UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Jorge Enrique Hernández González

Computational Strategies for Selective Inhibition of Falcipain-2

São José do Rio Preto
2020

Jorge Enrique Hernández González

Computational Strategies for Selective Inhibition of Falcipain-2

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: FAPESP – Proc 2016/24587-9
CAPES

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

Coorientador: Prof. Dr. Pedro Geraldo Pascutti

São José do Rio Preto

2020

H557c

Hernández-González, Jorge Enrique

Computational strategies for selective inhibition of falcipain-2 / Jorge Enrique Hernández-González. -- São José do Rio Preto, 2020
186 f. : il., tabs.

Tese (doutorado) – Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto.
Orientador: Vitor Barbanti Pereira Leite
Coorientador: Pedro Geraldo Pascutti

1. Biologia molecular. 2. Biofísica molecular. 3. Malária. 4. Plasmodium falciparum. 5. Dinâmica molecular. I. Título.

Jorge Enrique Hernández González

Computational Strategies for Selective Inhibition of Falcipain-2

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: FAPESP – Proc.. 2016/24587-9
CAPES – 031/2013 PRÓ-DEFESA 3

Comissão Examinadora

Prof. Dr. Vitor Barbanti Pereira Leite
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Sidney Jurado de Carvalho
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Alexandre Suman de Araujo
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Rodrigo Guerino Stabeli
USP – Câmpus de Riberão Preto

Prof. Dr. Gustavo Troiano Feliciano
UNESP – Câmpus de Araraquara

São José do Rio Preto

14 de fevereiro de 2020

A aquellos que aun en medio de las
adversidades mantienen viva la esperanza.

ACKNOWLEDGMENTS

I would like to thank God, our Lord. Without Him I would not be here today, without His love and mercy I would not have the strength and courage to move forward and face the obstacles of life.

I thank my daughter and my wife for their support, their company, their understanding and their true love. I am grateful to my mother for all the dedication and unconditional love she has given me throughout my life. My father, although not with us anymore, would be proud of his son's achievement. I am thankful for his support and guidance. Many thanks to my brother, my grandparents and the rest of the family. All of you dwell in a special place of my heart despite the distance.

I am very grateful to professor Pedro Pascutti for having trusted me when he barely knew me, and for having given me the opportunity to come to Brazil to pursue my career. I also thank professor Vitor Leite for having accepted me as his PhD student and for his constant support. Without both, this work would not have been possible.

Many thanks to my collaborators, especially Emir Salas, Pedro Valiente and Chris Oostenbrink for their contribution to this work. Thanks to my friends for making my life happier and worth living, and to my teachers, professors and former supervisors for having guided my first steps in Science.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001: 031/2013 PRÓ-DEFESA 3.

I also thank FAPESP, for the concession of the research grant No. 2016/24587-9, Sao Paulo Research Foundation (FAPESP), which provided financial support to my PhD project during the last three years.

RESUMO

A falcipaína-2 (FP-2) é uma hemoglobinase chave do *Plasmodium falciparum*, frequentemente selecionada como alvo para o planejamento de fármacos antimaláricos. Apesar dos esforços, nenhum inibidor da FP-2 tem entrado em fase de ensaios clínicos, pois a inibição cruzada de alvos humanos constitui um obstáculo. Neste trabalho, abordamos questões relacionadas à inibição seletiva da FP-2 através de diferentes técnicas computacionais, como simulações de dinâmica molecular, triagem virtual, ancoramento molecular, análise da comunidades e cálculos de volume de cavidades e energia livre. Ademais, realizamos ensaios de inibição *in vitro* para validar as predições. Apresentamos os compostos HTS07940 e HTS08262, que inibem a FP-2 e culturas de *P. falciparum* com boa seletividade respeito à catepsina K humana (hCatK) e à linha celular HeLa. Os inibidores foram identificados por meio de uma estratégia de triagem virtual que descartou ligantes com alta afinidade pela hCatK. Outras formas de atingir inibição seletiva são exploradas, como a busca por inibidores alostéricos, que ligam tipicamente sítios menos conservados. Analisamos uma região da FP-2 equivalente a um sítio alostérico da hCatK previamente caracterizado, o sítio 6, e predizemos vários ligantes potenciais. Após avaliação experimental, dois compostos, ZINC03225317 e ZINC72290660, são confirmados como inibidores não competitivos da FP-2, o que reforça a relevância deste sítio para o planejamento de fârmacos. A busca de cavidades alostéricas é expandida a outros sítios, baseada em informações experimentais prévias relacionadas a uma *E*-chalcona identificada acidentalmente como inibidor não competitivo da FP-2. Os resultados revelam a ocorrência de uma cavidade transiente numa região denominada sítio 3, que permanece maioritariamente ocluída pela cadeia lateral do resíduo K34. O modo de ligação predito da *E*-chalcona é consistente com os dados experimentais disponíveis e fornece informação sobre o mecanismo alostérico a nível molecular. Por fim, estudamos os determinantes moleculares da alta seletividade pela FP-2 de nitrilos que contêm substituintes de 3-piridina em P2, previamente estudados. Segundo as nossas predições, pontes de água envolvendo os resíduos I85 e D234 da FP-2 e o nitrogênio da piridina explicam os perfis experimentais de atividade. Portanto, inibidores seletivos da FP-2 podem ser planejados promovendo a formação de pontes de água no fundo do subsítio S2 e/ou introduzindo grupos químicos que substituem a molécula de água envolvida.

Palavras-chave: Malária. Plasmodium falciparum. Falcipaína-2. Inibidores. Seletividade. Alosteria. Dinâmica molecular. Triagem virtual. Energia livre. Ensaios de inibição. Cavidades transientes. Modo de ligação. Comunidade. Rota de sinalização. Ponte de água.

ABSTRACT

Falcipain-2 (FP-2) is a key hemoglobinase of *Plasmodium falciparum* that has been extensively targeted in antimalarial drug discovery projects. Despite the efforts, none of the existing FP-2 inhibitors has entered the clinical trials yet, as cross-inhibition of human off-targets remains a major concern. Here, we tackle issues related to selective FP-2 inhibition by employing different computational techniques, such as molecular dynamics simulations, virtual screening, docking, community analysis and pocket volume and free energy calculations. We also perform *in vitro* inhibition assays to validate the predictions. We report two compounds, HTS07940 and HTS08262, displaying inhibition against FP-2 and *P. falciparum* cultures with good selectivity with respect to human cathepsin K (hCatK) and HeLa cell line. The inhibitors were identified through a virtual screening strategy that ruled out ligands with high affinity for hCatK. Moreover, other ways to achieve selective inhibition are explored, such as the search for allosteric inhibitors, which bind to typically less conserved sites. We analyze an FP-2 region equivalent to a previously-characterized allosteric site of hCatK, termed site 6, and predicted various potential ligands. After experimental evaluation, two compounds, ZINC03225317 and ZINC72290660, are confirmed as non-competitive inhibitors of FP-2, thus reinforcing the suitability of site 6 as a druggable allosteric cavity. The search for potential cavities is expanded to other sites on the basis of previous experimental information related to a serendipitously-identified *E*-chalcone displaying non-competitive inhibition against FP-2. Our results reveal the occurrence of a transient pocket in a region termed site 3, which remains occluded most of the simulation time by the side-chain of residue K34. The predicted binding mode of the *E*-chalcone is consistent with the available experimental data and sheds light upon important features of the allosteric mechanism at molecular level. Finally, we study the molecular determinants of the high selectivity for FP-2 of previously-reported nitriles containing 3-pyridine substituents at P2. Our results reveal that water bridges involving residues I85 and D234 of FP-2, and the pyridine nitrogen, explain the experimental activity profiles. Therefore, selective FP-2 inhibitors can be designed by promoting the formation of water bridges at the bottom of the S2 subsite and/or by introducing substituents that replace the bridging water molecule.

Keywords: Malaria. Plasmodium falciparum. Falcipain-2. Inhibitors. Selectivity. Allostery. Molecular dynamics simulation. Virtual screening. Free energy. Inhibition assays. Transient pockets. Binding mode. Community. Signaling Pathway. Water bridge.

FIGURE LIST

TABLE LIST

ABBREVIATIONS AND SYMBOLS

2D	Two-dimensional
3D	Three-dimensional
AMC	7-amino-4-methyl coumarin
aMD	Accelerated Molecular Dynamics
APBS	Adaptive Poisson-Boltzmann Solver
CA	Community Analysis
CC_{50}	Half-Maximal Cytotoxic Concentration
Cpd48	Compound 48
Cpd66	Compound 66
CpHMD	Constant pH Molecular Dynamics
CQ	Chloroquine
DDT	Dithiothreitol
DMEM	Modified Eagle’s Medium
DMSO	Dimethyl sulfoxyde
E64	*trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane
EM	Energy Minimization
ENR	Enoyl Reductase
ESP	Electrostatic Potential
FBS	Fetal Bovine Serum
f_d	Fraction of Deprotonation
FEP	Free Energy Perturbation
ff14SB	ff14SB Amber Force-Field
FP-2	Falcipain-2
FP-3	Falcipain-3
FPs	Falcipains
gaff	Generalized Amber Force-Field
GC	Generalized Correlation
H-Bond	Hydrogen Bond
hCatB	Human Cathepsin B
hCatK	Human Cathepsin K

hCatL	Human Cathepsin L
hCatS	Human Cathepsin S
hCats	Human Cathepsins
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC_{50}	Half-Maximal Inhibitory Concentration
ID	Sequence Identity
InhA	Enoyl Reductase of *Mycobacterium tuberculosis*
K_a	Acid Dissociation Constant
K_i	Inhibition Constant
K_M	Michaelis-Menten Constant
K_s	Dissociation Constant of the Enzyme-Substrate complex
LINCS	Linear Constraint Solver
MD	Molecular Dynamics
MM-GBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
NADH	Reduced Nicotinamide Adenine Dinucleotide
Nit	Core structure of nitriles shown in Fig. 1.5, to which the variable P2 substituent (*X*) is bonded.
NPT	Isothermal-Isobaric Ensemble
N_{pyr}	Pyridine nitrogen
NVT	Isothermal-Isochoric Ensemble
PBS	Phosphate-Buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
PDB	Protein Data Bank
PfENR	*Plasmodium falciparum* Enoyl Reductase
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
POVME	Pocket Volume Measurer
Pyr	Pyridine
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation

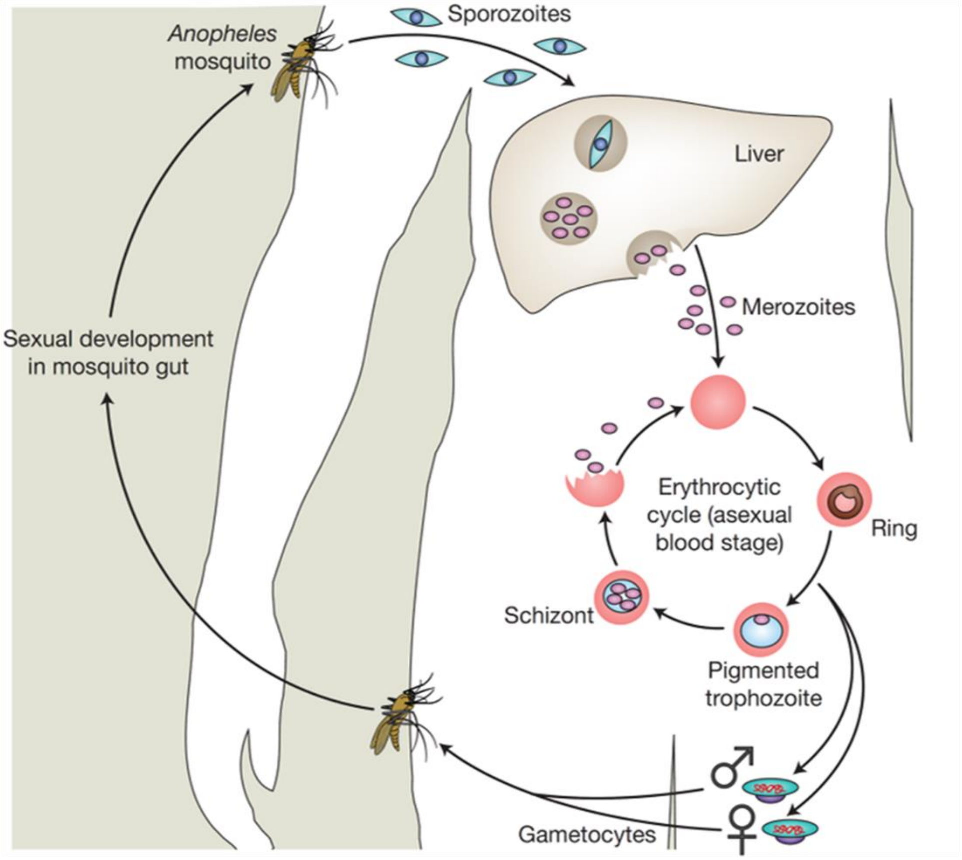
RPMI	Roswell Park Memorial Institute
SASA	Solvent Accessible Surface Area
SBVS	Structure-Based Virtual Screening
SEM	Standard Error of the Mean
S_{vina}	Autodock-Vina Energy Score.
TI	Thermodynamic Integration
US	Umbrella Sampling
WHO	World Health Organization
WISP	Weighted Implementation of Suboptimal Paths
α	Ratio of the dissociation constant between the enzyme:inhibitor complex and the substrate to that of the free enzyme and the substrate
ΔG	Gibbs Free Energy
ΔG°_{bind}	Absolute Binding Free Energy with correction of the standard state.
ΔG_{bind}	Binding Free Energy calculated through MM-GBSA
ΔG_{eff}	Effective Free Energy
ΔG_{R}	Free Energy of releasing the *xy* orthogonal restraints in the bound state
ΔS_{conf}	Configurational Entropy
$\Delta\Delta G$	Relative Free Energy

Table of Contents

1 INTRODUCTION

Malaria is still a major problem of public health. Nearly half of the world's population lives in areas where this disease is endemic, mainly in tropical low-income countries.[1, 2] According to the World Health Organization (WHO), approximately 290 million cases and 435 000 related deaths were reported in 2017.[2] The disease is caused in humans by five species of Plasmodia, i.e, *P. falciparum*, *P. ovale, P. vivax*, *P. malariae* and *P. knowlesi*, the former being responsible for the most lethal form of the disease.[3, 4] Unfortunately, after a period of success in controlling malaria worldwide, the progress has stalled over the last years.[2]

Figure 1.1. Life cycle of *P. falciparum*. The parasites have a complex life cycle involving female mosquitoes of the genus *Anopheles*, where they undergo a sexual development, and human beings, where the asexual blood stage inside the erythrocytes takes place.



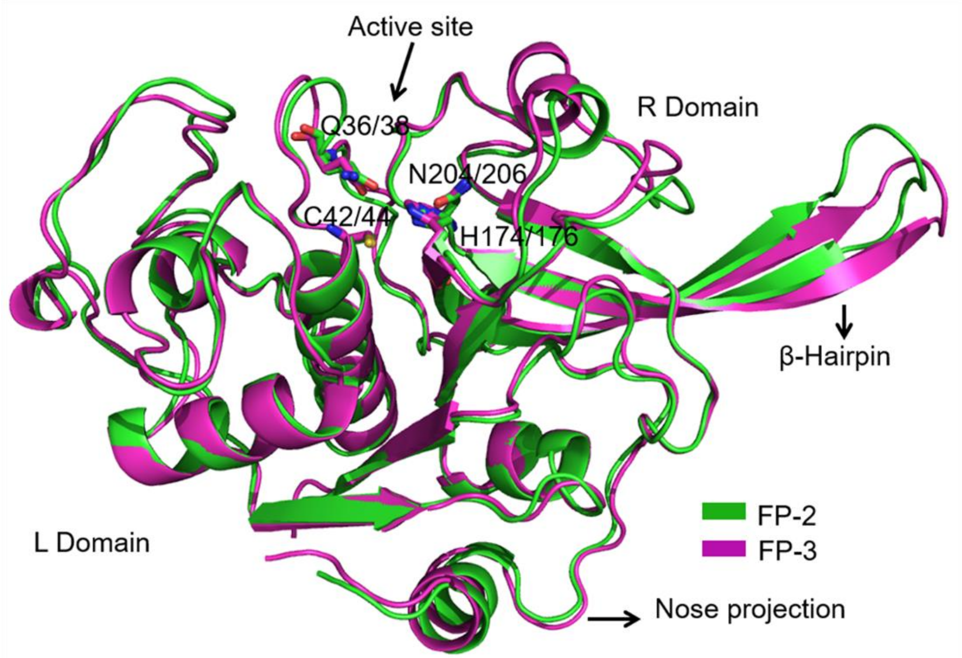
Source: Taken from ref. 5.

The life cycle of malaria parasites is complex and involves a number of different asexual and sexual developmental stages within both the vector and the vertebrate host (Fig. 1.1). In humans the disease is transmitted when bitten by an infected female *Anopheles* mosquito. The sporozoites are then injected into the dermal tissue and actively move into the circulatory system and colonize the liver.[3] Once in the liver, each sporozoite differentiates and divides into thousands of merozoite forms, which are released from hepatocytes into the bloodstream and invade the circulating erythrocytes. Within the new host cell, the parasite undergoes a maturation process from a ring-stage trophozoite to a pigmented trophozoite before finally undergoing mitotic nuclear divisions into daughter merozoites at the schizont stage. At this point, the infected erythrocytes rupture and release daughter merozoites into the bloodstream, which resume another round of asexual reproduction after having infected new erythrocytes.[3] Finally, some parasites differentiate into sexual forms, called gametocytes, which may be ingested by a feeding mosquito and undergo a sexual life cycle within the vector leading to parasite transmission to other human beings.[3]

The efforts to control malaria have been focused on two main approaches, i.e., killing the vector and killing the parasite. The first alternative comprises the extensive use of insecticides and habitat transformations, i.e., use of bed nets. Although this would be the most desirable method, its implementation has been largely unsuccessful due to the vast areas involved and the emergence of insecticide-resistant mosquito strains.[3] Alternatively, human beings can be immunized against malaria parasites. However, none of the developing vaccines is functional yet and there are no prospects for any of them to become available soon. In fact, the most promising vaccine to date, termed RTS,S, has shown, upon completion of phase III clinical trial, that its efficacy is limited and its effect depends on the geographic region.[6] Therefore, drug therapy remains an important approach to treat and prevent this infectious disease.[3, 7]

The discovery of antimalarials has been mostly serendipitous and the mechanism of action is still largely unknown for many of them.[3] Most drugs are directed against blood-stage parasites. Chloroquine and other quinine alkaloids derivatives were extensively used to treat malaria during the 20th century across the world. Antifolate drugs, such as pyrimethamine, an inhibitor of folate metabolism, constitute another group of antimalarials. New drugs based on the artemisinin scaffold have been discovered during the last decades.[3] Currently, however, the emergence of *P. falciparum* strains resistant to the available drugs is becoming a great concern, as it poses a tremendous challenge for disease control in most parts of the world. Therefore, the development of new antimalarials is urgently needed, as well as the concomitant search for new drug targets.[7] Regarding the latter aspect, a great progress has been made since the unveiling of *P. falciparum* genome in 2002.[3] Potential drug targets have been grouped into three main categories, based on their function during the parasite's life cycle: *i*) targets involved in macromolecular and metabolite synthesis, *ii*) targets associated with membrane transport and signaling and *iii*) targets involved in hemoglobin degradation.[3] Of note, process (*iii*) occurs within the food vacuoles of intraerythrocytic malaria parasites, and is carried out by a variety of proteases acting at an optimum pH within the range of 4.5-5.0, i.e., that of the food vacuole.[8] Amino acids derived from hemoglobin degradation are incorporated into parasite proteins or utilized as an energy source. It has also been suggested that removal of hemoglobin during its breakdown provides space in the erythrocyte for parasite growth and prevents early erythrocyte lysis.[8]

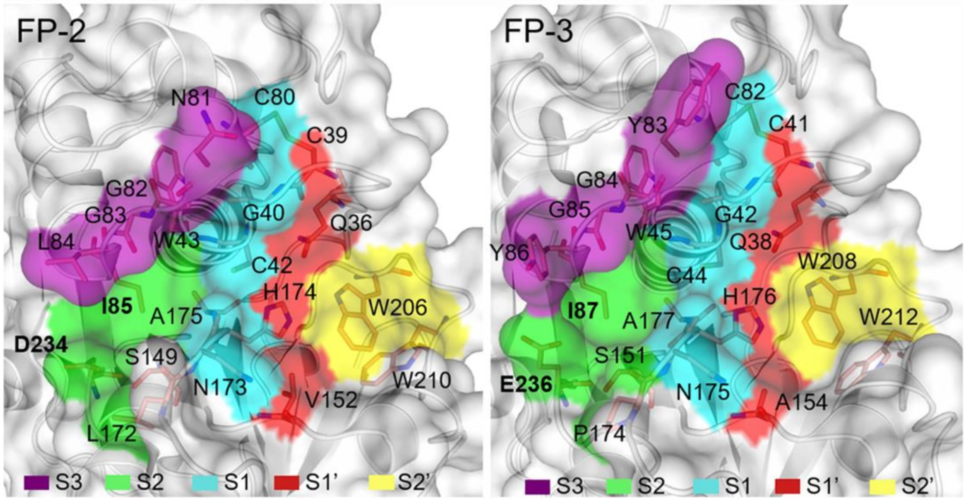
Figure 1.2. Structural features of FP-2 and FP-3. Both enzymes have the typical folding of C1 cysteine proteases, which consists of two domains. Mature FP-2 and FP-3 are made up of single polipeptides of 241 and 243 residues, respectively and share a 68% of sequence identity (ID).[4, 8] The catalytic residues of FP-2/3 are shown as sticks and labeled accordingly. The distinctive β-hairpin and nose-like insertions of both FPs are indicated by arrows. The crystal structures of FP-2 and FP-3 2OUL and 3BWK were taken for representation.



Source: Prepared by the author.

The abrogation of hemoglobin degradation with a variety of protease inhibitors leads to parasite death, thereby indicating the importance of hemoglobin-degrading enzymes as antimalarial drug targets. Among them, the *P. falciparum* C1 cysteine proteases, termed falcipain-2 (FP-2)[9] and falcipain-3 (FP-3)[10], are considered as particularly promising targets for antimalarial drug discovery.[4, 7, 8, 11] Both proteases are synthesized as membrane-bound proforms that are processed to soluble mature forms probably by auto-proteolysis after exiting the endoplasmic reticulum/Golgi network. Besides being associated with hemoglobinase activity, FP-2 and FP-3 are involved in the conversion of proplasmepsins, another group of hemoglobin-degrading aspartic proteases, into their mature active forms. FP-2 is also engaged in the degradation of erythrocyte membrane skeletal proteins at neutral pH, which suggests its involvement in erythrocyte rupture and release of mature merozoites into the bloodstream.[8] The disruption of FP-2 gene leads to trophozoites with swollen, dark-staining food vacuoles, consistent with markedly diminished hemoglobin degradation. However, more mature knockout parasites display normal phenotypes. In contrast to these results, repeated attempts to obtain FP-3 knockout parasites have been unsuccessful, although replacement of the FP-3 gene with a functional copy encoding a tagged protease was readily accomplished.[4, 8, 12] This indicates that FP-3 is essential to intraerythrocytic parasites and, thus, the development of cysteine protease inhibitors as antimalarial drugs should be directed against both FP-2 and FP-3.[8]

Figure 1.3. Active sites of FP-2 and FP-3. The active sites of both enzymes are made up of five subsites, from S3 to S2', which are colored differently according to the legend. The most important residues are shown as sticks and labeled accordingly. The representations correspond to the crystal structures 2OUL and 3BWK of FP-2 and FP-3, respectively.



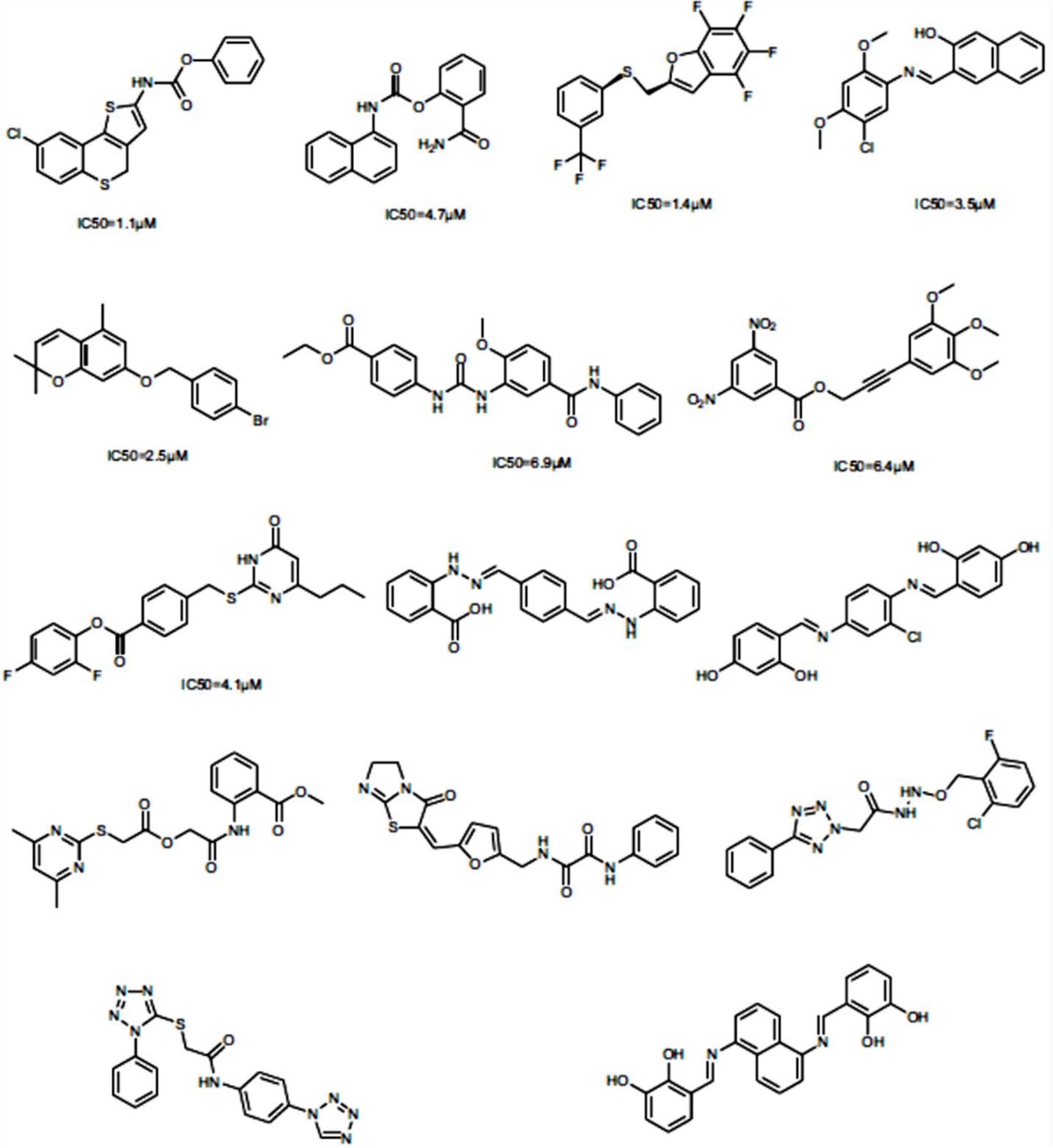
Source: Prepared by the author.

According to the information provided by the available crystal structures, FP-2 and FP-3 adopt the classical papain-like fold, which is divided into L (left) and R (right) domains and the active site is located in a cleft at the junction between them (Fig 1.2). Four conserved catalytic residues, i.e., Q36/38, C42/44, H174/176 and N204/206, are present in the active sites of FP-2/FP-3, respectively (Fig. 1.2).[13] In spite of the common features shared by both FPs with the other papain-like proteases, they possess some unusual structural characteristics, such as an 'arm-like' β-hairpin (near the C-terminus), which extends away from the protein surface, and a 'nose-like' projection (at the N terminus) (Fig.1.2). The arm-like structure may act as a distal site for hemoglobin binding, as suggested by mutagenesis studies carried out with FP-2.[14] It also displays a great flexibility due to solvent exposure as shown by the analysis of different FP-2 crystal structures.[8] On the other hand, the nose-like projection is essential to proper folding according to deletion analysis of its sequence in FP-2.[15] Because of the great structural similarity of FP-2 and FP-3, it is likely for the nose-like projection to play an analogous role in the latter.

The active sites of FPs are formed by five subsites: S3, S2, S1, S1’ and S2’ (Fig. 1.3). The substrate-specificity requirements of the different subsites of FP-2 and FP-3 have been already studied and, despite the sequence similarity of both FPs, some differences in substrate specificity were observed.[16] The S1 subsite of FP-2 preferentially accommodates Arg residues at the P1 site, while that of FP-3 prefers Leu. The S2 subsites of both FPs have strict preference for Leu at P2. The S3 subsite of FP2 is rather promiscuous, while Ala at P3 is preferred by the S3 subsite of FP-3. The S1’ subsite of FP-2 preferentially accommodates hydrophobic residues at P1’ and that of FP-3 shows specificity for Arg residues at this site. The specificity of a fifth subsite, S2’, was also assessed but it demonstrated no clear preference for any particular residue. These results can contribute to devise drug design strategies targeting both FPs.[16]

Over the past two decades, dozens of works have reported the discovery of FP-2 (and FP-3) inhibitors, some of them active against *P. falciparum* cultures.[7, 8, 11, 17] However, none of the existing candidates has entered clinical trials yet, and the search for novel inhibitors continues.[7] In general, the FP inhibitors can be grouped in three main categories according to their chemical structures: *i*) peptide-based, *ii*) peptidomimetics and *iii*) non-peptidic inhibitors.[4, 7, 8, 11] Remarkably, peptide-based inhibitors may display high affinities for FPs, as well as the capacity to inhibit the growth of malaria parasite cultures at sub-micromolar and even at nanomolar concentrations.[7, 8] In addition some of them have shown antimalarial effects *in vivo* using *P. vickei*-infected mice.[4] Despite the previous achievements, the use of peptide-based inhibitors in therapy is limited due to their susceptibility to protease degradation, low capacity to penetrate biological membranes and immunogenicity.[8, 18] Therefore, in an attempt to overcome the above drawbacks of peptide-based inhibitors, the search for other inhibitor classes, e.g., peptidomimetic and nonpeptidic compounds, has been conducted.[7, 8, 11] In this sense, structured-based virtual screening (SBVS) of compound databases has been successfully employed to identify many FP-2 inhibitors, especially of non-peptidic nature (Fig. 1.4).[7] This underscores the potential of computational approaches to discover new antimalarials.

