shots isopropanol molecules were able to reach the described cryptic pocket, but this opening caused the disruption

of the overall WPD-loop (**Fig. 2.60**). The same behavior was found for the halogen probes, except for the dichlorobenzene with EP-parametrization. No pocket opening was observed during the whole 100ns for chlorobenzene, dichlorobenzene and the EP-parametrized chlorobenzene systems.

However, for the EP-dichlorobenzene simulation, dichlorobenzene molecules were able to reach the cryptic pocket in both enzymes giving as a result an average volume of 200 $Å^3$ as can be observed in the **Figure 2.59**. In this case, even though both proteins present the same mean value for the volume, there are notable differences on the final opening of the pocket. By looking at the shape of the plot, the major population for PTPRZ1 is located around 200 $Å^3$, while in the case of PTPRG this major population is around 250 $Å^3$. By visual inspection, we observed that the first dichlorobenzene molecule reaches the cryptic pocket at approximately 10 ns in PTPRZ1 and 20 ns in PTPRG giving a pocket volume of about 120 $Å^3$ at that point of the simulation (**Fig. 2.61** and **Fig. 2.62**). This molecule reaches the pocket through the arc of the WPD-loop by interacting with, Tyr1864, Tyr1896, Tyr1907 and Pro1919 in the case of PTPRZ1 and Cys988, Tyr1023, Tyr1034 and Pro1037 in the case of PTPRG. Then, a second dichlorobenzene molecule was able to reach the cryptic pocket, this time at 60 ns of the simulation for both enzymes yielding a pocket volume about 250 $Å^3$. By this double dichlorobenzene occupancy, at the end of the MD simulation, the final volume of the pockets was 350 and 400 $Å^3$ for PTPRZ1 and PTPRG, respectively (**Fig. 2.61** and **Fig. 2.62**).

The final structure issued from the MD simulation was analyzed and compared by visual inspection to both PTPRZ1 and PTPRG (**Fig. 2.63** and **Fig. 2.64**). In PTPRZ1 the interactions of the first EP-dichlorobenzene are made up by hydrophobic contacts with Ile1866, Tyr1896, Trp1899, Val1911, Phe1914, Arg1939, Thr1942, Tyr1943 and Phe1984 and the second EP-dichlorobenzene interacts mainly with Tyr1864, Tyr1896, Trp1899, Tyr1907, Pro1910, Val1911, Phe1914 (**Fig. 2.63**).

For PTPRG, the two EP-dichlorobenzene molecules interacted within the cryptic pocket in a similar mode compared to PTPRZ1. However, the molecules explored deeper in the case of PTPRG into the pocket, which resulted in a bigger pocket volume at the end of the simulation. In this case, the main hydrophobic interactions made up by the first EP-dichlorobenzene are Ile950, Tyr1023, Trp1026, Val1038, Phe1041, Arg1066, Thr1069, Tyr1070 and for the second are Tyr1023, Val1031, Pro1032, Tyr1034, Pro1037 and Phe1111 (**Fig. 2.64**). Interestingly, both EP-dichlorobenzene molecule probes, there is a halogen bond between the EP dummy atom and the hydroxyl group from the Tyr1023 highlighted in.

The analysis of these results and the structures obtained show how the different amino acid composition in the arc of the WPD-loop could be implicated in the opening of the pocket. The different composition at position Cys988/Tyr1864 gives more flexibility to the WPD-loop in PTPRG compared to PTPRZ1. In PTPRZ1, the Tyr1864 is close to Pro1919 and Tyr1896, which seem to block the entrance to the WPD-loop. On the other side, Cys988, located at this same position in PTPRG presents weaker interactions, making the entrance inside the cryptic pocket more reachable (**Fig. 2.65**). In PTPRZ1 the EP-dichlorobenzene molecules have to displace the Tyr1864 in order to penetrate inside the pocket, while in PTPRG due to the small steric hindrance of the Cys988 EP-dichlorobenzene could reach the pocket easily. The comparison of these forced-opened structures with the corresponding superopen crystal structures of PTPRZ1 (**5H08**) and PTPRG (**3QCJ**) shows how the mechanism of the WPD-loop opening process takes place. In the case of the solvent exploration, the molecules enter through the arc of the WPD-loop exploring in depth the cryptic pocket and opening the WPD-loop, but the Trp1026/Trp1899 remain stable in their position. On the contrary, in the crystal structures **5H08** and **3QCJ**, the corresponding ligands **NX4** and **7WL** occupy the position of the side chains of Trp1026/Trp1899 and displace them, thus forcing the WPD-loop to move away from the active site. This opening induced by this type of ligands cannot be done through the arc of the WPD-loop due to their higher volume

and surface. As both ligands **NX4** and **7WL** present a negatively charged sulfonamide group it can be proposed that in a first step, the negatively charged small molecule interacts with the positively charged Arg1066/Arg1939 and the active site, followed by the entrance of the benzyl halide moiety inside the cryptic pocket. By this halide entrance and occupancy, the side chain of the tryptophan is displaced forcing the superopen conformation. This process is further discussed in next section (see section **2.5.14.**)

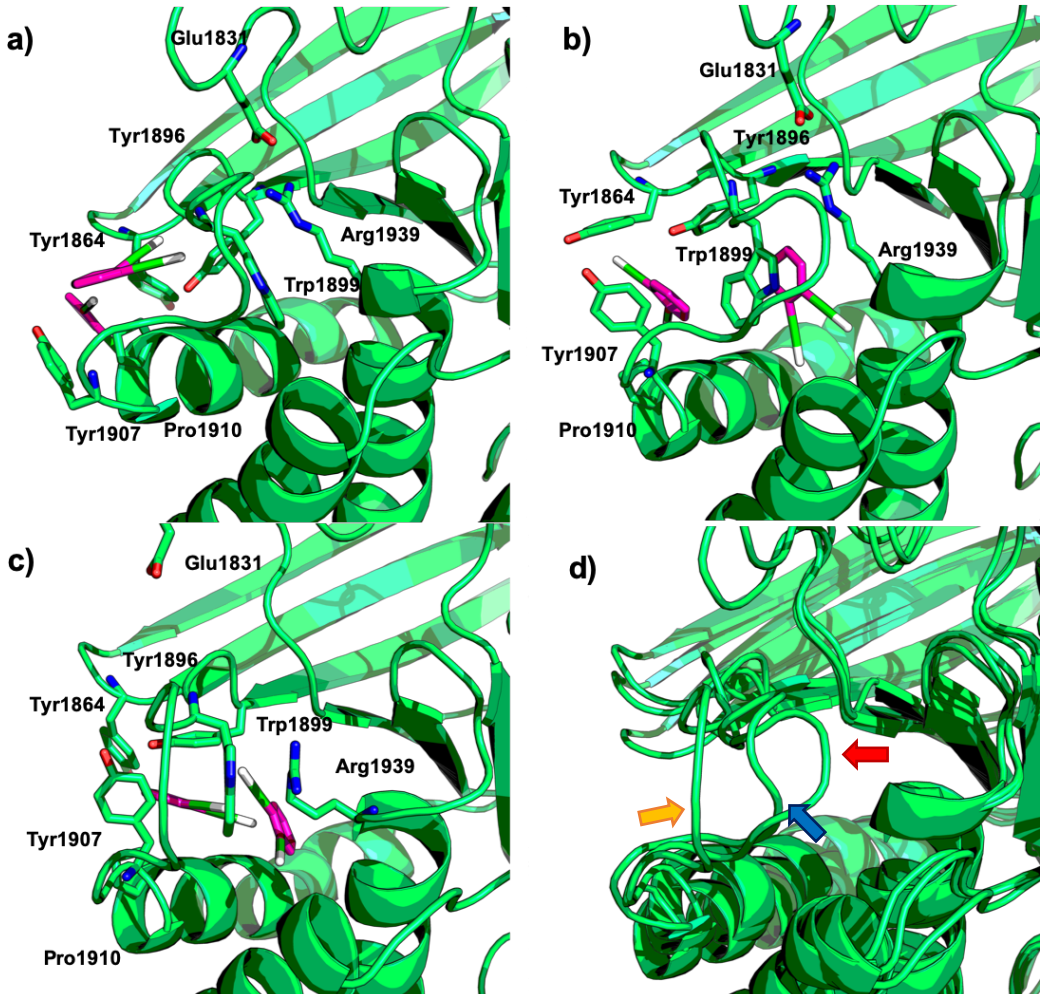


Figure 2.61. *The opening of the WPD-loop in PTPRZ1 through the MD simulations with EP-dichlorobenzene (pink).* ***a)*** *MD simulation at 20 ns with the first EP-Dichlorobenzene interacting within the arc of the WPD-loop.* ***b)*** *MD simulation at 60 ns showing the entrance of two EP-dichlorobenzene molecules.* ***c)*** *End of the simulation at 100 ns.* ***d)*** *Comparison of the WPD-loop at the same three snapshots: 20 ns (blue arrow), 60 ns (red arrow) and 100 ns (orange arrow).*

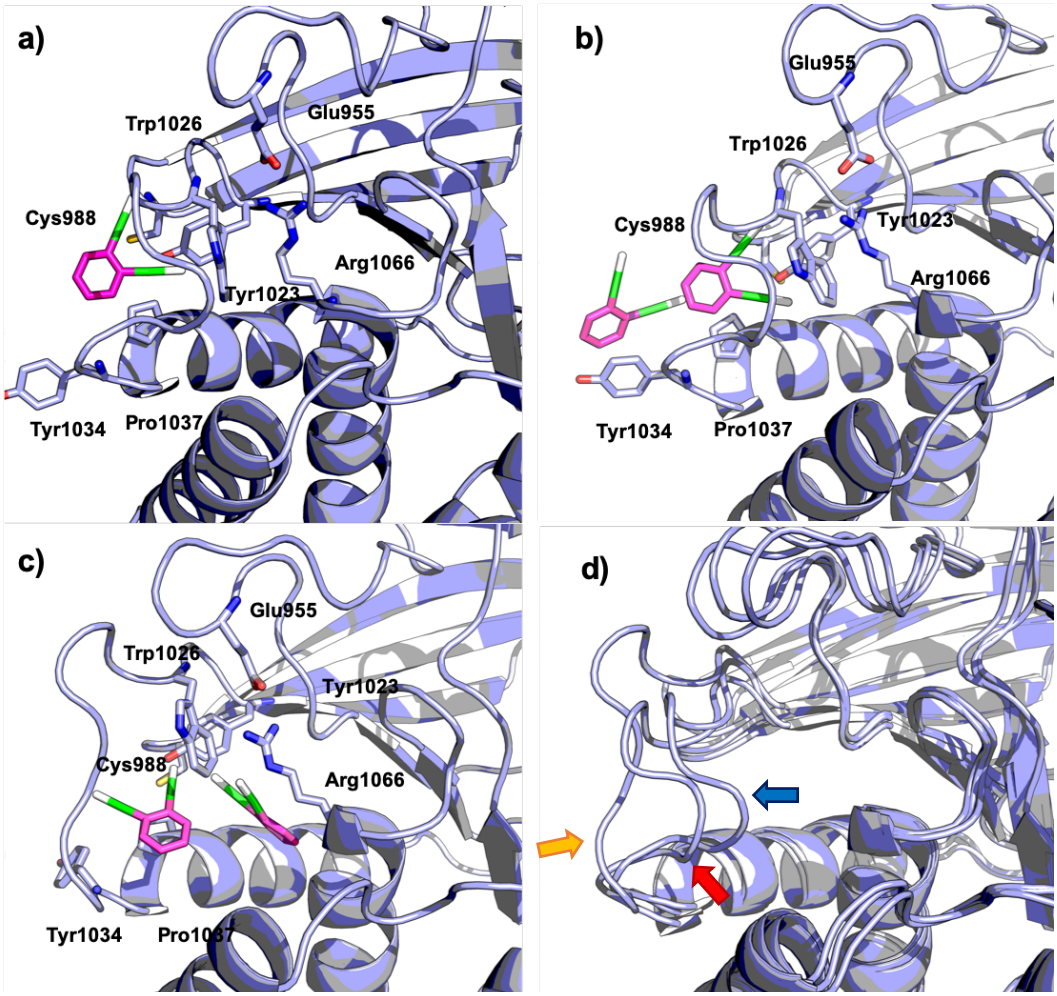


Figure 2.62. *The opening of the WPD-loop in PTPRG through the MD simulations with EP-dichlorobenzene (pink).* ***a)*** *MD simulation at 20 ns with the first EP-dichlorobenzene interacting within the arc of the WPD-loop.* ***b)*** *MD simulation at 60 ns showing the entrance of two EP-dichlorobenzene molecules.* ***c)*** *End of the simulation at 100 ns.* ***d)*** *Comparison of the WPD -loop state at the three snapshots 10 ns (blue arrow), 60 ns (red arrow) and 100 ns (orange arrow).*

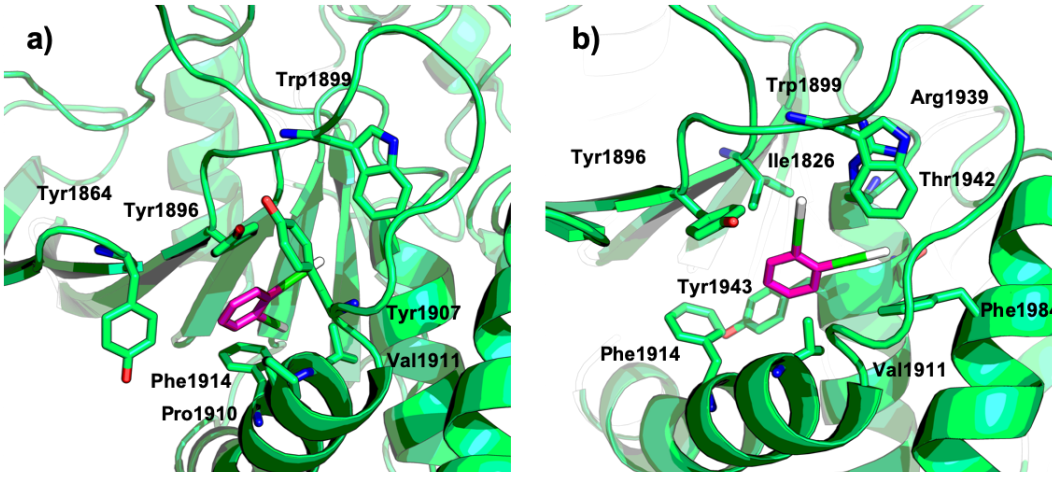


Figure 2.63. *Final snapshot of the 100 ns MD simulations in PTPRZ1 showing the interactions of the first EP-dichlorobenzene to reach the cryptic pocket* ***a)****, and the second EP-dichlorobenzene unit* ***b)****. Notice that a) is deeper inside the cryptic pocket than the b).*

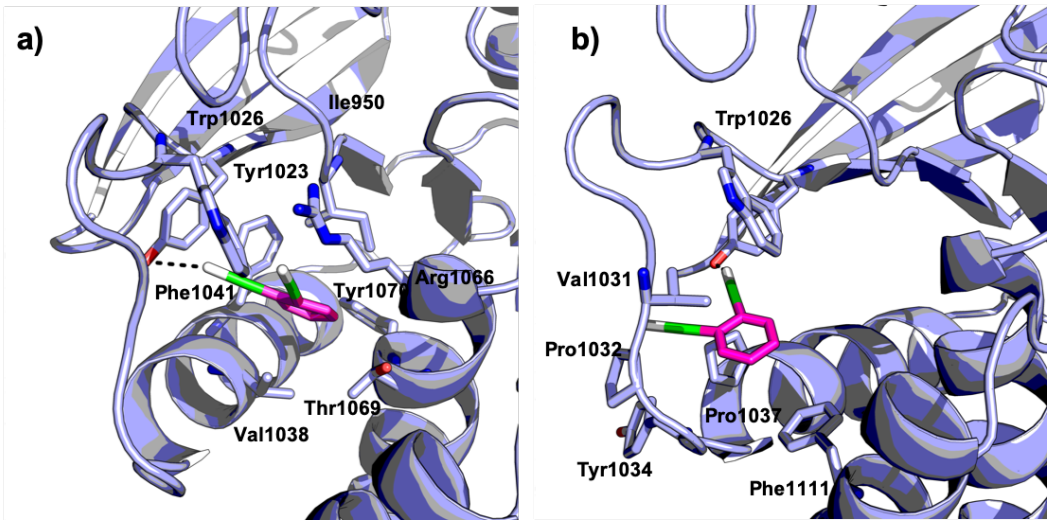


Figure 2.64. *Final snapshot of the 100 ns MD simulations in PTPRG showing the interactions of the first EP-dichlorobenzene to reach the cryptic pocket* ***a)*** *and the second EP-dichlorobenzene unit* ***b).*** *Notice that a) is deeper inside the cryptic pocket than the b). Black dashed lines show the halogen bonds.*

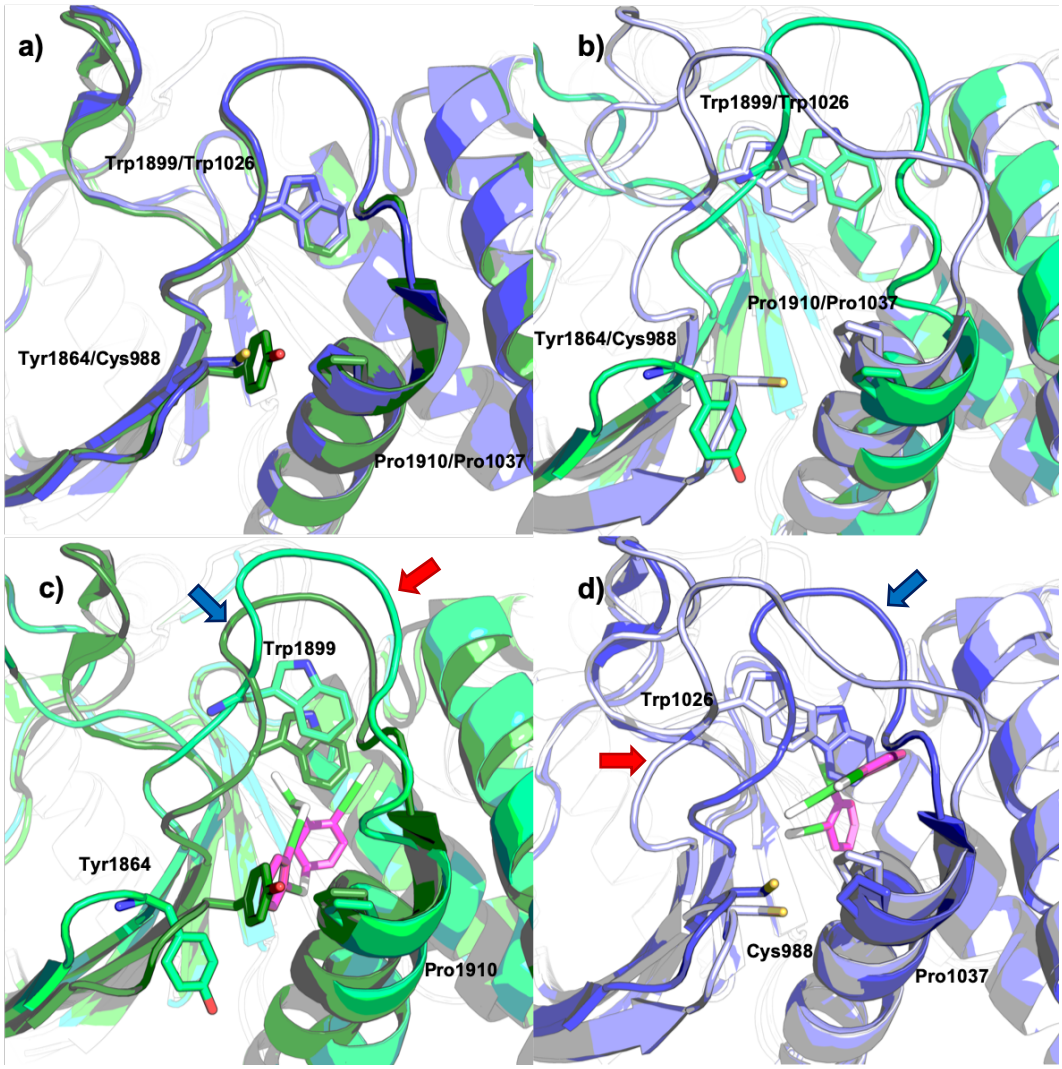


Figure 2.65. a) *Comparison of the WPD-loop conformation and amino acid interactions in the corresponding open crystal structures: PTPRZ1 (green,* ***5AWX****) and PTPRG (blue,* ***2PBN****).* ***b)*** *Final snapshots obtained through MD simulations with EP-dichlorobenzene (not shown) in PTPRZ1 (pale green) and PTPRG (cyan); notice the larger opening of the WPD-loop in PTPRG compared to PTPRZ1.* ***c)*** *The initial (blue arrow) and final (red arrow) snapshots in PTPRZ from the MD simulations with EP-dichlorobenzene, where the displacement of Tyr1864 after solvent exploration can be observed.* ***d)*** *The initial (blue arrow) and final (red arrow) snapshots in PTPRG from the MD simulations with EP-dichlorobenzene where the initial position of Cys988 remains stable during all the simulated time.*

2.12.3. The superopen conformation, a structural comparison

It wasn't until 2016 when the first crystal structure of the PD1 of PTPRZ1 in the open conformation was released in the PDB under code **5AWX** [101]. Our first task was to compare our built model with the crystal structure, and both were perfectly superimposable yielding an RMS of 0.342 Å upon alignment using PyMOL. Therefore, this fact, supported the previous homology modelling work for further docking studies and MD simulations along with the other two models, the closed and the superopen.

Focusing our attention on the novel PTPRZ1 crystal structure, it was obtained as a non-bound conformation even though the attempt from *Fujikawa et al.* [101] was to crystallize it in complex with ligand **SCB4380**. Citing their report "*In order to understand the structural basis of the inhibitory mechanism of SCB4380 in details, we challenged to determine the PTPRZ-SCB4380 complex structure. However, our attempts to soak **SCB4380** into the PTPRZ-D1 crystals, or to cocrystallize **SCB4380** with PTPRZ-D1 domain were not successful."*

A year later, in 2017 the first PTPRZ1 PD1 in complex with the small molecule **7WL** was released also by *this group.* This crystal structure was obtained in the superopen conformation as described previously for PTPRG. We compared this structure with our model, which was built, as described above, with code **3QCJ** as template due to the presence of a very similar small molecule **NX4** (**Fig. 2.66**). In this case, our PTPRZ1 model yielded an RMS value of 0.442 Å upon alignment using PyMOL with the recently reported crystal structure **5H08**. The overall general structure of the model and the crystal were superimposable, but slight differences came from the position of the WPD-loop, as the WPD-loop of the crystal structure **5H08** is slightly more opened than the one obtained in the built model. Focusing in this fact, we superimposed the PTPRZ1 model to the PTPRZ1 crystal structure in complex with ligand **7WL.** After visual inspection, we noticed that the ethoxy group in **7WL** causes a greater displacement in the Pro1905 and the backbone of the Glu1906 and Tyr 1907 (**Fig. 2.66**). This

small opening can be attributed to the fact that the arc of the WPD-loop in the crystal structure **5H08** is slightly more opened than the superopen crystal structures of PTPRG, and consequently than the PTPRZ1 superopen models built using those crystal structures as templates.

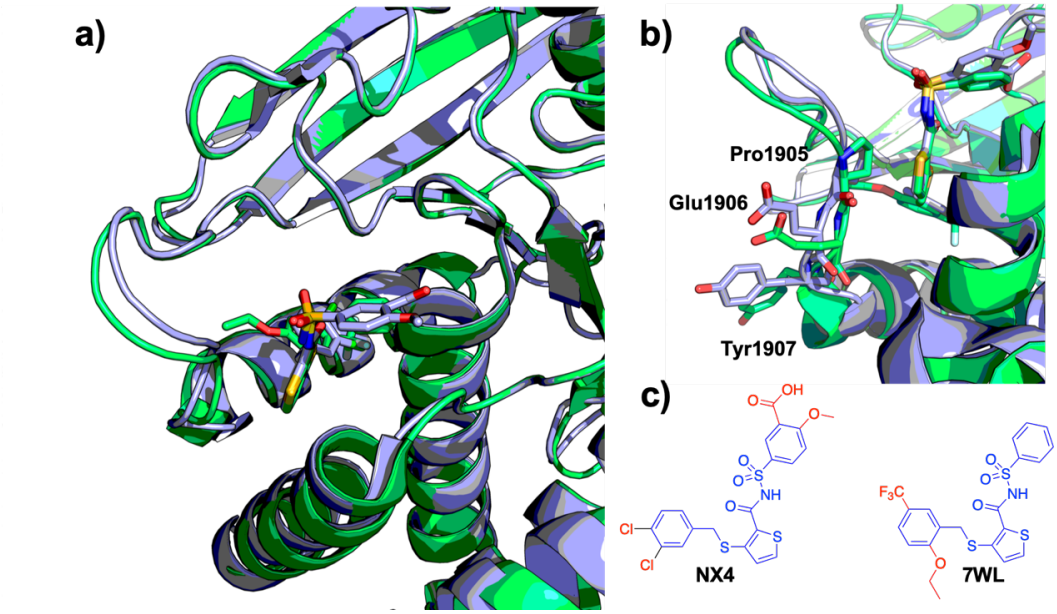


Figure 2.66. *Superimposition of crystal structures **5H08** (PTPRZ1, green) and **3QCJ** (PTPRG, blue). **a)** General position of compounds **NX4** and **7WL** inside PTPRG and PTPRZ1 respectively, showing the WPD-displacement. **b)** The most displaced amino acids after compound binding. **c)** Chemical structures of compounds **NX4** and **7WL**, showing in blue the common parts and red the different substituents.*

Additionally, the cryptic pocket volume of both crystal structures of PTPRZ1 and PTPRG (**5H08** and **3QCJ**, respectively) was also analysed by the Fpocket software (**Fig. 2.67**). The pocket volume in the superopen conformation for PTPRZ1 was 619 $Å^3$ and 456 $Å^3$ for PTPRG, showing the increases of the volume that this ethoxy group present in **7WL** can induce. All this information will be applied later on in order to rationalize a plausible and refined binding mode for the small molecules designed and synthesized in our research group.

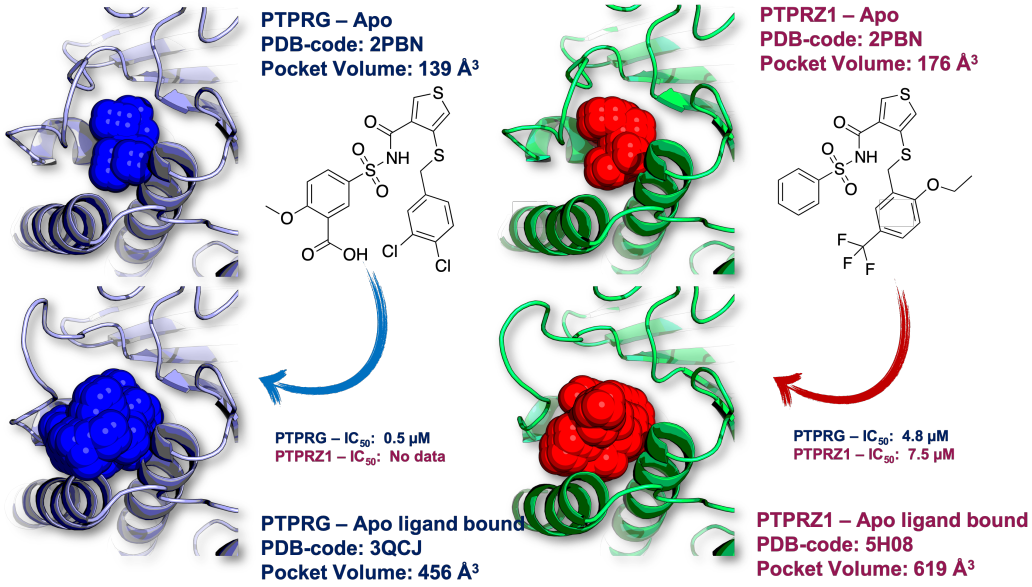


Figure 2.67. *Apo and apo ligand bound crystal structures of PTPRG (blue) and PTPRZ1 (green). The pocket volume in each structure has been calculated by the Fpocket software. The IC_{50} values from literature for* ***7WL*** *and* ***NX4*** *are also shown.*

2.12.4. The opening of the WPD-loop

Despite the fact that there is a common opening of the WPD-loop in the crystal structures bound to compounds **7WL** and **NX4** (**5H08** and **3QCJ**, respectively) and the structures issued from the MD simulation in solvent mixtures, there are significant differences in both systems. The greater difference arises from the position of the central Trp from the WPD-loop. The crystal structures present a conformational rearrangement of this amino acid not observed in the MD simulations carried out in solvent mixture with EP-dichlorobenzene. As reported by *Sheriff et al.*, once the inhibitor binds, there is a displacement of the Trp from the small hydrophobic pocket giving the novel "superopen" conformation. This small hydrophobic pocket that is partially occupied by the side chain of Trp in the open and closed conformations, is occupied by the bound ligands in the superopen conformation. After comparing the movement and displacement of the WPD-loop in all PTPRZ1 (PDB codes **5AWX** and **5H08**, and superopen conformation issued from MD simulation in organic solvent mixture) and PTPRG (PDB codes **2PBN** and **3QCJ**, and superopen conformation issued

from MD simulation in organic solvent mixture) systems, we noticed that organic solvent molecules fail to displace the Trp. In order to analyze the results, we have measured the dihedral angle in both types of systems using the D3 descriptor and established a new descriptor D4 that measures the dihedral made up by the atoms CA, CB, CG and CD1 of the Trp (**Fig. 2.68**).

As can be seen in **Table 2.9**, the D3 descriptor does not change significantly between the open and superopen structures for both enzymes PTPRZ1 and PTPRG. However, the D4 descriptor shows a difference of more than 180° between the open and superopen conformations of PTPRG crystal structures, but this displacement is less than 30° in the structures issued from the mixed solvent simulation. The results obtained for PTPRZ1 are in the same line, no apparent change can be seen in D3, but the D4 presents a similar behavior of the open conformation and the structure issued from the solvent mixture simulation. This is mainly due to the displacement brought about by the bound ligand in the superopen conformations, whereas the organic solvent is only able to occupy the pocket without displacing the Trp and therefore the conformation remains stable. In addition, the superopen conformations issued from the crystal structures present a bigger pocket volume compared to the one coming from the models obtained with the mixed solvent MD simulations. In the organic solvent models, there is only a slight pushing of the Trp, but the general conformation remains stable. Moreover, the superopen conformation from the crystal structures present a bigger pocket compared to the pocket obtained in the models through the MD simulations with solvents.

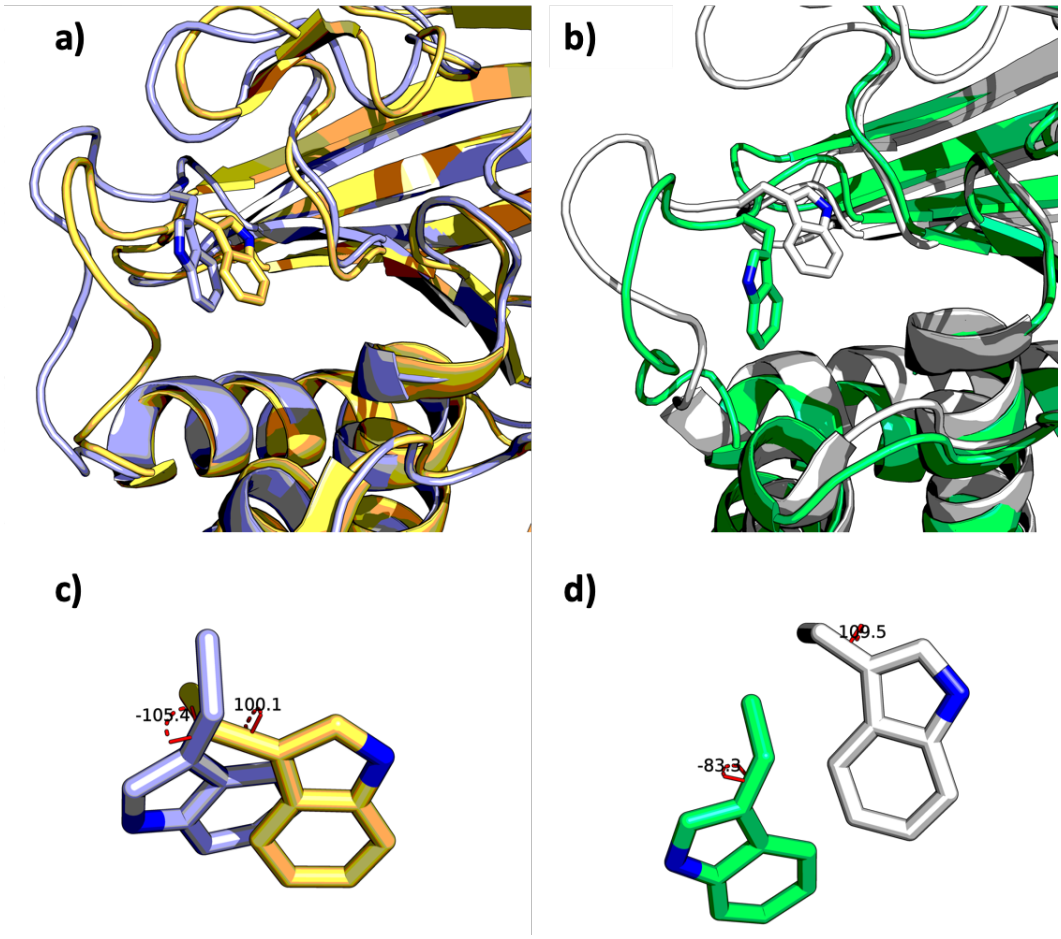


Figure 2.68. a-b) *Crystal structures **3QCJ** (cyan) and **5H08** (green) superimposed with the final snapshot from the EP-dichlorobenzene MD simulations (PTPRG gold and PTPRZ1 grey).* ***c- d)*** *Dihedral angle formed by CA, CB, CG, CD1 (D4 descriptor) in both structures of PTPRG structures and PTPRZ1 respectively. Red dashed lines show the D4 dihedral angle.*

Table 2.9. *D3 and D4 values (degrees) and the volume measured by Fpocket (Å^3) in the open and superopen conformations coming from the crystal structures and the MD simulations with organic solvents.*

System	D3 (°)	D4 (°)	Volume (Å^3)
PTPRG (2PBN open)	-110.3	-78.3	139
PTPRG (3QCJ superopen)	-116.3	100.1	456
PTPRG (model MD organic solvent)	-106.0	-105.3	502
PTPRZ1 (5AWX open)	-113.0	-82.6	176
PTPRZ1 (5H08 superopen)	-126.4	109.5	619
PTPRZ1 (model MD organic solvent)	-111.6	-83.3	427

2.13. Design and synthesis of PTPRZ1 inhibitors

As mentioned in **section 2.4.2.**, only one molecule (**TFMS-2)** described by *Cheng et al.* [103] showed moderate potency and selectivity against PTPRZ1. However, this molecule did not present the best ADME properties to reach the CNS (**Table 2.10**). For this reason, our research group selected this early hit as scaffold to carry out different chemical modifications in order to get a first hit with the desirable properties: small molecule that is able to cross the BBB and modulate the signaling pathway of PTN by inhibiting PTPRZ1. As summarized in **Figure 2.69**, the main modifications were the substitution in the aromatic rings present in **TFMS-2** by the exploration of the length of the linker between both aromatic rings, the substitution of the phenol oxygen by either amides, amines and triazoles of different length, and finally the substitution of the sulfoxides by their trifluorosulfide precursors. This last modification arises from the fact, that those sulfoxides present four hydrogen bond acceptors, making it very hydrophilic, with bad druglike properties non-desirable for crossing the BBB.

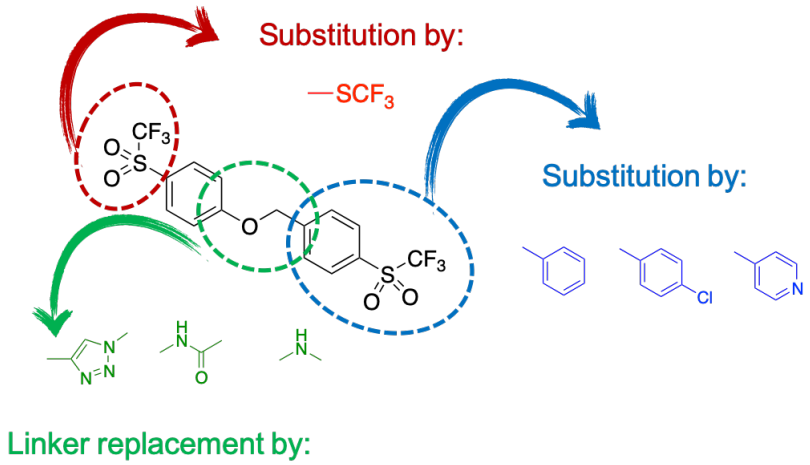


Figure 2.69. *Proposed chemical modifications for* ***TFMS-2.***

In adition, compound **NX4** was proposed as a novel PTPRPZ1 hit at the beginning of this project. Also, this compound was selected as a main scaffold in order to introduce modifications in synthetically accessible positions with the aim to explore the SAR in an attempt to increase the binding affinity. **NX4** was synthesized as reference compound for the in vitro

assays (in-house name **MPLNX4**); and, based on the experience in our research group, the trifluoromethyl-sulfur moiety was chosen as a main substituent due to its accessibility (**Fig. 2.70**). Finally, compound **7WL** was also synthetized (in-house name **MPL7WL**) in order to present another molecule to compare the biological results.

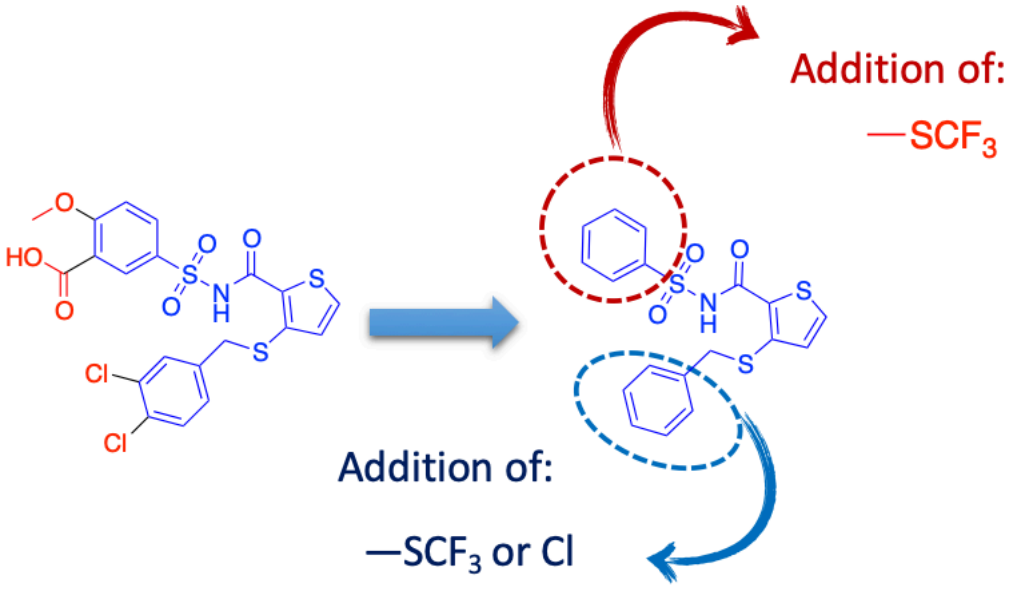


Figure 2.70. *Proposed chemical modifications for **NX4** in both terminal benzyl and phenyl groups.*

2.13.1. Synthesis

The chemical synthesis of 25 compounds was carried out by Dr. Miryam Pastor Fernández as part of her PhD Thesis. The synthesized compounds are summarized in **Table 2.10**. Some of the synthetic procedures and results are summarized in the published article "*Development of inhibitors of receptor protein tyrosine phosphatase β/ζ (PTPRZ1) as candidates for CNS disorders"* and all the details for the synthesis work is available in her PhD thesis [318]. On the other hand, all the synthetic work of the **NX4**-derivatives (called MPLs) has been carried out by Dr. Jose María Zapico and Dr. Mateusz Daśko. The synthethic route selected was the same to obtain compound **NX4**, changing the corresponding reactants to obtain the desired molecules [309].

2.13.2. Prediction of their drug-like properties by in silico tools

The SwissADME (www.swissadme.ch) web server was used to compute the physicochemical descriptors as well as ADME parameters of the compounds, that is: molecular weight (MW), Topological Polar Surface Area (TPSA) and WlogP (Wildman and Crippen Log P), the presence of PAINs, and the fulfillment of Lipinski's rules. The SwissADME tool stablishes the Lipinski's rule of five (Ro5) also known as the Pfizer's filter [319], as the fulfillment of the following premises predicts a molecule to be absorbed through the gastrointestinal tract: MW ≤ 500; MLOGP ≤ 4.15; N or O ≤ 10; and, NH or OH ≤ 5. The desirable molecule needs to accomplish at least three out of those four premises. Additionally, the predictive method known as Brain Or IntestinaL EstimateD permeation method (BOILED-Egg) was also used to predict if the set of compounds would be able to cross BBB [320]. Compounds that fall inside the egg are predicted to have good gastrointestinal absorption, and those inside the yolk are predicted to present also a good brain access. The results are collected in **Table 2.10**: For the first portfolio of designed compounds fulfilled the premises of Lipinski's rule, with the exception of **compound 13**, but due to its big volume it still was considered as a valid candidate. The BOILED-Egg results (**Fig. 2.71**) show that out of the 25 synthesized compounds, 8 molecules (**10c**, **12c**, **12d, 12e**, **16a**, **16c**, **17a** and **17c**) can be absorbed by the gastrointestinal tract, and out of those only compounds **10c**, **12c** and **12d** would be predicted to cross the BBB. In the case of the **NX4** and derivatives, compounds **MPL25**, **MPL27**, **MPL29**, **MPL30**, **MPL31** and **MPL47** violate the Lipinski rules due to their high MLOGP and MW. From this series, none of them presumable crossed the BBB according to the BOILED-Egg estimator. The PAIN prediction showed no potential PAINS neither other promiscuity alerts.

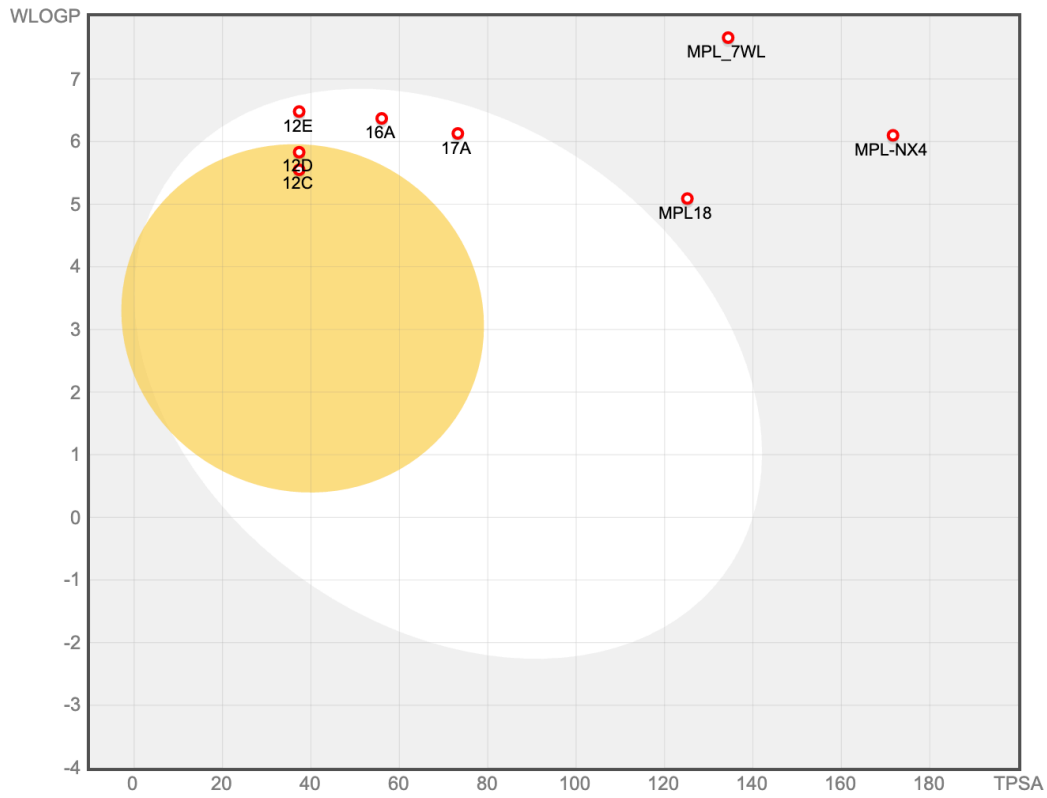
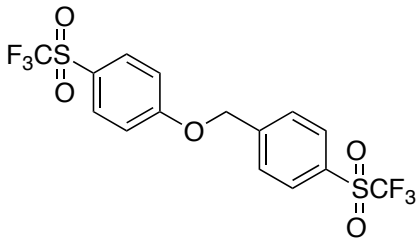
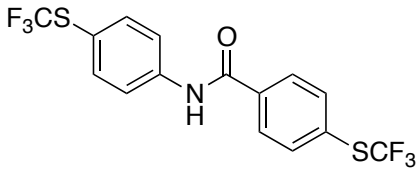
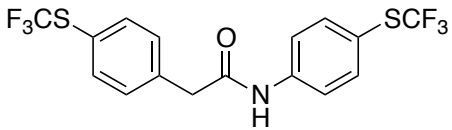
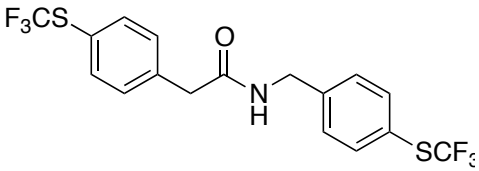
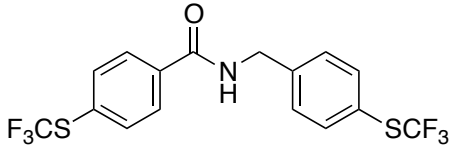
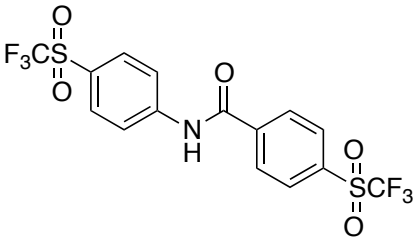
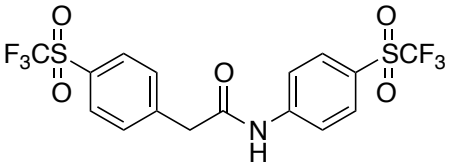
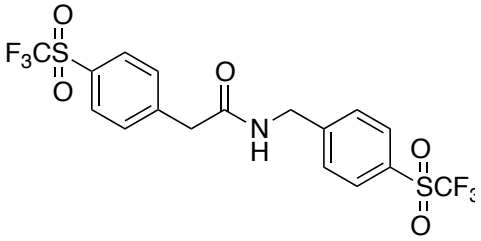
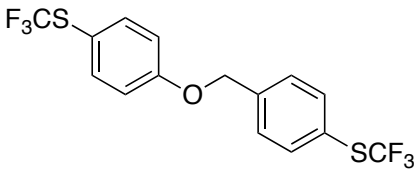
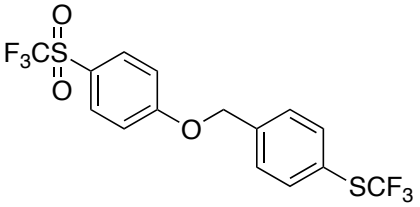
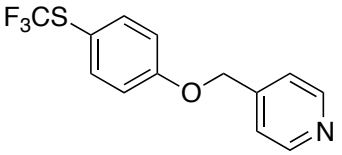


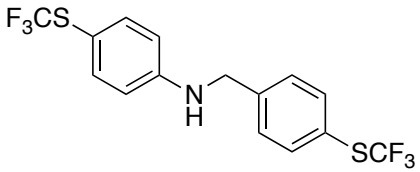
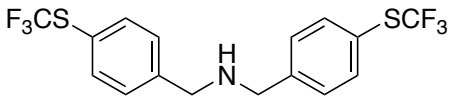
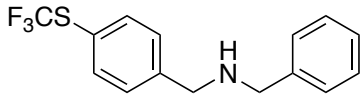
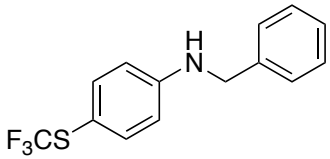
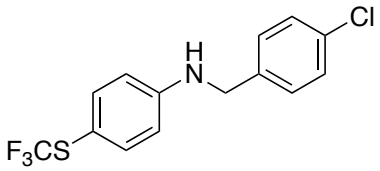
Figure 2.71. *Boiled-egg diagram obtained from our designed compounds (for clarity only few compounds were plotted).*

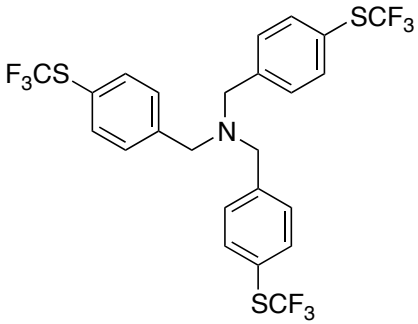
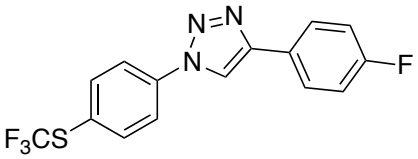
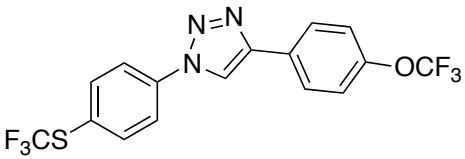
Table 2.10. *Chemical structure of the designed and synthesized compounds. Physicochemical properties such as MW, TPSA and MLogP are present. The Lipinski rules together with BBB permeant properties and PAINS are also shown. All data has been calculated by the SwissADME webserver.*

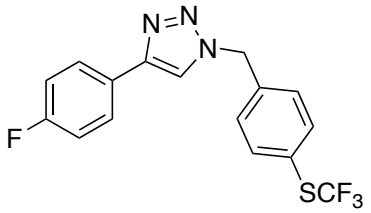
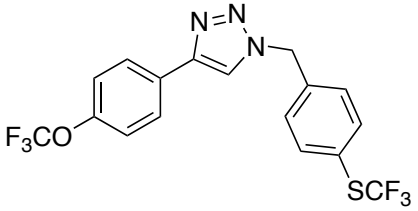
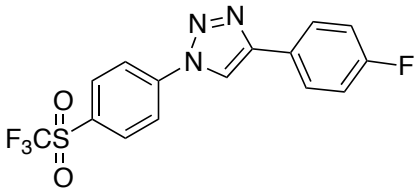
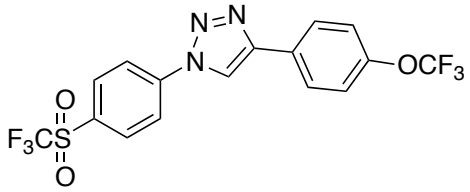
Compound	Structure	MW (g/mol)	TPSA (Å²)	MLopP	Lipinski	BBB permeant	PAINS
TFMS-2		448.36	94.27	2.96	Yes 0 violation	No	0 alert
4a		397.36	79.70	4.80	Yes; 1 violation: MLOGP>4.15	No	0 alert
4b		411.39	79.70	4.76	Yes; 1 violation: MLOGP>4.15	No	0 alert

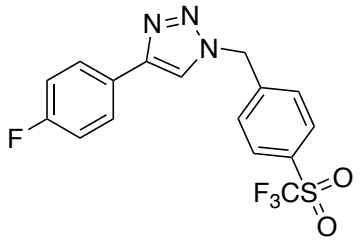
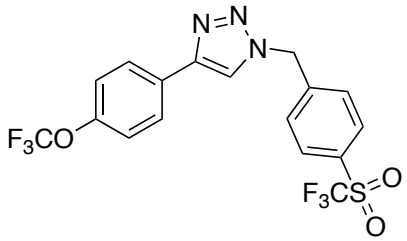
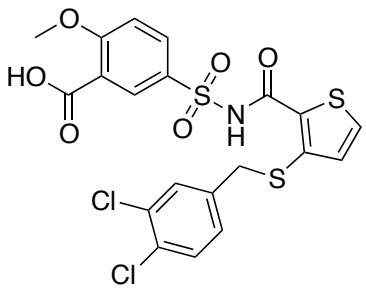
4c		425.41	79.70	4.72	Yes; 1 violation: MLOGP>4.15	No	0 alert
4d		411.39	79.70	4.76	Yes; 1 violation: MLOGP>4.15	No	0 alert
5a		461.36	114.14	2.75	Yes; 0 violation	No	0 alert
5b		475.38	114.14	2.71	Yes; 0 violation	No	0 alert

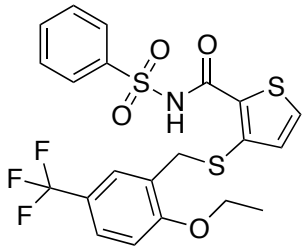
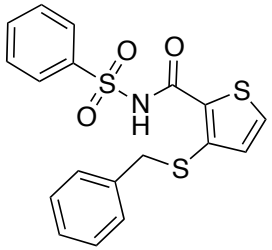
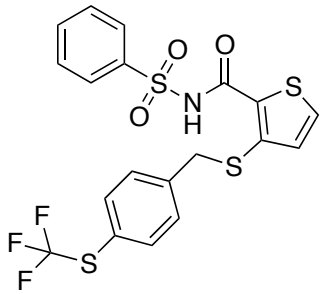
5c		489.41	114.14	2.67	Yes; 0 violation	No	0 alert
10a		384.36	59.83	5.13	Yes; 1 violation: MLOGP>4.15	No	0 alert
10b		416.36	77.05	3.98	Yes; 0 violation	No	0 alert
10c		285.28	47.42	2.69	Yes; 0 violation	Yes	0 alert

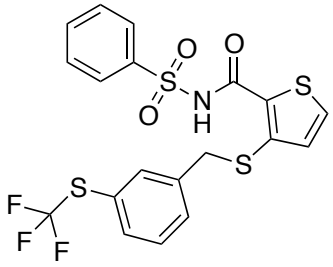
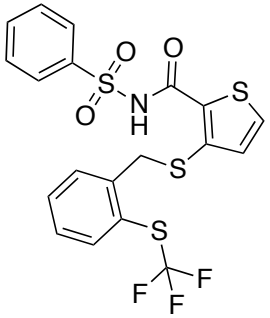
12a		383.37	62.63	5.13	Yes; 1 violation: MLOGP>4.15	No	0 alert
12b		397.40	62.63	5.10	Yes; 1 violation: MLOGP>4.15	No	0 alert
12c		297.34	37.33	4.24	Yes; 1 violation: MLOGP>4.15	Yes	0 alert
12d		297.34	37.33	4.26	Yes; 1 violation: MLOGP>4.15	Yes	0 alert
12e		317.76	37.33	4.77	Yes; 1 violation: MLOGP>4.15	No	0 alert

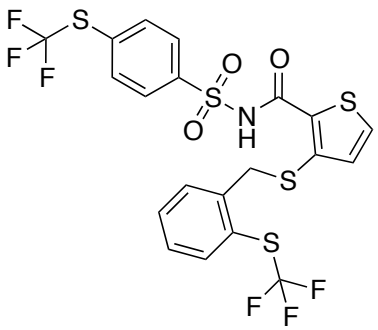
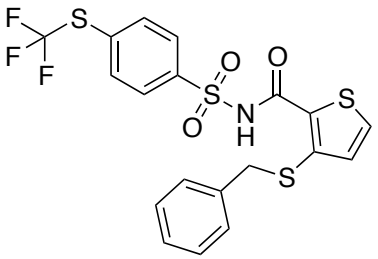
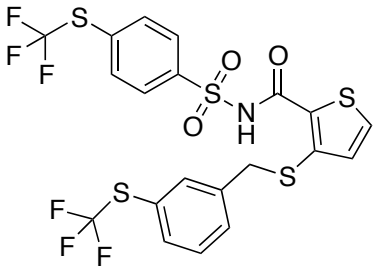
13		587.59	79.14	7.13	No; 2 violations: MW>500, MLOGP>4.15	No	0 alert
16a		339.31	56.01	4.12	Yes; 0 violations	No	0 alert
16b		405.32	65.24	3.74	Yes; 0 violations	No	0 alert

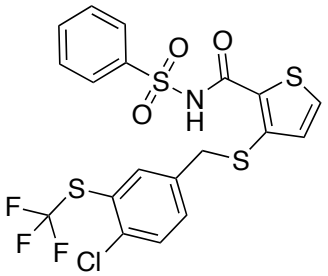
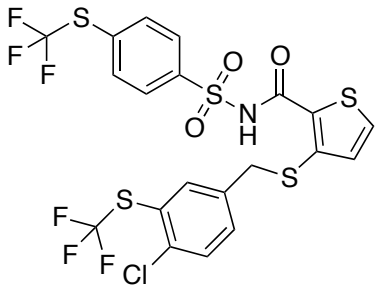
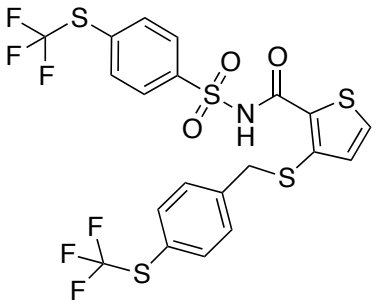
16c		353.34	56.01	4.09	Yes; 0 violations	No	0 alert
16d		419.34	65.24	3.70	Yes; 0 violations	No	0 alert
17a		371.31	73.23	3.11	Yes; 0 violations	No	0 alert
17b		437.32	82.46	2.76	Yes; 0 violations	No	0 alert

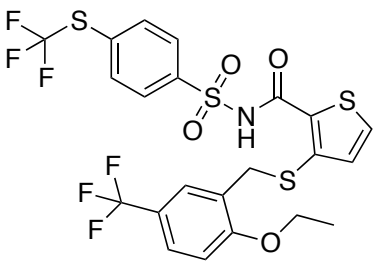
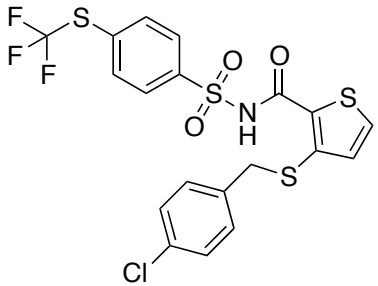
17c		385.34	73.23	3.08	Yes; 0 violations	No	0 alert
17d		451.34	82.46	2.72	Yes; 0 violations	No	0 alert
MPL-NX4		532.44	171.69	3.16	Yes; 1 violation: MW>500	NO	0 alert

MPL-7WL		501.56	134.39	3.60	Yes; 1 violation: MW>500	NO	0 alert
MPL18		389.51	125.16	2.88	Yes; 0 violation	NO	0 alert
MPL20		489.57	150.46	3.72	Yes; 0 violation	No	0 alert

MPL22		489.57	150.46	3.72	Yes; 0 violation	NO	0 alert
MPL24		489.57	150.46	3.72	Yes; 0 violation	NO	0 alert

MPL25		589.64	175.76	4.54	No; 2 violations: MW>500, MLOGP>4.15	NO	0 alert
MPL26		489.57	150.46	3.72	Yes; 0 violation	NO	0 alert
MPL27		589.64	175.76	4.54	No; 2 violations: MW>500, MLOGP>4.15	NO	0 alert

MPL29		524.02	150.46	4.21	No; 2 violations: MW>500, MLOGP>4.15	No	0 alert
MPL30		624.08	175.76	5.0	No; 2 violations: MW>500, MLOGP>4.15	NO	0 alert
MPL31		589.64	175.76	4.54	No; 2 violations: MW>500, MLOGP>4.15	NO	0 alert

MPL36		601.63	159.69	4.12	Yes; 1 violation: MW>500	NO	0 alert
MPL47		524.02	150.46	4.21	No; 2 violations: MW>500, MLOGP>4.15	NO	0 alert

2.14. Biological evaluation

The workflow followed in the biological evaluation was carried out by Rosalía Fernández-Calle and Marta Vicente, as part of their PhD Thesis projects in the area of pharmacology of this University, and under the supervision of Prof. Gonzalo Herradón.

2.14.1. In vitro evaluation

As a preliminary screening, all 25 compounds were tested using the general Phospho-Tyr assay. The total levels of phosphorylation of tyrosine residues in HeLa cells were treated with different compound concentrations (0.1, 1.0 and 10.0 μM) and the levels of phosphorylated residues were measured. Results are summarized in **Table 2.10**. Compounds that induced at least a 10% increase of Phospho-Tyr levels at any of the used concentrations were selected for further evaluation of the phosphatase activity inhibition. This preliminary assay was used as a pre-filter for the selection of those compounds that in a non-selective way blocked the cell phosphatase activity. One of the main limitations of this filter is that HeLa cells could compensate the phosphatase activity produced by the inhibition of PTPRZ1 by those compounds. Therefore, compounds with moderate activity against PTPRZ1 could be discarded as false negatives. On the other hand, promiscuous phosphatase inhibitors could be selected too. Despite these drawbacks and due to the limitations of the project to carry out direct in vitro assays in PTPRZ1, this process was chosen as a prefilter. From the total of the synthesized compounds, eight were discarded as they were Phospho-Tyr negative. It is interesting to note that our reference **TFMS-2** was Phospho-Tyr negative, this fact can be explained because of its moderate activity against PTPRZ1 that could be counterbalanced by other phosphatases. From the 17 Phospho-Tyr positives, only 12 were selected for the next in vitro assay; this time a single concentration (1.0 μM) of each of these compounds was tested directly in PTPRZ1. Compounds that at least induced a 10% inhibition of the phosphatase activity were finally

selected to obtain the IC_{50} value. Finally, compounds **4c**, **4d**, **5b**, **10a**, **12b** and **12d** achieved this criterion and were evaluated to obtain the concentration-response in PTPRZ1 and calculate the half-maximal inhibitory concentration (IC_{50}) value (**Table 2.11**). For compounds **16a-b**, **17a** and **17d**, only the two preliminary assays were performed as those bearing a triazole as a linker unit were not initially selected for further assays, nevertheless they are also currently under study. Among these compounds, the most active ones were **10a**, **12b** and **4c** with IC_{50} values ranging from 0.1 to 0.8 μM. In order to check the selectivity towards other phosphatases, PTP1B was selected as an off target for these compounds and the IC_{50} values were also obtained. In this assay, only **12b** and **12d** presented moderate activity against PTP1B, whereas compounds **4c** and **10a** proved to be extremely selective as no inhibitory activity was detected. From the **MPL4** series only compounds **MPLNX4**, **MPL18**, **MPL20**, **MPL24**, **MPL30** and **MPL31** were active in the Phospho-Tyr assay, interestingly compound **MPL7WL** was not active.

Table 2.11. *Results of the Phospho-Tyr that induced at least a 10% increase of Phospho-Tyr levels and PTPRZ1 inhibition test at least 10% of the activity at 1.0 μM. Calculated IC_{50} values (μM) of PTPRZ1 and PTP1B inhibition. ND = Non-Detected*

Compound	Phospho-Tyr	PTPRZ1 Inhibition (1.0 μM)	PTPRZ1 IC_{50} (μM)	PTP1B IC_{50} (μM)
TFMS-2	✗	-	-	-
4a	✓	✗	-	-
4b	✓	✗	-	-
4c	✓	✓	0.8	ND

4d	✓	✓	>100	-
5a	✗	-	-	-
5b	✓	✓	>100	ND
5c	✓	✗	-	-
10a	✓	✓	0.1	ND
10b	✓	-	-	-
10c	✓	✗	-	-
12a	✓	✗	-	-
12b	✓	✓	0.1	0.7
12c	✗	-	-	-
12d	✓	✓	5.0	1.0
12e	✗	-	-	-
13	✓	✗	-	-
16a	✓	✓	-	-
16b	✓	✓	-	-
16c	✗	-	-	-
16d	✗	-	-	-
17	✓	✓	-	-

17b	✓	-	-	-
17c	✗	-	-	-
17d	✓	✓	-	-
MPLNX4	✓	-	-	-
MPL7WL	✗	-	-	-
MPL18	✓	-	-	-
MPL20	✓	-	-	-
MPL22	✗	-	-	-
MPL24	✓	-	-	-
MPL25	✗	-	-	-
MPL26	✗	-	-	-
MPL27	✗	-	-	-
MPL29	✗	-	-	-
MPL30	✓	-	-	-
MPL31	✓	-	-	-
MPL36	✗	-	-	-
MPL47	✗	-	-	-

In addition, the remaining PTPRG enzyme from previous experiments [309] was kindly donated by Bristol Myers Squibb after the signature of an agreement with Dr. Brett R. Beno. The small quantity of PTPRG available allowed only performing one experiment for each selected compound at one concentration. From the MPL series only the positive compounds in the Phospho-Tyr assay were selected (**Table 2.11**), and the most active compounds against PTPRZ1 (**4c, 10a, 12b** and **12d**) were selected. For a better understanding of the results, the same in vitro assessment at the same concentration was carried out in PTPRZ1, for comparison purposes. The summary of these results with the compound concentration and the percentage of inhibition are summarized in the **Table 2.12.** and the results shows how compounds **4c, 10a, 12b** point to be more active in PTPRZ1 than in PTPRG. On the contrary, MPL series are considerably more active in PTPRG given even no activity in PTPRZ1.

Table 2.12. *Enzyme percentage inhibition in PTPRG and PTPRZ1 at one single concentration.*

Compound	Concentration (μm)	PTPRG Inhibition (%)	PTPRZ1 Inhibition (%)
4c	0.1	24.9	37.7
10a	0.1	17.7	50.0
12b	1	10.2	97.5
12d	0.1	24.9	14.1
MPLNX4	10	67.5	34.4
MPL18	10	27.0	X
MPL20	10	13.8	X
MPL24	10	29.3	-

MPL30	10	17.1	-
MPL31	10	27.4	X

2.14.2. Binding assay, Water LOGSY NMR spectra with PTPRG

In addition to the PTPRG enzymatic activity profile, we have performed binding assays using the Water LOGSY technique using the enzyme from the previously mentioned batch. These studies were performed by Dr. Leticia Ortí Pérez from the "Unidad de descubrimiento de Fármacos" from the Instituto de Investigación Sanitaria La Fe in Valencia, Spain. WaterLOGSY represents a powerful method for primary NMR screening in the identification of compounds interacting with macromolecules such as proteins. This methodology is particularly useful in the detection of novel scaffolds with micromolar range potency, which is precisely our case [321]. The main advantage of this method is its sensitivity, and the amount of material and time needed for the NMR screening process. The binding experiments were carried out with a relation of 2,5μM of PTPRG per 150μM of inhibitor, using 64 H^1 NMR scans and 512 WL scans with compounds **10a**, **MPLNX4**, **MPL18**, **MPL20**, **MPL24**, **MPL30** and **MPL31**.

Results show how compounds **MPL20, MPL24** and **MPL31** were the ones that presented the most potent binding signal. Unfortunately, compound **10a** was insoluble at this concentration (150μM) as well as **MPL30.** Moreover, reference compound **MPLNX4** showed low binding signal in comparison with the other actives.

Table 2.12. *Results obtained in the experimental NMR Water LOGSY. Calculated WLopP by Swiss ADME webserver, and the experimental experimental WL data obtained in relation with the potency signal.*

Compound	WLogP	Precipitated (aqueous solution)	Exp WL
10a	7,73	✓	No signal
MPLNX4	4,85/0,03	✗	+
MPL18	5,09	✗	++
MPL20	7,96	✗	+++
MPL24	7,96	✗	+++
MPL30	11,3	✓	+
MPL31	10,83	✗	+++

2.14.3. In vitro phosphorylation study of PTPRZ1 substrates

The most active and selective compounds (**10a** and **12b**) were screened in an in vitro biological assay to assess their impacts on the concentrations of phosphorylation of particular PTPRZ1 substrates. As mentioned in **section 2.4.1**, TrkA61 and ALK10 were chosen as appropriate PTPRZ1 substrates because they are engaged in PTN's pathway providing the neuroprotective effects. For this purpose, the neuroblastoma SH-SY5Y cells were selected, as they overexpress PTPRZ1 [322,323]. The SH-SY5Y cells were stimulated during 20 minutes with different concentrations of **10a** and **12b**

(1.0, 5.0 and 10.0 μM), and the phosphorylation of those specific tyrosine residues that are involved in activating both proteins TrkA (Tyr490) and ALK (Tyr1278) were evaluated using western blots. Western blots tested with anti-phospho-TrkA antibodies showed that stable concentrations of TrkA tyrosine phosphorylation improved 2-3-fold after incubation with compounds **10a** and **12b**. Western blots probed with anti-phospho-ALK antibodies demonstrated that both compounds caused a 2-fold increase in the phosphorylation of one isoform of ALK (**Fig. 2.71**)

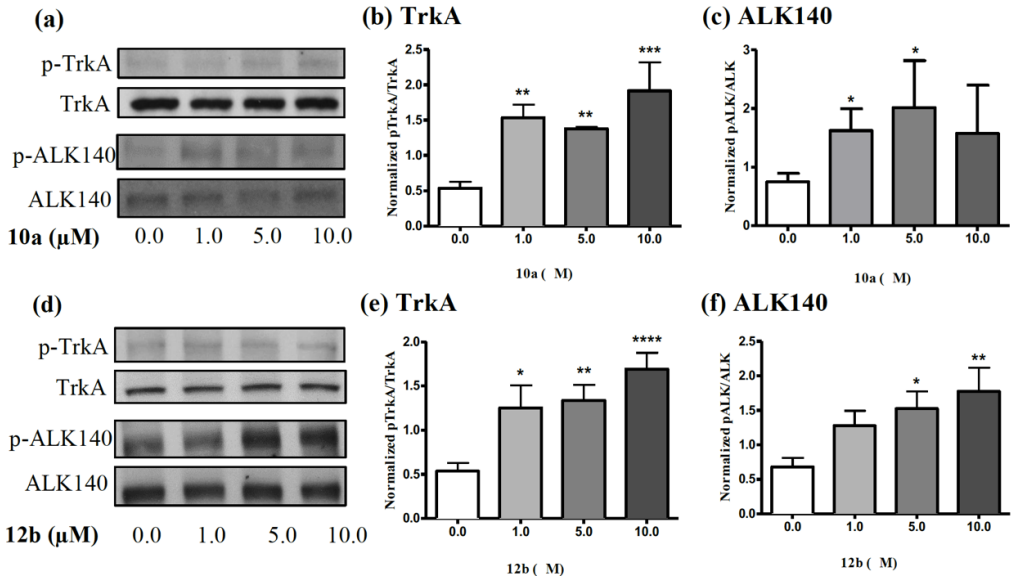


Figure 2.71. a) *Representative western blots showing* ***10a****-induced increases in phosphorylated TrkA (pTrkA) and the 140 kDa ALK isoform (pALK). Total TrkA and ALK western blots are shown below each phosphorylated protein blot for comparison.* ***b–c)*** *Quantification of western blots using ImageJ.* ***d)*** *Representative western blots showing* ***12b****-induced increases in pTrkA and pALK.* ***e-f)*** *Quantification of western blots using ImageJ. Data are presented as mean ± S.E.M. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 compared to vehicle controls (0.0 μM).*

These data demonstrate that the phosphorylation of important tyrosine residues of the PTPRZ1 substrates TrkA and ALK is considerably increased by two of the inhibitors, **10a** and **12b**, in accordance with their IC_{50} values (**Table 2.11**). TrkA is the receptor of the high affinity nerve growth factor (NGF). NGF activates the kinase activity of TrkA by enhancing the phosphorylation of Tyr490 in this protein, which is critical to cell survival and

the neuroprotective effects caused by NGF. Increased phosphorylation of Y1278 in ALK is also engaged in neuronal survival and differentiation. The information indicates that the capacity of PTPRZ1 inhibitors to boost the phosphorylation of the same proteins in TrkA and ALK will induce comparable neuroprotective impacts by inhibiting PTPRZ1's phosphatase activity on its substrates. Previous research carried out by Dr. Gonzalo Herradón demonstrated that PTN avoids in vitro and in vivo amphetamine-induced neuronal injury [239]. In order to confirm the protective roll of the PTN by the pharmacological modulation of PTPRZ1, the catecholaminergic PC12 cells were selected as they expressed readable levels of PTPRZ1. PC-12 cells were incubated during 24 hours with amphetamine (1 mM), and/or **10a** or **12b** (1.0 µM) to check that PTPRZ1 inhibitors were capable of protecting the cell cultures from amphetamine induced toxicity. The obtained results showed that compound **10a** substantially avoided the loss of the PC12 cell viability induced by amphetamine (**Fig. 2.72**) and **12b** (p = 0,06) was noted to have the same trend. These results show that PTPRZ1 inhibitors imitate the protective impacts of PTN on the toxicity in PC12 cells caused by amphetamines.

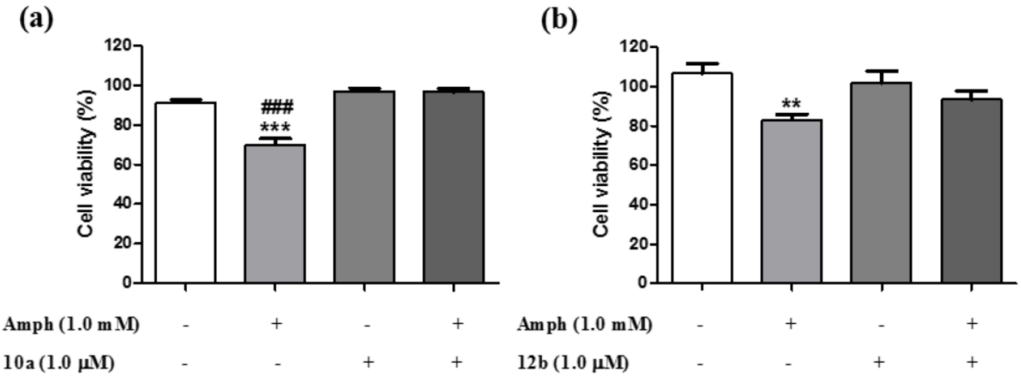


Figure 2.72. *Effects of **10a** and **12b** on amphetamine-induced toxicity in PC12 cells. PC12 cells cultured with media supplemented with amphetamine (Amph; 1 mM) and/or **10a** (1.0 µM) and/or **12b** (1.0 µM) for 24 hours. Cellular viability was assessed by the MTT test. Results are expressed as mean ± S.E.M. **p < 0.01, ***p < 0.001 vs. vehicle-treated (control) cells. ###p < 0.001 vs. Amph+10a-treated cells.*

2.14.4. Permeability Studies and GC-MS detection in brain tissue

The PAMPA assay and the detection by GC-MS were carried out by Dr. Myriam Pastor and Prof. Antonio Pineda-Lucena, respectively, in the Centro de Investigación Príncipe Felipe, Valencia.

Two parallel experiments were conducted to experimentally evaluate the capacity of the selected compounds to cross the BBB. On the one side, only compound **10a** was chosen for direct ex vivo identification in mice by measuring the compound levels in both plasma and brain tissue due to its potency and selectivity. On the other hand, the most active compounds (**4c**, **10a**, **12b** and **12d**) were chosen to perform the blood-brain barrier (PAMPA-BBB) assay [324]. This is a predictive, but more affordable, artificial membrane permeability in vitro assay. For the first experiment, samples from the sacrificed mice 1h post gavage with 60 mg/kg of **10a** were acquired for direct detection, and analyzed by GC-MS. Extracted ion chromatograms (EIC) demonstrated the presence of a peak in both plasma and brain samples of treated mice at the same retention time as a standard **10a** solution, whereas this peak was not present in untreated mice samples. The outcome of this ex vivo experiment shows that **10a** can cross the BBB and penetrate the CNS. Amitriptyline was used as a positive control for the PAMPA-BBB assay. In the acceptor plate, compounds were added at 100 μM (200 μM for **12b**) and the final concentration in the donor plate, acceptor plate and Membrane Retention (% MR) were measured by HPLC after 4 h of incubation at 37ºC (**Table 2.13**). The use of amitriptiline as a standard comes from the fact that it is a widely used antidepressant drug and has been selected as a standard due to its efficacy crossing the BBB. Compounds **4c** and **12d** display amitriptyline-like permeation through the artificial membrane. However, under these same conditions, compounds

10a and **12b** do not seem to be permeable. Considering that the in vitro permeability experiment mentioned above shows that **10a** reaches plasma and brain tissue, we were confident that at least compounds **4c**, **10a** and **12d** would be appropriate for the further development of oral CNS drugs.

Table 2.13. *PAMPA-BBB assay for selected compounds.*

Compound	Donor Plate (µM)	Acceptor Plate (µM)	%MR
4c	27.2	6.2	61
10a	17.5	0	78
12b	67.3*	0*	66*
12d	44.1	3.8	48
Amitrptiline	11.0	5.5	78

**200 µM concentration of the compound donor plate*

2.14.5. In vivo studies

Note: "All the animals used in this study were maintained in accordance with both the ARRIVE guidelines and the European Union Laboratory Animal Care Rules (86/609/ECC directive) or the National Institutes of Health Guide for the Care and Use of Laboratory Animals. All protocols were approved by the Animal Research Committee of USP-CEU or the Animal Care Committee of the University of Illinois at Chicago and procedures were used to minimize pain and suffering."

In vivo experiments were carried out by Prof. Gonzalo Herradón, Prof. Amy W. Lasek and their respective research groups.

Our best candidates **10a** and **12b** were selected to perform two in vivo assays related to alcohol consumption: a) the two-bottle drinking in the dark

(DID); and, b) the conditioned place preference (CPP) assay. The selection of compounds is based on the above described experiments that provided data of activity, selectivity and preliminary ADME properties.

As we mentioned during the introduction section, it is suggested that PTPRZ1 plays a significant role in regulating the impacts of drug abuse including AUD. In order to confirm by in vivo experiments the hypothesis that PTPRZ1 modulates the ethanol consumption, mice were treated previously with the selected PTPRZ1 inhibitors, and were subsequently tested for binge-like drinking using the DID protocol. The DID assay, as followed by *Dutton et al*.[325] consists of mice housed separately in the reverse-dark cycle spaces for 2 weeks before testing the use of ethanol. Three days before testing the use of ethanol, mice drank water from two pipes later used in the experiment in order to acclimatize them (**Fig. 2.73**).

Day	1	2	3	4
pre-treatment	-	-	10a (60mg/kg, p.o.) 1h before exposure to drink	10a (60mg/kg, p.o.) 1h before exposure to drink
Time exposed to drink	2h	2h	2h	2h + 2h
Drink	H_2O/EtOH	H_2O/EtOH	H_2O/EtOH	H_2O/EtOH

Figure 2.73. *Testing PTPRZ1 inhibitors **10a** and **12b** in binge-like drinking. Procedure followed on binge-like drinking using the drinking in the dark protocol in mice.*

In the sipper pipes, mice were provided an option of 20% ethanol or water three days later. By evaluating the quantity of fluid in the pipes, fluid consumption was evaluated daily for 4 days. The bottle position (left or right) has been altered to regulate side preference every day. Mice were provided access to the ethanol and water pipes for 3h in the dark cycle for a period of 2h on the first 3 days of testing. Mice had access to ethanol and water pipes for 4h on the fourth day and the volume consumed was evaluated at 4h. All mice received an oral gavage vehicle (0.1 mL) on days 1 and 2. Mice were administered on the third and fourth day of the DID exam with **10a**, **12b** or vehicle by oral gavage 1h prior to the drinking session. The results

show that mice treated with PTPRZ1 inhibitors drank less ethanol than controls, especially with **10a** (**Fig. 2.74**). Using the same dose, **10a** appears to be more efficient than **12b** in decreasing the use of ethanol, although a greater dose of **12b** may also be considerably efficient. Therefore, we can state that in the DID experiment, the **10a** RPTPZ1 inhibitor reduced binge-like drinking. The impacts of **10a** in ethanol place conditioning was evaluated in order to test the possibility that this RPTPZ1 inhibitor might reduce ethanol consumption for its capacity to block the rewarding impacts of ethanol. A two-compartment device was used for the CPP experiment; one compartment had black floor and walls, and the other one had black floor and white walls (**Fig. 2.75**). The phases of the experiment included preconditioning (Pre-C, day 1), conditioning (days 2-4) and testing (CPP, day 5). Mice were free to explore the two compartments during preconditioning for a 15-minute period; a monitoring video system was used to calculate the time spent in each compartment. The conditioning stage consisted of two sessions of conditioning per day. The first session started at 8 a.m. in the morning, where all mice got a single saline injection and were limited for 5 min to one room. The mice were injected with ethanol in the afternoon session starting at 3 p.m. and confined to the other compartment for 5 minutes. The compartments were randomly assigned to the ethanol or saline condition. The procedure used on days 3 and 4 was the same, but to avoid the influence of circadian variability, the order of the treatments (morning/evening) changed. For saline and ethanol conditioning, compound 10a or the vehicle were administered to the mice by oral gavage 1 hour before each of the sessions. Mice got a drug-free, 15-minute preference test in the exam stage on day 5. In this stage, as in the preconditioning stage, the animals moved freely throughout the device while the time spent in each room was registered. The distinction between the moment spent in this stage in the ethanol-paired compartment and the time spent in the preconditioning in the same room was deemed to indicate the degree of ethanol-induced conditioning. For the CPP control experiment in which we conditioned with **10a** in the lack of ethanol, compound or vehicle

were administered on days 2, 3 and 4, 1h before mice were injected with saline, and confined to one side of the device for 5 minutes (**Fig. 2.75**). During the testing stage on day 5, mice got a compound-free preference test of 15 minutes and the time they spent in the **10a** paired room measured.

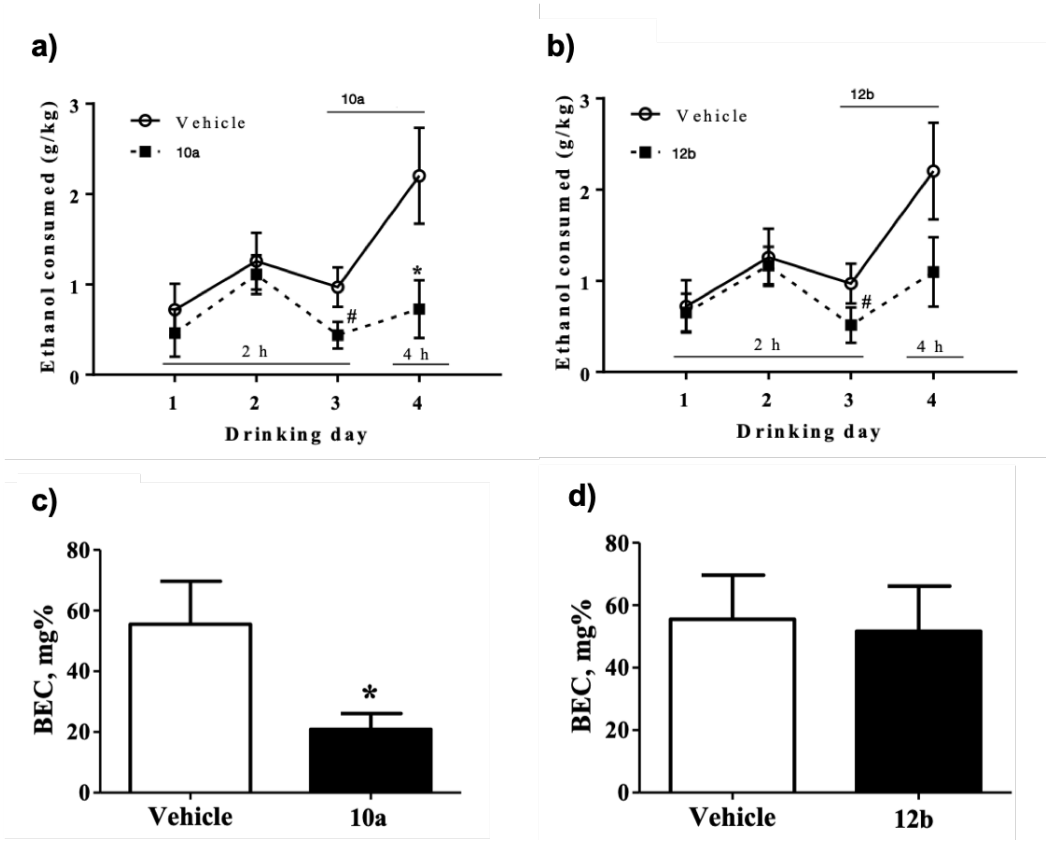


Figure 2.74. a) *Ethanol consumed (g/kg) during 2 h of drinking on days 1-4. On days 3 and 4, mice were administered 60 mg/kg* ***10a*** *or vehicle (n=12/group) 1 hour before each drinking session in the DID test. Inset bar graph shows the amount of ethanol consumed during 4 h on day 4.* ***b)*** *Ethanol consumed (g/kg) during 2 h of drinking on days 1-4. On days 3 and 4, mice were administered 60 mg/kg* ***12b*** *or vehicle (n=12/group) 1 hour before each drinking session in the DID test. Inset bar graph shows the amount of ethanol consumed during 4 h on day 4.* ***c-d)*** *Blood ethanol concentration (BEC, mg%) in vehicle and* ***10a*** *or* ***12b*** *treated mice after the final 4 h-drinking session on day 4.*

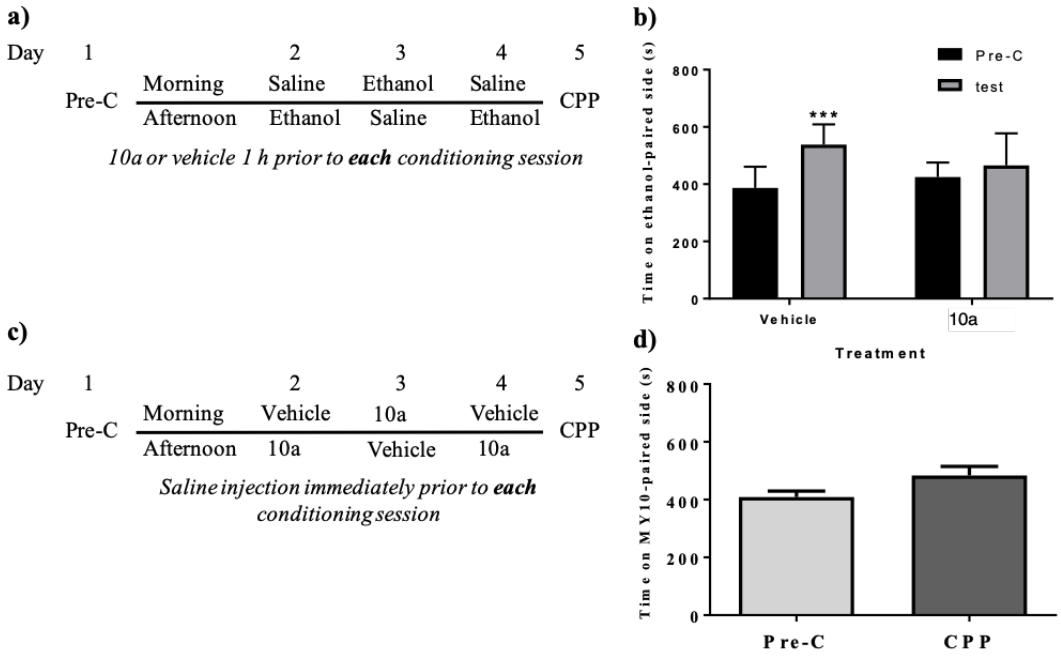


Figure 2.75. a) *Scheme showing treatment conditions for the ethanol CPP test.* ***b)*** *Time (seconds) spent on the ethanol-paired side before conditioning (Pre-C) and after conditioning (CPP) in mice treated with vehicle or 60 mg/kg* ***10a****. The data shows that* ***10a*** *prevents ethanol CPP.* ***c)*** *Scheme showing the treatment conditions for the* ***10a*** *CPP test (in the absence of ethanol).* ***d)*** *Time (seconds) spent on the* ***10a****-paired side before conditioning (Pre-C) and after conditioning (CPP) in mice conditioned with 60 mg/kg* ***10a*** *(n = 13) instead of ethanol.*

2.15. Molecular Modelling of complexes with PTPRZ1 and PTPRG

Molecular modelling techniques have been applied in order to rationalize the activity profile of the designed compounds for PTPRZ1 and PTPRG. In summary, docking and MD simulations were applied to propose a plausible binding mode of the **TFMS-2** and derivatives in complex with PTPRZ1 and PTPRG with the final goal to understand the in vitro results. This methodology was selected because **TFMS-2** derivatives present an unknown binding mode which has not been confirmed yet by other experimental techniques such as X-ray or NMR. Only complexes obtained with compounds **4c**, **4d**, **5b**, **10a**, **12b** and **12d** have been further studied by MD simulations as they are the only compounds that present IC_{50} values.

On the other hand, MPLs series were studied by FEP techniques because they arguably preserve the same binding mode than the corresponding compound **NX4** or **7WL** within the corresponding superopen crystal structures. As it was said before, due to the lack of crystal structures in PTPRZ1, models were built and used The complexes obtained inside PTPRZ1 and PTPRG with the aforementioned **TFMS-2** derivatives have been modelled with the available superopen conformations **3QCH, 5H08** and also in the model from the MD simulations with organic solvents. In those three, MD simulations have been carried out to shed light and propose a plausible binding mode of these molecules.

2.15.1. Molecular docking in PTPRZ1 models

At the beginning of this project, since our main target is PTPRZ1, we performed docking calculations inside the PTPRZ1 homology models that we built as an initial approach. This first step was carried out with the complete set of molecules designed against PTPRZ1 as a first and fast attempt to model these complexes. The PTPRZ1 closed, open and superopen conformations were selected as receptors and the compounds were docked inside using as center of the GRID box the phosphatase active site. The summary of the docking binding scores of the selected binding poses in all protein conformations of each compound are summarized in **Table 2.14.**

In general, the higher docking scores are found in the complexes with the superopen conformation, followed by those in the closed and the open ones. It is also relevant to highlight that those compounds with the trifluorosulfide group present higher scores than the sulfoxide derivatives, and that molecules with an extra carbon on the linker also have better scoring values. This last feature has a direct effect on the flexibility of the compound and therefore a better adjustment of the molecule to the receptor.

For the closed and open conformations of the receptor, the trifluorosulfide

and trifluorosulfoxide groups reach the active site interacting with Cys1939 and with the NH of the backbone. Moreover, in the closed conformation, these molecules interact with the WPD-loop (mainly with Met1902 and Arg1939), E-loop and P-Tyr loop each one in a different way (**Fig. 2.76-a**). For **TFMS-2**, the interaction is made-up by electrostatic interactions between the trifluoromethylsulfoxi group and the side chains of Lys1832, Arg1834 and Lys1836 present in the E-loop. Additionally, it establishes a T-displaced interaction with the side chain of Tyr1756, and a hydrogen bond with Asn1758, both present in the p-Tyr-loop.

In the apo conformation the interactions established with the WPD-loop are not as strong as this loop is farther away from the active site than in the closed conformation. These results in an overall lower scoring and weaker interactions. For instance, the previously mentioned interaction with Met1902 is not observed. As a consequence, molecules rearrange within the active site and the P-loop interacting mainly with the side chains of Tyr1756, Ile 1757 and Asn1758 from the pTyr-loop, and also with Gln1977 from the Q-loop (**Fig. 2.76-b**).

The superopen conformation gives accessibility to the cryptic pocket that is occupied by the trifluorosulfide or trifluorosulfoxide and the benzyl moieties. Among the amino acids that line-up the hydrophobic pocket, the main van der Waals interactions are established between the trifluoromethylthiobenzyl or trifluoromethylsulfoxibenzyl moieties and the aromatic side chains of Trp1989 and Phe1984, as well as with the aliphatic and cycloaliphatic side chains of Arg1939 and Pro1905, respectively. The rest of the molecule, in all cases, interacts with the surroundings of the WPD-loop and the active site, with no direct interaction with the catalytic site (**Fig. 2.76-c**).

Table 2.11. *Docking scores obtained in the different complexes with the PTPRZ1; closed, open and superopen models. Docking scores are expressed in Kcal·mol^{-1}*

Compound	Model Closed	Model Open	Model Superopen
TFMS-2	-5.28	-1.97	-5.44
4a	-3.43	-2.51	-7.45
4b	-2.12	-2.45	-7.76
4c	-2.68	-2.55	-7.36
4d	-2.51	-3.22	-8.71
5a	-3.28	-2.17	-2.36
5b	-3.38	-2.9	-6.60
5c	-5.36	-2.34	-6.1
10a	-3.72	-1.94	-8.25
10b	-4.54	-2.11	-7.23
10c	-3.67	-1.91	-6.74
12a	-3.72	-2.55	-7.28
12b	-3.72	-2.13	-6.16
12c	-4.30	-2.00	-4.97
12d	-3.992	-2.76	-8.21
12e	-3.61	-1.30	-8.31
13	-3.06	-1.02	-5.73
16a	-4.10	-1.87	-8.34
16b	-4.09	-1.74	-5.92
16c	-2.75	-2.04	-7.97
16d	-3.23	-2.48	-6.34
17a	-3.66	-1.88	-6.38
17b	-4.06	-1.69	-5.76
17c	-3.81	-1.74	-7.78
17d	-1.49	-1.95	-6.07

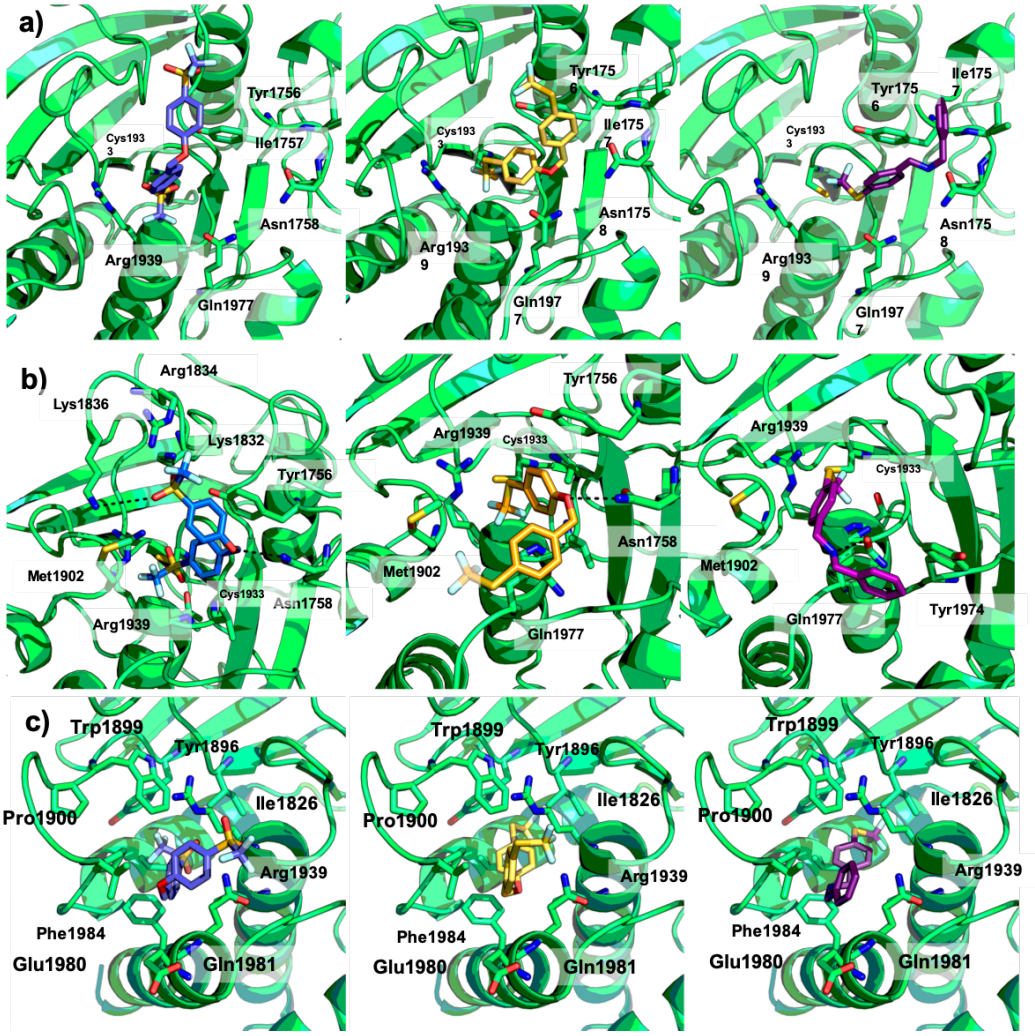


Figure 2.76. *Predicted binding modes for compounds* ***TFMS-2*** *(blue),* ***10a*** *(yellow) and* ***12c*** *(purple) docked inside PTPRZ1 closed a) open b) and superopen c) conformations.*

2.15.2. Complexes modelled inside PTPRZ1

The highest docking scores obtained in the initial PTPRZ1 models corresponded to the superopen conformation where the compounds explored the cryptic pocket (**Fig. 2.76** and **Table 2.11**). As explained at the beginning of this section (section **2.15**), we have three PTPRZ1 superopen conformations: the **3QCH** model, the **5H08** crystal structure, and the forced-opened structure obtained by MD simulations with organic solvents.

As mentioned before, the first and preliminary study in molecular modelling are the docking calculations. Nevertheless, the docking score is not the best choice to predict the most active compounds as the approximation does not

predict full protein flexibility. The disadvantages of this technique have been mentioned in Chapter I. For these reasons, MD simulations have been carried out in order to further study the proposed binding mode of the studied ligands and try to obtain a better correlation between the binding energy of the complexed compounds, and the stability along the simulation time. The 10 ns MD simulations were performed for the complexes with compounds **4c**, **4d**, **5b**, **10a**, **12b** and **12d** as their IC_{50} values had been obtained experimentally (**Table 2.11**). The stability of the complexes as evaluated by calculating the RMSD values of the ligands, and the total binding energies were measured by means of the MM-ISMSA program.

Therefore, the following sections will describe the docking results in the different PTPRZ1 conformations and their stability during the MD simulation.

2.15.2.1. Complexes with the PTPRZ1 3QCH model

The results of the MD simulations show that the previously described van der Waals interactions established with the hydrophobic pocket in the docking binding modes (**Fig. 2.78**) were maintained during most of the simulation. This can also be observed in the low RMSD values obtained, which remained under 3Å with the exception of compounds **5b** and **12b**; and, the structures of the most populated conformers (**Fig. 2.77** and **Fig. 2.79**). The interactions with the highest energy contributions are the hydrophobic, π–π stacking and T-shaped interactions established between the trifluoromethyl-thiobenzyl moieties of the ligands and the aromatic side chains of Ile1826, Phe1984, Trp1889 and Pro1905. All compounds remain stable inside the hydrophobic pocket and this binding mode is preserved during all simulated time. The differences come mainly by the interaction of the linkers and the intrinsic flexibility of the different compounds. For example, compound **4c** interacts strongly with the side chains of Pro1900, Val1904 and the aliphatic part of the side chain of Glu1980 of the WPD-loop. In addition, the amide moiety of **4c** stablishes alternating hydrogen bonding interactions with either the NH of the side chain of Gln1981 or the

backbone carbonyls of Pro1905 and Glu1980. However, in the complex of **10a** with PTPRZ1, the described hydrophobic interactions persist, whereas the electrostatic interactions are weaker as the previously described hydrogen bonds cannot be established.

Table 2.12. *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes **4c**, **4d**, **5b**, **10a**, **12b**, **12d**, inside the PTPRZ1 **3QCH** model.*

PTPRZ1 3QCH model	4c	4d	5b	10a	12b	12d
Docking score	-7.36	-8.71	-6.60	-8.25	-6.76	-8.21
RMSD	1.96 ±0.24	1.75 ±0.28	4.21 ±0.67	2.94 ±0.50	3.83 ±0.58	2.64 ±0.58
Total binding energy	-45.75	-39.49	-36.02	-41.30	-40.55	-36.24

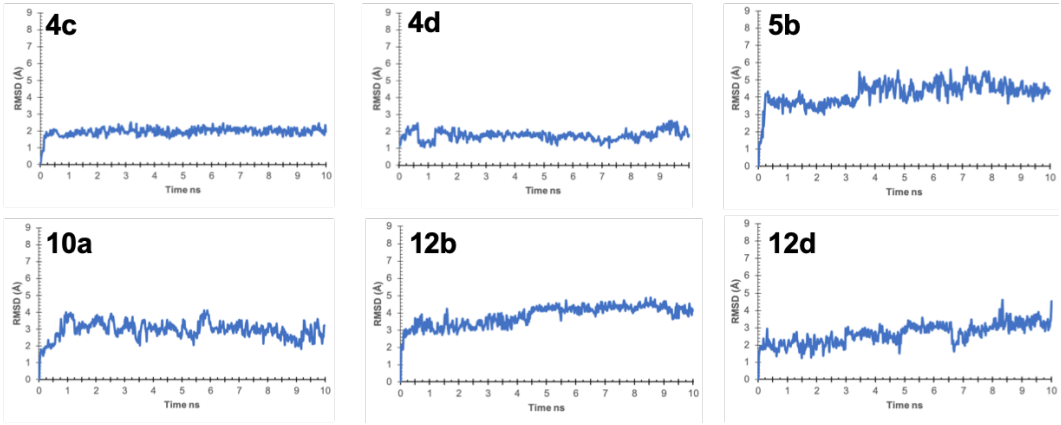


Figure 2.77. *Evolution of the RMSD (Å) for complexes of PTPRZ1 **3QCH** model and **4c**, **4d**, **5b**, **10a**, **12b** and **12d** along the 10 ns MD simulation.*

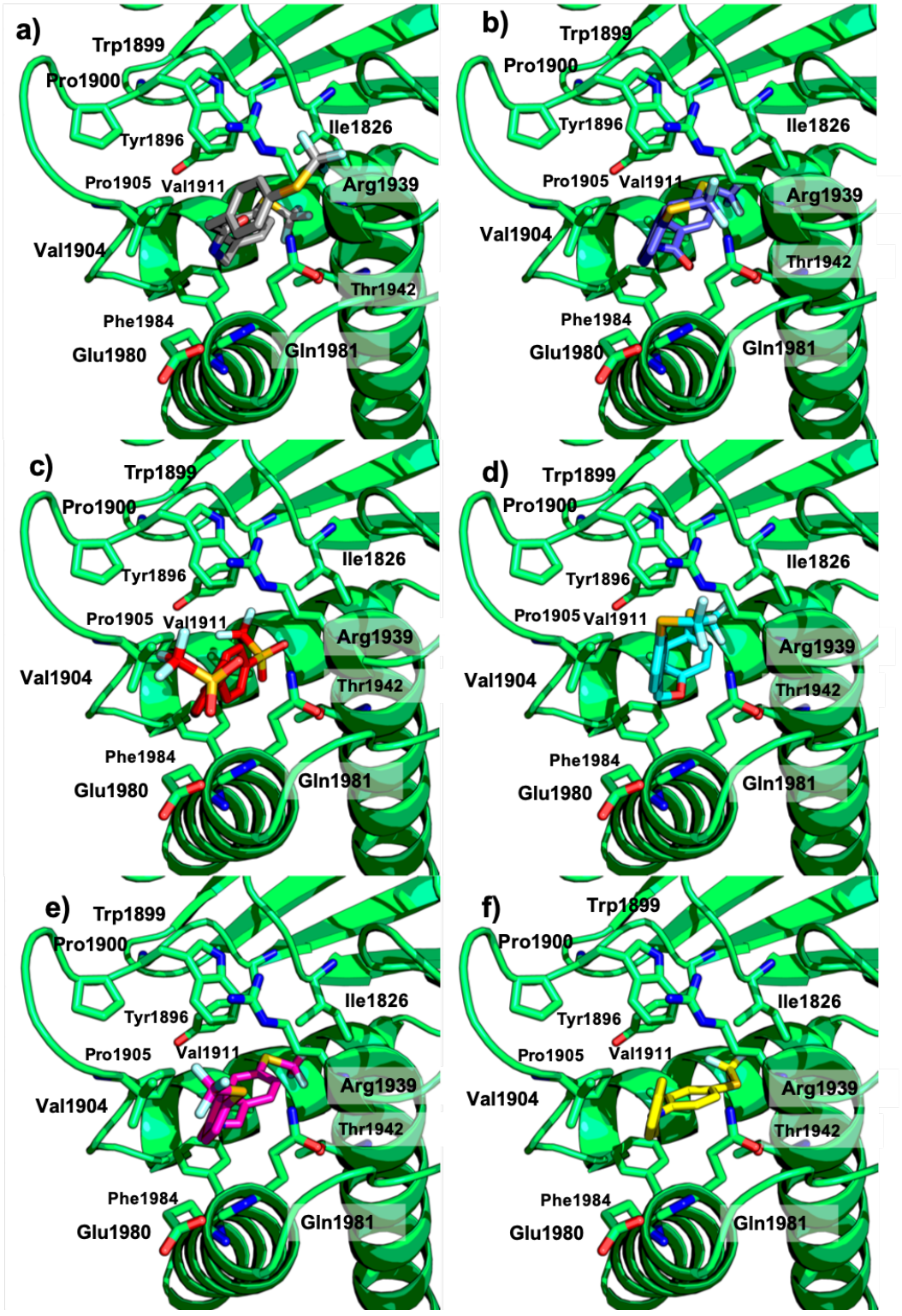


Figure 2.78. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *docked inside PTPRZ1 model* ***3QCH***, *superopen conformation.*

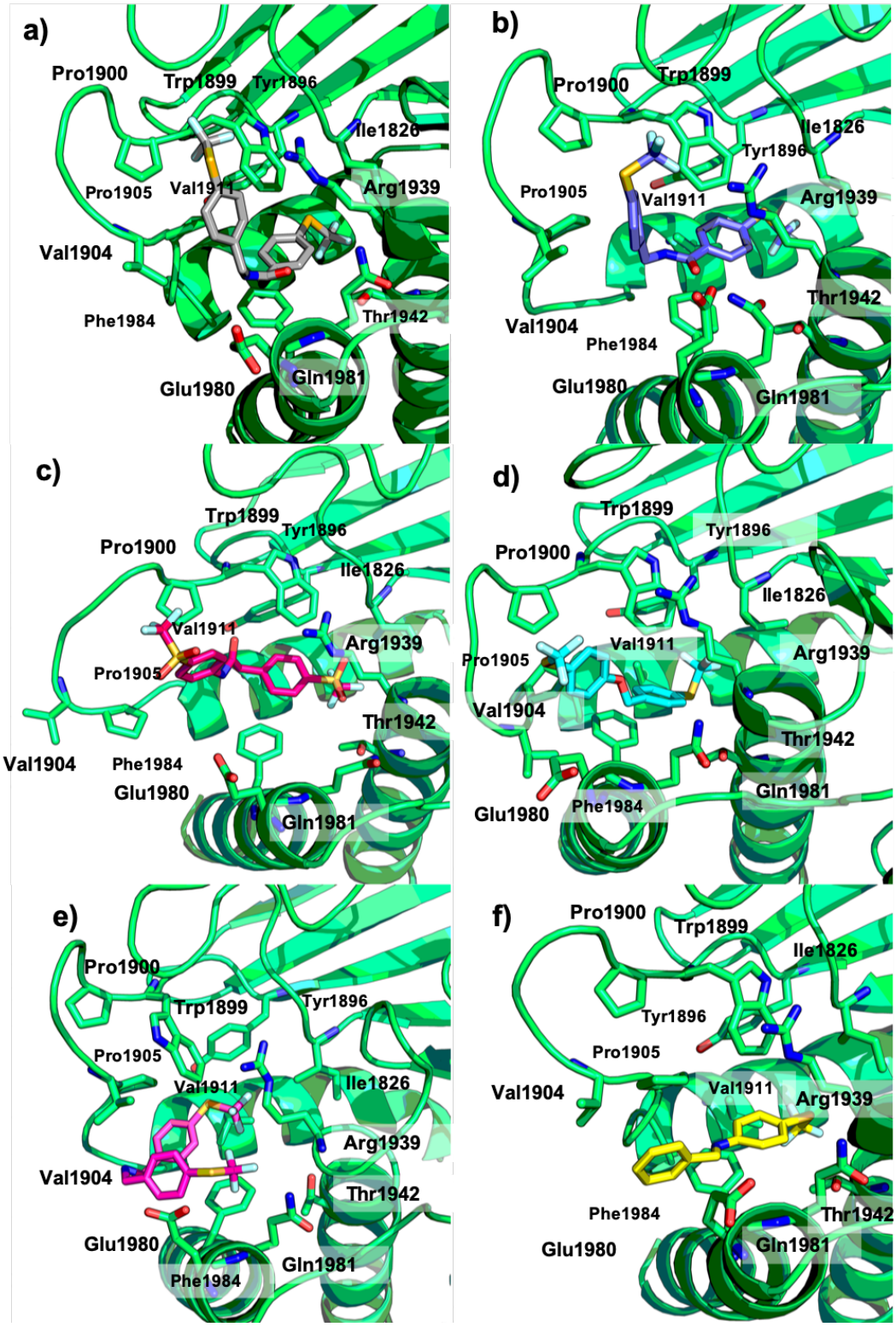


Figure 2.80. *Most populated conformers for complexes of* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *along the 10 ns MD simulation using as starting conformations the complexes obtained in the docking with the PTPRZ1 3QCH model structure.*

2.15.2.2. Complexes with the 5H08 PTPRZ1 crystal structure

The docking studies were carried out with the selected compounds: **4c**, **4d**, **5b**, **10a**, **12b** and **12d**, inside the cryptic pocket using the compound **7WL** as a reference (**Fig.2.83**). Interestingly, the binding modes obtained were totally different from the previous ones performed on a modeled structure based on **3QCH** (section **2.15.2.1**), because now the compounds reached the cryptic site but through the arc of the WPD-loop instead of through the aperture next to the active site. This new binding mode occurs because the crystal structure **5H08** is slightly more opened than the template used for the initial models (built with PDB code **3QCH** as template). This allowed the studied compounds to interact with the hydrophobic amino acids positioned surrounding the arc of the WPD-loop. The structure of **7WL** presents similar moieties compared to that of **10a**. In detail, the ethoxy-trifluoromethyl-benzyl moiety fits with the trifluoromethylsulfil-benzyl moiety (**Fig. 2.81**), and both moieties are superimposed in the docking, thus guiding the rest of the molecules through the arc of the WPD-loop. Moreover, in this new docking pose the interaction with Tyr1864, Gln1898, Pro1900 and Tyr1907 increases notably the scoring of the docking calculations (**Table 2.13**) compared with the previous docking results in the PTPRZ1 **3QCH** model (**Table 2.12)**. In all docking results, hydrogen bonds are established between the oxygen of the backbone of Gln1898 and NH group of **4c**, **4d**, **5b**, **12b** and **12d**; and, between the NH of the backbone of Glu1906 with an acceptor atom from the linker in **4c**, **4d**, **5b**, and **10a** (**Fig. 2.83**).

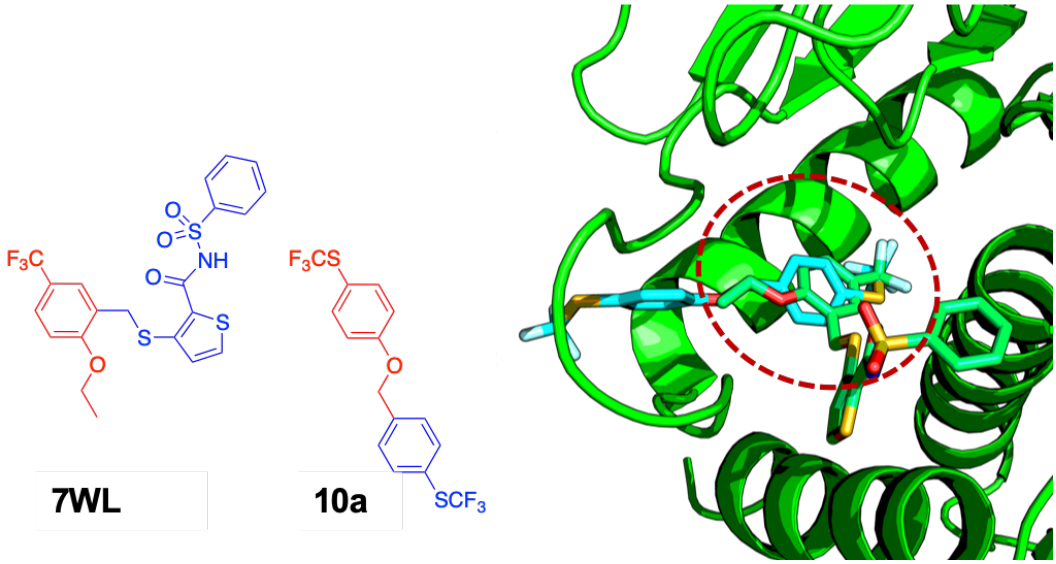


Figure 2.81. *2D structure of **7WL** and **10a** showing in red the similar moieties. Superimposition of the crystal structure **5H08** with **7WL** (green) and the docking result of **10a** (blue).*

The stability of the docking poses obtained in the PTPRZ1 **5H08** PDB structure was also tested by means of MD simulations out of which the RMSD values and the interaction energies were also computed (**Fig. 2.82**, **Fig. 2.84** and **Table 2.13**). For all compounds, the binding mode and the main interactions established with the arc of the WPD-loop and the cryptic pocket remained stable during the simulation. The most populated conformations extracted from the MD simulation showed the preserved binding mode for each complex (**Fig. 2.84**). The obtained mean RMSD values range from 3.7 to 1.83 Å for **5b** and **10a,** respectively (the less and most active compounds) (**Table 2.13**). Moreover, when comparing these results with those obtained for the PTPRZ1 **3QCH** model we noted that complexes with **5b**, **10a**, **12b** and **12b** presented lower RMSD values, which is indicative of higher stability inside this conformation of PTPRZ1. On the other hand, the energy decomposition analysis shows that **4c**, **10a** and **12b** present higher total binding energy in contrast with **4d**, **5b** and **12d** (**Table 2.13**). Moreover, comparing these data with the previous binding mode with the PTPRZ1 **3QCH** model we observed that **4c**, **10a** and **12b** presented higher binding energy, whereas **4d**, **5b** and **12d** lower. This fact, together with previous reported RMSD values, suggests that the proposed binding

mode in the PTPRZ1 **5H08** crystal structure seems to be more in agreement with the biological activity than the obtained with the PTPRZ1 **3QCH** model.

Table 2.13 *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes **4c**, **4d**, **5b**, **10a**, **12b**, **12d**, inside the **5H08** PTPRZ1 crystal structure.*

PTPRZ1 5H08	4c	4d	5b	10a	12b	12d
Docking score	-8.96	-9.13	-8.94	-8.71	-8.20	-8.60
RMSD	2.26 ±0.38	2.83 ±1.04	3.69 ±0.69	1.83 ±0.41	2.86 ±0.33	2.10 ±0.67
Total binding energy	-47.45	-31.85	-32.74	-42.70	-41.68	-33.90

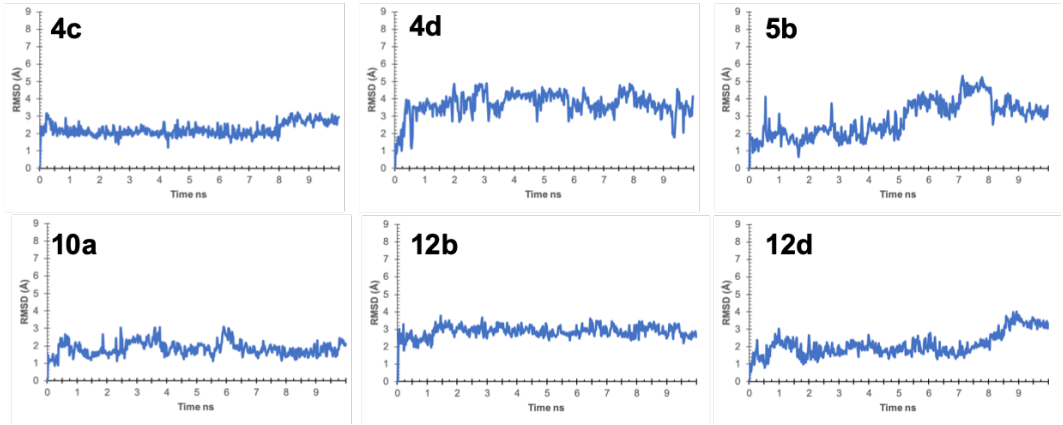


Figure 2.82. *Evolution of the RMSD (Å) for complexes of PTPRZ1 **5H08** crystal structure and **4c**, **4d**, **5b**, **10a**, **12b** and **12d** along the 10 ns MD simulation.*

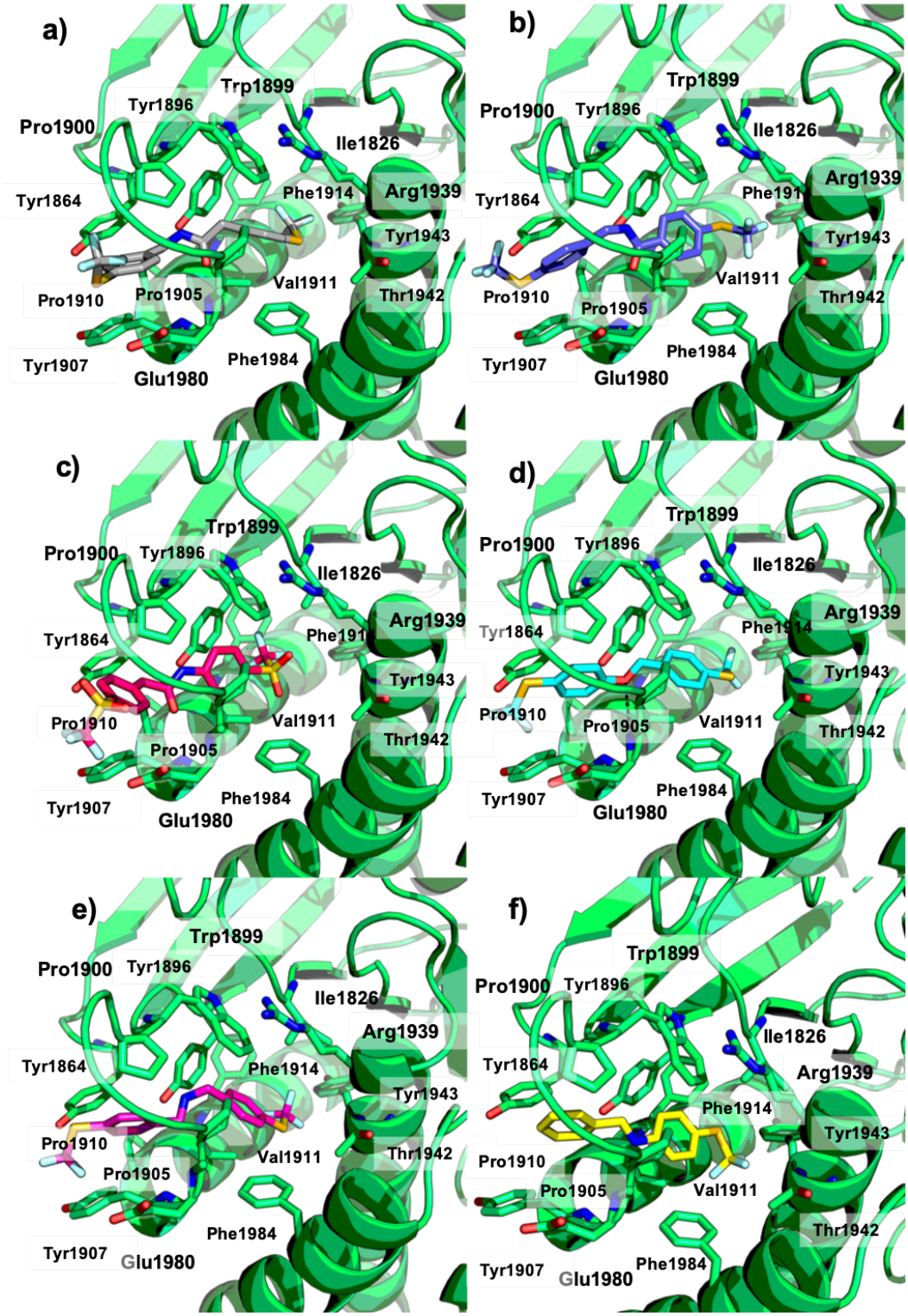


Figure 2.83. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)***, *docked inside the PTPRZ1* ***5H08*** *crystal structure.*

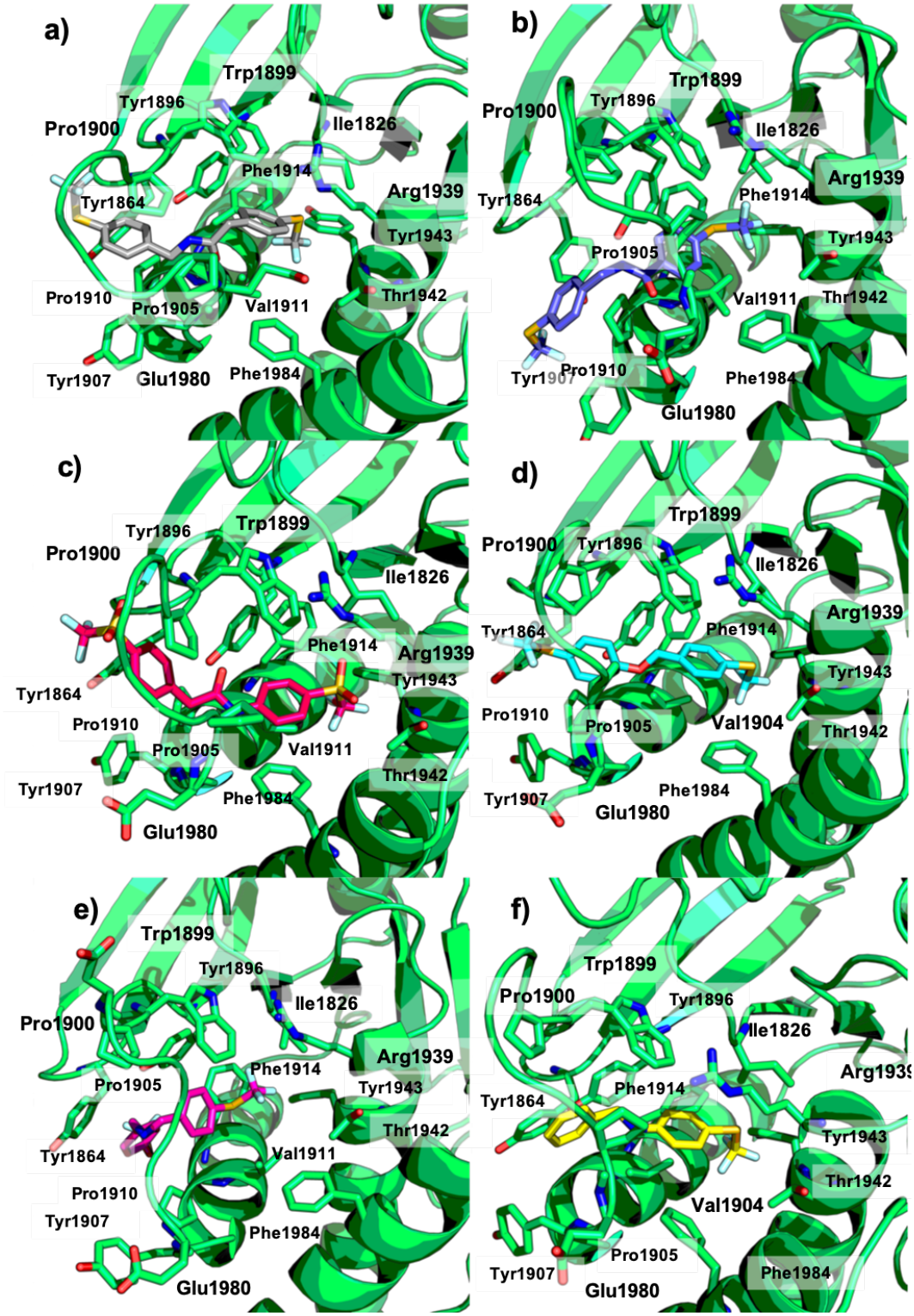


Figure 2.84. *Most populated conformations for complexes of PTPRZ1 and* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *along the 10 ns MD simulation using as starting conformations the complexes obtained in the docking with the* ***5H08*** *PTPRZ1 crystal structure.*

2.15.2.3. Complexes with the PTPRZ1 solvent model

The force-opened structure of PTPRZ1 issued from the last 10 ns of the EP-Dichlorobenzene/water MD simulations (see section **2.12.2**) was used to perform docking calculations of compounds **4c**, **4d**, **5b**, **10a**, **12b** and **12d**. The aim was to search for new plausible binding modes that could arise from these different protein conformations that were not brought about by the binding of inhibitors such as **NX4** and **7WL**. The force-opened structure of PTPRZ1 presented two dichlorobenzene molecules inside their cryptic pocket after the 100ns MD, so the grid box was centered on the deepest dichlorobenzene molecule and the protein was further stripped of any solvent.

Despite that, the docking results showed interactions with the arc of the WPD-loop and the cryptic pocket, they rendered different binding modes than those obtained with the PTPRZ1 **3QCH** model and the **5H08** crystal structure.

The set of compounds interacted strongly with a pull of tyrosines present in the corresponding helixes of the WPD-loop (Tyr1864, Tyr1865, Tyr1894, Tyr1896) and Tyr1907 and Tyr1943 from the cryptic pocket (**Fig. 2.85**). Additionally, hydrophobic interactions were established with Pro1910, Val1911, Phe1914, Arg1939, Phe1984 common to the cryptic pockets (**Fig. 2.87**).

The MD simulations showed high binding mode stability of the complexes with compounds **4d**, **5b** and **12b**; whereas, complexes with compounds **4c**, **10a** and **12b** suffered a rearrangement along the MD simulation until they found a binding mode common to all compounds where the interactions with Tyr1865 and Tyr1896 reach a stable π-π interaction (**Fig. 2.88**). The high binding energy values obtained for all complexes except for **12d**, and being higher for the most active compounds **4c**, **10a** and **12b** all compounds (**Table 2.214**). The total binding energy of these complexes is significantly higher than that of the ones obtained with the **5H08** crystal structure

because of the interaction with the pull of tyrosines. However, despite these higher energy values, the whole stability of the complexes is lower than that of the simulations carried out with the **5H08** structure (see section **2.15.2.3**). This can be observed from the higher RMSD values obtained except for **5b** (**Fig. 2.86**).

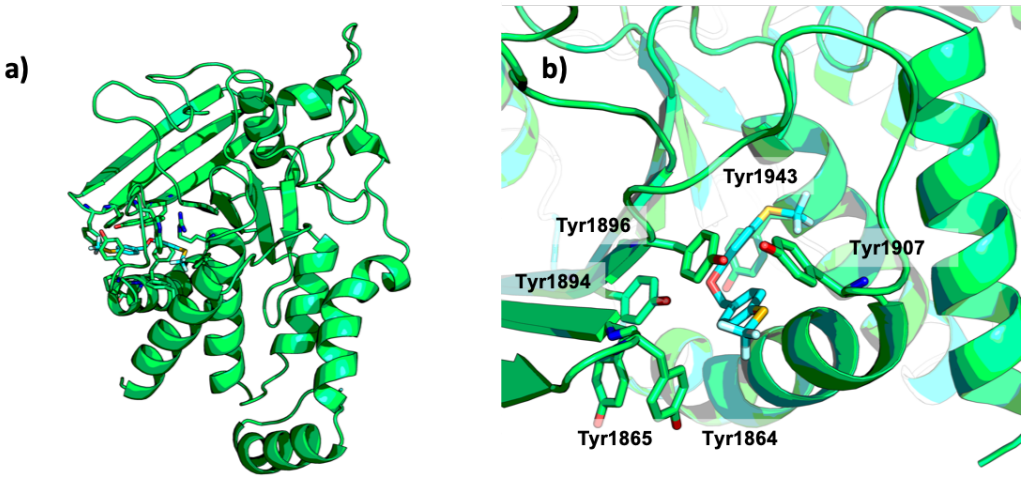


Figure 2.85. a) *The general binding mode inside PTPRZ1 using the structure obtained after MD simulations with EP-dichlorobenzene mixed in water.* ***b)*** *The interaction of compound* ***10a*** *with the 6 mentioned tyrosine residues.*

Table 2.14. *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes* ***4c***, ***4d***, ***5b***, ***10a***, ***12b***, ***12d****, inside the pocket formed by MD simulations with organic solvents in PTPRZ1.*

PTPRZ1 model	solvent	4c	4d	5b	10a	12b	12d
Docking		-10.25	-10.12	-10.14	-10.48	-8.62	-9.50
RMSD		3.11 ±0.45	1.91 ±0.39	2.18 ±1.13	4.06 ±0.44	2.47 ±0.25	3.96 ±0.74
Total binding energy		-48.39	-44.37	-40.02	-46.30	-45.67	-35.63

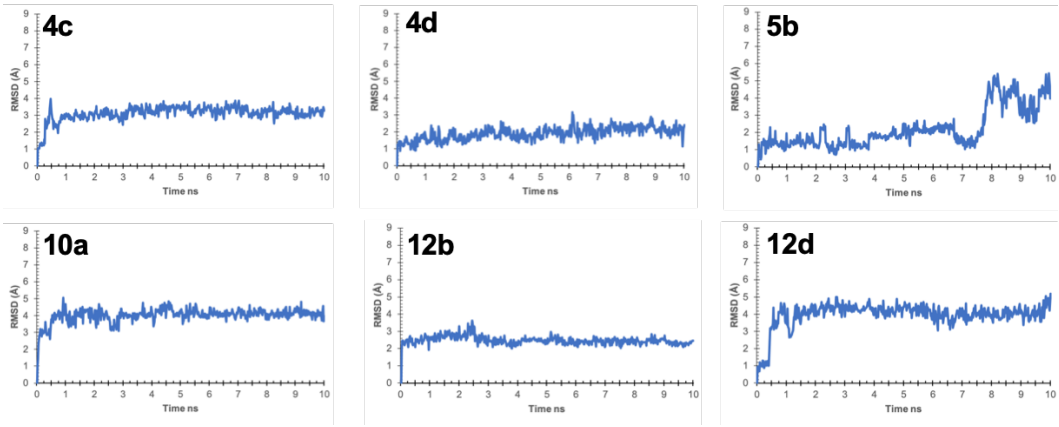


Figure 2.86. *Evolution of the RMSD for complexes of* ***4c***, ***4d***, ***5b***, ***10a***, ***12b*** *and* ***12d*** *with PTPRZ1 along the 10 ns MD simulations.*

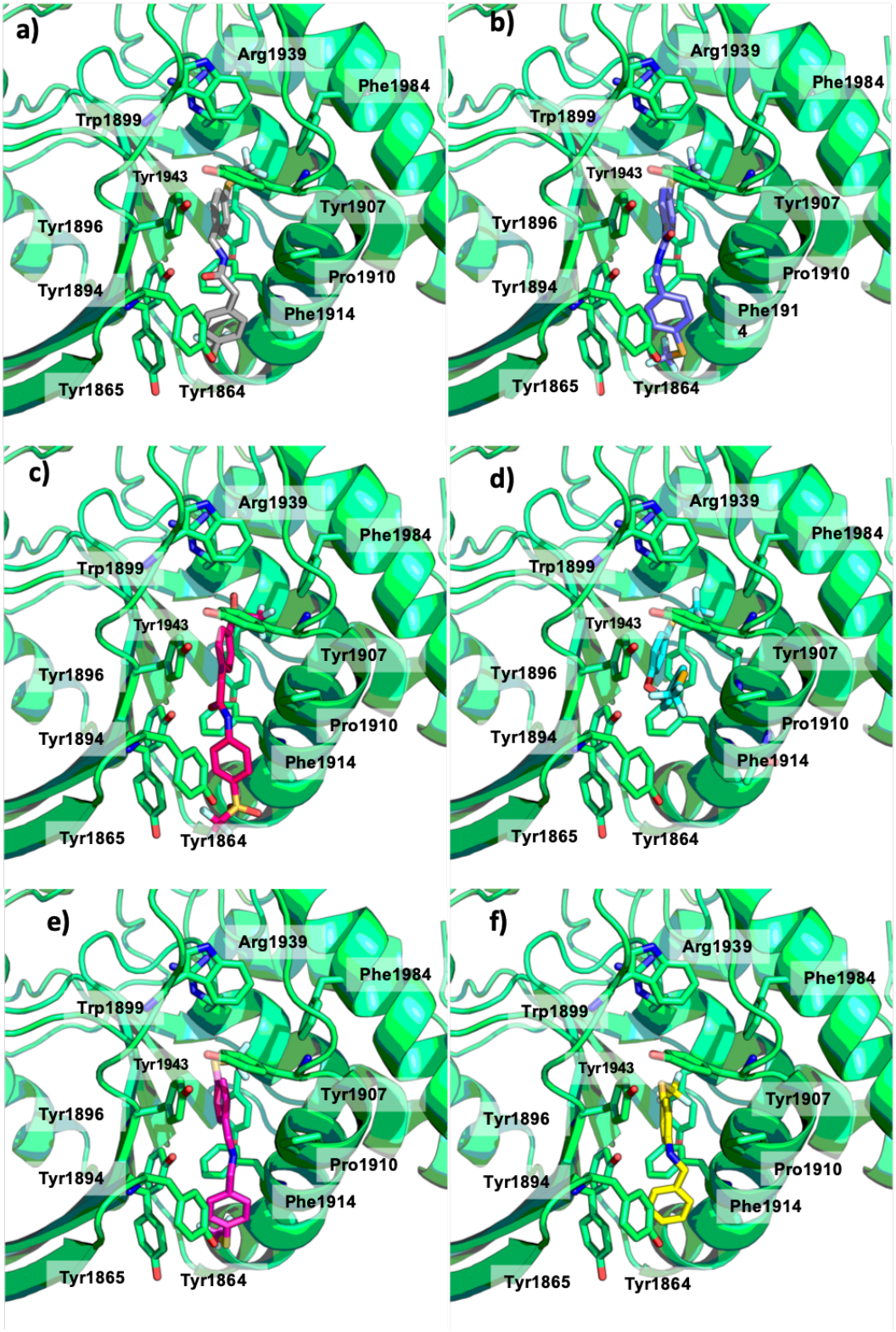


Figure 2.87. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)***, *docked inside the force-opened structure of PTPRZ1 issued from the MD simulation with organic solvents.*

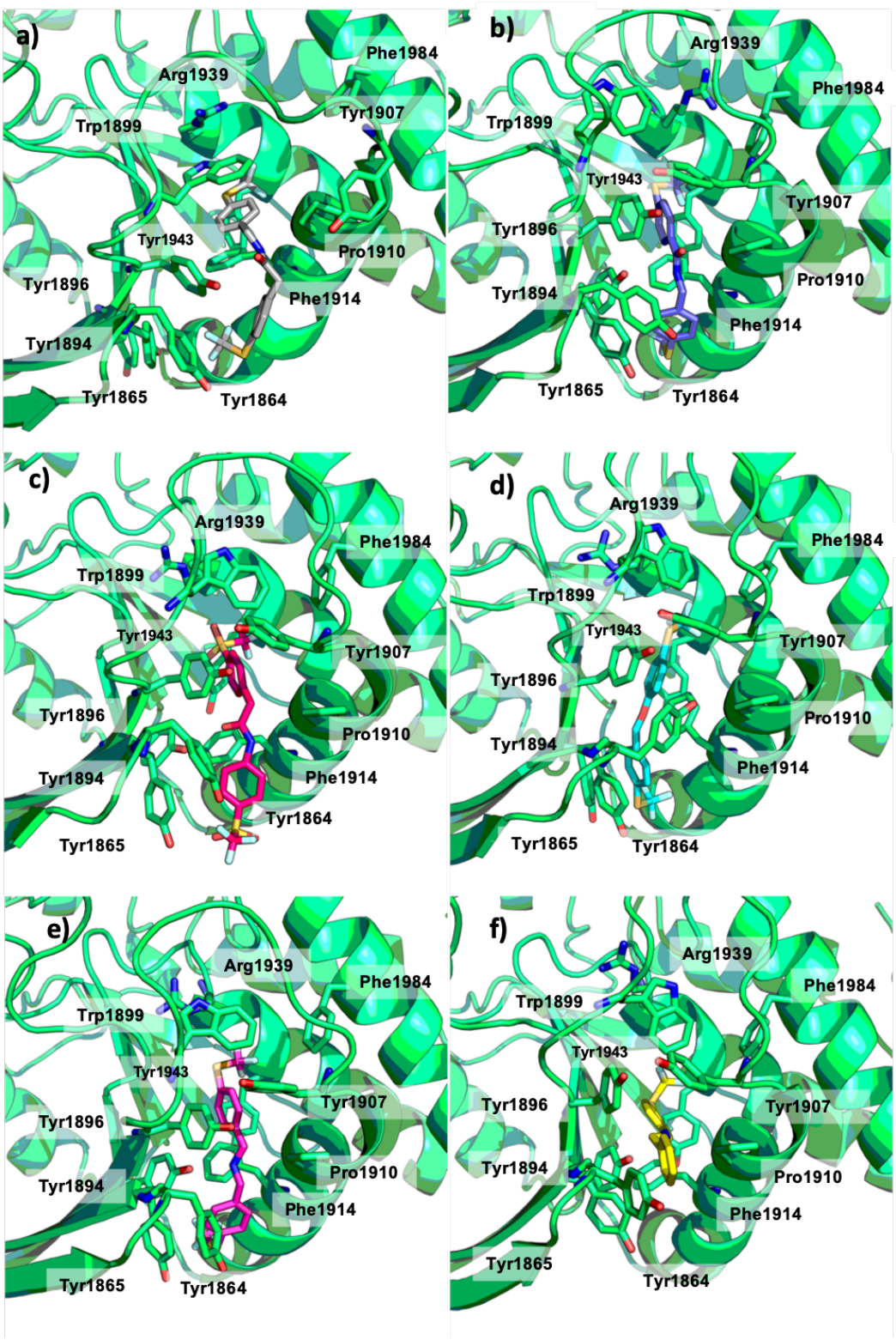


Figure 2.88. *Most populated conformations for complexes of PTPRZ1 and* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *along the 10 ns MD simulation using as starting conformations the complexes obtained in the docking with the force-opened structure of PTPRZ1 issued from the MD simulation with organic solvents. .*

2.15.3. Complexes modelled inside PTPRG

Aiming to rationalize the possible selectivity profile of the synthesized compounds, their most plausible binding mode to PTPRG was predicted. For that purpose, the PTPRG structures from PDB code **3QCH**, and the one issued from the MD simulation with EP-Dichlorobenzene/water mixture were used. However, to fully compare the results with those obtained with PTPRZ1, a homology model of the PD1 of PTPRG was built using the SWISS Model server [326] using the PTPRZ1 PDB code **5H08** as template. The sequence coverage was 75.58% and the QMEAN value -1.73, so the quality of the model was reliable. This way we obtained a PTPRG structure that shared the WPD-loop conformation of PTPRZ1 (**5H08**), exactly the same approach previously used for PTPRZ1.

2.15.3.1. Complexes with the PTPRG 3QCH crystal structure

The docking results obtained for **4c**, **4d**, **5b**, **12b** and **12d** inside the **3QCH** crystal structure were similar to those obtained in the previous dockings discussed in section **2.15.2.1** (where the PTPRZ1 model **3QCH** was used as a receptor) (**Fig. 2.90**). Following the same protocol, the docked complexes were subjected to MD simulations. The docking scores, RMSD values, and total binding energies are summarized in **Table 2.15**. The compounds interacted with the cryptic pocket formed by Ile950, Tyr1023, Val1031, Pro1032, Val 1038, Phe1041, Arg1066, Thr1069, Tyr1070, Phe1111 and Gln1108 mainly by hydrophobic interactions (**Fig. 2.90**). During the MD simulations the compounds remained inside the cryptic pocket, interacting in a similar manner as in the equivalent PTPRZ1 complexes, mainly by hydrophobic interactions with the side chains of Ile950, Tyr1023, Val1031, Pro1032, Val 1038, Phe1041, Arg1066, Thr1069, Tyr1070, Phe1111 and Gln1108 (**Fig. 2.91**). These RMSD values highlight a higher deviation compared to those in the PTPRZ1 **3QCH** model, this deviation seems to be mainly due to an easier movement of the molecules

inside the cryptic pocket of PTPRG than that of PTPRZ1 (**Fig. 2.89**). For example, the mean RMSD values of **12b** and **5b** range from 2.68 to 5.06 Å, respectively (**Fig. 2.89**). As we mentioned before, the cryptic pocket of PTPRG appears to be more flexible, and this flexibility would be the cause of the lower stability of these complexes. The total binding energy values calculated during the entire MD simulation provided lower values than in the PTPRZ1 **3QCH** model, which is a direct consequence of the higher mobility of the ligands inside the PTPRG cryptic pocket (**Table 2.15**).

Table 2.16. *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes **4c**, **4d**, **5b**, **10a**, **12b**, **12d**, inside the **3QCH PTPRG** crystal structure.*

PTPRG 3QCH	4c	4d	5b	10a	12b	12d
Docking Score	-7.42	-7.28	-7.12	-7.07	-7.32	-6.66
RMSD	3.91 ±0.89	3.07 ±0.69	5.06 ±0.93	4.39 ±0.82	3.05 ±0.37	2.68 ±0.49
Total binding energy	-38.67	-41.50	-30.00	-39.75	-36.79	-32.69

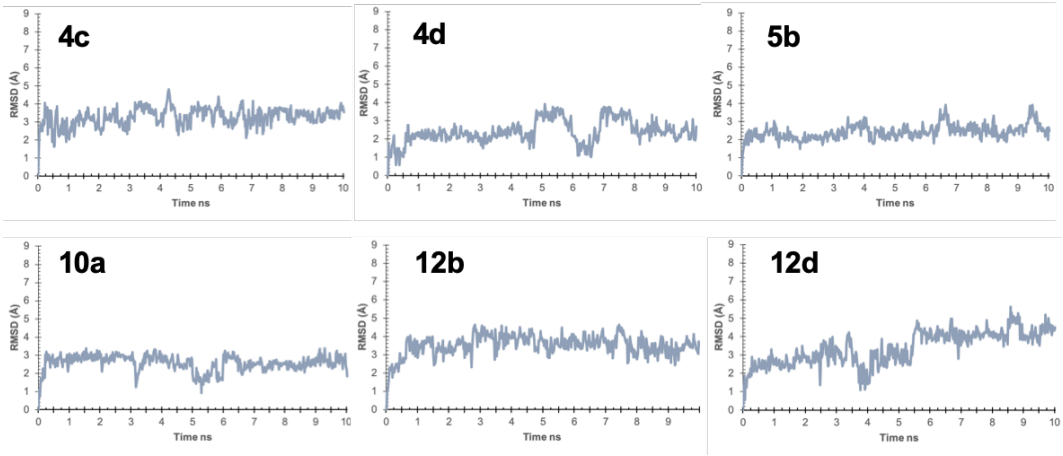


Figure 2.89. *Evolution of the RMSD for complexes of **4c**, **4d**, **5b**, **10a**, **12b** and **12d** with the PTPRG **3QCH** crystal structure along the 10 ns MD simulations.*

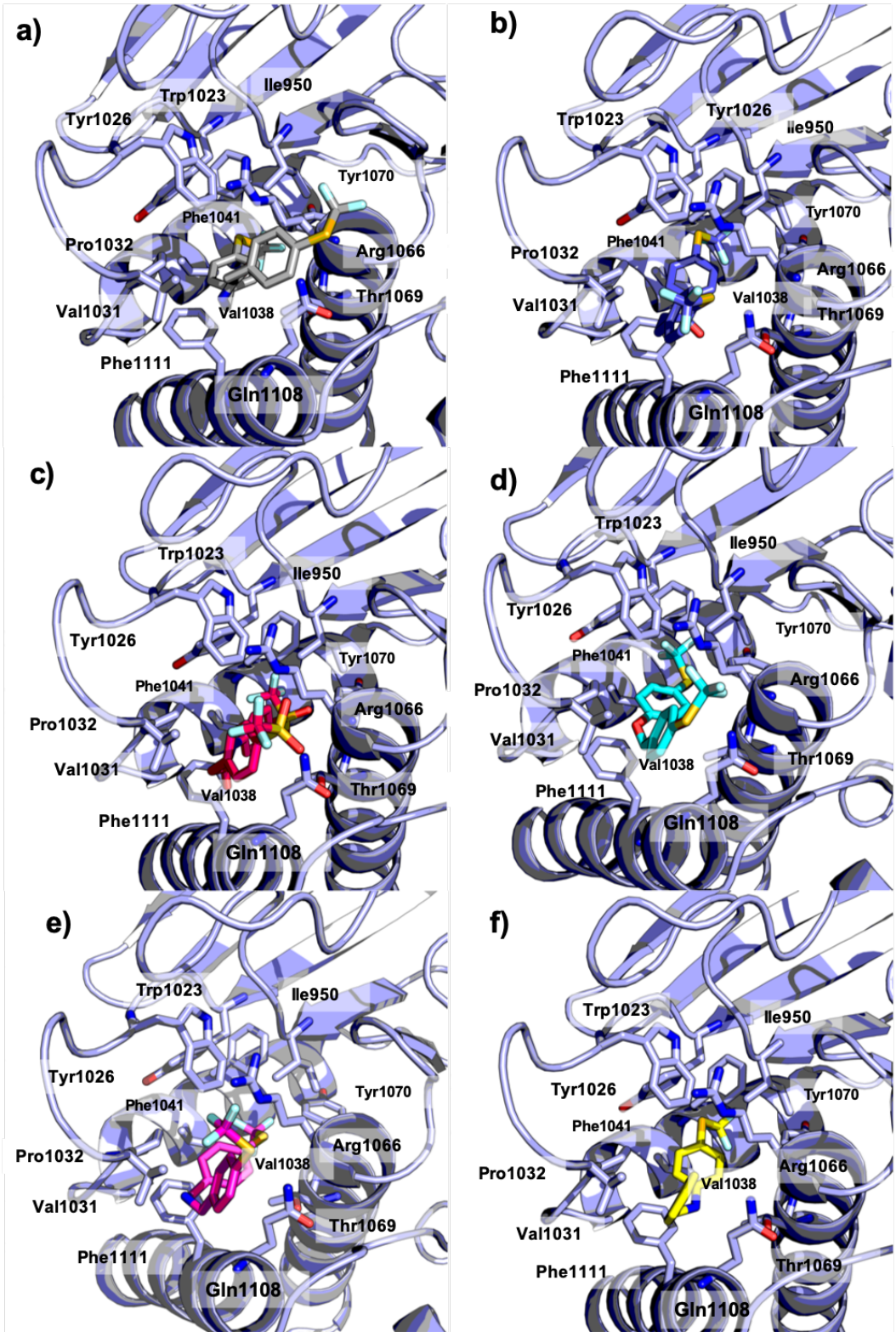


Figure 2.90. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *docked inside the PTPRG* ***3QCH*** *crystal structure.*

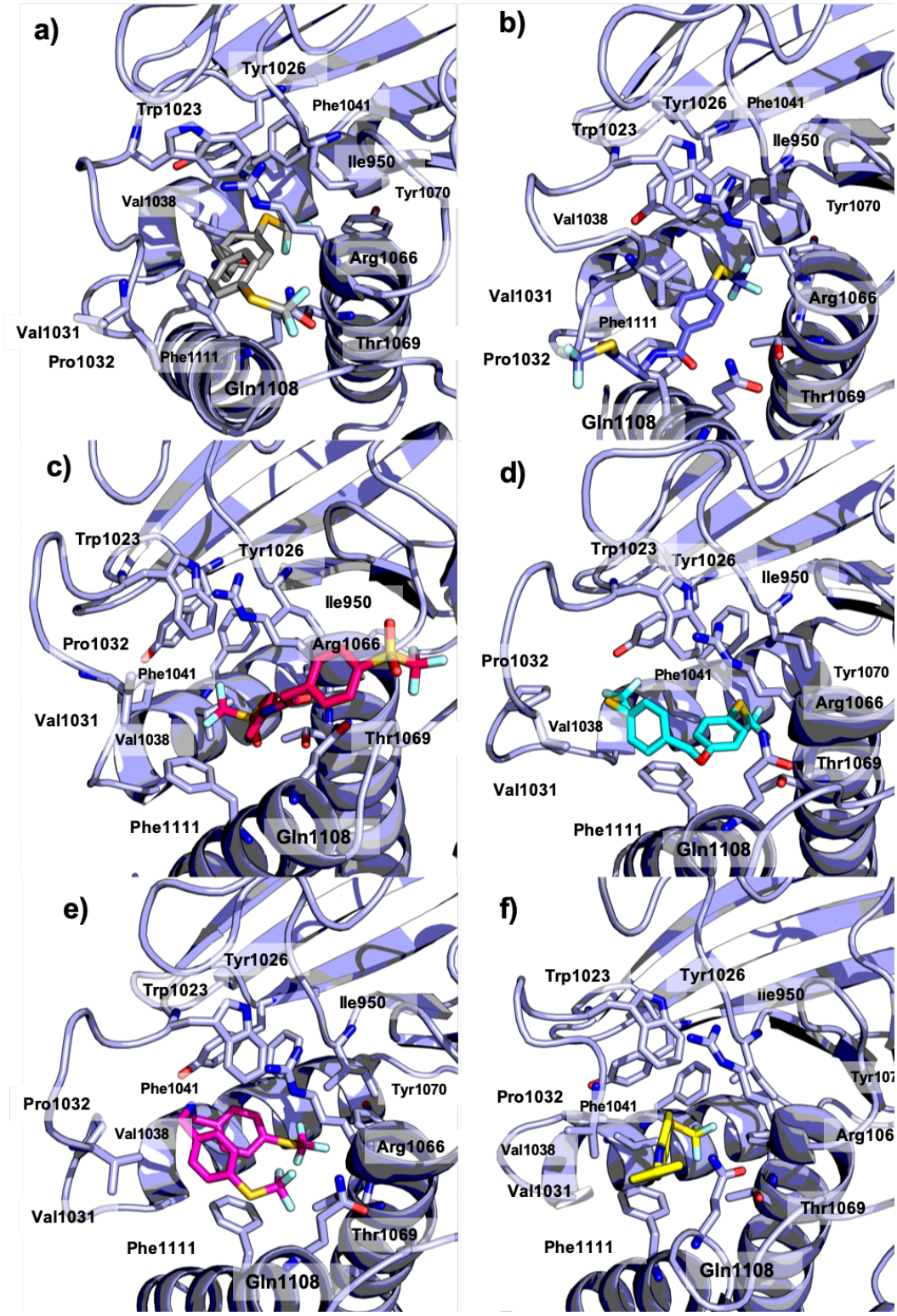


Figure 2.91. *Most populated conformers for complexes of* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *with the PTPRRG* ***3QCH*** *structure during the 10 ns MD simulation.*

2.15.3.2. Complexes with the PTPRG 5H08 model

The same approach has been carried out using the newly built PTPRG model. The results reproduced the binding poses previously obtained with the **5H08** PTPRZ1 crystal structure, however, the docking scores were comparatively lower (**Table 2.16** and **Fig. 2.93**). This would be mainly due to the difference in the amino acid composition in the arc of the WPD-loop between PTPRZ1 and PTPRG. The most interesting difference is the presence of Cys988 in PTPRG in the position occupied by Tyr1864 in PTPRZ1. The aromatic side chain of Tyr in PTPRZ1 allows a strong interaction with the set of compounds (discussed in **section 2.12.2, Fig. 2.65**) that is lost in PTPRG. The MD simulations issued from these docking poses also show a good stability of the complexes (**Fig. 2.92** and **Fig. 2.94**) maintaining the initially obtained pose. Interestingly, compound **12d** moves from the arc of the WPD-loop to reach for a better interaction with the side chain of Phe1111. This movement is possible as compound **12d** is one of the shortest compounds, which allows it to rearrange within the cryptic pocket. The RMSD values were generally higher than those obtained in the simulated complexes in which the **5H08** PTPRZ1 crystal structure was used as protein target (**Table 2.14** and **Fig. 2.88**). This confirms the higher flexibility of the WPD-loop of PTPRG, and the behavior of compound **12b** also highlights the bigger volume of the cryptic pocket. In line with these results, the total binding energy of the complexes is comparatively lower to those previously obtained. In summary, taking these computational results together we can conclude that the binding modes obtained by docking techniques were similar for both PTPRZ1 and PTPRG because they both present the same amino acid composition (with some exceptions), therefore docking studies cannot be of help to rationalize the in vitro results. MD simulations, that allow an evaluation of the complexes in an aqueous environment taking into account protein flexibility, afforded notable differences. The general behavior of our molecules inside PTPRG **5H08** model seems to be more stable compared to the original crystal structure

3QCH and this fact is because molecules are anchored through the WPD-loop reducing notably its mobility.