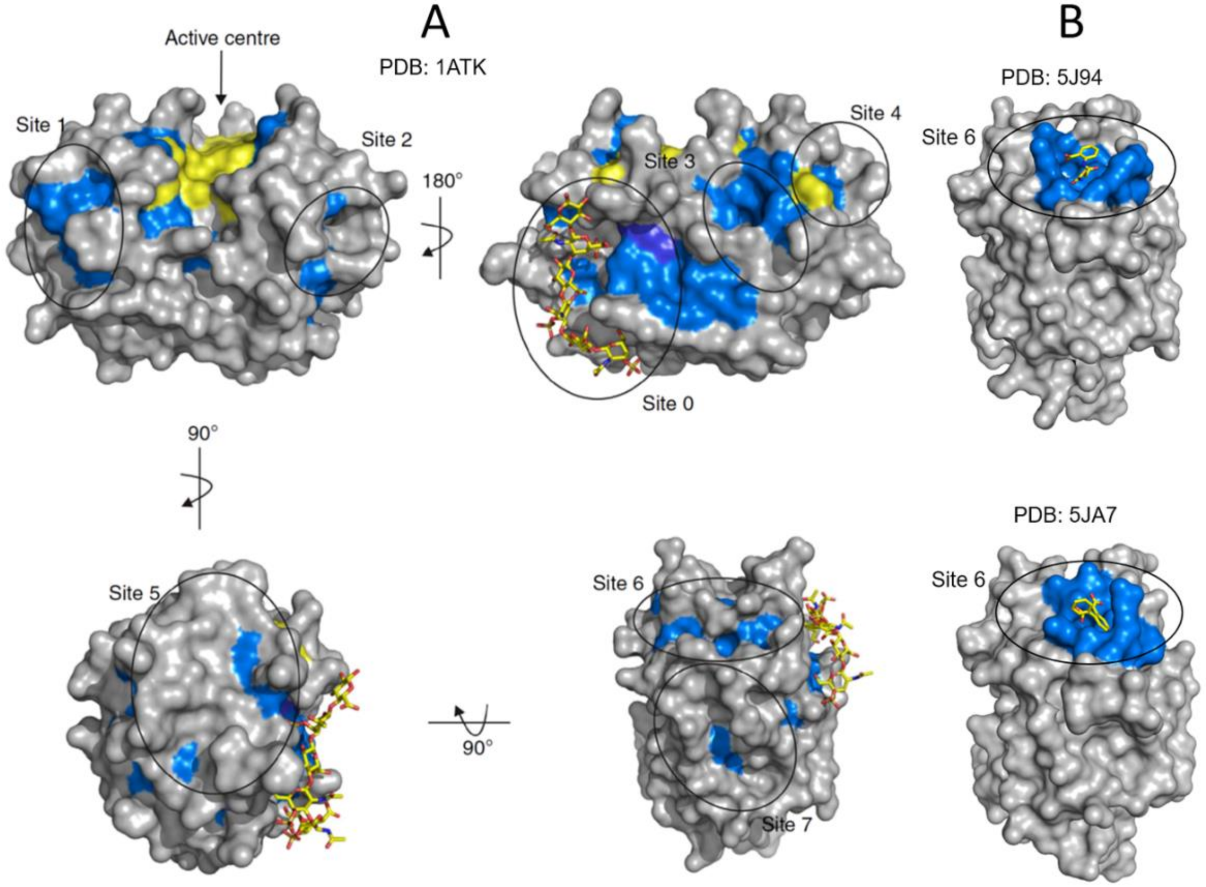
Figure 1.4. Some examples of non-peptidic FP-2 inhibitors identified through SBVS directed toowards the orthosteric site. The half maximal concentration (*IC₅₀*) values against the enzyme are provided when available.



Source: Taken from ref. 7.

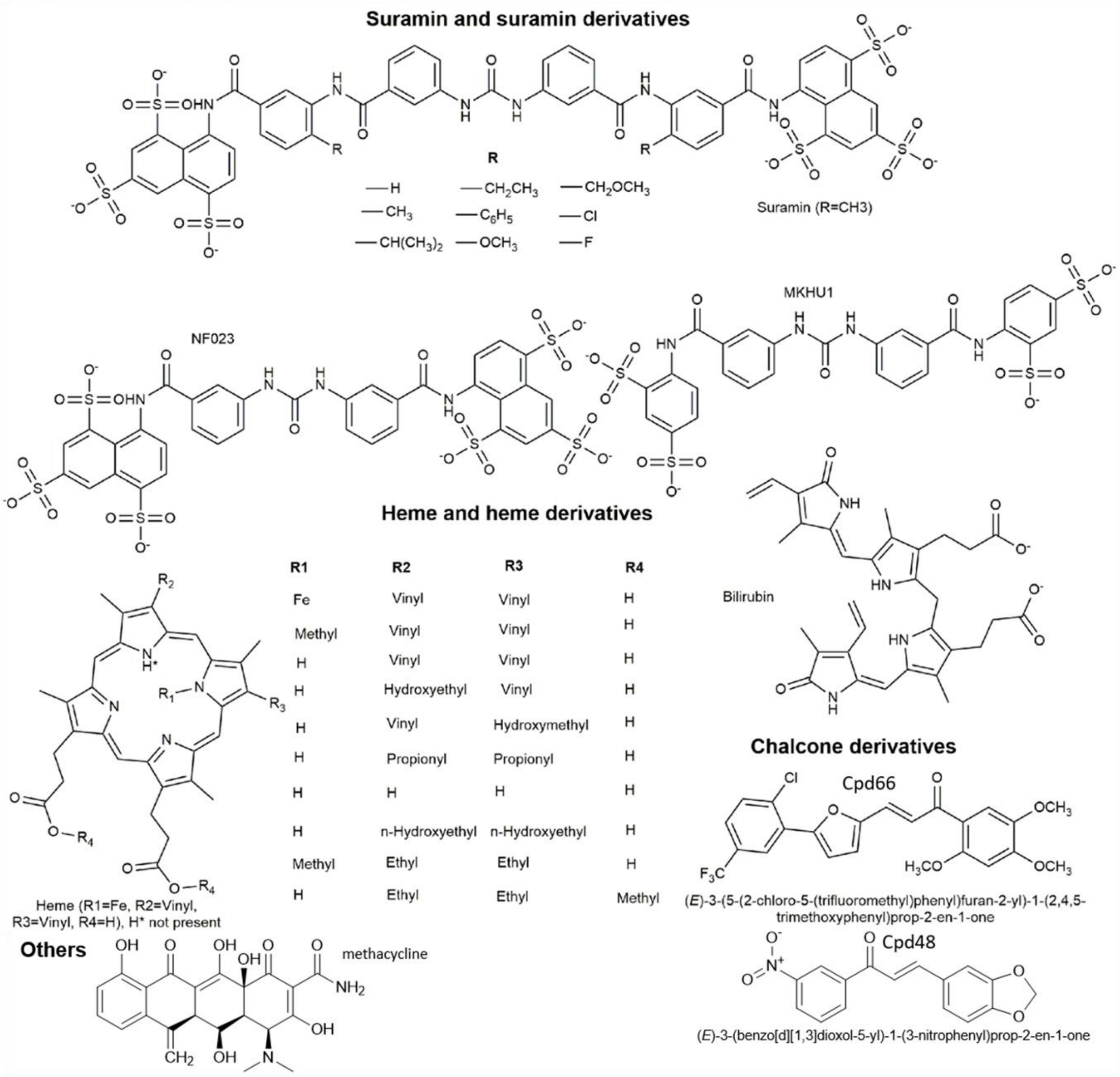
One of the main concerns regarding the identification of drugs against cysteine proteases is the likely occurrence of off-target effects due to lack of selectivity.[19] A promising way to tackle selectivity issues of molecules binding the orthosteric sites, i.e., the primary functional sites, of target proteins is the search for allosteric inhibitors, i.e., molecules that reduce the functional activity by binding to regions different from the orthosteric site.[20-23]. In fact, allosteric inhibitors typically bind to protein regions that are less subjected to evolutionary pressure; which, in turn, increases the sequence divergence in the allosteric sites of proteins of the same family if compared to that of the orthosteric sites.[23] Interestingly, allostery has been already reported for various papain-like proteases, such as human cathepsins K, B and S (hCatK, hCatB and hCatS, repectively), cathepsin L-like proteases of *Leishmania mexicana*, *Trypanosoma cruzi* and *Trypanosoma brucei*, papain and FP-2 itself.[24-38] Therefore, the search for allosteric inhibitors against these proteases is supported by abundant experimental findings.

Figure 1.5. Predicted and experimentally-confirmed allosteric sites in hCatK. A) Allosteric sites on the surface of hCatK predicted through statistical coupling analysis. Sector and well-conserved residues are depicted in blue and yellow, respectively, the remaining residues are shown in gray. The structure of chondroitin sulfate, an already-known allosteric modulator of hCatK (PDB: 3C9E) was superimposed on the protein structure. **B)** Representation of the crystal structures of the allosteric inhibitors NSC13345 (up) and NSC94914 (down). Note that both inhibitors are bound to site6 of hCatK (residues depicted in blue). Allosteric sites have been numbered for convenience.



Source: Fig. 1.5-A adopted from ref. 25 and Fig. 1.5-B prepared by the author.

Figure 1.6. Allosteric inhibitors of FP-2 reported in literature. The reported inhibitors can be grouped in four main scaffolds shown in the figure. Suramin and heme trigger an excess-substrate inhibition. The compounds belonging to the remaining scaffolds act as noncompetitive inhibitors.

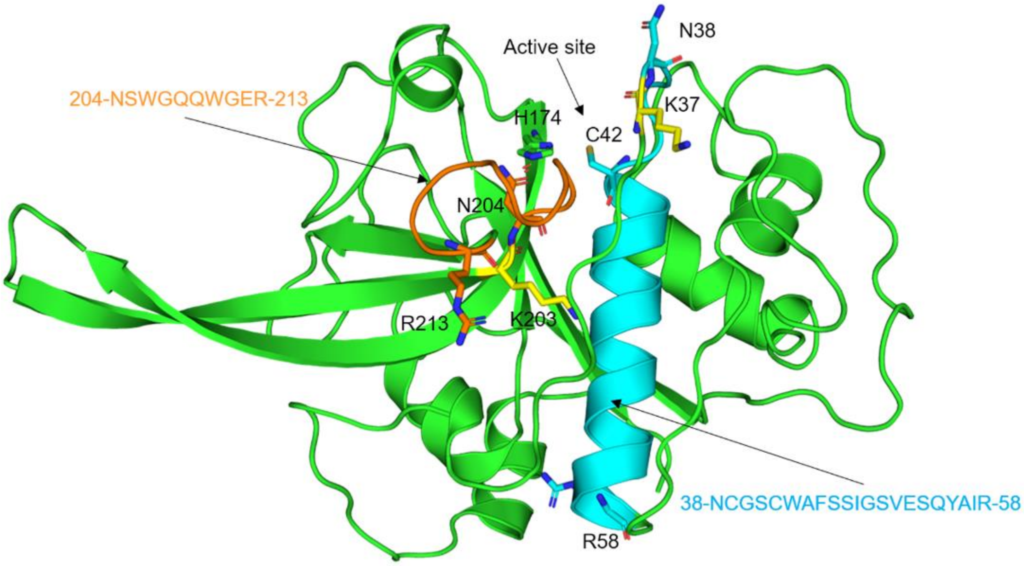


R1	R2	R3	R4
Fe	Vinyl	Vinyl	H
Methyl	Vinyl	Vinyl	H
H	Vinyl	Vinyl	H
H	Hydroxyethyl	Vinyl	H
H	Vinyl	Hydroxymethyl	H
H	Propionyl	Propionyl	H
H	H	H	H
H	n-Hydroxyethyl	n-Hydroxyethyl	H
Methyl	Ethyl	Ethyl	H
H	Ethyl	Ethyl	Methyl

Source: Prepared by the author.

Remarkably, a previous work has proposed potential allosteric cavities in hCatK by analyzing co-evolving residues in multiple sequence alignment of the C1 cysteine proteases (Fig. 1.5-A).[25] The authors then reported several non-competitive inhibitors of hCatK discovered through SBVS directed against the promising cavities, including two with available co-crystal structures targeting the so-called site 6 (Fig. 1.5-B).[25-27] This information is relevant for the identification of other allosteric inhibitors against cysteine proteases by means of structure-based computational approaches.

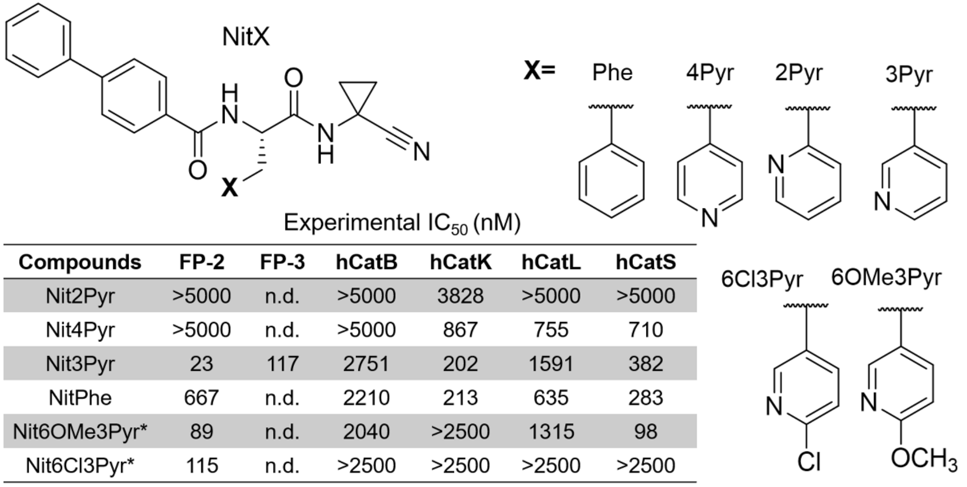
Figure 1.7. FP-2 tryptic peptides displaying differences in their hydrolysis rates in the presence of either a competitive or a noncompetitive inhibitor. The tryptic peptide shown in blue is protected from trypsin degradation in the presence of an orthosteric ligand, whereas the one depicted in orange becomes protected when the noncompetitive inhibitor Cpd66 binds the enzyme.[36] Trypsin cleavage sites are highlighted by depicting the residues forming the scissible peptide bond as sticks. In addition, the residues of the catalytic diad, C42 and H174, are represented as sticks. The primary sequences of the tryptic peptides are also included in the figure.



Source: Prepared by the author.

The currently known inhibitors of FP-2 can be classified in four scaffolds: *i*) suramin and derivatives, *ii*) heme and derivatives, *iii*) *E*-chalcones and *iv*) methacycline (Fig. 1.6).[34-36, 38] Heme and suramin are able to modulate the enzyme through an excess-substrate inhibition, which involves the interaction of a substrate molecule with a secondary binding site.[34, 35] The two *E*-chalcone derivatives termed Cpd48 and Cpd66 were reported to trigger a classical and mixed non-competitive inhibition against FP-2, respectively.[36] Cpd66 failed to protect from hydrolysis a peptide associated to the FP-2 orthosteric site, as determined by comparing the peptide mass fingerprints generated through trypsin digestion of the free enzyme and when bound to the inhibitor. However, the compound decreased in a concentration-dependent manner the generation of a tryptic peptide unrelated to the orthosteric binding site and comprising residues from N204 to R213 (loop$_{204-213}$, Fig. 1.7). These results were interpreted as a supporting evidence of the non-competitive inhibition mechanism of the compound.[36] A valuable clue to locate the FP-2 allosteric pocket that binds Cpd66 emerged from the previous experiment, which can be further exploited to propose a novel site for the modulation of the enzyme's catalytic activity. On the other hand, Cpd48 was co-crystalized with FP-2 and the solved structure indicates that it binds to a region comprising the prime subsites of the enzyme, which seems to be consistent with the mixed inhibition mechanism.[39] These findings reinforce the suitability of exploring the allosteric modulation of FP-2 in order to design novel antimalarials.

Figure 1.8. Chemical structures and experimental *IC_{50}* values of the studied compounds against FPs and hCats. Compounds marked with asterisks were tested as racemic mixtures in the inhibition assays instead of pure *S* enantiomers. Therefore, the actual *IC_{50}* values for these compounds are half of those reported by Nizi *et al.*, as the *R* enantiomers are inactive. N.d. stands for non-determined values. For all compounds, except for NitPhe, the values shown in the table correspond to apparent *IC_{50}'s*, since they undergo ionization in water. All experimental values were taken from the work by Nizi *et al.*[40]



Compounds	FP-2	FP-3	hCatB	hCatK	hCatL	hCatS
Nit2Pyr	>5000	n.d.	>5000	3828	>5000	>5000
Nit4Pyr	>5000	n.d.	>5000	867	755	710
Nit3Pyr	23	117	2751	202	1591	382
NitPhe	667	n.d.	2210	213	635	283
Nit6OMe3Pyr*	89	n.d.	2040	>2500	1315	98
Nit6Cl3Pyr*	115	n.d.	>2500	>2500	>2500	>2500

Source: Prepared by the author.

Despite the promising results that can be expected from allosteric compounds, their inhibitory activity is relatively modest if compared to that of orthosteric inhibitors. Instead, recent literature has drawn the attention to reversible covalent inhibitors of these enzymes, i.e., small molecules forming reversible covalent bonds with the catalytic Cys, as a good balance between affinity and selectivity can be reached by fine-tuning non-covalent interactions with the target and warhead reactivity.[19] Some studies have addressed this issue and have reported promising reversible covalent inhibitors of FP-2.[40-43] In this regard, a work by Nizi *et al.* is particularly relevant, as it demonstrates that it is possible to achieve selectivity against human cysteine cathepsins B, K, L and S (hCatB, hCatK, hCatL and hCatS, respectively) by focusing on the P2 position of peptidomimetic nitriles, i.e., peptide analogs bearing the cyano (-C≡N) warhead.[40] Interestingly, the authors reported noticeable changes in the activity and selectivity profiles of isomeric compounds with pyridine moieties at P2, in which the aromatic N atom (N_{pyr}) lies at different positions (Fig. 1.8).

Among the compounds with different P2 substituents tested by Nizi *et al.*, it was Nit3Pyr (*Nit* stands for the nitrile core structure and *3Pyr*, for the P2-pyridine with the N_{pyr} atom at position 3) the one showing the highest affinity for FP-2 (IC_{50}=23 nM) and selectivity against hCats (9 to 120-fold more active).[40] On the other hand, isomers Nit2Pyr and Nit4Pyr displayed no inhibitory activity against FP-2 (Fig. 1.8) up to 5 μM; whereas other Nit3Pyr derivatives, e.g., Nit6Cl3Pyr and Nit6OMe3Pyr, though still active, did not improve the affinity and selectivity of the parent compound (Fig. 1.8). Moreover, Nit3Pyr inhibits FP-3 (IC_{50}=117 nM) and outperformed the inhibition against FP-2 of compounds with structurally-similar yet more hydrophobic P2 moieties, such as NitPhe (IC_{50}=667 nM, Fig. 1.8).[40] Compound Nit3Pyr was further redesigned at P3 position, which led to better affinity and selectivity for FP-2, micromolar inhibition of *P. falciparum* growth *in vitro* and non-detectable cytotoxicity against human cell lines.[40] Despite these promising findings, currently there is no detailed structural knowledge on the complexes formed by the nitriles and proteases mentioned above. This lack of information precludes the rational optimization of P2-pyridine moieties and the structure-based design of novel inhibitors mimicking the Nit3Pyr properties.

On the basis of all the previous information, it becomes clear that the search for FP-2 inhibitors constitutes an active and relevant field. In spite of the current interest in this enzyme, there are some aspects requiring further study to foster the discovery of more selective FP-2 inhibitors. Therefore, in this work we intend to achieve the following general objective:

To propose new inhibitors, druggable allosteric sites and molecular determinants for selective orthosteric inhibition of FP-2.

To accomplish this objective, we will conduct the specific objectives shown below:

1. To identify novel competitive inhibitors of FP-2 through a combination of SBVSs and rescoring steps by taking into account the potential interaction with human off-targets, and to assess their inhibitory activity *in vitro* against FP-2, *P. falciparum* cultures, hCatK and HeLa cell-line.
2. To evaluate the suitability of site 6 of FP-2 as a druggable allosteric site by identifying ligands targeting this site through a combination of SBVSs and rescoring steps, and the subsequent determination of their inhibitory activity *in vitro* against FP-2 and their inhibition mechanism.

3. To predict the occurrence of transient allosteric cavities in FP-2 on the basis of experimental information related to the non-competitive inhibitor Cpd66 and the performance of computational analyses to determine the compound's binding mode and the inhibition mechanism at molecular level.
4. To identify the molecular determinants for the selective inhibition against FP-2 of nitriles containing 3-pyridine derivatives at position P2 through MD simulations and free energy calculations in order to enable the discovery of more selective FP-2 inhibitors.

The methodologies and experimental approaches employed here to carry out the previous specific objectives will be presented in the next chapter. Objectives 1 and 2 will be developed in Chapter 3, where four FP-2 inhibitors, two of them discovered through SVBSs against the active site, and the other two, against site 6, are presented. Chapters 4 and 5 show the results of the prediction of a novel allosteric site of FP-2 based on the experimental information related to Cpd66 and the role of water molecules in the selectivity for FP-2 of nitriles containing 3-pyridine derivatives at position P2, respectively. The main conclusions of this work are presented in Chapter 6, and four appendices (from A to D) provide supporting information to the methods and results.

2 MATERIAL AND METHODS

2.1 Computational section

2.1.1 Preparation of the protein structures

The structure of all proteins subjected to docking and MD simulations were retrieved from the Protein Data Bank (PDB). In all cases, the FP-2 and FP-3 PDB structures employed for the aforementioned computational analyses were 2OUL and 3BWK, respectively; whereas, for hCats, the particular structures will be specified in further sections. The protonation states of ionizable residues were predicted at pH 5.5 using the PDB2PQR webserver.[44, 45] Missing residues were added when necessary with Modeller v9.15.[46, 47]

2.1.2 Structure-based virtual screenings and rescoring steps

Given a small molecular weight compound and a target (usually a protein) of interest, docking is the computational answer to two basic questions: (1) What is the protein-bound conformation of the ligand? (2) What is the relative orientation of the ligand with respect to the target? This information is described by the proposed binding mode (pose) of the ligand for this particular target, illustrated by the Cartesian coordinates of the protein-bound ligand. To answer both questions, any molecular docking engine consists of two parts: a conformational search algorithm to sample all degrees of freedom (translational, rotational, and conformational) of the ligand, and a scoring function to rank putative poses for a compound and/or compounds from a large library.[48]

A typical docking flowchart is made up of four invariant steps whose importance is critical to the outcome of SBVSs, i.e., *i)* the ligand setup, *ii*) the protein setup, *iii*) the docking step and *iv*) the post-docking analysis. The first step usually involves the conversion of ligand libraries from two-dimensional (2D) formats to 3D formats and the usage of filters to enrich the library with potential drug-like compounds. For protein setup, it is advisable to start from a high-resolution X-ray structure, although homology models with a strong experimental support may also be used. It is convenient as well to select an ensemble of structures reflecting the conformational flexibility of the binding site. Missing hydrogen bonds must be added according to the protonation states of ionizable residues and tautomeric forms of His residues. Moreover, it is generally recommended to focus docking to a protein cavity of interest (catalytic site, inhibitor, or allosteric site) to limit the 3D space to sample. However, blind docking simulations, which scan the entire protein surface searching for potential binding sites, are also possible though less accurate. After protein and ligand setups, a docking algorithm has to be chosen to predict the binding modes and to rank the ligands according to their affinity for the target protein. Finally, the post-docking step is required to more accurately retrieve true positives from the virtual screening output. Many strategies are possible. The simplest consists in rescoring poses with additional scoring functions or more accurate methods such as Molecular Mechanics Generalized Born (Poisson Boltzmann) Surface Area (MM-GB(PB)SA) or Linear Interaction Energy (LIE). Topological filters can also be used to filter out poses exhibiting steric or electrostatic mismatches between the ligand and its target. Poses can also be minimized by a third-party software, clustered, and analyzed by machine learning methods.[48]

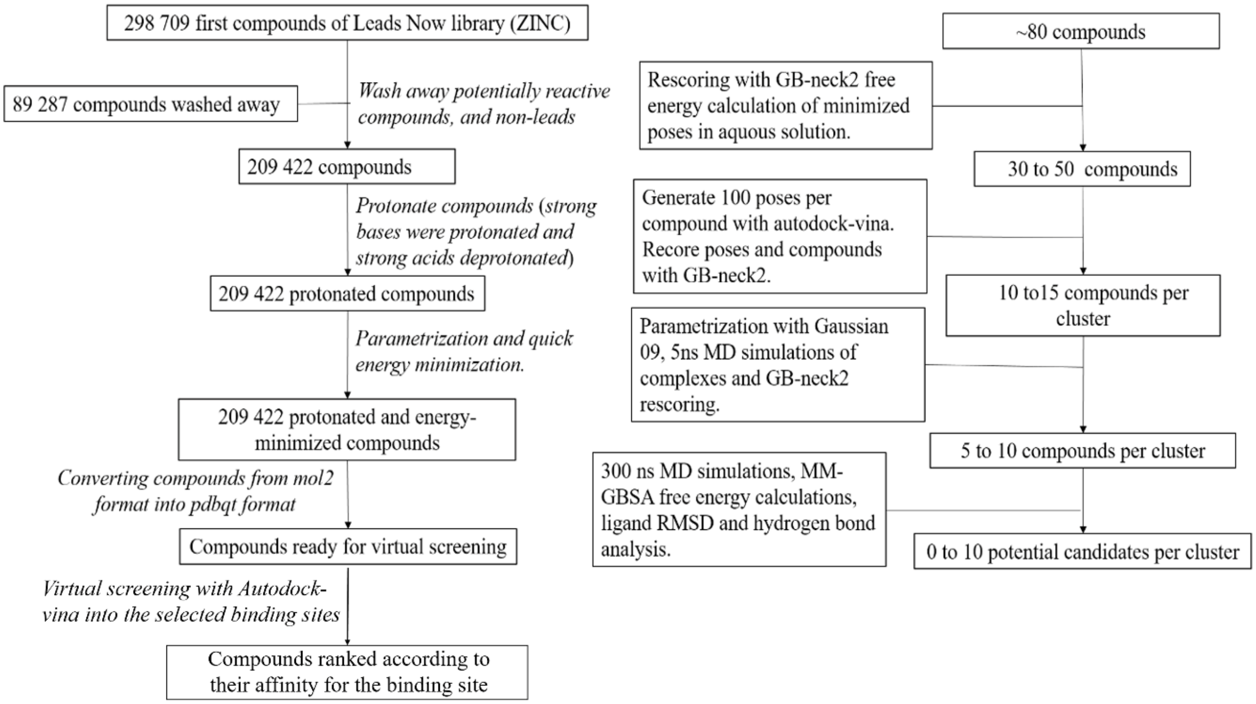
SBVSs of the Maybridge HitFinder^{TM} database, containing 14 400 drug-like compounds (http://www.maybridge.com/), were performed in parallel against of FP-2, FP-3 and hCatK using Autodock Vina 1.1.2.[49, 50] The FP-2, FP-3 and hCatK (PDB: 1ATK) crystal structures and the ligands from the database were prepared for the SBVSs with Autodock tools[51] and a box enclosing the active site of each enzyme was created in order to conduct the docking simulations.[50]

The results of the SBVSs were processed to select compounds fulfilling the following criteria: *i*) high affinity for both FPs (S_{vina}≤-8.0 kcal/mol) and *ii*) relative binding free energy ($\Delta\Delta G$) for at least one FP with respect to hCatK less than -1.0 kcal/mol ($\Delta S_{vina}=S_{vina}(\text{FP})-S_{vina}(\text{hCatK})<-1.0$ kcal/mol). The second condition sought to enhance the selectivity of the putative inhibitors towards the parasitic targets, thereby potentially reducing cross-inhibition of the human off-target. The compounds selected after the previous steps were purchased and subjected to experimental inhibition assays (see Section 2.2 for details).

The binding modes of the experimentally-identified best hits to FP-2, FP-3 and hCatK were predicted by conducting five independent random docking simulations with Autodock-vina, each generating 20 poses into the FP-2 active site.[52] Several docking poses, chosen by visual inspection, were then subjected to post-docking refinement steps as detailed below. The rationale behind the pose selection was to cover most of the conformational diversity generated by the docking simulations. Besides, we ensured that all the selected poses occluded, to some extent, the major selectivity pocket (S2 subsite) and the catalytic residues (S1 subsite) of FP-2. The prediction of the binding modes of the best hits to FP-3 and hCatK was focused only on the most favorable pose of each within the FP-2 active site. The initial complex structure was generated by structural superposition. Subsequent steps consisting of a combination of independent accelerated and conventional MD simulations were performed for each complex after appropriate equilibration steps; see Sections 2.1.5 and 2.1.6, and ref.50 for more details.

On the other hand, 209 422 non-reactive compounds from the Leads Now Library of ZINC12 database were prepared for SBVS with Autodock-vina in order to identify allosteric FP-2 inhibitors targeting site 6 (Fig. 2.1). Various site 6 conformations predicted through clustering analyses of molecular dynamics (MD) simulations conducted for free FP-2 were employed for parallel SBVSs. More details on the ZINC12 database and the preparation of the initial compound structures can be found elsewhere.[53] Nine poses per compound into the selected cavities were generated and the ligands were ranked according to the energy values of their respective top-scoring poses. Approximately 80 compounds per site 6 conformation were selected after this step. Then, the selected compounds were rescored based on MM-GBSA effective free energy (ΔG_{eff}) values, using the GB-neck2 implicit solvation model,[54] after subjecting each of the nine poses per the compound to energy minimization (EM) into a solvation box. The charges of the compounds were calculated with the AM1-BCC protocol available in antechamber program of Amber14 and the remaining parameters were taken from general amber force-field (gaff).[55]. Then, 30 to 50 compounds per site 6 conformation were selected and each of them was subjected to the generation of 100 poses with Autodock-vina in 10 independent docking simulations. Each of the 100 poses per compound was rescored based on the ΔG_{eff} values after EM in water. Subsequently, the 10 to 15 top-scoring compounds were chosen and subjected to 5 ns MD simulations after appropriate equilibration steps (see section 2.1.5). The charges of the compounds prior to the former MD simulations were re-calculated with Gaussian 09,[56] by following a protocol explained hereinafter (Section 2.1.4). The complexes having the most favorable ΔG_{eff} values during their respective 5 ns MD simulations were subjected to subsequent 300 ns MD simulation per selected compound. Those showing favorable ΔG_{eff} values and ligand Root Mean Square Deviation (RMSD) stability during the 300 ns MD simulation, in addition to prevalent hydrogen bonds (H-bonds) with the protein's residues, were proposed as potential allosteric inhibitors. Finally, the selected compounds were purchased form Enamine Ltd. company for further experimental evaluation, as explained in Sections 2.2.5 to 2.2.8.

Figure 2.1. Workflow employed for the identification of allosteric inhibitors against FP-2. The three initial steps sought to create a library of non-reactive compounds with energy-minimized structures ready for SBVSs and were carried out with the Molecular Operating Environment (MOE) platform.[57] The remaining steps involved the SBVSs and the refinement of docking poses through EM, MM-GBSA free energy calculations, AM1-BCC and quantum-mechanical parametrizations, and MD simulations (see main text for more details of each step).



Source: Prepared by the author.

2.1.3 Prediction of the structures of FP-2 in complex with Cpd66 and the studied nitriles

The prediction of the FP-2:Cpd66 complex structure was conducted as explained elsewhere.[58] Briefly, Cpd66 was docked using Autodock-vina v.1.1.2 into the central structure of the transient pocket occurring in the FP-2 site 3 region.[58] The poses generated in multiple parallel docking simulations were individually rescored by calculating the ΔG_{eff} values after appropriate EM.[59, 60] The pose corresponding to the lowest ΔG_{eff} value was then chosen for subsequent MD simulations. Further analyses required the presence of a peptide in the enzyme's active site (see ref. 58 for details).

The starting structures of the studied protease:nitrile complexes were obtained by transforming the ligand present in the crystal structure of hCatS bound to a similar nitrile (PDB 2FQ9) into each of the inhibitors of interest using *Avogadro*,[61] and by superimposing all other proteins to hCatS contained in the template co-crystal structure. The structures of FP-2, FP-3, hCatB, hCatK, hCatL and hCatS were retrieved from the PDB (2OUL, 3BWK, 1GMY, 4DMY, 5MAE and 2FQ9, respectively) and were prepared as mentioned in Section 2.1.1. Short EMs with the universal force-field (UFF) were carried out in *Avogadro* after ligand transformations in order to obtain suitable starting structures of the inhibitors.[61]

2.1.4 Compound parametrization for MD simulations

The 3D structures of the compounds of interest were minimized using the UFF of *Avogadro*.[61] Subsequently, the energy-minimized compounds were subjected to geometry optimization at B3LYP/6-31G(d,p) level using Gaussian 09 package.[56] The electrostatic potential (ESP) of each optimized structure was then generated by single-point calculations using the HF/6-31G(d) method and Merz-Kollman scheme. The partial atomic charges were fitted to the ESPs through the Restricted Electrostatic Potential (RESP) fit implemented in the program antechamber of Amber package (versions 14, 16 and 18 were used).[55, 62, 63] The remaining bonded and nonbonded parameters for the compounds identified from SBVSs and selected to conduct MD simulations, plus the allosteric inhibitor of FP-2, Cpd66, were assigned using gaff.[64]

The parametrization of nitriles covalently bonded to FP-2 followed a similar parametrization workflow. However, as the study comprising these inhibitors was completed recently, Amber18 was employed to conduct RESP charge derivations and parameters from gaff2 were assigned.[63] In addition, the charges obtained through RESP fits underwent certain modifications, so that all equivalent atoms across the studied compounds outside the thermodynamic integration (TI) regions had the same charge and those involved in alchemical transformations borne a 0 or +1 net charge (see Appendix A, Text A1 and figures cited therein for details).

2.1.5 Conventional MD Simulations

MD simulations constitute a powerful tool to study *in silico* the properties and structure-function relationships of macromolecules and, especially, of proteins. The prediction of protein 3D structures, the study of the interactions at the interfaces of protein-ligand complexes, the determination of macromolecule stabilities in different solvents, the prediction of protein folding pathways and the assessment of receptor-ligand affinities, etc., are just some applications of MD simulations.[65]

Essentially, an MD simulation is a computational experiment that allows to predict the time evolution of a system made up of a relatively-small number of particles by assessing the interacting forces among the particles of the system. This method requires solving the Newton's equation of motion for every particle in the system based on force-field equations, which evaluate the potential energy of the system.[65]

In general, the calculation of the total potential energy (E_{tot}) of a system depends on the energy associated with the interaction between covalently-bonded atoms (E_{bonded}) and the energy arisen from non-bonded interactions (E_{nbond}). In turn, E_{bonded} is further decomposed into three summands assessing the energy contributions of covalently-bonded atom pairs ($E_{tension}$), of bond angles (E_{ba}) and of proper and improper dihedral angles ($E_{dihedral}$). Finally, the E_{nbond} term comprises the potential electrostatic (E_{el}) and van der Waals (E_{vw}) energies.[66]

$$E_{tot} = E_{bonded} + E_{nbond} = (E_{tension} + E_{ba} + E_{dihedral}) + (E_{el} + E_{vw}) \tag{2.1}$$

This equation generally takes on the following form, once all summands are substituted by the respective formulae used in their calculation:

$$\begin{aligned} E_{tot}(r^N) = {} & \frac{1}{2}\sum_i k_{i,b}\left(b_i - b_{i,0}\right)^2 + \frac{1}{2}\sum_i k_{i,\theta}\left(\theta_i - \theta_{i,0}\right)^2 + \frac{1}{2}\sum_i k_{i,\xi}\left(\xi_i - \xi_{i,0}\right)^2 \\ & + \sum_i k_{i,\varphi}\left(1 + \cos\left(n\varphi_i - \varphi_{i,0}\right)\right) + \sum_i \sum_{j=i+1} \left(\left[\frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^{6}}{r_{ij}^{6}}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \end{aligned} \tag{2.2}$$

where $E_{tot}(r^N)$ stands for the total potential energy, which is a function of the positions of the N particles in the system. The first three terms in eq. 2.2 represent the harmonic potentials employed in the calculation of $E_{tension}$, E_{ba} and the contribution of improper dihedrals to $E_{dihedral}$, respectively. The fourth term evaluates the energy associated with proper dihedrals, which is modeled as a periodic function. In all cases, $k_{i,x}$ (x=b, θ ,ξ, φ) represents the force constant of each type of interaction and $x_{i,0}$, the equilibrium angles or positions. The fifth and sixth terms in eq. 2.2 stand for the van der Waals and electrostatic interaction energies, respectively. E_{vw} is modeled as the sum of an attractive term arisen from London's dispersion forces and a repulsive term caused mainly by repulsive Pauli-exclusive interactions. Various force fields, e.g., AMBER, GROMOS, CHARMM, OPLS-AA etc., use a Lennard-Jones (LJ) function to represent the van der Waals potential. In this function, C_{ij}^{12} and C_{ij}^{6} are the LJ interaction constants associated with the repulsion and attraction forces between atoms i and j, respectively, lying at a distance r_{ij}. Finally, the sixth term describes the electrostatic interaction energy between non-bonded atom pairs i,j lying at a distance r_{ij} through the well-known Coulomb's law equation, being q_i and q_j the partial charges of atoms i and j, respectively, and ε_0, the dielectric constant in vacuum. The different parameters contained in eq. 2.1 are estimated from experimental data or quantum mechanical calculations and their values may vary from one force field to another.[66]

For E_{vw} calculation, cutoff radii are established in order to reduce the computational time. This is possible since van der Waals interactions display fast sixth-power decay with respect to distance, thus, the energy contribution due to interactions at distances typically larger than 1.0 nm can be neglected. On the other hand, the electrostatic interactions are considered as long-range interactions due to their slow second-power decay and, therefore, they cannot be neglected at long distances. This explains the high computational cost associated with the evaluation of electrostatic interactions. The Particle Mesh Ewald (PME) method is nowadays one of the most popular and efficient algorithms to optimize the calculation of electrostatic energies during MD simulations. During the setup of a typical MD simulation, the macromolecule of interest is embedded in a unit cell filled with water molecules. In addition, periodic boundary conditions (PBCs) are settled in the limits of the unit cell to reduce surface artifacts.[67, 68]

In order to obtain the positions of atoms as a function of time, the Newton's equation of motion can be solved iteratively for every particle of the system:

$$\boldsymbol{F_i} = m_{\boldsymbol{i}}\boldsymbol{a_i} = m_i\frac{d^2\boldsymbol{r_i}}{dt^2} = -\frac{\partial E_{tot}(\boldsymbol{r_i})}{\partial \boldsymbol{r_i}} \tag{2.3}$$

where $\boldsymbol{F_i}$ is the force acting on particle i, m_i is the particle mass, $\boldsymbol{a_i}$ is its acceleration and $\boldsymbol{r_i}$, the position at time t. Eq. 2.3 can be solved by numerical integration employing different approaches, such as Verlet and leap-frog algorithms. Besides the initial positions, eq.2. 3 requires knowing the initial velocities of all atoms, which are generally assign at random and obeying a Maxwell-Boltzmann distribution.[67, 68]

The integration of Newton's equation of motion produces the microcanonical ensemble (NVE), in which the number of particles (N), the volume (V) and the total energy (E) of the system are conserved. On the other hand, in real life most experiments are typically conducted at constant temperature and/or pressure. Therefore, it is necessary to perform MD simulations in ensembles other than the NVE one, such as the canonical (NVT) and the isobaric-isothermal (NPT) ensembles. This is achieved by coupling thermostats and barostats to the system during the MD simulation.[67]

Thermostats can be broadly divided into global and local. The global ones act instantaneously with the same strength on all particles. Examples of such thermostats are the Nosé–Hoover and the Berendsen weak-coupling method. Local thermostats include the Langevin, Andersen and Dissipative Particle Dynamics methods. On the other hand, for pressure control, some of the most widely-used methods are the Andersen, Parrinello-Rahman and Berendsen barostats.[67]

In this work, all the systems subjected to MD simulations were prepared in a similar fashion using *tleap* of Amber (versions 14, 16 and 18 were used).[55, 62, 63] For the protein:nitrile complexes, the imine C atom of each ligand was covalently bonded to the catalytic Cys S atom using *bond* command of *tleap*.[63] In all cases, protein parameters were drawn from ff14SB force-field,[69] with the ligands being parametrized as described previously (Section 2.1.4). Subsequently, the complexes were embedded in octahedral boxes with edges spanning, at least, 10 Å from the solute surface, and containing TIP3P water molecules and appropriate number of counterions (Na^+ or Cl^-, depending on the solute net charge) using *tleap*. Crystallographic waters close to the proteases were kept during the setup of nitrile complexes. Each system was subjected to two consecutive EMs, followed by heating in the NVT ensemble using a linear temperature gradient from 10 to 298 K and subsequent NPT equilibration at T=298 K and p=1 bar. Each equilibration was performed for 500 ps and all heavy atoms were restrained by means of a 10 $kcal{\cdot}mol^{-1}{\cdot}Å^{-2}$ harmonic constant. Finally, production runs in the NVT ensemble of different lengths were carried out for each system. Replicate MD simulations were also performed when indicated in the result sections by assigning random initial velocities to the system atoms during the heating steps. More details concerning the EM protocols, treatment of nonbonded interactions, temperature and pressure control, timestep, etc., can be found elsewhere.[52] All productive MD simulations were carried out with *pmemd.cuda* (versions available in Amber14, 16 and 18 were employed).[55, 62, 63]

On the other hand, two replicate production runs of 200 ns for free FP-2 were performed with the Gromos 54a8 force-field in Gromcas v.5.1.4.[70, 71] These simulations were carried out in order to assess the impact of a different force-field on the dynamics of the potential allosteric site 6 of FP-2. A slightly different setup was employed in this case, in accordance with the particular features of Gromacs suite. The protein was embedded in a dodecahedral box filled with TIP3P waters and sufficient counter-ions after parametrization with the indicated force-field. The edges of the box were placed, at least, at 1.4 nm from the protein surface. A distance cut-off of 1.4 nm was set for electrostatic and van der Waals non-bonded interactions. PME was used to handle long-range electrostatics. PBCs were employed in all cases. The equation of motion was solved using the leap-frog algorithm with a Δt of 2 fs,[71] while constraining the bond lengths with the Linear Constraint Solver (LINCS).[72] The system was subjected to 5000 steps of steepest descents EM followed by 200 ps NVT heating, during which the temperature was linearly increased from 10 K to 298 K, while keeping the protein's heavy atoms restrained by a harmonic potential (spring constant of 1000 $\mathrm{kJ \cdot mol^{-1} \cdot Å^{-2}}$). Subsequently, a 200 ps NPT equilibration was performed at 298 K and 1 bar. Temperature, in the NVT and NPT equilibrations, and pressure, in the NPT equilibrations, were controlled using the Berendsen thermostat and barostat.[73] During the production runs, the velocity rescaling and Parrinello-Rahman algorithms were used to control temperature and pressure, respectively.[71, 74, 75]

2.1.6 Accelerated MD simulations

For most biomolecules, the energy landscape has multiple minima or potential energy wells separated by high free energy barriers. During typical MD simulations in the nanosecond time scale, the system is usually trapped in one or another local minimum for long periods of simulation time. Consequently, the conformational space of the system is under-sampled and rare events are not simulated. Accelerated molecular dynamics (aMD) is one of the currently-available methods aiming to enhance the conformational sampling in order to overcome the aforementioned problem of normal unbiased MD simulations.[76]

The basic idea behind the aMD simulation is to lower the height of potential energy well barriers to reduce the amount of computational time a given system remains trapped in a local minimum and to speed up the transitions to other local minima. To achieve this, a continuous non-negative bias boost potential function $\Delta V(\mathbf{r})$ is defined, such that when the true potential $V(\mathbf{r})$ is below a certain chosen value E, i.e., the boost energy, the simulation is performed on the modified potential $V^*(\mathbf{r})=V(\mathbf{r})+\Delta V(\mathbf{r})$, represented using dashed lines (Fig. 2.2), and when $V(\mathbf{r})$ is greater than E, the simulation is performed on the true potential $V^*(\mathbf{r})=V(\mathbf{r})$.[76] This leads to an enhanced escape rate for $V^*(\mathbf{r})$. The following step-wise function is employed in aMD simulations:[76]

$$\Delta V^*(\boldsymbol{r}) = \begin{cases} V(\boldsymbol{r})\,, if\ V(\boldsymbol{r}) \geq E \\ \Delta V^*(\boldsymbol{r}) = V(\boldsymbol{r}) + \Delta V(\boldsymbol{r}), \qquad if\ V(\boldsymbol{r}) < E \end{cases} \tag{2.4}$$

The bias potential increases the escape rate of the system from potential wells and the evolution of the system on the modified potential occurs at an accelerated speed with a non-linear time scale of Δt^{*}, where:

$$\Delta t^{*} = \Delta t e^{\beta \Delta V[\boldsymbol{r}(t_i)]} \tag{2.5}$$

In the previous equation, Δt is the time interval of the normal MD simulation and $\beta=1/(k_B T)$, where k_B is the Boltzmann constant and T, the temperature.[76].

The shape of the bias boost potential has been defined through the following equation:[55]

$$\Delta V(\boldsymbol{r}) = \frac{\left(E_P - V(\boldsymbol{r})\right)^2}{\alpha_P + \left(E_P - V(\boldsymbol{r})\right)} + \frac{\left(E_D - V_D(\boldsymbol{r})\right)^2}{\alpha_D + \left(E_D - V_D(\boldsymbol{r})\right)} \tag{2.6}$$

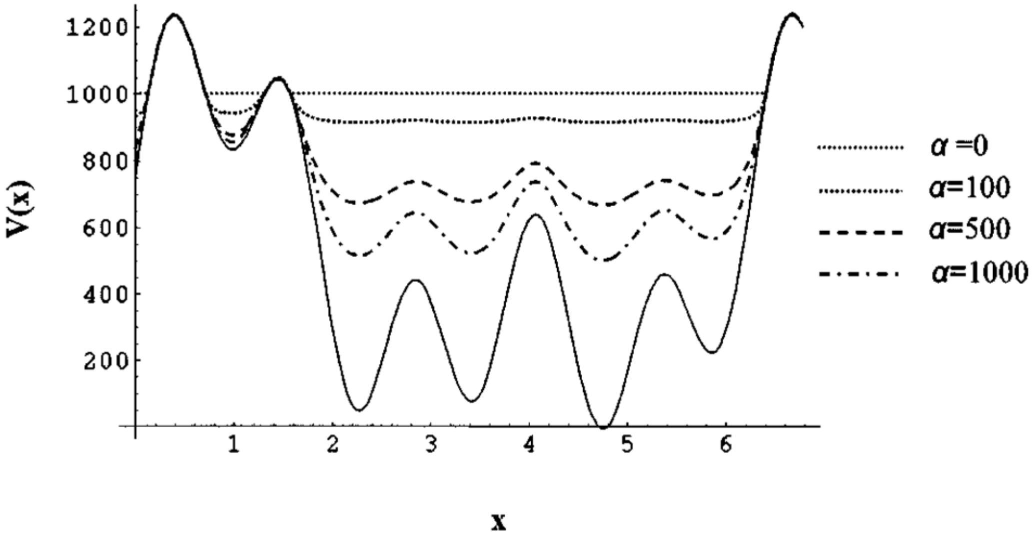
where $V(\boldsymbol{r})$ and $V_D(\boldsymbol{r})$ are the normal potential and the normal dihedral potential, respectively. E_P and E_D are threshold boost potential and dihedral energies, respectively, while α_P and α_D are tuning parameters determining how deep we want the potential energy well to be. The boosting potential preserves the underlying shape of the unbiased potential but the energy barriers are lowered, thus leading to enhanced conformational sampling (Fig. 2.2).[55] An outstanding property of the aMD method derived from the mathematical form of the modified potential is that it can yield, in principle, the unbiased canonical averages of a given observable, upon reweighting procedures, thereby allowing the determination of thermodynamic properties.[76] However, when the reweighting is applied to entire biomolecular systems, the calculation of accurate thermodynamic properties is difficult or practically impossible. Therefore, aMD simulations are more useful to explore the available conformational phase space than to estimate observable values.[77]

In addition, appropriate values for E_P, E_D, α_P and α_D can be obtained by applying empirical formulae, which, in turn, depend on previous knowledge of the system, e.g., the number of residues of the protein and the number of atoms of the whole system, and on other magnitudes easily acquirable by short normal MD simulations, e.g., average potential and dihedral energies.[55]

aMD simulations were employed here to enhance the phase space sampling of best hits from Maybridge HitFinder™ in complex with the target enzymes.[50] The dihedral boosting parameter values were calculated through empirical formulae published elsewhere.[55, 77] No total potential energy boost was applied here. Eleven independent 100 ns aMD simulations were performed for each studied complex using the same sets of boosting parameters. The generated trajectories were then independently clustered as explained in Section 2.1.14, and the central structure of each cluster was subjected to single-structure MM-GBSA free energy calculations after

EM[59, 60] in order to identify the lowest-energy ligand pose generated through aMD simulations. Finally, after the appropriate equilibration, 100 ns MD simulations were carried out for the identified lowest-energy complexes and mean MM-GBSA ΔG_{eff} values were computed.[50]

Figure 2.2: Schematic representation of a hypothetical potential energy function and several bias potentials. A threshold boost energy (E) of 1000 and increasing values of α have been chosen. Note that at α=0 the energy landscape becomes flat and the system undergoes a random walk leading to a low convergence. When progressively increasing the value of α, the basic shape of the bias potential becomes much more alike to that of the unmodified potential, being $\alpha=E-V_{min}=1000$ the optimum value.



Source: Taken from ref. 76.

2.1.7 Constant pH MD simulations

Constant pH MD (CpHMD) simulations constitute an approach that takes advantage of the ability of MD simulations to sample the conformational phase space while sampling at the same time the different protonation states according to a semigrand canonical ensemble.[78] The CpHMD simulations employed in this work were carried out through a process consisting of three repeating steps. Firstly, the MD simulation is performed at constant protonation states, which are set at the beginning of the simulation. The simulation is then halted periodically at fixed τ intervals, the solvent is stripped, the potential is switched to a Generalized Born (GB) model, and N protonation state changes, N being the user-defined whole set of titratable residues, are attempted through Metropolis Monte Carlo (MC) algorithms. If any of the protonation state change attempts were accepted, an equilibration process is conducted while freezing the solute in order to relax the explicit solvent around the new protonation states.[78] Upon completion of the solvent relaxation, the

MD simulation resumes using the velocities of the atoms prior to the relaxation step. From CpHMD simulations, fractions of deprotonation (f_d) can be determined for every titrated residue, and pK_a values can be calculated by conducting these simulations at different pHs. The following equation expressing the relationship between f_d and the pH is employed for the calculation of pK_a's:[78]

$$f_d = \frac{1}{1 + 10^{n(pK_a - pH)}} \tag{2.7}$$

where n is the Hill cooperativity index.

CpHMD simulations in explicit TIP3P water molecules were conducted here to titrate all the ASP, GLU, HIS and reduced CYS residues of FP-2 both in the free state and complexed with either Cpd66 or a peptide using *pmemd.cuda* of Amber16.[62] The respective central structures of the systems calculated from previous MD trajectories were used as starting frames for the CpHMD simulations. The systems were then subjected to EM, heating and NPT equilibration, using the same protocol explained in Section 2.1.5 for conventional MD simulations and without allowing titration. Subsequently, 5 ns CpHMD simulations were carried out at pH 5.5 in the NVT ensemble. More details on the CpHMD simulation setup can be found elsewhere.[58]

In order to calculate titration curves for the selected residues in all systems, we followed the protocol reported by Socher *et al.*,[79] but adapted to CpHMD simulations in explicit solvent.[58] The pH was gradually varied in a step-wise fashion using a ΔpH of 0.1 and covering a 1.0 to 12.0 pH range. Subsequently, 5 ns production runs were carried out for every pH value. To enhance the sampling, the CpHMD simulations were extended to 25 ns within the pH interval from 4.6 to 9.0, which led to better convergence of titration curves, especially of GLU and HIS residues. The fraction of deprotonation of every titrated residue at each pH was then calculated for all systems. Finally, pK_a's and cooperativity indices were calculated by fitting the titration curves. Mathematica v.7.0 was used to perform the non-linear fits, and standard errors were estimated by bootstrapping.[80]

2.1.8 MM-GBSA free energy calculations

MM-GB(PB)SA is an end-point method that allows the calculation of the binding free energy (ΔG_{bind}) or absolute free energies of molecules in solution.[81] Briefly, the ΔG_{bind} for the formation of the complex (RL) from its separate partners, i.e, ligand (L) and receptor (R), can be expressed as follows [81]:

$$\Delta G_{bind} = G(RL) - G(R) - G(L) \tag{2.8}$$

The MM-GB(PB)SA method calculates each $G(X)$ term through this equation:

$$G(X) = E_{MM}(X) + G_{solv}(X) - TS_{conf}(X) \tag{2.9}$$

The molecular mechanics energy term (E_{MM}) of eq. 2.9 lumps together all force-field derived energy components calculated in vacuum, i.e., internal energy (E_{int}), van der Waals energy (E_{vw}) and electrostatic energy (E_{el}), see eq. 2.10.[81]

$$E_{MM}(X) = E_{int}(X) + E_{vw}(X) + E_{el}(X) \tag{2.10}$$

It is worth noting that when using the single trajectory approach, i.e., the trajectories for the free ligand and the free receptor are extracted from that of the complex, the E_{int} terms for RL, R and L cancel in eq. 2.8.[81]

G_{solv} (eq. 2.9) stands for the solvation free energy of the solute molecules and is calculated through implicit-solvation models, e.g., GB and PB models, which typically decompose this term into the sum of the polar solvation ($G_{GB/PB}$) and the non-polar solvation (G_{SASA}) contributions (eq. 2.11).[81]

$$G_{solv}(X) = G_{GB/PB}(X) + G_{SASA}(X) \tag{2.11}$$

G_{GB} is calculated, in turn, through the following equation first proposed by Still *et al.*:[81, 82]

$$G_{GB}(X) = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_w}\right) \sum_{i,j \in X} q_i q_j \left[r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)\right]^{-\frac{1}{2}} \tag{2.12}$$

where ε_w is the dielectric constant of the solvent, e.g., water, r_{ij} is the distance between the solute atoms i and j, q_i and q_j are their partial charges, and R_i and R_j, their effective Born radii. Note that a pair-wise summation is carried out over all the atoms of solute X. Eq. 2.12 has been further modified to incorporate the effects of solvent ions and the solute dielectric constant (ε_{in}) value on G_{GB}. It is also worth saying that the exact evaluation of effective Born radii is the central issue of the GB model and defines its different variants.[55, 81] On the other hand, for PB models, the calculation of G_{PB} requires solving the well-known Poisson-Boltzmann differential equation describing the interaction of solute molecules embedded in a continuum solvent.[55, 81]

G_{SASA} is obtained by eq. 2.13, in which SASA represents the solvent accessible surface area of the solute molecules, while γ and β are empiric constants whose typical values for GB models are 0.0072 kcal·mol^{-1} and 0, respectively.[55]

$$G_{SA}(X) = \gamma SASA(X) + \beta \tag{2.13}$$

Finally, the third term of eq. 2.9 stands for the configurational entropy contribution, which is usually calculated either by normal modes analysis (NMA) or by quasi-harmonic analysis (QHA). When this term is neglected for each species, ΔG_{eff}, instead of ΔG_{bind}, is obtained from eq. 2.8. Of note, ΔG_{eff} frequently suffices for the calculation of relative affinities of similar systems, e.g., a series of ligands binding a given protein. This avoids the calculation of entropy, the most computationally-demanding and less accurate term of the MM-GB(PB)SA method.[55, 81]

Here, ΔG_{eff} values were calculated using the GB-neck (GB*n*) and/or GB*n*2 models available in MMPBSA.py program.[83] These models were selected as they yielded good correlations with experiment results in a previous work studying the interaction of FP-2 with peptide-based inhibitor.[52] The methods were applied to either trajectories or energy-minimized single structures during rescoring steps, as mentioned earlier. The bondi and mbondi3 sets of atomic radii were selected to conduct the MM-GBSA calculations with the GB*n* and GB*n2* models, respectively, and the single trajectory approach was used in all cases.[83] Default values were taken for all other parameters, except for the salt concentration, which was set to 0.1 M to mimic the effect of ionic strength in the free energy calculations.[52]

In addition, the configurational entropy (ΔS_{conf}) was calculated for the interaction of the best hits identified from the SBVSs of Maybridge HitFinder^{TM} database with the target proteases. The normal mode analysis implemented in MMPBSA.py[83] was employed to carried out the previous calculations.[83] More details on the configurational entropy calculations can be found elsewhere.[50]

MM-GBSA free energy calculations were also employed to predict the energy contribution of water bridges in the Nit3Pyr complexes. Firstly, the standard protocol, in which all water molecules are stripped from the trajectory, was applied and a $\Delta G_{eff(nw)}$ value was calculated. Secondly, the Nwat-MMGBSA approach was conducted with in-house scripts.[84, 85] In this case, the water molecule closest to the N_{pyr} atom in each frame was kept as part of the receptor and the remaining waters were stripped. The *closest* command of *cpptraj.cuda* was employed to perform the previous step.[63, 84, 85] Complex and receptor topologies were then created with ante-MMPBSA.py program[63] by keeping a single water molecule as part of the protein. From this procedure, a $\Delta G_{eff(w)}$ value was obtained for every system and the contribution of water bridges ($\Delta\Delta G_{eff(wb)}$) was estimated by subtracting the corresponding $\Delta G_{eff(w)}$ and $\Delta G_{eff(nw)}$ values.

2.1.9 Thermodynamic integration free energy calculations

The free energy variation ($\Delta G_{A\rightarrow B}$) during the transformation of a given system from state A into B can be calculated by connecting the end states through a series intermediate (hybrid) states scaled by a coupling parameter λ, which leads to the following equation:[63]

$$\Delta G_{A\rightarrow B} = \int_0^1 \langle \partial V/\partial \lambda \rangle_\lambda d\lambda = \int_0^1 \langle V_B - V_A \rangle_\lambda d\lambda \tag{2.14}$$

where V_A and V_B are the potential energies of the final states A and B, respectively, and $\langle ... \rangle_\lambda$ is the average of the enclosed magnitude, carried out for every λ intermediate. In turn, the potential energy for a λ state, $V(\lambda)$, can be obtained as follows:[63]

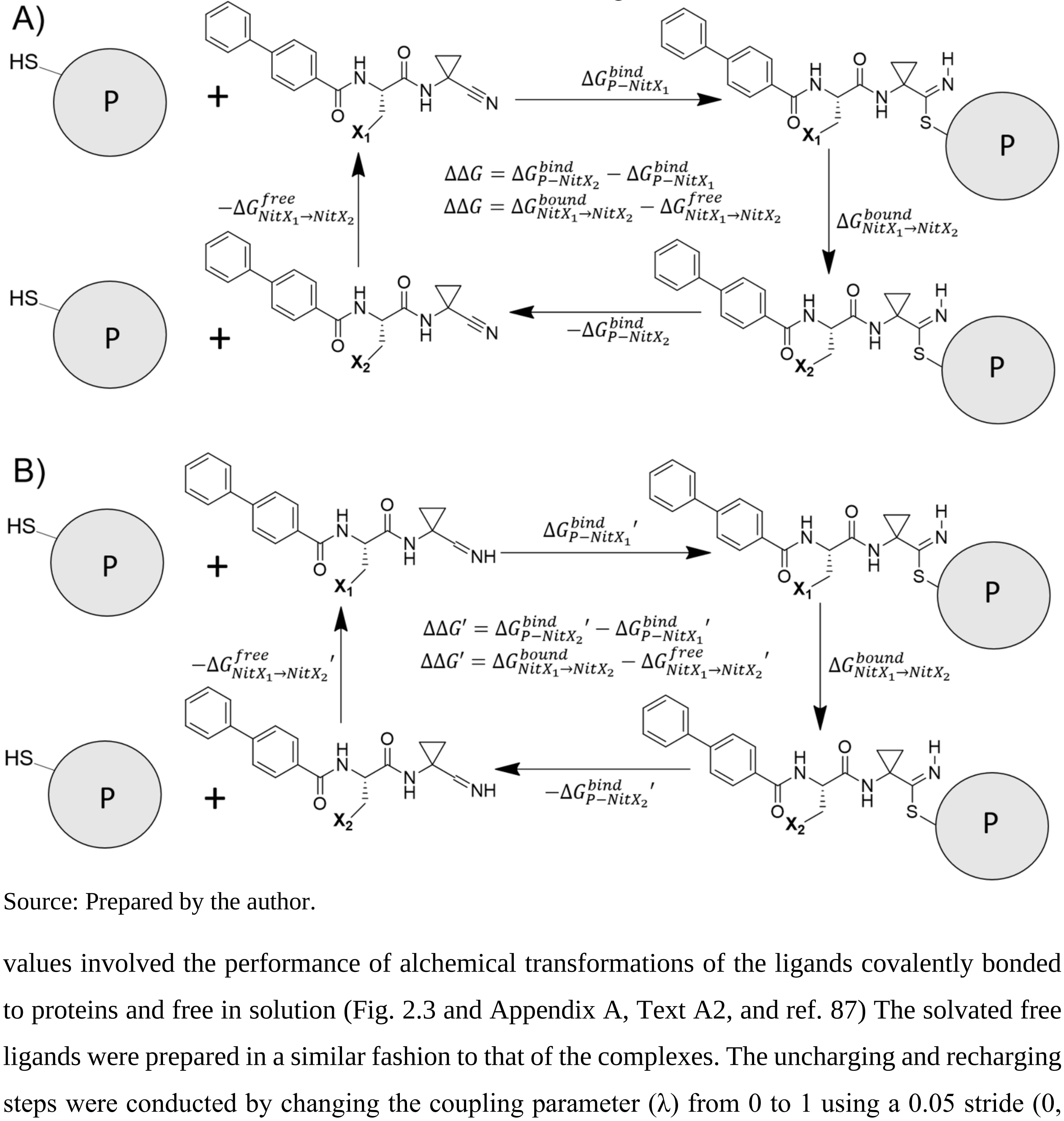
$$V(\lambda) = \lambda V_B + (1-\lambda)V_A \tag{2.15}$$

Note that the derivative with respect to λ in eq. 2.15 leads to the rightmost expression in eq. 2.14, as $\partial V/\partial \lambda = V_B - V_A$. Eqs. 2.14 and 2.15 constitute the basis of the so-called thermodynamic integration (TI) method. The MD simulation codes enabling TI calculations compute and store the potential energy difference between the end states for all frames sampled at every λ value. This allows the prediction of free energies through numerical integration of the $\langle V_B - V_A \rangle_\lambda$ values, e.g., by using the well-known trapezoid rule.[63]

TI free energy calculations were performed here through a three-step approach, which involves: *i*) an uncharging process, *ii*) a van der Waals transformation of atoms bearing zero partial charges and *iii*) a final recharging step of atoms with zero charges.[86] The atoms of the studied ligands included in the TI regions and the type of alchemical transformation that they underwent, i.e., either appearing/disappearing processes or linear transformations, are shown in Appendix A, Fig. A3. The uncharging and recharging steps were carried out through single-topology approach, while the van der Waals transformations were conducted using the dual-topology approach and soft-core potentials were applied to the atoms subjected to mutation to avoid singularities. Soft-core potentials were defined by the default parameters.[63]

The system setup for alchemical free energy calculations was similar to that presented earlier for conventional MD simulations. Representative structures of the complexes calculated from MD simulations and containing water molecules close (≤4 Å) to the ligands were embedded in octahedral boxes with edges placed, at least, 8 Å away from the solute surface and filled with TIP3P waters and sufficient counter-ions. The cut-off of nonbonded interactions was reduced to 8 Å to speed up of the MD simulations. The thermodynamic cycle employed to calculate the $\Delta\Delta G$

Figure 2.3. Thermodynamic cycles used to calculate ΔΔ*G* values for the studied covalent complexes. A) Thermodynamic cycle involving the nitriles in solution and covalently-bonded to the enzyme. The ΔΔ*G* value corresponding to this cycle is equivalent to the experimental one when conditions leading to eq. A6, Appendix A, are fulfilled. **B)** Same as A) but involving imines instead of the nitriles in solution. For simplicity, we kept in B) the same notation of the free energy terms related to equivalent processes in A), except for the addition of an apostrophe (*'*). The latter is missing from the term $\Delta G^{bound}_{NitX_1-NitX_2}$ of cycle B), as it corresponds to an alchemical transformation with identical end states to those of the process on the right vertical leg of cycle A). It can be fairly assumed that ΔΔ*G*=ΔΔ*G'*; therefore, the ΔΔ*G* values reported in this work are calculated using the thermodynamic cycle shown in B). X_1 and X_2 represent different P2 substituents. P stands for the protein. The free energies of the vertical legs are calculated through alchemical transformations in both, the bound and the free states of the ligands.