Table 2.16. *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes **4c**, **4d**, **5b**, **10a**, **12b**, **12d**, inside the **PTPRG** model **5H08.***

PTPRG MODEL 5H08	4c	4d	5b	10a	12b	12d
Docking Score	-8.55	-9.06	-8.28	-7.73	-8.26	-8.01
RMSD	3.27 ±0.51	2.04 ±0.63	2.41 ±0.42	2.56 ±0.44	3.51 ±0.56	3.37 ±0.91
Total binding energy	-37.41	-33.43	-37.62	-38.05	-28.69	-34.56

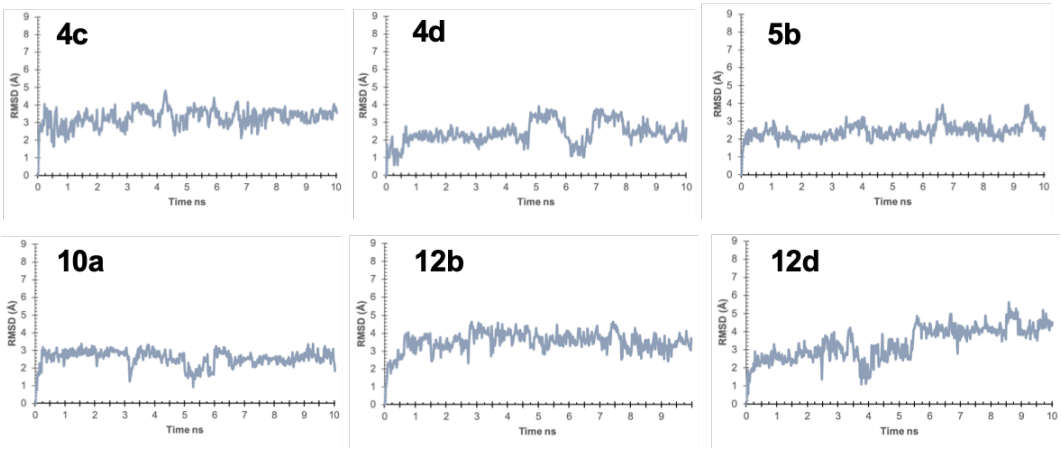


Figure 2.92. *Evolution of the RMSD for complexes of **4c**, **4d**, **5b**, **10a**, **12b** and **12d** with the PTPRG model **5H08**.*

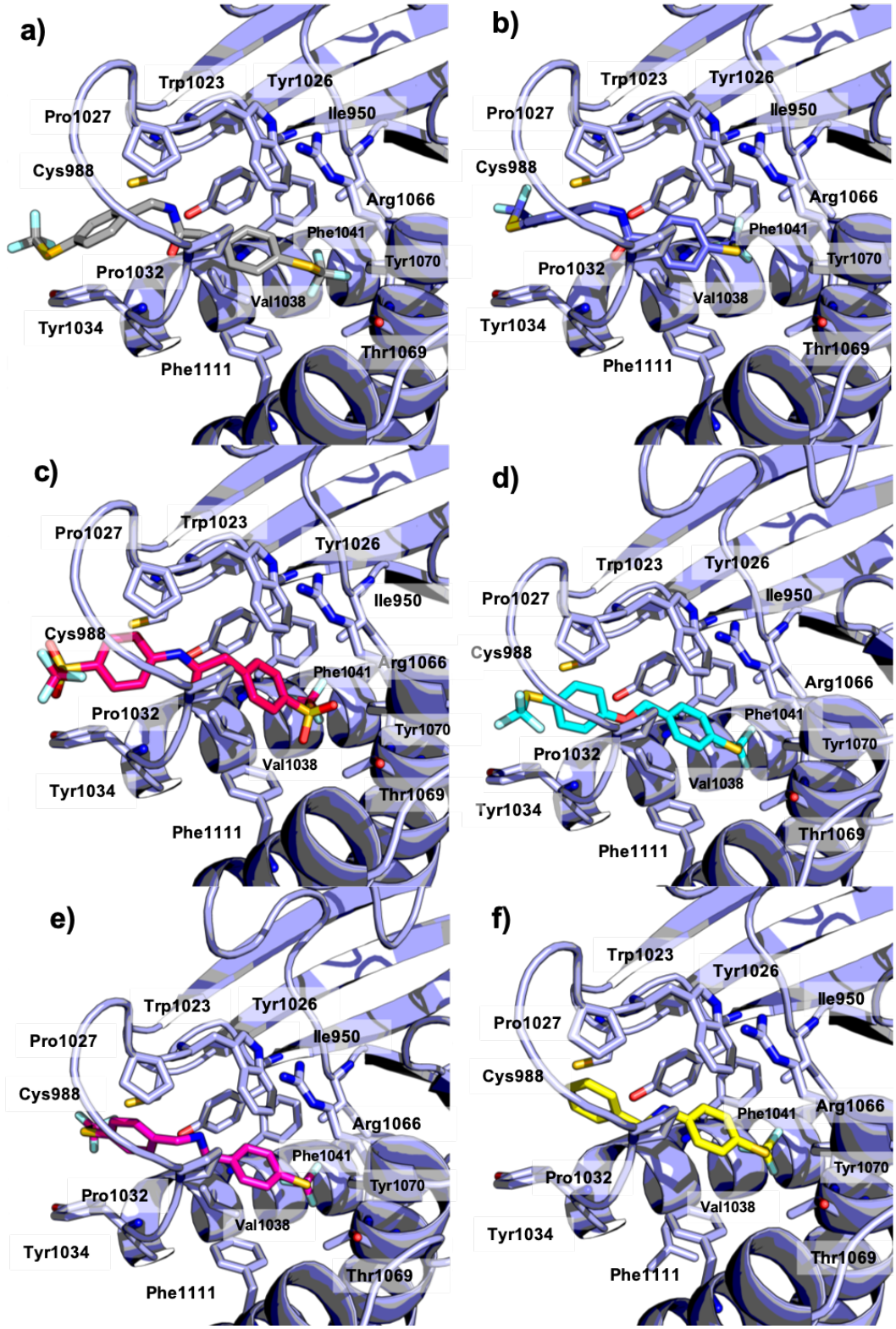


Figure 2.93. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)***, *docked inside the* ***PTPRG model 5H08***.

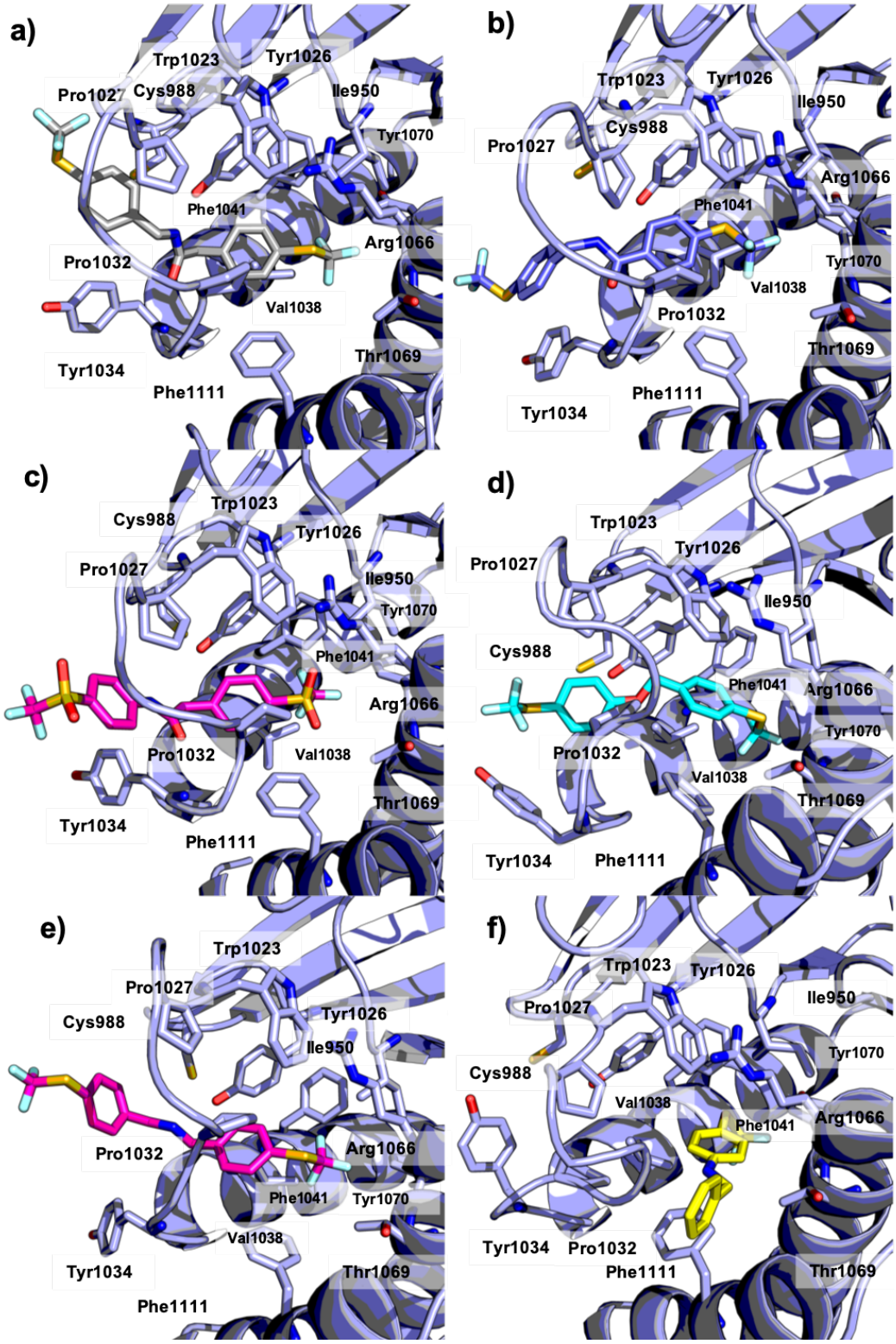


Figure 2.94. *Most populated conformers for complexes of* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *inside* ***PTPRG model 5H08*** *during the 10 ns MD simulation.*

2.15.3.3. Complexes inside PTPRG solvent model

The force-opened structure obtained by the MD simulation with EP-Dichlorobenzene/water mixture (see section **2.12.2**) was used as receptor for these docking calculations. It is important to mention, that the hydrophobic pocket obtained here was comparatively bigger than that of PTPRZ1.

In PTPRG the binding mode obtained present differences in comparison to those obtained for PTPRZ1 (**Fig 2.96**). For complexes with **4d**, **5b** and **10a** the binding mode was achieved through the arc of the WPD-loop and hydrophobic interactions established with Ile950, Tyr1023, Pro1027, Val1031, Pro1032, Tyr1034, Pro1037, Val 1038, Phe1041, Arg1066, Thr1069, Tyr1070 and Phe111. On the other hand, compounds **4c**, **12b** and **12d** interacted only in the cryptic pocket and the Q-loop establishing interactions with Ile950, Tyr1023, Val1031, Pro1032, Val 1038, Phe1041, Arg1066, Thr1069, Tyr1070, Phe1111 and Gln1108 (**Fig. 2.96**). The MD simulations carried out with these complexes showed lower stability of the binding modes in comparison to those analyzed in previously coming from the crystal structures **3QCH** and **5H08** model in PTPRG**,** this fact can be recognized by the higher mean RMSD values here compared to the two other previous systems (**Table 2.17**). The most unstable complexes in PTPRG were **10a** and **12b** that reached RMSD values even higher than 6 Å (**Fig. 2.95**). In some complexes **4c**, **4d** and **5b** the overall binding modes described above remained stable along the 10 ns simulations. In general, these small molecules moved around the artificially opened pocket, which leads to overall higher mean RMSD values and lower total binding energies (**Fig. 2.97** and **Table 2.17**).

Table 2.17. *Docking score (Kcal·mol^{-1}), mean RMSD (Å) with standard deviation and total binding energy calculated with MM-ISMSA during the 10 ns MD simulation (Kcal/mol) of the complexes **4c**, **4d**, **5b**, **10a**, **12b**, **12d**, inside the pocket formed by MD simulations with organic solvents in **PTPRG**.*

PTPRG solvent model	4c	4d	5b	10a	12b	12d
Docking	-9.05	-8.78	-7.66	-9.08	-7.17	-7.67
RMSD	2.69 ±0.36	3.1 ±0.53	2.12 ±0.85	5.46 ±0.67	4.97 ±0.96	2.88 ±0.66
Total binding energy	-34.22	-40-92	-40.47	-40.38	-41.0	-31.58

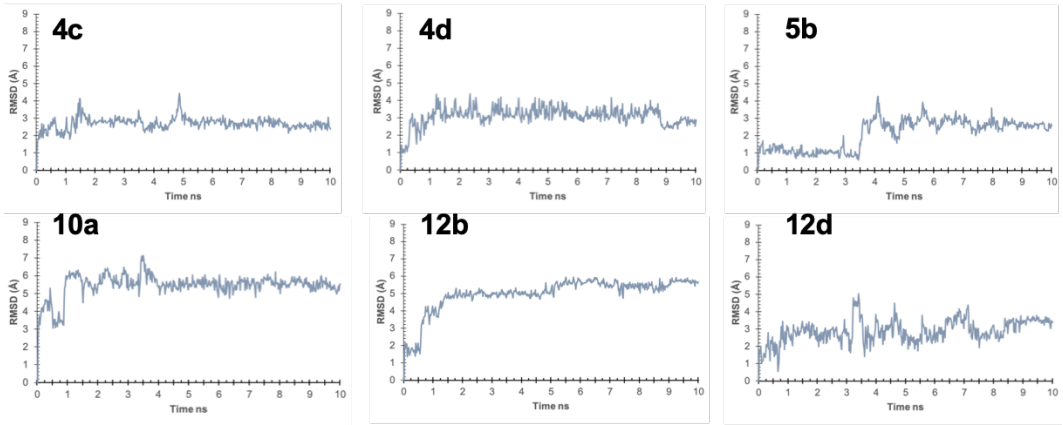


Figure 2.95. *Evolution of the RMSD for complexes of **4c**, **4d**, **5b**, **10a**, **12b** and **12d** with PTPRZ1 along the 10 ns MD simulations.*

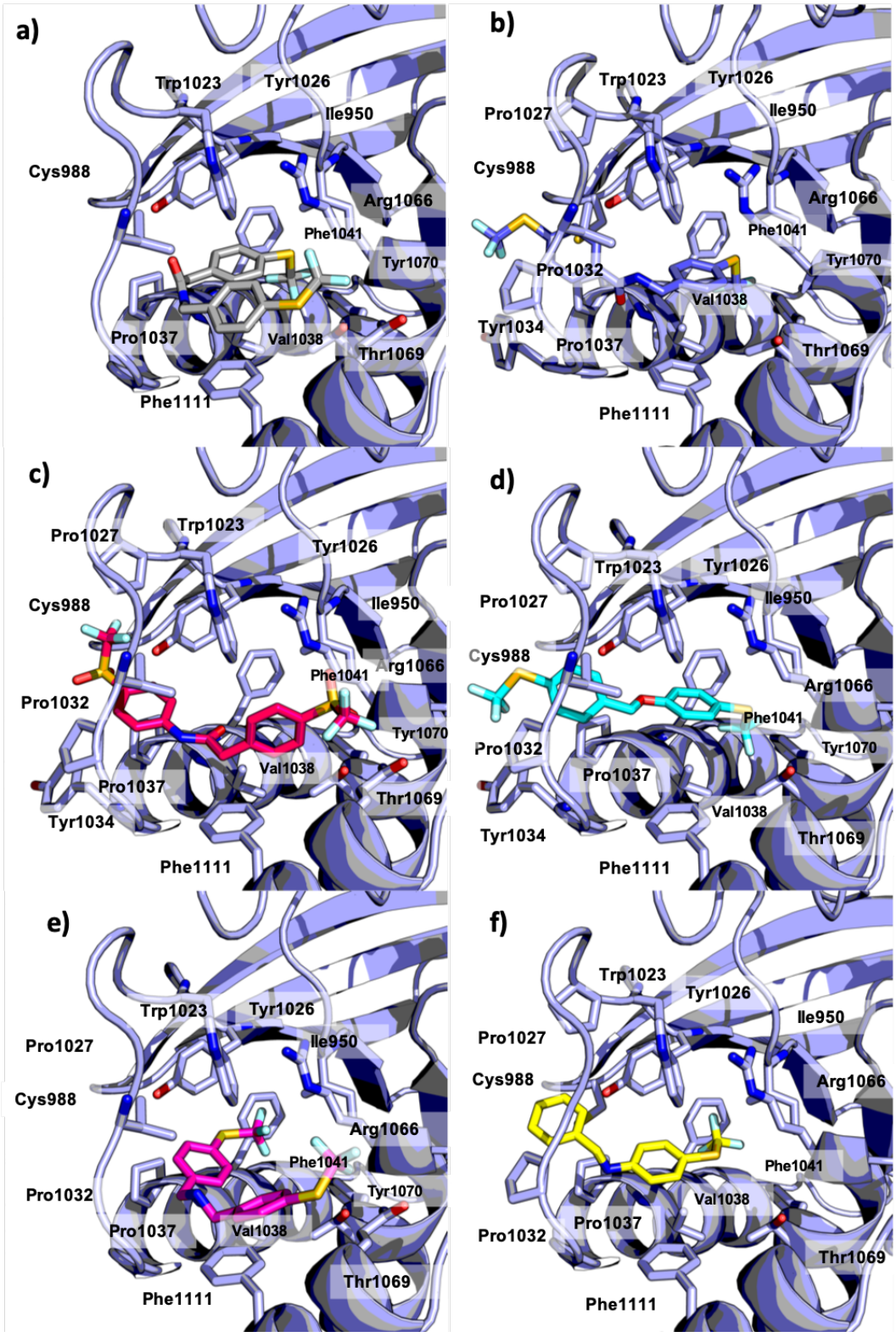


Figure 2.96. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)***, *docked inside. the force-opened structure of PTPRG issued from the MD simulation with organic solvents.*

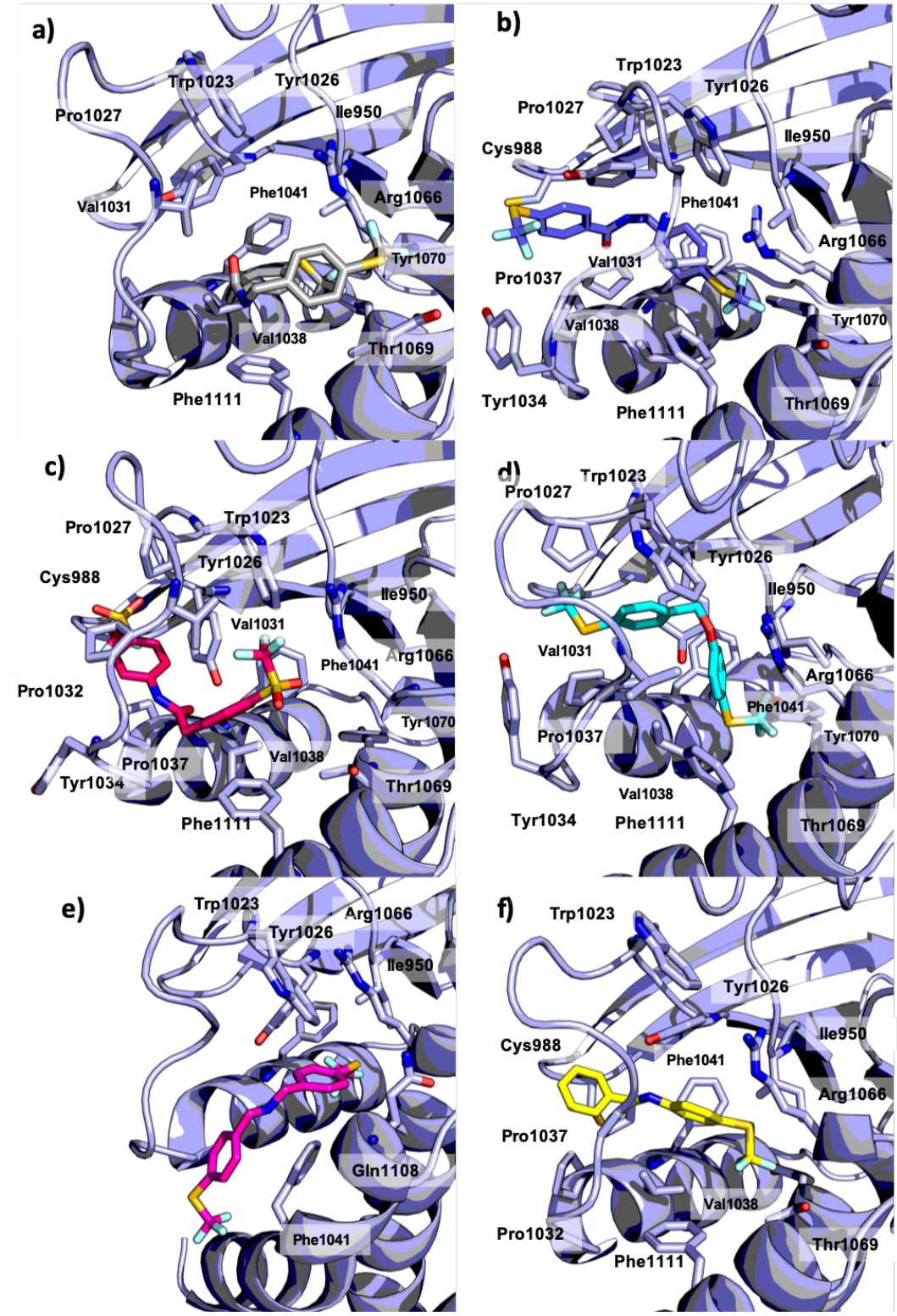


Figure 2.97. *Most populated conformers for complexes of* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)*** *and* ***12d (f)*** *along the 10 ns MD simulation using as starting conformations the complexes obtained in the docking with the force-opened structure of PTPRG issued from the MD simulation with organic solvents.*

2.15.4. Rational for the selective profile of PTPZ1 and PTPRG

The obtained data of the inhibition of PTPRZ1 and PTPRG summarized in the **Table 2.18** showed that compounds **4c**, **10a** and **12b** are more active against PTPRZ1 than PTPRG, which is indicative of a selectivity profile of those molecules. It is well established that the results obtained in an MD simulation are strongly dependent on the starting conformation [327,328].

The results shown in **sections 2.15.2** and **2.15.3** aimed to compare the different binding modes to rationalize the in vitro data. Of all of three PTPRZ1 and PTPRG conformations, the higher total binding energy come mainly from the complexes formed with the force-opened structures issued from the simulations in organic solvent mixtures (**Table 2.19**). This could be explained because the forced open conformation with organic solvent molecules, provides a bigger pocket where molecules can interact and accommodate with mote degrees of freedom. This would cause the bound compounds to explore and settle within the binding pocket. This is clearly visible for the two most active compounds **10a** and **12b** (IC_{50} values in **Table 2.9**) that exhibited the highest binding energies. Interestingly, total binding energy values are higher in PTPRZ1 than in PTPRG, which could account for the proposed selectivity of these compounds for PTPRZ1. From a structural point of view, this higher total binding energy could be explained by the interactions the compounds established with Tyr1864 in PTPRZ1. These hydrophobic interactions are not established in PTPRG as the equivalent position is occupied by Cys988. However, even if this extra interaction could account for the selectivity profile, more experimental evidence needs to be recollected to confirm this hypothesis.

The results obtained in the PTPRZ1 and PTPRG crystal structures (**5H08** and **3QCH**, respectively) showed higher binding energies for the PTPRG complexes (**Table 2.19**). This could be explained by the binding mode of the set of compounds as they were able to reach further inside the pocket than in PTPRZ1. However, the mean RMSD values and deviations suggest that complexes with PTPRZ1 are in general more stable along the

simulation compared to those of PTPRG and the solvent models (**Table 2.20**). This is probably due to the similarity of these binding modes to that of the ethoxy-trifluoromethyl-benzyl moiety of compound **7WL** crystalized inside the pocket of PTPRZ1. In this line, PTPRZ1 complexes with compounds **4c** and **10a** exhibit more stability than those with PTPRG, which is in line with the experimental data. In all, complexes with PTPRZ1 present higher stability than those with PTPRG, which could be explained by the fact that the WPD-loop seems to be more flexible in PTPRG than in PTPRZ1.

Table 2.18. *Enzyme percentage inhibition in PTPRG and PTPRZ1 at one single concentration.*

Compound	Concentration (μm)	PTPRG Inhibition (%)	PTPRZ1 Inhibition (%)
4c	0.1	24.9	37.7
10a	0.1	17.7	50.0
12b	1	10.2	97.5
12d	0.1	24.9	14.1

Table 2.19. *Total binding energy (Kcal·mol^{-1}) calculated by means of MM-ISMSA for each complex with PTPRZ1 and PTPRG.*

PTPRZ1	4c	4d	5b	10a	12b	12d
3QCH model	-45.75	-39.49	-36.02	-41.30	-40.55	-36.24
5H08	-47.45	-31.85	-32.74	-42.70	-41.68	-33.90
Solvent model	-48.39	-44.37	-40.02	-46.30	-45.67	-35.63
PTPRG	**4c**	**4d**	**5b**	**10a**	**12b**	**12d**
3QCH	-38.67	-41.50	-30.00	-39.75	-36.79	-32.69
5H08 model	-37.41	-33.43	-37.62	-38.05	-28.69	-34.56
Solvent model	-34.22	-40-92	-40.47	-40.38	-41.0	-31.58

Table 2.20. Mean RMSD values for each complex with PTPRZ1 and PTPRG.

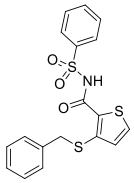
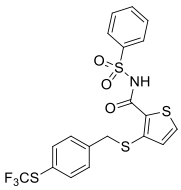
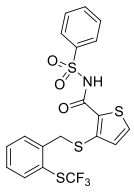
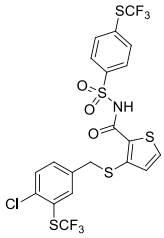
PTPRZ1	4c	4d	5b	10a	12b	12d
3QCH model	1.96	1.75	4.21	2.94	3.83	2.64
5H08	2.26	2.83	3.69	1.83	2.86	2.10
Solvent model	3.11	1.91	2.18	4.06	2.47	3.96
PTPRG	**4c**	**4d**	**5b**	**10a**	**12b**	**12d**
3QCH	3.91	3.07	5.06	4.39	3.05	2.68
5H08 model	3.27	2.04	2.41	2.56	3.51	3.37
Solvent model	2.69	3.10	2.12	5.46	4.97	2.88

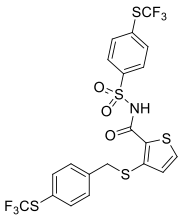
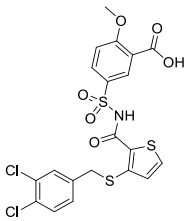
2.15.5. The selectivity profile of MPLs studied by FEP techniques

Compound **NX4** was proposed as a plausible PTPRPZ1 inhibitor at the beginning of this project. Therefore, this compound was selected as a main scaffold in order to introduce modifications in synthetically accessible positions with the aim to explore the SAR in an attempt to increase the binding affinity. **NX4** was synthesized as reference compound for the in vitro assays (**MPLNX4**) and based on the experience in our research group the trifluoromethyl-sulfur moiety was chosen as a main substituent due to its chemical accessibility and was intended to mimic the interactions established by the methoxy group present in **NX4**. All the synthetic work related to the **NX4**-derivatives has been carried out by Dr. Jose María Zapico and Dr. Mateusz Daśko. The synthesized molecules were tested in vitro against both PTPRG and PTPRZ1 at only one concentration of 10 μM. The chemical structures of the synthesized derivatives and the in vitro results are summarized in **Table 2.21**.

The results obtained for compound **MPLNX4** shows a distorted activity. In the reference publication by *Sheriff et al* this compound showed an IC_{50} of 0.5 μM in PTPRG [309]. However, in our experiments carried out at the same protein concentration but using 20 times more concentration of **MPLNX4** we only obtained a 67.5% percentage of inhibition. In spite of the different IC_{50} values, **NX4** shows better activity for PTPRG than PTPRZ1. Interestingly, compound **7WL** synthesized and evaluated by *Fujikawa et al*, also shows slightly higher activity in PTPRG than in PTPRZ1 [102].

Table 2.21. *Inhibition percentage of PTPRG and PTPRZ1 by MPL derivatives at 10uM*

Compound	Structure	PTPRG Inhibtion (%)	PTPRZ1 Inhibition (%)
MPL18		27.0	X
MPL20		13.8	X
MPL24		29.3	-
MPL30		17.1	-

MPL31		27.4	X
MPLNX4		67.5	34.4

None of the designed compounds presented activity against PTPRZ1, but they all exhibited moderate activity towards PTPRG, but always lower than the reference compound **NX4** in our experiments. After a careful inspection of the results, we assessed that the voluminous -SCF_3 moiety was not increasing the activity of the synthesized compounds as the activity of **MPL24** and **MPL31** (substituted at the position 2 or 4 of the benzyltioether, respectively) presented almost the same activity as compound **MPL18**. Interestingly, compounds **MPL20** and **MPL30** with a 4-SCF_3 and a 4-chlorine and 3-SCF_3 substitution, respectively, are the least potent ones of the series.

In order to rationalize these results, FEP calculations were performed on the series of compounds. The starting structures were the PTPRG **3QCJ** crystal structure in which PTPRG is bound to compound **NX4**, and a PTPRZ1 model, built using that same PDB code as template. Starting from these two structures PTPRG (**3QCJ**) and PTPRZ1 (**3QCJ** model), the growing/depleting of the different substituents was performed until the desired **MPL** molecule was achieved. The binding energies perturbations brought about by the change in the ligand structures are summarized in **Table 2.22**. Results show all MPL derivatives show a positive increment on the binding energy when compared to **NX4**, which is indicative of weaker

interactions with the target. These computational results are in line with the experimental results described above.

Table 2.22. *Percentage of inhibition activity in PTPRG and PTPRZ1 of MPL derivatives at 10uM; Increased free energy obtained by FEP methods.*

Compound	PTPRG Inhibition (%)	PTPRZ1 Inhibition (%)	PTPRG ΔG (Kcal·mol^{-1})	PTPRZ1 ΔG (Kcal·mol^{-1})
MPLNX4	67.5	34.4	-	-
MPL18	27.0	X	+18.61	+17.15
MPL20	13.8	X	+24.14	+25.82
MPL24	29.3	-	+19.78	+21.53
MPL30	17.1	-	+26.39	+25.21
MPL31	27.4	X	+18.27	+19.88

The resulting FEP analysis showed an increase of about 18 Kcal·mol^{-1} for those compounds with around 30% of inhibition (**MPL18**, **MPL24** and **MPL31**) and about 25 Kcal·mol^{-1} for those with an inhibition close to 15% (**MPL20** and **MPL30**). However, the correlation of the experimental and theoretical data is not completely linear as there is not enough experimental data to validate this comparison.

2.16. Molecular Modelling of complexes with PTP1B

Given the IC_{50} values of **4c**, **4d**, **5b**, **10a**, **12b** and **12d**, summarized in Table **2.23**, it is notable the lack of activity on PTP1B of some of the compounds (**4c**, **5b** and **10a**) while other present moderate activities (**12b** and **12d**). In order to rationalize these results molecular modelling techniques were applied as before. Firstly, these compounds were docked inside the three different conformations: closed from crystal structure 1PTY, open from

crystal 2HNP and superopen modelled from template 3QCH obtained from the previous modelling studies (section **2.12.1**).

Interestingly, for this enzyme the highest docking score was obtained for the complexes in which the compounds interacted with the closed conformation instead of the modelled superopen conformation (**Table 2.24**). Compounds bound to the closed conformation by interacting mainly with the active site by mimicking the phosphorylated substrate through the trifluoromethyl-thiol group, and by the interaction with the P-Tyr loop (Tyr46) through π-π interactions.

Table 2.23. *Calculated IC_{50} values (μM) of PTPRZ1 and PTP1B. ND = Non-Detected*

Compound	4c	4d	5b	10a	12b	12d
PTPRZ1	0.8	>100	>100	0.1	0.1	5.0
PTP1B	ND	-	ND	ND	0.7	1.0

Table 2.24*. Glide docking score (Kcal·mol^{-1}) for each compound docked into the three conformations (closed, open, superopen) of PTP1B.*

Compound	PTP1B CLOSED	PTP1B OPEN	PTP1B SUPEROPEN
4c	-5.21	-2.95	-4.27
4d	-4.71	-2.99	-4.34
5b	-4.59	-3.24	-3.65
10a	-4.84	-2.81	-4.82
12b	-4.96	-2.17	-3.94
12d	-4.94	-2.98	-4.17

Following the previous procedure carried out with PTPRZ1 and PTPRG, 10 ns MD simulations were carried out for the complexes obtained in this closed conformation (**Fig. 2.99** and **Fig. 2.100**). The corresponding total binding energy and RMSD values were obtained to further asses the stability of this complexes (**Table 2.25**).

The obtained RMSD values were relatively high, above 4Å with the exception of compound **4d**, which did not provide any detectable inhibition of PTP1B. Moreover, analyzing the RMSD plots of all the simulations it can be seen that the shifting of the RMSD values in PTP1B are comparatively higher than those obtained from PTPRZ1, with no stability of the complexes at the end of the simulation. (**Fig. 2.98**). The high selectivity displayed by **4c**, **10a**, **12b**, could be explained by the non-stability of the binding to PTP1B. The complex between **4c** and PTP1B shows weak interactions established with Tyr46, Asp48 and Val49 from the P-loop; Phe182 from the WPD-loop; Ser216 and Ala217 from the active site; and, Gln262 from the Q-loop (**Fig. 2.99**); all of which could account for the low stability along the simulation. The same behavior can be observed also for compounds **10a** and **12b** were the complexes are unstable during the simulated time. These findings reflect the reduced affinity for PTP1B exhibited by compounds **4c**, **10a** and **12b**, and can account for the selectivity profile outlined above.

Table 2.25. *Mean RMSD values with the standard deviation of compound complexes and average total binding energy (Kcal·mol^{-1}) calculated by MM-ISMSA for each compound in complex with PTP1B along the 10 ns MD simulations.*

PTP1B 1PTY Closed	4c	4d	5b	10a	12b	12d
RMSD	4.24 ±1.66	2.50 ±0.57	4.07 ±073	5.00 ±1.50	5.27 ±1.75	4.62 ±1.43
Total binding energy	-18.40	-29.28	-23.27	-22.37	-9.22	-8.79

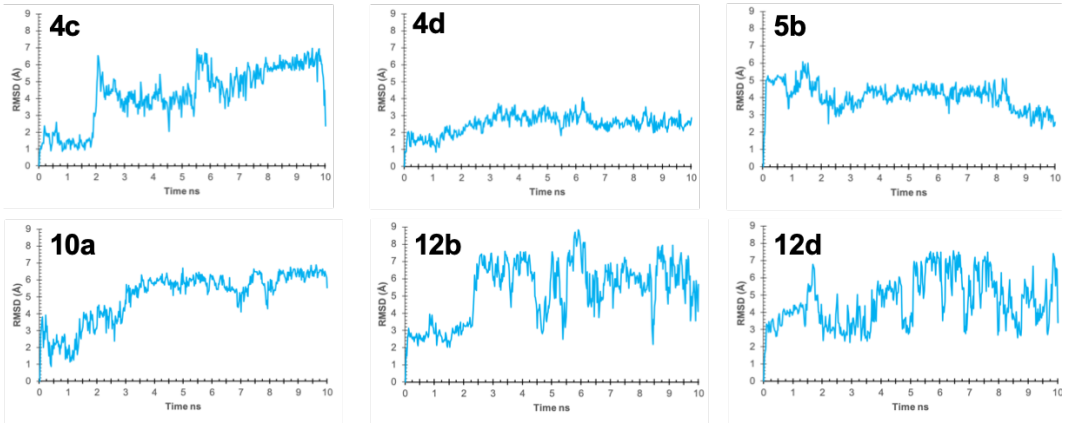


Figure 2.98. *Evolution of the RMSD for complexes of* ***4c****,* ***4d****,* ***5b****,* ***10a****,* ***12b*** *and* ***12d*** *in complex with PTP1B along the 10 ns MD simulations.*

Nevertheless, to find a plausible cause for the PTP1B inhibition shown by compounds **12b** and **12d**, the binding of the set of compounds to the allosteric pocket of PTP1B was also analyzed. Docking calculations were performed inside the allosteric pocket using as box center the **BB3** ligand of PDB code **1T48** (**Fig. 2.101**). Before that, modelling of this allosteric system was necessary due to the lack of some aminoacids missing in the crystal structure **1T48** (from D284 to D289) using Protein Preparation Wizard module from Maestro [329]. The docking score was between 2 to 3 Kcal·mol^{-1} higher in all complexes compared to the previously obtained results in the closed conformation (**Table 2.26**). The obtained binding mode was similar for all compounds, in which all compounds filled the allosteric pocket and mainly interacted through π-π stacking interactions with the side chains of Phe196 and Phe280, a T-shape interaction a with Trp291, and hydrophobic interactions within Leu192 and Val287 (**Fig. 2.103**). The MD simulations showed higher instability of these new binding modes with mean RMSD values higher than 3.3 Å for all complexes (**Fig. 2.102**). However, despite the instability shown by compounds **4c, 4d, 5b** and **10a**, compounds **12b** and **12d** rearranged and presented a stable binding mode at the end of the simulation (**Fig. 2.103**). The obtained binding pose for compound **12b** consists of a strong π-cation interaction with Phe196 and π-π interactions with Phe280 and Trp291. The same interactions were found for compound **12d** with the exception of a T-shaped interaction established with Phe196.

The total binding energy shows how the complex with compound **12b** is notably higher than the other complexes.

With this study, we have proposed a plausible binding mode for compounds **12b** and **12d** that could explain their moderate PTP1B inhibitory activity. Nevertheless, more experimental studies should be carried out to corroborate the allosteric inhibition by this type of compounds.

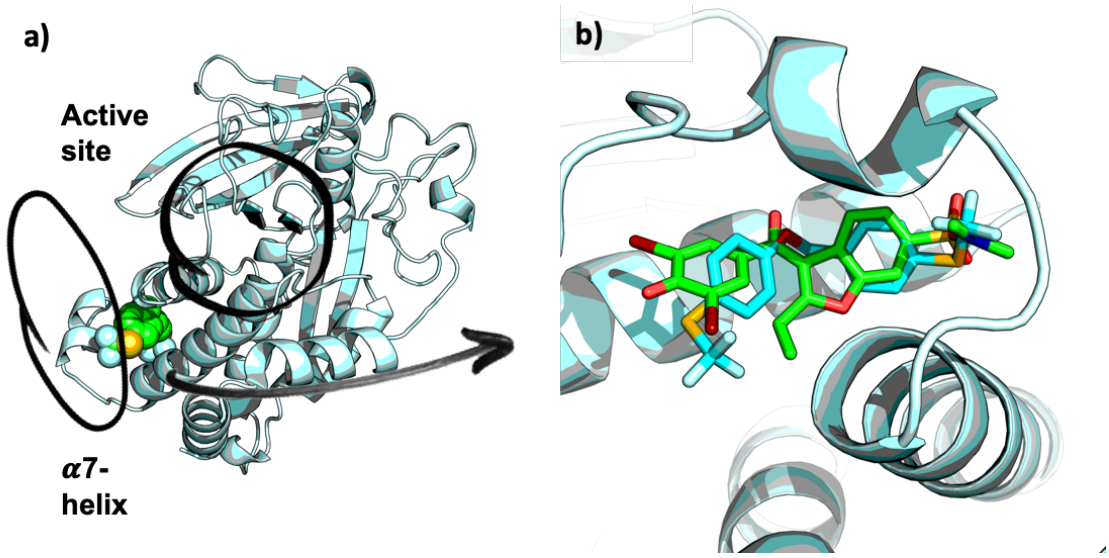


Figure 2.101. a) *General PTP1B structure showing in green the docked compound* ***10a****.* ***b)*** *Overlap of the docket compound* ***10a*** *and the reference compound* ***BB3*** *from the crystal structure* ***1T48*** *(missing aminoacids have been built).*

Table 2.26. *Glide docking scores (Kcal·mol-1) for each compound docked inside the allosteric PTP1B pocket. Mean RMSD values (in Å); Average total binding energy (Kcal.mol-1) calculated by means of MM-ISMSA for each complex.*

PTP1B allosteric 1T48	4c	4d	5b	10a	12b	12d
Docking	-8.52	-8.35	-7.81	-7.66	-8.33	-7.87
RMSD	3.7 ±0.97	4.84 ±0.96	3.62 ±0.72	3.78 ±0.57	4.99 ±0.84	3.3 ±0.57
Total Binding Energy	-37.40	-29.40	-31.96	-27.95	-40.96	-31.75

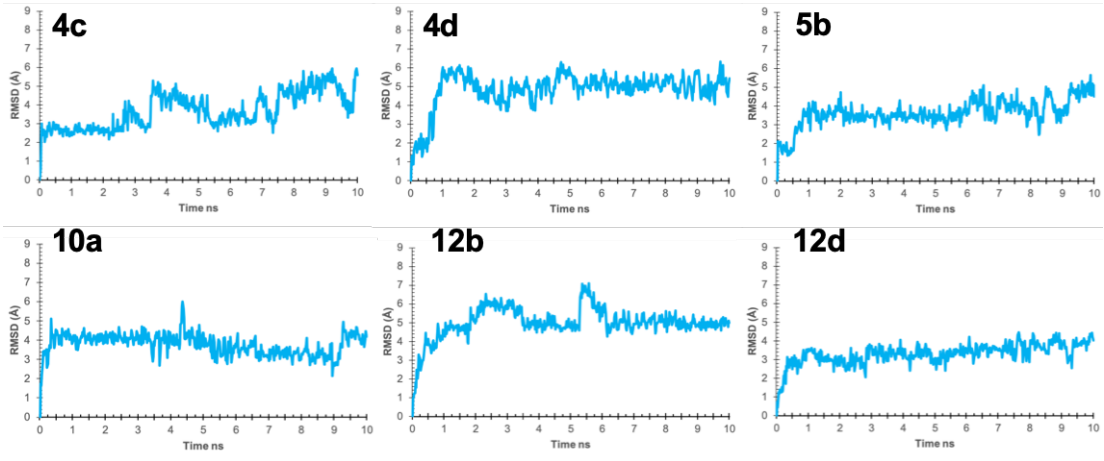


Figure 2.103. *Evolution of the root-mean-square deviations (RMSD) for complexes of* ***4c****,* ***4d****,* ***5b****,* ***10a****,* ***12b*** *and* ***12d*** *in complex with PTP1B along the 10 ns MD simulations.*

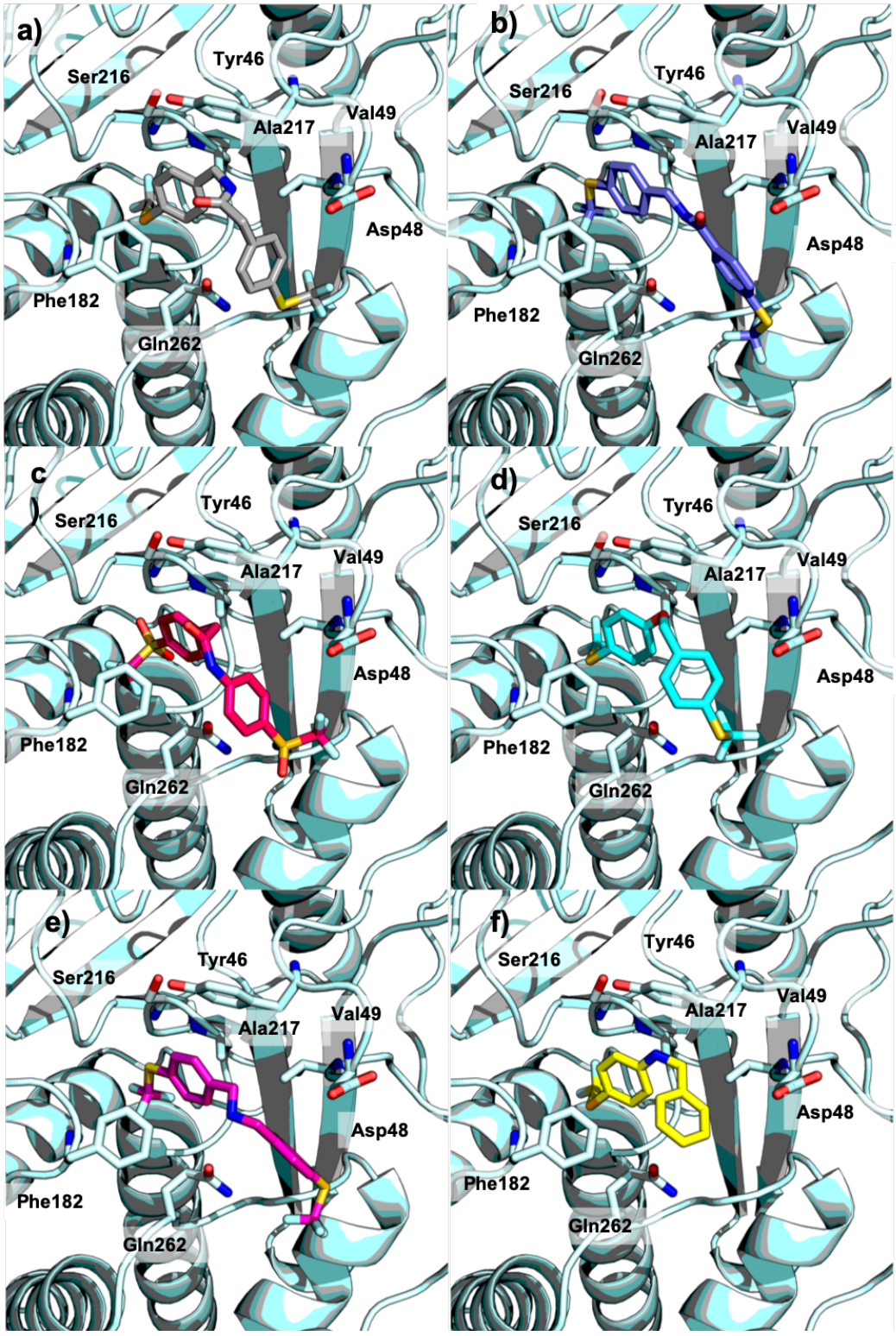


Figure 2.78. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *docked inside PTP1B* ***1PTY*** *crystal structure (closed conformation).*

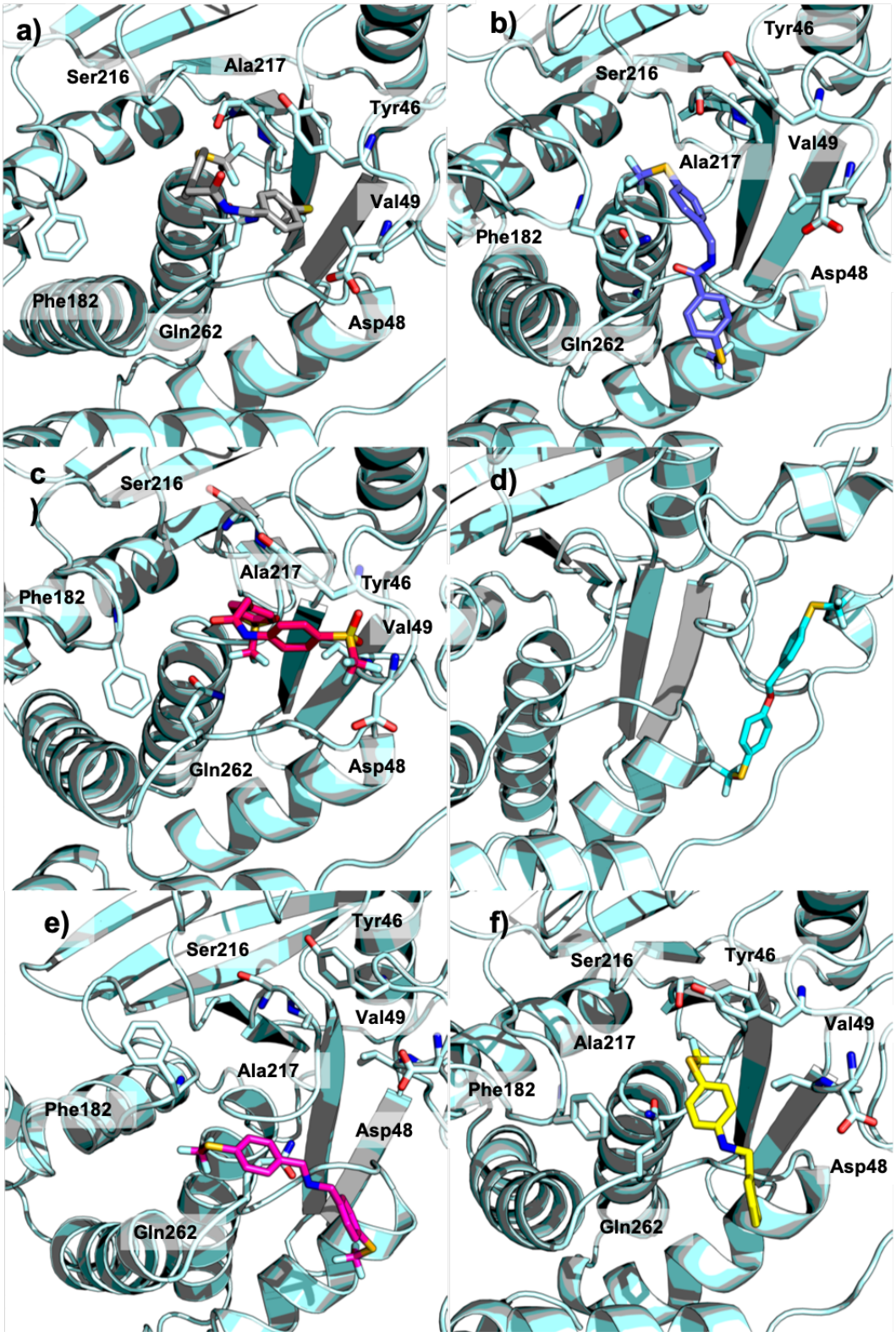


Figure 2.82. *Most populated conformers for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *inside PTP1B during the 10 ns MD simulation.*

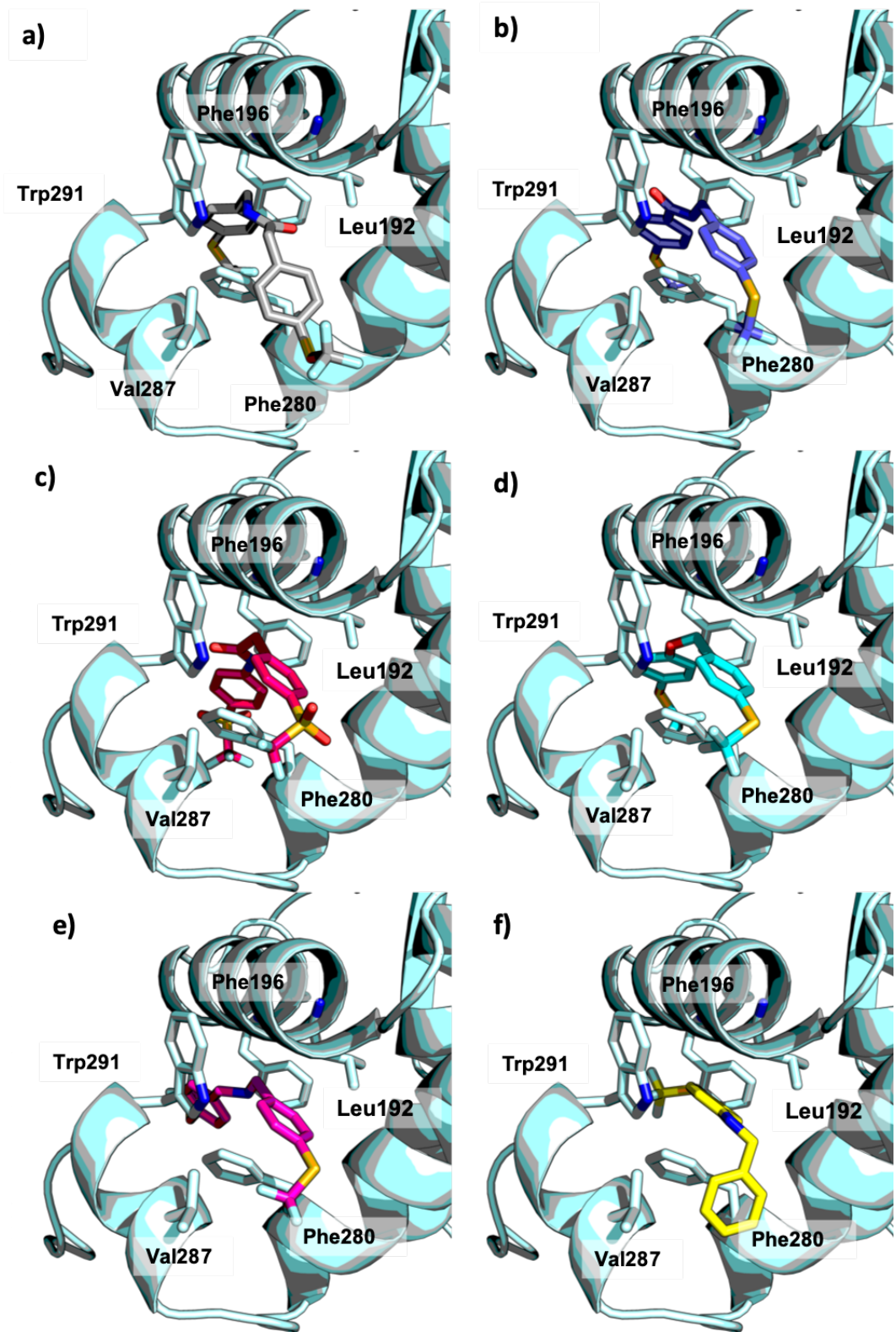


Figure 2.103. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *docked inside PTP1B from the model of* ***1T48***.

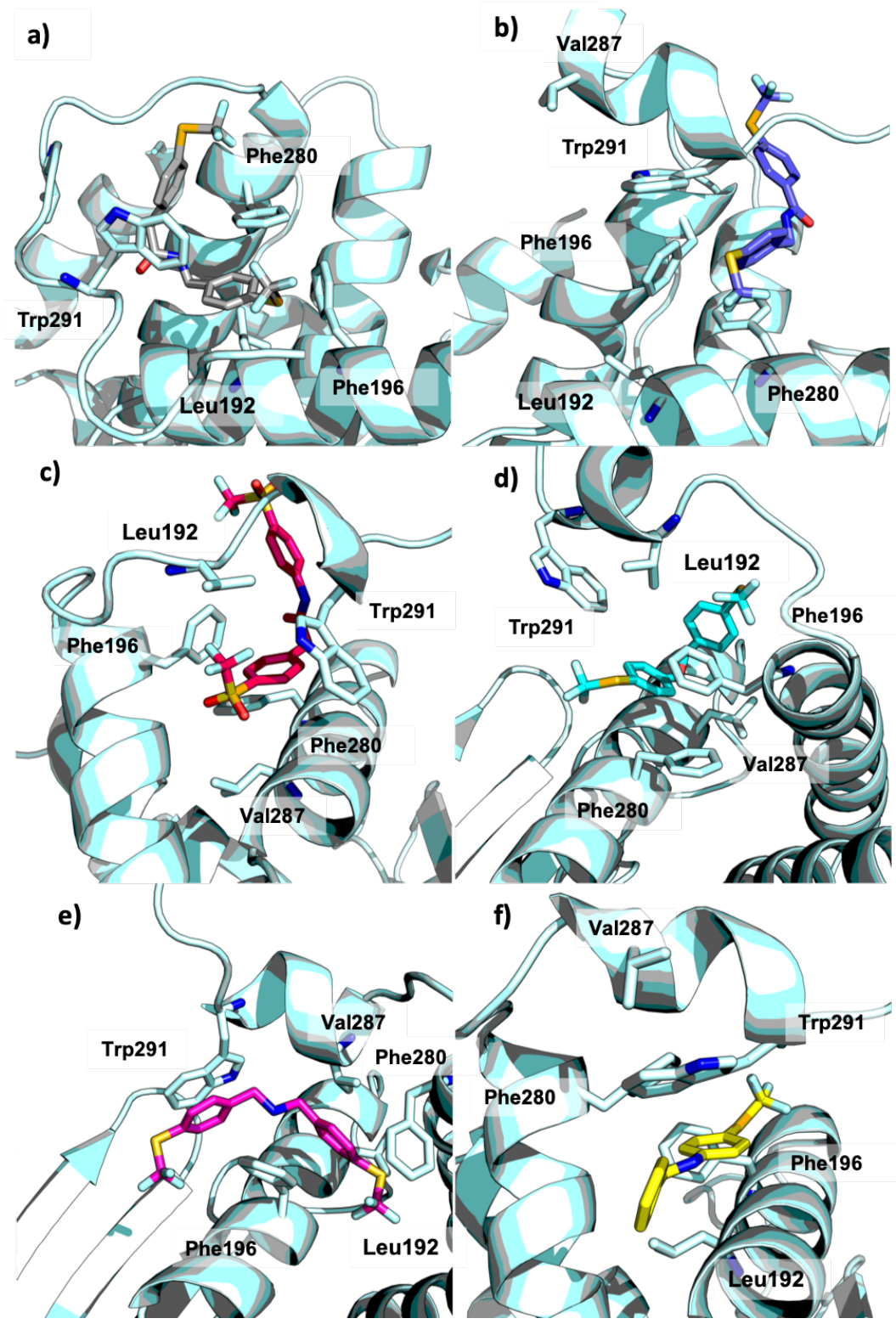


Figure 2.104. *Most populated conformers for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *inside the PTP1B* ***1T48*** *model during the 10 ns MD simulation.*

2.17. Conclusions

In this work we have demonstrated the relevance of PTPRZ1 as druggable target in the treatment of AUD. Our designed and synthesized a series of small molecules capable of mimicking the biological effect of PTN, by interacting with an allosteric site inside the intracellular domain PD1 of PTPRZ1. Our most active compounds **10a** and **12b** with and IC_{50} value of 0,1 μM are selective against PTP1B and PTPRG. These molecules produce a significant in vitro increase of the phosphorylation of proteins such as TrkA and ALK, two main PTPRZ1 substrates involved in neuronal survival and differentiation. Those substrates are involved in the mechanism of alcohol reinforcing effect. Moreover, the protective effect of these small molecules was tested in the amphetamine-induced toxicity assay carried out in PC12 cells, showing a remarkable cell survival. The capacity of these small molecules to cross the BBB was evaluated in ex vivo experiments. With all these data, in vivo experiments were carried out in mice to further support the activity of these molecules by modulating the PTPRZ1 in dopaminergic systems in a mice ethanol-drinking assay. Results show how these molecules reduce ethanol drinking and ethanol-induced conditioned place preference in mice model, confirming the role of PTPTZ1 in the reinforcing effects of alcohol and another drug abuse.

Molecular modelling techniques have allowed to propose a plausible binding mode of all these compounds inside PTPRZ1 and assess the stability of the corresponding complexes along the simulation time. More importantly, these methods have allowed explaining the remarkable PTPRZ1/PTP1B selectivity displayed by **4c** and **10a**, and the activity of **12b** and **12d** in both receptors. In addition, the selectivity profile of these molecules in the homologue PTPRG has been addressed by using different superopen conformations and models (**3QCH**, **5H08** and solvent model). MD simulations showing and pointing slight differences between PTPRZ1 and PTPRG assessed the behavior of the WPD-loop in PTPRZ1, PTPRG and PTP1B. This differences between PTPRZ1 and PTPRG, has been

highlighted during the MD simulations of the complexes and can explain the selectivity profile of our designed molecules.

In general, our compounds formed more stable complexes inside PTPRZ1 than in PTPRG. In case of the crystal structure **3QCH,** these differences were not so notable compared to the results from the PTPRZ1 **5H08** crystal structuredue to the amino acid difference Cys988/Tyr1864 (PTPRG/PTPRZ1) which can improve the interaction in favor of PTPRZ1.

The stability of the complexes obtained by using the PTPRZ1 and PTPRG structures issued from the MD simulations with organic solvents was relatively lower when compared to the original crystal structures (**3QCH or 5H08**) or the corresponding models. This fact comes mainly because the initial structures have been forced to open and in the following MD simulations, the ligands could move and rearrange within the large pocket.

Altogether, this work highlighted that presumable our molecules are able to interact in an allosteric binding mode inside PTPRZ1, occupying the hydrophobic pocket and interacting thought the arc of the WPD-loop. Nevertheless, more biological experiments need to be performed, in order to confirm this hypothesis.

Finally, the design and synthesized **NX4** analogues rendered selective PTPRG inhibitors that were confirmed by WaterLOGSY NMR binding assays. The selectivity of those molecules for PTPRG can be explained by the flexibility of the WPD-loop. In PTPRG the WPD-loop is much more flexible and can lodge larger molecules. The obtained in vitro data was correlated computationally by FEP techniques.

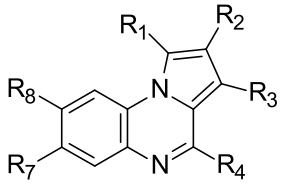
2.18. The allosteric inhibition of PTP1B by Pyrrolo[1,2–a]quinoxalines, an in silico approach

As part of a close collaboration established with the research group of Prof. Juan José Vaquero from the University of Alcalá, we have carried out molecular modeling studies on PTP1B.

This research group has synthesized and biologically evaluated a series of pyrrolo[1,2–a]quinoxalines as potent and selective inhibitors of PTP1B. Taking into account the biological data, these molecules have been proposed to act as allosteric inhibitors; therefore, our task was to validate and further rationalize the allosteric binding mode of this series of compounds. The results of this project have been submitted for publishing.

To summarize this work, we only present here the designed molecules and their inhibitory activity (**Table 2.27**), as well as the corresponding molecular modelling work. PTP1B had been crystalized with an allosteric inhibitor **BB3** and the structure deposited in the Protein Data Bank (PDB) under PDB code **1T48** (**Fig. 2.104**) [171]. This structure has been used as the target protein for docking purposes with the aim of evaluating the possibility that the series of compounds would bind the allosteric site of PTP1B. However, since PDB code **1T48** was not completely solved, as some amino acids were missing in the α7 helix, homology modelling techniques were necessary to obtain a complete target structure. The modelled structure was obtained with SWISS-MODEL web server using the **1T48** protein as template [326]. Compounds from series 1 were directly docked inside this PTP1B model.

Table 2.27. *Designed compounds and substituents added to the main scaffold (bottom). IC_{50} values for each compound are shown. Nd. Not determined.*



1

Compd.	R_1	R_2	R_3	R_4	R_7	R_8	IC_{50} ± SD (μM)
1a	H	H	H	Me	H	H	1.00 ± 0.01
1b	H	H	H	Me	Me	Me	> 10
1c	H	H	H	Me	Me	H	> 10[a]
1d	H	H	H	Me	Cl	Cl	0.57 ± 0.22
1e	H	H	H	Me	CF_3	H	> 10
1f	H	H	H	Me	OMe	H	> 10
1g	H	H	H	Me	H	Cl	0.60 ± 0.11
1h	H	H	H	Bn	H	H	0.24 ± 0.01
1i	H	H	H	Ph	H	H	0.62 ± 0.08
1j	H	H	H	H	H	H	0.88 ± 0.43
1k	Br	H	H	Me	Me	Me	Nd
1l	H	H	Br	Me	Me	Me	4.20 ± 0.98
1m	Br	Br	H	Me	Me	Me	Nd
1n	Br	H	H	Me	H	Cl	1.82 ± 0.94
1o	H	H	Br	Me	H	Cl	4.69 ± 0.43

Our results show that the tricyclic system present in the compounds occupies the site of the tribromo-benzoyl moiety of **BB3** (**Figure 2.105**), thus filling the hydrophobic pocket formed by helices α3, α6, and α7. The general binding mode was stabilized by stablishing π- π interactions with Phe196 (α3 helix) and Phe280 (α6 helix), and a parallel displaced interaction with Trp291(α7 helix). Moreover, van der Waals interactions are stablished with the side chains of Leu192 and Ile281, with a remarkable hydrogen bond between the side chain of Asn193 and the nitrogen of the quinoxaline moiety. Interestingly, compounds with halogens such as **1d**, **1m**, **1n** and **1o** occupy the same halogen positions as the reference compound (**BB3**).

To test the stability of the proposed binding modes and rationalize biological profiles (IC_{50}), 20 ns Molecular Dynamics (MD) simulations were carried out for all compounds. The ligand stability within the binding pocket, can be assessed through low RMSD (Root Mean Square Deviation) values (**Table 2.28** and **Fig. 2.105**) and agrees with a well stabilized, low energy predicted binding mode, mainly maintained by π-stacking and hydrophobic interactions. In addition, to further assess the allosteric binding mode, the stability of the helixes and the WPD-loop in the complexes was compared to that of the apo model of PTP1B and the crystal structure **2HNP** in which the α7 helix had been previously modelled. The RMSD values of the WPD-loop (**Table 2.28**) show an increased stability in most of the complexes when compared to the behaviour of the WPD-loop in the apo and **2HNP** structures. As expected for an allosteric binding mode, the WPD-loop is further stabilized upon ligand binding, and α7 remains with high mobility as experimentally observed in allosteric inhibitors crystal structures such as **1T48**, **1T49** and **1T4J** in which the atomic positions of the amino acids of the helix are not elucidated in the X-Ray structure [168]. Altogether, these results further support the with enzyme inhibition kinetics carried out for all the compounds while proposing an allosteric binding mode to PTP1B.

In order to rationalize the in vivo results, the physicochemical properties and ADME properties of the set of compounds were calculated using the SWISS-ADME web server [332] and pIC50 values from the web server tool

(www.sanjeevslab.org/tools.html) All compounds showed similar physicochemical properties and ADME properties (**Table 2.29**), with no remarkable difference between all of them, and additionally, no potential PAINS (pan assay interference compounds) were found [333]. The iLOGP value [334] was taken as reference and the results (**Table 2.29**) show that at higher value, improves the glucose uptake. Ligand efficiency index [335] were also calculated, using the binding efficiency index (BEI) and percentage efficiency index (PEI) based on the molecular weight (MW) and van der Waals polar surface area (PSA); unfortunately no correlated results were found between those values and the in vivo data.

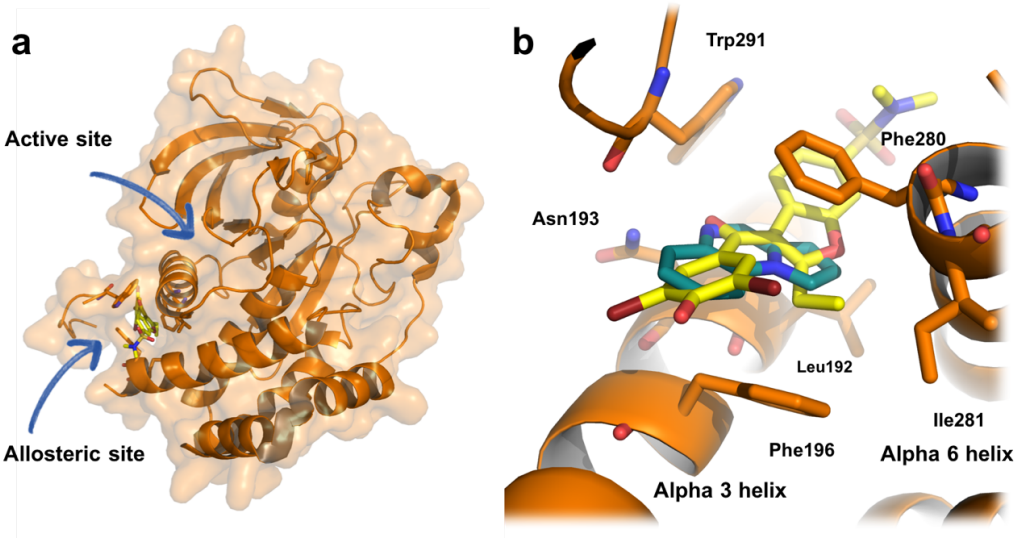


Figure 2.105. a) *Crystal structure* ***1T48*** *pointing the active and allosteric sites, also the lack of missing aminoacids can be observed.* ***b)*** *Superimposition of docked* ***1a*** *(pale blue) inside PTP1B model with* ***BB3*** *(yellow).*

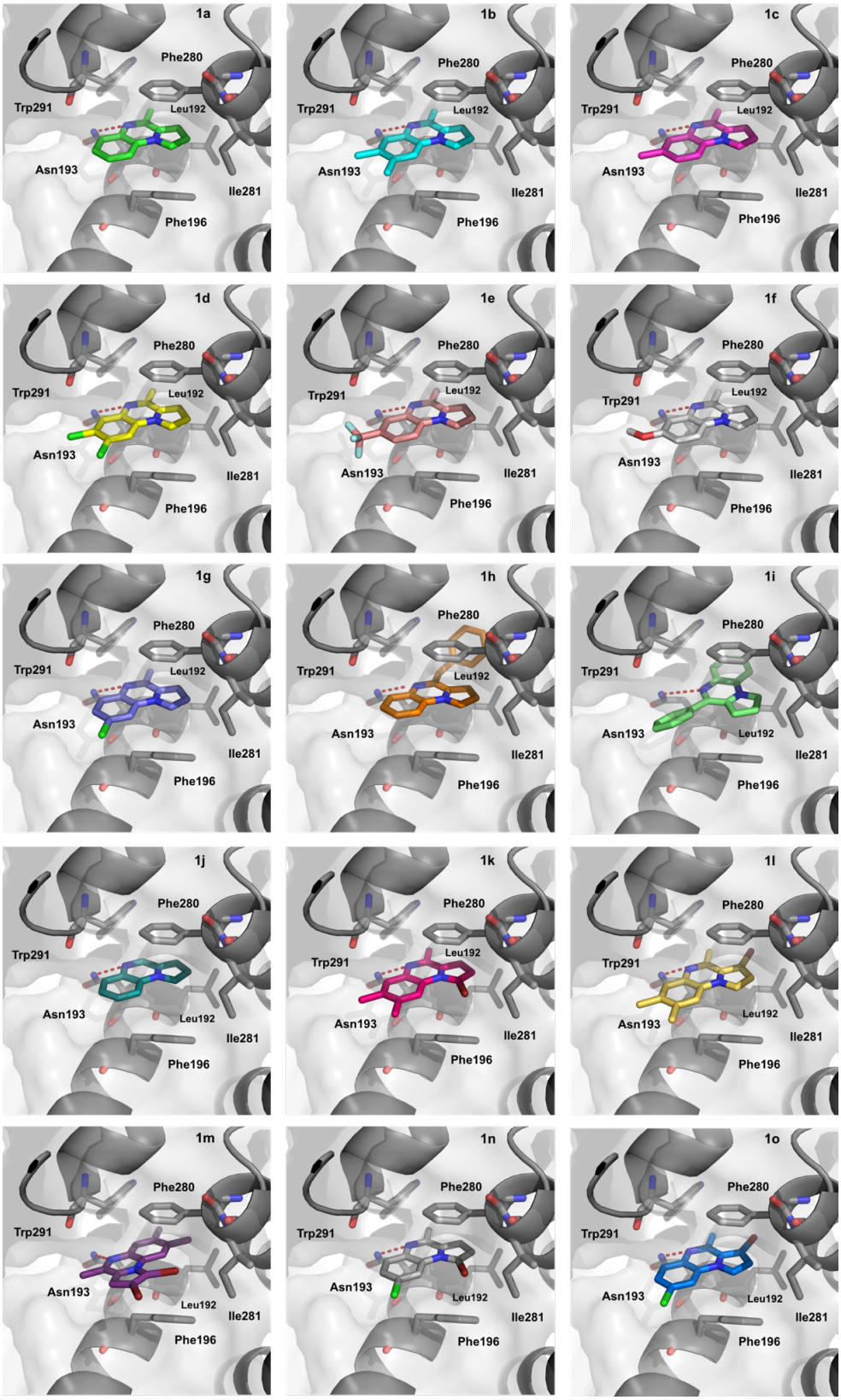


Figure 2.106. *Predicted binding modes for compounds* ***4c (a)***, ***4d (b)***, ***5b (c)***, ***10a (d)***, ***12b (e)***, ***12d (f)***, *docked inside the PTP1B* ***1T48*** *model.*

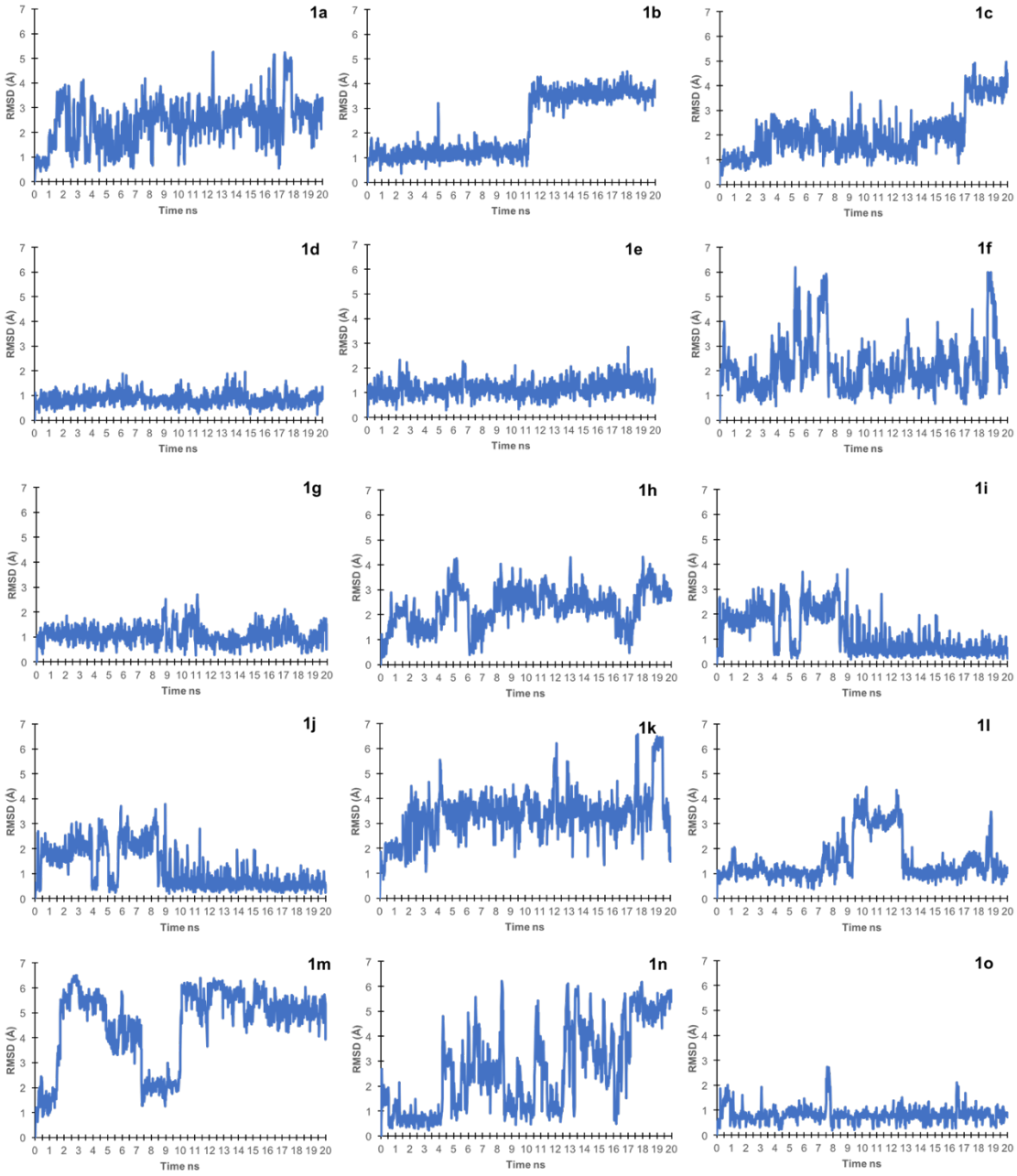


Figure 2.107. *RMSD evolution of the complexes of all compounds in complex with PTP1B along the 20-ns MD simulations.*

Table 2.28. *Mean RMSD values (in Å) for the helixes, the WPD-loop and the ligand along 20 ns MD simulation of all complexes and apo models. Calculated total binding energy (Kcal/mol) for all complexes by means of MM-ISMA.*

Compd.	Alpha 3 Helix	Alpha 6 Helix	Alpha 7 Helix	WPD-loop	Ligand	MM-ISMSA
1a	0.397	0.289	1.121	0.635	2.393	-27.19
1b	0.468	0.694	1.430	0.597	2.234	-31.01
1c	0.423	0.519	0.967	0.733	2.039	-29.7
1d	0.411	0.761	0.923	0.683	0.853	-32.60
1e	0.444	0.575	0.545	0.624	1.161	-29.46
1f	0.530	0.624	3.755	0.597	2.157	-23.86
1g	0.453	0.828	1.217	0.562	1.081	-29.41
1h	0.421	0.578	0.553	0.589	2.292	-33.31
1i	0.402	0.483	0.733	0.624	1.201	-34.51
1j	0.483	0.586	0.927	0.473	1.304	-24.60
1k	0.468	0.720	0.509	0.545	3.348	-30.62
1l	0.367	0.562	1.503	0.575	1.542	-33.88
1m	0.507	0.547	1.023	0.537	4.496	-36.15
1n	0.413	0.573	0.490	0.528	2.668	-29.55
1o	0.695	0.995	1.498	0.541	0.815	-34.10
1T48 mode	0.442	0.794	0.652	0.726	-	-
2HNP mode	0.3872	0.862	0.613	1.397	-	-

Table 2.29. *Docking scores from complexes and the physicochemical properties calculated. pIC50 values come from the IC50 value. BEI and SEI were calculated from the reference formula.*

Compound	Docking Score	MW (g/mol)	PSA (Å²)	iLogP	pIC50	BEI*	SEI*
1a	*-6,40*	*182.22*	*0.17*	*2.39*	*6*	*32.93*	*34.68*
1b	*-6.52*	*210.27*	*0.17*	*2.84*	*5*	*23.78*	*28.90*
1c	*-6.51*	*196.25*	*0.17*	*2.63*	*5*	*25.48*	*28.90*
1d	*-6.55*	*251.11*	*0.17*	*2.82*	*6.24*	*24.85*	*36.07*
1e	*-6.55*	*250.22*	*0.17*	*2.67*	*5*	*19.98*	*28.90*
1f	*-5.88*	*212.25*	*0.26*	*2.64*	*5*	*23.56*	*18.85*
1g	*-6.55*	*216.67*	*0.17*	*2.67*	*6.22*	*28.71*	*35.95*
1h	*-9.02*	*258.32*	*0.17*	*2.97*	*6.61*	*25.59*	*38.21*
1i	*-7.43*	*244.29*	*0.17*	*3.02*	*6.2*	*25.38*	*35.84*
1j	*-6.16*	*168.19*	*0.17*	*2.11*	*6.05*	*35.97*	*34.97*
1k	*-6.94*	*289.17*	*0.17*	*3.09*	*-*	*-*	*-*
1l	*-7.26*	*289.17*	*0.17*	*3.14*	*5.37*	*18.57*	*31.04*
1m	*-6.58*	*289.17*	*0.17*	*3.39*	*-*	*-*	*-*
1n	*-6.95*	*295.56*	*0.17*	*2.91*	*5.73*	*19.39*	*33.12*
1o	*-6.98*	*295.56*	*0.17*	*2.96*	*5.32*	*18.00*	*30.75*

**Reference values: BEI=27.0 and SEI=18.0*

2.19. References

1. Protein Hardware for Signaling. *Cell Biology*; Elsevier, 2017; p. 425–442.

2. Gonzalez-Sanchez, M.-B.; Lanucara, F.; Hardman, G. E.; Eyers, C. E. Gas-phase intermolecular phosphate transfer within a phosphohistidine phosphopeptide dimer. *Int. J. Mass Spectrom.* **2014**, *367*, 28–34.

3. Gonzalez-Sanchez, M.-B.; Lanucara, F.; Helm, M.; Eyers, C. E. Attempting to rewrite His tory: challenges with the analysis of histidine-phosphorylated peptides. *Biochem. Soc. Trans.* **2013**, *41*, 1089–1095.

4. Pertea, M.; Shumate, A.; Pertea, G.; Varabyou, A.; Breitwieser, F. P.; Chang, Y.-C.; Madugundu, A. K.; Pandey, A.; Salzberg, S. L. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **2018**, *19*, 208.

5. Willyard, C. New human gene tally reignites debate. *Nature* **2018**, *558*, 354–355.

6. Wang, T.; Birsoy, K.; Hughes, N. W.; Krupczak, K. M.; Post, Y.; Wei, J. J.; Lander, E. S.; Sabatini, D. M. Identification and characterization of essential genes in the human genome. *Science (80-.).* **2015**, *350*, 1096–1101.

7. Kim, M.-S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; et al. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581.

8. Savage, N. Proteomics: High-protein research. *Nature* **2015**, *527*, S6–S7.

9. Sacco, F.; Perfetto, L.; Castagnoli, L.; Cesareni, G. The human phosphatase interactome: An intricate family portrait. *FEBS Lett.* **2012**, *586*, 2732–2739.

10. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. Lo The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017**, *40*, 271–280.

11. Chen, S. C.-C.; Chen, F.-C.; Li, W.-H. Phosphorylated and Nonphosphorylated Serine and Threonine Residues Evolve at Different Rates in Mammals. *Mol. Biol. Evol.* **2010**, *27*, 2548–2554.

12. Lemmon, M. A.; Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **2010**, *141*, 1117–1134.

13. Manning, G. The Protein Kinase Complement of the Human Genome. *Science (80-.).* **2002**, *298*, 1912–1934.

14. Chen, M. J.; Dixon, J. E.; Manning, G. Genomics and evolution of protein phosphatases. *Sci. Signal.* **2017**, *10*, 1796–1813.

15. Tonks, N. K. Protein tyrosine phosphatases - from housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **2013**, *280*, 346–378.

16. Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 833–846.

17. Tonks, N. K. Pseudophosphatases: Grab and Hold on. *Cell* **2009**, *139*, 464–465.

18. Eyers, P. A.; Murphy, J. M. The evolving world of pseudoenzymes: proteins, prejudice and zombies. *BMC Biol.* **2016**, *14*, 98–104.

19. Firestein, R.; Cleary, M. L. Pseudo-phosphatase Sbf1 contains an N-terminal GEF homology domain that modulates its growth regulatory properties. *J. Cell Sci.* **2001**, *114*, 2921–2927.

20. Patterson, K. I.; Brummer, T.; O'brien, P. M.; Daly, R. J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* **2009**, *418*, 475–489.

21. Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. Protein Tyrosine Phosphatases in the Human Genome. *Cell* **2004**, *117*, 699–711.

22. Tonks, N. K.; Diltz, C. D.; Fischer, E. H. Characterization of the major protein-tyrosine-phosphatases of human placenta. *J. Biol. Chem.* **1988**, *263*, 6731–6737.

23. Barford, D.; Flint, A.; Tonks, N. Crystal structure of human protein tyrosine phosphatase 1B. *Science (80-.).* **1994**, *263*, 1397–1404.

24. Maheshwari, N.; Karthikeyan, C.; Trivedi, P.; Moorthy, N. S. H. N. Recent Advances in Protein Tyrosine Phosphatase 1B Targeted Drug Discovery for Type II Diabetes and Obesity. *Curr. Drug Targets* **2018**, *19*, 551–575.

25. Zhao, Y.; Wu, L.; Noh, S. J.; Guan, K.-L.; Zhang, Z.-Y. Altering the Nucleophile Specificity of a Protein-tyrosine Phosphatase-catalyzed Reaction. *J. Biol. Chem.* **1998**, *273*, 5484–5492.

26. Whittier, S. K.; Hengge, A. C.; Loria, J. P. Conformational Motions Regulate Phosphoryl Transfer in Related Protein Tyrosine Phosphatases. *Science (80-.).* **2013**, *341*, 899–903.

27. Brandão, T. A. S.; Hengge, A. C.; Johnson, S. J. Insights into the Reaction of Protein-tyrosine Phosphatase 1B. *J. Biol. Chem.* **2010**, *285*, 15874–15883.

28. Hobiger, K.; Friedrich, T. Voltage sensitive phosphatases: emerging kinship to protein tyrosine phosphatases from structure-function research. *Front. Pharmacol.* **2015**, *6*, 1–19.

29. Özcan, A.; Olmez, E. O.; Alakent, B. Effects of protonation state of Asp181 and position of active site water molecules on the conformation of PTP1B. *Proteins Struct. Funct. Bioinforma.* **2013**, *81*, 788–804.

30. Puius, Y. A.; Zhao, Y.; Sullivan, M.; Lawrence, D. S.; Almo, S. C.; Zhang, Z.-Y. Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: A paradigm for inhibitor design. *Proc. Natl. Acad. Sci.* **1997**, *94*, 13420–13425.

31. Kenny, P. W.; Newman, J.; Peat, T. S. Nitrate in the active site of protein tyrosine phosphatase 1B is a putative mimetic of the transition state. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *70*, 565–571.

32. Hale, A. J.; ter Steege, E.; den Hertog, J. Recent advances in understanding the role of protein-tyrosine phosphatases in development and disease. *Dev. Biol.* **2017**, *428*, 283–292.

33. Zhang, R.-Y.; Yu, Z.-H.; Zeng, L.; Zhang, S.; Bai, Y.; Miao, J.; Chen, L.; Xie, J.; Zhang, Z.-Y. SHP2 phosphatase as a novel therapeutic target for melanoma treatment. *Oncotarget* **2016**, *7*, 73817–73829.

34. LaRochelle, J. R.; Fodor, M.; Vemulapalli, V.; Mohseni, M.; Wang, P.; Stams, T.; LaMarche, M. J.; Chopra, R.; Acker, M. G.; Blacklow, S. C. Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. *Nat. Commun.* **2018**, *9*, 4508–4518.

35. Dong, L.; Yu, W.-M.; Zheng, H.; Loh, M. L.; Bunting, S. T.; Pauly, M.; Huang, G.; Zhou, M.; Broxmeyer, H. E.; Scadden, D. T.; et al. Leukaemogenic effects of Ptpn11 activating mutations in the stem cell microenvironment. *Nature* **2016**, *539*, 304–308.

36. Chen, P.-C.; Yin, J.; Yu, H.-W.; Yuan, T.; Fernandez, M.; Yung, C. K.; Trinh, Q. M.; Peltekova, V. D.; Reid, J. G.; Tworog-Dube, E.; et al. Next-generation sequencing identifies rare variants associated with Noonan syndrome. *Proc. Natl. Acad. Sci.* **2014**, *111*, 11473–11478.

37. Cordeddu, V.; Yin, J. C.; Gunnarsson, C.; Virtanen, C.; Drunat, S.; Lepri, F.; De Luca, A.; Rossi, C.; Ciolfi, A.; Pugh, T. J.; et al. Activating Mutations Affecting the Dbl Homology Domain of SOS2 Cause Noonan Syndrome. *Hum. Mutat.* **2015**, *36*, 1080–1087.

38. Quarmyne, M.; Doan, P. L.; Himburg, H. A.; Yan, X.; Nakamura, M.; Zhao, L.; Chao, N. J.; Chute, J. P. Protein tyrosine phosphatase–σ regulates hematopoietic stem cell–repopulating capacity. *J. Clin. Invest.* **2015**, *125*, 177–182.

39. Jiang, M.; Zheng, C.; Shou, P.; Li, N.; Cao, G.; Chen, Q.; Xu, C.; Du, L.; Yang, Q.; Cao, J.; et al. SHP1 Regulates Bone Mass by Directing Mesenchymal Stem Cell Differentiation. *Cell Rep.* **2016**, *16*, 769–780.

40. Lodeiro, M.; Alén, B. O.; Mosteiro, C. S.; Beiroa, D.; Nogueiras, R.; Theodoropoulou, M.; Pardo, M.; Gallego, R.; Pazos, Y.; Casanueva, F. F.; et al. The SHP-1 protein tyrosine phosphatase negatively modulates Akt signaling in the ghrelin/GHSR1a system. *Mol. Biol. Cell* **2011**, *22*, 4182–4191.

41. Rahimi, N. Vascular endothelial growth factor receptors: Molecular mechanisms of activation and therapeutic potentials. *Exp. Eye Res.* **2006**, *83*, 1005–1016.

42. Lee, S.; Rho, S.-S.; Park, H.; Park, J. A.; Kim, J.; Lee, I.-K.; Koh, G. Y.; Mochizuki, N.; Kim, Y.-M.; Kwon, Y.-G. Carbohydrate-binding protein CLEC14A regulates VEGFR-2– and VEGFR-3–dependent signals during angiogenesis and lymphangiogenesis. *J. Clin. Invest.* **2016**, *127*, 457–471.

43. Corti, F.; Simons, M. Modulation of VEGF receptor 2 signaling by protein phosphatases. *Pharmacol. Res.* **2017**, *115*, 107–123.

44. Zhang, Y.; Li, Q.; Youn, J. Y.; Cai, H. Protein Phosphotyrosine Phosphatase 1B (PTP1B) in Calpain-dependent Feedback Regulation of Vascular Endothelial Growth Factor Receptor (VEGFR2) in Endothelial Cells. *J. Biol. Chem.* **2017**, *292*, 407–416.

45. Zhang, J.; Li, L.; Li, J.; Liu, Y.; Zhang, C.-Y.; Zhang, Y.; Zen, K. Protein Tyrosine Phosphatase 1B Impairs Diabetic Wound Healing Through Vascular Endothelial Growth Factor Receptor 2 Dephosphorylation. *Arterioscler. Thromb. Vasc. Biol.* **2015**, *35*, 163–174.

46. Thiebaut, P.-A.; Besnier, M.; Gomez, E.; Richard, V. Role of protein tyrosine phosphatase 1B in cardiovascular diseases. *J. Mol. Cell. Cardiol.* **2016**, *101*, 50–57.

47. Shen, J.; Frye, M.; Lee, B. L.; Reinardy, J. L.; McClung, J. M.; Ding, K.; Kojima, M.; Xia, H.; Seidel, C.; Silva, R. L. e; et al. Targeting VE-PTP activates TIE2 and stabilizes the ocular vasculature. *J. Clin. Invest.* **2014**, *124*, 4564–4576.

48. Hayashi, M.; Majumdar, A.; Li, X.; Adler, J.; Sun, Z.; Vertuani, S.; Hellberg, C.; Mellberg, S.; Koch, S.; Dimberg, A.; et al. VE-PTP regulates VEGFR2 activity in stalk cells to establish endothelial cell polarity and lumen formation. *Nat. Commun.* **2013**, *4*, 1672–1680.

49. Brownlie, R. J.; Zamoyska, R.; Salmond, R. J. Regulation of autoimmune and anti-tumour T-cell responses by PTPN22. *Immunology* **2018**, *154*, 377–382.

50. Bae, S.-C.; Lee, Y. H. Association between the functional PTPN22

G788A (R263Q) polymorphism and susceptibility to autoimmune diseases: A meta-analysis. *Cell. Mol. Biol.* **2018**, *64*, 46–51.

51. Heneberg, P.; Kocková, L.; Čecháková, M.; Daňková, P.; Černá, M. Autoimmunity-Associated PTPN22 Polymorphisms in Latent Autoimmune Diabetes of the Adult Differ from Those of Type 1 Diabetes Patients. *Int. Arch. Allergy Immunol.* **2018**, *177*, 57–68.

52. Zhang, J.; Zhang, F.; Niu, R. Functions of Shp2 in cancer. *J. Cell. Mol. Med.* **2015**, *19*, 2075–2083.

53. Sur, S.; Agrawal, D. K. Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. *Mol. Cell. Biochem.* **2016**, *416*, 33–46.

54. Brenner, A.; Reikvam, H.; Rye, K.; Hagen, K.; Lavecchia, A.; Bruserud, Ø. CDC25 Inhibition in Acute Myeloid Leukemia–A Study of Patient Heterogeneity and the Effects of Different Inhibitors. *Molecules* **2017**, *22*, 446–464.

55. Kimelman, D. Cdc25 and the importance of G 2 control. *Cell Cycle* **2014**, *13*, 2165–2171.

56. Zacksenhaus, E.; Liu, J. C.; Granieri, L.; Vorobieva, I.; Wang, D.-Y.; Ghanbari-Azarnier, R.; Li, H.; Ali, A.; Chung, P. E. D.; Ju, Y.; et al. CDC25 as a common therapeutic target for triple-negative breast cancer - the challenges ahead. *Mol. Cell. Oncol.* **2018**, *5*, 1481814–1481817.

57. YE, W.; XUE, J.; ZHANG, Q.; LI, F.; ZHANG, W.; CHEN, H.; HUANG, Y.; ZHENG, F. MiR-449a functions as a tumor suppressor in endometrial cancer by targeting CDC25A. *Oncol. Rep.* **2014**, *32*, 1193–1199.

58. Guillem, V.; Amat, P.; Cervantes, F.; Alvarez-Larrán, A.; Cervera, J.; Maffioli, M.; Bellosillo, B.; Collado, M.; Marugán, I.; Martínez-Ruiz, F.; et al. Functional polymorphisms in SOCS1 and PTPN22 genes correlate with the response to imatinib treatment in newly diagnosed chronic-phase chronic myeloid leukemia. *Leuk. Res.* **2012**, *36*, 174–181.

59. Hebbring, S. J.; Slager, S. L.; Epperla, N.; Mazza, J. J.; Ye, Z.; Zhou, Z.; Achenbach, S. J.; Vasco, D. A.; Call, T. G.; Rabe, K. G.; et al. Genetic evidence of PTPN22 effects on chronic lymphocytic leukemia. *Blood* **2013**, *121*, 237–238.

60. Stephens, B. J.; Han, H.; Gokhale, V.; Von Hoff, D. D. PRL phosphatases as potential molecular targets in cancer. *Mol. Cancer Ther.* **2005**, *4*, 1653–1661.

61. Shen, J.; Zhang, Y.; Yu, H.; Shen, B.; Liang, Y.; Jin, R.; Liu, X.; Shi, L.; Cai, X. Role of DUSP1/MKP1 in tumorigenesis, tumor progression and therapy. *Cancer Med.* **2016**, *5*, 2061–2068.

62. Kang, Y.-S.; Seok, H.-J.; Jeong, E.-J.; Kim, Y.; Yun, S.-J.; Min, J.-K.; Kim, S. J.; Kim, J.-S. DUSP1 induces paclitaxel resistance through the regulation of p-glycoprotein expression in human ovarian cancer cells. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 403–409.

63. Teng, F.; Xu, Z.; Chen, J.; Zheng, G.; Zheng, G.; Lv, H.; Wang, Y.; Wang, L.; Cheng, X. DUSP1 induces apatinib resistance by activating the MAPK pathway in gastric cancer. *Oncol. Rep.* **2018**, 1203–1222.

64. Scott, L. M.; Chen, L.; Daniel, K. G.; Brooks, W. H.; Guida, W. C.; Lawrence, H. R.; Sebti, S. M.; Lawrence, N. J.; Wu, J. Shp2 protein tyrosine phosphatase inhibitor activity of estramustine phosphate and its triterpenoid analogs. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 730–733.

65. Zeng, L.-F.; Zhang, R.-Y.; Yu, Z.-H.; Li, S.; Wu, L.; Gunawan, A. M.; Lane, B. S.; Mali, R. S.; Li, X.; Chan, R. J.; et al. Therapeutic Potential of Targeting the Oncogenic SHP2 Phosphatase. *J. Med. Chem.* **2014**, *57*, 6594–6609.

66. Chen, Y.-N. P.; LaMarche, M. J.; Chan, H. M.; Fekkes, P.; Garcia-Fortanet, J.; Acker, M. G.; Antonakos, B.; Chen, C. H.-T.; Chen, Z.; Cooke, V. G.; et al. Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **2016**, *535*, 148–152.

67. Tai, W.-T.; Shiau, C.-W.; Chen, P.-J.; Chu, P.-Y.; Huang, H.-P.; Liu, C.-Y.; Huang, J.-W.; Chen, K.-F. Discovery of novel src homology region 2 domain-containing phosphatase 1 agonists from sorafenib for the treatment of hepatocellular carcinoma. *Hepatology* **2014**, *59*, 190–201.

68. Fan, L.-C.; Teng, H.-W.; Shiau, C.-W.; Lin, H.; Hung, M.-H.; Chen, Y.-L.; Huang, J.-W.; Tai, W.-T.; Yu, H.-C.; Chen, K.-F. SHP-1 is a target of regorafenib in colorectal cancer. *Oncotarget* **2014**, *5*, 6243–6251.

69. Liu, C.-Y.; Chen, K.-F.; Chao, T.-I.; Chu, P.-Y.; Huang, C.-T.; Huang, T.-T.; Yang, H.-P.; Wang, W.-L.; Lee, C.-H.; Lau, K.-Y.; et al. Sequential combination of docetaxel with a SHP-1 agonist enhanced suppression of p-STAT3 signaling and apoptosis in triple negative breast cancer cells. *J. Mol. Med.* **2017**, *95*, 965–975.

70. Chung, S.-Y.; Chen, Y.-H.; Lin, P.-R.; Chao, T.-C.; Su, J.-C.; Shiau, C.-W.; Su, Y. Two novel SHP-1 agonists, SC-43 and SC-78, are more potent than regorafenib in suppressing the in vitro stemness of human colorectal cancer cells. *Cell Death Discov.* **2018**, *4*, 82–94.

71. Dhillon, S. Regorafenib: A Review in Metastatic Colorectal Cancer. *Drugs* **2018**, *78*, 1133–1144.

72. Lazo, J. S.; Aslan, D. C.; Southwick, E. C.; Cooley, K. A.; Ducruet, A. P.; Joo, B.; Vogt, A.; Wipf, P. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. *J. Med. Chem.* **2001**, *44*, 4042–4049.

73. Guo, J.; Parise, R. A.; Joseph, E.; Lan, J.; Pan, S.-S.; Joo, B.; Egorin, M. J.; Wipf, P.; Lazo, J. S.; Eiseman, J. L. Pharmacology and antitumor activity of a quinolinedione Cdc25 phosphatase inhibitor DA3003-1 (NSC 663284). *Anticancer Res. 27*, 3067–3073.

74. Brezak, M.-C.; Valette, A.; Quaranta, M.; Contour-Galcera, M.-O.; Jullien, D.; Lavergne, O.; Frongia, C.; Bigg, D.; Kasprzyk, P. G.; Prevost, G. P.; et al. IRC-083864, a novel bis quinone inhibitor of CDC25 phosphatases active against human cancer cells. *Int. J. Cancer* **2009**, *124*, 1449–1456.

75. George Rosenker, K. M.; Paquette, W. D.; Johnston, P. A.; Sharlow, E. R.; Vogt, A.; Bakan, A.; Lazo, J. S.; Wipf, P. Synthesis and biological evaluation of 3-aminoisoquinolin-1(2H)-one based inhibitors of the dual-specificity phosphatase Cdc25B. *Bioorg. Med. Chem.* **2015**, *23*, 2810–2818.

76. Lund, G.; Dudkin, S.; Borkin, D.; Ni, W.; Grembecka, J.; Cierpicki, T. Inhibition of CDC25B Phosphatase Through Disruption of Protein–Protein Interaction. *ACS Chem. Biol.* **2015**, *10*, 390–394.

77. Johnston, P. A.; Foster, C. A.; Shun, T. Y.; Skoko, J. J.; Shinde, S.; Wipf, P.; Lazo, J. S. Development and Implementation of a 384-Well Homogeneous Fluorescence Intensity High-Throughput Screening Assay to Identify Mitogen-Activated Protein Kinase Phosphatase-1 Dual-Specificity Protein Phosphatase Inhibitors. *Assay Drug Dev. Technol.* **2007**, *5*, 319–332.

78. Vogt, A.; Tamewitz, A.; Skoko, J.; Sikorski, R. P.; Giuliano, K. A.; Lazo, J. S. The Benzo[c]phenanthridine Alkaloid, Sanguinarine, Is a Selective, Cell-active Inhibitor of Mitogen-activated Protein Kinase Phosphatase-1. *J. Biol. Chem.* **2005**, *280*, 19078–19086.

79. COLOMBO, M. PHARMACOLOGICAL ACTIVITIES OFCHELIDONIUM MAJUSL. (PAPAVERACEAE). *Pharmacol. Res.* **1996**, *33*, 127–134.

80. Pathak, M. K.; Dhawan, D.; Lindner, D. J.; Borden, E. C.; Farver, C.; Yi, T. Pentamidine is an inhibitor of PRL phosphatases with anticancer activity. *Mol. Cancer Ther.* **2002**, *1*, 1255–64.

81. Ahn, J. H.; Kim, S. J.; Park, W. S.; Cho, S. Y.; Ha, J. Du; Kim, S. S.; Kang, S. K.; Jeong, D. G.; Jung, S.-K.; Lee, S.-H.; et al. Synthesis and biological evaluation of rhodanine derivatives as PRL-3 inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2996–2999.

82. Daouti, S.; Li, W.; Qian, H.; Huang, K.-S.; Holmgren, J.; Levin, W.; Reik, L.; McGady, D. L.; Gillespie, P.; Perrotta, A.; et al. A Selective Phosphatase of Regenerating Liver Phosphatase Inhibitor Suppresses Tumor Cell Anchorage-Independent Growth by a Novel Mechanism Involving p130Cas Cleavage. *Cancer Res.* **2008**, *68*, 1162–1169.

83. Salamoun, J. M.; McQueeney, K. E.; Patil, K.; Geib, S. J.; Sharlow, E.

R.; Lazo, J. S.; Wipf, P. Photooxygenation of an amino-thienopyridone yields a more potent PTP4A3 inhibitor. *Org. Biomol. Chem.* **2016**, *14*, 6398–6402.

84. Zhang, Z.; Kozlov, G.; Chen, Y. S.; Gehring, K. Mechanism of thienopyridone and iminothienopyridinedione inhibition of protein phosphatases. *Medchemcomm* **2019**, *10*, 791–799.

85. Thura, M.; Al-Aidaroos, A. Q. O.; Yong, W. P.; Kono, K.; Gupta, A.; Lin, Y. Bin; Mimura, K.; Thiery, J. P.; Goh, B. C.; Tan, P.; et al. PRL3-zumab, a first-in-class humanized antibody for cancer therapy. *JCI Insight* **2016**, *1*, 87607–87622.

86. Fujikawa, A.; Sugawara, H.; Tanga, N.; Ishii, K.; Kuboyama, K.; Uchiyama, S.; Suzuki, R.; Noda, M. A head-to-toe dimerization has physiological relevance for ligand-induced inactivation of protein tyrosine receptor type Z. *J. Biol. Chem.* **2019**.

87. Bouyain, S.; Watkins, D. J. The protein tyrosine phosphatases PTPRZ and PTPRG bind to distinct members of the contactin family of neural recognition molecules. *Proc. Natl. Acad. Sci.* **2010**, *107*, 2443–2448.

88. Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D. Human Plasma N -Glycoproteome Analysis by Immunoaffinity Subtraction, Hydrazide Chemistry, and Mass Spectrometry. *J. Proteome Res.* **2005**, *4*, 2070–2080.

89. Onyango, P.; Lubyova, B.; Gardellin, P.; Kurzbauer, R.; Weith, A. Molecular Cloning and Expression Analysis of Five Novel Genes in Chromosome 1p36. *Genomics* **1998**, *50*, 187–198.

90. Fujiwara, K.; Horiguchi, K.; Maliza, R.; Tofrizal, A.; Batchuluun, K.; Ramadhani, D.; Syaidah, R.; Tsukada, T.; Azuma, M.; Kikuchi, M.; et al. Expression of the heparin-binding growth factor midkine and its receptor, Ptprz1, in adult rat pituitary. *Cell Tissue Res.* **2015**, *359*, 909–914.

91. Fukada, M.; Fujikawa, A.; Chow, J. P. H.; Ikematsu, S.; Sakuma, S.; Noda, M. Protein tyrosine phosphatase receptor type Z is inactivated by ligand-induced oligomerization. *FEBS Lett.* **2006**, *580*, 4051–4056.

92. Barr, A. J.; Ugochukwu, E.; Lee, W. H.; King, O. N. F.; Filippakopoulos, P.; Alfano, I.; Savitsky, P.; Burgess-Brown, N. A.; Müller, S.; Knapp, S. Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. *Cell* **2009**, *136*, 352–363.

93. Brembeck, F. H.; Rosário, M.; Birchmeier, W. Balancing cell adhesion and Wnt signaling, the key role of β-catenin. *Curr. Opin. Genet. Dev.* **2006**, *16*, 51–59.

94. Wolf, R. M.; Wilkes, J. J.; Chao, M. V; Resh, M. D. Tyrosine phosphorylation of p190 RhoGAP by Fyn regulates oligodendrocyte

differentiation. *J. Neurobiol.* **2001**, *49*, 62–78.

95. Taniguchi, S.; Liu, H.; Nakazawa, T.; Yokoyama, K.; Tezuka, T.; Yamamoto, T. p250GAP, a neural RhoGAP protein, is associated with and phosphorylated by Fyn. *Biochem. Biophys. Res. Commun.* **2003**, *306*, 151–155.

96. Pariser, H.; Herradon, G.; Ezquerra, L.; Perez-Pinera, P.; Deuel, T. F. Pleiotrophin regulates serine phosphorylation and the cellular distribution of -adducin through activation of protein kinase C. *Proc. Natl. Acad. Sci.* **2005**, *102*, 12407–12412.

97. Pariser, H.; Perez-Pinera, P.; Ezquerra, L.; Herradon, G.; Deuel, T. F. Pleiotrophin stimulates tyrosine phosphorylation of β-adducin through inactivation of the transmembrane receptor protein tyrosine phosphatase β/ζ. *Biochem. Biophys. Res. Commun.* **2005**, *335*, 232–239.

98. Kawachi, H.; Fujikawa, A.; Maeda, N.; Noda, M. Identification of GIT1/Cat-1 as a substrate molecule of protein tyrosine phosphatase / by the yeast substrate-trapping system. *Proc. Natl. Acad. Sci.* **2001**, *98*, 6593–6598.

99. Perez-Pinera, P.; Zhang, W.; Chang, Y.; Vega, J. A.; Deuel, T. F. Anaplastic Lymphoma Kinase Is Activated Through the Pleiotrophin/Receptor Protein-tyrosine Phosphatase β/ζ Signaling Pathway. *J. Biol. Chem.* **2007**, *282*, 28683–28690.

100. Lamprianou, S.; Chatzopoulou, E.; Thomas, J.-L.; Bouyain, S.; Harroch, S. A complex between contactin-1 and the protein tyrosine phosphatase PTPRZ controls the development of oligodendrocyte precursor cells. *Proc. Natl. Acad. Sci.* **2011**, *108*, 17498–17503.

101. Fujikawa, A.; Nagahira, A.; Sugawara, H.; Ishii, K.; Imajo, S.; Matsumoto, M.; Kuboyama, K.; Suzuki, R.; Tanga, N.; Noda, M.; et al. Small-molecule inhibition of PTPRZ reduces tumor growth in a rat model of glioblastoma. *Sci. Rep.* **2016**, *6*, 20473–20489.

102. Fujikawa, A.; Sugawara, H.; Tanaka, T.; Matsumoto, M.; Kuboyama, K.; Suzuki, R.; Tanga, N.; Ogata, A.; Masumura, M.; Noda, M. Targeting PTPRZ inhibits stem cell-like properties and tumorigenicity in glioblastoma cells. *Sci. Rep.* **2017**, *7*, 5609–5616.

103. Huang, P.; Ramphal, J.; Wei, J.; Liang, C.; Jallal, B.; McMahon, G.; Tang, C. Structure-Based design and discovery of novel inhibitors of protein tyrosine phosphatases. *Bioorg. Med. Chem.* **2003**, *11*, 1835–1849.

104. Xia, Z.; Ouyang, D.; Li, Q.; Li, M.; Zou, Q.; Li, L.; Yi, W.; Zhou, E. The Expression, Functions, Interactions and Prognostic Values of PTPRZ1: A Review and Bioinformatic Analysis. *J. Cancer* **2019**, *10*, 1663–1674.

105. Cressant, A.; Dubreuil, V.; Kong, J.; Kranz, T. M.; Lazarini, F.; Launay,

J.-M.; Callebert, J.; Sap, J.; Malaspina, D.; Granon, S.; et al. Loss-of-function of PTPR γ and ζ, observed in sporadic schizophrenia, causes brain region-specific deregulation of monoamine levels and altered behavior in mice. *Psychopharmacology (Berl).* **2017**, *234*, 575–587.

106. Fernández-Calle, R.; Vicente-Rodríguez, M.; Pastor, M.; Gramage, E.; Di Geronimo, B.; Zapico, J. M.; Coderch, C.; Pérez-García, C.; Lasek, A. W.; de Pascual-Teresa, B.; et al. Pharmacological inhibition of Receptor Protein Tyrosine Phosphatase β/ζ (PTPRZ1) modulates behavioral responses to ethanol. *Neuropharmacology* **2018**, *137*, 86–95.

107. Fernández-Calle, R.; Gramage, E.; Zapico, J. M.; de Pascual-Teresa, B.; Ramos, A.; Herradón, G. Inhibition of RPTPβ/ζ blocks ethanol-induced conditioned place preference in pleiotrophin knockout mice. *Behav. Brain Res.* **2019**, *369*, 111933.

108. Gaits, F.; Li, R. Y.; Ragab, A.; Selves, J.; Ragab-Thomas, J. M.; Chap, H. Implication of a protein-tyrosine-phosphatase in human lung cancer. *Cell. Mol. Biol. (Noisy-le-grand).* **1994**, *40*, 677–85.

109. Laczmanska, I.; Karpinski, P.; Gil, J.; Laczmanski, L.; Bebenek, M.; Sasiadek, M. M. High PTPRQ Expression and Its Relationship to Expression of PTPRZ1 and the Presence of KRAS Mutations in Colorectal Cancer Tissues. *Anticancer Res.* **2016**, *36*, 677–681.

110. Bourgonje, A. M.; Navis, A. C.; Schepens, J. T. G.; Verrijp, K.; Hovestad, L.; Hilhorst, R.; Harroch, S.; Wesseling, P.; Leenders, W. P. J.; Hendriks, W. J. A. J. Intracellular and extracellular domains of protein tyrosine phosphatase PTPRZ-B differentially regulate glioma cell growth and motility. *Oncotarget* **2014**, *5*, 8690–8702.

111. Norman, S. A.; Golfinos, J. G.; Scheck, A. C. Expression of a receptor protein tyrosine phosphatase in human glial tumors. *J. Neurooncol.* **1998**, *36*, 209–217.

112. Müller, S.; Kunkel, P.; Lamszus, K.; Ulbricht, U.; Lorente, G. A.; Nelson, A. M.; von Schack, D.; Chin, D. J.; Lohr, S. C.; Westphal, M.; et al. A role for receptor tyrosine phosphataseζ in glioma cell migration. *Oncogene* **2003**, *22*, 6661–6668.

113. Goldmann, T.; Otto, F.; Vollmer, E. A receptor-type protein tyrosine phosphatase PTP zeta is expressed in human cutaneous melanomas. *Folia Histochem. Cytobiol.* **2000**, *38*, 19–20.

114. Zhu, R.; Jian, X.-C.; Liu, D.-Y.; Zhou, C.; Wang, Y. Expression of PTPRZ1 in oral squamous cell carcinoma originated from oral submucous fibrosis and its clinical significance. *Shanghai J. Stomatol.* **2017**, *26*, 198–203.

115. Wu, C.-W.; Kao, H.-L.; Li, A. F.-Y.; Chi, C.-W.; Lin, W. Protein tyrosine-phosphatase expression profiling in gastric cancer tissues. *Cancer Lett.*

2006, *242*, 95–103.

116. Perez-Pinera, P.; Garcia-Suarez, O.; Menendez-Rodriguez, P.; Mortimer, J.; Chang, Y.; Astudillo, A.; Deuel, T. F. The receptor protein tyrosine phosphatase (RPTP)β/ζ is expressed in different subtypes of human breast cancer. *Biochem. Biophys. Res. Commun.* **2007**, *362*, 5–10.

117. FU, F.; XIAO, X.; ZHANG, T.; ZOU, Q.; CHEN, Z.; PEI, L.; SU, J.; YI, W. Expression of receptor protein tyrosine phosphatase ζ is a risk factor for triple negative breast cancer relapse. *Biomed. Reports* **2016**, *4*, 167–172.

118. Ma, Y.; Ye, F.; Xie, X.; Zhou, C.; Lu, W. Significance of PTPRZ1 and CIN85 expression in cervical carcinoma. *Arch. Gynecol. Obstet.* **2011**, *284*, 699–704.

119. Makinoshima, H.; Ishii, G.; Kojima, M.; Fujii, S.; Higuchi, Y.; Kuwata, T.; Ochiai, A. PTPRZ1 regulates calmodulin phosphorylation and tumor progression in small-cell lung carcinoma. *BMC Cancer* **2012**, *12*, 537–547.

120. Shang, D.; Xu, X.; Wang, D.; Li, Y.; Liu, Y. Protein tyrosine phosphatase ζ enhances proliferation by increasing β-catenin nuclear expression in VHL-inactive human renal cell carcinoma cells. *World J. Urol.* **2013**, *31*, 1547–1554.

121. Kuboyama, K.; Fujikawa, A.; Masumura, M.; Suzuki, R.; Matsumoto, M.; Noda, M. Protein Tyrosine Phosphatase Receptor Type Z Negatively Regulates Oligodendrocyte Differentiation and Myelination. *PLoS One* **2012**, *7*, 48797–48808.

122. Harroch, S.; Furtado, G. C.; Brueck, W.; Rosenbluth, J.; Lafaille, J.; Chao, M.; Buxbaum, J. D.; Schlessinger, J. A critical role for the protein tyrosine phosphatase receptor type Z in functional recovery from demyelinating lesions. *Nat. Genet.* **2002**, *32*, 411–414.

123. Huang, J. K.; Ferrari, C. C.; Monteiro de Castro, G.; Lafont, D.; Zhao, C.; Zaratin, P.; Pouly, S.; Greco, B.; Franklin, R. J. M. Accelerated Axonal Loss Following Acute CNS Demyelination in Mice Lacking Protein Tyrosine Phosphatase Receptor Type Z. *Am. J. Pathol.* **2012**, *181*, 1518–1523.

124. Kuboyama, K.; Fujikawa, A.; Suzuki, R.; Noda, M. Inactivation of Protein Tyrosine Phosphatase Receptor Type Z by Pleiotrophin Promotes Remyelination through Activation of Differentiation of Oligodendrocyte Precursor Cells. *J. Neurosci.* **2015**, *35*, 12162–12171.

125. Zwicker, S.; Bureik, D.; Bosma, M.; Martinez, G. L.; Almer, S.; Boström, E. A. Receptor-Type Protein-Tyrosine Phosphatase ζ and Colony Stimulating Factor-1 Receptor in the Intestine: Cellular Expression and Cytokine- and Chemokine Responses by Interleukin-34 and Colony Stimulating Factor-1. *PLoS One* **2016**, *11*, 167324–167340.

126. Herradon, G.; Ezquerra, L. Blocking Receptor Protein Tyrosine

Phosphatase β /ζ: A Potential Therapeutic Strategy for Parkinsons Disease. *Curr. Med. Chem.* **2009**, *16*, 3322–3329.

127. Buxbaum, J. D.; Georgieva, L.; Young, J. J.; Plescia, C.; Kajiwara, Y.; Jiang, Y.; Moskvina, V.; Norton, N.; Peirce, T.; Williams, H.; et al. Molecular dissection of NRG1-ERBB4 signaling implicates PTPRZ1 as a potential schizophrenia susceptibility gene. *Mol. Psychiatry* **2008**, *13*, 162–172.

128. Takahashi, N.; Sakurai, T.; Bozdagi-Gunal, O.; Dorr, N. P.; Moy, J.; Krug, L.; Gama-Sosa, M.; Elder, G. A.; Koch, R. J.; Walker, R. H.; et al. Increased expression of receptor phosphotyrosine phosphatase-β/ζ is associated with molecular, cellular, behavioral and cognitive schizophrenia phenotypes. *Transl. Psychiatry* **2011**, *1*, 8–18.

129. Ito, Y.; Yamada, S.; Takahashi, N.; Saito, S.; Yoshimi, A.; Inada, T.; Noda, Y.; Ozaki, N. No association between the protein tyrosine phosphatase, receptor-type, Z Polypeptide 1 (PTPRZ1) gene and schizophrenia in the Japanese population. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2008**, *147B*, 1013–1018.

130. Yahiro, K.; Niidome, T.; Kimura, M.; Hatakeyama, T.; Aoyagi, H.; Kurazono, H.; Imagawa, K.; Wada, A.; Moss, J.; Hirayama, T. Activation of Helicobacter pylori VacA Toxin by Alkaline or Acid Conditions Increases Its Binding to a 250-kDa Receptor Protein-tyrosine Phosphatase β. *J. Biol. Chem.* **1999**, *274*, 36693–36699.

131. Hirayama, T. Protein tyrosine phosphatase beta, a receptor for Helicobacter pylori vacA toxin. *Keio J. Med.* **2002**, *51 Suppl 2*, 20–23.

132. Thompson, H. C.; Osborne, C. E. Office records in the evaluation of quality of care. *Med. Care* **1976**, *14*, 294–314.

133. Yahiro, K.; Wada, A.; Yamasaki, E.; Nakayama, M.; Nishi, Y.; Hisatsune, J.; Morinaga, N.; Sap, J.; Noda, M.; Moss, J.; et al. Essential Domain of Receptor Tyrosine Phosphatase β (RPTPβ) for Interaction with Helicobacter pylori Vacuolating Cytotoxin. *J. Biol. Chem.* **2004**, *279*, 51013–51021.

134. Müller, S.; Lamszus, K.; Nikolich, K.; Westphal, M. Receptor protein tyrosine phosphatase ζ as a therapeutic target for glioblastoma therapy. *Expert Opin. Ther. Targets* **2004**, *8*, 211–220.

135. Foehr, E. D.; Lorente, G.; Kuo, J.; Ram, R.; Nikolich, K.; Urfer, R. Targeting of the Receptor Protein Tyrosine Phosphatase β with a Monoclonal Antibody Delays Tumor Growth in a Glioblastoma Model. *Cancer Res.* **2006**, *66*, 2271–2278.

136. Ulbricht, U.; Eckerich, C.; Fillbrandt, R.; Westphal, M.; Lamszus, K. RNA interference targeting protein tyrosine phosphatase ζ/receptor-type protein tyrosine phosphatase β suppresses glioblastoma growth in vitro and in vivo. *J. Neurochem.* **2006**, *98*, 1497–1506.

137. No Title.

138. Kostas, M.; Haugsten, E. M.; Zhen, Y.; Sørensen, V.; Szybowska, P.; Fiorito, E.; Lorenz, S.; Jones, N.; de Souza, G. A.; Wiedlocha, A.; et al. Protein Tyrosine Phosphatase Receptor Type G (PTPRG) Controls Fibroblast Growth Factor Receptor (FGFR) 1 Activity and Influences Sensitivity to FGFR Kinase Inhibitors. *Mol. Cell. Proteomics* **2018**, *17*, 850–870.

139. Lorenzetto, E.; Moratti, E.; Vezzalini, M.; Harroch, S.; Sorio, C.; Buffelli, M. Distribution of different isoforms of receptor protein tyrosine phosphatase γ (Ptprg-RPTP γ) in adult mouse brain: upregulation during neuroinflammation. *Brain Struct. Funct.* **2014**, *219*, 875–890.

140. Appiah, K. K.; Kostich, W. A.; Gerritz, S. W.; Huang, Y.; Hamman, B. D.; Allen, J.; Zhang, W.; Lanthorn, T. H.; Albright, C. F.; Westphal, R.; et al. A High-Throughput Screen for Receptor Protein Tyrosine Phosphatase-Selective Inhibitors. *J. Biomol. Screen.* **2011**, *16*, 476–485.

141. LaForgia, S.; Morse, B.; Levy, J.; Barnea, G.; Cannizzaro, L. A.; Li, F.; Nowell, P. C.; Boghosian-Sell, L.; Glick, J.; Weston, A. Receptor protein-tyrosine phosphatase gamma is a candidate tumor suppressor gene at human chromosome region 3p21. *Proc. Natl. Acad. Sci.* **1991**, *88*, 5036–5040.

142. van Doorn, R.; Zoutman, W. H.; Dijkman, R.; de Menezes, R. X.; Commandeur, S.; Mulder, A. A.; van der Velden, P. A.; Vermeer, M. H.; Willemze, R.; Yan, P. S.; et al. Epigenetic Profiling of Cutaneous T-Cell Lymphoma: Promoter Hypermethylation of Multiple Tumor Suppressor Genes Including BCL7a , PTPRG , and p73. *J. Clin. Oncol.* **2005**, *23*, 3886–3896.

143. Vezzalini, M.; Mafficini, A.; Tomasello, L.; Lorenzetto, E.; Moratti, E.; Fiorini, Z.; Holyoake, T. L.; Pellicano, F.; Krampera, M.; Tecchio, C.; et al. A new monoclonal antibody detects downregulation of protein tyrosine phosphatase receptor type γ in chronic myeloid leukemia patients. *J. Hematol. Oncol.* **2017**, *10*, 129–140.

144. Xiao, J.; Lee, S.-T.; Xiao, Y.; Ma, X.; Andres Houseman, E.; Hsu, L.-I.; Roy, R.; Wrensch, M.; de Smith, A. J.; Chokkalingam, A.; et al. PTPRG inhibition by DNA methylation and cooperation with RAS gene activation in childhood acute lymphoblastic leukemia. *Int. J. Cancer* **2014**, *135*, 1101–1109.

145. van Roon, E. H. J.; de Miranda, N. F. C. C.; van Nieuwenhuizen, M. P.; de Meijer, E. J.; van Puijenbroek, M.; Yan, P. S.; Huang, T. H.-M.; van Wezel, T.; Morreau, H.; Boer, J. M. Tumour-specific methylation of PTPRG intron 1 locus in sporadic and Lynch syndrome colorectal cancer. *Eur. J. Hum. Genet.* **2011**, *19*, 307–312.

146. Liu, M.; Yang, R.; Urrehman, U.; Ye, C.; Yan, X.; Cui, S.; Hong, Y.; Gu,

Y.; Liu, Y.; Zhao, C.; et al. MiR-19b suppresses PTPRG to promote breast tumorigenesis. *Oncotarget* **2016**, *7*, 64100–64108.

147. Liu, S.; Sugimoto, Y.; Sorio, C.; Tecchio, C.; Lin, Y. C. Function analysis of estrogenically regulated protein tyrosine phosphatase γ (PTPγ) in human breast cancer cell line MCF-7. *Oncogene* **2004**, *23*, 1256–1262.

148. Shu, S. T.; Sugimoto, Y.; Liu, S.; Chang, H.-L.; Ye, W.; Wang, L.-S.; Huang, Y.-W.; Yan, P.; Lin, Y. C. Function and regulatory mechanisms of the candidate tumor suppressor receptor protein tyrosine phosphatase gamma (PTPRG) in breast cancer cells. *Anticancer Res.* **2010**, *30*, 1937–1946.

149. Panagopoulos, I.; Pandis, N.; Thelin, S.; Petersson, C.; Mertens, F.; Borg, A.; Kristoffersson, U.; Mitelman, F.; Aman, P. The FHIT and PTPRG genes are deleted in benign proliferative breast disease associated with familial breast cancer and cytogenetic rearrangements of chromosome band 3p14. *Cancer Res.* **1996**, *56*, 4871–4875.

150. Cheung, A. K. L.; Lung, H. L.; Hung, S. C.; Law, E. W. L.; Cheng, Y.; Yau, W. L.; Bangarusamy, D. K.; Miller, L. D.; Liu, E. T.-B.; Shao, J.-Y.; et al. Functional Analysis of a Cell Cycle-Associated, Tumor-Suppressive Gene, Protein Tyrosine Phosphatase Receptor Type G, in Nasopharyngeal Carcinoma. *Cancer Res.* **2008**, *68*, 8137–8145.

151. Liang, F.; Lee, S.-Y.; Liang, J.; Lawrence, D. S.; Zhang, Z.-Y. The Role of Protein-tyrosine Phosphatase 1B in Integrin Signaling. *J. Biol. Chem.* **2005**, *280*, 24857–24863.

152. Stuible, M.; Tremblay, M. L. In control at the ER: PTP1B and the down-regulation of RTKs by dephosphorylation and endocytosis. *Trends Cell Biol.* **2010**, *20*, 672–679.

153. Frangioni, J. V; Oda, A.; Smith, M.; Salzman, E. W.; Neel, B. G. Calpain-catalyzed cleavage and subcellular relocation of protein phosphotyrosine phosphatase 1B (PTP-1B) in human platelets. *EMBO J.* **1993**, *12*, 4843–4856.

154. Moeslein, F. M.; Myers, M. P.; Landreth, G. E. The CLK Family Kinases, CLK1 and CLK2, Phosphorylate and Activate the Tyrosine Phosphatase, PTP-1B. *J. Biol. Chem.* **1999**, *274*, 26697–26704.

155. Ravichandran, L. V.; Chen, H.; Li, Y.; Quon, M. J. Phosphorylation of PTP1B at Ser 50 by Akt Impairs Its Ability to Dephosphorylate the Insulin Receptor. *Mol. Endocrinol.* **2001**, *15*, 1768–1780.

156. van Montfort, R. L. M.; Congreve, M.; Tisi, D.; Carr, R.; Jhoti, H. Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature* **2003**, *423*, 773–777.

157. Salmeen, A.; Andersen, J. N.; Myers, M. P.; Meng, T.-C.; Hinks, J. A.;

Tonks, N. K.; Barford, D. Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature* **2003**, *423*, 769–773.

158. Ferrari, E.; Tinti, M.; Costa, S.; Corallino, S.; Nardozza, A. P.; Chatraryamontri, A.; Ceol, A.; Cesareni, G.; Castagnoli, L. Identification of New Substrates of the Protein-tyrosine Phosphatase PTP1B by Bayesian Integration of Proteome Evidence. *J. Biol. Chem.* **2011**, *286*, 4173–4185.

159. Ide, R.; Maegawa, H.; Kikkawa, R.; Shigeta, Y.; Kashiwagi, A. High Glucose Condition Activates Protein Tyrosine Phosphatases and Deactivates Insulin Receptor Function in Insulin-Sensitive Rat 1 Fibroblasts. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 71–77.

160. Buckley, D. A.; Cheng, A.; Kiely, P. A.; Tremblay, M. L.; O'Connor, R. Regulation of Insulin-Like Growth Factor Type I (IGF-I) Receptor Kinase Activity by Protein Tyrosine Phosphatase 1B (PTP-1B) and Enhanced IGF-I-Mediated Suppression of Apoptosis and Motility in PTP-1B-Deficient Fibroblasts. *Mol. Cell. Biol.* **2002**, *22*, 1998–2010.

161. Choi, K.; Kim, Y.-B. Molecular Mechanism of Insulin Resistance in Obesity and Type 2 Diabetes. *Korean J. Intern. Med.* **2010**, *25*, 119–129.

162. Johnson, T. O.; Ermolieff, J.; Jirousek, M. R. Protein tyrosine phosphatase 1B inhibitors for diabetes. *Nat. Rev. Drug Discov.* **2002**, *1*, 696–709.

163. Myers, M. P.; Andersen, J. N.; Cheng, A.; Tremblay, M. L.; Horvath, C. M.; Parisien, J.-P.; Salmeen, A.; Barford, D.; Tonks, N. K. TYK2 and JAK2 Are Substrates of Protein-tyrosine Phosphatase 1B. *J. Biol. Chem.* **2001**, *276*, 47771–47774.

164. Zabolotny, J. M.; Bence-Hanulec, K. K.; Stricker-Krongrad, A.; Haj, F.; Wang, Y.; Minokoshi, Y.; Kim, Y.-B.; Elmquist, J. K.; Tartaglia, L. A.; Kahn, B. B.; et al. PTP1B regulates leptin signal transduction in vivo. *Dev. Cell* **2002**, *2*, 489–495.

165. St-Pierre, J.; Tremblay, M. L. Modulation of Leptin Resistance by Protein Tyrosine Phosphatases. *Cell Metab.* **2012**, *15*, 292–297.

166. Nottbohm, A. C.; Hergenrother, P. J. Phosphate Mimics: Cyclic Compounds. *Wiley Encyclopedia of Chemical Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; p. 1–16.

167. Black, E.; Breed, J.; Breeze, A. L.; Embrey, K.; Garcia, R.; Gero, T. W.; Godfrey, L.; Kenny, P. W.; Morley, A. D.; Minshull, C. A.; et al. Structure-based design of protein tyrosine phosphatase-1B inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2503–2507.

168. Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; et al. Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730–737.

169. Wrobel, J.; Sredy, J.; Moxham, C.; Dietrich, A.; Li, Z.; Sawicki, D. R.; Seestaller, L.; Wu, L.; Katz, A.; Sullivan, D.; et al. PTP1B Inhibition and Antihyperglycemic Activity in the ob/ob Mouse Model of Novel 11-Arylbenzo[b]naphtho[2,3- d]furans and 11-Arylbenzo[b]naphtho[2,3- d]thiophenes. *J. Med. Chem.* **1999**, *42*, 3199–3202.

170. Cui, W.; Cheng, Y.-H.; Geng, L.-L.; Liang, D.-S.; Hou, T.-J.; Ji, M.-J. Unraveling the Allosteric Inhibition Mechanism of PTP1B by Free Energy Calculation Based on Umbrella Sampling. *J. Chem. Inf. Model.* **2013**, *53*, 1157–1167.

171. Li, S.; Zhang, J.; Lu, S.; Huang, W.; Geng, L.; Shen, Q.; Zhang, J. The Mechanism of Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *PLoS One* **2014**, *9*, 97668–97678.

172. Krishnan, N.; Koveal, D.; Miller, D. H.; Xue, B.; Akshinthala, S. D.; Kragelj, J.; Jensen, M. R.; Gauss, C.-M.; Page, R.; Blackledge, M.; et al. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.* **2014**, *10*, 558–566.

173. Morishita, K.; Shoji, Y.; Tanaka, S.; Fukui, M.; Ito, Y.; Kitao, T.; Ozawa, S.; Hirono, S.; Shirahase, H. Novel Non-carboxylate Benzoylsulfonamide-Based Protein Tyrosine Phosphatase 1B Inhibitors with Non-competitive Actions. *Chem. Pharm. Bull. (Tokyo)*. **2017**, *65*, 1144–1160.

174. Ito, Y.; Fukui, M.; Kanda, M.; Morishita, K.; Shoji, Y.; Kitao, T.; Hinoi, E.; Shirahase, H. Therapeutic effects of the allosteric protein tyrosine phosphatase 1B inhibitor KY-226 on experimental diabetes and obesity via enhancements in insulin and leptin signaling in mice. *J. Pharmacol. Sci.* **2018**, *137*, 38–46.

175. Amit Koparde, A.; Chandrashekar Doijad, R.; Shripal Magdum, C. Natural Products in Drug Discovery. *Pharmacognosy - Medicinal Plants*; IntechOpen, 2019; p. 1–19.

176. Jiang, C.; Liang, L.; Guo, Y. Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades. *Acta Pharmacol. Sin.* **2012**, *33*, 1217–1245.

177. Chen, R. M.; Hu, L. H.; An, T. Y.; Li, J.; Shen, Q. Natural PTP1B inhibitors from Broussonetia papyrifera. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3387–3390.

178. Zhao, B. T.; Nguyen, D. H.; Le, D. D.; Choi, J. S.; Min, B. S.; Woo, M. H. Protein tyrosine phosphatase 1B inhibitors from natural sources. *Arch. Pharm. Res.* **2018**, *41*, 130–161.

179. Shi, D.; Xu, F.; He, J.; Li, J.; Fan, X.; Han, L. Inhibition of bromophenols against PTP1B and anti-hyperglycemic effect of Rhodomela confervoides extract in diabetic rats. *Sci. Bull.* **2008**, *53*, 2476–2479.

180. He, H.-B.; Gao, L.-X.; Deng, Q.-F.; Ma, W.-P.; Tang, C.-L.; Qiu, W.-W.; Tang, J.; Li, J.-Y.; Li, J.; Yang, F. Synthesis and biological evaluation of 4,4-dimethyl lithocholic acid derivatives as novel inhibitors of protein tyrosine phosphatase 1B. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 7237–7242.

181. Shechter, Y. Historic perspective and recent developments on the insulin-like actions of vanadium; toward developing vanadium-based drugs for diabetes. *Coord. Chem. Rev.* **2003**, *237*, 3–11.

182. Kiss, T.; Jakusch, T.; Hollender, D.; Dörnyei, Á.; Enyedy, É. A.; Pessoa, J. C.; Sakurai, H.; Sanz-Medel, A. Biospeciation of antidiabetic VO(IV) complexes. *Coord. Chem. Rev.* **2008**, *252*, 1153–1162.

183. Elberg, G.; Li, J.; Shechter, Y. Vanadium activates or inhibits receptor and non-receptor protein tyrosine kinases in cell-free experiments, depending on its oxidation state. Possible role of endogenous vanadium in controlling cellular protein tyrosine kinase activity. *J. Biol. Chem.* **1994**, *269*, 9521–9527.

184. Bernier, M.; Laird, D. M.; Lane, M. D. Effect of vanadate on the cellular accumulation of pp15, an apparent product of insulin receptor tyrosine kinase action. *J. Biol. Chem.* **1988**, *263*, 13626–13634.

185. Rehder, D. The role of vanadium in biology. *Metallomics* **2015**, *7*, 730–742.

186. Rehder, D. *Bioinorganic Vanadium Chemistry*; Inorganic Chemistry; John Wiley & Sons, Ltd: Chichester, UK, 2008.

187. Li, M.; Ding, W.; Smee, J. J.; Baruah, B.; Willsky, G. R.; Crans, D. C. Anti-diabetic effects of vanadium(III, IV, V)–chlorodipicolinate complexes in streptozotocin-induced diabetic rats. *BioMetals* **2009**, *22*, 895–905.

188. Thompson, K. H.; Orvig, C. Vanadium compounds in the treatment of diabetes. *Met. Ions Biol. Syst.* **2004**, *41*, 221–252.

189. Peters, K. G.; Davis, M. G.; Howard, B. W.; Pokross, M.; Rastogi, V.; Diven, C.; Greis, K. D.; Eby-Wilkens, E.; Maier, M.; Evdokimov, A.; et al. Mechanism of insulin sensitization by BMOV (bis maltolato oxo vanadium); unliganded vanadium (VO4) as the active component. *J. Inorg. Biochem.* **2003**, *96*, 321–330.

190. Scior, T.; Guevara-Garcia, J.; Do, Q.-T.; Bernard, P.; Laufer, S. Why Antidiabetic Vanadium Complexes are Not in the Pipeline of “Big Pharma” Drug Research? A Critical Review. *Curr. Med. Chem.* **2016**, *23*, 2874–2891.

191. Le Sommer, S.; Morrice, N.; Pesaresi, M.; Thompson, D.; Vickers, M. A.; Murray, G. I.; Mody, N.; Neel, B. G.; Bence, K. K.; Wilson, H. M.; et al. Deficiency in Protein Tyrosine Phosphatase PTP1B Shortens Lifespan and Leads to Development of Acute Leukemia. *Cancer Res.* **2018**, *78*, 75–87.

192. Mei, W.; Wang, K.; Huang, J.; Zheng, X. Cell Transformation by PTP1B Truncated Mutants Found in Human Colon and Thyroid Tumors. *PLoS One* **2016**, *11*, 166538–166553.

193. Wiener, J. R.; Kerns, B.-J. M.; Harvey, E. L.; Conaway, M. R.; Lglehart, J. D.; Berchuck, A.; Bast, R. C. Overexpression of the Protien Tyrosine Phosphatase PTP1B in Human Breast Cancer: Assocation With p185c-erbB-2 Protein Expression. *JNCI J. Natl. Cancer Inst.* **1994**, *86*, 372–378.

194. Soysal, S.; Obermann, E. C.; Gao, F.; Oertli, D.; Gillanders, W. E.; Viehl, C. T.; Muenst, S. PTP1B expression is an independent positive prognostic factor in human breast cancer. *Breast Cancer Res. Treat.* **2013**, *137*, 637–644.

195. Liu, X.; Chen, Q.; Hu, X.-G.; Zhang, X.-C.; Fu, T.-W.; Liu, Q.; Liang, Y.; Zhao, X.-L.; Zhang, X.; Ping, Y.-F.; et al. PTP1B promotes aggressiveness of breast cancer cells by regulating PTEN but not EMT. *Tumor Biol.* **2016**, *37*, 13479–13487.

196. Liao, S.; Li, J.; Yu, L.; Sun, S. Protein tyrosine phosphatase 1B expression contributes to the development of breast cancer. *J. Zhejiang Univ. B* **2017**, *18*, 334–342.

197. Lessard, L.; Labbe, D. P.; Deblois, G.; Begin, L. R.; Hardy, S.; Mes-Masson, A.-M.; Saad, F.; Trotman, L. C.; Giguere, V.; Tremblay, M. L. PTP1B Is an Androgen Receptor-Regulated Phosphatase That Promotes the Progression of Prostate Cancer. *Cancer Res.* **2012**, *72*, 1529–1537.

198. Pineda, M. M.; Urbina Ramírez, S.; Hinojosa, J.; De Anda Gonzalez, J.; Leon Rodriguez, E.; Bourlon, M. T. PTP1B expression and Gleason Score (GS) in localized prostate cancer (PC). *J. Clin. Oncol.* **2016**, *34*, 16566–16566.

199. Labbe, D. P.; Uetani, N.; Vinette, V.; Lessard, L.; Aubry, I.; Migon, E.; Sirois, J.; Haigh, J. J.; Begin, L. R.; Trotman, L. C.; et al. PTP1B Deficiency Enables the Ability of a High-Fat Diet to Drive the Invasive Character of PTEN-Deficient Prostate Cancers. *Cancer Res.* **2016**, *76*, 3130–3135.

200. Dubé, N.; Bourdeau, A.; Heinonen, K. M.; Cheng, A.; Lee Loy, A.; Tremblay, M. L. Genetic Ablation of Protein Tyrosine Phosphatase 1B Accelerates Lymphomagenesis of p53-Null Mice through the Regulation of B-Cell Development. *Cancer Res.* **2005**, *65*, 10088–10095.

201. Zabolotny, J. M.; Kim, Y.-B.; Welsh, L. A.; Kershaw, E. E.; Neel, B. G.; Kahn, B. B. Protein-tyrosine Phosphatase 1B Expression Is Induced by Inflammation in Vivo. *J. Biol. Chem.* **2008**, *283*, 14230–14241.

202. Gomez, E.; Vercauteren, M.; Kurtz, B.; Ouvrard-Pascaud, A.; Mulder, P.; Henry, J.-P.; Besnier, M.; Waget, A.; Hooft Van Huijsduijnen, R.; Tremblay, M. L.; et al. Reduction of heart failure by pharmacological inhibition or gene deletion of protein tyrosine phosphatase 1B. *J. Mol. Cell.*

Cardiol. **2012**, *52*, 1257–1264.

203. Nguyen, T. D.; Schwarzer, M.; Schrepper, A.; Amorim, P. A.; Blum, D.; Hain, C.; Faerber, G.; Haendeler, J.; Altschmied, J.; Doenst, T. Increased Protein Tyrosine Phosphatase 1B (PTP1B) Activity and Cardiac Insulin Resistance Precede Mitochondrial and Contractile Dysfunction in Pressure-Overloaded Hearts. *J. Am. Heart Assoc.* **2018**, *7*, 1–17.

204. Wade, F.; Belhaj, K.; Poizat, C. Protein tyrosine phosphatases in cardiac physiology and pathophysiology. *Heart Fail. Rev.* **2018**, *23*, 261–272.

205. Gogiraju, R.; Schroeter, M. R.; Bochenek, M. L.; Hubert, A.; Münzel, T.; Hasenfuss, G.; Schäfer, K. Endothelial deletion of protein tyrosine phosphatase-1B protects against pressure overload-induced heart failure in mice. *Cardiovasc. Res.* **2016**, *111*, 204–216.

206. Besnier, M.; Galaup, A.; Nicol, L.; Henry, J.-P.; Coquerel, D.; Gueret, A.; Mulder, P.; Brakenhielm, E.; Thuillez, C.; Germain, S.; et al. Enhanced angiogenesis and increased cardiac perfusion after myocardial infarction in protein tyrosine phosphatase 1B-deficient mice. *FASEB J.* **2014**, *28*, 3351–3361.

207. Forghieri, M.; Laggner, C.; Paoli, P.; Langer, T.; Manao, G.; Camici, G.; Bondioli, L.; Prati, F.; Costantino, L. Synthesis, activity and molecular modeling of a new series of chromones as low molecular weight protein tyrosine phosphatase inhibitors. *Bioorg. Med. Chem.* **2009**, *17*, 2658–2672.

208. Smith, A. M.; Maguire-Nguyen, K. K.; Rando, T. A.; Zasloff, M. A.; Strange, K. B.; Yin, V. P. The protein tyrosine phosphatase 1B inhibitor MSI-1436 stimulates regeneration of heart and multiple other tissues. *npj Regen. Med.* **2017**, *2*, 1–10.

209. Richards, R. I.; Robertson, S. A.; Kastner, D. L. Neurodegenerative diseases have genetic hallmarks of autoinflammatory disease. *Hum. Mol. Genet.* **2018**, *27*, 108–118.

210. de la Monte, S. M.; Wands, J. R. Alzheimer's Disease is Type 3 Diabetes—Evidence Reviewed. *J. Diabetes Sci. Technol.* **2008**, *2*, 1101–1113.

211. Ahmed, S.; Mahmood, Z.; Zahid, S. Linking insulin with Alzheimer's disease: emergence as type III diabetes. *Neurol. Sci.* **2015**, *36*, 1763–1769.

212. Bomfim, T. R.; Forny-Germano, L.; Sathler, L. B.; Brito-Moreira, J.; Houzel, J.-C.; Decker, H.; Silverman, M. A.; Kazi, H.; Melo, H. M.; McClean, P. L.; et al. An anti-diabetes agent protects the mouse brain from defective insulin signaling caused by Alzheimer's disease–associated Aβ oligomers. *J. Clin. Invest.* **2012**, *122*, 1339–1353.

213. Bonda, D. J.; Stone, J. G.; Torres, S. L.; Siedlak, S. L.; Perry, G.;

Kryscio, R.; Jicha, G.; Casadesus, G.; Smith, M. A.; Zhu, X.; et al. Dysregulation of leptin signaling in Alzheimer disease: evidence for neuronal leptin resistance. *J. Neurochem.* **2014**, *128*, 162–172.

214. Pandey, N. R.; Zhou, X.; Zaman, T.; Cruz, S. A.; Qin, Z.; Lu, M.; Keyhanian, K.; Brunel, J. M.; Stewart, A. F. R.; Chen, H.-H. LMO4 is required to maintain hypothalamic insulin signaling. *Biochem. Biophys. Res. Commun.* **2014**, *450*, 666–672.

215. Song, G. J.; Jung, M.; Kim, J.-H.; Park, H.; Rahman, M. H.; Zhang, S.; Zhang, Z.-Y.; Park, D. H.; Kook, H.; Lee, I.-K.; et al. A novel role for protein tyrosine phosphatase 1B as a positive regulator of neuroinflammation. *J. Neuroinflammation* **2016**, *13*, 86–100.

216. Vieira, M. N. N.; Lyra e Silva, N. M.; Ferreira, S. T.; De Felice, F. G. Protein Tyrosine Phosphatase 1B (PTP1B): A Potential Target for Alzheimer's Therapy? *Front. Aging Neurosci.* **2017**, *9*, 1–9.

217. Qin, Z.; Zhou, X.; Pandey, N. R.; Vecchiarelli, H. A.; Stewart, C. A.; Zhang, X.; Lagace, D. C.; Brunel, J. M.; Béïque, J.-C.; Stewart, A. F. R.; et al. Chronic Stress Induces Anxiety via an Amygdalar Intracellular Cascade that Impairs Endocannabinoid Signaling. *Neuron* **2015**, *85*, 1319–1331.

218. Tautz, L. PTP1B: a new therapeutic target for Rett syndrome. *J. Clin. Invest.* **2015**, *125*, 2931–2934.

219. Digenio, A.; Pham, N. C.; Watts, L. M.; Morgan, E. S.; Jung, S. W.; Baker, B. F.; Geary, R. S.; Bhanot, S. Antisense Inhibition of Protein Tyrosine Phosphatase 1B With IONIS-PTP-1B Rx Improves Insulin Sensitivity and Reduces Weight in Overweight Patients With Type 2 Diabetes. *Diabetes Care* **2018**, *41*, 807–814.

220. Sakaguchi, N.; Muramatsu, H.; Ichihara-Tanaka, K.; Maeda, N.; Noda, M.; Yamamoto, T.; Michikawa, M.; Ikematsu, S.; Sakuma, S.; Muramatsu, T. Receptor-type protein tyrosine phosphatase zeta as a component of the signaling receptor complex for midkine-dependent survival of embryonic neurons. *Neurosci. Res.* **2003**, *45*, 219–224.

221. Stoica, G. E.; Kuo, A.; Powers, C.; Bowden, E. T.; Sale, E. B.; Riegel, A. T.; Wellstein, A. Midkine Binds to Anaplastic Lymphoma Kinase (ALK) and Acts as a Growth Factor for Different Cell Types. *J. Biol. Chem.* **2002**, *277*, 35990–35998.

222. Maeda, N.; Ichihara-Tanaka, K.; Kimura, T.; Kadomatsu, K.; Muramatsu, T.; Noda, M. A Receptor-like Protein-tyrosine Phosphatase PTPζ/RPTPβ Binds a Heparin-binding Growth Factor Midkine. *J. Biol. Chem.* **1999**, *274*, 12474–12479.

223. Kadomatsu, K. Midkine and pleiotrophin in neural development and cancer. *Cancer Lett.* **2004**, *204*, 127–143.

224. Muramatsu, H.; Zou, K.; Sakaguchi, N.; Ikematsu, S.; Sakuma, S.; Muramatsu, T. LDL Receptor-Related Protein as a Component of the Midkine Receptor. *Biochem. Biophys. Res. Commun.* **2000**, *270*, 936–941.

225. Akihiro Fujikawa, Hajime Sugawara, Naomi Tanga, Kentaro Ishii, Kazuya Kuboyama, Susumu Uchiyama, Ryoko Suzuki, M. N. Structural basis for ligand-induced inactivation of protein tyrosine receptor type Z (PTPRZ): Physiological relevance of head-to-toe RPTP dimerization. *BioRxiv* **2019**.

226. Deuel, T. F.; Zhang, N.; Yeh, H.-J.; Silos-Santiago, I.; Wang, Z.-Y. Pleiotrophin: A Cytokine with Diverse Functions and a Novel Signaling Pathway. *Arch. Biochem. Biophys.* **2002**, *397*, 162–171.