Source: Prepared by the author.

values involved the performance of alchemical transformations of the ligands covalently bonded to proteins and free in solution (Fig. 2.3 and Appendix A, Text A2, and ref. 87) The solvated free ligands were prepared in a similar fashion to that of the complexes. The uncharging and recharging steps were conducted by changing the coupling parameter (λ) from 0 to 1 using a 0.05 stride (0, 0.5, …, 0.95, 1.00). The van der Waals transformations were carried out in a similar way, except for two more λ values (0.02 and 0.98) that were added close to the end states. In total, each $\Delta\Delta G$ value comprised 130 λ MD simulations, each of them subjected to a short EM, a 1 ns NVT heating process followed by a 12 ns production run in the NPT ensemble, from which the first 2 ns were discarded during free energy calculations. The heating process was performed as explained in Section 2.1.5. On the other hand, the pressure in the NPT MD simulations was controlled by means of the Monte Carlo barostat.[63] Moreover, complete force evaluation, i.e., including H atoms, was performed (ntf = 1) during the TI MD simulations.[63] All simulations were conducted with *pmemd.cuda* of Amber18.[63]

Remarkably, charge-related alchemical transformations involving charged ligands were performed following a single box/dual system approach in order to keep the simulation box neutral at all λ values.[88] The protein-ligand complex and the free ligand were placed in an 86x86x118 Å cuboid box with a minimal distance of 16 Å between them in the *z* axis. The interaction between the free ligand and the complex was prevented by applying 20 $\text{kcal·mol}^{-1}\text{·Å}^{-2}$ to the M177 Cα atom of FP-2 and the P2 Cα atom of the ligand. Two NaCl concentrations, 0 and 0.150 M, were tested in order to assess the impact of ionic strength on $\Delta\Delta G$ calculations involving charged ligands.[89]

MD simulations at every λ value were performed in triplicate both in the forward and backward directions.[90] Convergence was assessed through the calculation of hysteresis.[91] $dV/d\lambda$ vs λ curves were integrated to calculate ΔG values using the trapezoid rule.[86] Mean and standard errors of the mean (*SEM*s) were obtained by bootstrapping. Errors were propagated to calculate the total *SEM*s of the final $\Delta\Delta G$ values.[90] The predicted $\Delta\Delta G$ ($\Delta\Delta G_{calc}$) values were compared to the experimental ones ($\Delta\Delta G_{exp}$), which were derived from the reported IC_{50}'s[40] after conducting pH corrections due to the occurrence of ionization equilibria in the pyridine-containing nitriles (see Appendix A, Text A3).

2.1.10 Umbrella sampling free energy calculations

Umbrella sampling (US) is an equilibrium method that allows the obtaining of the potential of mean force (PMF) along a certain reaction coordinate (ξ), e.g., a Cartesian axis, an interatomic distance, an angle, and others. US involves the use of bias harmonic potentials restraining the system to specific ξ values that will ensure the correct sampling along ξ. The restrained ξ values define different windows, which must be properly weighted and combined in order to obtain the

PMF. The Weighted Histogram Analysis (WHAM) constitutes one of the more widely-used approaches to accomplish the aforementioned task.[92, 93]

Doudou *et al.* developed an US variant that allows the calculation of standard binding free energies (ΔG^{o}_{bind}) from the PMF, when the reaction coordinate is a Cartesian axis, say, the *z* axis, and restraints are applied on the orthogonal *x* and *y* axes.[94] Firstly, the authors noted that the binding free energy along the PMF (ΔG_{PMF}) is given by the following expression:[94]

$$\Delta G_{PMF} = -RTln\left(\frac{Q_{b,R}}{Q_{u,R}}\right) \tag{2.16}$$

where $Q_{b,R}$ and $Q_{u,R}$ are the partition functions for the bound and unbound regions of the PMF, respectively, and the subscript *R* indicates the use of orthogonal restraints. The ratio between both partition functions can be calculated as shown below:[94]

$$\frac{Q_{b,R}}{Q_{u,R}} = \frac{\int\left[exp\left(\frac{-W_R(z)}{RT}\right)dz\right]_{bound}}{\int\left[exp\left(\frac{-W_R(z)}{RT}\right)dz\right]_{unbound}} = \frac{l_b}{l_u}exp\left(\frac{-\Delta W_R(z)}{RT}\right) \tag{2.17}$$

where $W_R(z)$ is the PMF as a function of *z*, which becomes zero at its lowest point, i.e., when the ligand is bound. The boundary between the bound and unbound regions can be placed at the minimal value of z in which the PMF reaches saturation. The PMF depth $\Delta W_R(z)$ is defined as the difference between its lowest value, zero, and the exponential average of the PMF over the whole unbound region:[94]

$$\Delta W_R = RTln\left\{\left[\int exp\left(\frac{-W_R(z)}{RT}\right)dz\right]_{unbound}\Big/\int(dz)_{unbound}\right\} \tag{2.18}$$

The bound and unbound lengths in eq. 2.16 are configurational integrals of the PMF defined as follows:[94]

$$l_b = \int\left[exp\left(\frac{-W_R(z)}{RT}\right)dz\right]_{bound} \tag{2.19}$$

$$l_u = \int\left[exp\left(\frac{-(W_R(z)+\Delta W_R)}{RT}\right)dz\right]_{unbound} = \int(dz)_{unbound} \tag{2.20}$$

As the lowest values of $W_R(z)$ contribute more to the configurational integral in eq. 2.19, in practice, the l_b value becomes mostly insensitive to the chosen *z* cut-off. To determine the ΔG^{o}_{bind} value from that of ΔG_{PMF}, one must add the contributions to the free energy of releasing the orthogonal restraints and of changing from standard state volume (V^o=16661 Å^3) to that sampled in the unbound region ($V_{u,R}$), which are termed ΔG_R and ΔG_V, respectively:[94]

$$\Delta G^o_{bind} = \Delta G_{PMF} + \Delta G_V + \Delta G_R \tag{2.21}$$

ΔG_V can be calculated by the equation below, which is obtained after integrating out the area sampled by the ligand in the entire unbound state region upon the action of orthogonal restraints ($A_{u,R}$), defined, in turn, by the harmonic constant k_{xy}:[94]

$$\Delta G_V = -RTln\left(\frac{V_{u,R}}{V^o}\right) = -RTln\left(\frac{l_u A_{u,R}}{V^o}\right) = -RTln\left(\frac{l_u 2\pi RT}{k_{xy}V^o}\right) \tag{2.22}$$

The rightmost equality in eq. 2.21 derives from the following expressions: [94]

$$A_{u,R} = \int_{-\infty}^{+\infty} exp\left(\frac{-k_{xy}}{2RT}\right)dx \int_{-\infty}^{+\infty} exp\left(\frac{-k_{xy}}{2RT}\right)dy = \frac{2\pi RT}{k_{xy}} \tag{2.23}$$

$$V_{u,R} = l_u \times A_{u,R} \tag{2.24}$$

On the other hand, ΔG_R can be calculated through a single free energy perturbation (FEP) step as follows:[94]

$$\Delta G_R = RTln\, \langle exp\left(\frac{-k_{xy}(\Delta x^2 + \Delta y^2)}{2RT}\right)\rangle_{k_{xy}=0} \tag{2.25}$$

where Δx and Δy are the displacements with respect to the restraint minimum in the reference bound state, when no orthogonal restraints are applied. By substituting eqs. 2.16, 2.17, and 2.22 into eq. 2.21, one finally obtains:[94]

$$\Delta G^o_{bind} = \Delta W_R - RTln\left(\frac{2\pi l_b RT}{k_{xy}V^o}\right) + \Delta G_R \tag{2.26}$$

The previous equation constitutes the basis for the calculation of ΔG^o_{bind} values from unidimensional PMFs obtained through US.[94]

The above-explained US protocol was adapted to run the MD simulations in Gromacs v5.1.4.[71] $\Delta G°_{bind}$ values for the interaction of Cpd66 with FP-2 and with FP-2 complexed with a peptide substrate were calculated through this procedure. We used the previously-derived parameters of the ligand and the Amber ff99SB-ILDN force-field for the protein.[95] A slow pulling MD simulation of Cpd66 after carefully orienting the complex along the *z* axis, the Cartesian coordinate chosen as reaction coordinate, was performed to minimize clashes as the ligand leaves the deeply-buried cavity. Subsequently, 38 frames were extracted from the pulling simulation, each corresponding to a different distance between the protein and the ligand reference atoms and covering a 0 to 50 Å range along the reaction coordinate. The equilibrium *z* distances between the reference atoms were restrained by means of harmonic potentials. In addition, orthogonal restraints were applied during pulling and US simulations and the six external degrees of freedom of the protein were restrained. More details on the US protocol are provided in Appendix A, Text A4 and ref. 58) The data collected during the US productive simulations were processed with gmx wham program, and the PMF for each studied system was finally obtained.[93]

ΔG_R was calculated by performing five 25 ns replicas of unrestrained MD simulations for each system. The deviations from the equilibrium position along the x and y components of the distance vector were obtained from the previous simulations, after discarding the first 5 ns of each. ΔG_R values were calculated through eq. 2.24 and the mean over five replicas was computed.[58]

2.1.11 Pocket volume measurements

The volume of the transient pocket in the site 3 region was measured for every frame collected from the MD simulations using POVME 3.0.[96] It was established from visual inspection that the cavity occurring in the site 3 region of FP-2 and FP-3 was open when the transient pocket volume and the distance between the side-chains of key residues lying at the pocket entrance were both greater than certain respective thresholds.[58] The frames fulfilling these conditions were extracted and clustered.[58] The central structure of the largest cluster was employed for further analyses.

2.1.12 Community analysis and calculation of optimal and suboptimal paths

Our comprehension on allostery has recently undergone a change of paradigm, from a simplistic model involving the transition between two states (=conformations) to a more elaborate one comprising a shift in the equilibrium populations of multiple conformations, triggered by the modulator binding.[23] In this context, the idea of allosteric influence involving little structural change or even without well-defined structures has emerged.[97] The new vision on allostery makes the computational methods capable of exploring the phase space of molecules, such as MD and MC simulations, particularly suitable to study this phenomenon.[23] Among the various methods that rely on exploring the phase space to study allostery, community analysis (CA) constitutes a widely-used approach to dissect the underlying mechanisms triggered by allosteric modulators at molecular level.[23, 98]

CA is essentially a network-based representation of the protein, in which each residue is represented by a node located, for example, at its Cα [97]. Each pair of residues (i,j) in the network is connected by an edge whose length (d_{ij}) is calculated through the following expression:[97]

$$d_{ij} = -Log\left|C_{ij}\right| \tag{2.27}$$

where C_{ij} is the pairwise correlation between them. C_{ij} values can be computed from the ensemble of structures generated through MD simulations via the Pearson equation for linear correlations or a more sophisticated approach based on generalized correlations. The latter, though more computationally-demanding, can identify physically relevant allosteric connections that escape to the former approach.[97]

CA provides a coarse-grain representation of the network graph by clustering groups of residues into communities. Each community contains highly correlated residues; intercommunity connections being represented by edges that are weighted to account for the strength of their interactions.[97] Therefore, each community represents a module of correlated motions in the protein. The Girvan-Newman iterative algorithm can be employed to cluster the network graph.[99] Often, the network is filtered by ignoring low pairwise correlations and/or correlations between residues that are not in physical contact during most of MD simulations time, in order to reduce noise and increase robustness in the calculation of communities.[23, 97, 100]

The importance of an individual residue to the optimal dynamic networks in a system can be identified by means of its *betweenness centrality* (c_i):[97]

$$c_i = \frac{1}{N}\sum_{i\neq t, i\neq s} \sigma_i(s,t) \tag{2.28}$$

where N is the total number of pathways and s,t is every residue pair not including i. σ_i is defined as follows: it takes on the value 1 whenever the shortest pathway between residues t,s passes through residue i, and is zero otherwise.

The communication between the allosteric and the orthosteric sites occurs along paths involving specific residues, some of them, i.e., the hotspots, being critical to the process, as revealed by mutagenesis analyses.[97] To determine the most likely communication paths, a method called Weighted Implementation of Suboptimal Paths (WISP) has been devised, which uses the residue-residue distances d_{ij} defined in eq. 2.26 as the metric to determine the path length.[101] The latter is calculated as the sum of all d_{ij} between the interconnected pairs of residues along a certain path. WISP employs the Dijkstra's algorithm to search for the optimal (= minimal length) and suboptimal (= near optimal length) pathways between two nodes (= Cα's). Communication hotspots can be predicted on the basis of degeneracies, i.e., the number of times each residue occurs in the whole set of generated suboptimal paths.[101]

CA was performed for four systems: free FP-2, FP-2:peptide, FP-2:Cpd66 and FP-2:Cpd66:peptide.[98] The generalized Cα-Cα cross-correlation coefficients (*GC*) were calculated across the five replicate MD simulations conducted for each studied system, using g_correlation program of GROMACS v.3.3.4.[102] The *GC* matrices were then filtered using heavy atom contact maps calculated during the same set of MD trajectories,[100] in order to rule out the correlations between non-contacting residues. The Bio3d package implemented in R was employed for the previous calculations.[103] In addition to contact maps, we defined correlation cut-offs to eliminate noise, i.e., low and/or artificially-high correlations, and *GC* matrices were processed in a way that significant correlations were preserved, even when involving non-contacting residues. All relevant details concerning the processing of *GC* matrices is explained in details in ref. 58. The Guirvan-Newman algorithm was employed to conduct CA.[99] The final number of communities for every system was chosen as the minimum one yielding a near-optimal modularity based on the average filtered *GC* matrix of every system.[100] Betweenness centralities were also calculated in order to determine the overall relevance of each residue in signal propagation across the protein.[97] All the previous steps were carried out with Bio3d package of R.

Finally, optimal and suboptimal signalling paths between source and sink residues linking the allosteric pocket and the active site were calculated using WISP.[101] In total, 500 suboptimal paths were calculated between the source and sink. Residue degeneracies, as well as path length distributions, were obtained for all the studied systems.[58]

2.1.13 Principal Component Analysis

To extract the motions relevant for the function of a given protein from an MD simulation is a non-trivial task. Principal component analysis (PCA) or essential dynamics (ED) is a statistical method allowing the determination of uncorrelated collective coordinates, which describe the structural fluctuation of a protein during an MD simulation. Moreover, the determination of the system collective coordinates allows the identification of a small subset of dimensions or degrees of freedom accounting for the most important structural changes in a protein (the so-called essential subspace). PCA is based on the diagonalization of the covariance matrix built from atomic fluctuations in a trajectory from which the overall translation and rotations have been removed:[104]

$$C_{ij} = \langle (X_i - X_{i,0})(X_j - X_{j,0}) \rangle \tag{2.29}$$

in which X are the separate x, y, and z coordinates of the atoms fluctuating around their average positions X_0. $\langle \dots \rangle$ represents the average values calculated from the entire trajectory. Frequently, to construct the protein covariance matrix, the Cα atom trajectory is used. Indeed, it has been shown that the Cα atom contains all the information for a reasonable description of the protein's large concerted motions. Upon the covariance matrix diagonalization, a set of eigenvalues and eigenvectors, also called principal components (PCs), is obtained. The motions along a single eigenvector correspond to concerted fluctuations of atoms. On the other hand, the eigenvalues represent the total mean square fluctuation of the system along the corresponding eigenvectors.[104]

PCA was conducted for all the studied systems, i.e., free FP-2, FP-2:Cpd66, FP-2:peptide and FP-2:Cpd66:peptide complexes as explained elsewhere.[58] Briefly, the covariance matrix for the Cα atoms of free FP-2 concatenated MD simulation was obtained and PCs were subsequently calculated. Per-residue contribution to the main PCs, extreme structures and 2D representations of the protein motions along pairs of PCs were also carried out.[58] The *gmx covar* and *gmx anaeig* programs of Gromacs v5.1.4 were used to conduct all the above-mentioned tasks.[71]

2.1.14 Miscellaneous trajectory analyses

Root mean squared deviations (RMSDs), interatomic distances and root mean square fluctuations (RMSFs) values along the MD trajectories were calculated with *cpptraj* module of Amber (versions 14, 16 and 18).[62, 63] The central structures for the MD simulations of all the studied complexes generated with *pmemd.cuda*, except for those involving nitriles, were calculated using the average linkage algorithm implemented in the *cpptraj*, taking into account the RMSD values for the ligands and the residues lying within a 4 Å cut-off. As a rule of thumb, and unless otherwise stated, five predefined clusters were set, and the central structure corresponding to the most populated one was chosen for subsequent analyses. All figures displaying structural data were created with Pymol 2.1.0,[105] except for those showing the suboptimal paths connecting the active and allosteric sites in the studied complex, which were created with VMD 1.9.1.[106] Electrostatic representations of protein surfaces were obtained with the Adaptive Poisson-Boltzmann Solver (APBS) plugin of Pymol 2.1.0.[105] Direct H-bonds and water bridges during the MD simulations were calculated using *hbond* command.[63] A donor-hydrogen-acceptor angle ≥ 120° and a donor-acceptor distance ≤ 3.5 Å were set as geometric criteria to define the occurrence of an H-bond.[107]

Central structures of the protein:nitrile complexes were obtained through two clustering steps in order to account for different orientations of the pyridine ring within the S2 subsite. Firstly, trajectories were clustered based on distance profiles between the N_{pyr} atom of the ligand and a key residue of the protein (see Sections 5.1.1 and 5.1.2). The number of clusters for each system was taken equal to the number of peaks in the corresponding distance distribution and trajectories were split accordingly. Secondly, a single central structure per newly-generated trajectory was calculated by clustering the latter with respect to the RMSD of the interface heavy atoms as explained before. The central structure obtained from RMSD clustering of the most populated distance-based cluster was taken for TI free energy calculations. SASAs were calculated with *molsurf* command using the default probe radius (1.4 Å) and atomic radii (mbondi set).[63]

Conformational clusters corresponding to the FP-2 simulations run with Gromos 54a8 force-field were carried out with the *gmx cluster* program of Gromacs v.5.1.4. As a metric for this clustering analysis, we selected the RMSD of the heavy atoms of residues within site 6. Various RMSD cut-offs were tested and, finally, 0.12 nm was deemed suitable, as it yielded a handful number of clusters (less than 10) covering more than 80% of the analyzed MD trajectories.

2.2 Experimental section

The experimental procedures involving enzymatic assays and determination of the inhibition mechanisms were conducted by Dr. Emir Salas Sarduy (Universidad Nacional de San Martín, Argentina). Antiplasmodial and cytotoxicity *in vitro* assays were performed by Dr. Luisa F Hernández and Dr. Adriana Pabón (Universidad de Antioquia, Colombia), and Dr. María J Pascual and Dr. Diego E Álvarez (Universidad Nacional de San Martín, Argentina), respectively.

2.2.1 Expression and purification of recombinant enzymes

FP-2 was expressed as inclusion bodies in BL21(DE3) *Escherichia coli* strain, purified under denaturing conditions (final purity: 91%) and refolded to active enzyme [108]. A *Pichia pastoris* strain expressing a Ser149Ala (glycosylation) hCatK mutant was donated by Dr. Dietter Bromme (University of British Columbia, Vancouver, Canada). The enzyme was expressed, activated and purified (final purity: >85%) as previously described.[109]

2.2.2 *In vitro* enzymatic assays of compounds selected from SBVSs targeting the active site

The nine top-scoring compounds selected from the Maybridge HitFinder^{TM} SBVSs were purchased for experimental assessment. Inhibition assays against FP-2 and hCatK were conducted as described by Bertoldo *et al*.[36] Briefly, stock solutions of the compounds were prepared at 10-20 mM in dimethyl sulfoxide (DMSO). The enzymatic activity of both enzymes was measured by fluorimetric assays in a buffer (100 mM NaOAc, 10 mM dithiothreitol (DTT), pH=5.5) employing Z-FR-AMC (15 µM) as enzymatic substrate, and fixed concentrations of FP-2 (0.7 nM) and hCatK (1 nM). The enzymes and compounds were incubated 15 minutes prior to the addition of substrate and the release of 7-amino-4-methyl coumarin (AMC) was monitored for 600 s at 30 °C with a Thermo Spectronic Aminco Bowman Series 2 spectrofluorometer (excitation at λ=355 nm; emission at λ=460 nm). All scans were corrected from the corresponding blanks, and controls (FP-2/hCatK+DMSO 2%, Z-FR-AMC+DMSO 2% and FP-2/hCatK+Z-FR-AMC+DMSO 2%). IC_{50} values were determined from dose-response curves varying the inhibitor concentration, while keeping fixed the concentrations of the enzyme and the substrate. The curves were fitted using GraphPad Prism version 5.03.[110] All measurements were conducted in triplicate.[50]

2.2.3 *In vitro* antiplasmodial activity assay

The antimalarial activities of the compounds selected from Maybridge HitFinder^{TM} SBVSs were tested *in vitro* against cultures of erythrocytes infected with a multi-resistant strain of *P. falciparum* (FCR3).[50] The parasite cultures were incubated at 37 °C in a pure gas mixture of 5% O_2, 5% CO_2, and balanced N_2 environment, in RPMI 1640 medium supplemented with 25 mM HEPES, 5% (w/v) $NaHCO_3$, 0.1 mg/mL gentamicin, and 10% A+ heat-inactivated human serum, as previously described.[111].

Several concentrations of each compound (from 200 µM to 1.5 µM) were evaluated. Diphosphate salt of chloroquine (CQ), ≥ 98%, SIGMA C6628, evaluated in a range of 0.0023 to 2.0 µM, and *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane (E64), 0.3125 to 20.0 µM, were used as a treatment control in each assay. Culture medium in the absence of the compounds was used as a growth control. A suspension of parasitized erythrocytes with a hematocrit of 2%, a total parasitemia of 1%, and with a predominance of young forms, was prepared. The cultures with the treatments were incubated at 37 °C for 48 h in the presence of 5% CO_2, 5% O_2 and 90% N_2. [112] After the incubation period, 50µL of 0.4% SYBR Green was added to each well and incubated for

10 min, so that the fluorochrome was intercalated into the DNA of the parasite. After a 10 min incubation, fluorescence emission detection was performed using the BD AccuriTM C6 flow cytometer (excitation at λ=485 nm excitation, emission at λ=530 nm).[113] Each concentration was evaluated in duplicate, and two or more independent assays were performed. The percentages of parasitemia were analyzed using a non-linear slope-dependent regression with GraphPad Prism™ version 5.03 to estimate the IC_{50} values.[110]

2.2.4 Cytotoxicity assays

In addition to antiplasmodial activity, cytotoxicity assays were performed as described previously.[114] Briefly, HeLa cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% FBS for 24 h at 37 ºC under a 5% CO_2 atmosphere. Serial dilutions of each compound were prepared in DMEM containing 2% Phosphate-Buffered Saline (PBS) and 1% DMSO. Cells were incubated with compound dilutions for 72 h at 37 ºC under a 5% CO_2 atmosphere. After the incubation period, viable cells were stained with crystal violet and absorption was measured at λ=590 nm. The percentage of viability was calculated as follows:

$$\%Vialbility = \left[(A_{590nm})_{sample}/(A_{590nm})_{control}\right] \times 100\% \tag{2.30}$$

where A_{590nm} stands for the absorbance at the indicated wavelength. The half-maximal cytotoxic concentration (CC_{50}) for each compound was estimated by nonlinear regression from dose-response curves using GraphPad Prism version 5.03.[50, 110]

2.2.5 Compounds from ZINC12 database

Compounds identified from the SVBSs against the site 6 conformations of FP-2 were purchased from Eanime Ltd. Stock solutions (25 mM) of investigational compounds were prepared in DMSO. Due to its lower solubility, stock solutions for some compounds were prepared at 10 mM.

FP-2 activity was assayed fluorometrically with Z-LR↓AMC (Bachem) as substrate, according to a previously-described procedure,[38] to which only minor changes were made. Assays (final reaction volume ~80 µL in Corning solid black 384 well plates) were performed at 30ºC in 100 mM acetate buffer pH 5.5 containing 5 mM DTT and 0.01 % Triton X-100. Unless otherwise stated, the fluorogenic substrate was added at final concentration of 5µM (~1 x K_M) to match balanced assay conditions.[115] The AMC release was monitored continuously for 60 min with a

FilterMax F5 Multimode Microplate Reader (Molecular Devices) using standard 360 nm excitation and 465 nm emission filter set. Enzyme activity was estimated as the linear region slope of the resultant progress curves. Under the described conditions, the Selwyn test[116] indicated that enzyme remained stable during the whole assay.

2.2.6 Activity assays of potential allosteric inhibitors against FP-2

For preliminary assessment of FP-2 inhibition, four 10-fold serial dilutions (25 mM, 2.5 mM, 0.25 mM and 0.025 mM) were prepared in DMSO from stock solutions of each compound. One microliter of each compound dilution, E-64 (Sigma Aldrich) (10 µM in DMSO) or DMSO were dispensed into each well, followed by 20 µL of activity buffer. Then, 40 µL of activity buffer containing FP-2 (6.6 nM) were added to each well, plates were homogenized (30 seg, orbital, medium intensity) and each well subjected to a single autofluorescence read (exc/ems = 360/465nm). Plates were incubated in darkness for 30 min at 30 ºC and then 20 µL of Z-LR-AMC (20 µM in assay buffer) were added to each well to start the reaction. After homogenization (30 seg, orbital, medium intensity), the fluorescence of AMC (exc/ems = 360/465nm) was acquired kinetically for each well (12 read cycles, one cycle every 300 s). Fluorescence measurements were used to determine the slope (dF/dt) of progress curves by linear regression and inhibition percentage (%*Inh*) was calculated for each compound according to:

$$\%Inh = 100\% \times [1 - ((dF/dt)_{well} - \mu C_-)/(\mu C_+ - \mu C_-)] \qquad (2.31)$$

where $(dF/dt)_{well}$ represents the slope of each compound well and μC_+ and μC_-, the average of FP-2+DMSO (no-inhibition) and substrate (no-enzyme) controls, respectively. The compounds showing %*Inh* > 20% at the higher concentration and dose-response inhibition were selected for a secondary screening.

The selected compounds were re-tested in a dose-response manner (final concentration ranging from 47 pM to 1562.5 µM) using similar assay conditions. 10 µL of compound stock solutions (25 mM in DMSO), E-64 (10 µM in DMSO), and DMSO were added to the first wells (column 1), followed by addition of 30 µL of activity buffer. After addition of 20 µL of buffer to subsequent wells, 16 serial 2-fold dilutions were made horizontally. Then, 40 µL of activity buffer containing FP-2 (6.6nM) were added to each well, except for those corresponding to C-; completed with 40 µL of activity buffer. After homogenization, incubation, and autofluorescence measurement, 20 µL of Z-LR-AMC substrate (20 µM in activity buffer) were added. Data collection and processing were performed exactly as described above.

The percentage of FP-2 residual activity (%*Res.Act*) was calculated for each condition according to:

$$\%Res.Act = 100\% \times \frac{((dF/dt)_{well} - \mu C_{-})}{\mu C_{+} - \mu C_{-}} \tag{2.32}$$

IC_{50}'s and Hill slope parameters were estimated by fitting experimental data from dose-response curves to the four-parameter Hill equation using GraphPad Prism v5.03:[110]

$$y = \ y_{min} + \frac{y_{max} - y_{min}}{\left(1+10^{(LogIC_{50}-x)HillSolpe}\right)} \tag{2.33}$$

where y_{max} and y_{min} are the maximum and minimal percentage of residual activities, x stands for the different concentrations of the compound and *HillSlope* is a coefficient controlling the shape of the curve.

2.2.7 Centrifuge counter-screening

To identify which of the potentially allosteric compounds form aggregates under the assayed conditions (potential nonspecific effects on FP-2 activity), a centrifuge counter-screen was performed as described elsewhere,[117] with minor changes. In brief, 2-fold serial dilutions of compounds were initially prepared in DMSO (final volume 25 µL). Then, 225 µL of activity buffer were added to each tube, vortexed and incubated for 5 minutes at room temperature. Then, samples were divided: one half was incubated 30 minutes at 4ºC (non-centrifuged) and the other half centrifuged at 13000 rpm, 30 minutes, 4ºC (centrifuged). Finally, both samples were assayed for dose-dependent inhibition of FP-2 as indicated in Section 2.2.6.

2.2.8 Determining mode of inhibition and K_i value of active compounds

FP-2 assay was performed as described earlier, with the following modifications: activity buffer was supplemented with 16% DMSO to achieve maximal solubility of investigational compounds. In addition, the final enzyme concentration was increased to 10.56 nM and the assay volume was reduced to 50 µL. Under these conditions, FP-2 showed an apparent K_M value of 25 µM for Z-LR-AMC.

The identification of the mode of inhibition was performed as described elsewhere. [115] In brief, FP-2 activity was determined for eight different substrate concentrations (ranging from 1.56

to 200 μM) in the presence of six fixed doses of the active inhibitors (ZINC03225317: 0, 25, 50, 100, 200 and 400 μM; ZINC72290660: 0, 400, 800, 1200, 2600 and 4000 μM). To visually infer the mode of inhibition, classical Lineweaver-Burk plots were constructed for each compound. Finally, kinetic data were arranged in the form of Michaelis curves (dF/dt vs. $[\text{Z-LR-AMC}]_0$) and globally fitted to the corresponding (mixed inhibition) equation present in GraphPad Prism v5.03[110] for the simultaneous estimation of α and K_i.

3 IDENTIFICATION OF FALCIPAIN-2 INHIBITORS THROUGH AN INTEGRATED *IN SILICO* AND EXPERIMENTAL APPROACH

3.1 Identification of FP-2 inhibitors targeting the enzyme active site

The results presented in this section have been already published in the scientific journal *Biochimica et Biophysica Acta -General Subjects.*[50]

3.1.1 *In silico* identification of selective FP-2 inhibitors from Maybridge HitFinder^{TM} Database

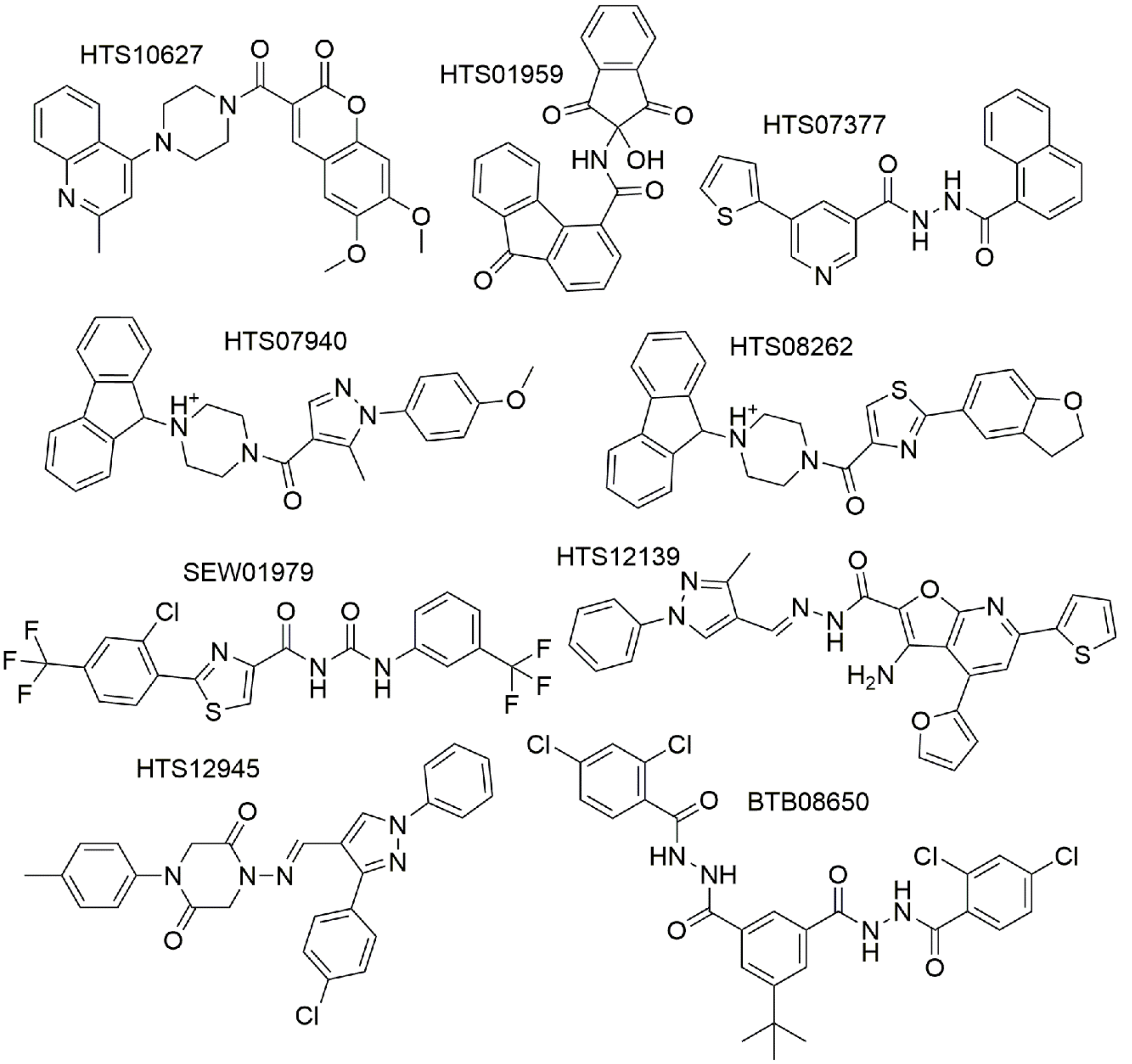
The Maybridge Hitfinder^{TM} database was screened to identify putative selective inhibitors of FPs. In order to potentially enhance the selectivity of the compounds, we took into account the interaction with human off-targets. Among the human cathepsins analyzed in this work, hCatK was found as the one sharing the largest ID with FP-2 in the active site region (see Appendix B, Fig. B1 and Table B1). In fact, the ID percentages (%IDs) of this protease with respect to FP-2 and FP-3 are 53.85% and 61.54%, respectively (see Appendix B, Fig. B1 and Table B1). Other proteases, such as hCatL and hCatB, were ranked in second and third places, respectively, with %IDs ranging from 46.15% to 48.72% with respect to both FPs (see Appendix B, Fig. B1 and Table B1). Furthermore, FP-2, FP-3 and hCatK share the same specificity for Leu at P2; whereas hCatB and hCatL prefer Phe in that position.[16, 118-123] On the basis of the previous analyses, we selected hCatK to evaluate the potential off-target inhibition caused by the compounds selected from the SBVS results.

Table 3.1. Autodock-Vina scores for the selected compounds[a]

Compound	S_{vina}(FP3) (kcal/mol)	ΔS_{vina}(FP3-hCatK) (kcal/mol)	ΔS_{vina}(FP2) (kcal/mol)	ΔS_{vina}(FP2-hCatK) (kcal/mol)
HTS01959	-8.9	-1.2	-8.0	-0.3
SEW01979	-8.4	-1.4	-9.1	-2.1
HTS10627	-8.7	-1.5	-8.8	-1.6
BTB08650	-9.0	-1.1	-8.0	-0.1
HTS07377	-8.6	-1.3	-8.2	-0.9
HTS08262	-9.4	-1.1	-9.6	-1.3
HTS12139	-8.5	-1.4	-8.4	-1.3
HTS12945	-8.9	-1.3	-8.4	-2.1
HTS07940	-8.8	-0.9	-9.2	-1.3

[a]Compounds were selected according to the conditions stated in Section 2.1.2.
Source: Prepared by the author.

Figure 3.1. Chemical structures, and identifier of the compounds selected from SBVSs. The identifier of each compound is shown on top of its structure.



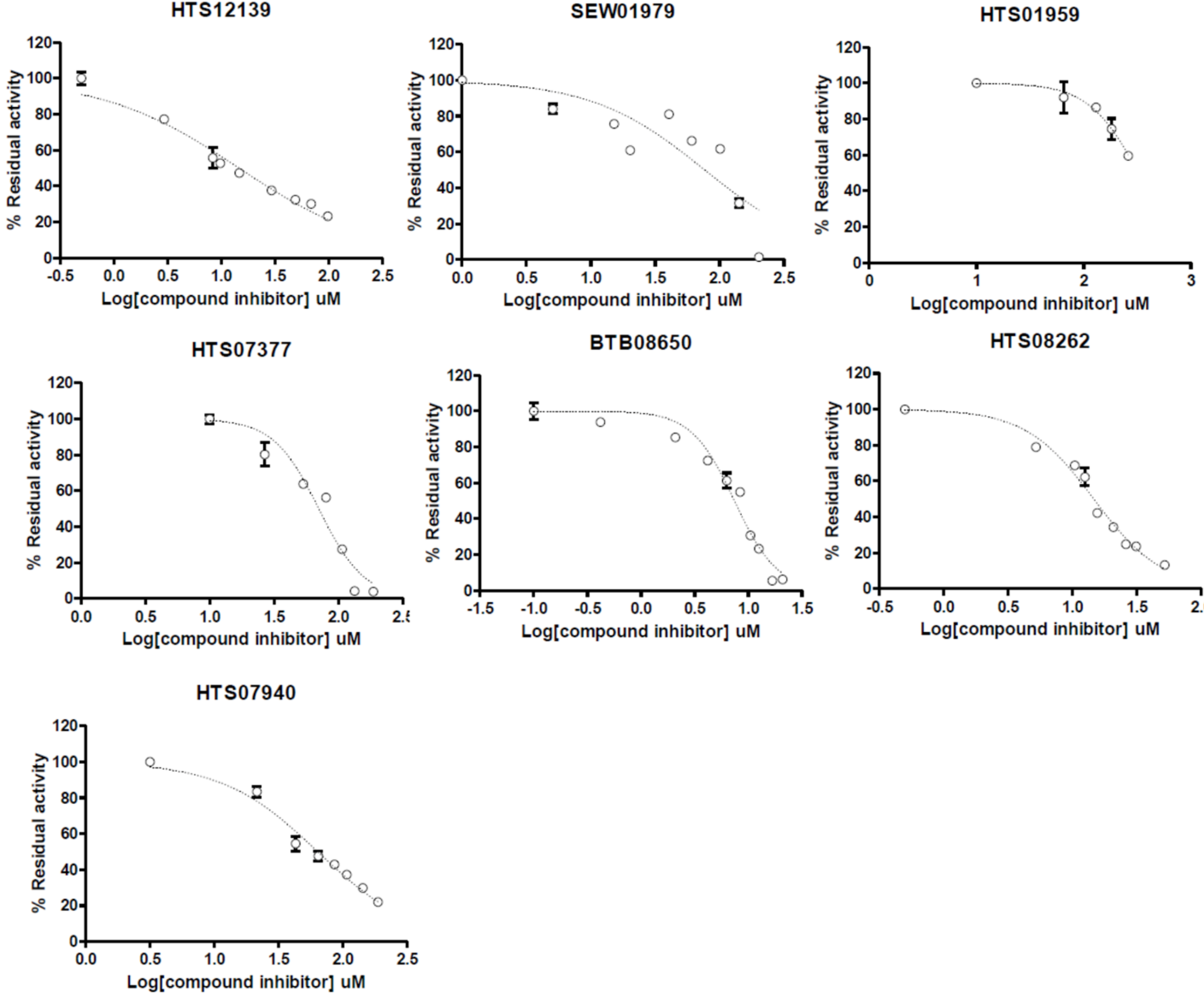
Source: Prepared by the author.

Nine compounds were chosen from the results of the SBVSs that were carried out in parallel for both FPs and hCatK (Table 3.1 and Fig. 3.1), and were subsequently filtered using the PAINS-Remover webserver (http://cbligand.org/PAINS) to exclude potential false positive compounds with non-specific reactivity.[124] All the compounds shown in the table fulfill the conditions of affinity and selectivity stated in Section 2.1.2.

3.1.2 Experimental evaluation of the selected compounds

The compounds shown in Table 3.1 were purchased and subjected to experimental evaluation in order to determine their ability to inhibit the FP-2 enzymatic activity *in vitro*. In parallel, their inhibitory activity against hCatK was assessed in cross-inhibition assays. Seven compounds inhibited FP-2 in a dose-dependent manner (Fig. 3.2), with IC_{50} values ranging within 10^{-4} to 10^{-6} M (Table 3.2). On the other hand, none of these compounds was active against hCatK (Table 3.2). In fact, at the highest concentration of each compound tested in the cross-inhibition assays, only 8-25% of hCatK maximum inhibition was attained, thus clearly indicating that the compounds are more active against FP-2.

Figure 3.2. Dose-response curves for seven FP-2 inhibitors identified by SBVS. The percentages of FP-2 residual activity were measured using at least eight increasing concentrations of the inhibitors. IC_{50} values were determined by fitting experimental data to the four-parameter model of GraphPad Prism (eq. 2.33).



Source: Prepared by the author.

Subsequently, the antiplasmodial activities of the seven compounds displaying FP-2 inhibition were assessed, using E64 and CQ as positive controls. Four compounds (HTS07377, BTB08650, HTS08262 and HTS07940) inhibited the *P. falciparum* growth *in vitro*, with IC_{50}

values within the 10^{-4} to 10^{-6} M range (Table 3.2). The cytotoxicity of the four active compounds against HeLa cells was also measured and it was found that two of them, HTS07940 and HTS08262, possessed promising selectivity indexes (46 and 10, respectively, Table 3.2). Remarkably, the IC_{50} value of HTS07940 against *P. falciparum* cultures is similar to that of E64, the prototypic inhibitor of C1 cysteine proteases,[118] and only one order-of-magnitude lower than that of CQ (Table 3.2). In general, the antiplasmodial activities of HTS07940 and HTS08262 against *P. falciparum* cultures are in the typical range displayed by many non-peptidic FP-2 inhibitors reported so far in literature.[5]. However, the most promising feature of the identified hits is their suitable selectivity indices (Table 3.2).

Table 3.2. IC_{50} values and selectivity indices for the selected compounds

Compound	IC_{50} (FP-2) (μM)	IC_{50} (hCatK) (μM)	Selectivity Index[a]	IC_{50} (*Pf*)[b] (μM)	CC_{50} (HeLa) (μM)	Selectivity Index[c]
HTS01959	326(56)[d]	>261	>1	Inactive	-	-
SEW01979	77(25)	>142	>1	Inactive	-	-
HTS10627	n.d.[e]	n.d.	n.d.	Inactive	-	-
BTB08650	7.4(0.5)	>314	>42	29(5)	34(5)	1.17(0.03)
HTS07377	71(8)	>469	>6	163(6)	<33.5	<0.2
HTS08262	14.7(0.8)	>521	>35	34(7)	350(14)	10(2)
HTS12139	15(2)	>294	>20	Inactive		-
HTS12945	n.d.	n.d.	-	Inactive	-	-
HTS07940	64(5)	>565	>8	2.91(0.04)	133(10)	46(3)
E64	-[f]			1.3(0.1)	-	-
CQ				0.14(0.01)	-	-

[a] Selectivity index for FP-2 calculated as follows: Selectivity index=IC_{50}(hCatK)/IC_{50}(FP-2).
[b] *Pf* stands for a culture of *P. falciparum*-infected erythrocytes.
[c] Selectivity index for parasite cells calculated as follows: Selectivity index=CC_{50}(HeLa)/IC_{50}(*Pf*).
[d] Standard deviations calculated from three independent measurements are shown between parentheses.
[e] IC_{50} values were not measured due to low water solubility of the compounds.
[f] Measurement not performed.
Source: Prepared by the author.

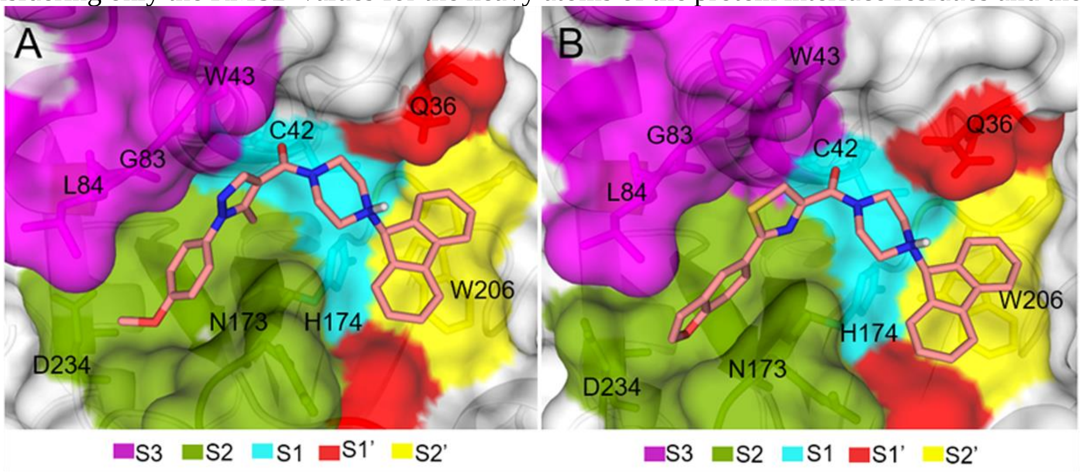
Of note, HTS07940 and HTS08262 have a similar scaffold consisting of (4-(9H-fluoren-9-yl) piperazin-1-yl) *R*-yl methanone, where *R* stands for derivatives of a core structure that contains a five-membered aromatic ring attached by a single covalent bond to a six-membered aromatic ring (Fig. 3.1). Incidentally, Genz10850, an analog compound in which *R* is an indole group, has been already described in previous works as an inhibitor of reduced nicotinamide adenine dinucleotide (NADH)-dependent enoyl (acyl carrier protein) reductases (ENR) of *Mycobacterium tuberculosis* (InhA, $IC_{50} = 0.16$ μM) and *P. falciparum* (PfENR, $IC_{50} = 18$ μM).[125] This compound also shows inhibitory activity *in vitro* against several *P. falciparum* strains, with IC_{50} values ranging from 14 to 31 μM.[125] The crystal structure of InhA in complex with Genz10850 (PDB: 1P44) has been also reported. Therefore, we used the bound conformation of the compound to build a model of a putative PfENR:HTS07940:NAD$^+$ ternary complex (see Appendix B, Text B1 and Fig. B2). Our results suggest that HTS07940 (and, presumably, HTS08262) interact with PfENR. Hence, the observed inhibition of *P. falciparum* growth exerted by these two compounds (Table 3.2) could stem from a combined action on PfENR and FP-2. This prediction requires further experimental validation and remains beyond the scope of this work.

3.1.3 Prediction of the binding modes of HTS08262 and HTS07940 to FP-2

The binding modes to FP-2 of the two most promising compounds, HTS08262 and HTS07940, were predicted after subjecting docking poses to MD simulations. As both compounds share a common structure, similar binding modes to FP-2 are expected for them. Therefore, we conducted a thorough refinement of HTS07940 docking poses into the FP-2 active site to determine its more likely bound conformation, while the inferred lowest-energy pose of HTS08262 was subjected to similar analyses. 100 ns MD simulations were performed for a set of fourteen HTS07940 docking poses, clustered into six groups (from A to F) on the basis of ligand RMSD values (Appendix B, Figs. B3 and B4). Two stable binding modes, i.e., A3 and F1, were identified by analyzing the final frames of the generated trajectories (Appendix B, Fig. B4). Subsequent MM-GBSA free energy calculations corresponding to the last 3 ns of each trajectory confirmed that pose A3 is the most stable one (Appendix B, Table B2).

The phase space sampling of HTS07940 pose A3 was enhanced by eleven frames obtained from the previous 100 ns MD simulation of the complex. MM-GBSA ranking of the central structures calculated for each independent aMD simulation led to the conclusion that pose A3b is the most stable binding mode of HTS07940 to FP-2 (Appendix B, Fig. B5). Fig. 3.3-A depicts the predicted interface of the FP-2:HTS07940 complex, obtained after clustering a 100 ns MD simulation of the most stable bound conformation (A3b). In addition, the binding mode of HTS08262 to FP-2, which was inferred from that of HTS07940, is shown in Fig. 3.3-B. The bound conformations of both compounds into the FP-2 active site fulfill previously-described determinants for the interaction with this enzyme. The condensed aromatic rings are close to the indole group of the conserved residue W206, thus allowing the formation of π-π stacking interactions characteristic of C1 cysteine protease complexes.[126] Furthermore, the 2,3-dihydrobenzofuran-5-yl and *p*-methoxyphenyl moieties of HTS08262 and HTS07940, respectively, lie within the S2 subsite, the major specificity determinant of FP-2 and most C1 cysteine proteases,[8, 126, 127] in agreement with preference of FP-2 for hydrophobic (aliphatic and aromatic) amino acid side chains at P2 position.[8, 9, 16] Finally, it can be observed that the carbonyl groups of both inhibitors occupy the S1 subsite, consistently with the well-known proneness of electrophilic groups to interact with the catalytic residue (C42).[128]

Figure 3.3. Structures of FP-2 in complex with compounds HTS07940 and HTS8262. A) Three-dimensional (3D) representation of the FP2:HTS07940 complex interface. **B)** 3D representation of the FP2:HTS8262 complex interface. FP-2 subsites are colored according to the legend. The depicted interfaces correspond to the central structures generated by clustering each 100 ns MD trajectory into a single cluster, and considering only the RMSD values for the heavy atoms of the protein interface residues and the ligand.



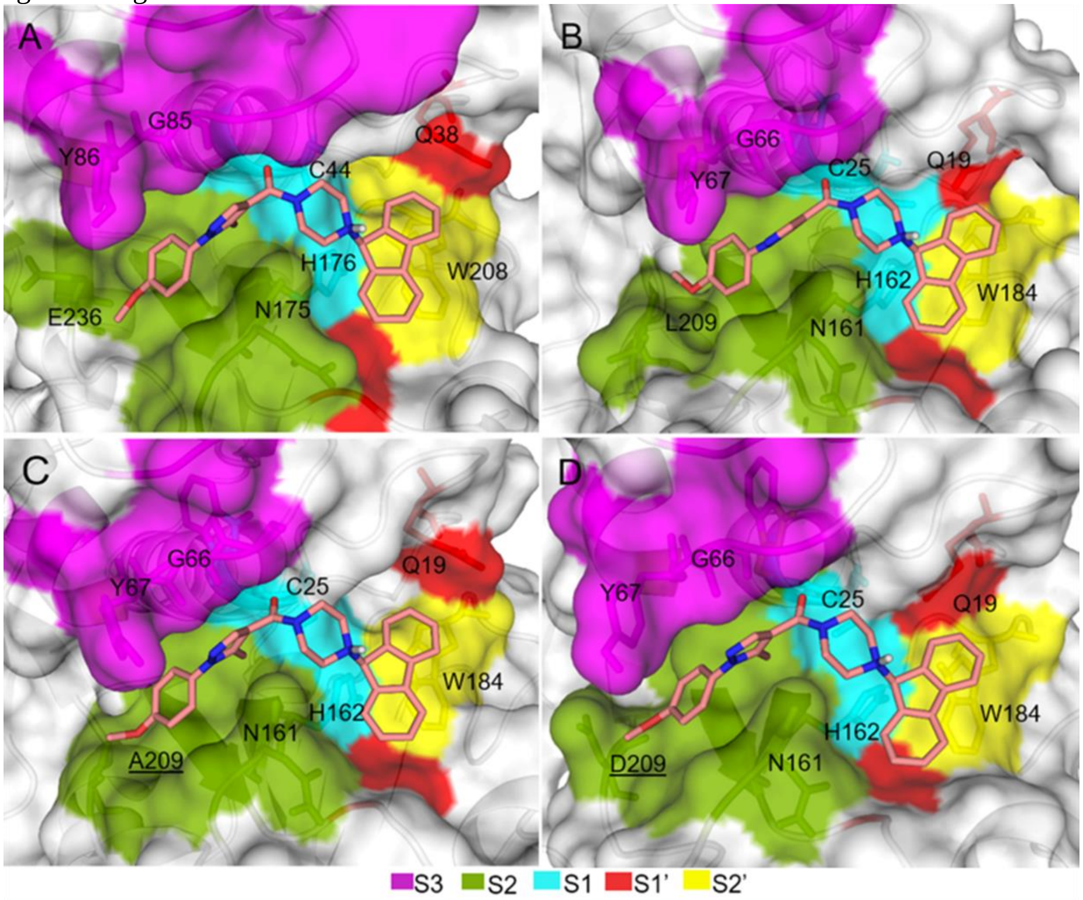
Source: Prepared by the author.

Remarkably, the proposed binding mode of HTS07940 (Fig. 3.3-A) was not found among the docking poses; nor was it obtained after the subsequent 100 ns MD simulation (Appendix B, Figs. B3 and B4). In fact, we resorted to aMD simulations to generate more stable poses (Appendix B, Fig. B5). Further structural and PCA results[50] suggest that HTS07940 can bind to pre-existing free FP-2 conformations, thus following a conformational selection mechanism.[129] The previous conclusion is also extensible to compound HTS08262.[50]

3.1.4 Prediction of the binding modes of HTS07940 to FP-3 and to hCatK

We also studied the interaction of HTS07940 with FP-3 by subjecting pose A3’, which was, in turn, derived from the superposition of pose A3 in complex with FP-2 onto the FP-3 structure, to multiple aMD simulations (Appendix B, Fig. B6). Interestingly, the binding mode to FP-3 predicted from these simulations (pose A3’b) is similar to that obtained for the FP-2:HTS07940 complex. The refinement of pose A3’b through a 100 ns MD simulation led to the central structure represented in Fig. 3.4-A, which was obtained from clustering analysis. A similar approach was followed to predict the structure of the hCatK:HTS07940 complex (Fig. 3.4-B), although some further steps detailed elsewhere were needed in this case (Appendix B, Text B2 and figures cited therein).

Figure 3.4. Structural representation of the interfaces of compound HTS07940 bound to FP-3, hCatK and hCatK mutants. A) FP-3 in complex with HTS07940. **B)** hCaK in complex with HTS07940. **D)** hCatK/L209D in complex with HTS07940. **E)** hCatL in complex with HTS07940. Central structures were generated by clustering each 100 ns MD trajectory into a single cluster. Enzyme subsites are colored according to the legend.



Source: Prepared by the author.

The bottom of hCatK S2 subsite is delimited by the side-chains of residues Y67 and L209, which form a steric wall that is likely to prevent the accommodation of large P2 moieties. Our predictions suggest that the aforementioned ‘wall’ displays a certain degree of flexibility, and that the Y67 and L209 side-chains can act as a gate leading to a distal region of hCatK S2 subsite

(Appendix B, Text B2 and figures cited therein). Similar 'gating' processes involving equivalent residues of the papain-like proteases *Fasciola hepatica* cathepsin L3 and cruzain have also been proposed.[130, 131] Nevertheless, the eventual opening of Y67-L209 gate in hCatK has an associated energetic penalty of roughly 2.6 kcal/mol, which decreases the binding affinity for the ligand (Appendix B, Table B3). On the other hand, residues Y86 and E236 of FP-3 adopt a more open conformation in the crystal structure of the enzyme in complex with an organic inhibitor (PDB: 3BWK) (Appendix B, Fig. B9). In fact, E236 side-chain is not oriented towards the S2 subsite as L209 of hCatK does, but bends away to form an H-bond with Y238(OH) in the crystal structure and during 72% of the MD simulation time of the FP-3:HTS07940 complex (Appendix B, Fig. B9).

A close inspection of the predicted hCatK:HTS0940 structure (Fig. 3.4-B) reveals that, even in the open state of the 'gate', L209 is likely to sterically hinder a deeper accommodation of the *p*-metoxyphenyl moiety of the compound into the S2 bottom. In agreement with this observation, some authors have attributed a key role to this residue in preventing the accommodation of bulky groups within the S2 subsite of hCatK.[132] We further investigated this hypothesis by mutating L209 to Ala and Asp, respectively, and by subsequently performing structural and energetic analyses on the mutant variants of hCatK in complex with HTS07940. The rationale behind the L209A mutation was to reduce significantly the steric hindrance by replacing an isobutyl side-chain of Lue by the smaller methyl group of Ala. As for the L209D mutation, we sought to assess the impact of the similarly-shaped but smaller Asp residue. The central structures of hCatK/L209A and hCatK/L209D in complex with HTS07940 are shown in Figs. 3.4-D and E, respectively. Both mutations seem to increase the space available to accommodate the P2 moiety at the S2 bottom of hCatK, as suggested by visual inspection of the predicted structures (Figs. 3.4-D and E). The previous effect is particularly more evident for the hCatK/L209A:HTS07940 complex. According to the results presented so far and the information available in literature,[132-134] small residues in position 205 (papain numbering scheme, position 358 in Appendix B, Fig. B1) may allow the accommodation of bulky moieties at the S2 bottom of the cysteine proteases. Overall, the presence of a Leu residue at the bottom of the S2 subsite of hCatK could be one of the main factors determining the lower affinity of this enzyme for HTS07940 and HTS08262 with respect to that of FP-2 (Table 3.2).

3.1.5 Calculation of the MM-GBSA free energies for the studied complexes

MM-GBSA free energy calculations for the studied complexes were subsequently conducted using two GB models (Table 3.3). Of note, the method correctly predicted the higher affinity of HTS08262 for FP-2 with respect to that of HTS07940 (Tables 3.2 and 3.3), and the calculated $\Delta\Delta G_{eff}$ value was close to the experimental measurement (-0.67 kcal/mol and $\Delta\Delta G_{exp}$=-0.87 kcal/mol, respectively). The inclusion of the configurational entropy energy contribution (-$T{\cdot}\Delta S_{conf}$) does not lead to a qualitative change of the predictions. However, the calculated relative affinity is now overestimated and reaches an unrealistic value of 3.1 or 3.4 kcal/mol, depending on the GB model (Table 3.3). This result is not surprising in light of previous reports showing that the inclusion of entropy in MM-GBSA calculation may undermine the correlation between experimental and calculated free energies.[81, 135, 136]

The free energy calculations were also carried out for FP-3 and hCatK complexed with HTS07940 (Table 3.3). According to the results obtained with both GB models, the compound displays a lower affinity for FP-3 than for FP-2. The previous conclusion holds even when the -$T{\cdot}\Delta S_{conf}$ term is included in the calculations. Therefore, FP-2 is predicted to be a more important target of HTS07940 in *P. falciparum* than its close homolog FP-3. Moreover, the MM-GBSA results reveal that FP-2 binds to HTS07940 more tightly than its off-target hCatK, with relative affinities ranging from 0.9 to 2.7 kcal/mol (Table 3.3). Remarkably, the previous results are consistent with the experimental relative affinities of HTS07940 for hCatK and FP-2. (>1.2 kcal/mol favorable to the latter enzyme, see selectivity index in Table 3.2).

Finally, we calculated the affinities of HTS07940 for two hCatK mutants employing the same MM-GBSA approach (Table 3.3). Higher affinities of the compound for the mutated variants when compared to the native enzyme were predicted. Therefore, the substitution of L209 by smaller residues (Ala or Asp) improved the ligand binding, as suggested by the structural analysis presented before. These results are also in agreement with other authors that have pointed out the important role of this residue in hampering the accommodation of bulky moieties into the S2 subsite.[132] It is worth noting that previous works have also demonstrated that the double mutant hCatK/Y67L/L209A was able to accommodate Phe at the S2 subsite.[133, 134] However, the single-point mutants hCatK/L209A and hCatK/Y67L were not studied those works, thereby leaving the role of the above-mentioned residues in an individual fashion as an open question. In this regard,

Table 3.3 MM-GBSA free energy values for the studied complexes

Energy components[a] (kcal/mol)	Complex					
	FP-2:HTS08262	FP-2:HTS07940	FP-3:HTS07940	hCatK:HTS07940	hCatK/L209A: HTS07940	hCatK/L209D: HTS07940
ΔG_{npol}[b]	-55.21	-54.94	-54.98	-54.12	-55.66	-57.01
ΔG_{pol} (GB*n*)[c]	17.76	18.16	20.10	20.06	18.75	20.20
ΔG_{pol} (GB*n2*)[c]	18.45	19.20	21.09	20.34	19.86	20.75
$-T\cdot\Delta S_{conf}$[d]	21.3(0.4)[i]	23.6(0.4)	24.2(0.4)	22.6(0.3)	-23.6(0.4)	-23.5(0.4)
ΔG_{eff} (GB*n*)[e]	-37.45(0.05)	-36.78(0.08)	-34.88(0.05)	-34.06(0.04)	-36.91(0.06)	-36.81(0.05)
ΔG_{eff} (GB*n2*)[e]	-36.76(0.04)	-35.74(0.08)	-33.89(0.05)	-33.78(0.04)	-35.80(0.06)	-36.26(0.05)
ΔG_{bind} (GB*n*)[f]	-16.2(0.4)	-13.1(0.4)	-10.7(0.4)	-11.5(0.3)	-13.3(0.4)	-13.3(0.4)
ΔG_{bind} (GB*n2*)[f]	-15.5(0.4)	-12.1(0.4)	-9.7(0.4)	-11.2(0.3)	-12.2(0.4)	-12.8(0.4)
$\Delta\Delta G_{eff}$ (GB*n*)[g]	-0.67(0.09)	-	1.90(0.09)	2.72(0.09)	-2.85(0.07)	-2.75(0.06)
$\Delta\Delta G_{eff}$ (GB*n2*)[g]	-1.02(0.09)	-	1.85(0.09)	1.96(0.09)	-2.02(0.07)	-2.48(0.06)
$\Delta\Delta G_{bind}$ (GB*n*)[h]	-3.1(0.6)	-	2.4(0.6)	1.6(0.5)	-1.8(0.5)	-2.1(0.5)
$\Delta\Delta G_{bind}$ (GB*n2*)[h]	-3.4(0.6)	-	2.4(0.6)	0.9(0.5)	-1.0(0.5)	-1.6(0.5)

[a] All energy components were calculated based on 100 ns MD simulations of the complexes.

[b] Non-polar free energy calculated as follows: $\Delta G_{npol}=\Delta E_{vw}+\Delta G_{SASA}$, where ΔE_{vw} and ΔG_{SASA} are the variations of the van der Waals energy and of the surface tension, respectively, associated to the ligand binding.

[c] Polar free energy calculated as follows: $\Delta G_{npol}=\Delta E_{el}+\Delta G_{GB}$, where ΔE_{el} and ΔG_{GB} are the variations of the electrostatic energy in vacuum and of the polar solvation energy calculated with the indicated GB models, respectively.

[d] Contribution of configurational entropy at T=298 K to the binding free energy calculated from Normal Mode Analysis.

[e] Effective free energy calculated by the following equation: $\Delta G_{eff}=\Delta G_{pol}+\Delta G_{npol}$. The ΔG_{pol} term was calculated using the GB*n* and GB*n2* models.

[f] Binding free energy calculated as follows: $\Delta G_{bind}=\Delta G_{eff}+(-T\cdot\Delta S_{conf})$

[g] Relative effective free energy calculated as follows: $\Delta\Delta G_{eff}=\Delta G_{eff}$(X)-$\Delta G_{eff}$(FP-2:HTS07940), where X stands for the FP-2:HTS08262, hCatK:HTS07940 and hCatL:HTS07940 complexes. The $\Delta\Delta G_{eff}$ values for the complexes involving hCatK mutants were calculated with respect to that of native hCatK, i.e., $\Delta\Delta G_{eff}=\Delta G_{eff}$(Y)-$\Delta G_{eff}$(hCatK:HTS07940), where Y stands for hCatK/L2049:HTS07940 and hCatK/L209D:HTS07940, respectively.

[h] Same as before ([g]), but using the ΔG_{bind} values instead of the ΔG_{eff} ones.

[i] *SEM*s written between parentheses.

Source: Prepared by the author.

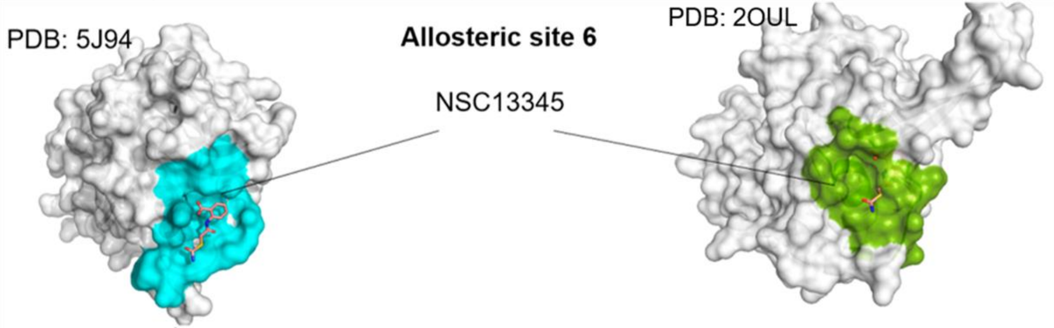
the predicted affinities of HTS07940 for both hCatK mutants were higher than that for FP-2 (Table 3.3). Consequently, residue Y67 of hCatK does not seem to impair the accommodation of the P2 moiety (keep into account that FP-2 bears a Leu residue at the equivalent position; see Appendix B, Fig. B1, aligned position 107).

3.2 Identification of non-competitive FP-2 inhibitors

3.2.1 *In silico* identification of potential allosteric inhibitors of FP-2 targeting site 6

Previous experimental results involving hCatK has revealed that site 6 is an attractive site to target when seeking for allosteric inhibitors of the enzyme.[25, 27] We observed the occurrence of a cavity in the FP-2 crystal structure (PDB: 2OUL) lying in an equivalent position to that of hCatK site 6 (Fig. 3.5). Subsequent MD simulations with Gromos 54A8 force-field showed that the site 6 region in free FP-2 adopts distinct conformations, as can be noticed from the cavity shapes depicted in Fig. 3.6. An in-depth analysis revealed that the different conformations of site 6 arise from fluctuations of the flanking loops (Appendix B, Fig. B11). On the contrary, the MD simulations carried out with ff14SB force-field predict basically a single conformation for the site 6 region of FP-2 (Fig. 3.7), as the flanking loops remain stable during the whole trajectories (Appendix B, Fig. B12).