227. Erlandsen, H.; Ames, J. E.; Tamkenath, A.; Mamaeva, O.; Stidham, K.; Wilson, M. E.; Perez-Pinera, P.; Deuel, T. F.; MacDougall, M. Pleiotrophin Expression during Odontogenesis. *J. Histochem. Cytochem.* **2012**, *60*, 366–375.

228. Weng, T.; Liu, L. The role of pleiotrophin and β-catenin in fetal lung development. *Respir. Res.* **2010**, *11*, 80–90.

229. Muramatsu, T. Midkine: a promising molecule for drug development to treat diseases of the central nervous system. *Curr. Pharm. Des.* **2011**, *17*, 410–423.

230. Yoshida, Y.; Sakakima, H.; Matsuda, F.; Ikutomo, M. Midkine in repair of the injured nervous system. *Br. J. Pharmacol.* **2014**, *171*, 924–930.

231. Herradón, G.; Pérez-García, C. Targeting midkine and pleiotrophin signalling pathways in addiction and neurodegenerative disorders: recent progress and perspectives. *Br. J. Pharmacol.* **2014**, *171*, 837–848.

232. Muramatsu, T. Structure and function of midkine as the basis of its pharmacological effects. *Br. J. Pharmacol.* **2014**, *171*, 814–826.

233. MURAMATSU, T. Midkine, a heparin-binding cytokine with multiple roles in development, repair and diseases. *Proc. Japan Acad. Ser. B* **2010**, *86*, 410–425.

234. Herradon, G.; Ramos-Alvarez, M. P.; Gramage, E. Connecting Metainflammation and Neuroinflammation Through the PTN-MK-RPTPβ/ζ Axis: Relevance in Therapeutic Development. *Front. Pharmacol.* **2019**, *10*, 377–390.

235. Fernández-Calle, R.; Vicente-Rodríguez, M.; Gramage, E.; Pita, J.; Pérez-García, C.; Ferrer-Alcón, M.; Uribarri, M.; Ramos, M. P.; Herradón, G. Pleiotrophin regulates microglia-mediated neuroinflammation. *J. Neuroinflammation* **2017**, *14*, 46–56.

236. Salama, R. H. M.; Muramatsu, H.; Shimizu, E.; Hashimoto, K.; Ohgake,

S.; Watanabe, H.; Komatsu, N.; Okamura, N.; Koike, K.; Shinoda, N.; et al. Increased midkine levels in sera from patients with Alzheimer's disease. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2005**, *29*, 611–616.

237. Yasuhara, O.; Muramatsu, H.; Kim, S. U.; Muramatsu, T.; Maruta, H.; McGeer, P. L. Midkine, a novel neurotrophic factor, is present in senile plaques of Alzheimer disease. *Biochem. Biophys. Res. Commun.* **1993**, *192*, 246–251.

238. Vicente-Rodríguez, M.; Fernández-Calle, R.; Gramage, E.; Pérez-García, C.; Ramos, M. P.; Herradón, G. Midkine Is a Novel Regulator of Amphetamine-Induced Striatal Gliosis and Cognitive Impairment: Evidence for a Stimulus-Dependent Regulation of Neuroinflammation by Midkine. *Mediators Inflamm.* **2016**, *2016*, 1–11.

239. Vicente-Rodríguez, M.; Pérez-García, C.; Ferrer-Alcón, M.; Uribarri, M.; Sánchez-Alonso, M. G.; Ramos, M. P.; Herradón, G. Pleiotrophin differentially regulates the rewarding and sedative effects of ethanol. *J. Neurochem.* **2014**, *131*, 688–695.

240. Gramage, E.; Rossi, L.; Granado, N.; Moratalla, R.; Herradón, G. Genetic inactivation of Pleiotrophin triggers amphetamine-induced cell loss in the substantia nigra and enhances amphetamine neurotoxicity in the striatum. *Neuroscience* **2010**, *170*, 308–316.

241. Soto-Montenegro, M. L.; Vicente-Rodríguez, M.; Pérez-García, C.; Gramage, E.; Desco, M.; Herradón, G. Functional neuroimaging of amphetamine-induced striatal neurotoxicity in the pleiotrophin knockout mouse model. *Neurosci. Lett.* **2015**, *591*, 132–137.

242. Gramage, E.; Putelli, A.; Polanco, M. J.; González-Martín, C.; Ezquerra, L.; Alguacil, L. F.; Pérez-Pinera, P.; Deuel, T. F.; Herradón, G. The neurotrophic factor pleiotrophin modulates amphetamine-seeking behaviour and amphetamine-induced neurotoxic effects: evidence from pleiotrophin knockout mice. *Addict. Biol.* **2010**, *15*, 403–412.

243. Gombash, S. E.; Manfredsson, F. P.; Mandel, R. J.; Collier, T. J.; Fischer, D. L.; Kemp, C. J.; Kuhn, N. M.; Wohlgenant, S. L.; Fleming, S. M.; Sortwell, C. E. Neuroprotective potential of pleiotrophin overexpression in the striatonigral pathway compared with overexpression in both the striatonigral and nigrostriatal pathways. *Gene Ther.* **2014**, *21*, 682–693.

244. Gombash, S. E.; Lipton, J. W.; Collier, T. J.; Madhavan, L.; Steece-Collier, K.; Cole-Strauss, A.; Terpstra, B. T.; Spieles-Engemann, A. L.; Daley, B. F.; Wohlgenant, S. L.; et al. Striatal Pleiotrophin Overexpression Provides Functional and Morphological Neuroprotection in the 6-Hydroxydopamine Model. *Mol. Ther.* **2012**, *20*, 544–554.

245. Jeanblanc, J.; Rolland, B.; Gierski, F.; Martinetti, M. P.; Naassila, M. Animal models of binge drinking, current challenges to improve face validity. *Neurosci. Biobehav. Rev.* **2019**, *106*, 112–121.

246. NIH Curriculum Supplement Series *The Essence of Drug Addiction*; Bethesda (MD), NIH, 2007.

247. Marballi, K.; Genabai, N. K.; Blednov, Y. A.; Harris, R. A.; Ponomarev, I. Alcohol consumption induces global gene expression changes in VTA dopaminergic neurons. *Genes, Brain Behav.* **2016**, *15*, 318–326.

248. Abrahao, K. P.; Salinas, A. G.; Lovinger, D. M. Alcohol and the Brain: Neuronal Molecular Targets, Synapses, and Circuits. *Neuron* **2017**, *96*, 1223–1238.

249. Most, D.; Ferguson, L.; Harris, R. A. Molecular basis of alcoholism.; 2014; p. 89–111.

250. Tulisiak, C. T.; Harris, R. A.; Ponomarev, I. DNA modifications in models of alcohol use disorders. *Alcohol* **2017**, *60*, 19–30.

251. Bourguet, E.; Ozdarska, K.; Moroy, G.; Jeanblanc, J.; Naassila, M. Class I HDAC Inhibitors: Potential New Epigenetic Therapeutics for Alcohol Use Disorder (AUD). *J. Med. Chem.* **2018**, *61*, 1745–1766.

252. Warnault, V.; Darcq, E.; Levine, A.; Barak, S.; Ron, D. Chromatin remodeling — a novel strategy to control excessive alcohol drinking. *Transl. Psychiatry* **2013**, *3*, 231–231.

253. Pradhan, A. A.; Tipton, A. F.; Zhang, H.; Akbari, A.; Pandey, S. C. Effect of Histone Deacetylase Inhibitor on Ethanol Withdrawal-Induced Hyperalgesia in Rats. *Int. J. Neuropsychopharmacol.* **2019**, *22*, 523–527.

254. Ma, H.; Zhu, G. The dopamine system and alcohol dependence. *Shanghai Arch. Psychiatry* **2014**, *26*, 61–68.

255. Maleki, N.; Tahaney, K.; Thompson, B. L.; Oscar-Berman, M. At the intersection of alcohol use disorder and chronic pain. *Neuropsychology* **2019**, *33*, 795–807.

256. Enoch, M.-A. The role of GABAA receptors in the development of alcoholism. *Pharmacol. Biochem. Behav.* **2008**, *90*, 95–104.

257. Enoch, M.-A. Pharmacogenomics of alcohol response and addiction. *Am. J. Pharmacogenomics* **2003**, *3*, 217–232.

258. Storvik, M.; Haukijarvi, T.; Tupala, E.; Tiihonen, J. Correlation between the SERT binding densities in hypothalamus and amygdala in cloninger type 1 and 2 alcoholics. *Alcohol Alcohol.* **2007**, *43*, 25–30.

259. Johnson, B. A. Role of the Serotonergic System in the Neurobiology of Alcoholism. *CNS Drugs* **2004**, *18*, 1105–1118.

260. Engleman, E. A.; Rodd, Z. A.; Bell, R. L.; Murphy, J. M. The role of 5-HT3 receptors in drug abuse and as a target for pharmacotherapy. *CNS Neurol. Disord. Drug Targets* **2008**, *7*, 454–467.

261. Cannady, R.; Rinker, J. A.; Nimitvilai, S.; Woodward, J. J.; Mulholland, P. J. Chronic Alcohol, Intrinsic Excitability, and Potassium Channels: Neuroadaptations and Drinking Behavior.; 2018; p. 311–343.

262. McGuier, N. S.; Griffin, W. C.; Gass, J. T.; Padula, A. E.; Chesler, E. J.; Mulholland, P. J. Kv7 channels in the nucleus accumbens are altered by chronic drinking and are targets for reducing alcohol consumption. *Addict. Biol.* **2016**, *21*, 1097–1112.

263. Rinker, J. A.; Fulmer, D. B.; Trantham-Davidson, H.; Smith, M. L.; Williams, R. W.; Lopez, M. F.; Randall, P. K.; Chandler, L. J.; Miles, M. F.; Becker, H. C.; et al. Differential potassium channel gene regulation in BXD mice reveals novel targets for pharmacogenetic therapies to reduce heavy alcohol drinking. *Alcohol* **2017**, *58*, 33–45.

264. Mitchell, J. M.; Teague, C. H.; Kayser, A. S.; Bartlett, S. E.; Fields, H. L. Varenicline decreases alcohol consumption in heavy-drinking smokers. *Psychopharmacology (Berl).* **2012**, *223*, 299–306.

265. Fucito, L. M.; Toll, B. A.; Wu, R.; Romano, D. M.; Tek, E.; O'Malley, S. S. A preliminary investigation of varenicline for heavy drinking smokers. *Psychopharmacology (Berl).* **2011**, *215*, 655–663.

266. Mason, B. J.; Quello, S.; Goodell, V.; Shadan, F.; Kyle, M.; Begovic, A. Gabapentin Treatment for Alcohol Dependence. *JAMA Intern. Med.* **2014**, *174*, 70–78.

267. Mula, M.; Cavanna, A. E.; Monaco, F. Psychopharmacology of topiramate: from epilepsy to bipolar disorder. *Neuropsychiatr. Dis. Treat.* **2006**, *2*, 475–488.

268. Patsalos, P. N. The mechanism of action of topiramate. *Rev. Contemp. Pharmacother.* **1999**, *10*, 147–153.

269. Guglielmo, R.; Martinotti, G.; Quatrale, M.; Ioime, L.; Kadilli, I.; Di Nicola, M.; Janiri, L. Topiramate in Alcohol Use Disorders: Review and Update. *CNS Drugs* **2015**, *29*, 383–395.

270. de Beaurepaire, R. Suppression of Alcohol Dependence Using Baclofen: A 2-Year Observational Study of 100 Patients. *Front. Psychiatry* **2012**, *3*, 1–7.

271. Mayor, S. Baclofen is largely ineffective for alcohol use disorders, finds study. *BMJ* **2018**, 930.

272. Reynaud, M.; Aubin, H.-J.; Trinquet, F.; Zakine, B.; Dano, C.; Dematteis, M.; Trojak, B.; Paille, F.; Detilleux, M. A Randomized, Placebo-Controlled Study of High-Dose Baclofen in Alcohol-Dependent Patients—The ALPADIR Study. *Alcohol Alcohol.* **2017**, *52*, 439–446.

273. Knapp, C. M.; O'Malley, M.; Datta, S.; Ciraulo, D. A. The Kv7

potassium channel activator retigabine decreases alcohol consumption in rats. *Am. J. Drug Alcohol Abuse* **2014**, *40*, 244–250.

274. Zwierzyńska, E.; Krupa, A.; Pietrzak, B. A pharmaco-EEG study of the interaction between ethanol and retigabine in rabbits. *Am. J. Drug Alcohol Abuse* **2015**, *41*, 153–160.

275. Simon-O'Brien, E.; Alaux-Cantin, S.; Warnault, V.; Buttolo, R.; Naassila, M.; Vilpoux, C. The histone deacetylase inhibitor sodium butyrate decreases excessive ethanol intake in dependent animals. *Addict. Biol.* **2015**, *20*, 676–689.

276. Jeanblanc, J.; Lemoine, S.; Jeanblanc, V.; Alaux-Cantin, S.; Naassila, M. The Class I-Specific HDAC Inhibitor MS-275 Decreases Motivation to Consume Alcohol and Relapse in Heavy Drinking Rats. *Int. J. Neuropsychopharmacol.* **2015**, *18*, 1–9.

277. Jeanblanc, J.; Bourguet, E.; Sketriené, D.; Gonzalez, C.; Moroy, G.; Legastelois, R.; Létévé, M.; Trussardi-Régnier, A.; Naassila, M. Interest of new alkylsulfonylhydrazide-type compound in the treatment of alcohol use disorders. *Psychopharmacology (Berl).* **2018**, *235*, 1835–1844.

278. Laksitorini, M.; Prasasty, V. D.; Kiptoo, P. K.; Siahaan, T. J. Pathways and progress in improving drug delivery through the intestinal mucosa and blood–brain barriers. *Ther. Deliv.* **2014**, *5*, 1143–1163.

279. Banks, W. A.; Gertler, A.; Solomon, G.; Niv-Spector, L.; Shpilman, M.; Yi, X.; Batrakova, E.; Vinogradov, S.; Kabanov, A. V. Principles of strategic drug delivery to the brain (SDDB): Development of anorectic and orexigenic analogs of leptin. *Physiol. Behav.* **2011**, *105*, 145–149.

280. Copeland, R. A.; Harpel, M. R.; Tummino, P. J. Targeting enzyme inhibitors in drug discovery. *Expert Opin. Ther. Targets* **2007**, *11*, 967–978.

281. Guex, N.; Peitsch, M. C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **2009**, *30*, 162–173.

282. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27*, 221–234.

283. Siu, S. W. I.; Pluhackova, K.; Böckmann, R. A. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *J. Chem. Theory Comput.* **2012**, *8*, 1459–1470.

284. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided. Mol. Des.* **2007**, *21*, 681–691.

285. Schrödinger, LLC LigPrep 2019.

286. Schrödinger, LLC Glide 2019.

287. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

288. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

289. Gaussian Inc. Gaussian 03 2003.

290. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

291. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.

292. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

293. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

294. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

295. Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, Á.; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: An Ultrafast and Accurate Scoring Function for Protein–Protein Docking. *J. Chem. Theory Comput.* **2012**, *8*, 3395–3408.

296. Alvarez-Garcia, D.; Barril, X. Relationship between Protein Flexibility and Binding: Lessons for Structure-Based Drug Design. *J. Chem. Theory Comput.* **2014**, *10*, 2608–2614.

297. Seco, J.; Luque, F. J.; Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **2009**, *52*, 2363–2371.

298. Martínez, J. M.; Martínez, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **2003**, *24*, 819–825.

299. Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A

package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

300. Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564–2574.

301. Ibrahim, M. A. A.; Hasb, A. A. M.; Mekhemer, G. A. H. Role and nature of halogen bonding in inhibitor···receptor complexes for drug discovery: casein kinase-2 (CK2) inhibition as a case study. *Theor. Chem. Acc.* **2018**, *137*, 38–48.

302. Riley, K. E.; Merz, K. M. Insights into the Strength and Origin of Halogen Bonding: The Halobenzene−Formaldehyde Dimer. *J. Phys. Chem. A* **2007**, *111*, 1688–1694.

303. Ibrahim, M. A. A. Molecular mechanical perspective on halogen bonding. *J. Mol. Model.* **2012**, *18*, 4625–4638.

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

305. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276–3285.

306. Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and flexible gpu accelerated binding free energy calculations within the amber molecular dynamics package. *J. Comput. Chem.* **2018**, *39*, 1354–1358.

307. Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.* **2013**, *9*, 4131–4139.

308. Kamerlin, S. C. L.; Rucker, R.; Boresch, S. A targeted molecular dynamics study of WPD loop movement in PTP1B. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 1161–1166.

309. Sheriff, S.; Beno, B. R.; Zhai, W.; Kostich, W. A.; McDonnell, P. A.; Kish, K.; Goldfarb, V.; Gao, M.; Kiefer, S. E.; Yanchunas, J.; et al. Small Molecule Receptor Protein Tyrosine Phosphatase γ (RPTPγ) Ligands That Inhibit Phosphatase Activity via Perturbation of the Tryptophan–Proline–Aspartate (WPD) Loop. *J. Med. Chem.* **2011**, *54*, 6548–6562.

310. Almo, S. C.; Bonanno, J. B.; Sauder, J. M.; Emtage, S.; Dilorenzo, T. P.; Malashkevich, V.; Wasserman, S. R.; Swaminathan, S.; Eswaramoorthy, S.; Agarwal, R.; et al. Structural genomics of protein phosphatases. *J. Struct. Funct. Genomics* **2007**, *8*, 121–140.

311. Lins, L.; Thomas, A.; Brasseur, R. Analysis of accessible surface of residues in proteins. *Protein Sci.* **2003**, *12*, 1406–1417.

312. Ung, P. M. U.; Ghanakota, P.; Graham, S. E.; Lexa, K. W.; Carlson, H. A. Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case. *Biopolymers* **2016**, *105*, 21–34.

313. Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 14257–14263.

314. Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845–5859.

315. Tan, Y. S.; Spring, D. R.; Abell, C.; Verma, C. The Use of Chlorobenzene as a Probe Molecule in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 1821–1827.

316. Uehara, S.; Tanaka, S. Cosolvent-Based Molecular Dynamics for Ensemble Docking: Practical Method for Generating Druggable Protein Conformations. *J. Chem. Inf. Model.* **2017**, *57*, 742–756.

317. Arcon, J. P.; Defelipe, L. A.; Modenutti, C. P.; López, E. D.; Alvarez-Garcia, D.; Barril, X.; Turjanski, A. G.; Martí, M. A. Molecular Dynamics in Mixed Solvents Reveals Protein–Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. *J. Chem. Inf. Model.* **2017**, *57*, 846–863.

318. Pastor, M.; Fernández-Calle, R.; Di Geronimo, B.; Vicente-Rodríguez, M.; Zapico, J. M.; Gramage, E.; Coderch, C.; Pérez-García, C.; Lasek, A. W.; Puchades-Carrasco, L.; et al. Development of inhibitors of receptor protein tyrosine phosphatase β/ζ (PTPRZ1) as candidates for CNS disorders. *Eur. J. Med. Chem.* **2018**, *144*, 318–329.

319. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337–341.

320. Daina, A.; Zoete, V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, *11*, 1117–1121.

321. Huang, R.; Bonnichon, A.; Claridge, T. D. W.; Leung, I. K. H. Protein-ligand binding affinity determination by the waterLOGSY method: An optimised approach considering ligand rebinding. *Sci. Rep.* **2017**, *7*, 43727.

322. Abbott, K. L.; Matthews, R. T.; Pierce, M. Receptor Tyrosine Phosphatase β (RPTPβ) Activity and Signaling Are Attenuated by Glycosylation and Subsequent Cell Surface Galectin-1 Binding. *J. Biol. Chem.* **2008**, *283*, 33026–33035.

323. Ambjørn, M.; Dubreuil, V.; Miozzo, F.; Nigon, F.; Møller, B.; Issazadeh-Navikas, S.; Berg, J.; Lees, M.; Sap, J. A Loss-of-Function Screen for Phosphatases that Regulate Neurite Outgrowth Identifies PTPN12 as a

Negative Regulator of TrkB Tyrosine Phosphorylation. *PLoS One* **2013**, *8*, 65371–65387.

324. Müller, J.; Esső, K.; Dargó, G.; Könczöl, Á.; Balogh, G. T. Tuning the predictive capacity of the PAMPA-BBB model. *Eur. J. Pharm. Sci.* **2015**, *79*, 53–60.

325. Dutton, J. W.; Chen, H.; You, C.; Brodie, M. S.; Lasek, A. W. Anaplastic lymphoma kinase regulates binge-like drinking and dopamine receptor sensitivity in the ventral tegmental area. *Addict. Biol.* **2017**, *22*, 665–678.

326. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.

327. Ramírez, D.; Caballero, J. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules* **2018**, *23*, 1038–1055.

328. Bhakat, S.; Åberg, E.; Söderhjelm, P. Prediction of binding poses to FXR using multi-targeted docking combined with molecular dynamics and enhanced sampling. *J. Comput. Aided. Mol. Des.* **2018**, *32*, 59–73.

329. Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2019.

330. Imai, K.; Takaoka, A. Comparing antibody and small-molecule therapies for cancer. *Nat. Rev. Cancer* **2006**, *6*, 714–727.

331. Thura, M.; Al-Aidaroos, A. Q.; Gupta, A.; Chee, C. E.; Lee, S. C.; Hui, K. M.; Li, J.; Guan, Y. K.; Yong, W. P.; So, J.; et al. PRL3-zumab as an immunotherapy to inhibit tumors expressing PRL3 oncoprotein. *Nat. Commun.* **2019**, *10*, 2484.

332. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717.

333. Baell, J.; Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **2014**, *513*, 481–483.

334. Daina, A.; Michielin, O.; Zoete, V. iLOGP: A Simple, Robust, and Efficient Description of n -Octanol/Water Partition Coefficient for Drug Design Using the GB/SA Approach. *J. Chem. Inf. Model.* **2014**, *54*, 3284–3301.

335. ABADZAPATERO, C.; METZ, J. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **2005**, *10*, 464–469.

CHAPTER III

3. Dual CK2-HDAC inhibitors

One of the primary goals of our research group is the design of multi-target modulators aimed at distinct tumor disease mechanisms. Multi-targeted drugs (MTDs) for cancer treatment is an unconventional approach consisting of drug combinations in a single molecule [1,2]. The rationale for the design of dual targeting agents is to overcome incomplete efficacy and drug resistance that often occurs in individual drugs [3]. Combination cancer therapy achieves the desirable outcomes through distinct mechanisms of action in a concerted way, but presents also some disadvantages such as the potential for drug-drug interactions, pharmacokinetic complexity and patient compliance [4]. The aim of the MTDs is to overcome these difficulties through the advantages expected from this approach:

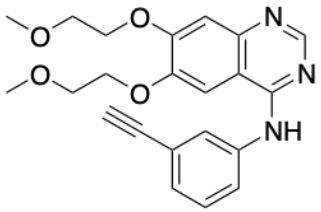
1) Depleting the drug-drug interactions with only one molecule.
2) Presence of only one pharmacokinetic profile.
3) Only one pharmaceutical form, so the dosage schedule is easier to follow.

The usual paradigm in drug discovery is based in one target, one drug, but although drugs are designed with this final aim, very often the outcome is not the expected one. Small molecules inside the body can interact with other targets called off targets. The challenge of MTDs is to obtain molecules from various chosen targets with suitable and balanced affinities.

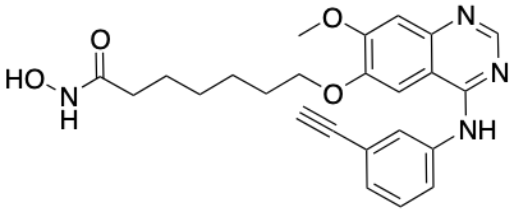
As an example, Imatinib, was originally intended as a selective fusion protein inhibitor for BCR-Abl, however, it was later shown to inhibit non-oncogenic C-Abl kinase in ordinary cells. This extra activity has been revealed to be important to the effectiveness of imatinib against chronic myeloid leukemia, with the polypharmacological profile of the drug being the foundation of its therapeutic utility [5].

In our research group, two validated targets involved in the development and progression of tumor processes have been selected for our multi-target approach: protein kinase-CK2 and histone deacetylases (HDACs). CK2 is

an active serine/threonine kinase that is constitutive and crucial for cell viability. It presents proliferative and anti-apoptotic effects that generate a favorable tumor progression and maintenance of cell environment, thus constituting an interesting target for cancer treatment. On the other side, HDACs are epigenetic regulators expected to be one of the most promising targets for cancer therapy in drug development. Protein kinase CK2 is an appealing choice for a secondary target due to its role in HDAC post-translational modification and regulation; and, to its ability to activate HDAC1 in hypoxia related tumors (**Fig. 3.1**) [6,7]. The relationship between the two enzymes is a significant element that supports our hypothesis that HDAC1/CK2 dual inhibitors can be a fresh and valid approach in the quest for new antitumor agents (**Fig. 3.2**). This is not the first time that a dual inhibitor has been designed using as inspiration a developed kinase inhibitor. For instance, Erlotinib has been used as a scaffold for the design of CUDC-101, a dual inhibitor of HDAC/EFGR with potent activity against both enzymes [8].



ERLOTINIB



CUDC-101

Figure 3.1. Erlotinib *is an EGFR inhibitor with IC_{50} of 2 nM, >1000-fold more sensitive for EGFR than human c-Src or v-Abl. Erlotinib is under clinical assays in phases 1/2/3 for different treatments.* ***CUDC-101*** *is a potent multi-targeted inhibitor against HDAC, EGFR and HER2 with IC_{50} of 4.4 nM, 2.4 nM, and 15.7 nM, respectively and inhibits class I/II HDACs, but not class III, Sir-type HDACs. It is currently under phase 1 clinical trials in patients with head and neck squamous cell carcinoma.*

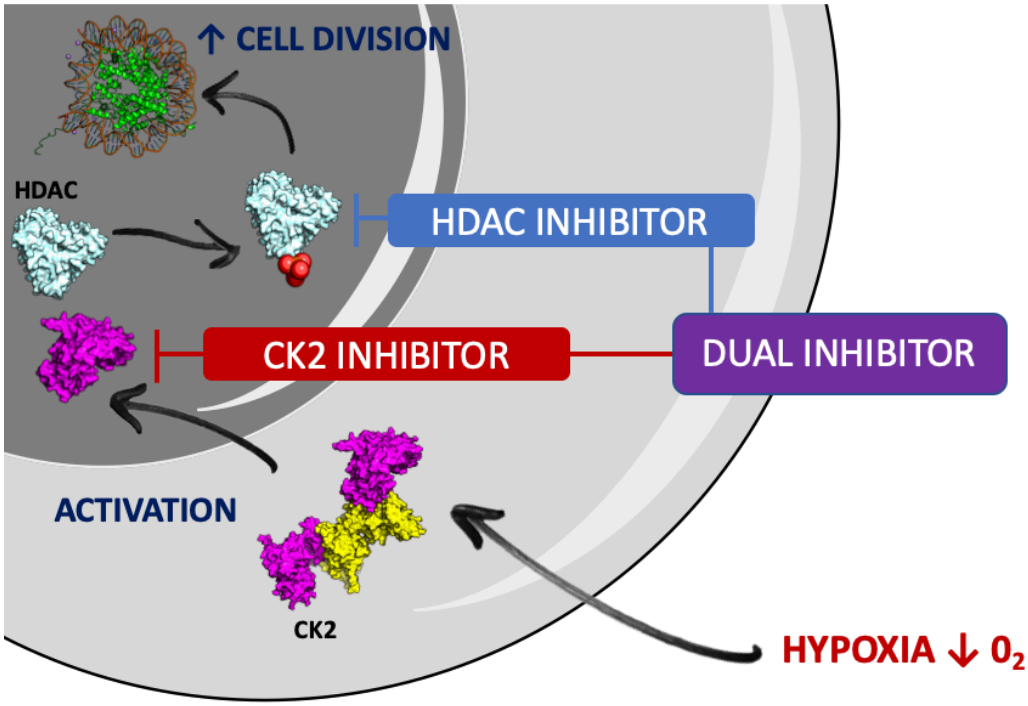


Figure 3.2. *General mechanism of action of a dual HDAC/CK2 inhibitor. The synergic effect occurs by blocking two different pathways with only one molecule.*

3.1. Histone deacetylases (HDACs)

Histone deacetylases (EC 3.5.1.98, HDACs) are a class of enzymes that remove acetyl groups from the π-N-acetyl lysine on a histone, enabling the histones to wrap the DNA more closely. They are also referred to as lysine deacetylases (KDAC) to define their role rather than their goal, including non-histone proteins. The activity of HDACs is contrary to that of histone acetyltransferases (EC 2.3.1.48, Histone acetyltransferase). In order to fit within the nucleus and participate in the formation of chromosomes, nuclear DNA is extremely condensed and wrapped around histones. Acetylation and de-acetylation regulate the expression of DNA. A histone modification is a posttranslational covalent modification (PTM) that can result in a functional shift in gene expression or on the structure of chromatin [9]. HDACs, are classified in four classes depending on sequence homology to the yeast original enzymes and domain organization (**Table 3.1**) [10].

Table 3.1. HDAC classification.

Class	I	IIa	IIb	III	IV
	HDAC 1,2,3 & 8	HDAC 4,5,7 & 9	HDAC 6 & 10	(SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, SIRT7)	HDAC11

They are zinc-dependent enzymes that require the presence of a zinc ion (Zn^{2+}) in the active site to carry out the reaction. The overall structure of the catalytic domain consists of a globular domain with a hydrophobic tunnel lined by two phenylalanine residues that give access to the catalytic Zn^{2+}. The deacetylation reaction is common to all HDACs and starts with the penetration of the acetylated lysine into the active site through the tunnel. The carbonyl of the acetylated lysine chelates the Zn^{2+} while a water molecule interacts strongly with the imidazole side chains of two histidine residues. The water molecule is then activated by deprotonation by one of those histidines and assaults the carbonyl of the substrate, resulting in a tetrahedral intermediate stabilized by Zn^{2+} and a tyrosine residue. The collapse of the tetrahedral intermediate releases the acetate and the lysine. The protonated histidine acts as a general acid to protonate the group that leaves the amine (**Fig. 3.3**).

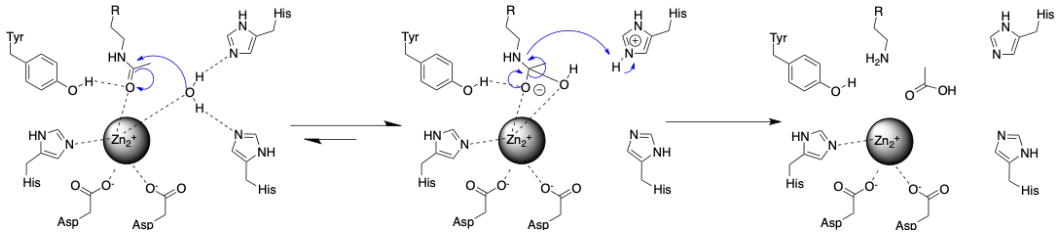


Figure 3.3. Mechanism of the lysine deacetylation by HDACs.

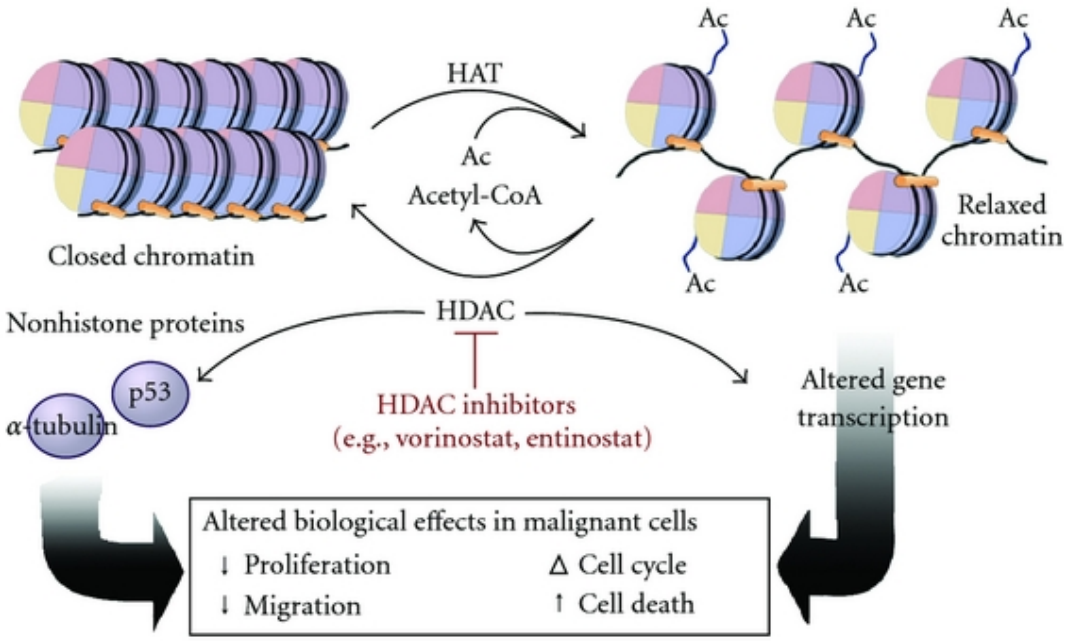


Figure 3.4. *General mechanism of chromatin regulation by HDAC and HAT and the effect in cells.*

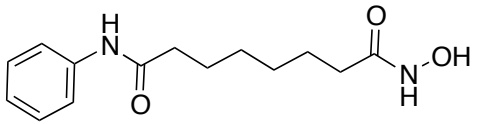
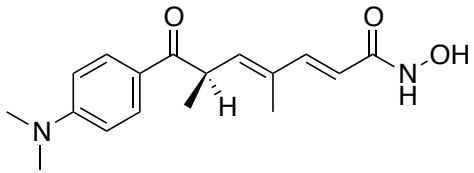
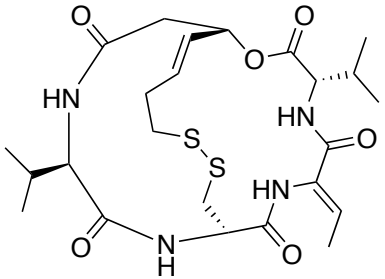
HDACs are engaged in several cell cycle pathways (cell growth and death), environmental information processing, signal transduction, and human diseases (**Fig. 3.4**). These diseases range from neurodegenerative processes (including amyotrophic lateral sclerosis & alcoholism) [11], rare diseases, and cancer. Different HDAC family members are implicated in different types of cancers, such as: HDAC1 in gastric, prostate, colon, and breast cancer; HDAC2 in cervical and gastric cancers, and in colorectal carcinoma; and HDAC3 and HDAC6 in colon and breast cancer [12].

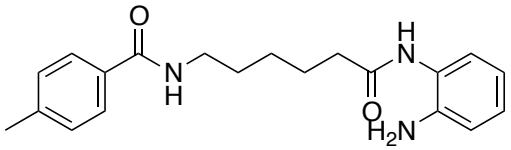
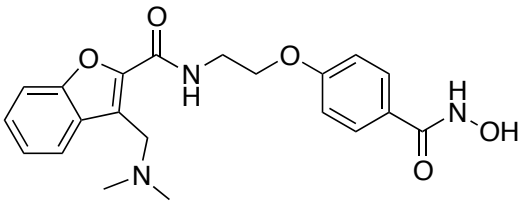
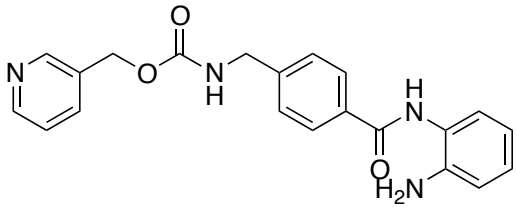
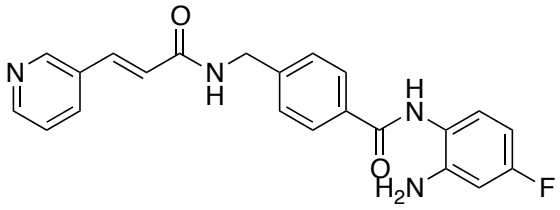
In many cancers, hypoacetylation is an intrinsic characteristic that has been related to HDAC silencing of tumor suppressor genes (such as p21). Clinical studies have demonstrated that the inhibition of HDACs could lead to therapeutic advantages [13]. Moreover, HDACs are not exclusive to the control of the transcription level of DNA; they also regulate non-histone proteins such as p53, E2F1, RelA, YY1, TFIIE, BCL6 and TFIIF.

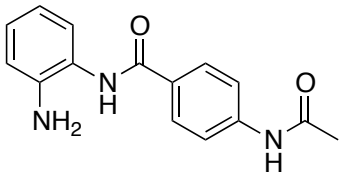
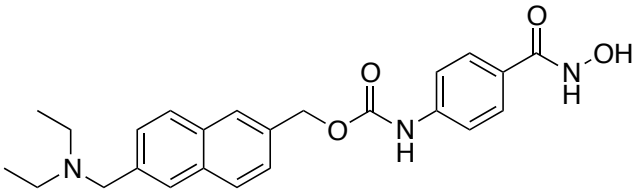
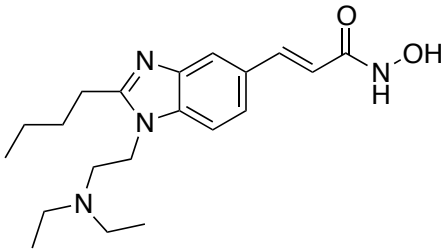
3.1.1. Known HDAC inhibitors.

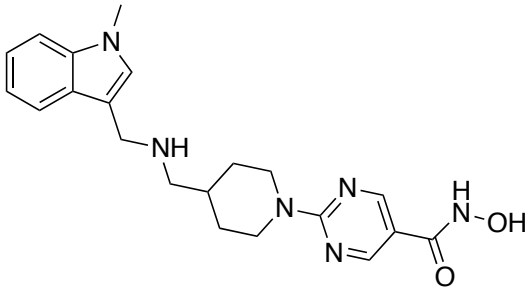
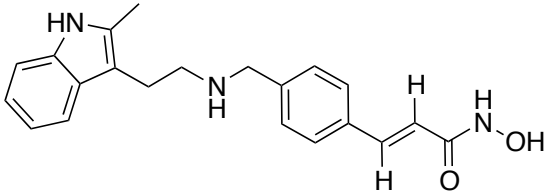
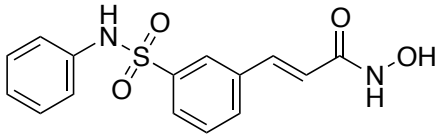
Numerous and approved HDAC inhibitors (HDIs) have been developed. A brief summary is shown in **Table 3.2**. Amongst those, valproic acid that has a lengthy history of use in psychiatry and neurology as a mood stabilizer and antiepileptic is considered to be a HDIs. This fact has recently led to the research directed to the development of HDIs as a neurodegenerative disease therapy. Additionally, a clinical trial investigates the impact of valproic acid in infected individuals on the latent pools of HIV due to its capacity to deplete this reservoir [14]. In the last years, several attempts have been undertaken to broaden the arsenal for cancer therapy with HDIs. In 2006, Vorinostat (SAHA®) was approved by the FDA to treat skin manifestations in patients with Cutaneous T Cell Lymphoma (CTCL) who have failed prior treatments; and a second HDI for CTCL patients, Romidepsin (Istodax®), was approved in 2009 [15]. The precise mechanisms by which these compounds function remains uncertain, but several proposals point towards epigenetic pathways. Based on in vitro synergy, HDIs are presently being studied in conjunction with DNA methylation inhibitors as chemosensitizers for cytotoxic chemotherapy or radiation therapy. Isoform selective HDIs have been designed and synthesized to help elucidate the function of individual HDAC isoforms. In parallel, HDIs also have effects on non-histone proteins that are modulated by acetylation altering the degree of acetylation and increasing or repressing their activity. Thus, HDIs have shown to alter the activity of many transcription factors, including ACTR, cMyb, E2F1, EKLF, FEN 1, GATA, HNF-4, HSP90, Ku70, NFκB, PCNA, p53, RB, Runx, SF1 Sp3, STAT, TFIIE, TCF and YY1 [16].

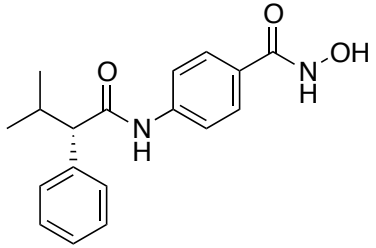
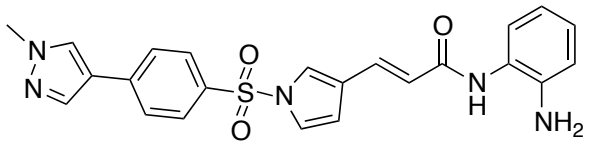
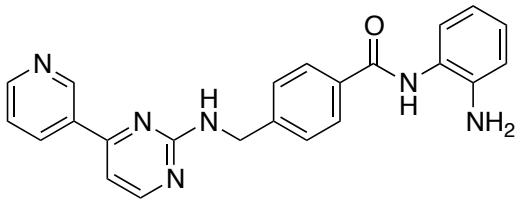
Table 3.2. *Relevant HDIs. Biological activity and clinical relevance.*

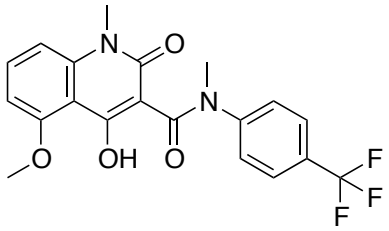
Compound	Chemical structure	Biological activity and clinical relevance
Vorinostat (SAHA, MK0683)		HDAC inhibitor with IC50 of ~10 nM in a cell-free assay. FDA approved
Trichostatin A (TSA)		HDAC inhibitor with IC50 of ~1.8 nM in cell-free assays.
Romidepsin (FK228, Depsipeptide)		HDAC1 and HDAC2 inhibitor with IC_{50} of 36 nM and 47 nM in cell-free assays, respectively. FDA approved.

RG2833 (RGFP109)		HDAC inhibitor with IC_{50} of 60 nM and 50 nM for HDAC1 and HDAC3 in cell-free assays, respectively. Brain-penetrant
Abexinostat (PCI-24781)		Pan-HDAC inhibitor mostly targeting HDAC1 with Ki of 7 nM, modest potency against HDACs 2, 3, 6, and 10 and greater than 40-fold selectivity against HDAC8. Phase 1/2.
Entinostat (MS-275)		HDAC1 and HDAC3 with IC_{50} of 0.51 μM and 1.7 μM in cell-free assays, compared with HDACs 4, 6, 8, and 10. Phase 3.
Tucidinostat (Chidamide)		HDAC1, 2, 3, and 10, the HDAC isotypes well documented to be associated with the malignant phenotype with IC_{50} values of 95, 160, 67, 78 nM for HDAC1, 2, 3, 10, respectively.

Tacedinaline (CI994)		HDAC inhibitor with IC_{50} of 0.9, 0.9, 1.2, and >20 μM for human HDAC 1, 2, 3, and 8, respectively. Phase 3.
Givinostat (ITF2357)		HDAC inhibitor for maize HD2, HD1B and HD1A with IC_{50} of 10 nM, 7.5 nM and 16 nM in cell-free assays. Phase 2.
Pracinostat (SB939)		pan-HDAC inhibitor with IC_{50} of 40-140 nM with exception for HDAC6. It has no activity against the class III isoenzyme SIRT I. Phase 2.

Quisinostat (JNJ-26481585)		HDAC inhibitor with highest potency for HDAC1 with IC_{50} of 0.11 nM in a cell-free assay, modest potent to HDACs 2, 4, 10, and 11; greater than 30-fold selectivity against HDACs 3, 5, 8, and 9 and lowest potency to HDACs 6 and 7. Phase 2.
Panobinostat (LBH589)		HDAC inhibitor with IC_{50} of 5 nM in a cell-free assay (HT29). Phase 3.
Belinostat (PXD101)		HDAC inhibitor with IC_{50} of 27 nM in a cell-free assay, with activity demonstrated in cisplatin-resistant tumors.

AR42		HDAC inhibitor with IC50 of 30 nM (cell-free assay). Phase 1.
Domatinostat (4SC-202)		HDAC inhibitor with IC_{50} of 1.20 μM, 1.12 μM, and 0.57 μM for HDAC1, HDAC2, and HDAC3, respectively. Also displays inhibitory activity against lysine specific demethylase 1 (LSD1). Phase 1.
Mocetinostat (MGCD0103)		HDAC inhibitor with most potency for HDAC1 with IC50 of 0.15 μM in a cell-free assay, 2- to 10- fold selectivity against HDAC2, 3, and 11, and no activity to HDAC4, 5, 6, 7, and 8. Phase 2.

Tasquinimod		Orally active antiangiogenic agent by allosterically inhibiting HDAC4 signaling. Phase 3.

Despite their chemical diversity shown in **Table 3.2**, most of these agents have a prevalent pharmacophore consisting of the following three structural patterns (**Fig. 3.5**):

1) Zinc binding group (ZBG) designed to chelate the catalytic zinc ion.
2) Aliphatic or aromatic linker that interacts with the hydrophobic tunnel. This linker has to be long enough for the ZBG to reach the catalytic zinc ion.
3) A CAP group that interacts with residues of the surface of the protein. This group normally presents an aromatic moiety.

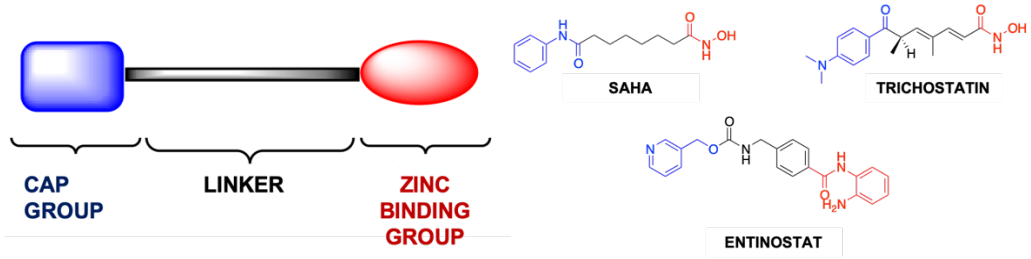


Figure 3.5. General HDAC Inhibitor design.

Our research group has been working mainly with two types of HDACs: HDAC1 and HDAC6. The comparative amino acid sequence between both HDAC 1 & 6 is shown below in **Figure 3.6**. The total identity percentage is 25.73%, but the active site and surrounding sites are quite similar. The hydrophobic tunnel that gives access to the catalytic zinc ion and the surrounding space present only a few different amino acids, making the binding of any orthoestetic inhibitor similar in all cases.

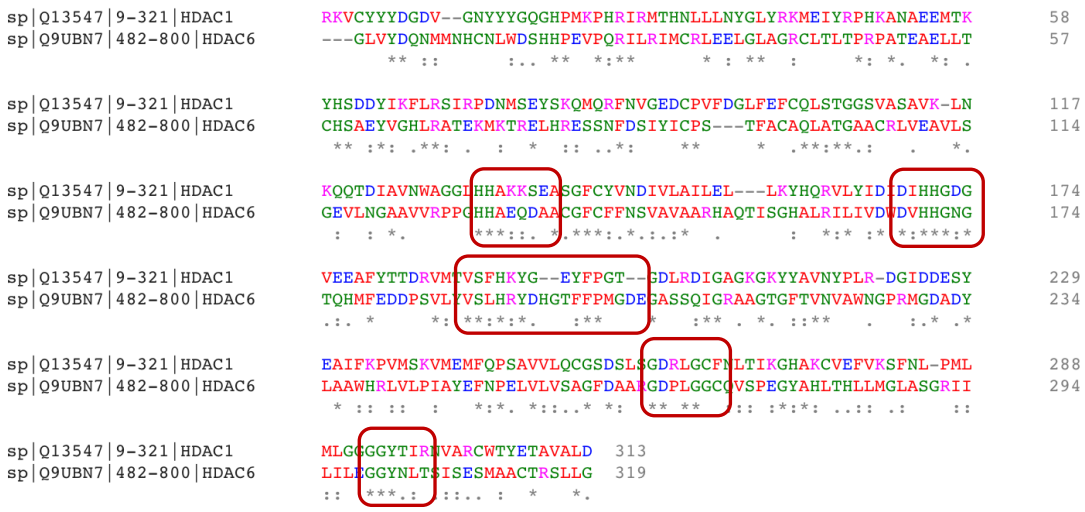


Figure 3.6. *Sequence alignment carried out with BLAST. The red boxes mark the active site and its surroundings with high percentage of sequence similarity.*

3.1.2. HDAC1

HDAC1 (UniProtKB - Q13547) is a metalloenzyme monomer formed by 482 amino acids that can be phosphorylated by CK2 in Ser421 and Ser423 [17]. The mutation of either of these two residues to alanine, or the disruption of the CK2 consensus sequence in charge of directing the phosphorylation, reduces the enzymatic activity and complex formation of HDAC1. It is in charge of the deacetylation of SP proteins, mainly SP1 and SP3, regulating their function [18]. It is also a main component of the BRG1-RB1-HDAC1 complex, which negatively regulates the CREST-mediated transcription in resting neurons. Upon calcium stimulation, HDAC1 is released from the complex and CREBBP is recruited, which facilitates transcriptional activation. Moreover, HDAC1 is required for the transcriptional repression of circadian target genes, such as PER1, mediated by the large PER complex or CRY1 through histone deacetylation [19].

A core of eight-stranded parallel ß sheets, created by a compact α/β fold domain, forms the structure of HDAC1. Additional α helices are found near the β sheet's C-terminal end. A portion of these helixes together with loops

connecting to the secondary structural elements form a hydrophobic channel on with a depth of 10 to12 Å that includes a catalytic Zn^{2+} atom that conforms the polar catalytic nucleus at the bottom (**Fig. 3.7**). In addition to the catalytic Zn^{2+}, HDAC1 presents two K^{+} that seem to present structural function and to support the activity of Zn^{2+} [20].

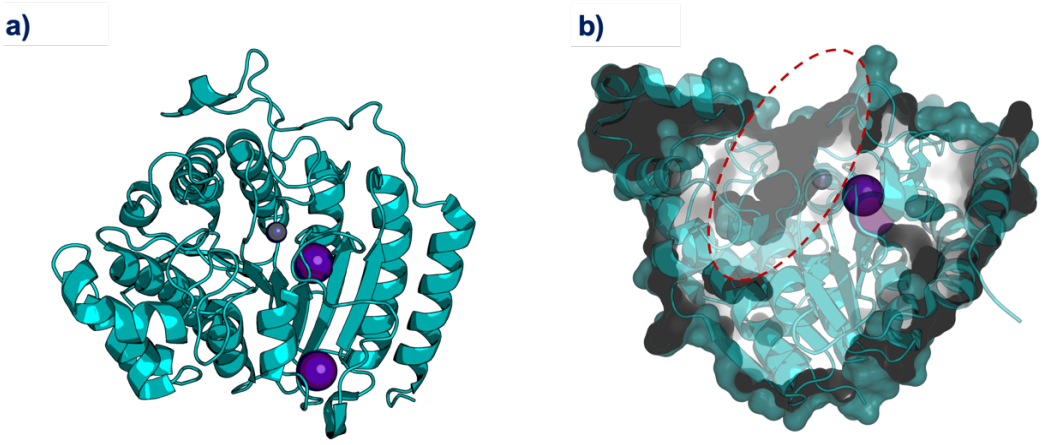


Figure 3.7. a) *General structure of HDAC1 showing the catalytic Zn^{2+} (grey) and relevant K_+ (purple).* ***b)*** *The highlighted catalytic tunnel.*

The PDB structures available for HDAC1 are only two obtained by X-ray crystallography and shown in **Table 3.3**.

Table 3.3. *The PDB structures available for HDAC1.*

PDB Entry	Domain	Associate Protein (Complex)	PDB ID of the bound Ligand
4BKX	1-482 aa	MTA1	ACT acetic acid
5ICN	1-376 aa	MTA1	I6P Inositol 1,2,3,4,5,6-hexakisphosphate and 6A0 (2S)-2-amino-8-(hydroxyamino)-8-oxooctanoic acid

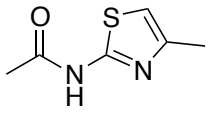
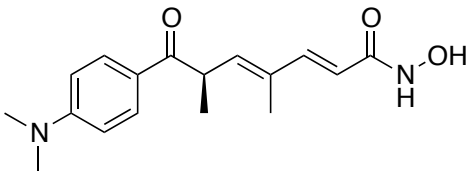
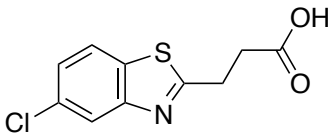
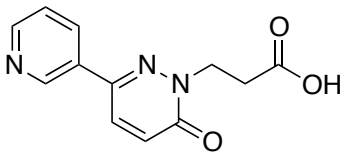
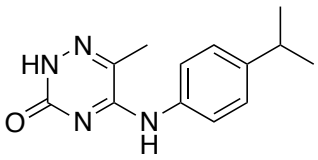
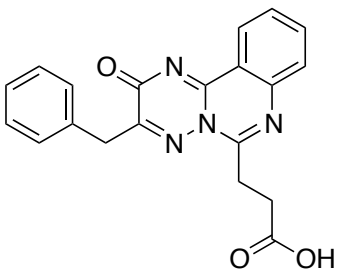
3.1.3. HDAC6.

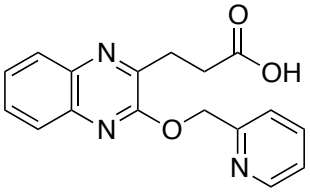
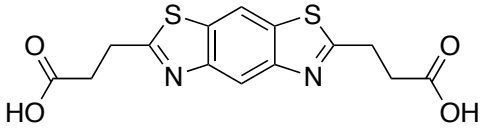
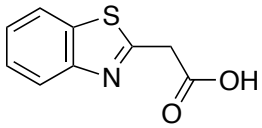
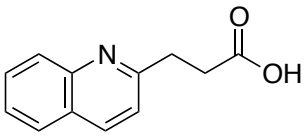
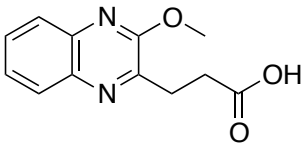
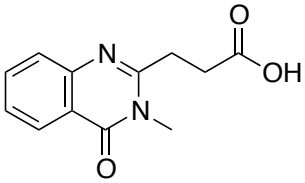
HDAC6 (UniProtKB - Q9UBN7) presents a similar structure to HDAC1 in the main core that is constituted of two catalytic domains (1 and 2), giving a total structure of 1215 amino acids. The last 200 amino acids of this protein present a zinc-finger ubiquitin-binding domain that has been the target of a recent drug design process by soaking techniques [21]. This ubiquitin-binding domain is involved in cell motility processes via tubulin deacetylation [22]. It is also involved in the MTA1-mediated epigenetic regulation of ESR1 expression in breast cancer [23]. Even though the specific regulation mechanism of this HDAC is not well known, there is a number of serine and threonine residues that can be phosphorylated thus enhancing its activity [24].

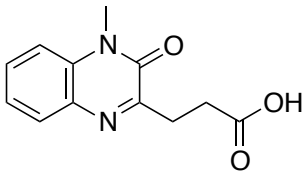
Regarding the crystal structures available for HDAC6, the vast majority of them (16 structures) solve the portion between the amino acids 1109 to 1215, which represent the Histone Deacetylase 6 Zinc-Finger Domain (**Table 3.4**). Only one PDB code, **5EDU** represents the pure Histone deacetylase domain 2. This crystal structure is in complex with **Trichostatin A (Table 3.2-TSA)**. TSA is an antibiotic that serves as an antifungal and selectively inhibits the class I and II mammalian HDAC family of enzymes, but not class III HDACs.

Table 3.4. *The PDB structures available for HDAC6.*

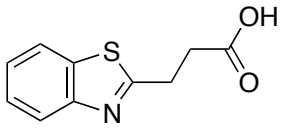
PDB entry	Domain	PDB ID of the bound Ligand
3C5K	1109-1215	-
3GV4	1109-1215	-
3PHD	1109-1215	-
5B8D	1109-1213	**6T4**

		
		TSN
5EDU	479-835	
		6U6
5KH3	1109-1213	
		6T7
5KH7	1109-1213	
		6T5
5KH9	1109-1213	
		9ZV
5WBN	1108-1213	

		B8P
5WPB	1109-1208	
		EYP
6CE6	1109-1213	
		EYV
6CE8	1109-1213	
		EYY
6CEA	1109-1213	
		EY7
6CEC	1109-1213	
		EYA
6CED	1109-1213	
6CEE	1109-1213	**EYM**



EYJ

6CEF	1109-1213	

3.2. CK2 Casein kinase II subunit alpha

Protein kinases (PKs) phosphorylate protein residues acting as counterparts of protein phosphatases. PKs transduce most of the signals inside eukaryotic cells by modifying the activity of the substrate and regulating a number of other cellular procedures such as metabolism, transcription, cycle development, rearrangement and motion of the cytoskeleton, apoptosis and differentiation. CK2 is a non-specific serine/threonine protein kinase, which is found to be a cancer driver since it has been identified as a proto-oncogene [25]. Several studies highlighted that CK2 is involved in the regulation of basic cell processes in eukaryotic cells like intercellular and intracellular signaling pathways by transducing, amplifying or integrating upstream signals. In fact, the catalytic activity of CK2 can be ascribed to a substantial percentage of the cell phosphoproteoma [26]. This omnipresent PK can be defined as a key enzyme, regulating the crosstalk among several signaling processes critical to cell differentiation, proliferation and survival [27].

CK2 can be found in a multitude of cell compartments in which hundreds of substrates are phosphorylated. These significant characteristics are connected to its many tasks and roles inside the cell:

1) Implicated in cell cycle progression and cell damage [28].
2) RNA-polymerase I, II, III and IV regulation [29].

3) Regulation of numerous transcription factors including NF-kappa-B, STAT1, CREB1, IRF1, IRF2, HDACs, ATF1, SRF, MAX, JUN, FOS, MYC and MYB [30].
4) Regulation of chaperone function by phosphorylation of Hsp90 and its co-chaperones FKBP4 and CDC37 [31].
5) It has been implicated in neuronal processes and several diseases such as inflammatory, neurodegenerative, vascular; and several types of tumors [32].

The relevance of CK2 in drug discovery can be highlighted by the huge number of crystal structures deposited in the PDB. Until now, there are up to 160 structures available, and it is worth to mention that **5ZN0** was obtained by Neutron imaging with 1.10 Å resolution [33]. Since the analysis of all these structures would be unaffordable, only few structures depicting relevant differences such as a novel allosteric pocket will be mentioned in this work. In any case, the kinase druggability has been extensively studied leading to the discovery of some common allosteric pockets amongst all family members (kinase-atlas.bu.edu) [34].

The general structure of kinases presents two lobes: a small N-terminal lobe and large C-terminal lobe. The N-lobe is constituted of five stranded antiparallel β-sheets (β1- β5) and a long α-helix called α-C-helix; whereas the C-terminal domain is mainly composed by various α-helixes, making a more robust tertiary structure. The region connecting the N and C lobes is known as the hinge region [35].

The main pocket in any kinase is the nucleotide-binding site located at the interface of the two lobes. In this hydrophobic cavity, both lobes contribute to the substrate recognition, but most of the interactions involve the N-lobe. The nitrogenated base of the nucleotide binds in the pocket making specific hydrogen bonds with the backbone of the amino acids of the hinge region, whereas the triphosphate moiety of the nucleotide stretches towards the activation loop and the amino acid residues responsible for the catalytic

activity and phosphate transfer to the substrate protein. The activation loop, also known as A-loop, is present in all kinases but the phosphorylation requirements for it to adopt an active conformation vary depending on the kinase [36]. Inside the nucleotide-binding site, the end of the A-loop is constituted of a conserved three-amino acid pattern called the DFP motif (Asp-Phe-Gly). This pattern is in charge of the regulation of the enzymatic turnover as the side chain of the aspartic residue is member of the Mg^{2+} coordination sphere and plays a key role in the phosphate transfer to the substrate. The side chain of the central phenylalanine is located inside a hydrophobic pocket located between the N-lobe and the C-lobe in the conformation referred to as DFG-in conformation. Upon substrate phosphorylation, the side chain of the aspartic acid is no longer able to coordinate the Mg^{2+} and rotates towards the hydrophobic pocket [37]. This rotation forces the change of Φ and Ψ angles of phenylalanine so that the side chain exits the pocket and pushes the diphosphate nucleotide out of the binding site. This conformation is known as the DFG-out conformation and leaves the back pocket unoccupied, a fact that has been extensively explored in drug design [38]. Access to this back pocket is provided by an amino acid located at the back of the nucleotide-binding site and known as gatekeeper that changes amongst kinases [34]; and, inhibitors that explore the back pocket whilst interacting with the gatekeeper have achieved high kinase selectivity [39].

Despite being member of a much-conserved family of enzymes, CK2 exhibits certain structural differences that makes it outstand. Apart from the fact that CK2 is able to recognize both ATP and GTP [40], the constitutive activation without the need of phosphorylation of the A-loop of CK2 changes the way this kinase is regulated. Therefore, CK2 presents a quaternary structure with two catalytic domains (α and α') and two regulatory β subunits (β and β') that modulate the substrate binding by protein-protein interactions [41]. The enzyme turnover mechanism brought about by the DFG domain is different, as in CK2 the phenylalanine is mutated to tryptophan thus giving a DWG motif [42]. There is no account of the flexibility of this loop and no

crystal structure has been obtained in an out conformation of this motif.

Moreover, CK2 is also constrained by three intrinsic elements:

1) The N-terminal domain that stabilizes and fixes the activation segment and the α-C-helix [42].
2) Trp in the DWG loop that replaces the central Phe residue of the canonical DFG motif. Trp is bulkier than Phe and the additional hydrophobic interactions disfavor the "DFG-out" conformation observed in other kinases [43].
3) Structurally conserved water and chloride atoms observed, supports contact between the N-terminal segment and the activation loop [33].

Additionally, a novel cryptic pocket located near the nucleotide-binding site has been elucidated by X-Ray crystallography and soaking techniques in a project carried out by D. R. Springb and M. Hyvönen in the search for specific CK2 inhibitors [44]. The outcome of this project led to the determination of three main druggable pockets in CK2: i) the nucleotide-bidding site; ii) the MPP pocket [34]; and, iii) the α-D-pocket exclusive to CK2 [44]. The MPP pocket is located on the N-terminal domain, and various PDB codes present small molecules bound inside this pocket (**Fig. 3.8**). The α-D-pocket has been discovered by soaking CK2 with 3,4-dichlorophenethylamine derivatives that were able to displace Phe121 and Tyr125 to occupy the pocket (**Fig. 3.9**).

As a result, this research group has deposited more than 54 structures, exploring the α-D-pocket that led to the design and synthesis of selective compounds **CAM4066** and **CAM4712** with IC_{50} values of 320nM and 7 μM, respectively (**Fig. 3.10**) [45].

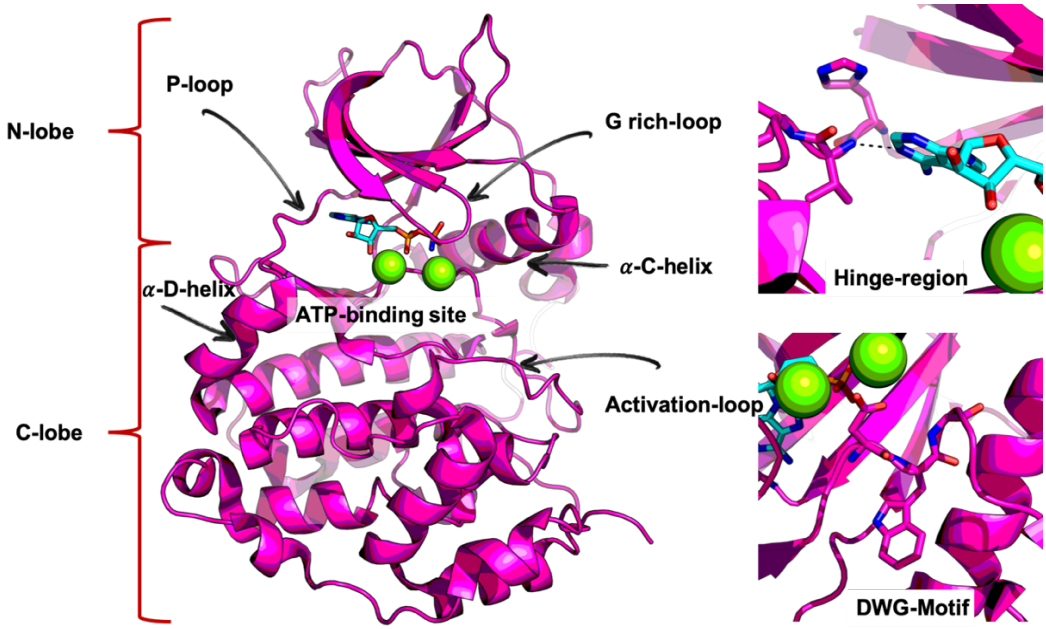


Figure 3.8. *Catalytic subunit of CK2 bound to a slow-hydrolysis ATP analogue (**3NSZ**). The main components of the kinase are indicated, and the hinge region and DWG-motif are zoomed. ATP analogue is colored in blue and magnesium ions in green.*

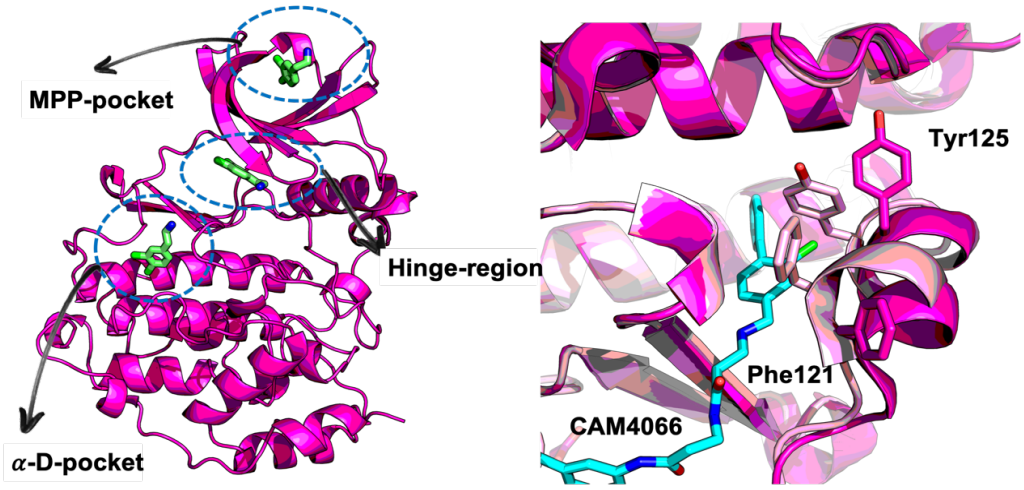


Figure 3.9. Left. *Main druggable pockets in CK2. Right: the α-D-pocket filled by **CAM4066** where the displacement of the Phe121 and Tyr125 is highlighted by the superimposition of the apo conformation in pink (**1NA7**) and α-D-pocket-bound conformation brought about by **CAM4066** in magenta (**5CU4**).*

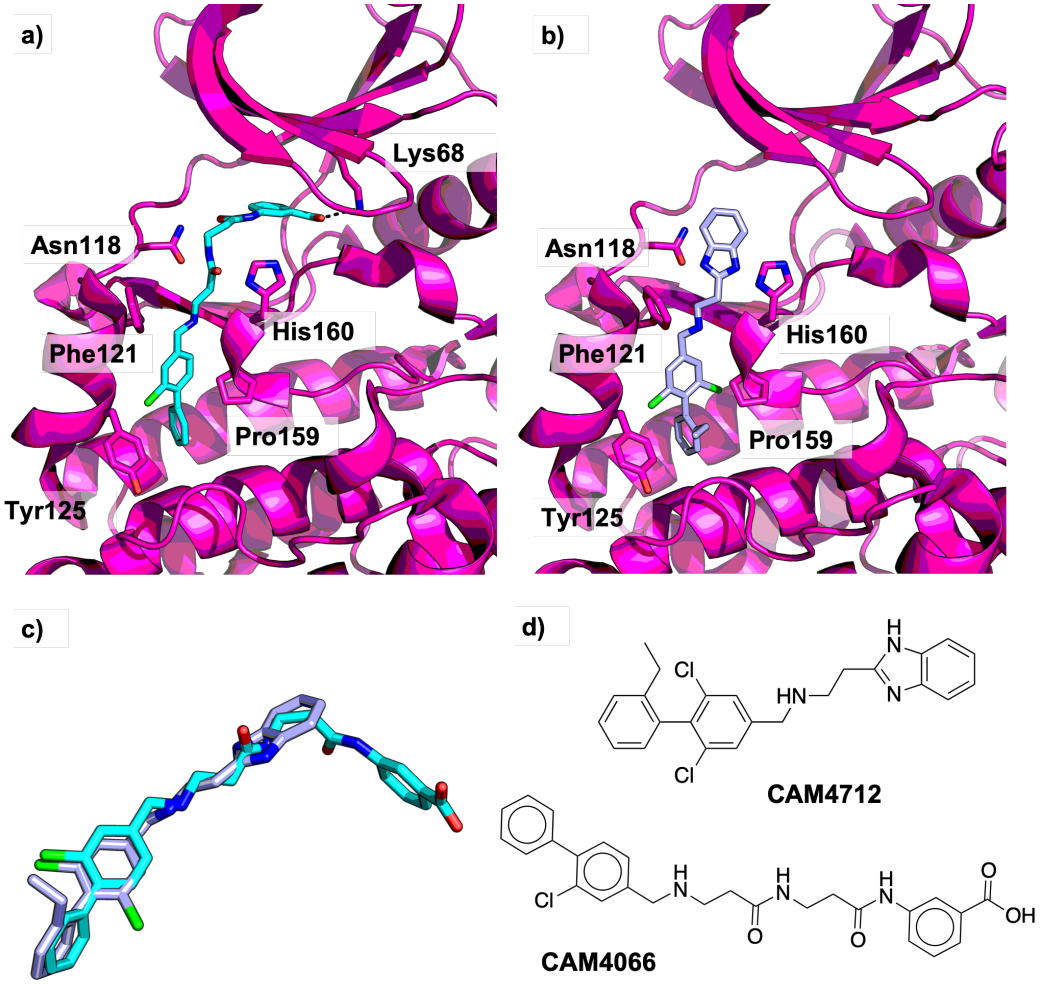


Figure 3.10. ***a)*** *Crystal structure o fCK2 (**5CU4**) bound to **CAM4066**.* ***b)*** *Crystal structure of CK2 (**5OTY**) bound to **CAM4712**.* ***c)*** *Superimposition of **CAM4066** and **CAM4712**.* ***d)*** *Chemical structures of CK2 α-D-pocket inhibitors **CAM4066** and **CAM4712***

3.2.1. Known CK2 inhibitors

ATP-competitive inhibitors (type I inhibitors): They are able to reach the ATP-binding site with the ability to compete directly with the ATP (hence competitive-inhibitors) (**Fig. 3.11**). The structure of this type of inhibitors is quite diverse, but amongst all of them is a common interaction with the hinge region by the establishment of either hydrogen or halogen bonds with the backbone of Val116. Of all of them, **CX4945** (Silmitasertib) deserves a particular mention as it is the only one currently under clinical assays (NCT02128282 & NCT02128282). On the other hand, both **TBB** (4,5,6,7-tetrabromo-1*H*-benzotriazole) and its analogue **DMAT** (2-dimethylamino-

4,5,6,7-tetrabromo-1*H*-benzimidazole) have also been considered important CK2 inhibitors because of their potency and selectivity [46]. In these molecules, the four bromine atoms of the benzene ring are able to interact with the bulky side chains inside the ATP-binding site as well as with Val116 through halogen bonds. Selectivity arises from the fact that these bulky molecules can only fit in CK2 [47,48].

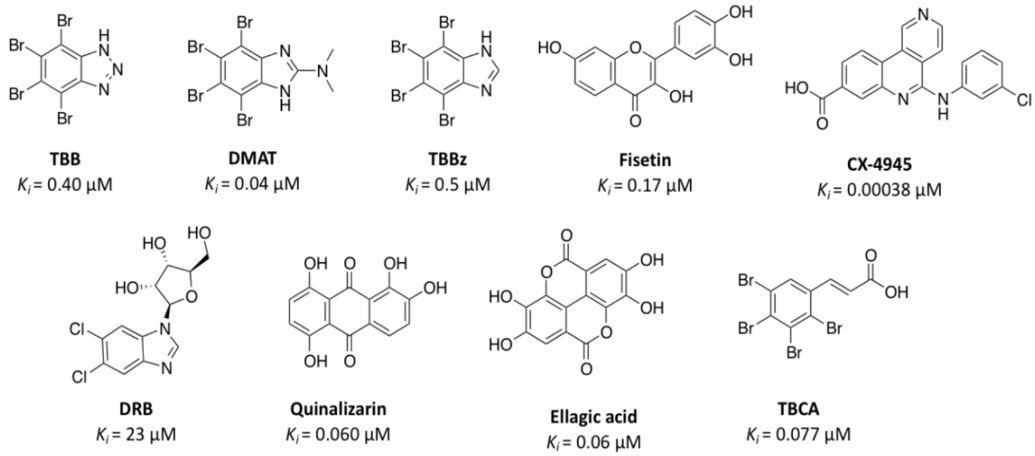


Figure 3.11. *CK2 ATP-competitive inhibitors and their inhibitory constant (Ki).*

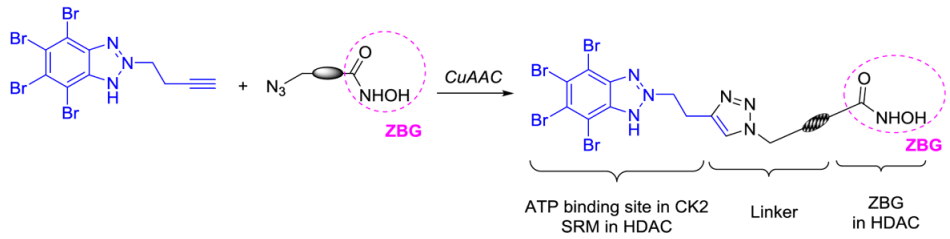
Non-competitive ATP inhibitors (allosteric inhibitors or CAM derivatives) (**Fig. 3.10**) [44,45]. As we mentioned before, CAM derivatives demonstrated that it is possible to inhibit CK2 by reaching the α-D-pocket through additional interactions outside the active site and without establishing any interaction with the hinge region.

3.3. Objectives

In this chapter, both enzymes have been selected as targets for our MTDs project. The CK2 ATP-competitive inhibitors **TBB** and **DMAT** were chosen for this purpose as they present high affinity for CK2. These molecules were functionalized with a linker with a terminal alkyne. A hydroxamic acid, known to be a good ZBG intended to bind the catalytic site of HDACs, is also functionalized with a linker bearing an azide. Both functionalized molecules were subjected to the Cu(I) alkyne azide cycloaddition reaction (CuAAC),

also known as “click reaction” (**Scheme 1**) [49,50]. The resulting molecules are deemed to interact with CK2 through the TBB and DMAT moieties whereas the hydroxamic acid stabilizes the binding with additional interactions; and, with HDACs through the hydroxamic ZBG whereas the TBB and DMAT moieties interact with the residues in the rim of the active site of HDACs, acting as a surface recognition motif (SRM).