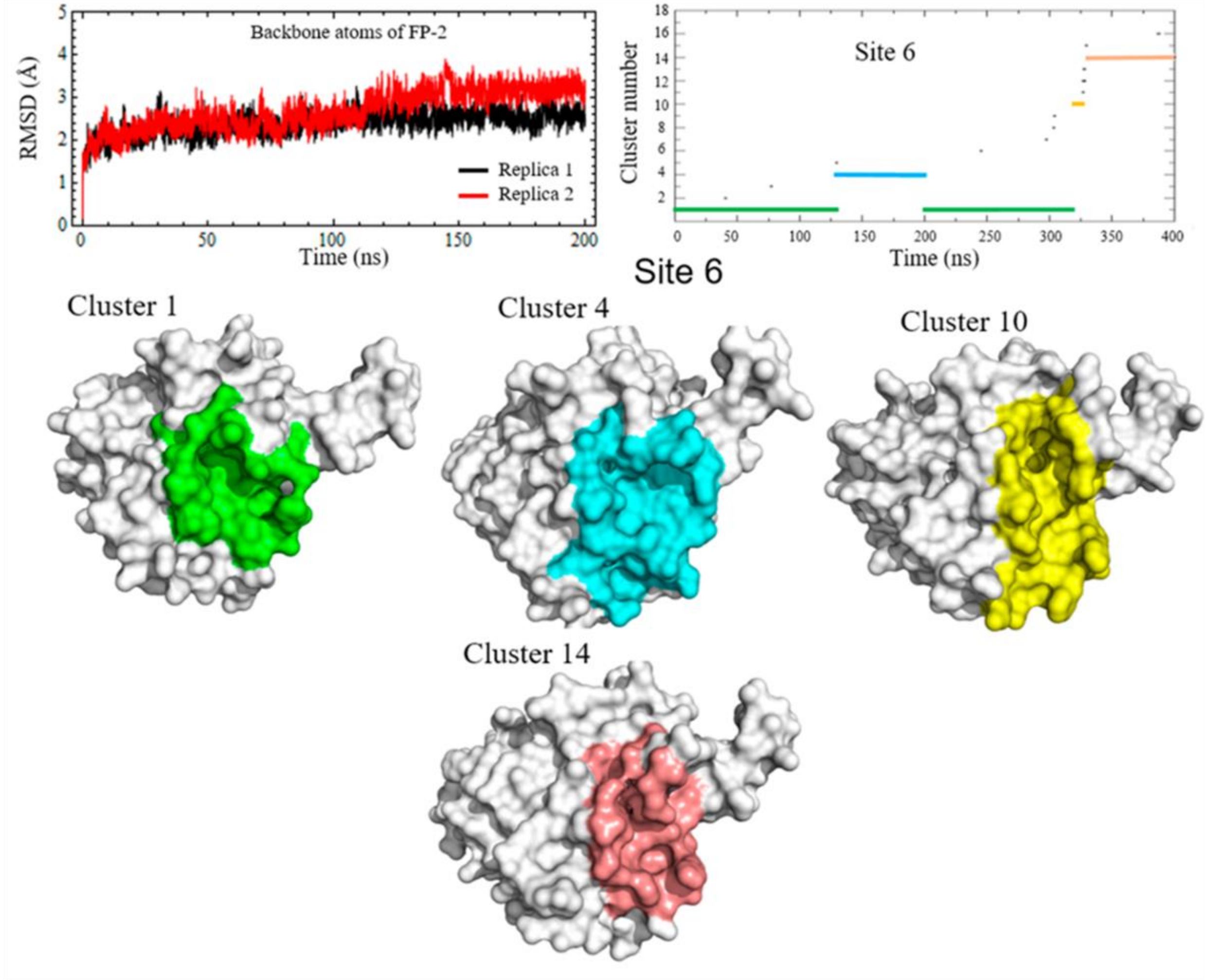
Figure 3.5. Mapping of hCatK site 6 onto the crystal structure of FP-2. The hCatK:NSC13345 complex (PDB: 5J94, left) was structurally aligned with the FP-2 crystal structure (PDB: 2OUL,right). The equivalent regions corresponding to site 6[12] are colored in cyan (hCatK) and green (FP-2).



Source: Prepared by the author.

The central structures depicted in Figs. 3.6 and 3.7 were used to conduct SBVSs aiming to find compounds with the ability to bind site 6. By following a strategy that included various refinement steps after the SBVSs, we selected 18 compounds on the basis of their stability during their respective 300 ns MD simulations and the calculated ΔG_{eff} values (Table 3.4). Finally, 12 compounds were purchased to conduct experimental validation (Table 3.4). These compounds were selected by taking into account their cost, availability, structural diversity and predicted affinities.

Figure 3.6: Conformational variation of putative allosteric site 6 of FP-2 during GROMOS 54a8 MD simulations. Two replicate 200 ns MD simulations were conducted for the free enzyme. Four main structural clusters were found for site 6.

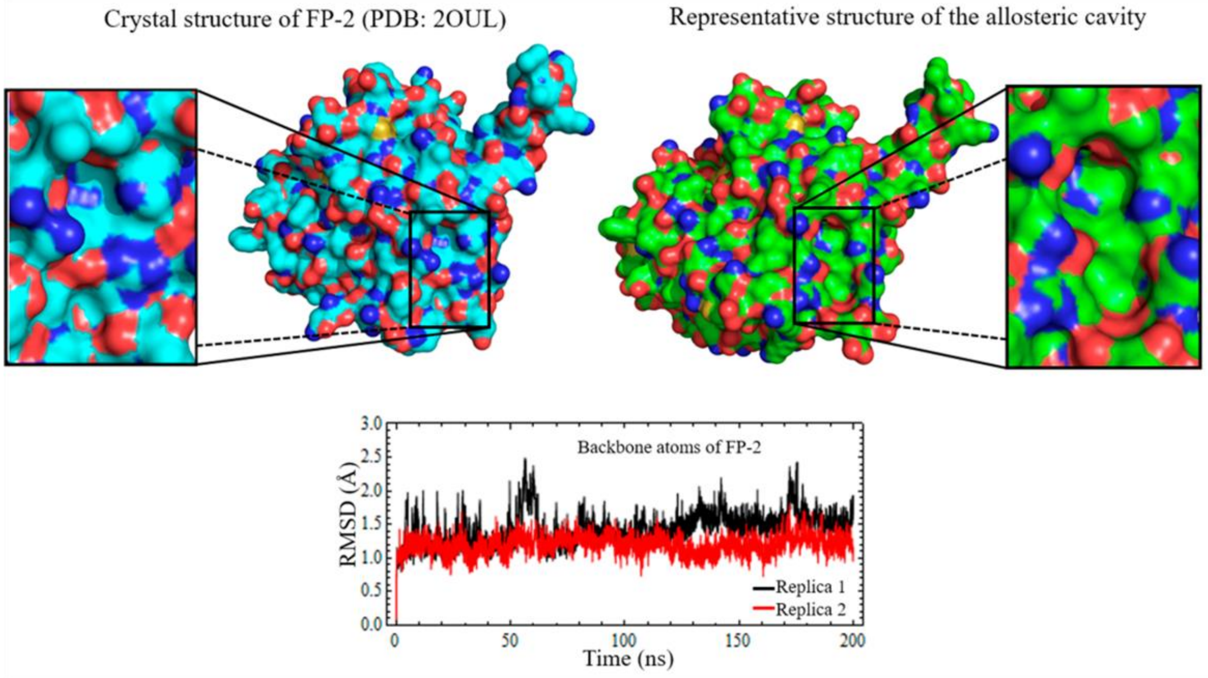


Source: Prepared by the author.

3.2.2 Experimental characterization of the potential allosteric inhibitors of FP-2.

Six compounds showed inhibition of FP-2 activity in a concentration-dependent fashion during the secondary screening (Fig. 3.8). The relative potency of the compounds was rather low and their insolubility at high concentrations precluded, in most cases, the obtaining of complete IC_{50} curves. In addition, we observed the formation of aggregates in the wells when conducting the activity assays of several compounds, which can create interference in the fluorescence measurement and, consequently, lead to spurious results. Therefore, a centrifuge counter-screen was carried out to assess the impact of aggregation on the fluorescence measurements. As shown in Fig. 3.9, only two compounds, ZINC03225317 and ZINC72290660, preserve their inhibitory activity after the centrifugation step. For the remaining compounds, the artifacts created by aggregates during the fluorescence measurements precluded further kinetic characterization of their inhibitory activity and mechanism. Due to these issues, it was not possible to determine whether the discarded compounds could exert inhibitory effects on FP-2, apart from inducing aggregation.

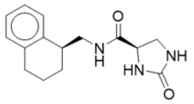
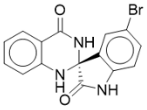
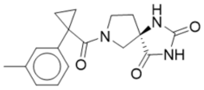
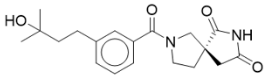
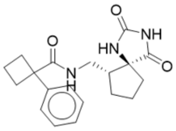
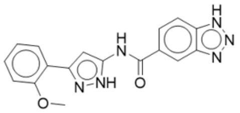
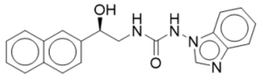
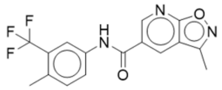
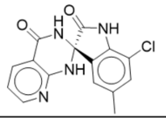
Figure 3.7. Conformational variation of putative allosteric site 6 of FP-2 during Amber ff14SB MD simulations. Two replicate 200 ns MD simulations were conducted for the free enzyme. Only one major cluster was found for site 6 during these simulations.

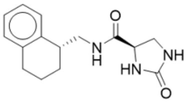
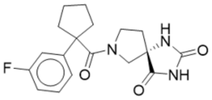
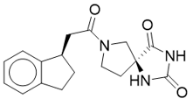
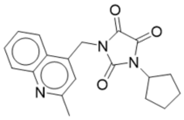
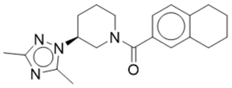
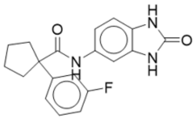
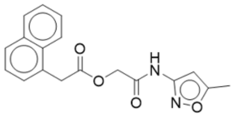
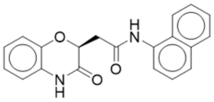
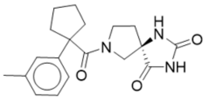


Source: Prepared by the author.

Compounds ZINC03225317 and ZINC72290660 were subjected to further kinetic characterization to determine their inhibition mechanisms. Figs. 3.10 and 3.11 show the graphs obtained after conducting inhibition assays at various concentrations of the inhibitors. Lineweaver-Burk plots and direct fit of the Michaelis-Menten hyperboles to different inhibition models demonstrate that both compounds behave as mixed inhibitors. This term is used for non-competitive inhibitors with a finite alpha (α), and $\alpha \neq 1$, with α being equal to the ratio of the dissociation constant between the enzyme:inhibitor complex and the substrate (K'_S) to that of the

Table 3.4. Compounds selected after refinement of SBVSs against different conformations of FP-2 site 6

Identifier	Structure	ΔG_{eff}^{a} (kcal/mol)
ZINC89858266[b]		-26.47
ZINC03225317		-31.07
ZINC72290717		-27.56
ZINC77455016		-30.06
ZINC84525109		-25.47
ZINC58223135		-23.20
ZINC67455750		-35.47
ZINC42649291		-30.58
ZINC84128950		-26.04

Identifier	Structure	ΔG_{eff}^{a} (kcal/mol)
ZINC89858270		-25.78
ZINC72290660		-27.96
ZINC72290743		-33.45
ZINC69949197		-26.87
ZINC72235263		-35.59
ZINC40159524		-33.81
ZINC03508120		-30.86
ZINC05069514		-34.07
ZINC72290621		-25.74

[a] The free energy values were calculated for the respective 300 ns trajectories of the complexes using the GB*n2* implicit solvation model.

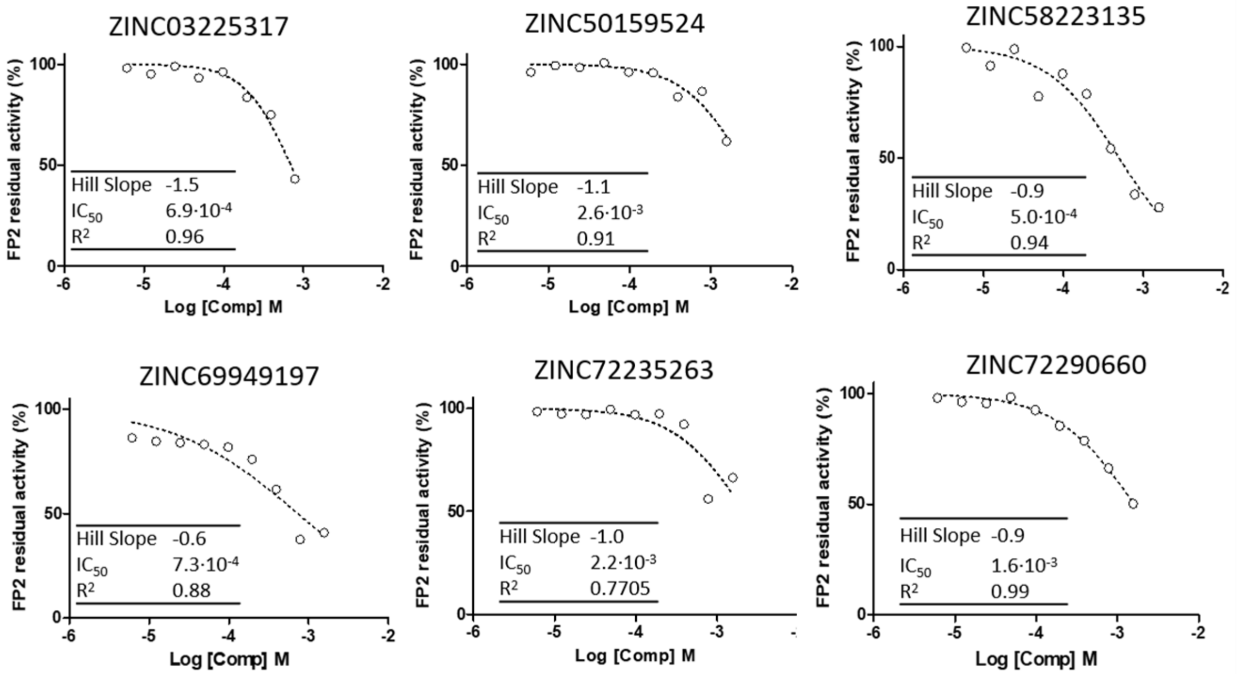
[b] Underlined compounds were purchased for subsequent experimental validation.

Source: Prepared by the author.

free enzyme and the substrate (K_s), i.e., $\alpha = K'_s / K_s$.[137] Poor quality fits were obtained when using competitive and uncompetitive models (Figs. 3.10 and 3.11).

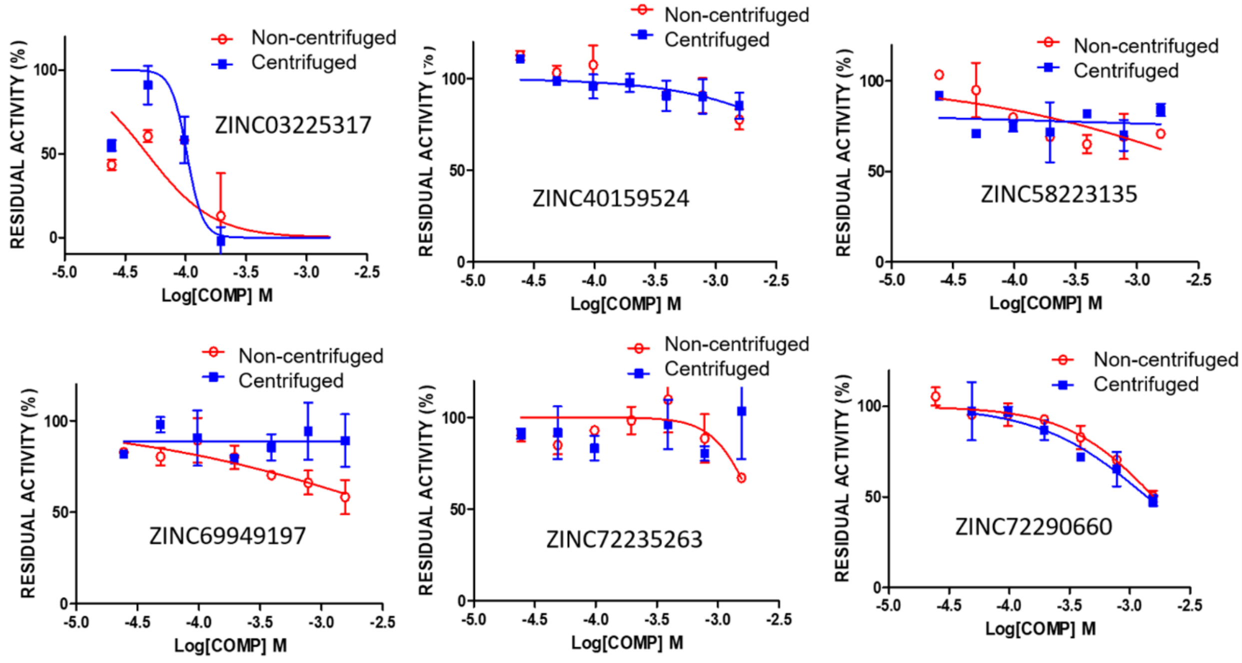
The calculated K_i values for the two active non-competitive inhibitors are in the millimolar-submillimolar range, which indicates that they are weak binders of FP-2 (Table 3.5). However, these hits are a relevant proof of concept that non-competitive inhibition of FP-2 can be achieved

Figure 3.8. Dose-response curves for the compounds displaying inhibitory activity against FP-2. Experimental data were fitted to eq. 2.33. For each compound, the estimated values of Hill Slope, IC_{50} and R^2, which measures the goodness of fit, are provided.



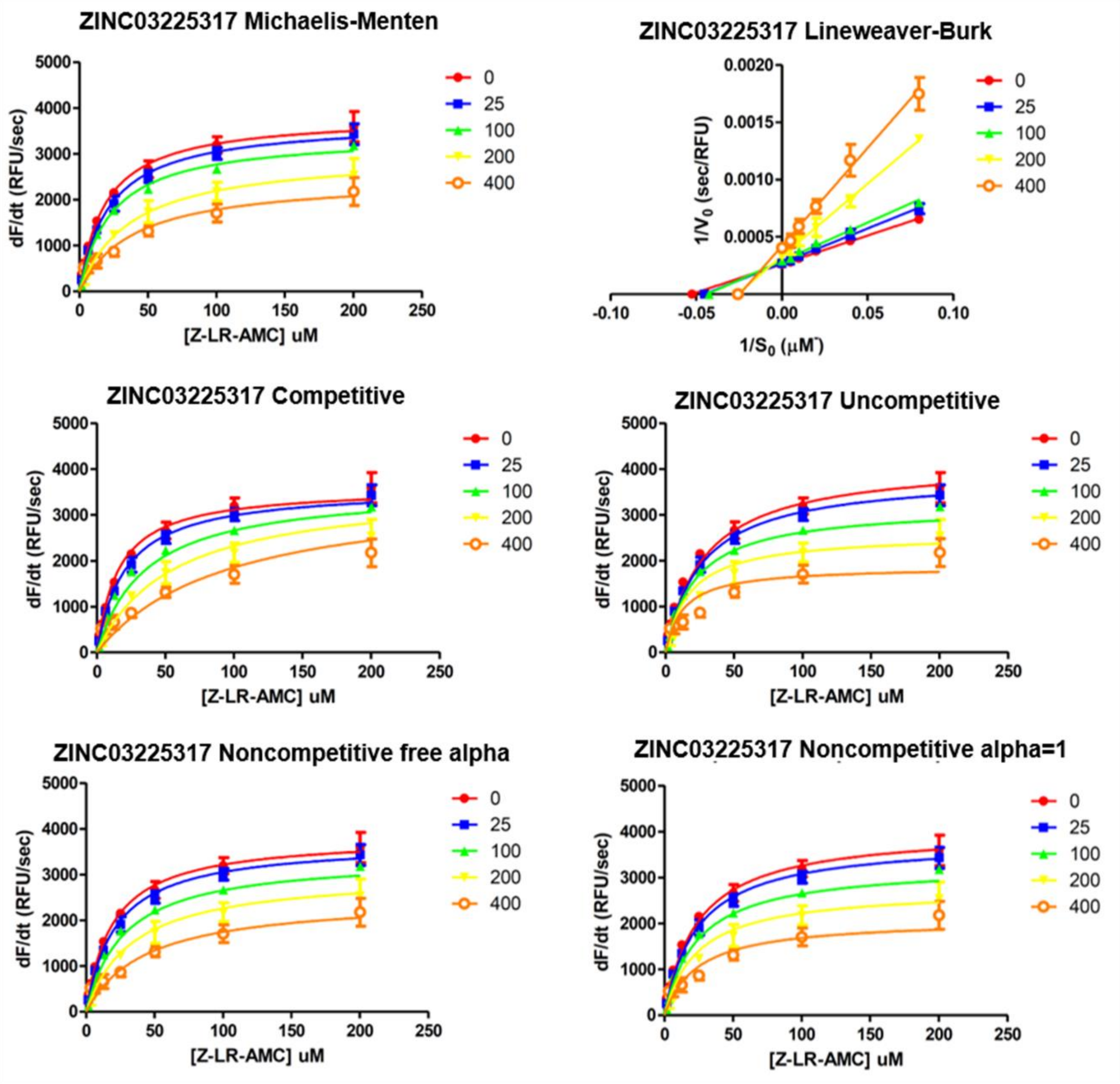
Source: Prepared by the author.

Figure 3.9. Dose-response curves of the selected compounds obtained from the centrifuge counter-screen against FP-2. Error bars correspond to standard deviations of two independent measurements.



Source: Prepared by the author.

Figure 3.10. Determination of the inhibition mechanism of ZINC03225317 against FP-2. Curves corresponding to the compound concentrations indicated beside each graph are colored accordingly. Four different concentrations of the inhibitor were assayed in addition to the non-inhibited system. Error bars obtained from triplicate measurements are added to each point.

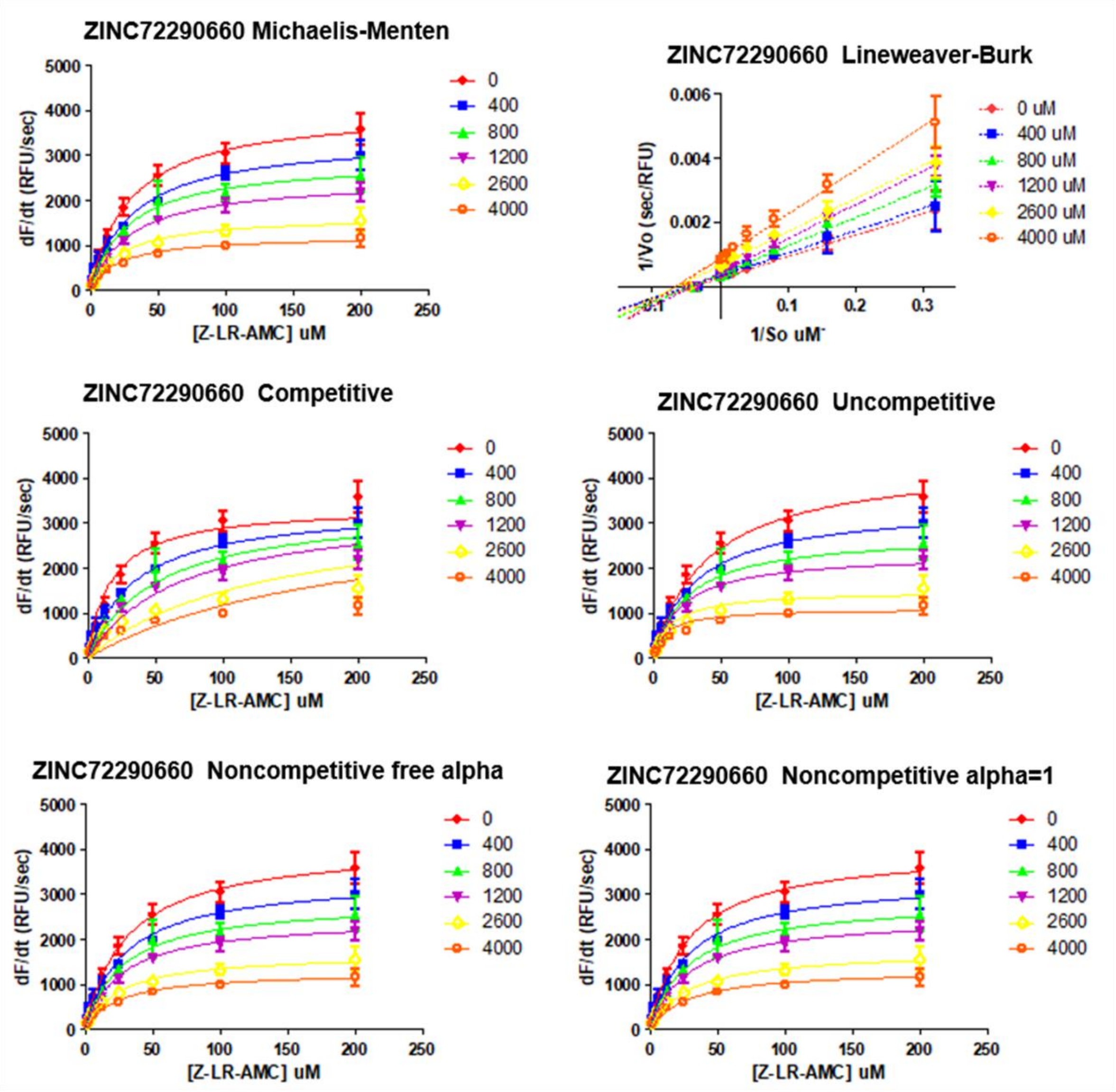


Source: Prepared by the author.

by targeting site 6. Further structural optimization would be required in order to improve the affinity of the ligands. In addition, the α value of ZINC03225317 indicates that its binding to the enzyme reduces the affinity of the latter for the substrate. Therefore, this compound is expected to induce some changes in the active site leading to the previous effect. An opposite impact on the affinity for the substrate arises from the binding of ZINC72290660. However, the α value in this case is close to 1 (α =0.7); hence, the compound act nearly as a classical noncompetitive inhibitor.

In fact, through visual inspection one can see that the non-competitive model with α= 1 fits well the experimental data corresponding to ZINC72290660 (Fig. 3.11).

Figure 3.11. Determination of the inhibition mechanism of ZINC72290660 against FP-2. Curves corresponding to the compound concentrations indicated beside each graph are colored accordingly. Four different concentrations of the inhibitor were assayed in addition to the non-inhibited system. Error bars obtained from triplicate measurements are added to each point.



Source: Prepared by the author.

3.2.3 Structures of FP-2 in complex with the active non-competitive inhibitors

The structures of FP-2 in complex with ZINC03225317 and ZINC72290660 were predicted as described in section 2.1.2. The central structures of each complex calculated from the respective

Table 3.5. ***K_i*** **and alpha values for the noncompetitive inhibitors of FP-2**

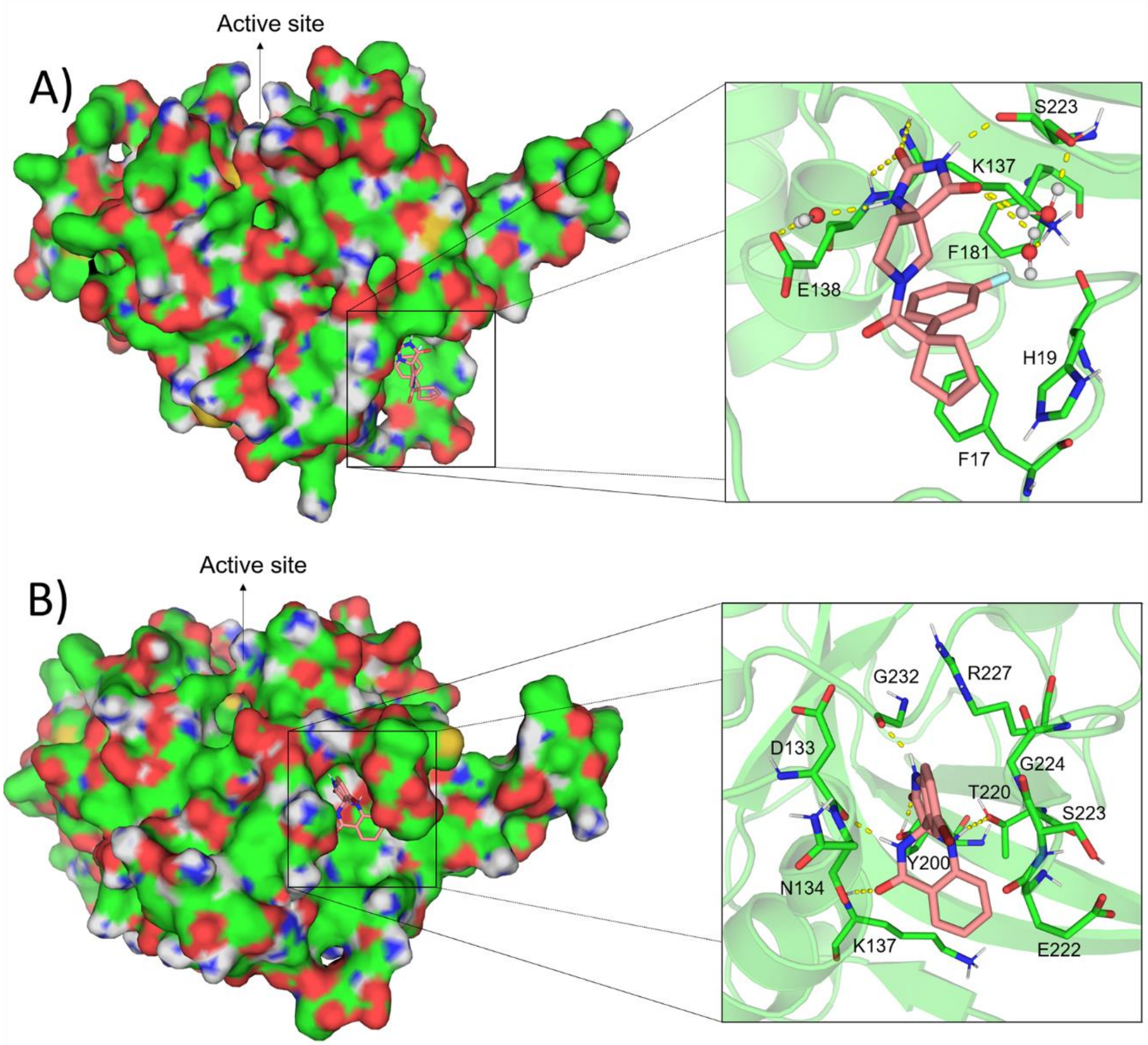
Compounds	K_i (mM)	A	R^2 [a]
ZINC72290660	2.6 (0.8)[b]	0.7 (0.2)	0.98
ZINC03225317	0.16 (0.02)	4 (1)	0.97

[a]Non-linear regression coefficient used to quantify the goodness of fit.

[b]Standard deviations are shown in parentheses.

Source: Prepared by the author.

Figure 3.12. Predicted structures of FP-2 in complex with two non-competitive inhibitors. A) FP-2 in complex with ZINC72290660. **B)** FP-2 in complex with ZINC03225317. FP-2 is shown in surface representation on the left. The allosteric site region is enclosed in a rectangle and the active site is indicated with an arrow. A more detailed representation of the allosteric site interface and the H-bonds (yellow dashed lines) formed between the inhibitor (salmon sticks) and the protein interacting residues (green sticks) are provided on the right. The (*S*) ZINC72290660 and (*R*) ZINC03225317 enantiomers of the compounds are depicted.



Source: Prepared by the author.

300 ns MD simulations are shown in Fig. 3.12. As can be seen in the figure, both compounds bind different conformations of the broad region that was initially termed site 6. In particular, ZINC72290660 was identified in the SBVS against the central structure of site 6 determined from the MD simulations carried out with ff14SB force-field (Fig. 3.7); whereas ZINC03225317 was identified from the SBVS against the largest cluster (cluster 1) of the Gromos 54a8 MD simulations (Fig. 3.6). Both compounds establish various H-bonds with FP-2 residues that contribute to the complex stability (Fig. 3.12). Remarkably, some water bridges were also detected at the interface of the FP-2:ZINC72290660 complex (Fig. 3.12-A).

As can be noted from Table 3.4, each of the active compounds, ZINC72290660 and ZINC03225317, has a chiral center and was purchased as racemic mixture, according to the information provided by the vendor. However, each enantiomer is not expected to bind FP-2 with the same affinity, given the asymmetric shape of the binding sites. It is worth noting that these compounds were analyzed in the enantiomeric forms that occur in ZINC12 database, which are (*S*) ZINC72290660 and (*R*) ZINC03225317. Additionally, the former compound could have two different conformations with respect to the amide bond (-CO-NH-), which must be analyzed independently, as Autodock-Vina freezes rotations around amide (=peptide) bonds by default. We applied the workflow devised for pose refinement outlined in Section 2.1.2 to all other possible conformations around the amide bond and enantiomeric forms of the previous compounds (data not shown). Our results indicate that the enantiomers shown in Fig. 3.12 are by far the most stable in terms of ΔG_{eff} values of the energy-minimized poses. This, in turn, suggests that a two-fold decrease in the true K_i values of ZINC72290660 and ZINC03225317 can be expected, as half of their molecules are inactive in the racemic mixture. For ZINC72290660, the occurrence of two conformations around the amide bond with likely different relative abundances further reduces the amount of the active forms of the compound in solution, thus leading to a weaker apparent affinity.

3.3 Discussion

We have successfully identified two promising FP-2 inhibitors, HTS07940 and HTS08262, capable of arresting the growth of *P. falciparum* trophozoites *in vitro* and displaying low toxicity for HeLa cells. These compounds were selected from the results of SBVSs against the active sites of FP-2 and FP-3. The *in silico* strategy employed here sought to improve the selectivity of the hits by excluding the ligands with potentially high affinities for hCatK. HTS07940 and HTS08262 possess a common scaffold comprising a (4-(9H-fluoren-9-yl) piperazin-1-yl) methanone moiety attached to a variable substituent whose core structure is made up of two aromatic rings with five and six members each and linked together by a single covalent bond. Interestingly, the common (4-(9H-fluoren-9-yl) piperazin-1-yl) methanone moiety is present in Genz10850, a compound previously characterized as an inhibitor of PfENR and of *P. falciparum* growth *in vitro*.[125] However, the remaining structural features of the compounds identified here and Genz10850 are different, as the latter bears an indole group instead of the core structure described for the former two.

Due to the structural similarities shared by Genz10850 and the FP-2 inhibitors, the possibility that HTS07940 might also inhibit PfENR was raised. On the basis of MD simulations, we proposed that the a putative PfENR:HTS07940 is likely to be stable, a conclusion also valid, in principle, for HTS08262. Hence, these two compounds could halt the *P. falciparum* growth *in vitro* by inhibiting two enzymes involved in orthogonal metabolic pathways, i.e., the hemoglobin hydrolysis (FP-2) and the biosynthesis of fatty acids (PfENR).[9, 138] Remarkably, PfENR has been described as a promising drug target for the discovery of selective antimalarials, as it is involved in a metabolic pathway not found in humans.[125] Further experimental assays will be required to assess the ability of the compounds identified here to inhibit PfENR.

Interestingly, the 9H-fluoren-9-yl moiety present in HTS07940 and HTS08262 can be found in previously-identified FP-2 inhibitors, such as compound 19 (IC_{50}=44.94 μM) reported by Shah *et al.*[128] Compound 19 also bears a piperazinyl ring, which occurs in various FP-2 inhibitors.[8] Note, however, that this substituent is not directly bonded to the 9H-fluoren-9-yl moiety as in the compounds reported here. Apart from the abovementioned similarities, compound 19 differs appreciably from HTS07940 and HTS08262. Overall, it can be concluded that the 9H-fluoren-9-yl moiety is a promising structural feature for the interaction with FP-2, especially with residue W206, something first highlighted by us elsewhere.[50]

Our results showed that the IC_{50} values of HTS07940 and HTS08262 against *P. falciparum* cultures are similar to those measured against FP-2. This seems counterintuitive, as the compound bioavailability is likely to be reduced inside the parasite food vacuoles, due to the diffusion to other parts of the cell and the presence of membranes, which act as barriers. However, the high efficiency displayed by the compounds can stem from different factors. For example, the interaction with multiple parasitic targets, such as FP-3 and/or other cysteine proteases, and PfENR. Furthermore, HTS07940 and HTS08262 have amine groups that become protonated at the acid pH of the food vacuole. Interestingly, it has been reported that positively-charged compounds can accumulate in acidic cell bodies, such as lysosomes and reservosomes.[139] This phenomenon can increase the potency of the identified FP-2 inhibitors in the live parasites.

The structural and energetic analyses carried out here suggest that the presence of bulky residues at the bottom of the S2 subsite of cysteine proteases can be detrimental to the affinity for HTS07940. This prediction is in agreement with previous experimental evidence.[132-134] Remarkably, we observed that L209D and L209A mutations in hCatK led in both cases to increased affinities for the compound. These predictions provide valuable insight for the design of more selective inhibitors against FP-2 by exploiting the structural differences displayed by this enzyme and human cysteine cathepsins at positions 67 and/or 205 (papain numbering scheme).

On the other hand, we reported the identification of two non-competitive inhibitors, ZINC03225317 and ZINC72290660, by conducting SBVSs against site 6 of FP-2. The low K_i values of the hits (millimolar and submillimolar range) indicate that they are not suitable as drug candidates and, thus, further characterization in *P. falciparum* cultures was deemed unnecessary. However, our discovery reinforces the possibility of inhibiting FP-2 in an allosteric fashion, in agreement with previous works.[34-38] Additionally, on the basis of our results, site 6 is proposed as an attractive cavity to target when designing allosteric inhibitors against FP-2. The fact that hCatK can also be inhibited by compounds that bind site 6, which initially inspired our search in the equivalent site of FP-2, indicates that allosteric modulation follows similar patterns in members of the same protein family. Indeed, this property constitutes the basis of the computational analysis employed by Novinec *et al.* to propose potential allosteric cavities in hCatK.[25]

Interestingly, both active non-competitive inhibitors were predicted to bind different conformations of site 6. According to the predicted binding mode of ZINC03225317, this compound lies closer to the enzyme active site than ZINC72290660. This can explain why the former was found to exert a more appreciable effect on the active site conformation, measured in terms of the parameter α, than the latter. Therefore, the conformational variability of site 6 can lead to appreciable structural differences in the ligands targeting this site and, in turn, to dissimilar non-competitive inhibition mechanisms.

Our work constitutes the first attempt to identify allosteric inhibitors of FP-2 by using an SBVS approach. The small number of hits obtained through this strategy (2 out of 12 purchased compounds) underlines the intrinsic difficulties of identifying compounds targeting allosteric sites with conventional computational techniques usually employed to predict orthosteric ligands. Of note, when dealing with allosteric compounds, one cannot establish a direct correlation between the binding affinity and the inhibition potency, as the binding of these compounds does not necessarily block the enzymatic activity. Indeed, partial inhibition is, in principle possible, as the bound compound might not effectively lead to a sharp decrease in catalytic activity.[25, 140] Even worse, the compounds can still bind the protein without causing any effect on the protein function, thus acting as inactive ligands. All these issues, together with those associated with the characterization of relevant cavity conformations to be targeted, can complicate the discovery of allosteric inhibitors.

4 PROBING A NOVEL ALLOSTERIC SITE OF FALCIPAIN-2 THROUGH THE PREDICTION OF A NON-COMPETITIVE INHIBITOR BINDING MODE

The results presented in this chapter have been published in the *Journal of Physical Chemistry B*.[58]

4.1 Structural analyses

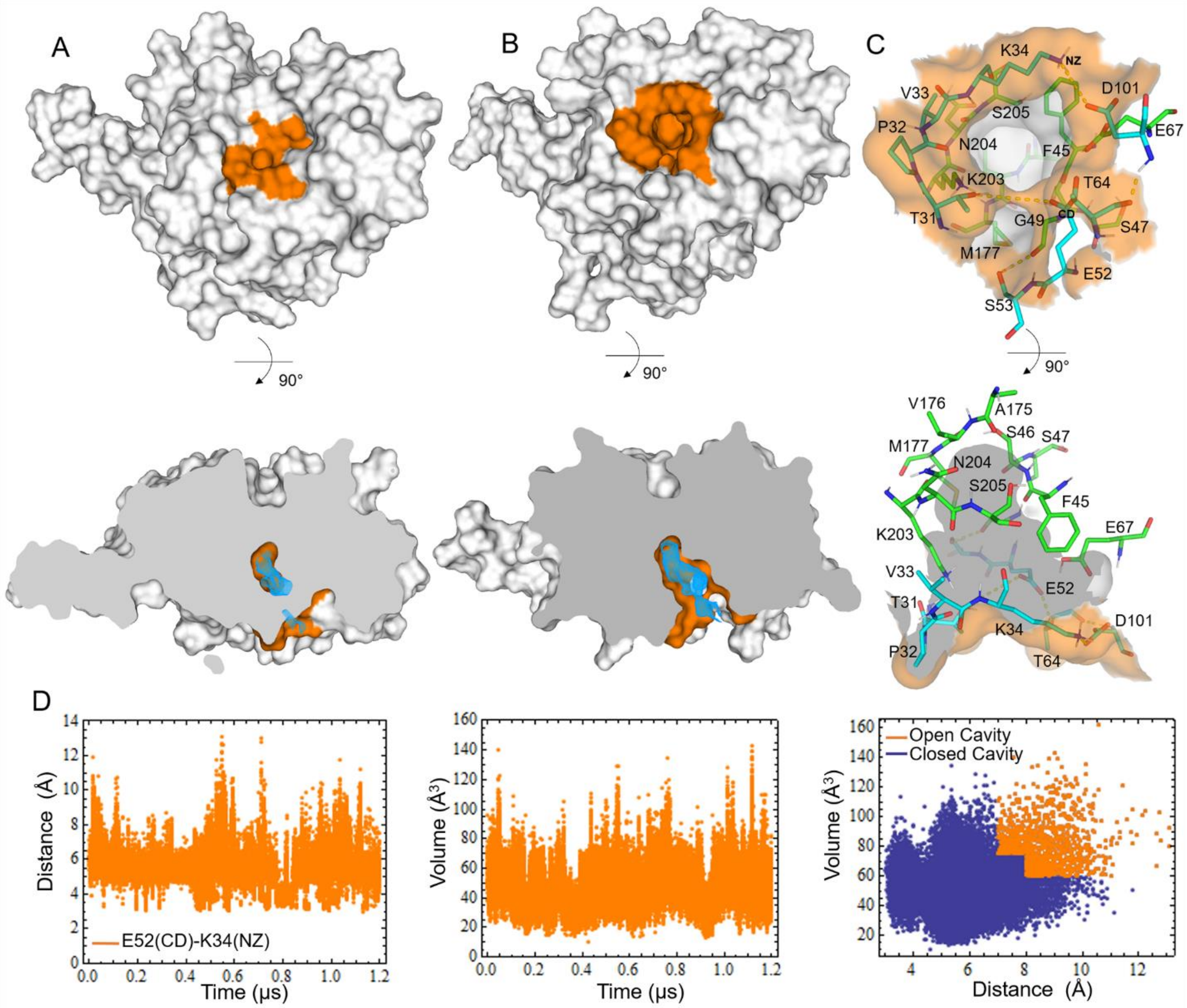
4.1.1 Formation of a transient pocket at the site 3 region of FP-2

Previous experiments have shown that Cpd66 protects loop$_{204\text{-}213}$ from tryptic hydrolysis upon binding to FP-2 (Fig. 1.3). However, it can be observed that there is no cavity in the region enclosing the aforementioned loop in the available FP-2 crystal structures (PDBs: 2OUL, 3BPF, 2GHU, 1YVB and 3PNR). Remarkably, during the MD simulations of free FP-2, a tunnel-like cavity was transiently formed in a region equivalent to site 3 of hCatK,[25] whose inner surface is made up of resides F45, C26, C27, G49, E67, A175, V176, M177, K203, N204 and S205, the latter two being present in loop$_{204\text{-}213}$ (Fig. 4.1). As the formation of the tryptic peptide that is affected by Cpd66 binding requires the cleavage of K203-N204 peptide bond, the presence of the compound in the detected transient pocket could explain the experimental observations.

According to our predictions, the rearrangement of certain residues at the surface of site 3 region unveils the cryptic pocket that remains occluded in the FP-2 crystal structures by opening a gate that leads to the inner space (Figs 4.1-A, B and C). Specifically, we observed that only when the H-bonds and salt bridge between residues E52 and K34 are disrupted, the entrance of the cryptic pocket opens (Figs. 4.1 and C1, Appendix C). Of note, the distance between E52(CD) and K34(NZ) fluctuates within a wide range of values, 3-13 Å (Fig. 4.1-D), which allows the occurrence of multiple opening and occlusion events at the pocket entrance. Residue K34 is likely to act, thus, as a lid that controls the access to site 3 cavity. Apart from the large E52(CD)-K34(NZ) distance fluctuations, we also observed that the pocket inner volume varies appreciably during the MD simulations, from 10 Å^3 to 160 Å^3 (Fig. 4.1-D). The discrimination of the pocket state, i.e., closed or open, was found to depend on the E52(CD)-K34(NZ) distances and the corresponding internal volumes. As a rule of thumb, we established that the pocket is open when the E52(CD)-K34(NZ) distance and the volume are simultaneously larger than 7 Å and 75 Å^3, respectively; or 8 Å and 60

Å^3, respectively (Fig. 4.1-D). Approximately (1.5±0.4)% of the frames collected during the MD simulations possess an open pocket, according the abovementioned geometric criteria. Most of the time, the site 3 entrance remains closed, as observed in the FP-2 crystal structures.

Figure 4.1. Formation of a transient pocket in the site 3 region of FP-2. A) Surface representation of the crystal structure of FP-2 (PDB: 2OUL). **B)** Surface representation of the central structure of FP-2 having an open pocket conformation. In both, A) and B), the surface region corresponding to the transient pocket entrance is depicted in orange. A protein slice after 90° rotation shows in both cases the inner space of the pocket. The pocket internal volume density is represented in blue. **C)** Detailed structural representation of the open pocket conformation. Residues lying at the pocket entrance are depicted as cyan sticks surrounded by a translucid orange surface. Residues forming the pocket internal wall are shown as green sticks. The internal surface of the pocket is colored in gray. H-bonds occurring at the pocket entrance are indicated with yellow dashed lines. **D)** From left to right, graphs showing the E52(CD)-K34(NZ) distance and site 3 internal volume time profiles during the concatenated MD simulations of FP-2, and the volume versus E52(CD)-K34(NZ) distance plot. Dots depicted in orange in the third graph correspond to frames bearing an open pocket.



Source: Prepared by the author.

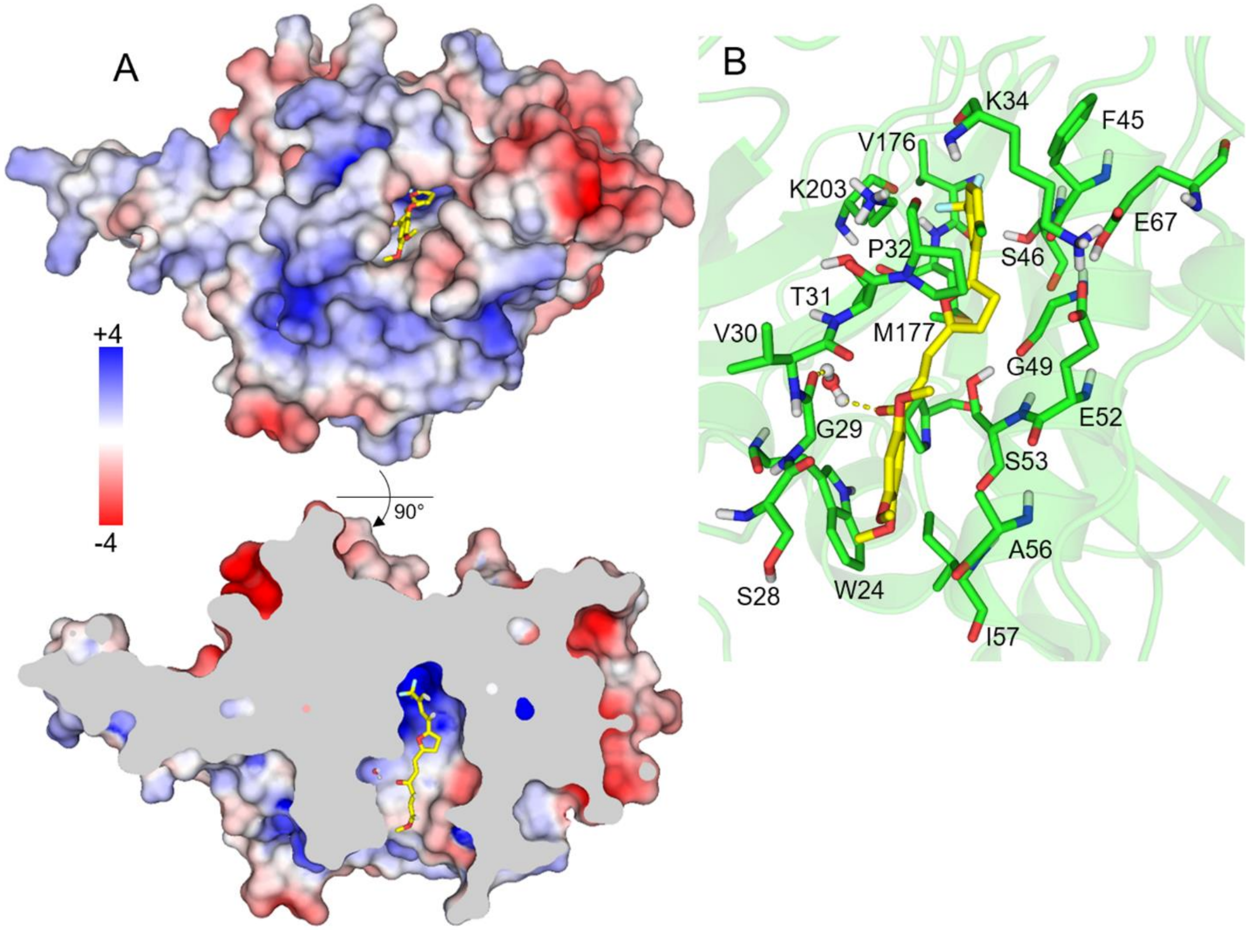
In parallel, we performed similar analyses for FP-3, and the results showed that a cavity in the site 3 region of this enzyme also occurs transiently (Appendix C, Figs. C2-A and C2-B). However, in this case, both, the average pocket volume and the opening probability, are smaller than those of FP-2 (compare Figs. 4.1-D and C2-C and D, Appendix C). Indeed, only (0.5±0.2)% of the FP-3 frames possess an open pocket. Among the various factors underlying the different pocket opening probabilities in both FPs, one can mention the residue composition at the pocket entrance and the motions of neighboring loops. We noted that the FP-3 pocket entrance is made up of bulkier residues, e.g., L66 (T64 in FP-2) and Q103 (D101 in FP-2). Furthermore, other conserved residues, e.g., D37 (D35 in FP-2) and Y108 (Y106 in FP-2), adopt different average conformations in these enzymes, being present in the pocket entrance of FP-3 but not in that of FP-2 (Figs. 4.1-C and C2-B, Appendix C). Remarkably, due to the strong propensity of residues K34 and D101 of FP-2 to form a strong salt bridge and H-bond, the side-chains of these residues lie in a strikingly different position when compared to that of the equivalent FP-3 residues, K36 and Q103 (Figs. 4.1-C and C2-B, Appendix C). This can influence the shape of the pocket entrance in both FPs and may also determine the opening probabilities.

4.1.2 Prediction of Cpd66 binding mode

Cpd66 was initially docked into the site 3 open pocket of FP-2 and energy-minimized poses were rescored using MM-GBSA free energy calculations. The pose yielding the lowest ΔG_{eff} value (Appendix C, Fig. C3) was subjected to subsequent replicate MD simulations (Appendix C, Fig. C4). The most stable binding mode was finally identified by analyzing the ΔG_{eff} values and ligand RMSD values along the MD simulations (Appendix C, Fig. C4-A). Of note, the proposed structure of the FP-2:Cpd66 complex differs appreciably from that predicted through docking, as the compound's main axis became aligned to that of the FP-2 central α-helix during the MD simulation, with the concomitant disruption of T31 and E52 H-bond (Appendix C, Fig. C5).

The central structure of the FP-2:Cpd66 complex is shown in details in Fig. 4.2. As can be observed, the trifluoromethyl group is deeply buried into the positively-charged bottom of site 3 cavity (Fig. 4.2-A). Hence, favorable electrostatic interactions can occur in this region, due to the high electronegativity of the fluorine atoms. In addition, the hydrophobic phenyl ring favors the insertion of the compound into the pocket, mainly formed by aliphatic and aromatic groups (side-chains of residues F46, M177 and the hydrocarbon side-chains of K34 and K203, Fig. 4.2-B). The

Figure 4.2. Predicted binding mode of Cpd66 into the site 3 pocket of FP-2. A) Structural representation of FP-2 bound to Cpd66. The protein surface is colored according to the electrostatic potential, expressed in k_bT/e units, where k_b, T and e stand for the Boltzmann's constant, the temperature (298.15 K) and the electron charge, respectively. The lower view corresponds to a protein slice revealing the internal space of the cavity in the presence of the ligand. **B)** Structural representation of the FP-2:Cpd66 complex interface. Residues contacting the compound at a distance ≤4.0 Å are shown as green sticks. In all cases, the ligand is depicted as yellow sticks. A water molecule mediating a water-bridge interaction between G29(O) and the carbonyl oxygen atom of Cpd66 is shown as spheres and sticks.



Source: Prepared by the author.

propensity of the 2-chloro-4-trifluoromethylphenyl moiety to accommodate into the site 3 cavity may explain why Cpd66 is the only one among the various active *E*-chalcones reported by Bertoldo *et al.*, with the ability to interact with $\text{loop}_{204\text{-}213}$.[36] The remaining compounds are smaller in size, as they have only one aromatic ring or a naphthyl group at the B-ring position; whereas Cpd66 bears two aromatic rings, i.e., the furanyl and the 2-chloro-4-trifluoromethylphenyl moieties, connected through a single covalent bond (Fig. 1.2).[36] As can be straightforwardly deduced, the shorter *E*-chalcones are unable to reach the bottom of site 3 cavity, on the assumption that their common carbonyl groups lie at the pocket entrance as in Cpd66. This, in turn, is likely to be detrimental to the affinity for this pocket.

Incidentally, Cpd48, reported as a mixed inhibitor by Bertoldo *et al.*,[36] was subsequently found to bind the prime region of FP-2 active site through X-ray crystallography.[39] The observed binding mode leaves the catalytic residues exposed in agreement with the proposed inhibition mechanism, while explaining the ability of Cpd48 to protect from hydrolysis a tryptic peptide related to the active site.[36, 39] Moreover, the *E*-chalcones bearing the bulkier naphthyl groups at B-ring position lack halogen substituents that favor the interaction with the positively-charged environment within the site 3 cavity inner space (Fig. 4.2-A). Remarkably, the 2-chloro-4-trifluoromethylphenyl group of Cpd66 remains close to K203 during the MD simulations of the FP-2:Cpd66 complex (Appendix C, Fig. C6), which can be related to the ability of the compound to hinder the formation of the tryptic peptide 204-NSWGQQWGER-213. On the other hand, note that the trimethoxyphenyl substituent of Cpd66 contacts residue W24, thus modifying the native chemical environment of the latter residue. This is likely to lead to the variation in the Trp intrinsic fluorescence emission of FP-2 observed in the experiments.[36] Interestingly, the link between microenvironment disruption and the modification of Trp intrinsic fluorescence emission has been established elsewhere.[141]

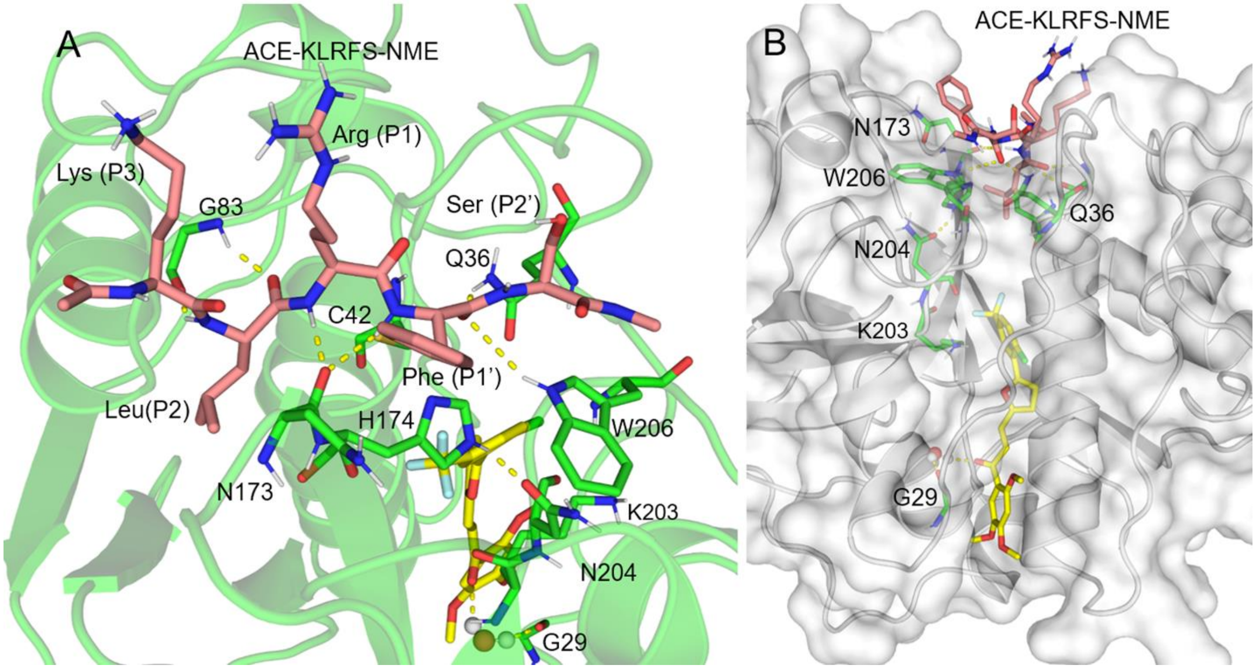
Finally, it was found that the carbonyl oxygen atoms of Cpd66 and residue G29 are involved in the formation of a prevalent water bridge (occurring roughly 70% of the MD simulation time), which contributes to the stability of the ligand within the pocket (Fig. 4.2-B). Remarkably, the water molecule mediating the aforementioned interaction readily exchanges with the bulk solvent instead of remaining trapped within the interface cavity. Fig. C7 in Appendix C shows how the exchange events occur, which require the transient opening of a hole involving the R25, T31 and F215 side-chains, thus allowing the water molecules to diffuse in both directions (Appendix C, Fig. C7-B).

4.2 Energetic analyses

4.2.1 Impact of Cpd66 binding on the affinity of FP-2 for the substrate

Previous experimental measurements confirm that the Cpd66 binding to FP-2 does not alter the affinity of the latter for the substrate, as the measured value of α was 1.[36] The influence of Cpd66 binding on the interaction between FP-2 and a peptidic substrate was assessed by building

Figure 4.3. Predicted structure of FP-2 in complex with a peptide and Cpd66. A) Interface of the FP-2 active site in the presence of the peptide ACE-KLRFS-NME. The peptide residues are labeled using the three-letter code and their positions in the active site are indicated between parentheses. **B)** A different view of the ternary complex showing the binding modes of the substrate and the inhibitor, and the protein surface. In all cases, the intermolecular H-bonds are shown as yellow dashed lines. Residues forming H-bonds with the peptide and the inhibitor, or involved in the catalytic mechanism, are shown as green sticks. The peptide and Cpd66 are represented as salmon and yellow sticks, respectively. For brevity's sake, the interface of the FP-2:peptide complex is not shown, as it is identical to that of the ternary complex.



Source: Prepared by the author.

two systems containing a peptide in the enzyme's active site, i.e., FP-2:peptide and FP-2:Cpd66:peptide complexes. Note that, instead of modelling the fluorogenic substrate employed in the activity assays,[36] we decided to build a true peptide, whose parameters are already contained in the ff14SB force-field, thus avoiding additional parametrization steps. The sequence of the peptide was chosen on the basis of the work by Cotrin *et al.*, that reported the amino acid preference of FP-2 at different subsites (S3 to S2').[16] We built an optimal capped pentapeptide with the sequence ACE-KLRFS-NME, where ACE and NME stand for the acetyl and N-methyl caps, respectively, to conduct the structural and energetic analyses. Of note, residues at P4 and P3' were excluded from the peptide, as the enzyme displays no preference at these positions, according to the results by Cotrin *et al*,[16] thus indicating that those positions are largely irrelevant for the binding process. The most likely binding mode of the peptide into the FP-2 active site was predicted by performing docking and refinement steps consisting of MD simulations and MM-GBSA free calculations, as explained elsewhere.[52] As can be seen in Fig. 4.3-A, the peptide establishes well-known interactions with FP-2 interface residues, e.g., the H-bonds involving residues Q36, G83, N173, and W206.[52] Additionally, the Phe side-chain at P1' adopts a conformation that has been previously predicted for various peptide-based inhibitors of FP-2 bearing this substituent.[52] Finally, note that the Lys side-chain at P3 is exposed to the solvent, in resemblance to the Arg moiety of E64 in the FP-2:E64 crystal structure (PDB: 3BPF).

We conducted four replicate 300 ns MD simulations for each analyzed complex, i.e., FP-2:peptide, FP-2:Cpd66 and FP-2:peptide:Cpd66. The RMSD profiles for the heavy atoms of the peptide and Cpd66 indicate that both remained in stable binding modes during all the MD simulations (Appendix C, Fig. C8). Subsequent MM-GBSA free energy calculations based on the collected frames show that the non-polar interactions are responsible for the favorable ΔG_{eff} values in all the analyzed systems, in accordance with the fact that the allosteric pocket and the active site are mainly made up of hydrophobic residues (Table 4.1). Remarkably, the predicted ΔG_{eff} values confirm that the Cpd66 binding to FP-2 does not change the affinity of the enzyme for the peptide, as indicated by the experimental measurements.[36] Furthermore, this result implies that the shape of the active site is not modified upon the Cpd66 binding, as variations in the relative positions of the interface residues would have led to ΔG_{eff} differences. On the other hand, our predictions show a slight decrease in the affinity of the FP-2:peptide complex for the inhibitor. However, on theoretical grounds,[137] the change in the affinities of FP-2 for the peptide and for Cpd66 must be equal. Hence, the abovementioned discrepancies are likely to stem from inaccuracies associated to the MM-GBSA calculations.

Table 4.1: MM-GBSA free energy values for the studied complexes [a]

Energy (kcal/mol)	**FP-2:Cpd66**	**FP-2:peptide**	**FP-2:Cpd66:peptide**[b]	
			Cpd66	**Peptide**
ΔG_{npol}[c]	-69.5±0.3	-74.6±0.2	-70.0±0.2	-74.8±0.2
ΔG_{pol}[d]	25.4±0.5	16.8±0.6	26.7±0.2	16.9±0.5
ΔG_{eff}	-44.1±0.3	-57.8±0.5	-43.3±0.2	-57.9±0.6

[a] Mean values ± *SEM*s in kcal/mol for each energy component are shown in the table.

[b] The free energy components for the binding of Cpd66 or the peptide in the ternary complex were calculated in the presence of the other ligand.

[c] Non-polar free energy component calculated as the sum of the van der Waals interaction energy and the variation of surface tension upon ligand binding.