***Scheme 1.**: Synthesis of dual HDAC1/CK2 or HDAC6/CK2 based on TBB.*

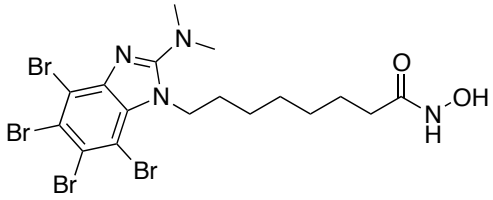
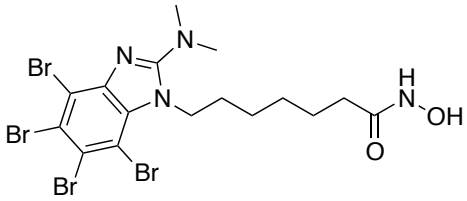
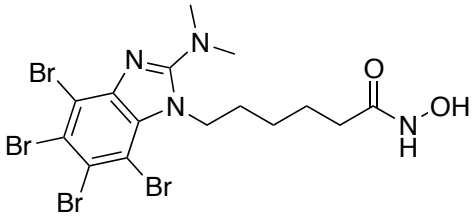
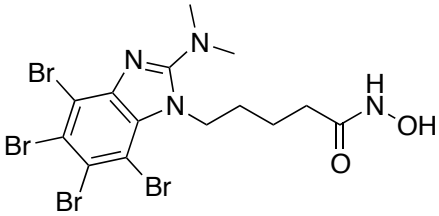
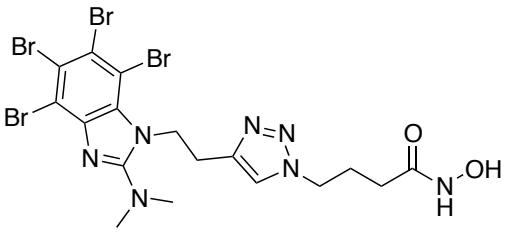


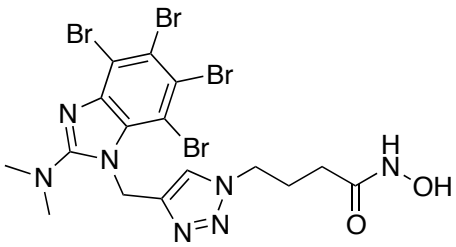
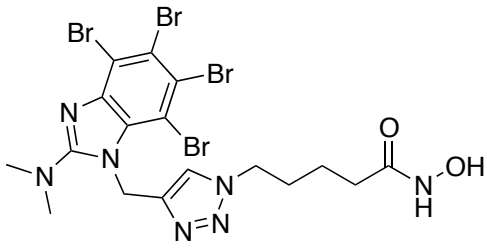
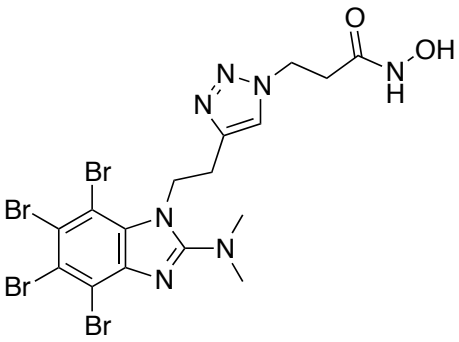
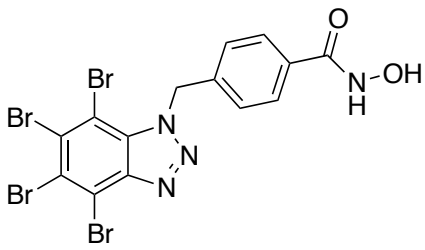
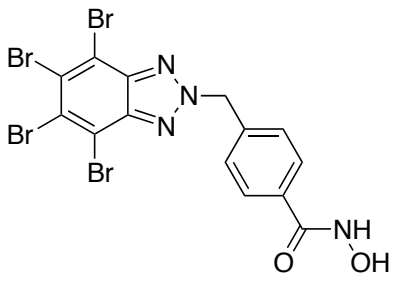
3.4. Synthesis of dual inhibitors and in vitro results

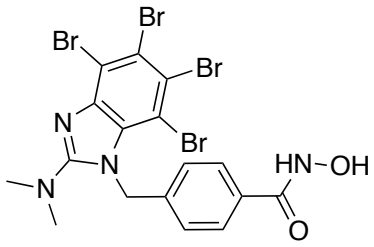
The total synthesis of the set of compounds was carried out by Dr. Regina Martinez and Dr. Myriam Pastor at the Universidad San Pablo CEU; and the biological evaluation was carried out by Prof. Rostyslav Panchuk and Dr. Maciej Maslyk at Institute of Cell Biology, NAS of Ukraine and The John Paul II Catholic University of Lublin, respectively.

The general aim of this work is the design, synthesis and biological evaluation of a series of dual HDAC1/CK2 inhibitors with the aim of finding a novel strategy for the development of antitumor agents. The design of these structures is based on the general structure of HDIs, that is: a cap group, an aliphatic linker, and a ZBG. As the hydroxamic acid moiety is a well-known ZBG, it will be attached though different length aliphatic linkers to the CK2 inhibitors **TBB** and **DMAT**. These CK2 inhibitors are both ATP-competitive with high affinity for the nucleotide-binding site of CK2 and will mimic the CAP group of the HDIs. The synthesized molecules and their IC_{50} values against each enzyme are shown in **Table 3.5**.

Table 3.5. Developed dual inhibitors and their IC_{50} values in μM.

Compound	Structure	HDAC1 IC_{50}	HDAC6 IC_{50}	CK2 IC_{50}
11d		13.7	8.98	5.89
11c		1.77	1.13	16.6
11b		1.46	0.66	3.67
11a		62.0	-	34.2
7d		10.2	-	53.8

7a		1.43	-	85.1
7b		2.25	2.41	50.1
7c		13.3	30.7	12.5
15b		4.22	2.70	136
15a		8.50	11.0	15.0

19		9.45	20.6	7.15

3.5. In vitro cell evaluation

The cytotoxic activity of the most potent CK2-HDAC multitarget inhibitors (**7c**, **11b**, **11c**, **11d**, **15a** and **19**) was evaluated in human leukemia (Jurkat) and carcinoma (MCF-7, HCT-116) cell lines using trypan blue exclusion assay [51,52]. The potential anti-cancer activity was tested, and selectivity of these compounds was compared in the human embryonic kidney cells of HEK293. (**Fig. 3.12**).

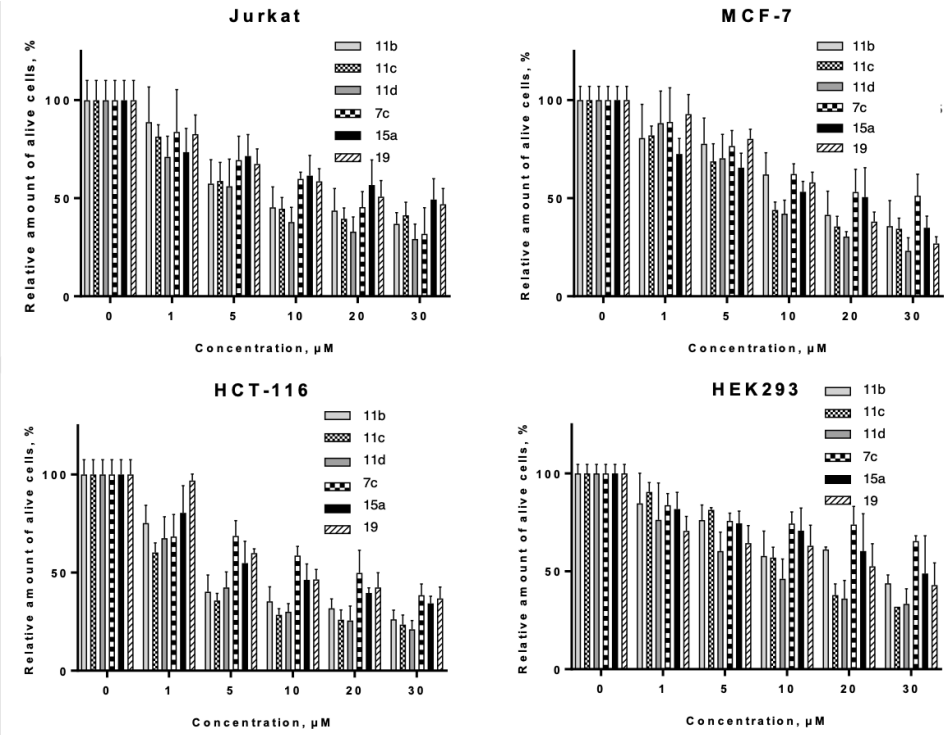


Figure 3.12. *Comparison of cytotoxic activity of selected CK2-HDAC dual inhibitors (**7c, 11b, 11c, 11d, 15a** and **19**) towards human leukemia (a), carcinoma (b and c) and pseudonormal (d) cell lines. Trypan blue assay, 24-hour incubation. Means and SD of at least three experiments in triplicate are shown.*

LC_{50} doses (lethal concentration of drug needed to kill 50% cells) of all compounds was calculated and is summarized in **Table 3.6.** To reveal potential mechanisms of cell death induction by selected CK2-HDAC inhibitors, Annexin V-PI double staining assay was addressed. The Annexin V corresponding signal provides a very sensitive method for detecting cellular apoptosis, while propidium iodide (PI) is used to detect necrotic or late apoptotic cells, characterized by the loss of the integrity of the plasma and nuclear membranes. (**Fig. 3.13**). Compounds with the highest cytotoxic activity (**11b-d**) also led to 2-fold increase in number of annexin V-positive (apoptotic) cells in concentration-dependent mode, when compared to **15a** and **19**. Introduction of a triazole in the linker (**7c**) led to a 3-fold decrease in the amount of AnV(+) cells compared to **11d**, indicating a negative impact of this group both on cytotoxic and pro-apoptotic activity of **7c**. However, all tested compounds demonstrated potent inhibitory effect on cell cycle progression in human leukemia cells, depleting cell population in S-phase (**Fig. 3.13**). In particular, **11b-d** led to 2-fold decrease in the number of cells in S-phase at 5 μM dose, while at 10 μM concentrations these compounds already induced massive apoptosis (measured by significant increase in the number of pre-G1 cells). Compounds with weaker cytotoxic potential (**7c, 15a** and **19**) possessed weak proapoptotic activity even at 30 μM dose. Besides lowering the number of cells in S-phase these inhibitors also increased cell population in G1 phase, a phenomenon that was not typical for **11c** and **11d** due to their high proapoptotic activity (**Fig. 3.13**). As G1/S cell arrest is considered a typical feature of HDAC inhibitors, including **SAHA** the observed phenomenon is yet another confirmation of HDAC targeting by novel dual inhibitors in cell culture.

Table 3.6. LC_{50} *values for* ***7c***, ***11a-d*** *and* ***19*** *towards human tumor and pseudonormal cell lines in vitro.*

Compound	LC_{50} values of compounds for cell line, μM (M±SD)			
	Jurkat	MCF-7	HCT-116	HEK293
7c	10,69±1,50	15,66±2,54	4,22±0,46	23,33±4,44
11b	10,04±1,15	9,97±0,99	1,90±0,24	13,61±0,76
11c	5,63±0,78	9,02±0,90	3,10±0,37	8,41±1,28
11d	14,17±1,83	27,75±6,06	16,87±3,90	>30
15a	>30	13,66±2,67	8,67±1,02	>30
19	21,80±3,17	13,55±0,78	11,30±1,01	24,47±8,335

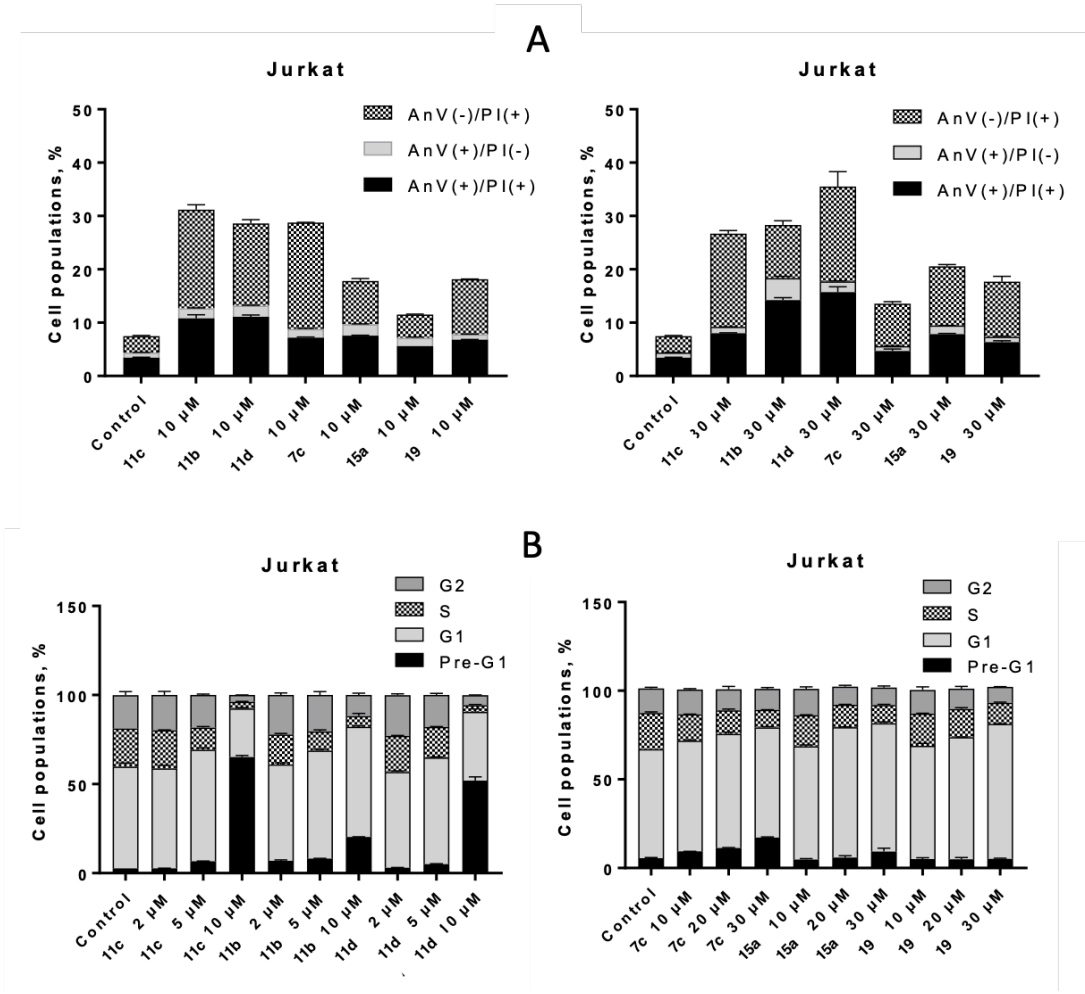


Figure 3.13. *Cell death induction and cell cycle arrest induced by dual CK2-HDAC inhibitors.* ***A)*** *Evaluation of the impact of selected dual CK2-HDAC inhibitors (10 µM and 30 µM, 24 hours) on phosphatidylserine externalization in Jurkat T-leukemia cells determined by 21 FACS using FITC-labeled annexin V and PI staining.* ***B)*** *Impact of the indicated compound doses on the distribution of Jurkat T-leukemia cells in the different phases of the cell cycle was determined by FACS of propidium iodide-stained cells after 24 hours of continuous exposure. One of three experiments delivering comparable data is shown.*

3.6. Molecular modelling

Computational modeling techniques were applied in order to understand the general binding mode of these compounds and rationalize the interesting in vitro results shown in the previous **Table 3.5**. For this purpose, docking techniques were used as a first approach, followed by Molecular Dynamics (MD) simulations with the aim to propose a plausible binding mode for each

complex on CK2, HDAC1 and HDAC6 and to evaluate the stability of the ligand-target complexes in an aqueous environment. The general binding mode of known selective inhibitors such as **TBB**, **TBI** and **DMAT** in CK2 has been established through X-ray studies. Upon binding in the nucleotide-binding site, these drugs occupy the nitrogenated base cavity through strong and selective interactions. The specific nature of **TBB**, **TBI** and **DMAT** against CK2, comes from the halogen bonds established with the hinge (Glu114, His115 and Val116) plus the side chain of Asp175 and the van der Waals interactions with the side chains of Leu45, Val 53, Val66, Phe113 from the N-lobe and Ile95, Met163 and Ile174 from de C-lobe (**Fig. 3.15**). Docking calculations performed in PDB code **5CQU** as protein target with our set of compounds predicted an overall similar binding mode for the TBB and TBI moieties that differed very little from the reference TBB moiety of the bound ligand **JRJ** (4-[4-[2-[4,5,6,7-tetrakis(bromanyl)benzotriazol-2-yl]ethyl]-1,2,3-triazol-1-yl]butan-1-amine). The main differences arouse from the orientation of the linker and hydroxamic acid of the set of molecules. In this line, the general binding mode of this series is the hydroxamic acid pointing towards the catalytic loop (**Fig. 3.15**) establishing different hydrogen bonds with the side chains of Lys158, Asn161, and Asp175. The only exception is compound **7b** with the hydroxamic (**Fig. 3.15**) pointing to the αD-pocket and establishing hydrogen bond interactions with the side chains of Asn118 and Asp120 and the backbone of His160. The difference between the predicted binding mode of compounds **11c, 11b, 11a, 11d, 7d, 7a, 7c, 15a, 15b** and **19** comes from the length and adjustment of the aliphatic or aromatic linkers inside the ATP binding site. Compounds **15a,** and **15b,** were only able to reach the catalytic loop through the slight rotation of the tetrabromo benzyl moiety, but **11c, 11b, 11a, 11d, 7d, 7a, 7c** and **19** preserve the position of all bromine atoms compared to **JRJ** from the reference crystal structure. In order to better understand the stability of the different binding modes and rationalize the in vitro results, all complexes were submitted to 20ns MD simulations and the ligand RMSD values were monitored along the simulation time (**Fig. 3.14**). The binding

mode of compounds **7a, 7b, 7c, 11d, 11a, 11b** and **11c** is preserved along the MD simulation, giving a mean RMSD value lower than 2.5 Å with slight variations between them (**Table 3.7**). These compounds establish a hydrogen bond between the protonated hydroxamic moiety and the side chain of Asp155 from the catalytic loop (**Fig. 3.16**). Finally, compounds **15b, 15a** and **19** present the lowest RMSD values, under 1.6 Å, even though their biological activity was not so significant (**Table 3.7**). Compound **19** that contains a DMAT unit, locates the four bromine atoms perfectly positioned like the corresponding **JRJ** positions in the crystal structure. However, compound **15a**, has to twist the TBB moiety in order to fit in the binding site and compound **15b** only fixes three bromine atoms comparedto **JRJ** therefore losing the bromine orientation, which could be in agreement with their lower activity.

In HDAC1, crystal structure **5ICN**, which contains an inositol-6-phosphate unit, and a modified peptide inhibitor with a hydroxamic acid as a ZBG, was selected as a target protein for modeling purposes. The general binding mode has been established by docking techniques and further studied by MD simulations. For the series **11a, 11b, 11c,** and **11d**, the aliphatic chain was long enough to reach the metal site. (**Fig. 3.17**) The binding poses for compounds **11a**, **11b** and **11c** were similar, the DMAT moiety established π-π interactions with the side chains of Tyr204 and Phe205, while **11c** interacted with the side chains of Leu271 and Phe150, all on the surface of the protein. Compounds bearing the triazol linker **7a, 7b, 7c** and **7d** establish a double π-π interaction with the side chains of Phe205 and Phe150, both of which line the access tunnel to the metal site. (**Fig. 3.17**) The DMAT moiety of those compounds interact alternatively with the side chain of the hydrophobic amino acids from the HDAC1 surface, Pro29, Tyr204 and Leu271. Finally, compounds **15b, 15a** and **19** were able to reach the metal site, establishing strong double π-π interactions between the benzyl linker and the side chains of Phe205 and Phe150, while the TBB moiety interacts either with Phe205 or Leu271 of the surface of the protein. (**Fig. 3.17**) During the MD simulations the mean RMSD value for

compounds **11a, 11b, 11c** and **11d** remained under 4 Å, showing small rearrangements from the starting conformation but preserving the initial interactions described above (**Table 3.7**). The series **7a, 7b 7c** and **7d** present the biggest mean RMSD value along the simulation mainly due to the flexibility and mobility of the aliphatic triazole fraction, thus establishing hydrophobic interactions with the Pro29, Phe150, Tyr204, Phe205 and Leu271. Compounds **7a** and **7c** with the shortest chain between the triazol ring and the DMAT fraction where the most stable during the MD simulation, which is in agreement with biological results (**Fig. 3.19**). Compounds **15b**, **15a** and **19** were stable during the MD simulation, giving mean RMSD values under 2 Å after small rearrangements but preserving the previously mentioned interactions which is in accordance with their HDAC1 activity (**Fig. 3.14**).

In HDAC6, the only crystal structure available of the histone deacetylase domain is 5EDU, that was selected as target protein for modeling purposes. For the **7** series, compounds **7b** and **7d** reproduce the Trichostatin pose and compounds **7a** and **7c** turn because of the interaction between the DMAT fraction and the sidechains of Phe679 and Leu749 (**Fig. 3.18**). For the **11** series**,** due to the flexible aliphatic chain, more variety of binding modes were obtained. Compounds **11a** and **11d**, present interactions with Phe679 and Met682 while **11b** mainly with Leu749. Compound **11c**, due to its larger aliphatic chain, can reach distant amino acids, in a similar way as Trichostatin, interacting with His500, Pro501 and Leu749. Compounds **15a**, **15b** and **19** established π-π interactions between their phenyl moiety with both Phe620 and Phe680; additionally, **15a** interacted with Phe679 and Met682, while **15a** interacte with Pro501 and Leu749, and **19** just with Leu749.

MD simulations were carried out for all complexes and the general low RMSD values (**Table 3.7**) account for the high stability of these designed ligands in complex with HDAC6 (**Fig. 3.20**). For the **7** series, the most stable complexes were **7b** and **7d** with a mean RMSD value lower than 2.19Å,

whereas **7a** and **7c** show higher values and increased standard deviations. In the case of **11** series**,** due to the high flexibility of the aliphatic chain, the mean RMSD values increase even though the complexes remain highly stable along the simulation time. Compounds **11b** and **11c** preserved the docking pose along the entire simulation without significant changes. Finally, compounds **15a**, **15b** and **19** present mean RMSD values ranging between 0.55 Å and 2.12Å preserving the above-mentioned interactions during the entire simulation. In general, this lower RMSD values are in agreement with the in vitro data, and confirm the proposed binding mode by the previous docking studies (**Fig. 3.14**).

The extremely high similarity of the active site of both HDACs leads to slight differences between the modelling results. As shown in **Table 3.7**, the mean RMSD values are in general lower in HDAC6 than in HDAC1, this can be explained by the crystal structures used in these simulations. HDAC1 has been crystalized with a peptide bound to the active site, thus giving a more open conformation than HDAC6 that has been crystalized with a small molecule. As a consequence, ligands bound to HDAC1 require more simulation time to rearrange in the complex. Given that the amino acids implicated in the interactions with the ligands in both complexes are the same, selectivity between HDAC1 and HDAC6 inhibitions remains a great challenge.

3.7. Methods

For CK2 and HDAC6 we have deleted water and co-crystalized molecules, keeping the ligand as box center for the later GRID preparation. In HDAC1, the bound ligand to the holo structure was removed as well as water molecules. Moreover, protonated states of the side chains of titratable groups were established by the Propka module of Maestro [53], fixing the His140 at the protonated estate. The Glide Grid (Glide, Schrödinger, LLC, New York, NY, 2018) module was used to get the grid-box with a 20 Å length, using as center JRJ compound for CK2, TSN (Trichostatin A) for

HDAC6, and the catalytic Zn^{2+} for HDAC1, applying the versatile metal binding restriction to the zinc ion in both HDACs. Note must be taken that when using the ligand as box center, the program automatically deletes the ligand for the grid generation. The set of described ligands were prepared with the LigPrep module of Maestro at pH 7 ± 4, to obtain all possible hydroxamic protonation states and the metal binding site identification. Ligands with a protonated hydroxamic acid were used for docking on CK2, whereas ligands with the deprotonated hydroxamate were docked in HDAC1 and HDAC6. The docking calculations were performed with the Glide module using the XP algorithm [54], applying a van der Waals radii scale factor of 1.0/0.8 and establishing halogen atoms as acceptors. Finally, visual inspection was used to select the most plausible ligand conformation on the receptor, which was subsequently used for MD simulations.

The geometry optimization and charge distributions for CK2 ligands were calculated quantum mechanically (RHF/6-31+G**) with Gaussian 09 (Esto es una referencia), using the general AMBER force field 2 (GAFF2) to assign bonded and nonbonded parameters [55]. In HDAC1 and HDAC6, binding metal sites as well as the ligand were parametrized with the MCPB module embedded into the MTK++ software package in AMBER16 [56], as this methodology has been widely used to facilitate the modeling of metal effects on metalloproteins [57]. The optimized zinc and potassium metal ions coordination sphere was obtained by geometry optimization in the gas phase using Gaussian09 at B3LYP/6-31G* level, getting the equilibrium values of bond lengths, angles and force constants of the atoms bound to the metal. Classical molecular dynamics (MD) simulations were performed on all complexes, with the AMBER16 program (http://ambermd.org/). The classic ff14SB AMBER force field [58] for the protein parametrization was applied together with previous active site and ligand parameter values. After that, systems were minimized in vacuum, in order to release possible undesired interactions. Complexes were then embedded in a TIP3P water octahedron of approximately 10 000 to 13 000 water molecules. System neutrality was forced by adding sodium and chlorine ions in all systems.

Water molecules and counter ions were minimized and then the embedded system was heated to 300 K for 25 ps at constant volume keeping protein restrained to the initial positions using quadratic harmonic restraints with a constant force of 50 kcal·mol^{-1}Å^{-2}. In all systems the hydrogen bond lengths were kept at their equilibrium distance by means of the SHAKE algorithm [59]. Atom pair distance cutoffs were applied at 10.0 Å to compute the van der Waals interactions, while long-range electrostatics were computed by means of Particle-Mesh Ewald (PME) method [60]. Finally, MD simulation production was performed up to 20 ns using the thermostat NPT ensemble at 300 K, generating snapshots each 20 ps for further analysis of both systems. The trajectories of all complexes were collected and analyzed by the cpptraj module of AMBER16 [61,62].

Table 3.7. *Mean RMSD values and standard deviations (Å) obtained during the 20 ns MD simulation for complexes with HDAC1, HDAC6 and CK2.*

Compound	HDAC1	HDAC6	CK2
11d	3.74 ± 0.21	3.11 ± 0.67	2.37 ± 0.61
11c	2.78 ± 0.29	1.62 ± 0.97	1.96 ± 0.28
11b	1.94 ± 0.25	2.14 ± 0.30	1.87 ± 0.28
11a	1.71 ± 0.47	2.43 ± 0.47	1.59 ± 0.30
7d	4.82 ± 0.74	1.94 ± 0.25	1.89 ± 0.31
7a	2.77 ± 0.64	3.12 ± 0.81	2.18 ± 0.27
7b	3.84 ± 0.30	2.19 ± 0.61	2.60 ± 0.57
7c	1.90 ± 0.38	2.2 ± 1.08	2.12 ± 0.61
15b	2.93 ± 0.67	0.51 ± 0.15	0.61 ± 0.19
15a	0.94 ± 0.24	1.71 ± 0.69	1.43 ± 0.25
19	1.43 ± 0.35	2.12 ±0.80	1.54 ± 0.26

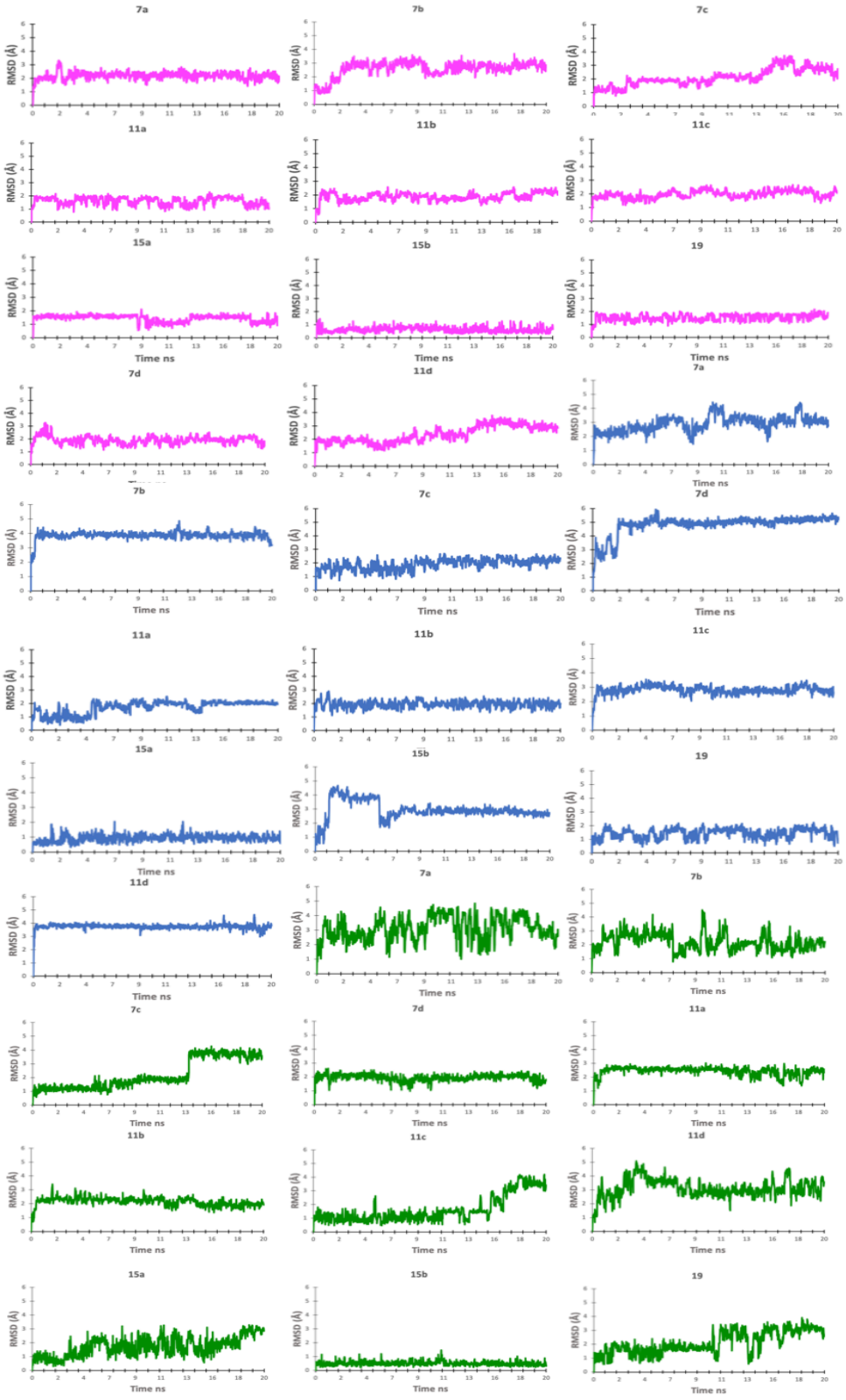


Figure 3.14. *RMSD evolution of the CK2 (magenta), HDAC1 (cyan) and HDAC6 (green) complexes with* ***7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b*** *and* ***19*** *along the 20 ns MD simulation.*

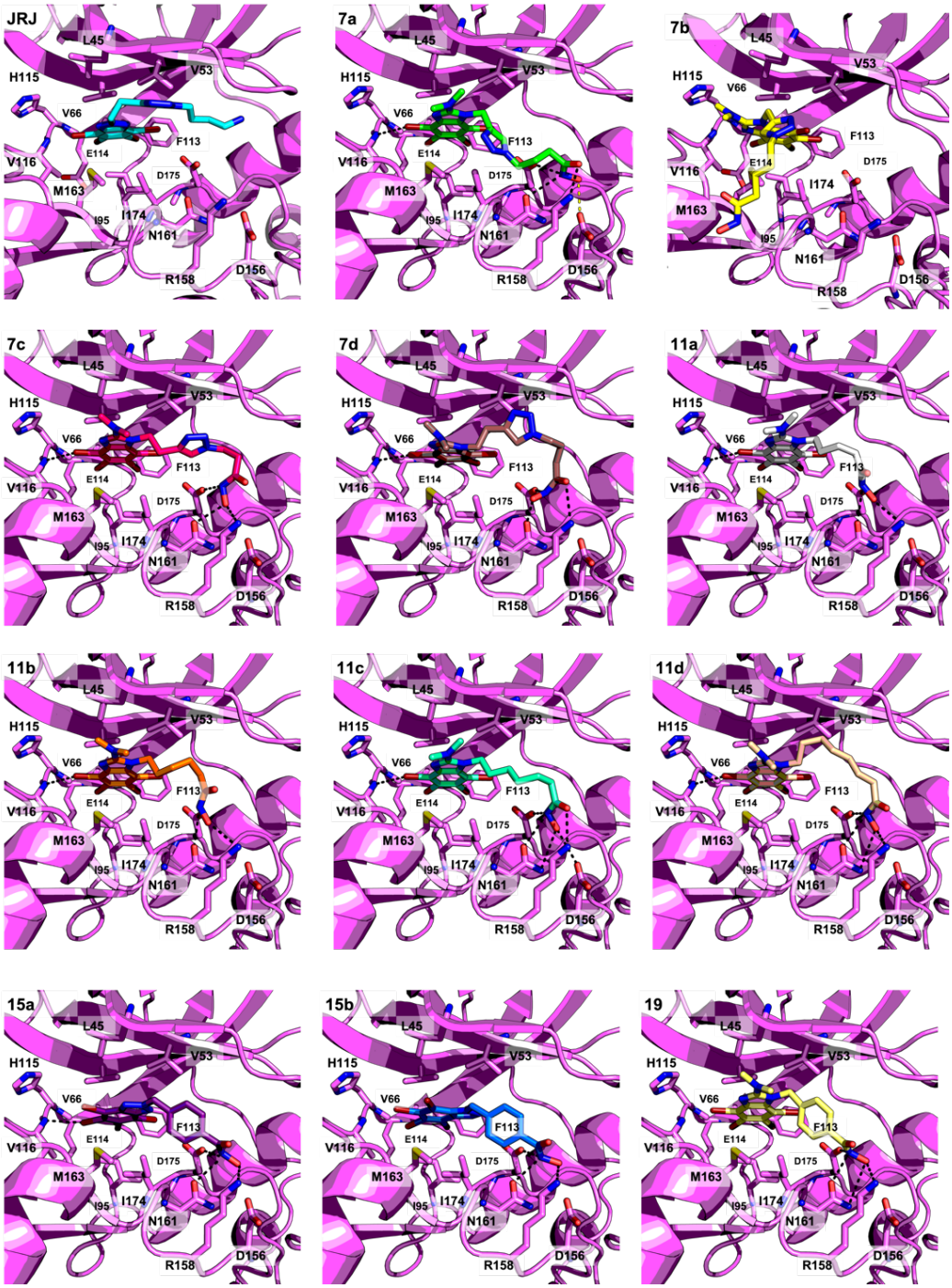


Figure 3.15. *Docking poses obtained in CK2 (**5CQU**) for compounds **7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b** and **19.** Reference compound **JRJ** is also shown for comparison.*

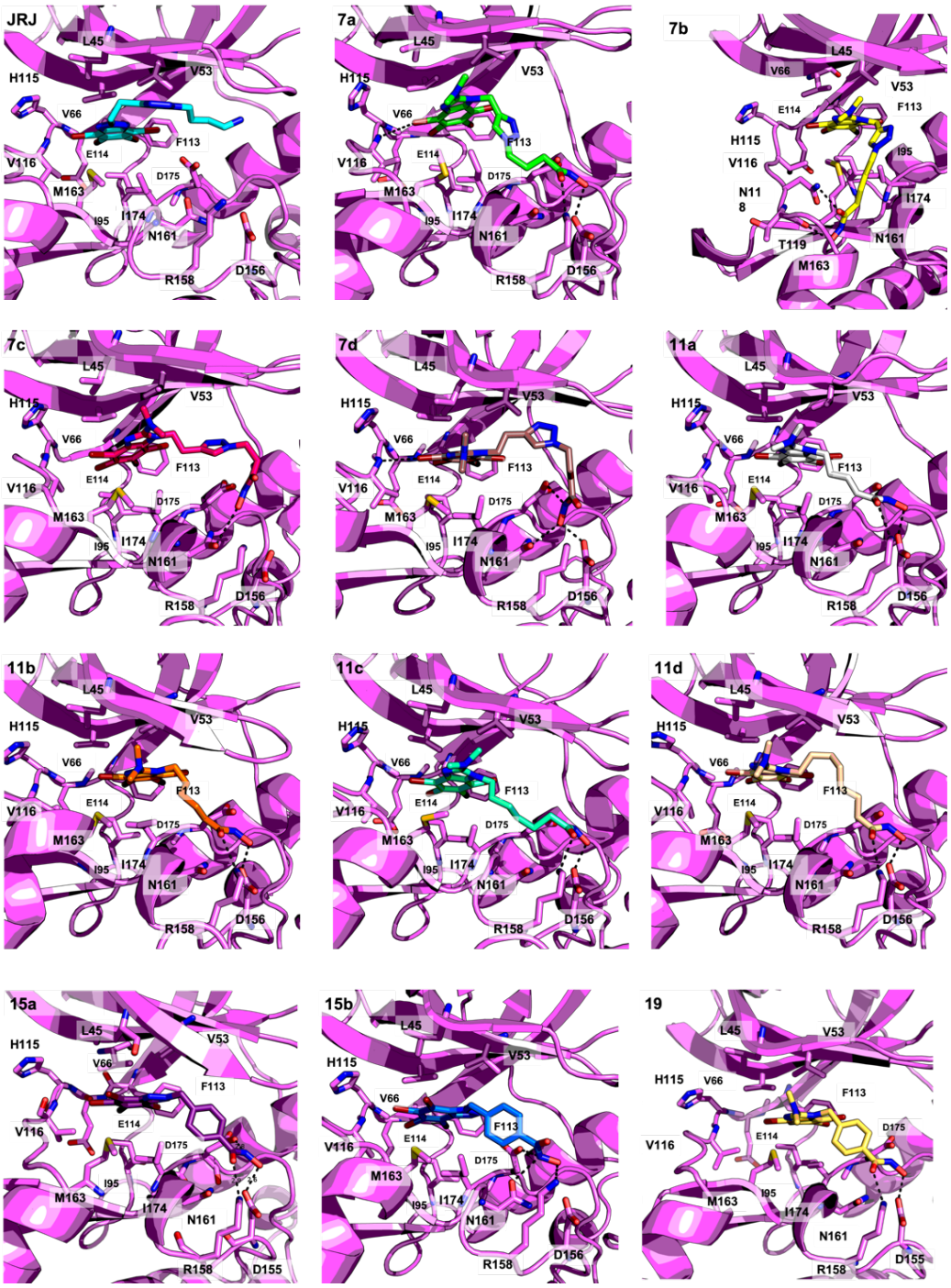


Figure 3.16. *Most populated conformers of CK2 complexes with* ***7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b*** *and* ***19*** *extracted from the 20 ns MD simulations.*

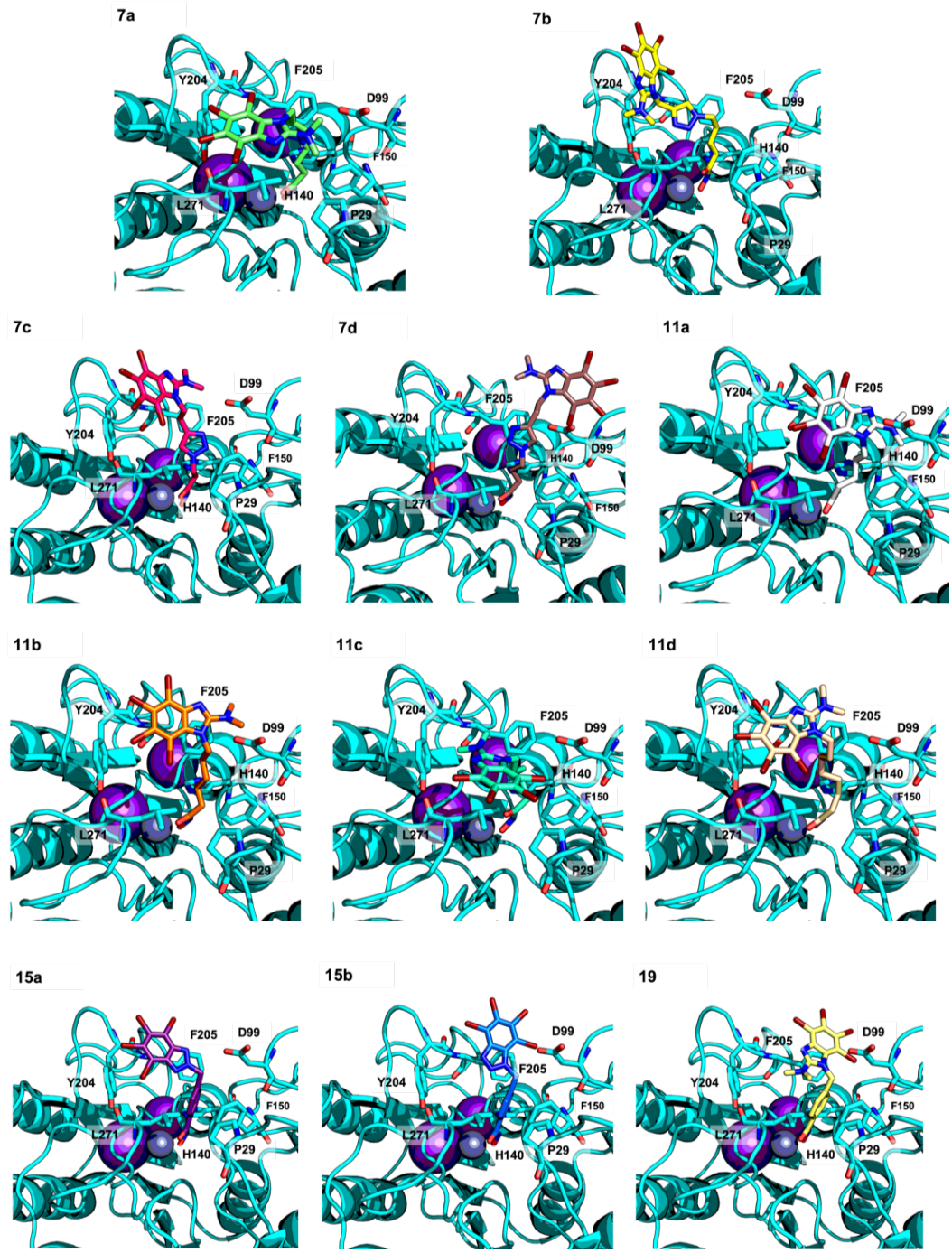


Figure 3.17. *Docking poses obtained in HDAC1 (5CQU) with compounds* ***7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b*** *and* ***19.***

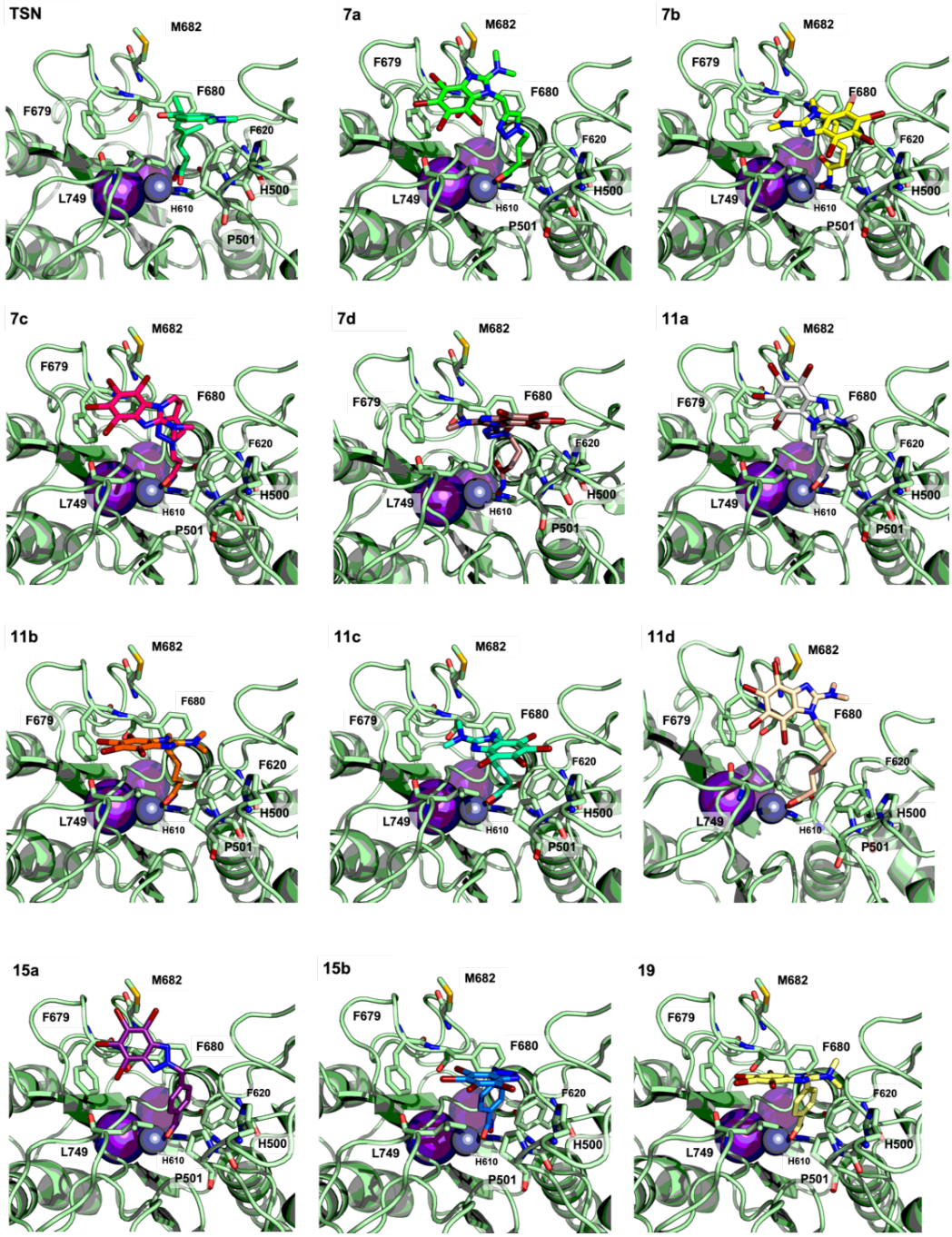


Figure 3.18. *Docking poses obtained in HDAC6 (**5EDU**) with compounds **7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b** and **19.** Reference compound Trichostatin (**TSN**) is also shown for comparison.*

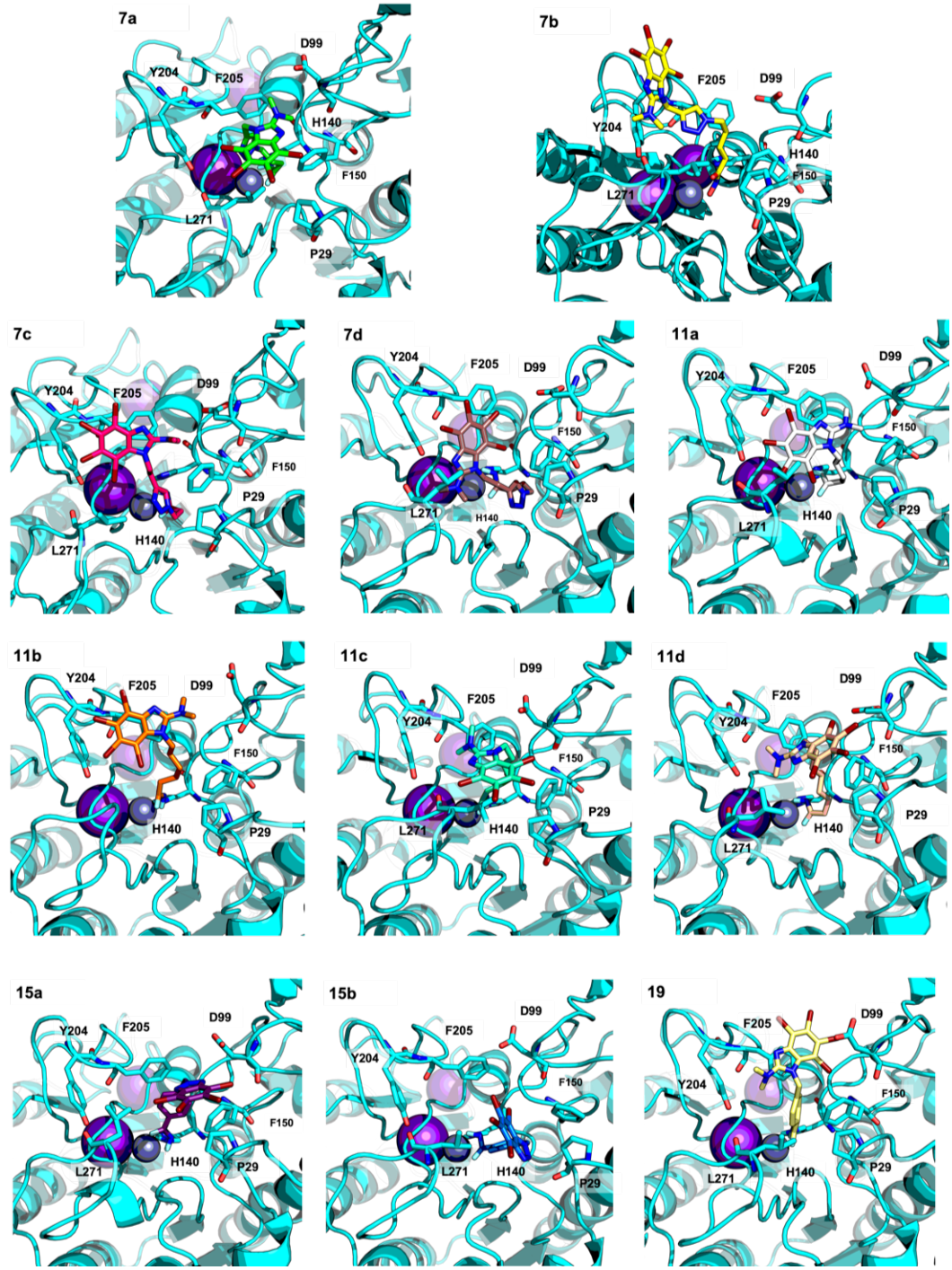


Figure 3.19. *Most populated conformers of HDAC1 complexes with* ***7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b*** *and* ***19*** *extracted from the 20 ns MD simulations.*

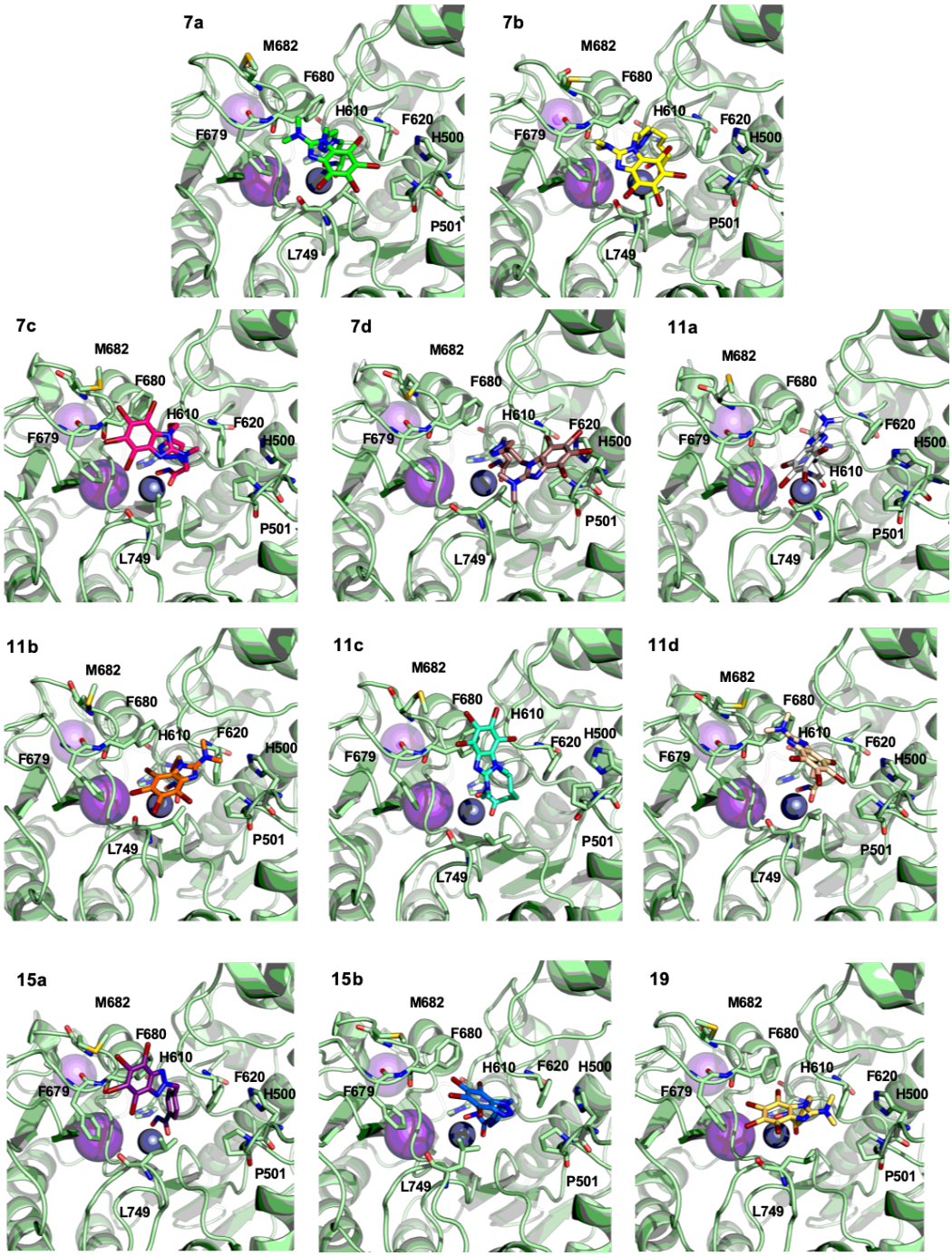


Figure 3.20. *Most populated conformers of HDAC6 complexes with* ***7a, 7b, 7c, 7d, 11a, 11b, 11b, 11c, 11d, 15a, 15b*** *and* ***19*** *extracted from the 20 ns MD simulations.*

3.8. Conclusions

In this project, we have demonstrated that a multitarget approach can provide interesting lead anticancer compounds. In silico studies showed plausible binding modes for all the designed compounds in spite of their different dynamic behaviors. Moreover, the proposed binding modes and MD simulations of these compounds relies in the well-defined stability of those; specifically the TBB moiety that remains stable in CK2, and the hydroxamic acid that is also specifically attached to the catalytic zinc atom. Nevertheless, among all of them, there is not an accurate correlation between the stability in MD simulation by the RMSD values and the in vitro activity obtained. The best result for a dual CK2/HDAC1 inhibitor in the enzymatic assays was found for **11b** (CK2 IC_{50} = 1.46 μM; HDAC1 IC_{50} = 3.67 μM), a DMAT derivative with a five-carbon atom chain connecting the hydroxamate moiety. This fact highlights the exact length of the chain to match both moieties (DMAT and hydroxamic acid) and suggests the lack of activity upon adding the triazol moiety. Furthermore, this compound was found to be even more active against HDAC6 (IC_{50} = 0.86 μM). These compounds showed a promising in vitro antiproliferative activity, with LC_{50} in the order of 4-23 μM in several tumor cell lines. Differences in LC_{50} values for tested compounds were observed. DMAT derived hydroxamates (**7c** and **11b-d**) showed 2-fold higher activity towards various types of tumor cells compared to aromatic containing linker derivatives (**15a** and **19**). Moreover, the length of the aliphatic chain in **11b-d** inhibitors influenced cytotoxic activity towards tumor cells. In particular, **11b** with a five carbon aliphatic chain was the weakest compound (LC_{50} = 4.22-23.33 μM depending on cell line), while **11c** (six-carbon chain) demonstrated higher activity (LC_{50} in 1.90-16.08 μM), and **11d** (seven-carbon chain) was the most active among all tested compounds (LC_{50} in 3.10-8.41 μM). Introduction of a triazole ring in the linker (**7c**), led to a decreased antiproliferative activity in all cell lines tested. HEK293 cell line was found to be more resistant to the action of selected compounds. Studies on the mechanism of cell death induction for **11b** showed apoptotic activity and G1/S growth arrest in Jurkat T-leukemia

cells. Compound **11d,** showing lower enzymatic activity against the three protein targets (CK2 IC_{50} = 5.89 μM; HDAC1 IC_{50} = 13.7 μM; HDAC6 IC_{50} = 8.98 μM), unexpectedly provided the best cellular activity with LC_{50} in 3-8 μM range. This difference in cellular behavior may be attributed to the increase in lipophilicity brought about by the longer seven-membered chain of **11d** compared to the five-carbon linker of **11b**. With all these data, we can conclude that both compounds **11b** and **11d** constitute promising candidates as anticancer agents and have been selected for further in vivo studies on animal tumor models.

Altogether, this work further highlighted that this multitarget approach could be a valid strategy for drug discovery in the field of anticancer drug development.

3.9. References

1. Bolognesi, M. L.; Cavalli, A. Multitarget Drug Discovery and Polypharmacology. *ChemMedChem* **2016**, *11*, 1190–1192.

2. Bolognesi, M. L. Harnessing Polypharmacology with Medicinal Chemistry. *ACS Med. Chem. Lett.* **2019**, *10*, 273–275.

3. Ramsay, R. R.; Popovic-Nikolic, M. R.; Nikolic, K.; Uliassi, E.; Bolognesi, M. L. A perspective on multi-target drug discovery and design for complex diseases. *Clin. Transl. Med.* **2018**, *7*, 3–17.

4. Lu, J.-J.; Pan, W.; Hu, Y.-J.; Wang, Y.-T. Multi-Target Drugs: The Trend of Drug Research and Development. *PLoS One* **2012**, *7*, 40262–40268.

5. Nagar, B. c-Abl Tyrosine Kinase and Inhibition by the Cancer Drug Imatinib (Gleevec/STI-571). *J. Nutr.* **2007**, *137*, 1518–1523.

6. Pluemsampant, S.; Safronova, O. S.; Nakahama, K.; Morita, I. Protein kinase CK2 is a key activator of histone deacetylase in hypoxia-associated tumors. *Int. J. Cancer* **2008**, *122*, 333–341.

7. Eom, G. H.; Cho, Y. K.; Ko, J.-H.; Shin, S.; Choe, N.; Kim, Y.; Joung, H.;

Kim, H.-S.; Nam, K.-I.; Kee, H. J.; et al. Casein Kinase-2α1 Induces Hypertrophic Response by Phosphorylation of Histone Deacetylase 2 S394 and its Activation in the Heart. *Circulation* **2011**, *123*, 2392–2403.

8. Wu, L.; Yu, J.; Chen, R.; Liu, Y.; Lou, L.; Wu, Y.; Huang, L.; Fan, Y.; Gao, P.; Huang, M.; et al. Dual Inhibition of Bcr-Abl and Hsp90 by C086 Potently Inhibits the Proliferation of Imatinib-Resistant CML Cells. *Clin. Cancer Res.* **2015**, *21*, 833–843.

9. de Castro, I. J.; Amin, H. A.; Vinciotti, V.; Vagnarelli, P. Network of phosphatases and HDAC complexes at repressed chromatin. *Cell Cycle* **2017**, *16*, 2011–2017.

10. Dokmanovic, M.; Clarke, C.; Marks, P. A. Histone Deacetylase Inhibitors: Overview and Perspectives. *Mol. Cancer Res.* **2007**, *5*, 981–989.

11. Wang, W.-Y.; Pan, L.; Su, S. C.; Quinn, E. J.; Sasaki, M.; Jimenez, J. C.; Mackenzie, I. R. A.; Huang, E. J.; Tsai, L.-H. Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. *Nat. Neurosci.* **2013**, *16*, 1383–1391.

12. Ropero, S.; Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Mol. Oncol.* **2007**, *1*, 19–25.

13. Xu, W. S.; Parmigiani, R. B.; Marks, P. A. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene* **2007**, *26*, 5541–5552.

14. Matalon, S.; Rasmussen, T. A.; Dinarello, C. A. Histone Deacetylase Inhibitors for Purging HIV-1 from the Latent Reservoir. *Mol. Med.* **2011**, *17*, 466–472.

15. Duvic, M.; Vu, J. Vorinostat: a new oral histone deacetylase inhibitor approved for cutaneous T-cell lymphoma. *Expert Opin. Investig. Drugs* **2007**, *16*, 1111–1120.

16. Li, W.; Sun, Z. Mechanism of Action for HDAC Inhibitors—Insights from Omics Approaches. *Int. J. Mol. Sci.* **2019**, *20*, 1616–1637.

17. Pflum, M. K. H.; Tong, J. K.; Lane, W. S.; Schreiber, S. L. Histone

Deacetylase 1 Phosphorylation Promotes Enzymatic Activity and Complex Formation. *J. Biol. Chem.* **2001**, *276*, 47733–47741.

18. Ammanamanchi, S.; Freeman, J. W.; Brattain, M. G. Acetylated Sp3 Is a Transcriptional Activator. *J. Biol. Chem.* **2003**, *278*, 35775–35780.

19. Qiu, Z.; Ghosh, A. A Calcium-Dependent Switch in a CREST-BRG1 Complex Regulates Activity-Dependent Gene Expression. *Neuron* **2008**, *60*, 775–787.

20. Ruijter, A. J. M. de; Gennip, A. H. van; Caron, H. N.; Kemp, S.; Kuilenburg, A. B. P. van Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* **2003**, *370*, 737–749.

21. Ferreira de Freitas, R.; Harding, R. J.; Franzoni, I.; Ravichandran, M.; Mann, M. K.; Ouyang, H.; Lautens, M.; Santhakumar, V.; Arrowsmith, C. H.; Schapira, M. Identification and Structure–Activity Relationship of HDAC6 Zinc-Finger Ubiquitin Binding Domain Inhibitors. *J. Med. Chem.* **2018**, *61*, 4517–4527.

22. Hubbert, C.; Guardiola, A.; Shao, R.; Kawaguchi, Y.; Ito, A.; Nixon, A.; Yoshida, M.; Wang, X.-F.; Yao, T.-P. HDAC6 is a microtubule-associated deacetylase. *Nature* **2002**, *417*, 455–458.

23. Kang, H.-J.; Lee, M.-H.; Kang, H.-L.; Kim, S.-H.; Ahn, J.-R.; Na, H.; Na, T.-Y.; Kim, Y. N.; Seong, J. K.; Lee, M.-O. Differential Regulation of Estrogen Receptor Expression in Breast Cancer Cells by Metastasis-Associated Protein 1. *Cancer Res.* **2014**, *74*, 1484–1494.

24. Bian, Y.; Song, C.; Cheng, K.; Dong, M.; Wang, F.; Huang, J.; Sun, D.; Wang, L.; Ye, M.; Zou, H. An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J. Proteomics* **2014**, *96*, 253–262.

25. Fleuren, E. D. G.; Zhang, L.; Wu, J.; Daly, R. J. The kinome „at large“ in cancer. *Nat. Rev. Cancer* **2016**, *16*, 83–98.

26. Bortolato, A.; Cozza, G.; Moro, S. Protein kinase CK2 inhibitors: emerging anticancer therapeutic agents? *Anticancer. Agents Med. Chem.*

2008, *8*, 798–806.

27. Cozza, G.; Meggio, F.; Moro, S. The Dark Side of Protein Kinase CK2 Inhibition. *Curr. Med. Chem.* **2011**, *18*, 2867–2884.

28. Pinna, L. A. Protein kinase CK2: a challenge to canons. *J. Cell Sci.* **2002**, *115*, 3873–3878.

29. Johnston, I. M.; Allison, S. J.; Morton, J. P.; Schramm, L.; Scott, P. H.; White, R. J. CK2 Forms a Stable Complex with TFIIIB and Activates RNA Polymerase III Transcription in Human Cells. *Mol. Cell. Biol.* **2002**, *22*, 3757–3768.

30. Keller, D. M.; Zeng, X.; Wang, Y.; Zhang, Q. H.; Kapoor, M.; Shu, H.; Goodman, R.; Lozano, G.; Zhao, Y.; Lu, H. A DNA Damage–Induced p53 Serine 392 Kinase Complex Contains CK2, hSpt16, and SSRP1. *Mol. Cell* **2001**, *7*, 283–292.

31. Kimura, Y.; Rutherford, S. L.; Miyata, Y.; Yahara, I.; Freeman, B. C.; Yue, L.; Morimoto, R. I.; Lindquist, S. Cdc37 is a molecular chaperone with specific functions in signal transduction. *Genes Dev.* **1997**, *11*, 1775–1785.

32. Götz, C.; Montenarh, M. Protein kinase CK2 in development and differentiation. *Biomed. Reports* **2017**, *6*, 127–133.

33. Shibazaki, C.; Arai, S.; Shimizu, R.; Saeki, M.; Kinoshita, T.; Ostermann, A.; Schrader, T. E.; Kurosaki, Y.; Sunami, T.; Kuroki, R.; et al. Hydration Structures of the Human Protein Kinase CK2α Clarified by Joint Neutron and X-ray Crystallography. *J. Mol. Biol.* **2018**, *430*, 5094–5104.

34. Yueh, C.; Rettenmaier, J.; Xia, B.; Hall, D. R.; Alekseenko, A.; Porter, K. A.; Barkovich, K.; Keseru, G.; Whitty, A.; Wells, J. A.; et al. Kinase Atlas: Druggability Analysis of Potential Allosteric Sites in Kinases. *J. Med. Chem.* **2019**, *62*, 6512–6524.

35. Niefind, K. Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. *EMBO J.* **2001**, *20*, 5320–5331.

36. Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H. Ten things you should

know about protein kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **2015**, *172*, 2675–2700.

37. Vijayan, R. S. K.; He, P.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *J. Med. Chem.* **2015**, *58*, 466–479.

38. Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.

39. Chen, F.; Fang, Y.; Zhao, R.; Le, J.; Zhang, B.; Huang, R.; Chen, Z.; Shao, J. Evolution in medicinal chemistry of sorafenib derivatives for hepatocellular carcinoma. *Eur. J. Med. Chem.* **2019**, *179*, 916–935.

40. Niefind, K.; Pütter, M.; Guerra, B.; Issinger, O. G.; Schomburg, D. GTP plus water mimic ATP in the active site of protein kinase CK2. *Nat. Struct. Biol.* **1999**, *6*, 1100–1103.

41. Valero, E.; De Bonis, S.; Filhol, O.; Wade, R. H.; Langowski, J.; Chambaz, E. M.; Cochet, C. Quaternary Structure of Casein Kinase 2. *J. Biol. Chem.* **1995**, *270*, 8345–8352.

42. Niefind, K.; Raaf, J.; Issinger, O.-G. Protein Kinase CK2 in Health and Disease. *Cell. Mol. Life Sci.* **2009**, *66*, 1800–1816.

43. Raaf, J.; Issinger, O.-G.; Niefind, K. First inactive conformation of CK2 alpha, the catalytic subunit of protein kinase CK2. *J. Mol. Biol.* **2009**, *386*, 1212–1221.

44. Brear, P.; De Fusco, C.; Hadje Georgiou, K.; Francis-Newton, N. J.; Stubbs, C. J.; Sore, H. F.; Venkitaraman, A. R.; Abell, C.; Spring, D. R.; Hyvönen, M. Specific inhibition of CK2α from an anchor outside the active site. *Chem. Sci.* **2016**, *7*, 6839–6845.

45. Iegre, J.; Brear, P.; De Fusco, C.; Yoshida, M.; Mitchell, S. L.; Rossmann, M.; Carro, L.; Sore, H. F.; Hyvönen, M.; Spring, D. R. Second-generation CK2α inhibitors targeting the αD pocket. *Chem. Sci.* **2018**, *9*, 3041–3049.

46. Pagano, M. A.; Bain, J.; Kazimierczuk, Z.; Sarno, S.; Ruzzene, M.; Di Maira, G.; Elliott, M.; Orzeszko, A.; Cozza, G.; Meggio, F.; et al. The selectivity of inhibitors of protein kinase CK2: an update. *Biochem. J.* **2008**, *415*, 353–365.

47. Prudent, R.; Cochet, C. New Protein Kinase CK2 Inhibitors: Jumping out of the Catalytic Box. *Chem. Biol.* **2009**, *16*, 112–120.

48. Cozza, G. The Development of CK2 Inhibitors: From Traditional Pharmacology to in Silico Rational Drug Design. *Pharmaceuticals* **2017**, *10*, 26–49.

49. Hotha, S.; Anegundi, R. I.; Natu, A. A. Expedient synthesis of 1,2,3-triazole-fused tetracyclic compounds by intramolecular Huisgen ('click') reactions on carbohydrate-derived azido-alkynes. *Tetrahedron Lett.* **2005**, *46*, 4585–4588.

50. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chemie Int. Ed.* **2002**, *41*, 2596–2599.

51. Strober, W. Trypan Blue Exclusion Test of Cell Viability. *Current Protocols in Immunology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.

52. Strober, W. Trypan Blue Exclusion Test of Cell Viability. *Current Protocols in Immunology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; p. 1–3.

53. Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinforma.* **2005**, *61*, 704–721.

54. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

55. Vassetti, D.; Pagliai, M.; Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1983–1995.

56. Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599–604.

57. Torras, J.; Maccarrone, M.; Dainese, E. Molecular dynamics study on the Apo - and Holo -forms of 5-lipoxygenase. *Biotechnol. Appl. Biochem.* **2018**, *65*, 54–61.

58. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

59. Yoneya, M.; Berendsen, H. J. C.; Hirasawa, K. A Non-Iterative Matrix Method for Constraint Molecular Dynamics Simulations. *Mol. Simul.* **1994**, *13*, 395–405.

60. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

61. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N. AMBER 2016 Reference Manual. *Univ. Calif. San Fr. CA, USA* **2016**, 1–923.

62. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

CHAPTER IV

4. A Rational Design for CK2-Directed PROTACS

4.1. Introduction

A PROteolysis TArgeting Chimera (PROTAC) is a heterofunctional small molecule able to deplete the target protein through selective ubiquitination (**Fig. 4.1**). The exceptional nature of a PROTAC molecule lies in its capacity to degrade the target instead of inhibiting it, as the majority of drugs do. This is a very attractive characteristic, as it can make almost any target druggable. As *Alessio Ciulli*, a biochemist at the University of Dundee, said "We're breaking the rules of what we thought would be druggable" [1].

PROTACs are made-up by two covalently linked protein-binders: one specifically designed to bind the E3 ligase responsible for the start of the ubiquitination; and, the other to the target protein meant for degradation, also known as Protein Of Interest (POI). So, the PROTAC mechanism of action is to favor the binding between the POI and the E3 ligase in charge of its polyubiquitination so it is recognized by the degradome

The E3 ligase is part of the Ubiquitin-Proteasome system (UPS) which is made-up by three ligases E1, E2 and E3. The activating enzyme E1 binds ubiquitin and transfer it to E2. The E2 activated with ubiquitin binds to the E3 ligase forming a complex that identifies the POI and adds a polyubiquitin tag to it. Finally, the E2-E3-POI complex will be degraded by 26S proteasome (**Fig. 4.2**) [2–5].

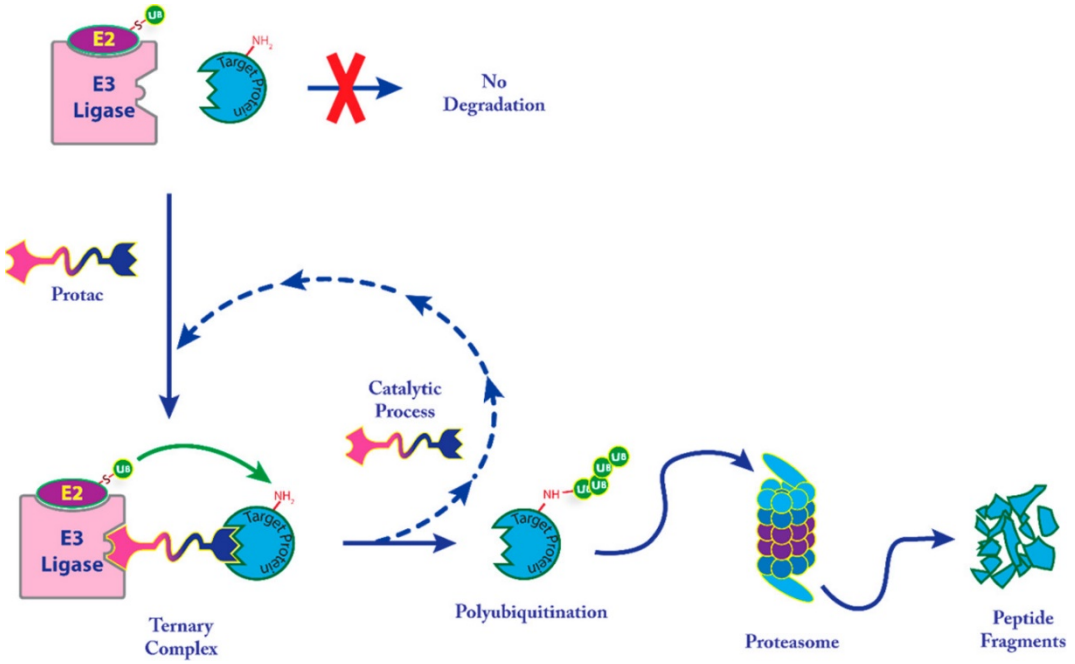


Figure 4.1. *General PROTAC mechanism of action. Figure taken from J. Med. Chem.* ***2018****, 61, 2, 444-452.*

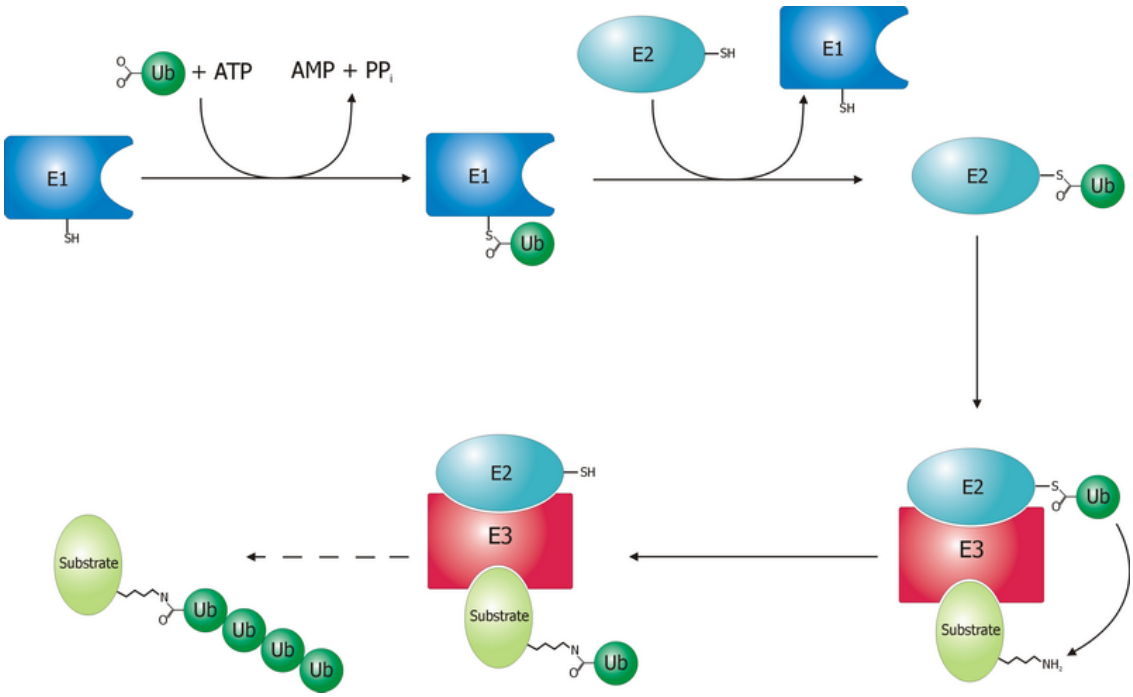


Figure 4.2. *Schematic diagram of the ubiquitination system Created by Roger B. Dodd from Wikipedia.*

4.2. PROTAC Design and examples

The rational design of small molecule PROTACs has four main concerns: the POI binder, the ligase binder, the linker, and the ADME properties.

The POI binder. Commonly during the PROTAC design, the POI ligand is already well known and characterized [6], this has led to a vast majority of reported PROTACs based in competitive inhibitors; nevertheless, allosteric and non-competitive PROTACs have also been described [7].

The E3 ligase binder. It has been estimated that there are about 500-1000 E3 ligases in humans, which impart substrate specificity onto different types of E1 and E2 ligases [8]. However, amongst all of them, only four groups of E3 ligases have shown to present potential as PROTAC targets. Interestingly, each group of E3 ligase presents different binders (**Fig. 4.3**), which is very useful for achieving ligase selectivity:

- CRBN or Cereblon (UniProtKB - Q96SW2-CRBN_HUMAN) has know binders such as thalidomide, lenalidomide, pomalidomide, and derivatives.
- MDM2 (UniProtKB - Q00987-MDM2_HUMAN) that binds to nutlin.
- c-IAP1 or Baculoviral IAP repeat-containing protein 2 (UniProtKB - Q13490-BIRC2_HUMAN) with binders such as Bestatin and IAP-7.
- VHL or von Hippel-Lindau disease tumor suppressor (UniProtKB - P40337-VHL_HUMAN) with the known ligand VHL-9.

Even though there are so many E3 ligase ligands, the vast majority of PROTACs are based on VHL or CRBN binders and the ones targeting MDM2 and cIAP1 are not effective in protein depletion [9,10].

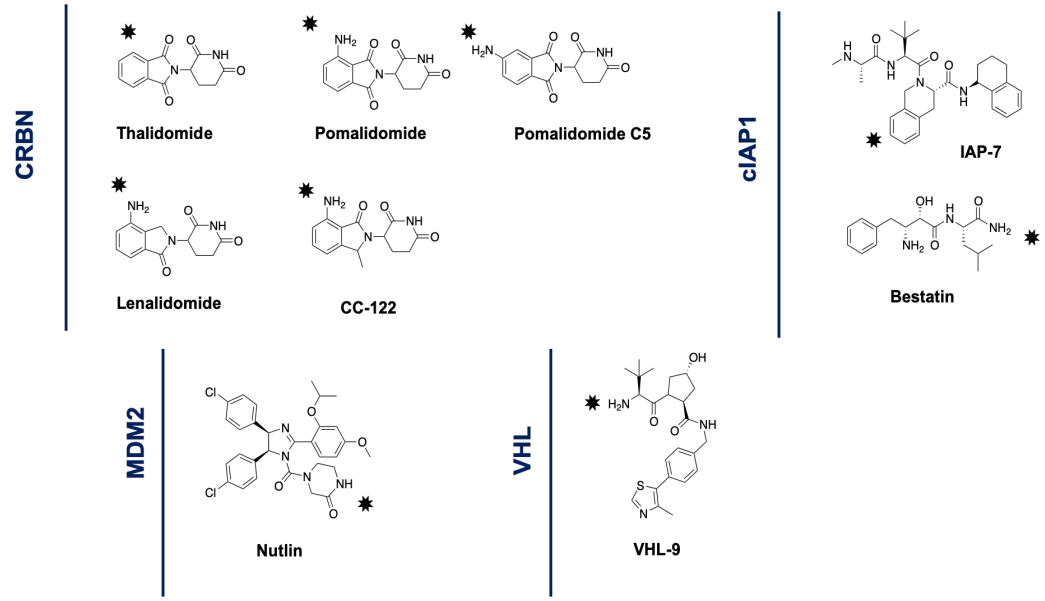


Figure 4.3. The chemical structure of E3 ligase ligands used for PROTAC design. The asterisk shows the attachment point for the linker. Figure taken from Comput Struct Biotechnol J. ***2019****, 17, 160–176.*

The linker. This part covalently joins both binders. The size, orientation and composition of the linker is of equal importance and plays an important role in PROTAC efficiency. The linker positions the POI and the E3 ligase in a way that the surface complementarity increases the ppi and the stability of the complex. There are a large number of linkers published in the literature that contain either ether, amide, esther or triazol moieties. The selection of the length and composition depends on the type of the E3 ligase, POI ligand and the POI ligand-binding site location with respect to the E3 ligase [11].

The ADME properties. The design of PROTACs remains quite challenging not only for the choice of ligands and linker, but also because the resulting compound must satisfy the suitable physicochemical properties in order to reache its target. A recent publication highlighted the most suitable physicochemical parameters for PROTACs: MW = 614–1413, clog P = −2.7–9, Hydrogen Bond Donors (HBD) = 1–10, HBA = 8–23, number of rotatable bonds (NRotB) = 6–49, number of aromatic rings (Nar) = 1–7, TPSA = 124–389 $Å^2$ [12]. These numbers give us a comprehensive idea about the limits in PROTAC design.

To date, only one PROTAC (ARV-110) has been reported to reach clinical trials. It has been developed by the company Arvinas and deposited in patent. ARV-110 and is currently in phase I clinical trial (NCT03888612). This is an orally bioavailable agent designed to selectively target and degrade the androgen receptor (AR) and has been designated for the treatment of patients with metastatic castration-resistant prostate cancer (mCRPC) [13,14].

4.3. Kinase-targeted PROTACs

Kinases are one of the top druggable targets for both pharmaceutical industry and academia. Since the majority of the kinase inhibitors are intended to treat cancer, it was hypothesized that PROTACs would be more effective in cancer because of the normally increased kinase concentration in cancer cell lines [15][16]. As an example, the approved kinase inhibitors imatinib, bosutinib and dasatinib have been used as POI ligands in the design of PROTACs by conjugating them with either a VHL or CRBN binds through a range of different linkers and yielding promising in vitro results [17]. PROTACs against ERK, bromodomain and extra-terminal (BET) proteins, SMARCA FLT3, CDK9, CDK8, PI3K, HDAC, and ALK, have also been reported [16,18–26].

From a structural point of view, the interaction of CK1α (UniProtKB - P48729 KC1A_HUMAN), a very closely related to CK2 kinase, with the E3 ligase CRBN has been obtained by X-Ray crystallography and deposited in the PDB (**5FQD**) [27]. The aim of the corresponding publication and the obtained crystal structure was to shed light on the structural mechanism by which thalidomide and its derivatives, lenalidomide and pomalidomide (also known as immunomodulatory drugs, IMiDs) by themselves are able to deplete CK1α and other lymphoid transcription factors such as IKZF1 and IKZF3 [28,29]. The **5FQD** crystal structure, with 2.45 Å resolution, in which

DDB1–CRBN bound to lenalidomide interacts with CK1α, provides the first structural insight into how lenalidomide stabilizes the contact between both proteins leading to ubiquitination and degradation. DDB1, also known as DNA damage-binding protein 1, together with CUL4A recruits the E2 ligase, and the DDB1-CRBN complex is necessary to recruit the rest of the proteasome system [30].

The binding of lenalidomide inside the CRBN thalidomide-binding site induces a conformational change recognized by the β-hairpin loop (aa 35-41) of CK1α stabilizing the ppi (**Fig. 4.4**). In this way, CRBN and lenalidomide together provide the interaction surface for CK1α where the β-hairpin loop can selectively bind to CRBN-thalidomide [31]. It is also shown that the binding of IMiDs produce different ranges of protein depletion, being the one produced by lenalidomide the most potent. Interestingly, a very recent publication by *Cheng et al.* [32] highlighted the discovery of a moderately active CK2-targeting PROTAC, in which the potent inhibitor CX4945 (Chapter III) was used as ligand for the POI and linked to pomalidomide by click chemistry thus obtaining a series of compounds [32] (**Fig. 4.4.**). This series of compounds were able to degrade CK2 in a time- and dose-dependent manner, although at relatively high concentrations (in the range of 10 μM) (**Fig. 4.5.**). This specific CK2 depletion also resulted in some downstream changes, such as phosphorylation of Akt; and, reduction and upregulation of the P_{53} tumor suppressor. The results of this work highlight two main facts: i) **Cheng2** is the most potent of all the set of PROTACS; ii) a two-fold increase in compound concentration (from 5 μM to 10 μM) leads to a remarkable increase in protein degradation (from 60% to more than 90%). Both facts suggest that there is a correlation between the length of the linker and the activity as well as for the incubation concentration.

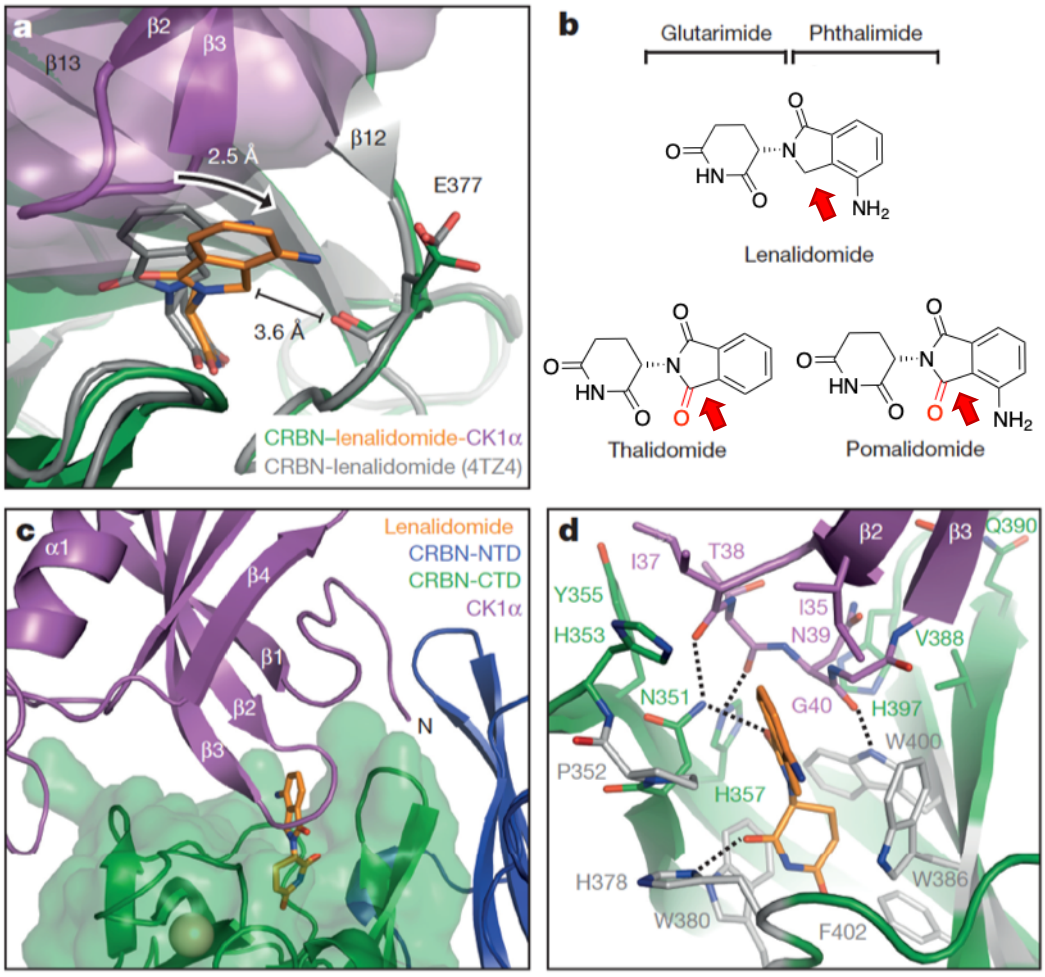


Figure 4.4. *Interactions of CK1α with CRBN and lenalidomide.* ***a)****, Superimposition of CRBN–lenalidomide–CK1α and human CRBN bound to lenalidomide (****4TZ4****).* ***b)****, Chemical structures of lenalidomide, pomalidomide and thalidomide. Arrows indicate C3 position of the phthalimide ring.* ***c)*** *CRBN binds the glutarimide moiety in a hydrophobic pocket. The phthalimide ring of lenalidomide is surface exposed. A β-hairpin loop of CK1α binds CRBN on top of its lenalidomide-binding pocket.* ***d)****, Side-chain interactions between CK1α, CRBN and lenalidomide. Lenalidomide provides van der Waals interactions with CK1α. Figure taken from Nature.* ***2016****, 532, 127–130*

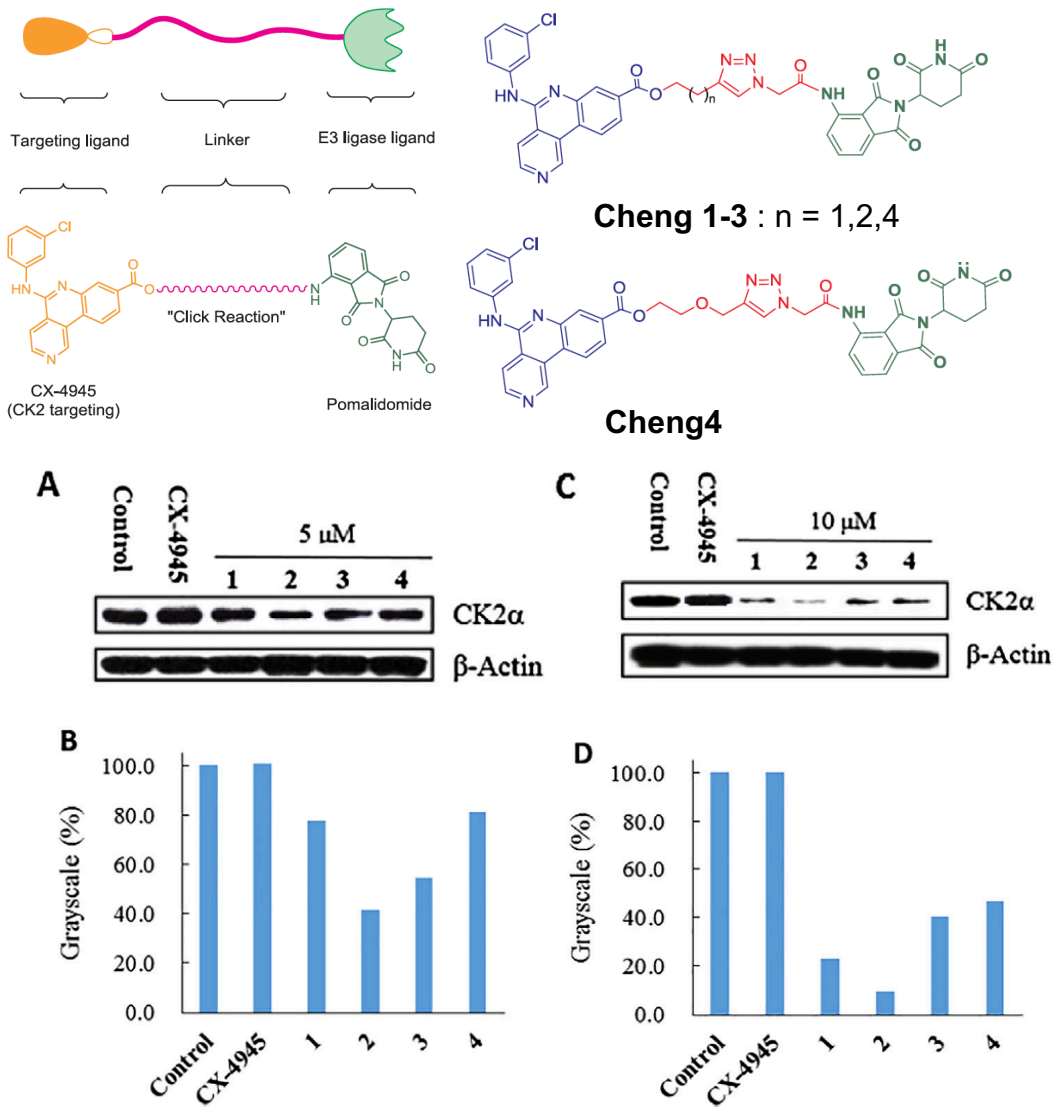


***Figure 4.5**. **Top:** Chemical structures of the PROTAC designed and synthesized by Cheng et al.*[32] ***Bottom:** Biological evaluation, CK2 protein degradation in MDA-MB-231 cells. The levels of CK2α after treatment with **CX4945** and compounds Cheng1–4 for 24 hours at 5 μM (A, B) and 10 μM (C, D), respectively. Figure taken from Bioorganic Chemistry, **2018**, 81, 536–544.*

4.4. Objectives

With the experience on the design and synthesis of dual inhibitors by our research group [33] and more recently the dual HDAC-CK2 agents (in **Chapter III**), we decided to continue and increase the research on CK2 modulation by designing CK2-targeting PROTACS. To achieve this objective:

- We have to first propose a plausible model for the interaction of CK2 with CRBN E3 ligase using as template the CRBN-lenalidomide-CK1α complex obtained by X-Ray crystallography and deposited (**5FQD**).
- Then, based on the work carried out by *H. Cheng et al.*, we aim to propose a plausible binding mode for the most active CK2-targeting PROTAC identified in this work (**Cheng2**).
- Based on our proposed models we will design a novel CK2-targeting PROTACs.

4.5. Methods

The methods applied in this Chapter for protein preparation, ligand preparation, docking calculations, and MD simulations are the same as those used in **Chapter III** with the following exceptions:

The CK1α-lenalidomide-CRBN system was obtained from the crystal structure under PDB code **5FQD**. On the other hand, the CK2-lenalidomide-CRBN system was obtained using PDB code **5FQD** as template for the superimposition of the CK2 catalytic domain to CK1α using the PyMOL alignment tool. (PDB code **3PE1** and **6GIH** without ligand bind in the binding site).

1. The linker built between the recruiting moieties has been manually built using the 3D-building module of MAESTRO [34].

2. MD simulations have been carried out during 100 ns simulations.
3. MD analysis also includes the total binding energy calculation carried out with the MM-ISMSA software [35].
4. The RMSD values were taken from the Cα for proteins and complexes and all atoms were considered for the ligand using as reference the initial conformation.

4.6. Results and Discussion

4.6.1. Modelling of CK2-CRBN complex

The first question that we had to address was if CK2 could interact with the CRBN E3 ligase just as CK1α did in the PDB code **5FQD**. For this purpose, the amino acid sequence of both CK2 and CK1α was aligned using the Clustal Omega web server. These proteins only present a 24.61% of sequence identity (**Fig. 4.6**). Nevertheless, an in-depth analysis of the β-hairpin loop responsible for the ppi of both proteins proved that the sequence was almost the same ("INITNGE" for CK1α, and "INITNNGE" for CK2). This means that the β-hairpin loop composition is almost identical with the exception of the insertion of an asparagine that displaces the position of the glycine. However, this minor change of amino acid sequence could change the entire recognition process. A recent article by *Buhimschi et al.* [36] highlighted a common motif recognized by IMiDs and CRBN, which he called the Zinc Finger Degrome. Amongst all proteins depleted by IMiDs (GSPT1, CK1α, IKZF1 ZF2 and ZNF692 ZF4) there is a common loop similar to the β-hairpin loop where the aforementioned glycine is always present in the same exact position (**Fig. 4.7**).

```
sp|CK1A|P48729|2-337    ----------------------ASSSGSKAEFIVGGKYKLVRKIGSGSFGDIYLAINIT  37
sp|CK2|P68400|1-391     MSGPVPSRARVYTDVNTHRPREYWDYESHVVEWGNQDDYQLVRKLGRGKYSEVFEAINIT  60
                                               . ..  .*:   ..*:*****:* *.:.:::: *****

sp|CK1A|P48729|2-337    NGEEVAVKLESQKARHPQLLYESKLYKILQGGVGIPHIRWYGQEKD--YNVLVMDL-LGP  94
sp|CK2|P68400|1-391     NNEKVVVKILKPVKK-KKIKREIKILENLRGGPNIITLADIVKDPVSRTPALVFEHVNNT  119
                        *.*:*.**  .   :  ::  * *: : *:** .*  :    ::       .**::   .

sp|CK1A|P48729|2-337    SLEDLFNFCSRRFTMKTVLMLADQMISRIEYVHTKNFIHRDIKPDNFLMGIGRHCNKLFL  154
sp|CK2|P68400|1-391     DFKQLY----QTLTDYDIRFYMYEILKALDYCHSMGIMHRDVKPHNVMID--HEHRKLRL  173
                        .:::*:    : :*   : :   :::. ::* *: .:::**:**.*.::.  :. .** *

sp|CK1A|P48729|2-337    IDFGLAKKYRDNRTRQH---IPYREDKNLTGTARYASINAHLGIEQSRRDDMESLGYVLM  211
sp|CK2|P68400|1-391     IDWGLAEFYHPGQEYNVRVASRYFKGPELL--VDYQMYD--------YSLDMWSLGCMLA  223
                        **:***: *: .:  :      * :. :*   . *   :           ** *** :*

sp|CK1A|P48729|2-337    YFNRTSLPW-QGLKAATKKQKYEKISEKK-------MSTPVEVLCKGFPAEFAMYLNYCR  263
sp|CK2|P68400|1-391     SMIFRKEPFFHGH------DNYDQLVRIAKVLGTEDLYDYIDKYNIELDPRFNDILGRHS  277
                         :   . *: :*       ::*::: .          :   ::      :  .*   *.

sp|CK1A|P48729|2-337    GLRFE-----------EAPDYMYLRQLFRI-------LFRTLNHQYDYTFDWTMLKQKAA  305
sp|CK2|P68400|1-391     RKRWERFVHSENQHLVSPEALDFLDKLLRYDHQSRLTAREAMEHPYF----YTVVKDQAR  333
                          *:*          .     :* :*:*          .:::* *     :*::*::*

sp|CK1A|P48729|2-337    QQAASSSGQGQQAQTPTGKQTDKTKSNMKGF------------------------------  336
sp|CK2|P68400|1-391     MGSSS---------MPGGSTPVSSANMMSGISSVPTPSPLGPLAGSPVIAAANPLGMPVP  384
                          ::*          * *.   .: . *.*:

sp|CK1A|P48729|2-337    -------        336
sp|CK2|P68400|1-391     AAAGAQQ        391
```

Figure 4.6. *Sequence alignment of kinase domain of CK1α and CK2. Red box highlights the β-hairpin loop.*

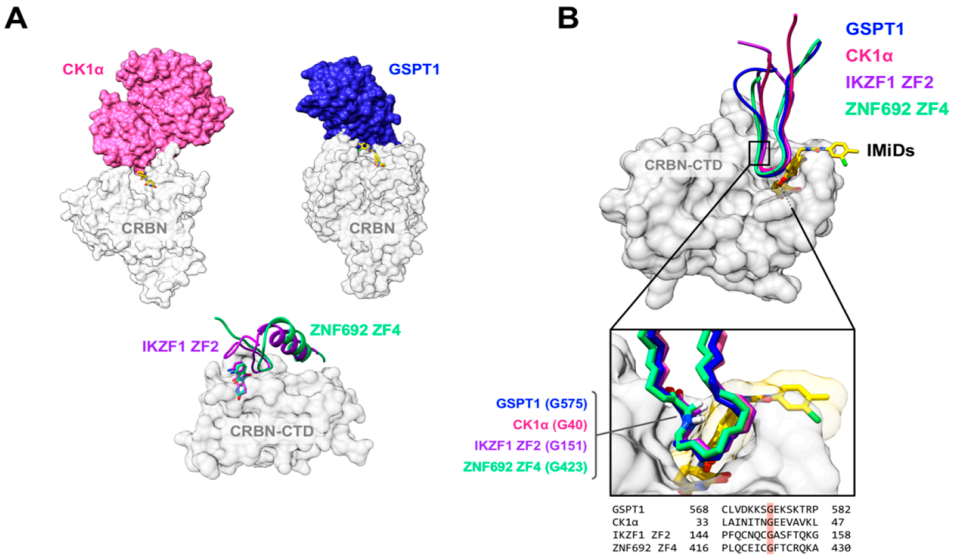


Figure 4.7. *Available crystal structures of neo-substrate−IMiD−CRBN complexes.* ***A.*** *(clockwise) CK1α in complex with lenalidomide and CRBN (**5FQD**), GSPT1 in complex with CC-885 and CRBN (**5HXB**), alignment of IKZF1 ZF2 (**6H0F**) and ZNF692 ZF4 (**6H0G**) in complex with pomalidomide and CRBN.* ***B.*** *Alignment of surface turn within all neo-substrates proposed to account for IMiD-induced degradation specificity. Inset illustrates the position of a critical glycine within neo-substrates relative to IMiDs' surfaces. Figure taken from Biochemistry. **2019**, 58, 861−864.*

Considering all this information, we decided to test if the interaction of CK2 with CRBN would be stable in the presence of lenalidomide. For this purpose the CK1α-lenalidomide-CRBN complex extracted from the crystal structure, and the CK2-lenalidomide-CRBN model were subjected to 100 ns MD simulations; and the RMSD values of the CK1-α of the kinases, CRBN and β-hairpin loop, as well as lenalidomide were monitored in order to analyze the stability of both complexes (**Table 4.1**).

These results show how the interaction between CK1α and CRBN seems to be more stable than for the CK2 complex. The mean RMSD values along the simulation for the CK2-lenalidomide-CRBN complex is 6.31 Å, whereas for the CK1-lenalidomide-CRBN complex is 2,91 Å a very long difference which can be clearly appreciated in **Figure 4.8** pointing towards the instability of this complex when compared to the analogue CK1. However, a detailed and visual inspection allowed us to observe that the ppi between both kinases and CRBN is maintained, and the β-hairpin loop remains stable. We have also observed during the simulation that in the case of the CK2-lenalidomide-CRBN complex, there is a remarkable conformational change in the tail of CRBN not related with the interaction with CK2 (**Fig. 4.9.**). This is not probably the origin of RMSD values in comparison with the CK1α-CRBN system. Nevertheless, during the MD simulations the mean RMSD values of the β-hairpin loop are similar, as well as for the lenalidomide. Additionally, MM-ISMSA analysis of the trajectories were performed to account for the effect of the different amino acid composition in the β-hairpin loop in the total binding energy. The MM-ISMSA analysis showed that the total binding energy of the ppi was similar and comparable in both systems along the simulation. From these results, we can conclude that these two models are comparable and present a similar behavior; therefore, this model was used to find a plausible binding mode for the published PROTAC **Cheng2.**

Table 4.1. *Mean RMSD values (Å) from the 100 ns MD simulations from CK1α-lenalidomide-CRBN and CK2-lenalidomide-CRBN complexes. The total binding energies values (Kcal·mol^{-1}) come from the ppi between the corresponding kinase and CRBN and LVY with the entire complex.*

System	RMSD values					Total Binding Energy	
	Complex	CRBN	Kinase	LVY	Loop	ppi	LVY
CK1α	2.91 ± 0.47	2.59 ± 0.58	2.01 ± 0.22	0.68 ± 0.16	0.20 ± 0.05	-82.55	-49.31
CK2	6.31 ± 1.67	4.44 ± 1.04	2.07 ± 0.21	1.02 ± 0.23	0.30 ± 0.06	-68.82	-45.29

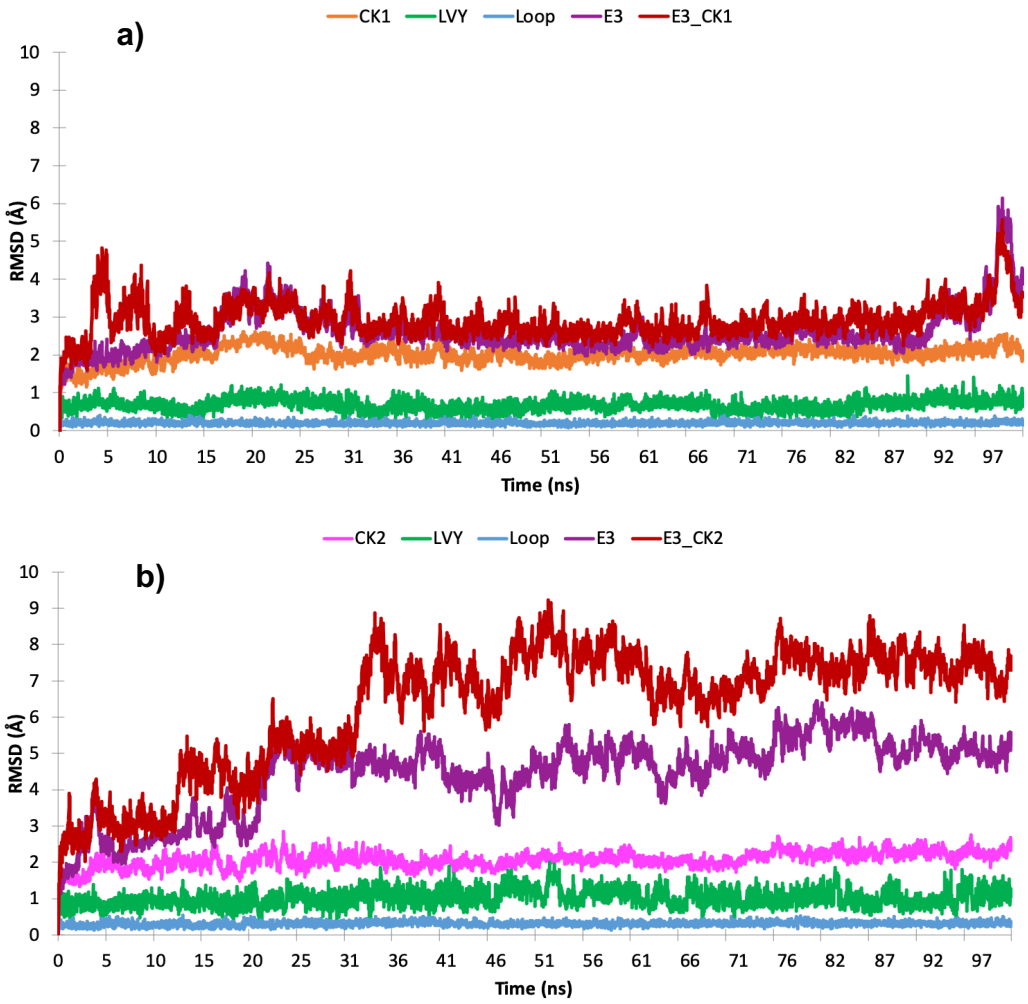


Figure 4.8. *Monitoring of the RMSD values obtained along the 100 ns MD simulation* ***a)*** *CK1α-lenalidomide-CRBN and* ***b)*** *CK2-lenalidomide-CRBN complexes.*

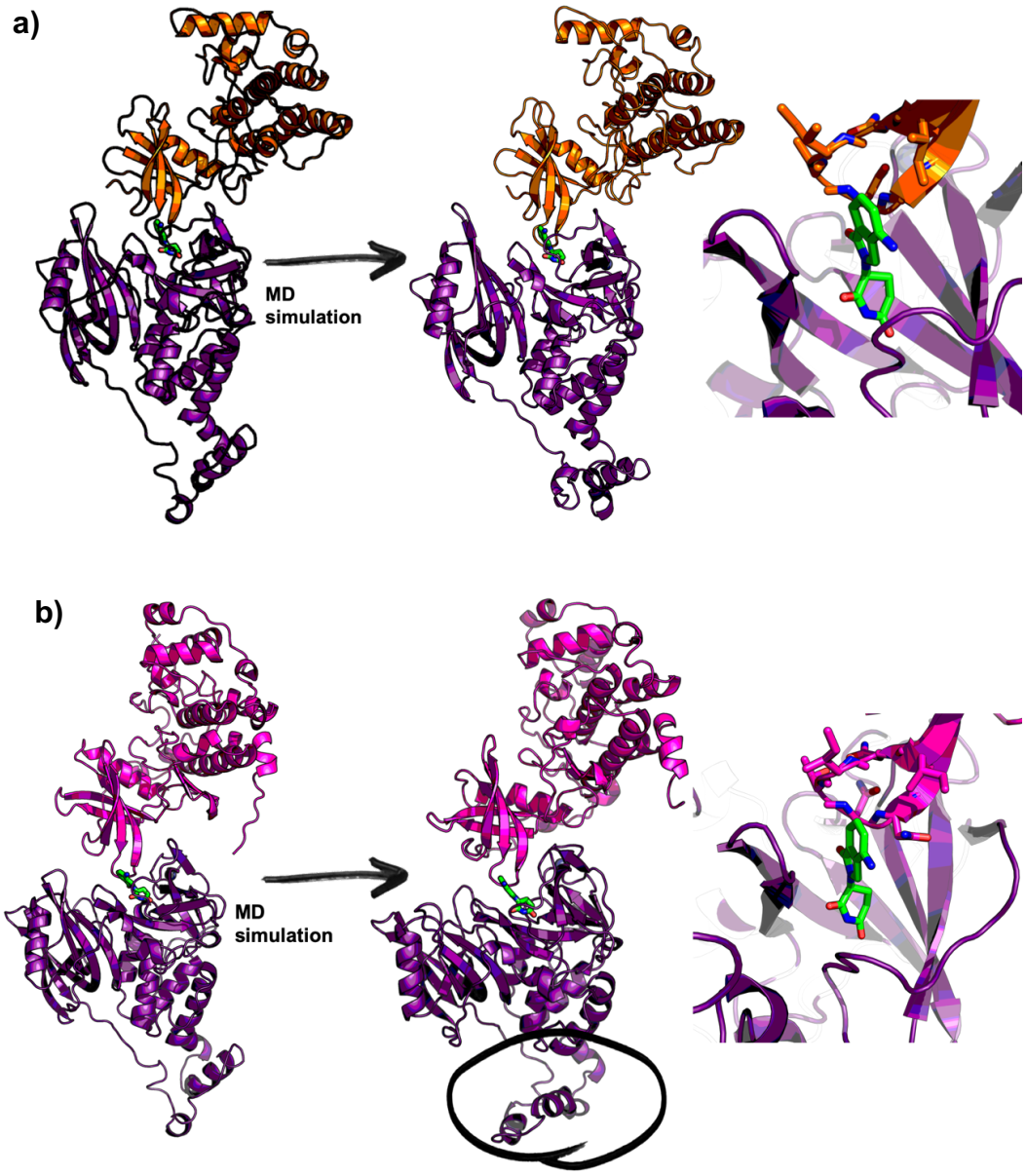


Figure 4.9. a) *CK1α(orange)-lenalidomide(green)-CRBN(purple) and* ***b)*** *CK1α(magenta)-lenalidomide(green)-CRBN(purple) and complexes. Left panel initial system, on the right after 100 ns MD simulation showing the stability of the β-hairpin loop and lenalidomide. The corresponding loop highlighted with the black circle shows the most mobile region in the in the CK2-lenalidomide-CRBN complex.*

4.6.2. Proposal of binding mode for PROTAC Cheng2 inside CK2-CRBN model

As described in Chapter III, **CX4945** is an ATP-competitive inhibitor and has been crystallized in complex with CK2 and deposited under PDB code **3PE1**. Upon superimposition of crystal structure **3PE1** on the CRBN-lenalidomide-CK2 model, it became clear that both ligand binding sites (lenalidomide and **CX4945**) were at least 20Å distant from each other (**Fig. 4.11-a**). Additionally, the carboxylic moiety of **CX4945** that is involved in a key ionic interaction with Lys68 has been used as the anchor point to add the 10 Å PROTAC linker by esterification in **Cheng2 (Fig. 4.10**). Therefore, the first concern was that if **Cheng2** bound to CK2 in the same way as **CX4945,** the linker and the CRBN-recruiting moiety pomalidomide would be projected in the opposite direction from where the lenalidomide binding site was. Docking calculations followed by MD simulations were carried out to shed light on the binding of the POI recruiting moiety in CK2 and analyze if the position of the pomalidomide moiety would be able to recruit the CRBN (**Fig. 4.11**).

Results show that, even though the proposal binding mode of **Cheng2** inside the ATP-binding site, the pomalidomide moiety seems to remain inside the cavity of CK2 with no clear solvent exposure, thus reducing the possibility to attach the corresponding E3 ligase (**Fig. 4.11**). However, these modelling results confirm the capability of this PROTAC to inhibit CK2, which is in agreement with the published biological results [32]. All these results suggest that according to our proposed model, the PROTAC effect that is depletion of CK2, brought about by **Cheng2** should not come from the binding to the ATP-binding site. Upon further inspection of different CK2 crystal structures and superimposition with our model (CK2-CRBN), we found that the MPP allosteric pocket described in **Chapter III** lies within a 10Å distance of the lenalidomide-binding site of the CRBN E3 ligase, while the distance to the ATP-binding site is more than 20Å (**Fig. 4.12**). There is a total of 21 crystal structures that present a small molecule bound to the MPP pocket of CK2 (**Table 4.2**). It is noteworthy that the majority of these structures present a chlorophenyl moiety similar to that present in **CX4945**. However, of the two CK2 structures complexed with **CX4945** (**3PE1** and

6HMB) none of them presented binding to the MPP pocket. For this reason, we have docked **CX4945** inside the MPP pocket, and the results show a similar binding mode of the chlorophenyl moiety compared to the ligand **42J** (**Fig. 4.13**). The obtained complex was superimposed to the CK2-lenalidomide-CRBN model where the lenalidomide was modeled into pomalidomide. Then, the linker was manually built from the carboxylic moiety of **CX4845** to the amine of pomalidomide (**Fig. 4.14-a**). The resulting complex was subjected to 100ns MD simulation in order to check its stability by means of the RMSD analysis and total binding energies calculations. These data are summarized in **Table 4.3.** and **Figure 4.15**. Results show that despite the initial mobility of **Cheng2**, by the end of the simulation the PROTAC rearranges and was able to find a stable binding mode highlighted by the total binding energies between **Cheng2** and the proteins (**Fig. 4.14-b**). Additionally, the mean RMSD values as well as the binding energy values of CK2 and the β-hairpin loop indicate that they remain stable in a similar way as in the previous simulation of the CK2-lenalidomide-CRBN model.

Giving the biological activity reported by *Cheng et al.* it is undeniable that **Cheng2** presents a remarkable CK2 inhibitory activity, for this reason the favorable binding mode of **Cheng2** is inside the ATP binding site. Nevertheless, this binding mode does not further support the depletion of CK2. Only the non-bonded molecules of **Cheng2** could interact with the proposed binding site after occupying the CK2 active site. This fact could be supported by the remarkable increase of the activity after a two fold increase the concentration of the PROTAC [32]**.**

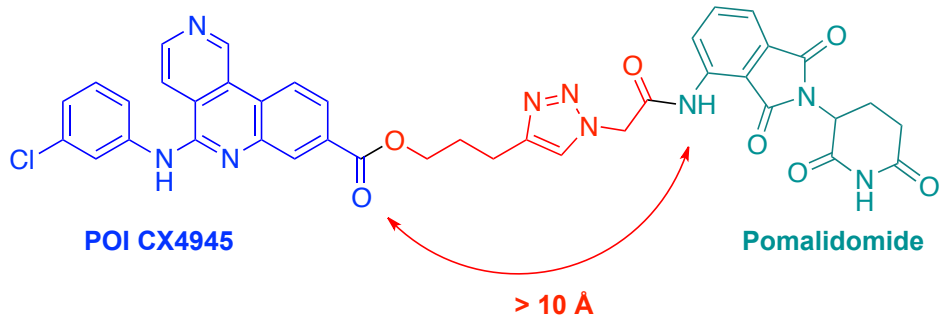


Figure 4.10. *Chemical structure of the most active PROTAC (**Cheng2**) reported by Cheng et al.*[32]

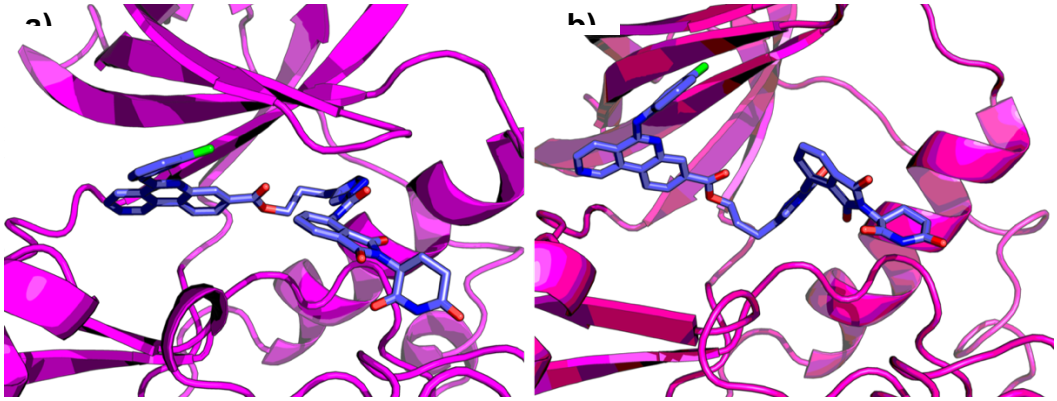


Figure 4.11. a) PROTAC Compound 2 *docked inside CK2 active site.* ***b)*** *Most populated conformation from 100ns MD simulation.*

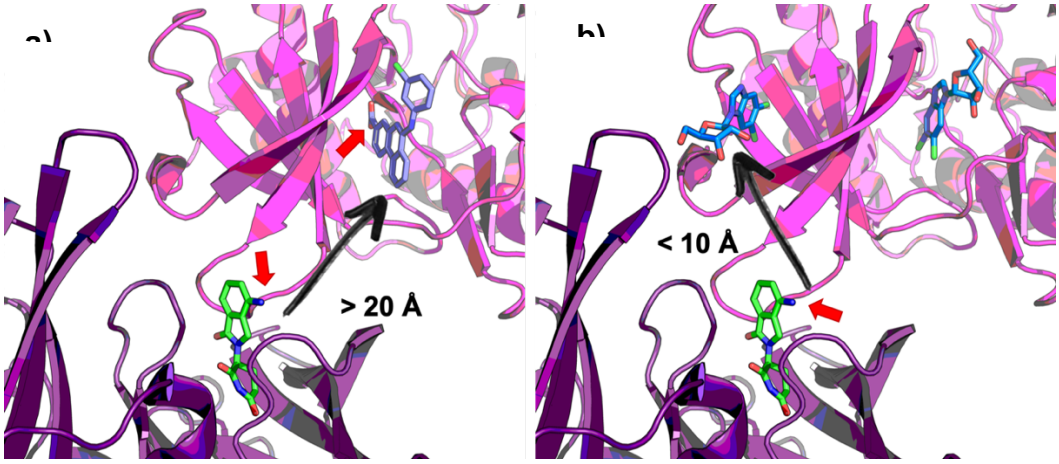
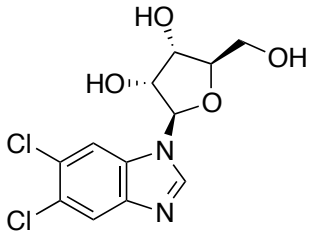
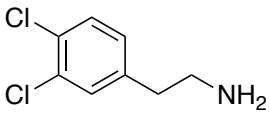
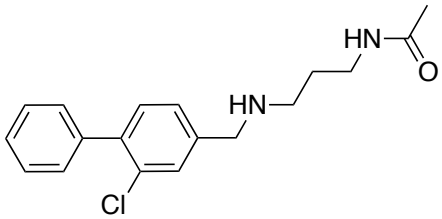
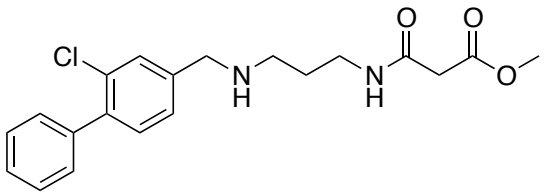
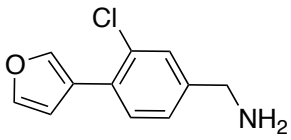
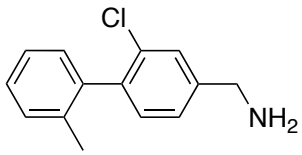
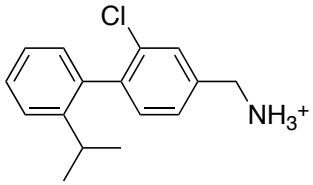
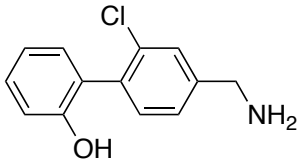
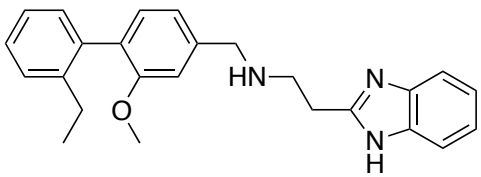
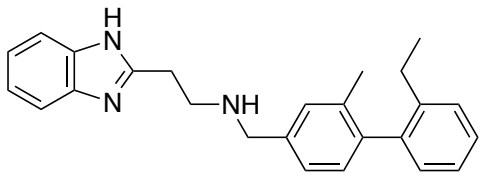
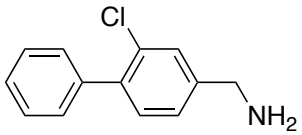
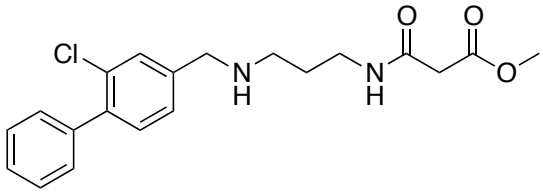
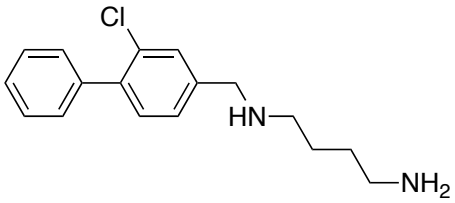
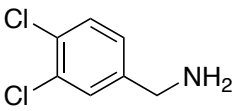
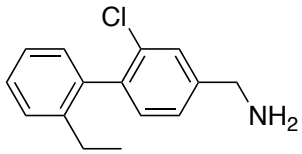


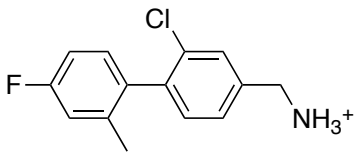
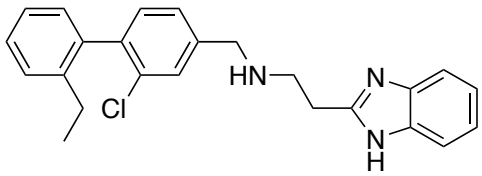
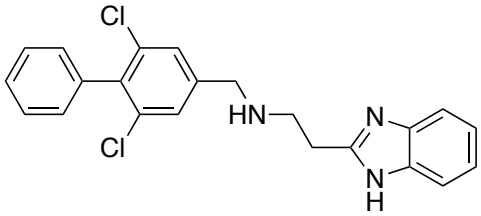
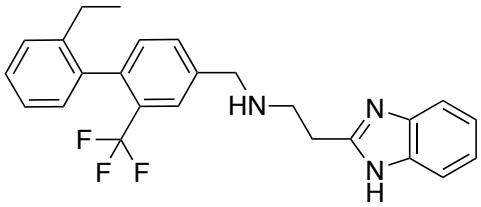
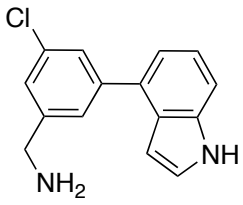
Figure 4.12. a) *Alignment of 3PE1(magenta-CK2) and **5FQD** (purple-CRBN) showing the distance between both ligands lenalidomide (green) and **CX4945** (blue).* ***b)*** *Alignment of 3H30 (magenta-CK2) and **5FQD** (purple-CRBN) showing the distance between both ligands lenalidomide (green) and **RFZ** (blue). Red arrows mark the anchor points.*

Table 4.12. *Analysis of the CK2 structures which present small molecules (cosolvent molecules were not taken into account) crystallized inside the MPP pocket. The ligand can also bind other sites such as the ATP-binding site or the allosteric α-D pocket.*

PDB ID	LIGAND	ATP Binding pocket	MPP Pocket	α-D Pocket
3H30	RFZ 	✓	✓	X
5CLP	42J 	✓	✓	✓
5CU0	54R 	X	✓	✓
5CX9	551	X	✓	✓

				
5MOE	**OQC** 	**X**	✓	✓
5ORH	**A4N** 	**X**	✓	✓
5OS7	**A8Q** 	**X**	✓	✓
5OSL	**A9K** 	**X**	✓	✓
5OTQ	**AUH** 	✓	✓	✓

5OYF	**B4Q** 	✓	✓	✓
5CSH	**54E** 	**X**	✓	✓
5CU2	**551** 	**X**	✓	✓
5MMR	**H83** 	**X**	✓	✓
5NQC	**1KP** 	✓	✓	**X**
5ORJ	**A4Q** 	**X**	✓	✓

5OS8	**J27** 	**X**	✓	✓
5OSZ	**AHK** 	**X**	✓	✓
5OTS	**AUD** 	**X**	✓	✓
6EHU	**B5E** 	**X**	✓	✓
6GIH	**EZN** 	**X**	✓	**X**

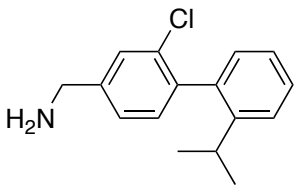
6GMD	**A8Q** 	**X**	✓	✓

Table 4.3. *Mean RMSD values (Å) from the 100 ns MD simulations from CK2-Cheng2-CRBN complex. The total binding energies values (Kcal·mol^{-1}) come from the ppi between the CK2 and CRBN and between* ***Cheng2*** *with the entire complex.*

System	RMSD values					Total Binding Energy	
	Complex	CRBN	Kinase	Cheng2	Loop	ppi	Cheng2
CK2	5.50 ± 1.88	3.76 ± 0.67	2.19 ± 0.37	3.43 ± 0.76	0.72 ± 0.16	-46.83	-83.24

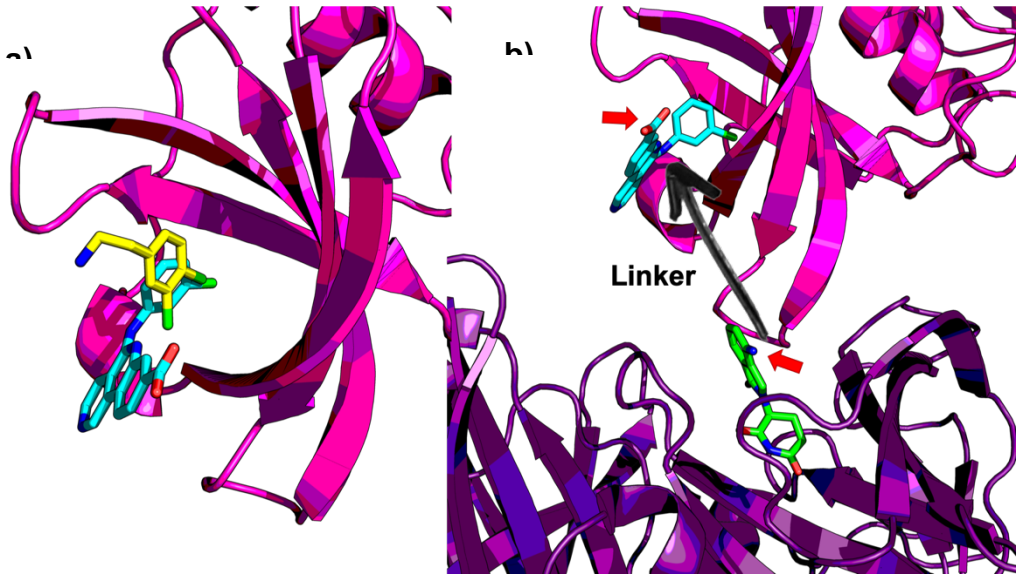


Figure 4.13. a) CX4945 *(blue) docked inside CK2 MPP pocket, superimposed with* ***42J*** *ligand present in* ***5CLP****.* ***b) CX4945*** *(blue) docked inside the model CK2(magenta)-lenalidomide(green)-CRBN (purple). The arrows mark the putative linker to be manually designed. Red arrows mark the anchor points.*

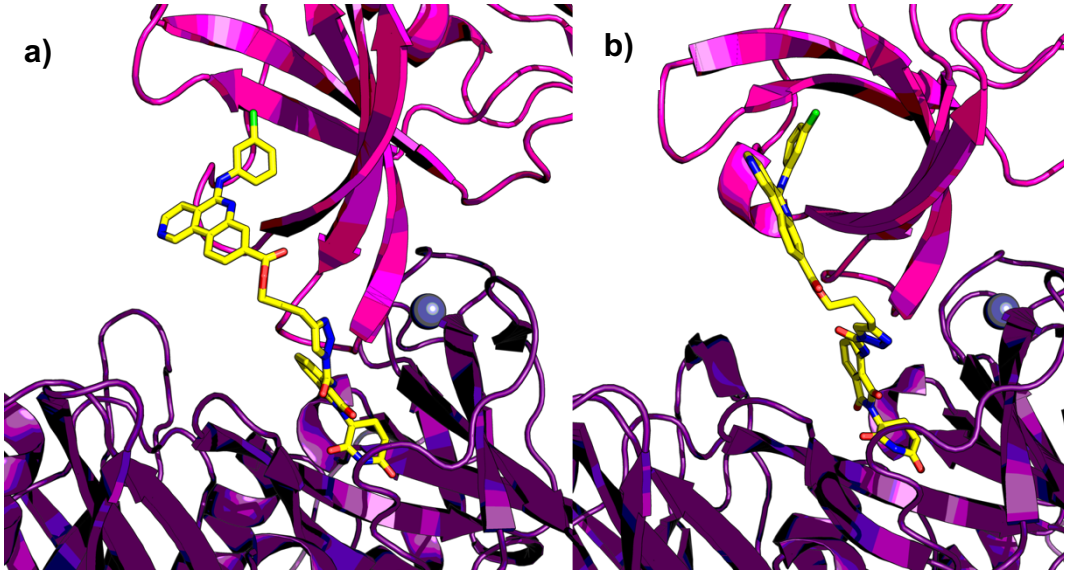


Figure 4.14. a) *Energy minimized structure of the complex between* ***Cheng2*** *docked inside the modelled CK2(magenta)-CRBN (purple) complex.* ***b)*** *The most populated conformed from 100ns MD simulation of the complex.*

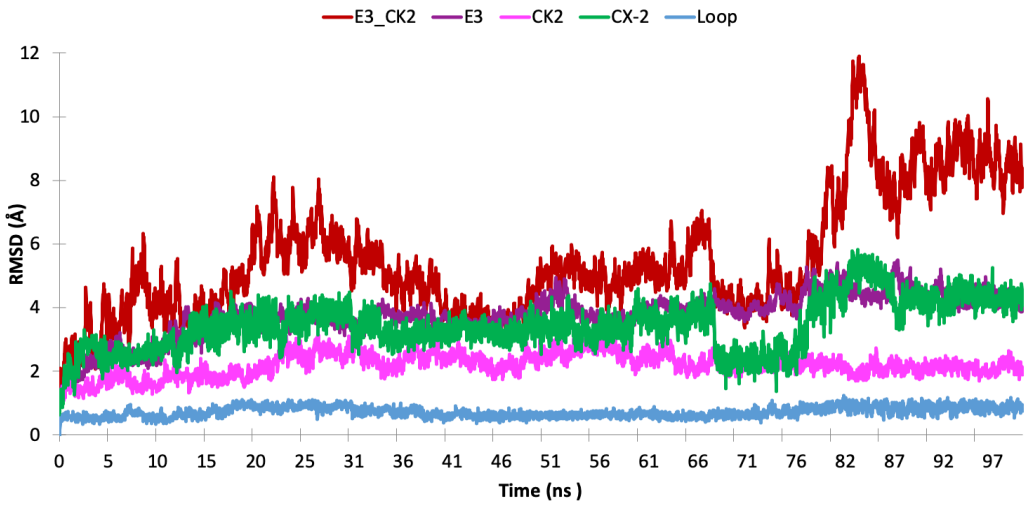


Figure 4.15. *Monitoring of the RMSD values obtained along the 100 ns MD simulation in CK2-Cheng2-CRBN complex.*

4.6.3. Design of a MPP site based CK2 PROTAC

Based on the proposal for the allosteric binding mode of **Cheng2**, we revisited the 21 CK2 crystal structures with small molecules bound to the

MPP pocket collected in (**Table 4.2**). The **6GIH** crystal structure is the only one that presents a ligand (**EZN**) that does not bind to the ATP binding neither to the α-D pocket, which suggests a high level of selectivity and affinity for the MPP pocket. Therefore, the first choice as POI binder in the design of an allosteric PROTAC was compound **EZN**. As for the E3 ligase-recruiting moiety, pomalidomide was used as the crystal structure used as starting point was bound to this IMiD; however, other molecules such as lenalidomide could be used as well. The proposed linker length must accomplish the distance between both molecules. Specifically, the distance between the attachment points on our CK2-CRBN model. For the attachment points between moieties, amide groups were selected, and the linker was designed using chains of poly-ethylene glycol (PEG) until satisfaction of the required distance (**Fig. 4.16**).

Crystal structure **6GIH** was superimposed to the previously obtained CK2-lenalidomide-CRBN model in order to obtain the binding mode of compound **EZN** to manually build the PEG linker until obtaining the desired compound (**EZN-P**) inside the CK2-CRBN complex (**Fig. 4.16.**). In order to analyze the stability and binding energy of the complex, a 100 ns MD simulation was carried out. The results show an increased stability of the entire complex along the simulation time compared to the CK2-**Cheng2**-CRBN complex (**Fig. 4.18**). CK2 and CRBN are more stable presenting lower RMSD values; additionally, the β-hairpin loop also presents lower RMSD values (**Table 4.4.**). Compound **EZN-P** preserved all the interactions between the EZN moiety and the POI, and those between pomalidomide and CRBN (**Fig. 4.17**). The only observable change comes from the linker, which rearranges during the simulations increasing the interactions with the proteins. Computationally, the stability of the complex with **EZN-P** seems to be higher than the complex with **Cheng2** for both the ligand-receptor complex and the ppi.

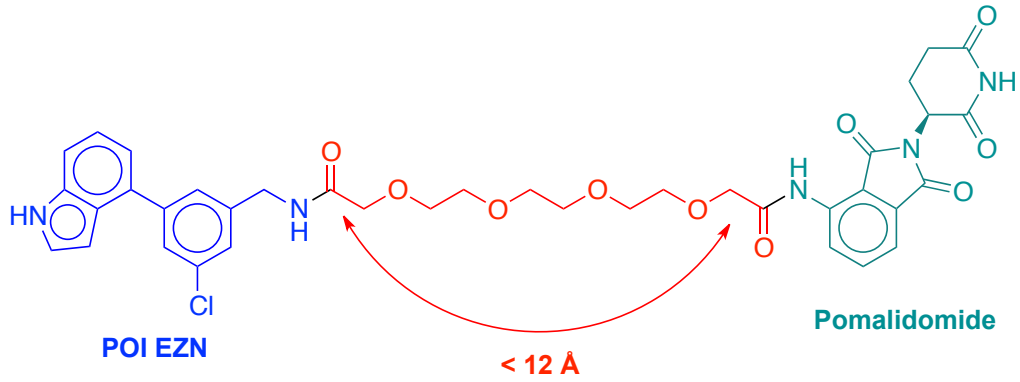


Figure 4.16. *Chemical structure of the designed allosteric PROTAC* ***EZN-P****.*

Table 4.4. *Mean RMSD values (Å) from the 100 ns MD simulations from CK2-EZN-P-CRBN complex. The total binding energies values (Kcal·mol^{-1}) come from the ppi between the CK2 and CRBN and between* ***EZN-P*** *with the entire complex.*

System	RMSD values					Total Binding Energy	
	Complex	CRBN	Kinase	EZN-P	Loop	ppi	EZN-P
CK2	3.70 ± 0.56	3.34 ± 0.48	1.86 ± 0.40	2.03 ± 0.29	0.27 ± 0.07	-69.48	-93.85

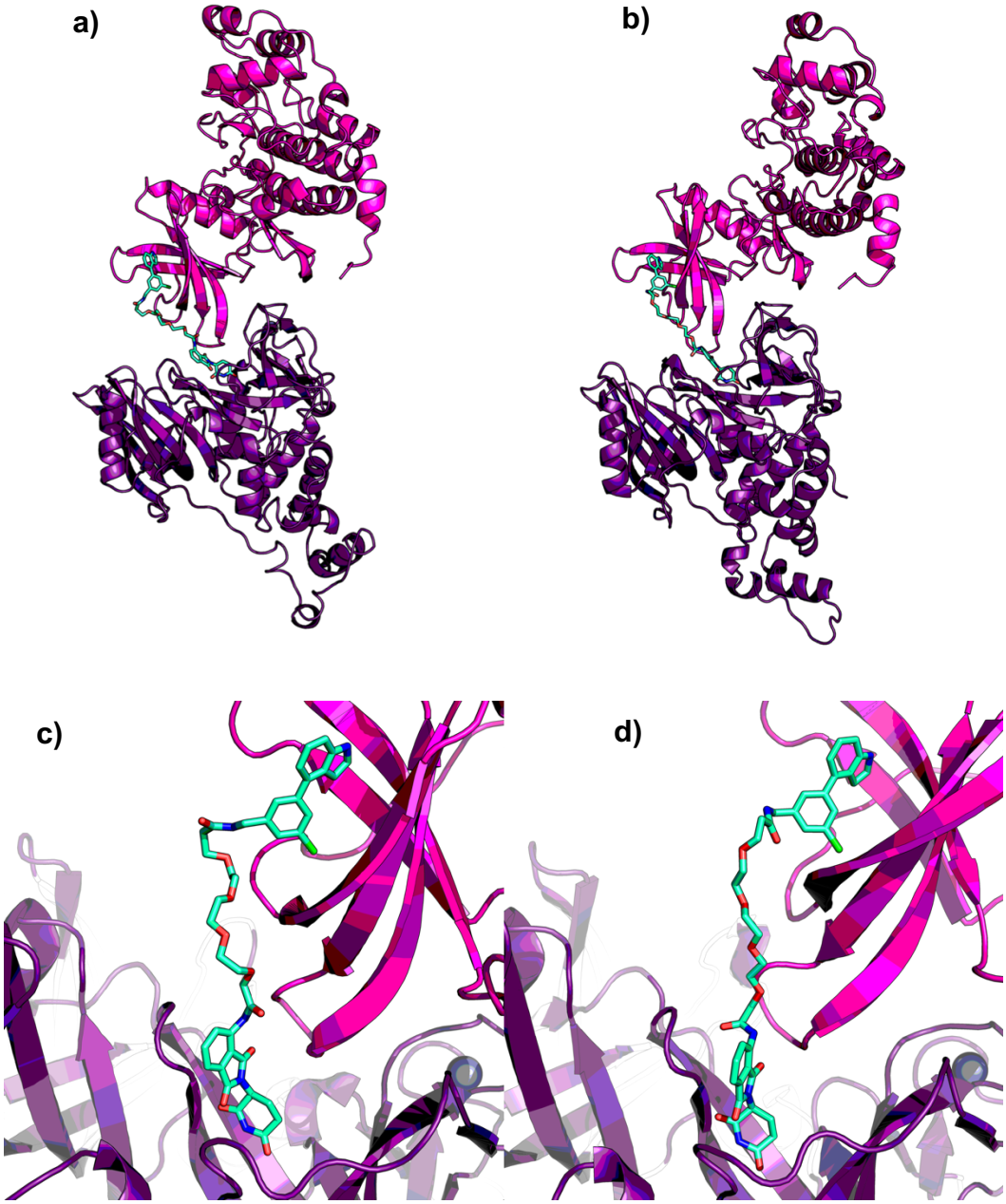


Figure 4.17. a) *CK2- **EZN-P**-CRBN complex built.* ***b)*** *The most populated conformation of CK2- **EZN-P**-CRBN complex after 100 ns MD simulation.* ***c-d)*** *Zoom to **EZN-P** before and after the MD simulation.*

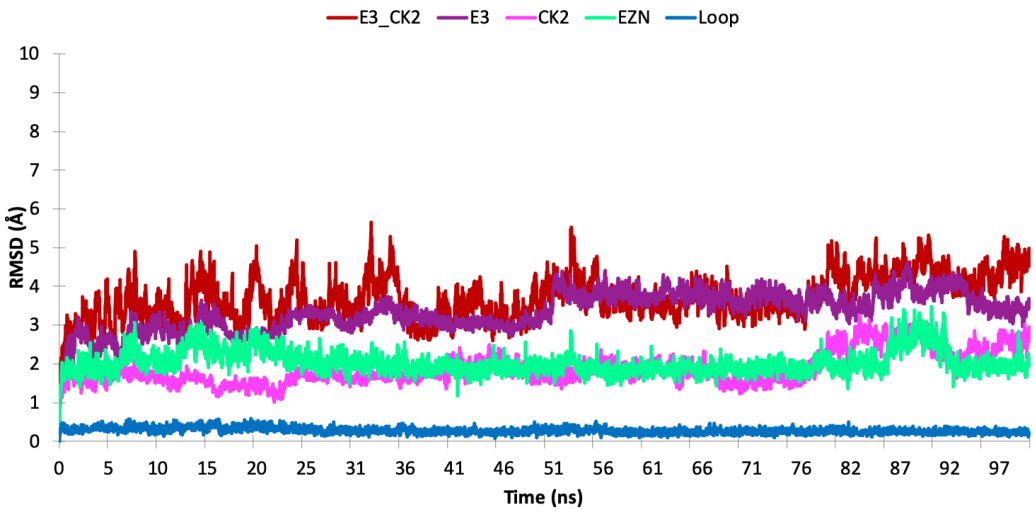


Figure 4.18. *Monitoring of the RMSD values obtained along the 100 ns MD simulation in CK2-**EZN-P**-CRBN complex.*

4.7. Conclusions

PROTACs as heterobifunctional molecules are perhaps the most significant type of protein degraders offering a rational chemical tool for protein degradation. The design of PROTACs involve mainly the selection of the POI-recruiting moiety, the E3 recruiter, the linker design, and suitable ADME properties.

We have proposed a plausible CK2-CRBN complex model supported by MD simulation in which the ppi between the CRBN E3 ligase and CK2 could be the same as in the CK1-CRBN crystal structure, even though in CK2 the β-hairpin loop does not present the selective glycine.

Additionally, we have proposed two plausible binding modes for the most active compound reported by *Cheng et al*. (**Cheng2**). The binding mode of **Cheng2** in the ATP-binding site could account for the CK2 inhibitory activity of this PROTAC, this model puts forward the impossibility of the pomalidomide to reach its binding site in the CRBN. Additionally, we have proposed an alternative binding mode in which the POI-recruiting moiety binds to the MPP pocket and preserves similar ppi as reported for CRBN and CK1. Additionally, the proposed allosteric binding mode, corroborates

the low PROTAC activity reported for **Cheng2** at low concentrations (5μM) while increase at high concentrations (10μM).

Based on the previous results, we have designed a PROTAC based on a MPP-binding site (**EZN-P**) in an attempt to increase the CK2 depletion. To the best of our knowledge this could be the first time, that an allosteric PROTAC is computationally designed. Nevertheless, some issues need to be taken into account, such as the length of the linker, its composition, or, the anchor points. Moreover, it is noteworthy that pomalidomide can be replaced by lenalidomide, which seems to be more effective in the recruitment of CRBN.

4.8. References

1. Scudellari, M. Protein-slaying drugs could be the next blockbuster therapies. *Nature* **2019**, *567*, 298–300.

2. Scheffner, M.; Nuber, U.; Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **1995**, *373*, 81–83.

3. Yamao, F. Ubiquitin System: Selectivity and Timing of Protein Destruction. *J. Biochem.* **1999**, *125*, 223–229.

4. Teixeira, L. K.; Reed, S. I. Ubiquitin Ligases and Cell Cycle Control. *Annu. Rev. Biochem.* **2013**, *82*, 387–414.

5. Koepp, D. M. Cell Cycle Regulation by Protein Degradation. *Methods Mol. Biol.* **2014**, *1170*, 61–73.

6. Scheepstra, M.; Hekking, K. F. W.; van Hijfte, L.; Folmer, R. H. A. Bivalent Ligands for Protein Degradation in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 160–176.

7. Shimokawa, K.; Shibata, N.; Sameshima, T.; Miyamoto, N.; Ujikawa, O.; Nara, H.; Ohoka, N.; Hattori, T.; Cho, N.; Naito, M. Targeting the Allosteric Site of Oncoprotein BCR-ABL as an Alternative Strategy for Effective Target

Protein Degradation. *ACS Med. Chem. Lett.* **2017**, *8*, 1042–1047.

8. Nakayama, K. I.; Nakayama, K. Ubiquitin ligases: cell-cycle control and cancer. *Nat. Rev. Cancer* **2006**, *6*, 369–381.

9. Schneekloth, A. R.; Pucheault, M.; Tae, H. S.; Crews, C. M. Targeted intracellular protein degradation induced by a small molecule: En route to chemical proteomics. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5904–5908.

10. Okuhira, K.; Ohoka, N.; Sai, K.; Nishimaki-Mogami, T.; Itoh, Y.; Ishikawa, M.; Hashimoto, Y.; Naito, M. Specific degradation of CRABP-II via cIAP1-mediated ubiquitylation induced by hybrid molecules that crosslink cIAP1 and the target protein. *FEBS Lett.* **2011**, *585*, 1147–1152.

11. Cyrus, K.; Wehenkel, M.; Choi, E.-Y.; Han, H.-J.; Lee, H.; Swanson, H.; Kim, K.-B. Impact of linker length on the activity of PROTACs. *Mol. BioSyst.* **2011**, *7*, 359–364.

12. Maple, H. J.; Clayden, N.; Baron, A.; Stacey, C.; Felix, R. Developing degraders: principles and perspectives on design and chemical space. *Medchemcomm* **2019**.

13. Mullard, A. First targeted protein degrader hits the clinic. *Nat. Rev. Drug Discov.* **2019**.

14. Neklesa, T.; Snyder, L. B.; Willard, R. R.; Vitale, N.; Raina, K.; Pizzano, J.; Gordon, D.; Bookbinder, M.; Macaluso, J.; Dong, H.; et al. Abstract 5236: ARV-110: An androgen receptor PROTAC degrader for prostate cancer. *Clinical Research (Excluding Clinical Trials)*; American Association for Cancer Research, 2018; p. 5236–5236.

15. Jones, L. H. Small-Molecule Kinase Downregulators. *Cell Chem. Biol.* **2018**, *25*, 30–35.

16. Huang, H.-T.; Dobrovolsky, D.; Paulk, J.; Yang, G.; Weisberg, E. L.; Doctor, Z. M.; Buckley, D. L.; Cho, J.-H.; Ko, E.; Jang, J.; et al. A Chemoproteomic Approach to Query the Degradable Kinome Using a Multi-kinase Degrader. *Cell Chem. Biol.* **2018**, *25*, 88–99.

17. Lai, A. C.; Toure, M.; Hellerschmied, D.; Salami, J.; Jaime-Figueroa, S.; Ko, E.; Hines, J.; Crews, C. M. Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL. *Angew. Chemie Int. Ed.* **2016**, *55*, 807–810.

18. Lebraud, H.; Wright, D. J.; Johnson, C. N.; Heightman, T. D. Protein Degradation by In-Cell Self-Assembly of Proteolysis Targeting Chimeras. *ACS Cent. Sci.* **2016**, *2*, 927–934.

19. Qin, C.; Hu, Y.; Zhou, B.; Fernandez-Salas, E.; Yang, C.-Y.; Liu, L.; McEachern, D.; Przybranowski, S.; Wang, M.; Stuckey, J.; et al. Discovery of QCA570 as an Exceptionally Potent and Efficacious Proteolysis Targeting Chimera (PROTAC) Degrader of the Bromodomain and Extra-Terminal (BET) Proteins Capable of Inducing Complete and Durable Tumor Regression. *J. Med. Chem.* **2018**, *61*, 6685–6704.

20. Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steurer, S.; Karolyi-Oezguer, J.; Riedmueller, C.; et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **2019**, *15*, 672–680.

21. Olson, C. M.; Jiang, B.; Erb, M. A.; Liang, Y.; Doctor, Z. M.; Zhang, Z.; Zhang, T.; Kwiatkowski, N.; Boukhali, M.; Green, J. L.; et al. Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. *Nat. Chem. Biol.* **2018**, *14*, 163–170.

22. Hatcher, J. M.; Wang, E. S.; Johannessen, L.; Kwiatkowski, N.; Sim, T.; Gray, N. S. Development of Highly Potent and Selective Steroidal Inhibitors and Degraders of CDK8. *ACS Med. Chem. Lett.* **2018**, *9*, 540–545.

23. Li, W.; Gao, C.; Zhao, L.; Yuan, Z.; Chen, Y.; Jiang, Y. Phthalimide conjugations for the degradation of oncogenic PI3K. *Eur. J. Med. Chem.* **2018**, *151*, 237–247.

24. Yang, K.; Song, Y.; Xie, H.; Wu, H.; Wu, Y.-T.; Leisten, E. D.; Tang, W. Development of the first small molecule histone deacetylase 6 (HDAC6) degraders. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2493–2497.

25. Zhang, C.; Han, X.-R.; Yang, X.; Jiang, B.; Liu, J.; Xiong, Y.; Jin, J. Proteolysis Targeting Chimeras (PROTACs) of Anaplastic Lymphoma Kinase (ALK). *Eur. J. Med. Chem.* **2018**, *151*, 304–314.

26. Powell, C. E.; Gao, Y.; Tan, L.; Donovan, K. A.; Nowak, R. P.; Loehr, A.; Bahcall, M.; Fischer, E. S.; Jänne, P. A.; George, R. E.; et al. Chemically Induced Degradation of Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* **2018**, *61*, 4249–4255.

27. Petzold, G.; Fischer, E. S.; Thomä, N. H. Structural basis of lenalidomide-induced CK1α degradation by the CRL4 CRBN ubiquitin ligase. *Nature* **2016**, *532*, 127–130.

28. Kronke, J.; Udeshi, N. D.; Narla, A.; Grauman, P.; Hurst, S. N.; McConkey, M.; Svinkina, T.; Heckl, D.; Comer, E.; Li, X.; et al. Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. *Science (80-.).* **2014**, *343*, 301–305.

29. Krönke, J.; Fink, E. C.; Hollenbach, P. W.; MacBeth, K. J.; Hurst, S. N.; Udeshi, N. D.; Chamberlain, P. P.; Mani, D. R.; Man, H. W.; Gandhi, A. K.; et al. Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. *Nature* **2015**, *523*, 183–188.

30. Vittal, V.; Stewart, M. D.; Brzovic, P. S.; Klevit, R. E. Regulating the Regulators: Recent Revelations in the Control of E3 Ubiquitin Ligases. *J. Biol. Chem.* **2015**, *290*, 21244–21251.

31. Bhogaraju, S.; Dikic, I. A peek into the atomic details of thalidomide's clinical effects. *Nat. Struct. Mol. Biol.* **2014**, *21*, 739–740.

32. Chen, H.; Chen, F.; Liu, N.; Wang, X.; Gou, S. Chemically induced degradation of CK2 by proteolysis targeting chimeras based on a ubiquitin–proteasome pathway. *Bioorg. Chem.* **2018**, *81*, 536–544.

33. Zapico, J. M.; Puckowska, A.; Filipiak, K.; Coderch, C.; de Pascual-Teresa, B.; Ramos, A. Design and synthesis of potent hydroxamate inhibitors with increased selectivity within the gelatinase family. *Org. Biomol. Chem.* **2015**, *13*, 142–156.

34. Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2019.

35. Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, Á.; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: An Ultrafast and Accurate Scoring Function for Protein–Protein Docking. *J. Chem. Theory Comput.* **2012**, *8*, 3395–3408.

36. Buhimschi, A. D.; Crews, C. M. Evolving Rules for Protein Degradation? Insights from the Zinc Finger Degrome. *Biochemistry* **2019**, *58*, 861–864.

CHAPTER V

5. In Silico Study of Chemical Reactions

Quantum Mechanics and Density Functional Theory can be used to study the Energy Profile (EP) of a chemical reaction. In this section, two chemical reactions will be studied by means of these techniques. Therefore, we shall first address the general scheme of an EP and its important hallmarks. For this purpose, the general scheme of a classical S_N2 reaction shown in **Figure 4.1** will be used as an example [1–3].

The first point (**R**) corresponds to the reactants, and for each of them the single point energies are calculated and used as baseline for the reaction. When the reactants are put together and minimized thus forming the Reactive Complex (**RC**), the resulting energy ($\mathbf{\Delta E^{cmpx}}$) corresponds to the energy necessary to combine the reactants. In this example, an S_N2 reaction where one bond is broken and another is formed, the highest potential energy along this reaction coordinate corresponds to the Transition State (**TS‡**, which is marked with the double dagger symbol, ‡). In the **TS‡**, one bond is cleaved and the other is formed at the same time, so this point cannot be chemically isolated. However, the **TS‡** can be determined by searching for first-order saddle points on the Potential Energy Surface (**PES**) by Quantum Mechanical such as DFT [4]. The energy necessary to reach the **TS‡** starting from the **RC** is the $\mathbf{\Delta E^{\ddagger,centr}}$; but the overall energy barrier of the reaction is $\mathbf{\Delta E^{\ddagger,ovr}}$ that relates **R** and the **TS‡**. Once the reactants have passed through the transition state, they continue to form the Product Complex (**PC**) and then the separate products (**P**). The $\mathbf{\Delta E^{react}}$ energy corresponds to the difference between **R** and **P**, pointing to an exothermic reaction if negative, or and endothermic reaction if positive.

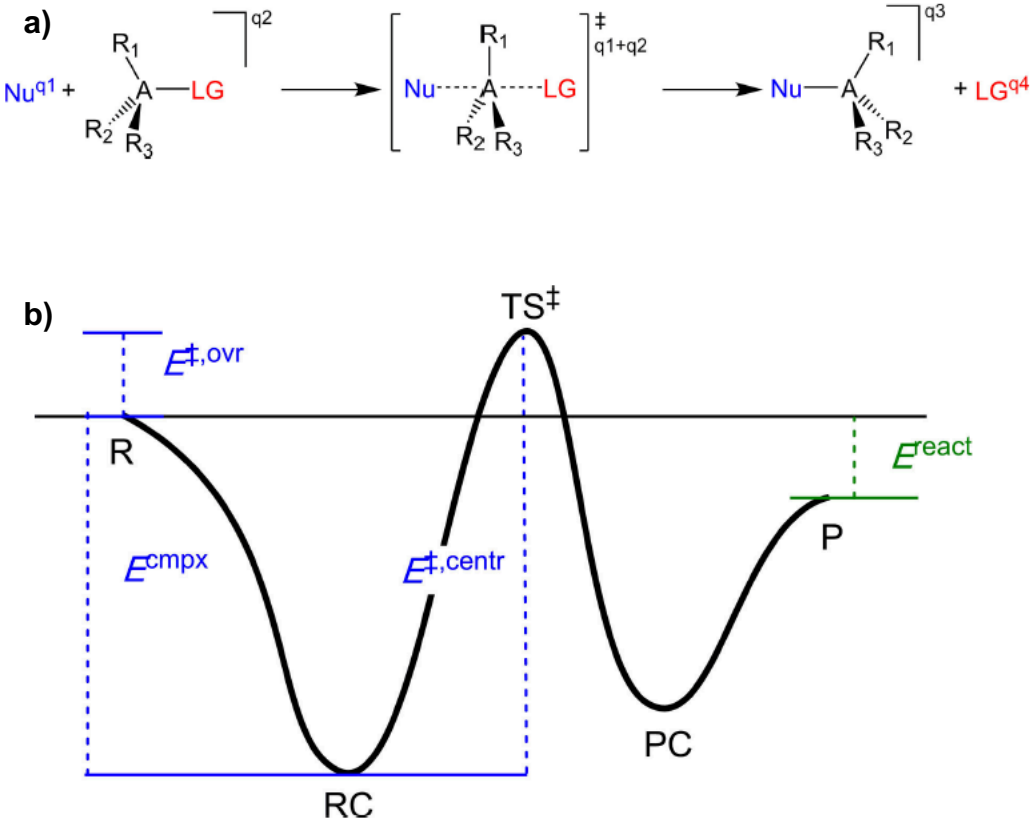


Figure 5.1. a) *General SN_2 reaction scheme, where Nu is the nucleophile, A the central atom, and LG the leaving group (q1 to q4 is the charge of each species: q1+q2=q3+q4).* ***b)*** *Energy profile of a typical SN_2 reaction: the reactants (**R**), reactant complex (**RC**), transition state (**TS≠**), product complex (**PC**), and products (**P**). The central barrier $\Delta E^{\ddagger,centr}$ relates the **TS** to the **RC**, and an overall barrier $\Delta E^{\ddagger,ovr}$ relates the **TS** to **R**. ΔE^{react} relates **R** with the separated products, and ΔE^{cmpx} relates **R** with **RC**. Figure taken from Chem. Phys. Chem.* ***2018****,19, 1315–1330.*

5.1. Sulfonium ylide cyclopropanation

During this thesis, a four-month placement was carried out under the supervision of Dr. Lubomír Rulíšek in the Institute of Organic Chemistry and Biochemistry (IOCB) in Prague, Czeck Republic (www.uochb.cz). The work carried out during this period was in the field of thermochemistry and QM studies within the DFT field. In detail, this placement focused on the study of the reaction mechanism of the stereoselective cyclopropanation

synthesis by DFT methods with the aim to understand the experimental results obtained by the research group of Prof. Javier Pérez-Castells [5]. The stereoselective sulfonium ylide cyclopropanation results gave us the opportunity to use in silico tools with the aim to learn the methodology. The chemical reaction is proposed to carry on as an addition-elimination reaction (**Fig. 5.2**):

i) The initial attack of the sulfur ylide to the electrophilic center of the Michael acceptor forming a betaine (intermediate).

ii) An intramolecular nucleophilic displacement giving the final cyclopropane (product).

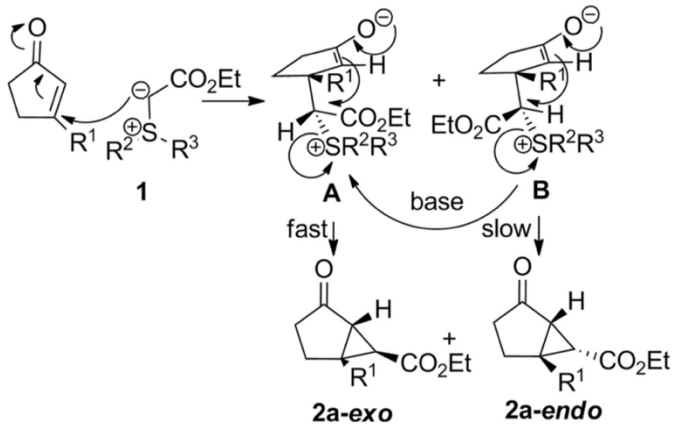


Figure 5.2. *Proposed mechanism of reaction for the ylide cyclopropanation* [5].

In the first step of the reaction, the chiral sulfur ylide is the nucleophile that attacks the gamma position of the Michael acceptor producing the betain (A or B) reaction intermediate. This process, in which the first C-C bond is formed, is considered the determinant step of the reaction. The second step is the intramolecular cycloaddition that occurs after nucleophilic displacement.

This reaction yields either the exo or endo configuration adducts depending if the reaction proceeds through betain A or B, respectively. However, both intermediates can be exchanged by the presence of a base such as DBU (1,8-Diazabicyclo[5.4.0]undec-7-ene). Overall, the reaction is

stereoselective with a preference for the exo cycloaddition yielding a higher ratio for the exo product. The entire EP for both pathways, exo and endo, in vacuum and in a chloroform implicit solvent model were obtained (**Fig. 5.5**) following the scheme summarized in **Figure 5.3** and **Figure 5.4.** The results of this work summarized in **Table 5.1.** are in agreement with experimental data and other similar chemical reactions studied by DFT methods [6–8]. In summary this work has been useful as a tool to understand the application of the DFT methods and the methodology applied with the software Gaussian09 [9].

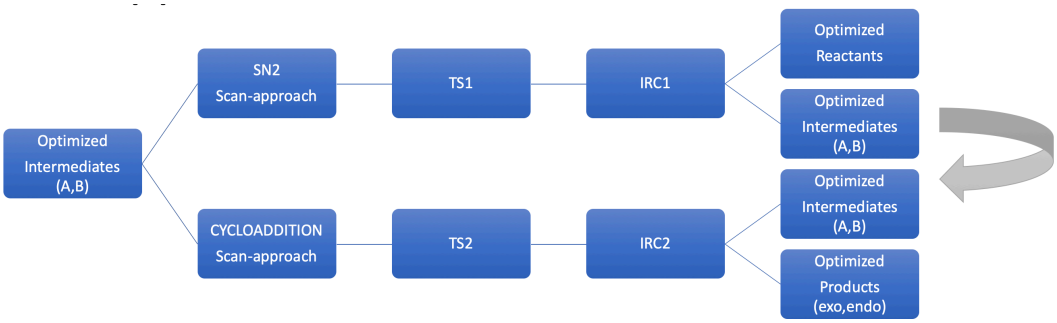


Figure 5.3. *The approach used to obtain the corresponding energy profile for the entire ylide cyclopropanation. The optimized intermediates must present the same structure and energy.*

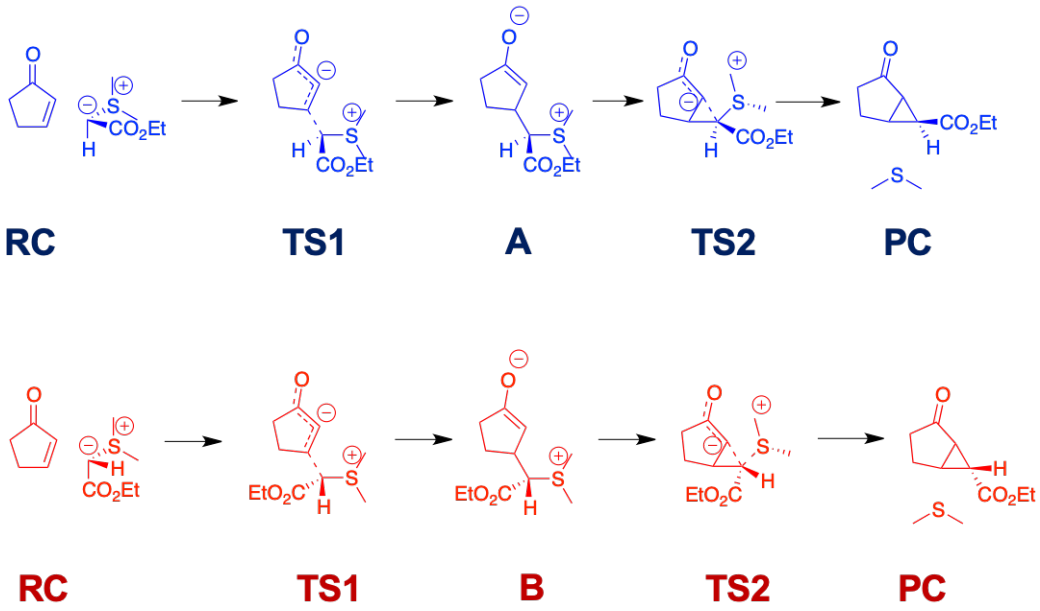


Figure 5.4. *Chemical structure of **RC**, **TS1**, intermediates **A** and **B**, **TS2** and **PC** for both pathways exo (A) and endo (B) (blue and red, respectively).*

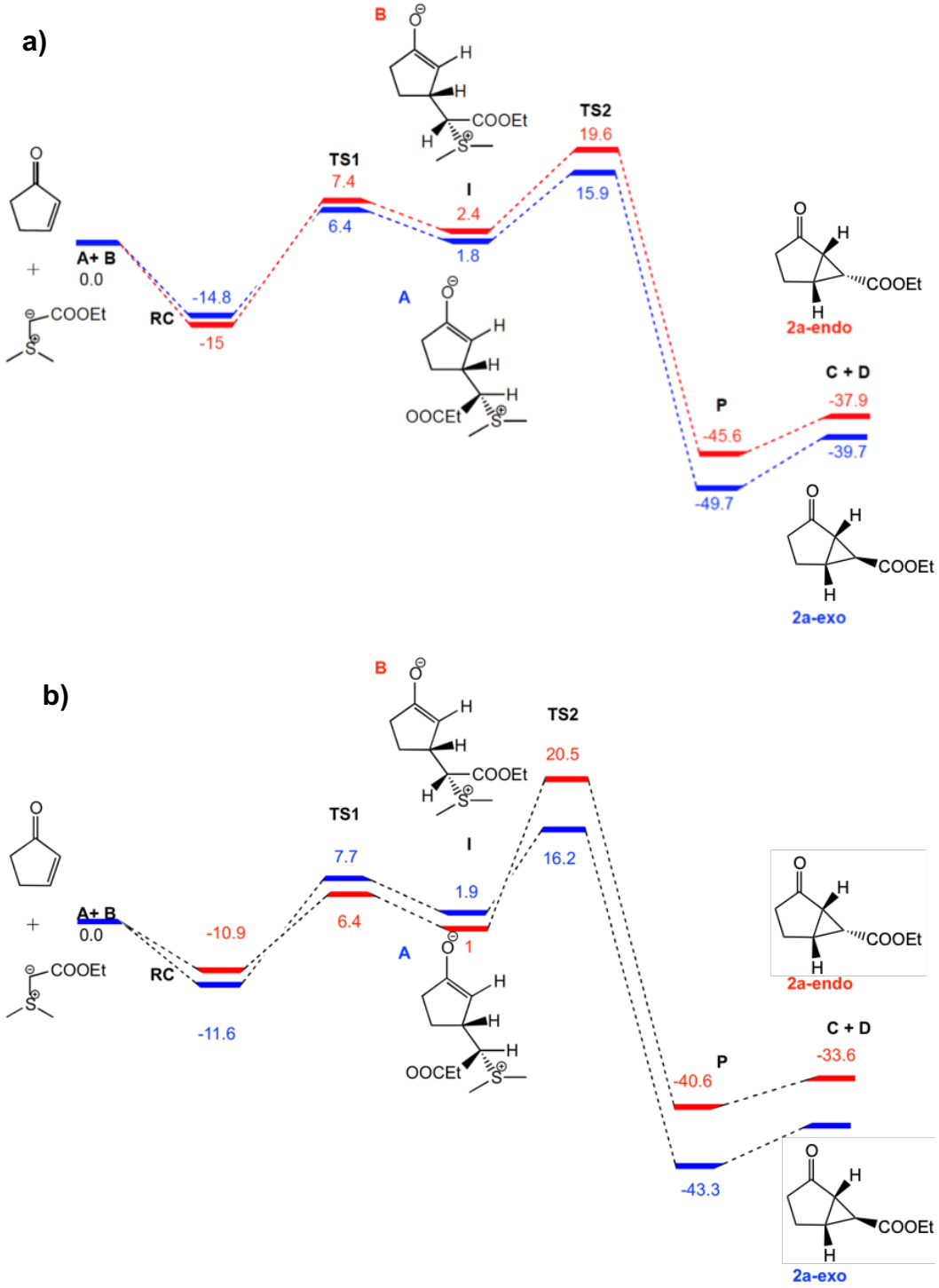


Figure 5.5. *Total Energy Profile of the ylide cyclopropanation using B3LYP-D3/6-31+G(d,p)* ***a)*** *At vacuum.* ***b)*** *With chloroform as implicit solvent. Calculated energies are Kcal·mol^{-1}.*

Table 5.1. *Relevant calculated barriers for both pathway exo (A) and endo (B). Calculated energies are Kcal·mol^{-1}.*

Pathway	ΔE^{cmpx}	ΔE^{react}	$\Delta E^{\ddagger,centr}$ (TS1)	$\Delta E^{\ddagger,centr}$ (TS2)	$\Delta E^{\ddagger,ovr}$
A-Vacuum	-14.8	-39.7	21.2	14.1	15.9
B-Vacuum	-15	-37.9	22.4	17.2	19.6
A-Chloroform	-11.6	-35.8	19.3	14.3	16.2
B-Chloroform	-10.9	-33.6	17.3	19.5	20.5

5.1.1. Methodology

The Gaussian 09 and Gview software were used throughout the project [9]. The calculations were carried out using the B3LYP method and the 6-31G(d,p) basis set, and empirical dispersion GD3 (D3 version of Grimme's dispersion with the original D3 damping function was added at vacuum) [10]. Cyclopent-2-enone and (2-ethoxy-2-oxoethyl)dimethylsulfonium bromide were modelled as reactants. Moreover, the implicit solvent method was applied using the conductor-like polarizable continuum model (CPCM) [11]. Chloroform was selected as implicit solvent as it presented highest experimental yields than DCM, toluene or H_2O/DMSO [5]. Scan approach was applied using a 0.1 Å scale between each step with optimization (**Fig. 5.6**). The TS were calculated using the option for the force constants and No Berny optimization (noeigentest) was applied; and the lowest obtained imaginary frequency was selected for TS.

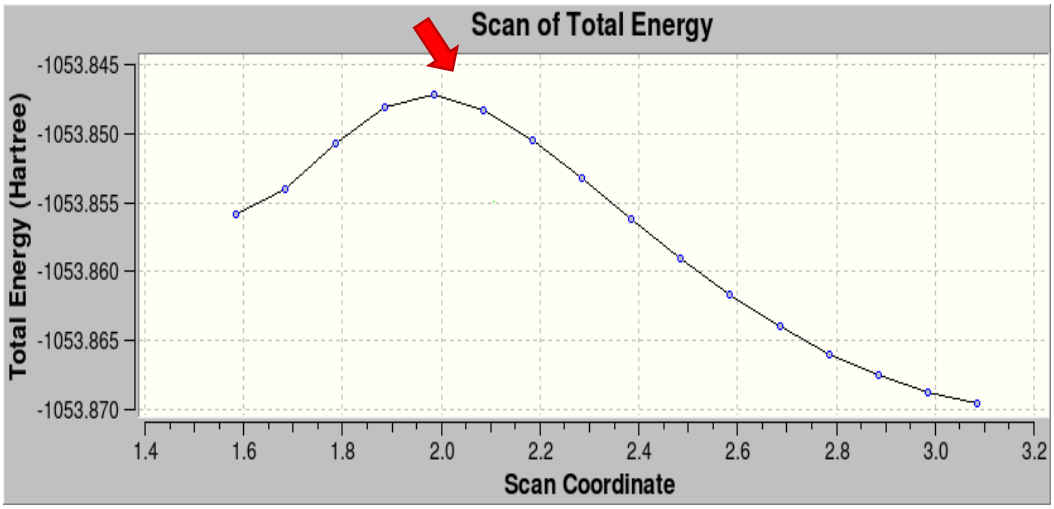


Figure 5.6. *Scan example obtained from G09 showing the coordinates explores (in Å) and total Energy (expressed in Hartrees). The red arrow shows the presumable TS obtained that must be confirmed by the corresponding imaginary frequency and IRC calculations.*

5.2. TBB alkylation

The extensive clinical use of triazole-based drugs has promoted an increased effort to develop new benzotriazole derivatives, which has resulted in anticancer benzotriazole-based compounds such as the aromatase inhibitor Vorozole®; or, ATP-competitive CK2 inhibitors 4,5,6,7-tetrabromo-1H-benzotriazole (TBB), 4,5,6,7-tetrabromo-1H-benzimidazole (TBI) and 2-dimethylamino-4,5,6,7-tetrabromo-1H-benzimidazole (DMAT), among others [12]. As seen in Chapter III, the CK2 inhibitor TBB has been extensively used in our group for the developing of novel molecules with multitarget activity against CK2 and HDACs. For this purpose, TBB was alkylated with two different bromo-alkyl derivatives, leading to two different major regioisomers (N1-substituted or N2-substituted) [13]. This reaction follows a SN2 mechanism of action, where the nucleophile corresponds to the N1 or N2 from the non-protonated TBB and the leaving group is the bromine atom from the 3-bromoprop-1-yne or 4-bromobut-1-yne. In summary, the alkylation carried out with 3-bromoprop-1-yne (**Fig. 5.7-a**) yielded the alkylated product in position N1; on the other hand, alkylation carried out with 4-bromobut-1-yne (**Fig. 5.7-b**) yielded the N2 substitution.

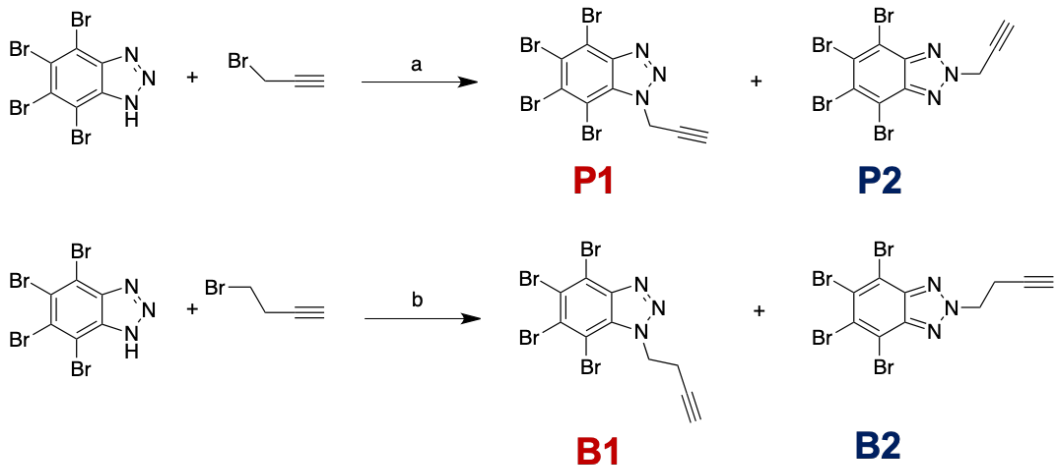


*Figure 5.7. TBB alkylation. Reagents: (a) K_2CO_3, acetone, MW, 150 ºC, 1 min; (b) K_2CO_3, CH_3CN, RT, 24 h. Image taken from RSC Adv., **2016**, 6, 66595-66608*

5.2.1. Objectives

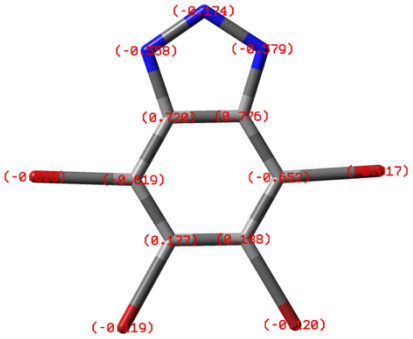
We have applied the same methodology discussed in section **5.1.1.** in order to shed light on the regioselectivity of this reaction. The total energy profile of this alkylation reaction will be obtained using different solvation models and basis sets. Additionally, calculations have been also performed using explicit solvent molecules.

5.2.2. Methodology

We have started our calculations through optimized product complexes, and by using the corresponding scans, the alkyl groups have been slightly separated until reaching the TS of the SN2 reaction. Then the calculated TS has been corroborated by the IRC calculation and RC and final P were also optimized. All calculations have been carried out by means of B3LYP/6-31G(d,p) method including empirical dispersion. In the first approach we have used implicit solvent method (CPCM with acetone); in the second approach we included three explicit acetone molecules surrounding the triazol moiety.

5.2.3. Results and discussion

The main calculated barriers are summarized in **Table 5.2**. The total EP obtained with the implicit solvent model are shown in **Fig. 5.8** and **Fig. 5.9** and with the explicit solvent in **Fig. 5.11** and **Fig. 5.12**. 3D optimized structures of the reaction are shown in. **Fig. 5.10** for the implicit solvent and **Fig. 5.13** for the explicit solvent.



One of the first concerns in this alkylation was to understand the nucleophilicity of the triazol ring present in TBB. In order to shed light on this issue, Restrained Electrostatic Potential (RESP) charges were calculated [14]. With this basis set, results show that the N1 presents more negative charge (between -0.579 and -0.558) than N2 (-0.174), suggesting that the N1 presents a higher nucleophilicity profile than N2. This fact suggests that the N1 would be the main position to be substituted in this reaction. In the implicit solvent model, the total EP displayed in **Fig. 5.8** and **Fig. 5.9** and the energies summarized in **Table 5.2** show that in all pathways the **RC** is the same for N1 or N2 alkylation resulting in the same ΔE^{cmpx} for P1, P2, B1 and B2. However, the **TS** constitutes the rate-limiting step as the $\Delta E^{\ddagger,centr}$ is lower for the N1 substitution in both cases: $\Delta E^{\ddagger,centr}$**P1** < $\Delta E^{\ddagger,centr}$ **P2** and $\Delta E^{\ddagger,centr}$**B1** < $\Delta E^{\ddagger,centr}$**B2** as can be seen in **Table 5.2**. However, this general approach cannot explain the regioselectivity of the butynyl bromide for the N2 position, but this fact could be explained by analyzing the 3D optimized structures. In the alkylation of the N2 position neither the propynyl nor the butynyl moieties are stabilized as they are too far away to make any contact within the TBB reactant (**Fig. 5.10**). For this reason, the explicit solvent model was applied.

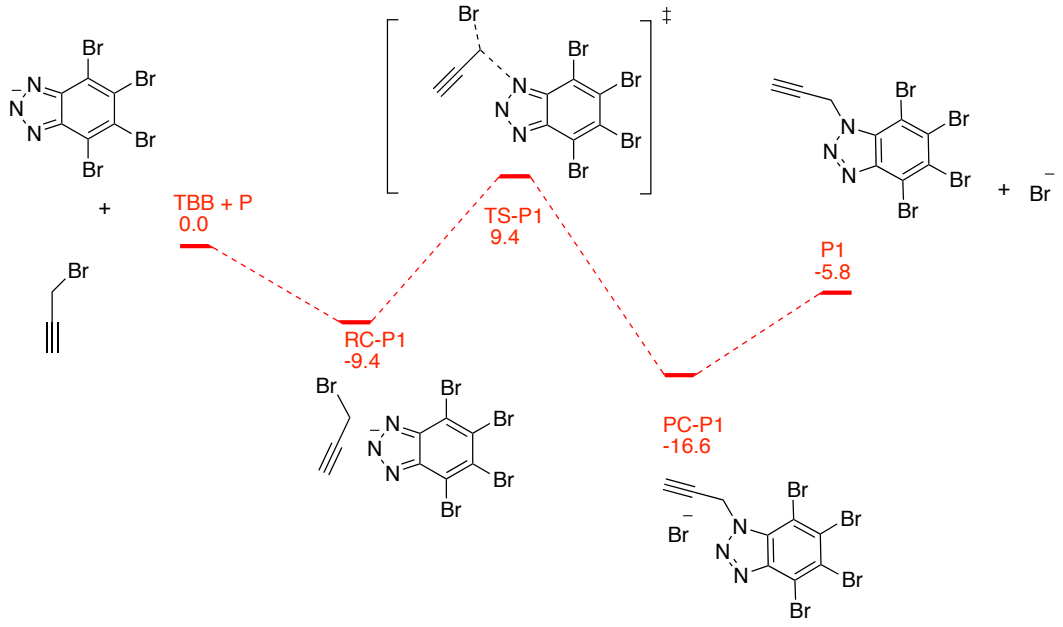
In the explicit solvent model, given the symmetry of the TBB reactant, only three molecules of acetone have been randomly added to one side of the complex near the triazol moiety (**Fig. 5.13**). The total **EP** for this reaction

displayed in **Fig. 4.13** and **Fig. 14**, and the energies summarized in **Table 5.2.** show that the propynyl bromide reaction is favored in the N1 as $\Delta E^{\ddagger,centr}$-***P2*** > $\Delta E^{\ddagger,centr}$-***P1***, while the butynyl bromide reaction is favored in the N2 by almost 2 Kcal·mol^{-1}. These results are in complete agreement with the experimental data. The presence of the explicit solvent can account for the observed difference in the calculated energy. After the optimization of the **RC** of both complexes, the acetone molecules pointed their lipophilic groups towards the TBB and the corresponding alkynes favoring an optimal orientation for the N2 substitution (**Fig. 5.13)**. Hence, in this explicit solvent model approach, the N2 alkylation is further stabilized compared to the implicit solvent model; and, in both cases, the alkyne reactant interacts with the solvent giving different EP compared to the implicit solvent calculations. Moreover, in this model, the $\Delta E^{\ddagger,ovr}$ is also favorable for the N2 position, highlighting that in this case the alkylation at N2 is energetically more efficient than N1, which is in agreement with the experimental data.

Table 5.2. *Relevant calculated barriers for both implicit and explicit solvent calculations. Calculated energies are Kcal·mol^{-1}.*

Implicit Solvent	ΔE^{cmpx}	ΔE^{react}	$\Delta E^{\ddagger,centr}$ *(TS)*	$\Delta E^{\ddagger,ovr}$
P1-Pathway	-9.4	-5.8	18.8	9.4
P2-Pathway	-9.4	-7.1	20	10.6
B1-Pathway	-10.5	-8.5	22.3	11.8
B2-Pathway	-10.5	-9.2	23.1	12.6
Explicit Solvent	ΔE^{cmpx}	ΔE^{react}	$\Delta E^{\ddagger,centr}$ *(TS)*	$\Delta E^{\ddagger,ovr}$
P1	-13.8	-5.4	20.3	6.5
P2	-13.8	-4.7	24.9	11.1
B1	-11.3	-10.7	23.3	12.0
B2	-11.3	-10.6	21.9	10.6

a) Chemical route leading to P1



b) Chemical route leading to P2

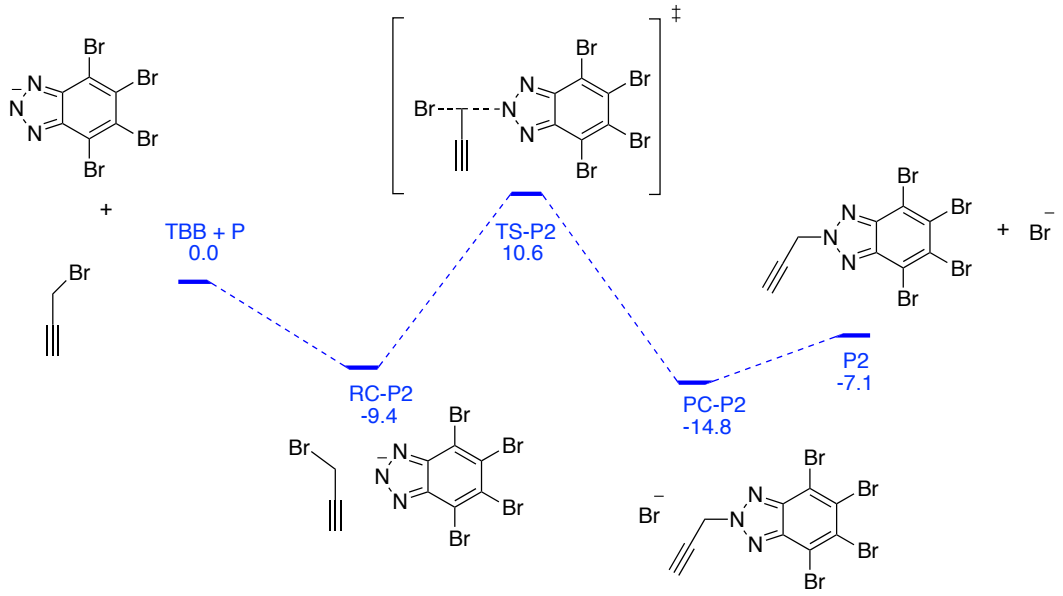


Figure 5.8. *Total energy profile of the alkylation of TBB with propynyl bromide using the B3LYP-D3/6-31G(d,p) method in implicit solvent model.* ***a)*** *N-1 alkylation.* ***b)*** *N-2 alkylation. Calculated energies are Kcal·mol^{-1}*

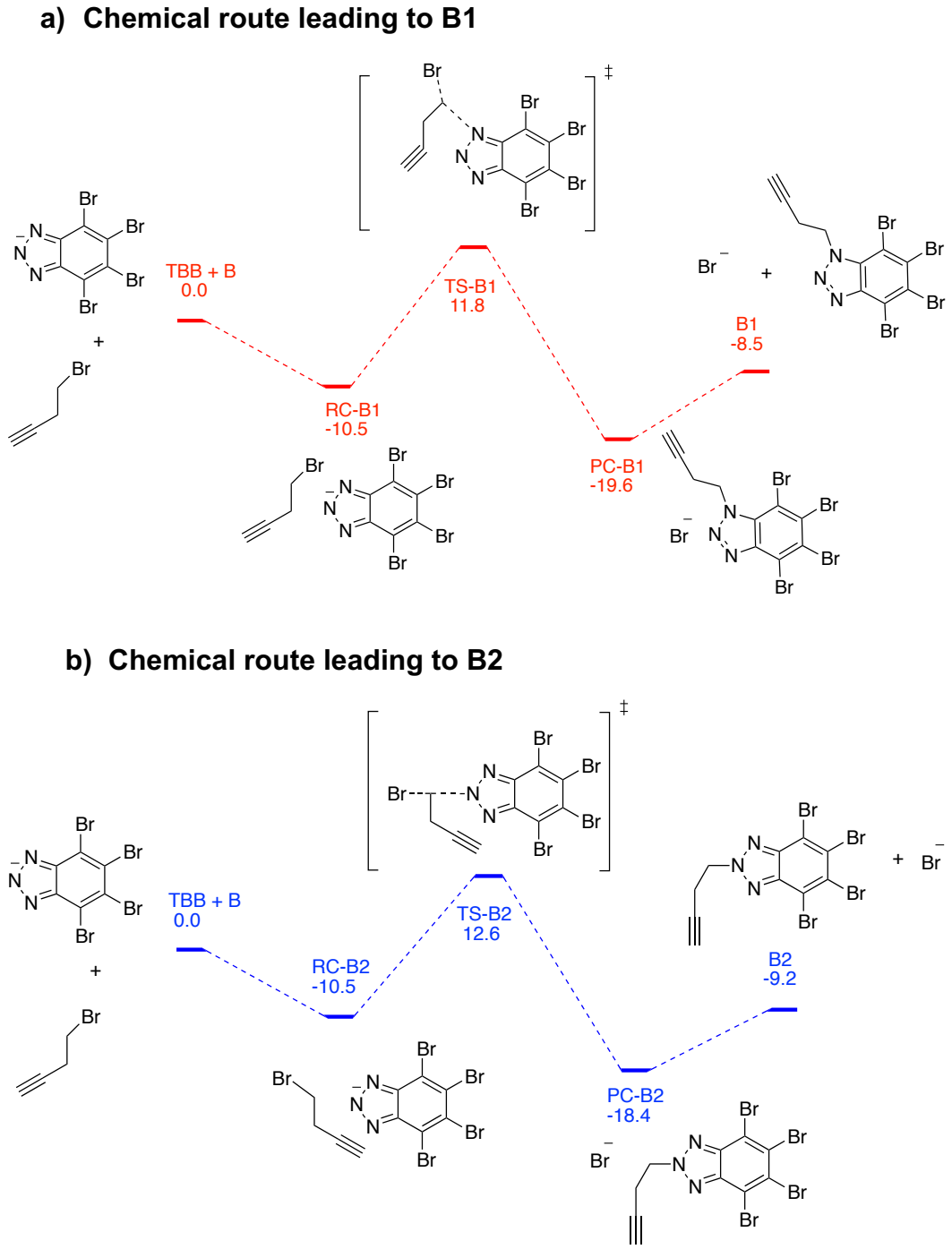


Figure 5.9. *Total energy profile of the alkylation of TBB with butynyl bromide using the B3LYP-D3/6-31G(d,p) method in the implicit solvent model.* ***a)*** *N-1 alkylation.* ***b)*** *N-2 alkylation. Calculated energies are Kcal·mol^{-1}.*

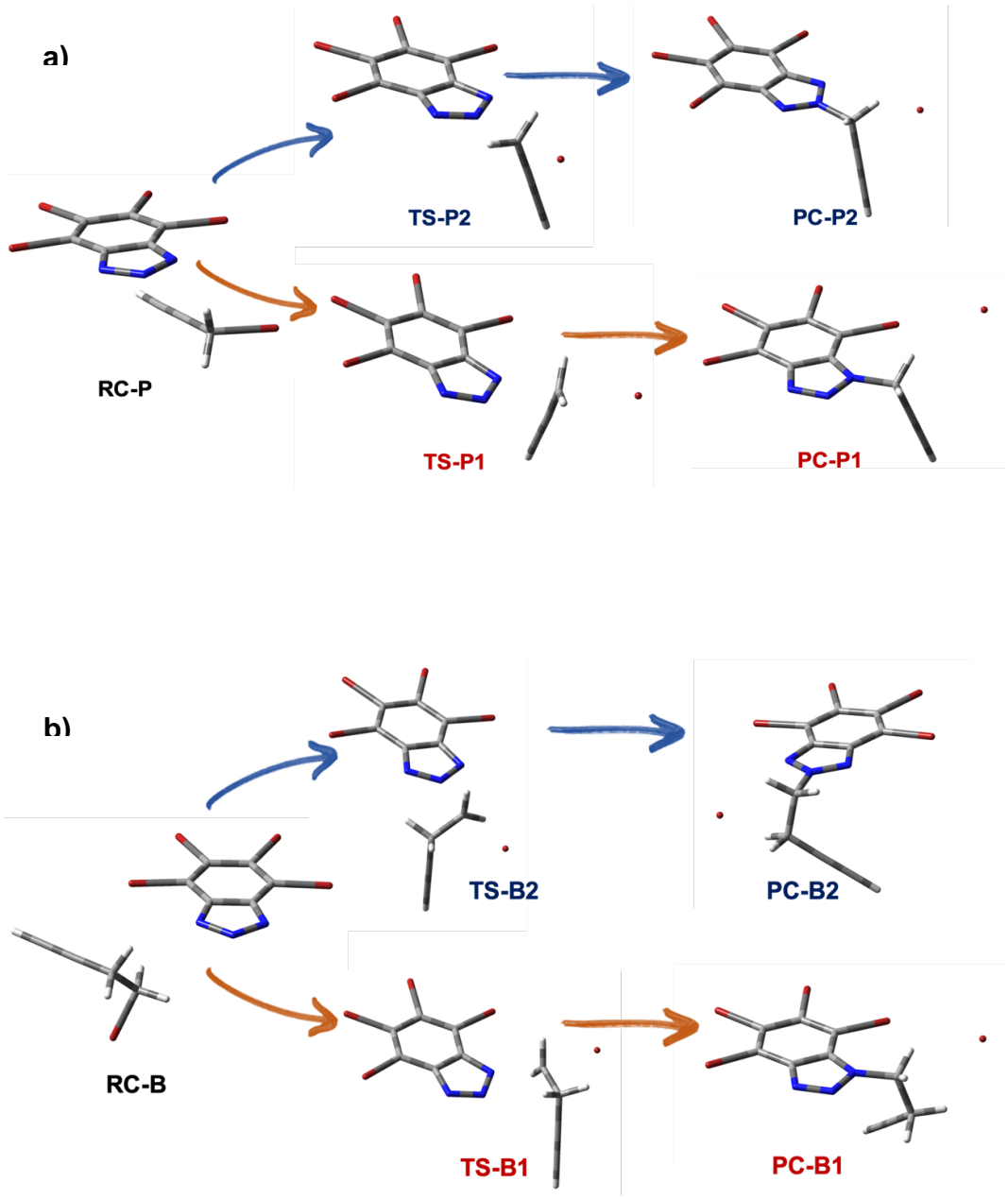
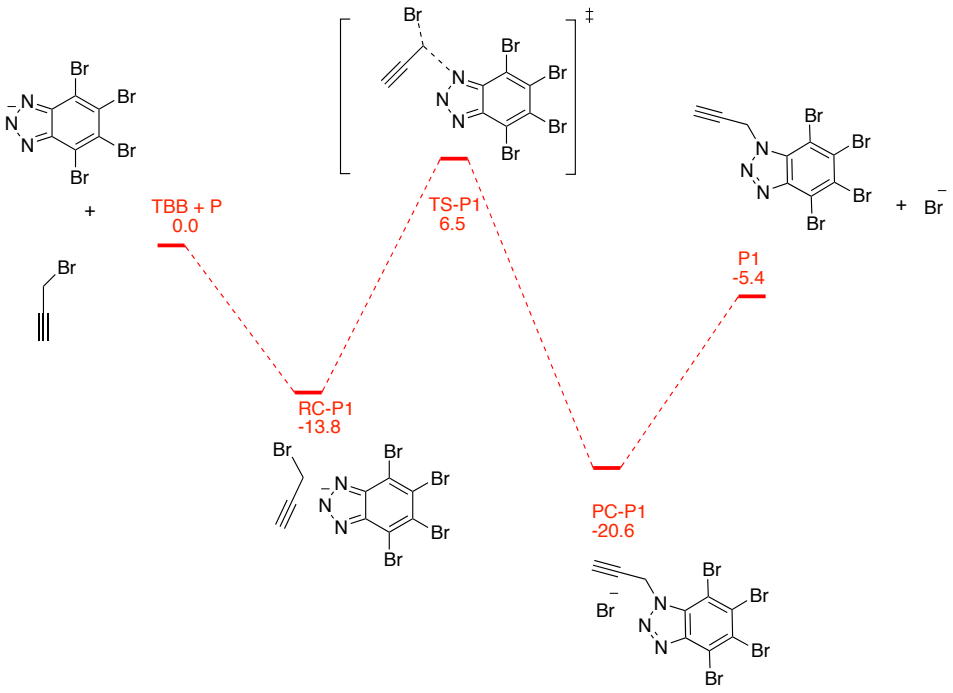


Figure 5.10. *3D optimized structures of **RC**, **TS** and **PC** for the TBB alkylation in the implicit solvent model. **a)** with propynyl bromide and **b)** butynyl bromide in both N1 and N2 positions.*

a) Chemical route leading to P1 including three acetone molecules



b) Chemical route leading to P1 including three acetone

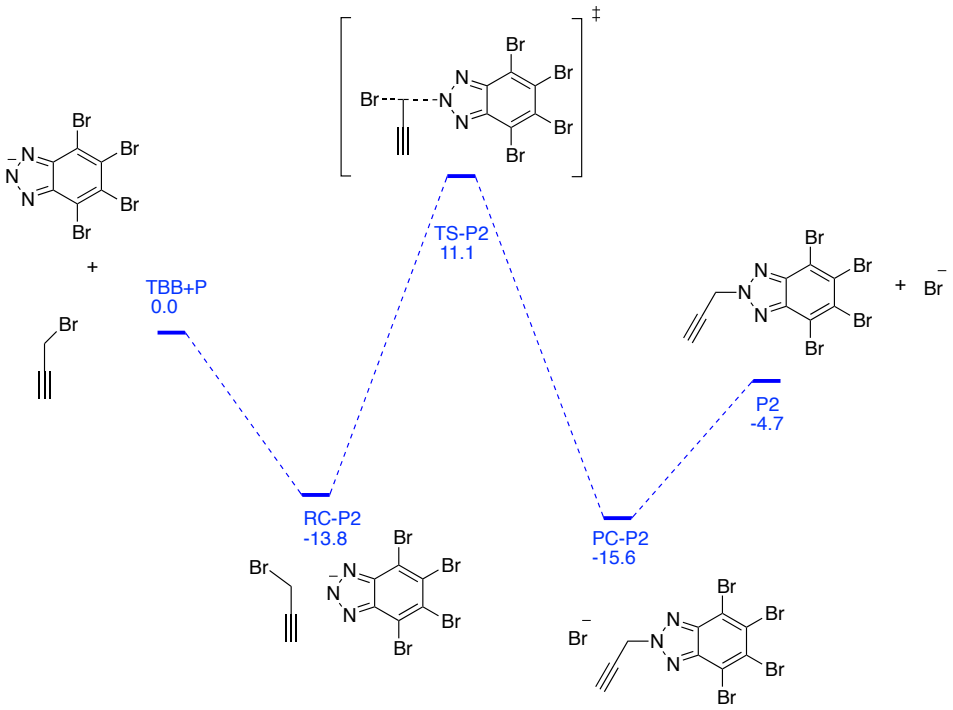
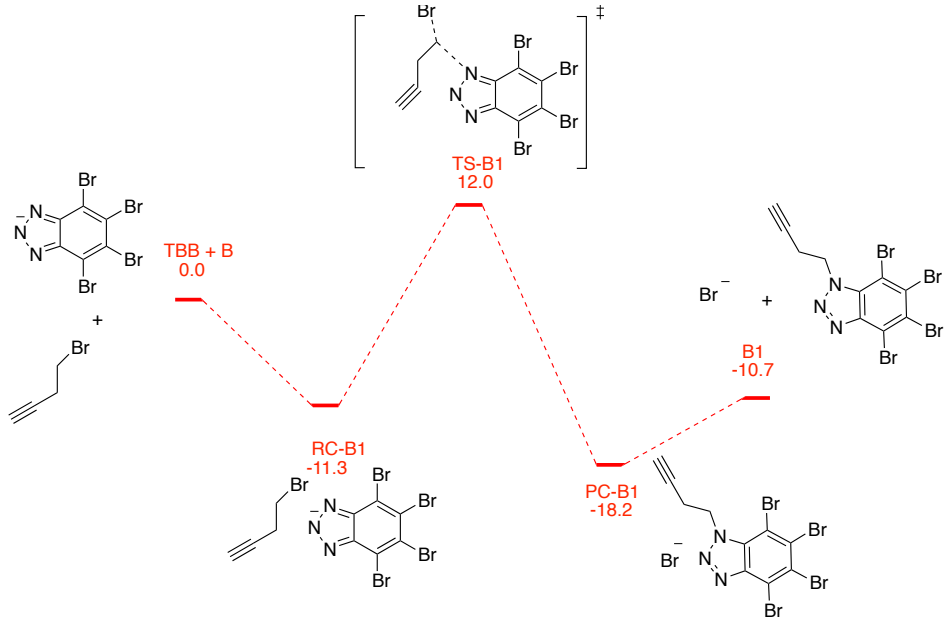


Figure 5.11. *Total energy profile of the alkylation of TBB with propynyl bromide using the B3LYP-D3/6-31G(d,p) method in the presence of three acetone molecules (these*

molecules are excluded from the scheme for clarity. ***a)*** *N-1 alkylation.* ***b)*** *N-2 alkylation. Calculated energies are Kcal·mol^{-1}.*

a) Chemical route leading to B1 including three acetone molecules



b) Chemical route leading to B2 including three acetone molecules

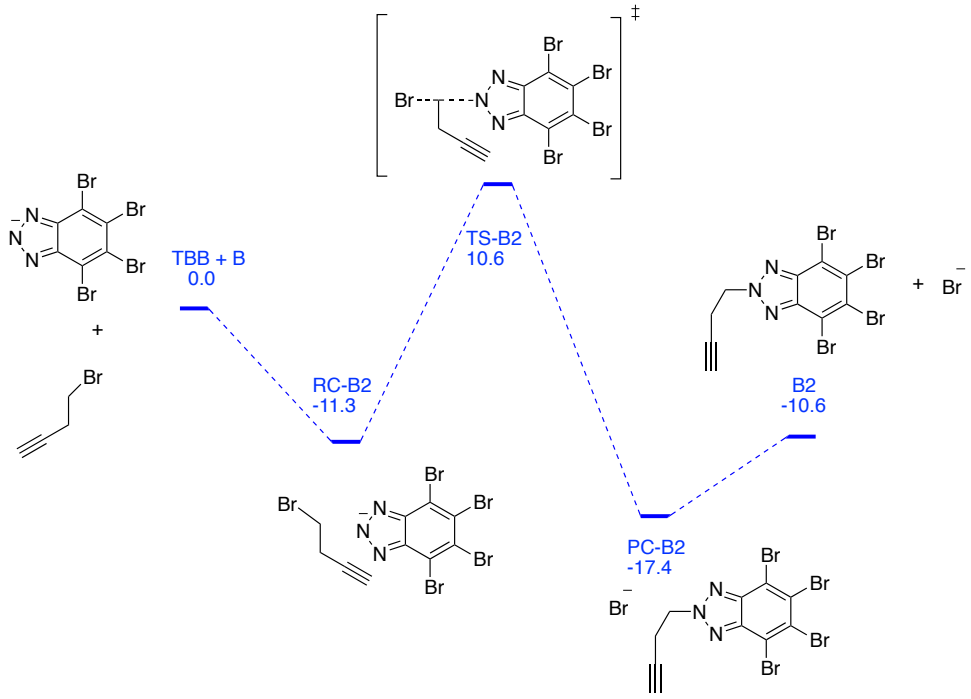


Figure 5.12. *Total energy profile of the alkylation of TBB with butynyl bromide using the B3LYP-D3/6-31G(d,p) method in the presence of three acetone molecules (these molecules are excluded from the scheme for clarity.* ***a)*** *N-1 alkylation.* ***b)*** *N-2 alkylation. Calculated energies are Kcal·mol^{-1}.*

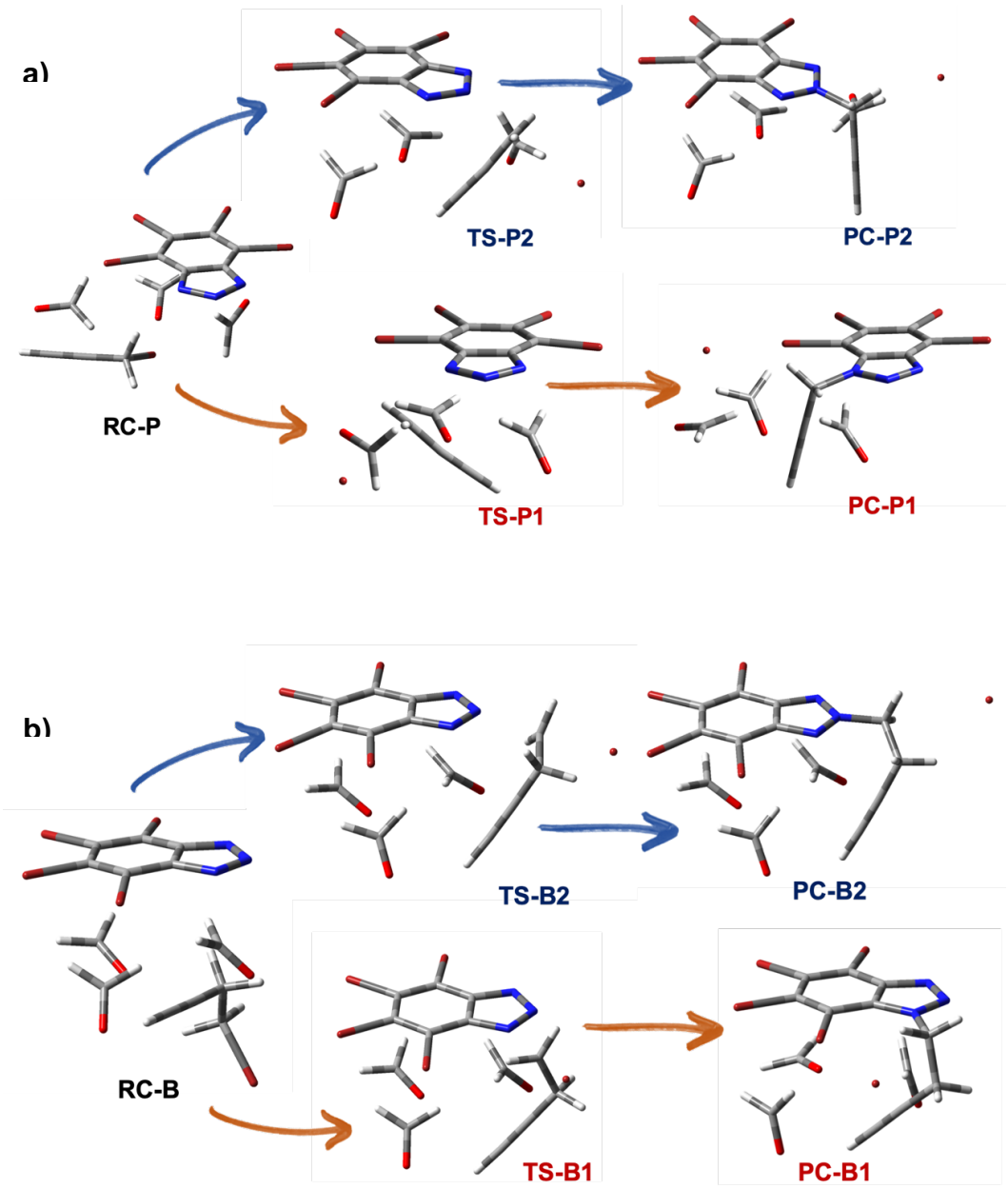


Figure 5.13. *3D optimized structures of **RC**, **TS** and **PC** for the TBB alkylation in the presence of three acetone molecules. **a)** Propynyl bromide and **b)** Butynyl bromide alkylation in both N1 and N2 positions.*

5.2.4. Conclusions

The total energy profile was obtained first by using B3LYP-D3/6-31G(d,p) with an implicit solvent method and then with explicit solvent molecules. In the explicit solvent approach, the N1 alkylation by propynyl bromide derivatives was correlated with the experimental data, but N2 alkylation by butynyl bromide was not. This fact is due two main reasons: i) the N2 position according to our calculations is less nucleophilic than the N1; ii) and the TS-B2 where the alkyne fraction does not interact with other section of the TBB molecule.

The explicit solvent approach, by addition of three molecules of acetone to the system provides further stabilization of the 4-bromobut-1-yne thus, lowering the TS-B2 energies. In this explicit solvent model, the relative energies favor the alkylation in the N1 position by the 3-bromoprop-1-yne, and in the N2 position by the 4-bromobut-1-yne, which is in complete agreement with the experimental data.

This is not the first time a reaction has been rationalized and predicted more accurately by using explicit solvent molecules as some Diels-Alder reactions have been described by adding hydrogen-bonding solvents to the model systems [15]. In adition, other studies show that implicit solvent methods are not accurate for calculating binding free energies in comparison to the explicit models [16–18]. Nevertheless, this study is not fully finished, since the explicit solvent molecules do not interact with the total surface of the reactants. Systems that are more complex should be built by adding more solvent molecules and DFT methods for further obtaining a global understanding of this chemical reaction.

5.3. Study of the Conformational Stability by Intramolecular Sulfur Bonds

Conformational restriction is a key issue in drug design as while fixing the active conformation reduces the entropy component of the binding process. Compounds reported by *Sheriff et al.* [19] not only present very interesting conformational requirements; but also, require the presence of a thiophen and thioether moiety for the activity as the substitution of those by nitrogen or oxygen atoms leads to a decrease in the inhibitory activity. In this chapter, we want to give an explanation to the results reported by *Sheriff et al.* and computationally study how the sulfur atoms stabilize the active conformations.

As we mentioned in the introduction, sulfur atoms present sigma holes that can interact within donor atoms such as oxygen or nitrogen atoms that can establish intramolecular bonds, further stabilizing the desired conformation and therefore increase the activity of the molecule (**Fig. 5.14**) [20]. The 3D structure of compounds **7WL** and **NX4**, crystallized in complex with PTPRG, has been studied by QM methods. Interestingly, both compounds present a sulfonamide amide and two additionally sulfur atoms, on one inside a thiophene ring and on the other one in a thioether fraction.

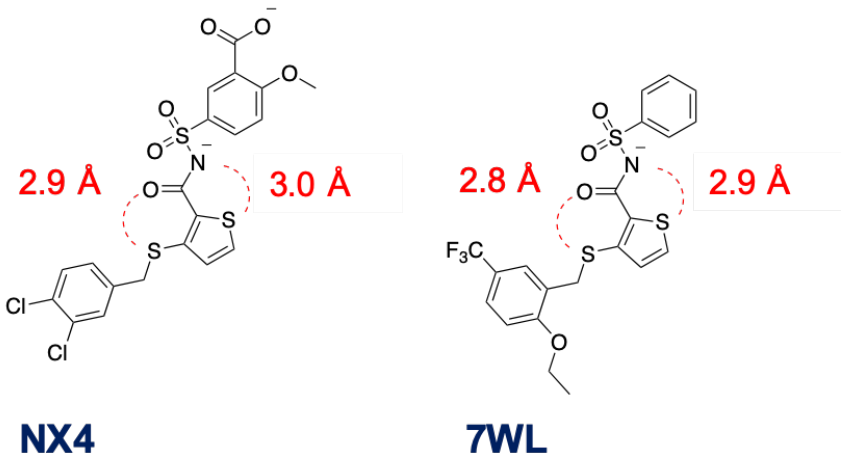


Figure 5.14. *Chemical structure of compounds **NX4** and **7WL**. The sulfur bonds are highlighted in red dashed lines (in Å) highlighted the intramolecular bonds.*

5.3.1. Objectives

We aim to predict and further study the intramolecular bonds involving the two sulfur atoms present in **NX4** and **7WL.**

5.3.2. Methods

The initial structure of **7WL** was taken from the crystal structure **5H08** and the **NX4** from **3QCJ**. The protonated states were calculated using the web server chemicalize.com.

QM methods have been applied to obtain the molecular electrostatic potential (MEP) where V_{max} and V_{min} can be calculated. The V_{max} values are calculated for the sulfur atoms and represent the sigma holes; while the V_{min} values are calculated for the nitrogen and oxygen atoms (intramolecular bonded atoms) as a representation of the free electron pairs. In addition, each point of the electrostatic energy can be calculated for further understanding of the bond strength.

Optimization and total electron density were obtained by using the B3LYP method and the basis set cc-pVQZ with the software G09. This large basis set was selected in order to correctly predict the low-lying σ* orbitals of the C−S bond [21]. The MEP minima has been used for the prediction and calculation of the hydrogen bond basicity (pK_{BHX}), while the maxima has been used to obtain the pKa values and the sigma-holes in halogen atoms [22–26]. Herein, the maximum and minimum means of the wave functional algorithm from the MultiWFN program version 3.7. taking an isodensity value of 0.001 a.u. [27].

5.3.3. Results and discussion

The sulfonamide was predicted to be deprotonated at pH = 7.2 with a pKa value for the nitrogen of 4.02 and 4.11 for **7WL** and **NX4**, respectively. After that, DFT studies were performed in order to understand the role of this sulfur atoms. The optimized structures preserve almost the same 3D structure compared to the original crystal structure (**Fig. 5.15.**). In the case

of **7WL**, the superimposition shows no displacement, while in the case of **NX4** there is a slight displacement of the benzoic acid moiety. This is probably due to the fact that the starting conformation was a result of the induced fit that takes place upon protein binding and the minimization was carried out in vacuum.

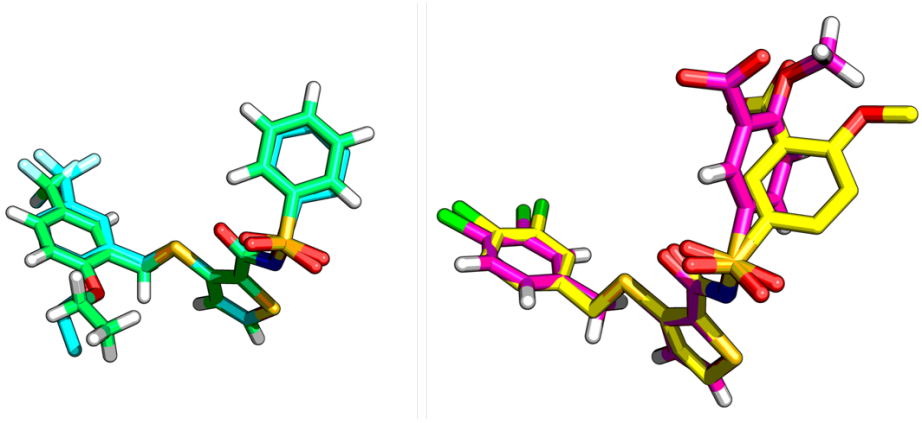


Figure 5.15. *Original crystal structure **7WL** (cyan) and **NX4** (yellow) superimposed with the optimized QM structures (pale green and magenta, respectively).*

The calculated V_{max} and V_{min} values for both molecules are represented in **Fig. 5.16.**, while the energy values obtained for the V_{max} of the sulfur atom and the corresponding V_{min} from the intramolecular bonded atom (the nitrogen or oxygen) have been summarized in **Table 4.6**.

The calculated V_{max} Enery values present positive values, while V_{min} present negatives values. The V_{max} value is 30Kcal·mol^{-1} higher for **7WL** compared to **NX4**; on the contrary, the V_{min} is more negative in **NX4** than in **7WL**.This is because **NX4** presents a total charge of -2 and **7WL** of -1 thus making the MEP more negative for **NX4** than for **7WL**. In both cases, the interaction between the V_{max} and V_{min} is strong, which means that this intramolecular bond seems to be extremely stable (**Table 5.3**). Moreover, the V_{max} and V_{min} calculated in both molecules have shown some interesting features. In the case of **NX4** the chlorine atoms present a V_{max} parallel to the halogen which represent the σ-holes (**Fig. 5.16-f**). The carboxylate group presents also V_{min} points that are related with the unpaired electrons. In **7WL** it is interesting to mention that the -CF_3 moiety presents three V_{min}, each

corresponding to a fluorine atom (**Fig. 5.16-c**) and one V_{max} in the middle of all three. In addition, V_{min} points can be observed in the center of the benzyl, which that clearly represents the electronic cloud in the aromatic rings.

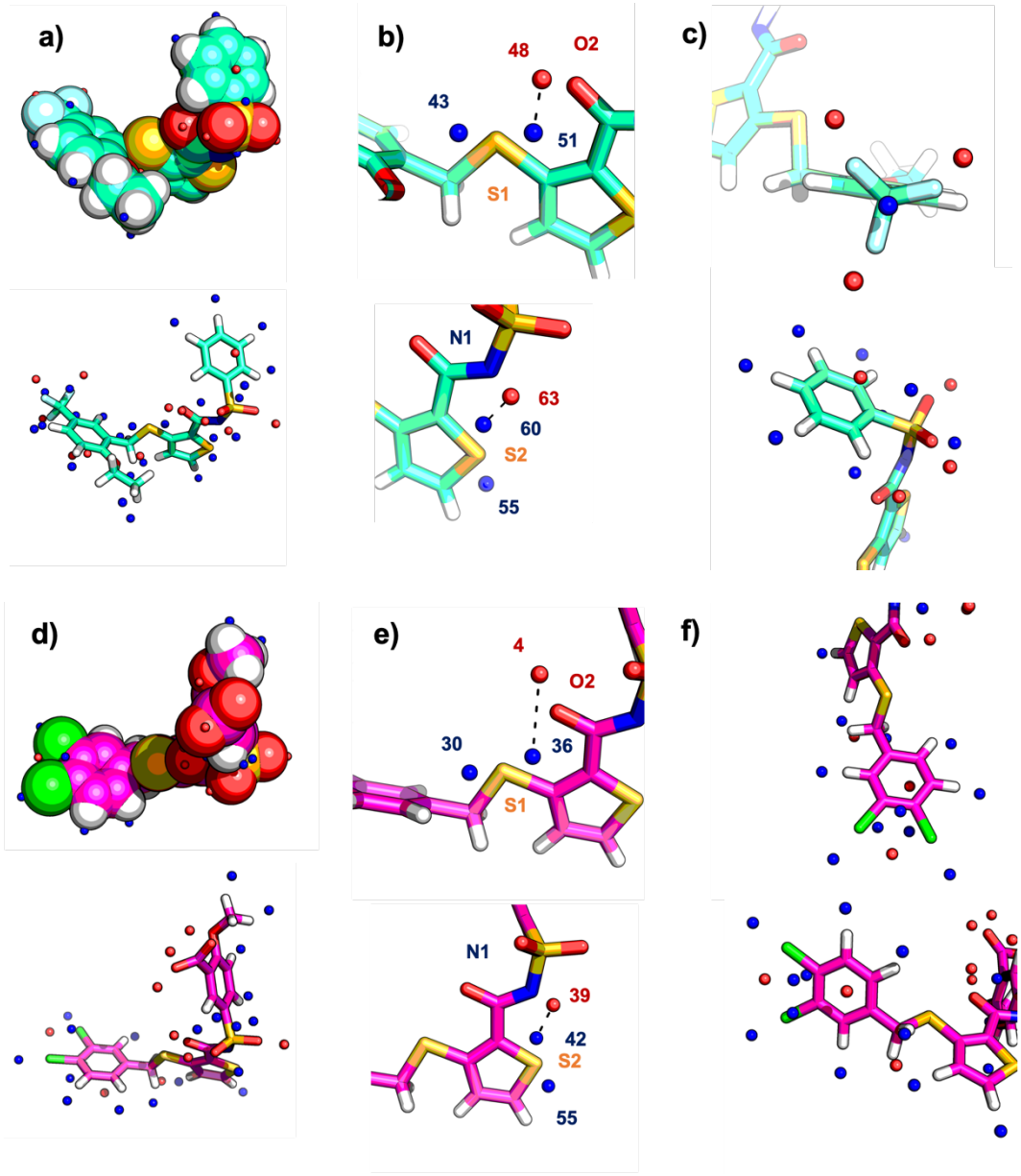


Figure 5.16. a) *&* ***d)*** *General view of compound* ***7WL*** *(pale green) and* ***NX4*** *(magenta) with the corresponding V_{min} and V_{max} calculated points (red and blue dots, respectively).* ***b)*** *&* ***d)*** *In detail sulfur σ* orbitals calculated interacting with the corresponding electron pair from nitrogen or oxygen atoms.* ***c)*** *and* ***f)*** *In detail V_{max} and V_{min} from the trifluoromethyl moiety, dichlorobenzene rings and aromatic rings.*

Table 5.3. *Calculated* V_{max} *and* V_{min} *from* ***7WL*** *and* ***NX4*** *in the sulfur atoms and the corresponding intramolecular atom bond.*

Compound	Atom	V_{max} (point)	Kcal·mol^{-1}
7WL	S1	43	83.46
		51	69.47
	S2	56	99.03
		60	68.49
	Atom	**V_{min} (point)**	**Kcal·mol^{-1}**
	O2	48	-187.45
	N1	63	-196.33
NX4	**Atom**	**V_{max} (point)**	**Kcal·mol^{-1}**
	S1	30	50.80
		36	35.26
	S2	42	30.63
		39	62.63
	Atom	**V_{min} (point)**	**Kcal·mol^{-1}**
	O2	4	-199.03
	N1	53	-240.60

5.3.4. Conclusions

Both optimized structures preserve the overall 3D conformation from the original crystal structure. The calculated molecular electrostatic potential (MEP) together with the obtained V_{max} and V_{min} values have been useful to understand the intramolecular interactions that stabilize the conformations in the 3D structures. Both molecules present two high energy N···S and O···S bonds, but higher in the case of **7WL** than **NX4**. Moreover, the V_{max} and V_{min} can be also useful for calculating and predicting the sigma-holes in halogen atoms and the pK_{BHX} from hydrogen bond acceptors or donors. These calculations account for the experimental results. Substitution of the sulfur atoms by nitrogen or oxygen leads to a complete loss of the activity, most probably due to these molecules adopting a non active conformation. Altogether, it is possible to calculate V_{max} and V_{min} by QM methods in order to predict the interaction with the receptor or the stabilization of intramolecular bonds.

5.4. References

1. Hamlin, T. A.; Swart, M.; Bickelhaupt, F. M. Nucleophilic Substitution (S N 2): Dependence on Nucleophile, Leaving Group, Central Atom, Substituents, and Solvent. *ChemPhysChem* **2018**, *19*, 1315–1330.

2. Riveros, J. M.; José, S. M.; Takashima, K. Gas-phase Nucleophilic Displacement Reactions.; 1985; p. 197–240.

3. Bento, A. P.; Solà, M.; Bickelhaupt, F. M. E2 and SN2 Reactions of X− + CH3CH2X (X = F, Cl); an ab Initio and DFT Benchmark Study. *J. Chem. Theory Comput.* **2008**, *4*, 929–940.

4. Manz, T. A.; Sholl, D. S. A dimensionless reaction coordinate for quantifying the lateness of transition states. *J. Comput. Chem.* **2009**, *31*, 1528–1541.

5. López-Rodríguez, A.; Domínguez, G.; Pérez-Castells, J. Microwave-Mediated Sulfonium Ylide Cyclopropanation. Stereoselective Synthesis of

Cyclopropa[c]pentalenes. *ChemistrySelect* **2017**, *2*, 2565–2568.

6. Jaccob, M.; Venuvanalingam, P. Mechanism and diastereoselectivity of arsenic ylide mediated cyclopropanation: a theoretical study. *RSC Adv.* **2013**, *3*, 17793.

7. Janardanan, D.; Sunoj, R. B. Chemo-, regio-, and diastereoselectivity preferences in the reaction of a sulfur ylide with a dienal and an enone. *Org. Biomol. Chem.* **2011**, *9*, 1642.

8. Janardanan, D.; Sunoj, R. B. Density Functional Theory Investigations on Sulfur Ylide Promoted Cyclopropanation Reactions: Insights on Mechanism and Diastereoselection Issues. *J. Org. Chem.* **2007**, *72*, 331–341.

9. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, and D. J. F. (Gaussian, Inc., Wallingford CT, 2009). *Gaussian 09*.

10. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

11. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

12. Ren, Y. Recent Development of Benzotriazole-based Medicinal Drugs. *Med. Chem. (Los. Angeles).* **2014**, *4*.

13. Purwin, M.; Hernández-Toribio, J.; Coderch, C.; Panchuk, R.; Skorokhyd, N.; Filipiak, K.; de Pascual-Teresa, B.; Ramos, A. Design and synthesis of novel dual-target agents for HDAC1 and CK2 inhibition. *RSC Adv.* **2016**, *6*, 66595–66608.

14. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

15. Yang, Y.-F.; Yu, P.; Houk, K. N. Computational Exploration of Concerted and Zwitterionic Mechanisms of Diels–Alder Reactions between 1,2,3-Triazines and Enamines and Acceleration by Hydrogen-Bonding Solvents. *J. Am. Chem. Soc.* **2017**, *139*, 18213–18221.

16. Zhang, J.; Zhang, H.; Wu, T.; Wang, Q.; van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **2017**, *13*, 1034–1043.

17. Hamid, A.; Roy, R. K. Solvent effect on stabilization energy: An approach based on density functional reactivity theory. *Int. J. Quantum Chem.* **2019**, *119*, 25909.

18. Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angew. Chemie Int. Ed.* **2018**, *57*, 4170–4176.

19. Sheriff, S.; Beno, B. R.; Zhai, W.; Kostich, W. A.; McDonnell, P. A.; Kish, K.; Goldfarb, V.; Gao, M.; Kiefer, S. E.; Yanchunas, J.; et al. Small Molecule Receptor Protein Tyrosine Phosphatase γ (RPTPγ) Ligands That Inhibit Phosphatase Activity via Perturbation of the Tryptophan–Proline–Aspartate (WPD) Loop. *J. Med. Chem.* **2011**, *54*, 6548–6562.

20. Beno, B. R.; Yeung, K.-S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A. A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. *J. Med. Chem.* **2015**, *58*, 4383–4438.

21. Denis, P. A. Basis Set Requirements for Sulfur Compounds in Density Functional Theory: a Comparison between Correlation-Consistent, Polarized-Consistent, and Pople-Type Basis Sets. *J. Chem. Theory Comput.* **2005**, *1*, 900–907.

22. Kenny, P. W.; Montanari, C. A.; Prokopczyk, I. M.; Ribeiro, J. F. R.; Sartori, G. R. Hydrogen Bond Basicity Prediction for Medicinal Chemistry Design. *J. Med. Chem.* **2016**, *59*, 4278–4288.

23. Caballero-García, G.; Mondragón-Solórzano, G.; Torres-Cadena, R.;

Díaz-García, M.; Sandoval-Lira, J.; Barroso-Flores, J. Calculation of VS,max and Its Use as a Descriptor for the Theoretical Calculation of pKa Values for Carboxylic Acids. *Molecules* **2018**, *24*, 79–94.

24. Thapa, B.; Schlegel, H. B. Density Functional Theory Calculation of p K a ’s of Thiols in Aqueous Solution Using Explicit Water Molecules and the Polarizable Continuum Model. *J. Phys. Chem. A* **2016**, *120*, 5726–5735.

25. Kenny, P. W. Prediction of hydrogen bond basicity from computed molecular electrostatic properties: implications for comparative molecular field analysis. *J. Chem. Soc. Perkin Trans. 2* **1994**, 199–202.

26. Ibrahim, M. A. A. Molecular mechanical perspective on halogen bonding. *J. Mol. Model.* **2012**, *18*, 4625–4638.

27. Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592.