[d] Polar interaction energy component calculated as the sum of the electrostatic and the Generalized-Born polar solvation energies.

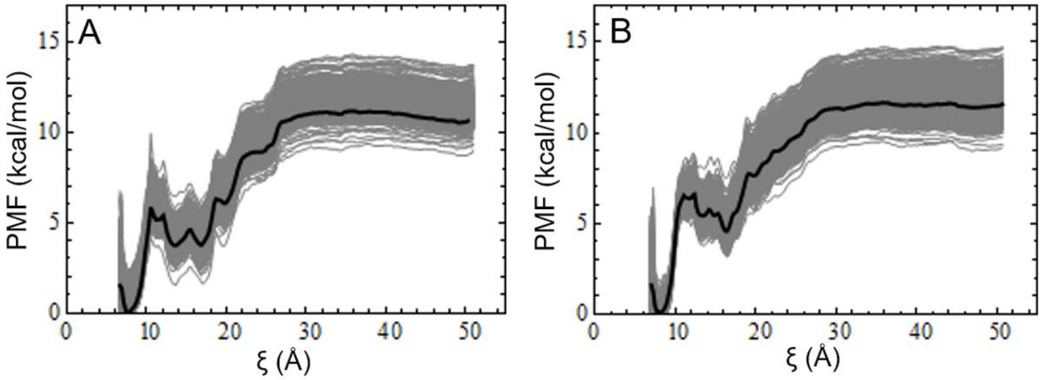
Source: Prepared by the author.

4.2.2 Calculation of standard binding free energies for FP-2 and Cpd66 in the presence and absence of the peptide

The US method was also employed to obtain the $\Delta G°_{bind}$ values for the interaction of Cpd66 with free FP-2 and with the FP-2:peptide complex (Table 4.2). These calculations allowed the assessment of the impact of the peptide on the affinity of FP-2 for Cpd66, thus shedding light onto the inhibition mechanism. For reasons fully explained elsewhere, an older version of the protein force-field (ff99SB-ILDN) was employed in the US protocol.[58]

The PMFs for the dissociation of Cpd66 from FP-2 and from the FP-2:peptide complex reach their respective plateaus at distances larger than 25 Å, where the inhibitor is distant enough to interact with the protein, and at similar free energy values (Fig. 4.4). The $\Delta G°_{bind}$ values predicted from the PMFs and the subsequent correction of the ligand orthogonal restraints, -7.5 kcal/mol and -7.7 kcal/mol for FP-2:Cpd66 and FP-2:Cpd66:peptide complexes, respectively, are close to the experimental affinity, -7.0 kcal/mol (Table 4.2). Therefore, these results show that the affinity of the enzyme for Cpd66 is not significantly changed when the peptide is bound, in agreement with the classical noncompetitive inhibition mechanism reported for this inhibitor.[36] Finally, it is worth saying that no calculations of the $\Delta G°_{bind}$ values for the interaction of the peptide with FP-2 in the presence or absence of Cpd66 were conducted, due to the poor quality results expected for very flexible ligands.

Figure 4.4. PMFs for the dissociation of Cpd66 from FP-2 in the absence and presence of the substrate. A) FP-2:Cpd66 complex. **B)** FP-2:peptide:Cpd66 complex. In both cases, Cpd66 was pulled away from the protein. The reaction coordinate ξ stands for the *z* Cartesian component of the distance vector between the carbonyl carbon of I146 and the alkenyl carbon of Cpd66 bonded to the furanyl ring (Fig. 1.3). The gray lines represent each of the 1000 PMFs obtained through bootstrapping. The thick black line corresponds to the average PMF in each graph.



Source: Prepared by the author.

4.3 Community, pathway and pK_a analyses

4.3.1 Analysis of community rearrangements in the studied systems

According to the results shown in AppendixC, Text C1, Cpd66 induces small perturbations on the FP-2 conformational ensemble, e.g., shifts in RMSD distribution means and widths (Appendix C, Figs. C9-A and B), changes in local flexibility of certain loops (Appendix C, Figs. C9-C and D), and variations in the amplitude of collective motions along the main PCs (Appendix C, Fig. C10). Subsequently, we performed CA to get in-depth knowledge on the allosteric mechanism triggered by Cpd66 binding to FP-2 from a network-based perspective.

Table 4.2: Results of the $\Delta G°_{bind}$ calculations through US for the interaction of Cpd66 with free FP-2 and the FP-2:peptide complex[a]

	FP-2:Cpd66	**FP-2:pept:Cpd66**
k_{xy} (kcal·Å^{-2}·mol^{-1})[b]	9.56	9.56
l_b (Å)[c]	1.15	1.16
ΔW_R (kcal/mol)[d]	-11.2±0.6	-11.6±0.8
ΔG_R (kcal/mol)[e]	-1.2±0.1	-1.0±0.2
$\Delta G°_{bind,calc}$ (kcal/mol)[f]	-7.5±0.6	-7.7±0.8
$\Delta G°_{bind, exp}$ (kcal/mol)[g]	-7.0±0.2	-7.0±0.2[h]

[a]Mean values ± *SEM*s are shown for the free energy terms.

[b]Orthogonal distance restraint constant applied on the *x* and *y* axes.

[c]Configurational integral of the PMF in the bound region, calculated through eq. 2.19.

[d]PMF depth defined by eq. 2.18.

[e]Free energy associated to orthogonal *xy* restraint removal when the ligand is bound, calculated through eq. 2.25.

[f]Calculated standard free energy for the binding of Cpd66 to free FP-2 and to FP-2 in complex with the substrate.

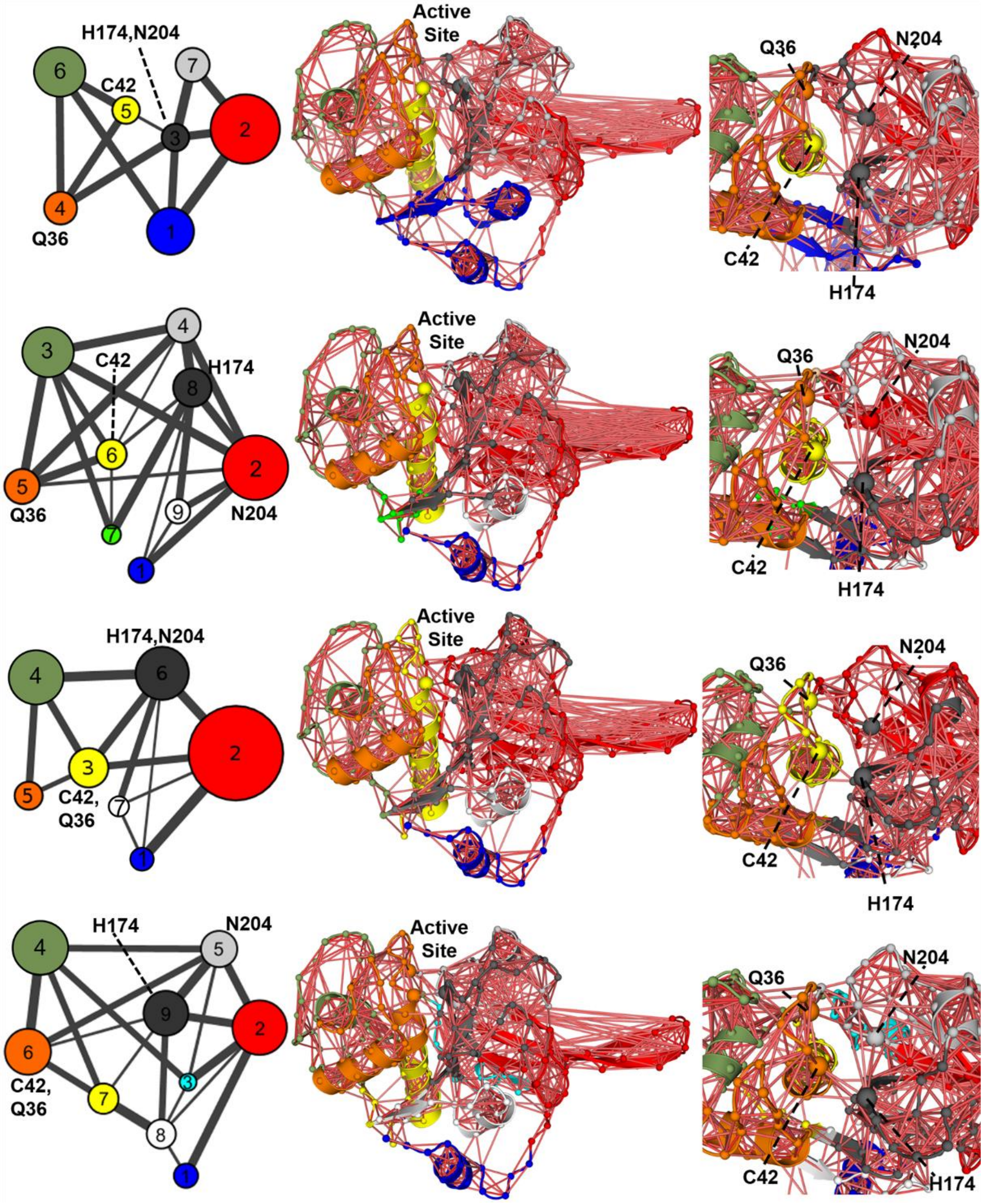
[g]Experimental absolute free energy for the binding of Cpd66 to free FP-2 and to FP-2 in complex with the substrate taken from ref. 36.

[h]It is assumed that the peptide modeled here does not alter the affinity of FP-2 for Cpd66, just as the fluorogenic peptide used in the experiments does.[36]

Source: Prepared by the author.

The raw matrices of *GC* coefficients (Appendix C, Fig. C11) for the studied systems were filtered to reduce noise by using contact maps and *GC* cut-offs, as suggested elsewhere.[100] Initially, two *GC* cut-offs were tested, i.e.,0.5 and 0.6, but the first value was deemed more suitable to conduct the CA, as it preserves meaningful moderate to strong long-ranged *GC*s, i.e., correlations between residues that are not contacting each other during the MD simulations (Appendix C, Fig.

Figure 4.5. Community analysis for the studied systems. A) FP-2, **B)** FP-2:Cpd66, **C)** FP-2:peptide and **D)** FP-2:peptide:Cpd66. 2D diagrams showing the organization of communities for each system are depicted on the left. Communities are represented by circles whose diameters are proportional to the number of residues contained within each. The lines account for intercommunity correlations, which are proportional to the line width. On the right, two views of the community organization represented on the 3D structure of FP-2 are provided for each system, the rightmost one highlighting the active site region. The spheres and red lines stand for the network nodes (Cα atoms) and their pairwise *GC*s, respectively. Dashed lines, when necessary, and labels Q36, C42, H174 and N204 indicate the communities containing catalytic residues (leftmost panel) or the respective nodes of these residues (large spheres on rightmost panel). The same coloring scheme was employed throughout the different community representations of each system.



Source: Prepared by the author.

C12). On the contrary, we observed that most pairs of residues having a $GC \geq 0.6$ are in contact, thus indicating that this cut-off is too high for the systems analyzed here and, consequently, leads to the loss of relevant information (Appendix C, Fig. C12). Finally, CA was performed using the contact map-filtered *GC* matrices of each system (Fig. 4.5 and Appendix C, Table C1), with the optimal number of communities being determined from the graphs shown in Appendix C, Fig. C13.

The CA results show that the residues of free FP-2 are grouped into seven communities (Fig. 4.5-A), which increase up to nine upon the inhibitor binding (Fig. 4.5-B). Communities 1, 3 and 7 are considerably reshaped in the presence of Cpd66, as their residues are redistributed among pre-existing and newly-generated communities (Table C1 and Figs. 4.5-A and B). It is worth noting that community 3 of free FP-2 comprises two key residues that participate in the catalytic process, i.e., H174 and N204.[3] However, these residues split into two different communities in the FP-2:Cpd66 complex. We also observed appreciable changes in the overall pattern of couplings involving communities that contain catalytically-relevant residues (Q36, C42, H174 and N204). Therefore, Cpd66 is able to significantly rearrange the basal community organization in FP-2, especially active site related communities, a phenomenon that could be associated with its ability to inhibit the enzymatic activity.

We noticed that the CA results for the FP-2:peptide and FP-2:peptide:Cpd66 complexes were similar, except for some minor differences, to those presented above. Remarkably, in the FP-2:peptide complex, the key catalytic residues occur only in two communities, 3 and 6 (Fig. 4.5-C). Therefore, upon the peptide binding, Q36 and C42, that belong to different communities in free FP-2, are grouped into the same community; whereas H174 and N204 are not affected by the presence of the peptide (Figs. 4.5-A and C). On the other hand, the Cpd66 binding significantly modifies communities 2, 3, 5 and 6 of the FP-2:peptide complex, and increases the total number of communities from seven to nine (Figs. 4.5C and D, and Appendix C, Table C1). Moreover, residues H174 and N204 split again into two communities in the presence of Cpd66, while the Q36 and C42 remain in the same community (Figs. 4.5-C and D). Overall, our results suggest that the segregation of residues H174 and N204 into separate communities is a key feature of the allosteric mechanism triggered by Cpd66, regardless of the substrate binding. This conclusion relies on the assumption that the native community organization is essential for the proper catalytic activity of FP-2.

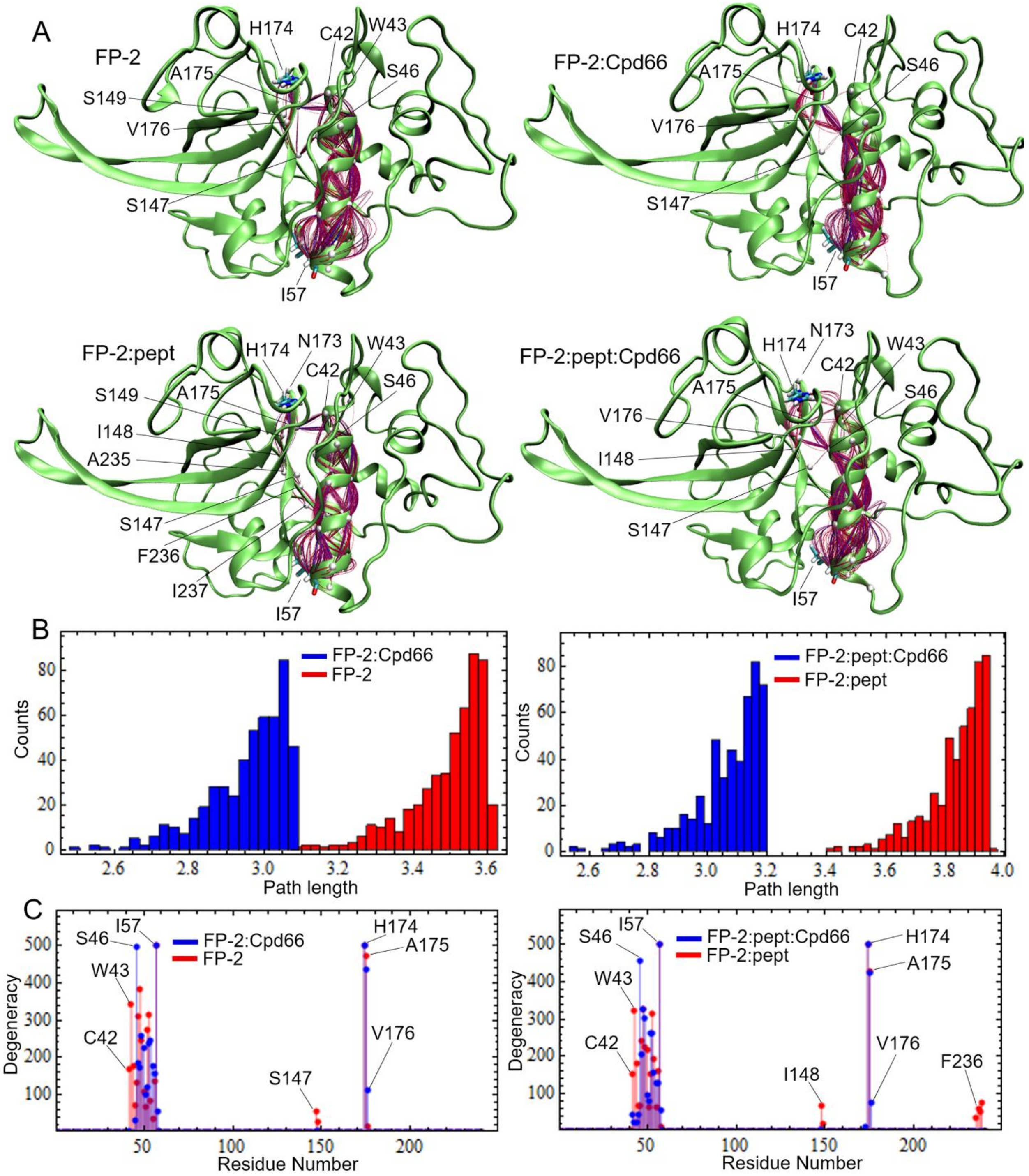
Finally, we would like to stress that the main results derived from the CA remain qualitatively the same even if the distance cut-off used to define the residue-residue contacts is lowered from 5 to 4.5 Å (Appendix C, Fig. C14 and Table C2). Furthermore, note that the former value was deemed more appropriate to conduct the CA, as it leads to a smaller number of communities in each analyzed system when compared to the latter value.[97]

4.3.2 Perturbations of signal propagation patterns across the FP-2 structure caused by Cpd66

The betweenness centrality (hereinafter simply referred to as centrality) is a network property that measures, for a given node, the number of shortest paths connecting every pair of nodes in the network through the former node, provided that the latter is not included in any of those pairs (eq. 2.27).[97] Hence, residues (=nodes) having high centralities can be considered as hot-spots for the signal propagation across the protein (=network). When conducting the analysis of centralities in the studied systems, we detected that Cpd66 is able to modify the distribution of moderate and high centrality nodes in both, free FP-2 and the FP-2:peptide complex (Appendix C, Text C2 and Fig. C15). Remarkably, the centrality profiles of the inhibited systems, i.e., FP-2:Cpd66 and FP-2:peptide:Cpd66, share more similarities between each other that with respect to the active states, i.e., FP-2 and FP-2:peptide; the opposite being also true. Consequently, the centrality perturbations caused by the Cpd66 binding are largely insensitive to the presence of the peptide in the FP-2 active site.

The signal propagation involving the allosteric pocket and the active site was also studied.[101] This analysis requires the selection of an allosteric pocket residue as a source, which was taken from the central α-helix, given the extensive interactions that it establishes with Cpd66 (Fig. 4.2-B) and the presence of the catalytic residue C42 at one of its ends. We finally selected I57 at the opposite end of the central α-helix as source, because it lies far from the active site, which facilitates the detection of signaling pathway differences between the active and inhibited FP-2 systems through visual inspection. It is worth noting, however, that any other residue contained within the allosteric pocket would be, in principle, suitable to perform the pathway analysis. On the other hand, we selected H174 as the sink residue for three main reasons: *i*) it is involved in the catalytic mechanism, *ii*) the community containing this residue and N204 undergoes a significant rearrangement upon Cpd66 binding (Fig. 4.5) and *iii*) H174 and I57 belong to different domains. The latter structural feature allows the detection of differences between inter-domain communication patterns in the studied systems.

Figure 4.6. Communication pathways linking the allosteric and the active sites in four FP-2 systems. A) Structural representation of optimal and suboptimal signaling pathways emerging form the source residue I57 and reaching the active site sink residue H174 in the indicated systems. Nodes (Cα atoms) and paths are represented by spheres and splines, respectively. Shorter paths are depicted in blue. **B)** Path length statistical distributions in the four systems. **C)** Residue degeneracies within the set of pathways of each system. Path length is defined as the sum of d_{ij} values along the given path (Section 2.1.12)



Source: Prepared by the author.

Our results show that the optimal (=shortest) paths connecting the source and the sink residues in free FP-2 and in the FP-2:peptide complex are I57 → E52 → I48 → S47 → W43 → A175 → H174 (length=3.08) and I57 → S53 → G49 → I48 → A44 → W43 → A175 → H174 (length=3.42), respectively. Of note, in both systems, residues W43 and A175 are involved in the inter-domain communications along the optimal paths. In addition, the communication between the allosteric pocket and the active site in free FP-2 and the FP-2:peptide complex can occur via suboptimal paths that emerge from I57 and propagate up to the top of central α-helix, where they finally reach H174 after traversing the inter-domain space (Fig. 4.6-A). Remarkably, the inter-domain communication through most of the suboptimal paths also involves residues W43 and A175, but a small fraction of them can alternately involve C42 and A175 in both systems (Fig. 4.6-A).

In addition to the similar features between the calculated pathways in free FP-2 and the FP-2:peptide complex, there are certain noteworthy differences. Firstly, it can be observed that some suboptimal paths connect S46 and S147 of the adjacent domain in free FP-2, and continue towards H174 by traversing various alternate residues, i.e., S149, V176 and A175 (Fig. 4.6-A). On the other hand, a group of suboptimal pathways in the FP-2:peptide complex propagates from the central α-helix to neighboring β-strands that contains residues P238, I237, F236, A235, S147, I148 and S149, and finally reaches H174. Most of the previously-mentioned residues possess higher centrality values in the FP-2:peptide complex than in free FP-2, which indicates their overall relevance in the signal transduction upon the peptide binding (Appendix C, Fig. C15-A).

The Cpd66 binding to FP-2 and to the FP-2:peptide complex modifies appreciably the basal topology of signaling paths between the allosteric pocket and the active site (Fig. 4.6-A). The optimal paths in the FP-2:Cpd66 and FP-2:peptide:Cpd66 complexes are: I57 → S53 → G49 → S46 → A175 → H174 (length=2.49) and I57 → E52 → I48 → S46 → A175 → H174 (length=2.56), respectively. As can be noted, these two paths are shorter than those of the active states in terms of both, their lengths and the number of residues participating in the signal propagation. Furthermore, the inter-domain communication along the optimal paths of the inhibited states is mediated by residues S46 and A175. Hence, instead of traveling up to the top of the central α-helix, as in the active states, the optimal paths take a shortcut and traverse the inter-domain space at the height of residue S46 (Fig. 4.6-A), thus strengthening the coupling between the allosteric pocket and the active site. Most of the suboptimal paths of both inhibited states also propagate to the adjacent domain through S46, but, subsequently, they bifurcate and continue through either V176 or A175. Interestingly, the hydroxyl group of S46 side-chain is able to form stable H-bonds with the carbonyl oxygen of V176 (occupancies of ~85% in all systems). Therefore, S46 constitutes a key interacting residue at the interface between the central α-helix and the adjacent domain, whose role in inter-domain communication is greatly enhanced by the inhibitor binding, as previously detected through centrality analyses (Appendix C, Text C2 and Fig. C15). Also note that, in closer resemblance to the active states, some paths of the ternary complex propagate through the central α-helix up to residues C42 and W43 and then traverse to the adjacent domain through A175 and H174 (Fig. 4.6-A).

Form the results shown in Fig. 4.6-A, it becomes clear that the inter-domain communication between active site residues is more favorable in the active states than in the inhibited ones, i.e., more suboptimal paths in the former states traverse the inter-domain space via active site residues. The disruption of this basal inter-communication caused by Cpd66 binding might be related to the loss of enzymatic activity. Conversely, the active site and the allosteric pocket are more tightly correlated in the presence of the inhibitor (Fig. 4.6-B). Therefore, the signal propagation between these two regions of the protein become more coherent upon Cpd66 binding, which reflects an entropy loss along the signaling paths.[101]

Finally, the degeneracies of residues in the suboptimal paths were calculated for all the studied systems (Fig. 4.6-C). As expected from the results presented so far, residue S46 and, to a lesser extent, V176 have larger degeneracies in the inhibited complexes than in the active states. Conversely, the active site residues C42 and W43 occur more frequently in the suboptimal paths of the latter systems. Of note, A175 was found to be a dominant residue of the calculated suboptimal paths in all cases (Fig. 4.6-C). Therefore, it is expected for this residue to play a relevant role in the signal transduction between the allosteric pocket and the active site, and also within the whole protein, as revealed by its high centralities (Fig. C15-A).

4.3.3 pK_a shifts in various FP-2 ionizable residues upon Cpd66 binding

An ionizable residue can undergo a pK_a shift when its local environment is modified. Therefore, the calculation of pK_a values can reveal the occurrence of conformational changes in proteins related to the binding of allosteric modulators.[142] Here, we performed the titration of all ionizable residues in free FP-2 and the FP-2:Cpd66 complex by employing CpHMD simulations in explicit solvent.[78, 79] However, some deeply-buried residues, e.g., D23, D91, D109, D221, E52, E67 and C42, did not sample sufficient protonation state transitions during the CpHMD simulations, which precluded the reliable calculation of their pK_a values and, thus, were ruled out from the analyses.[79] Furthermore, we also found residues whose titration curves were well-behaved only in one of the studied systems (Appendix C, Fig. C16), thus probably reflecting the strong impact of Cpd66 binding on the chemical environments of such residues. The reliably-titrated residues that underwent the largest pK_a shifts due to the inhibitor binding are shown in Appendix C, Fig. C17.

The structural representation of all FP-2 residues displaying large pK_a shifts or noticeable changes in their titration curves in the presence of Cpd66 reveals that the latter can perturb the protein's dynamics in regions that are distal to the allosteric pocket (Appendix C, Fig. C18). The affected ionizable residues lie mainly in loops that were shown to change their motions upon Cpd66 binding through RMSF calculations and PCA (Appendix C, Figs. C9-C, C10 and C18). Hence, the results of the CpHMD simulations confirm the perturbation of FP-2 conformational ensemble previously predicted by means of other computational techniques.

Interestingly, the pK_a of the catalytic residue H174 increases from 6.0 to 6.67 upon Cpd66 binding (Appendix C, Fig. C17). The stronger basicity of this key residue in the presence of the inhibitor might be related with the loss of enzymatic activity. In terms of the relative abundance of its protonation states, the previous pK_a difference predicts a two-fold increase in the population of deprotonated H174 in free FP-2 with respect to that of the FP-2:Cpd66 complex. On the other hand, we could not titrate C42, as mentioned earlier, because it remains protonated during nearly the whole CpHMD simulations at every pH. The protonation state of this residue is likely to affect that of H174 due to the proximity of both. Therefore, the inability to fully sample the protonation/deprotonation equilibrium of C42 is likely to lead to inaccuracies in the predicted pK_a of H174. However, the H174 pK_a shift caused by Cpd66 binding is more reliable than the absolute pK_a values of the residue due to the cancellation of systematic errors.

Finally, we attempted to conduct the titration of ionizable residues of the FP-2:peptide and FP-2:peptide:Cpd66 complexes. Unfortunately, the peptide dissociated during the CpHMD simulations conducted for the former complex at pH < 4.8 (Appendix C, Fig. C19), thus precluding the obtaining of complete titration curves in its presence. Interestingly, the experimental enzymatic activities of natural and recombinant FP-2 against the synthetic substrate Z-FR-AMC at different pHs, sharply decrease at pH<5.0 and pH>6.0.[9] Therefore, our results suggest that the FP-2 activity decrease at pHs below 5.0 can stem from the lower affinity of the enzyme for the peptidic substrates. Furthermore, it was predicted that the peptide remains bound to the active site at basic pHs. Hence, the activity decrease observed in the experiments is likely to be caused by the inability of the enzyme to cleave the P1-P1' peptide bond of the substrate.

4.4 Discussion

We have studied here the interaction between FP-2 and the allosteric inhibitor Cpd66. The predicted structure of the complex is consistent with several experimental findings, e.g., the protection of the peptide 204-NSWGQQWGER-213 from tryptic hydrolysis, the micromolar K_i, the negligible impact of the Cpd66 on the affinity of FP-2 for the substrate and the disruption of Trp microenvironment due to the compound binding.[36] According to our predictions, Cpd66 binds a cryptic pocket in the site 3 region,[25] whose entrance remains occluded in the available FP-2 crystal structures and in most frames collected during the MD simulations of the free enzyme. Of note, several lines of evidence support the binding of allosteric inhibitors to cryptic or transient pockets, thus reinforcing the importance of MD simulations to explore the occurrence of such hidden cavities.[23, 143]

Interestingly, the entrance of the site 3 pocket occurs in the open state in the crystal structures of hCatK (e.g., 1ATK, 5TUN), the enzyme for which this pocket was first proposed as a druggable allosteric site.[25] Differences in the composition of residues lying at the pocket entrance are likely to determine variable opening probabilities. In fact, hCatK bears residues P50 and E84 in equivalent positions to E67 and D101 of FP-2, respectively. This favors the accommodation of the K17 side-chain (K34 in FP-2) closer to E35 of hCatK (E52 in FP-2), thus creating an open entrance between the former residue and P50 that leads to site 3 internal cavity. On the contrary, residues E52, E67 and D101 of FP-2 stabilize the K34 side-chain in a position that occludes the pocket entrance. The opening events in this case require the disruption of stable electrostatic interactions, which explains why the open pocket is rarely observed during MD simulations.

We also studied the potential formation of an open cavity at the site 3 region of FP-3. Our results showed we that the pocket entrance opens transiently as well in this enzyme, although its opening probability and the average cavity volume are both lower than those of FP-2. The importance of site 3 as a druggable cavity for the design of specific inhibitors is supported by these findings, as variations in the cavity shape and stability occur in close homologs such as FP-2 and FP-3. On the basis of previous considerations, it is likely for the differences between the FP site 3 cavities and those of human cysteine cathepsins to become even more appreciable. Therefore, low selectivity issues might be addressed by focusing on the identification of novel compounds targeting site 3.

As a non-competitive inhibitor with α=1,[36] Cpd66 does not alter the affinity of FP-2 for its substrate and, hence, no variations of the active site shape are expected to occur. This experimental finding was reproduced by the free energy calculations based on the predicted binding mode of Cpd66 to FP-2. However, perturbations in the conformational ensembles of FP-2 were detected upon Cpd66 binding by means of RMSD, RMSF profiles and PCA. Such subtle changes emerge only when taking into account the whole conformational populations of active and inhibited states, but remain mostly elusive from pairwise structural comparisons. In accordance with our predictions, the allosteric inhibition of FP-2 does not fit into the classic model of allostery requiring two well-defined states, the tense and relaxed forms, but one involving a dynamic shift of equilibrium populations induced by the modulator, mainly through an entropy-driven process.[23, 144, 145] It is worth noting that the occurrence of allosteric modulation in the absence of appreciable conformational changes in a second site is a well-established phenomenon.[145-147] Based on the crystal structures of hCatK in complex with two allosteric inhibitors, NSC13345 and NSC94914 (PDBs: 5JA7 and 5J94, respectively), it can be thought that these compounds exert the inhibition through a mechanism apparently fitting into the aforementioned model of allostery, as no significant conformational changes arise upon their binding to the enzyme.[25, 27] However, an *in silico* study by Novinec does show the occurrence of some conformational changes in the hCatK active site due to the binding of NSC13345 and NSC94914 to site 6.[26] The apparent contradiction between the *in silico* results and the crystal structures can be reconciled by arguing that structural changes are probably lost during the crystallization process.[26, 147]

On the other hand, the CA conducted here provided meaningful insights into the dynamical changes caused by Cpd66 binding. In this sense, we noted that the inhibitor induces a fragmentation of the basal organization of communities in free FP-2 and in the FP-2:peptide complex. A similar phenomenon was observed in modeled complexes of cruzain and potential allosteric ligands targeting site 3.[53] Furthermore, the catalytic residues H174 and N204, which pertain to the same community in the active forms, split into different communities in the presence of Cpd66.

Likewise, catalytic residues of thrombin were found to separate into different communities upon the binding of the allosteric inhibitor hirugen.[97] Hence, the rearrangement of active site communities might be considered a relevant feature of allosteric modulation in some systems. Remarkably, the similar effects caused by Cpd66 on the community organization of free FP-2 and the FP-2:peptide complex is consistent with the non-competitive inhibition mechanism, since the substrate is not able to reverse the changes triggered by the allosteric inhibitor.

Centrality and pathway analyses revealed further changes triggered by Cpd66 in the signal propagation within the FP-2 structure. We observed that the communication between the allosteric pocket and the active site becomes stronger in terms of residue-residue correlations in the presence of the inhibitor. Similar effects have been reported for other systems, including the thrombin:hirugen and predicted cruzain:compound complexes, and the bacterial amidotransferase HisH-HisF.[53, 97, 101] Moreover, the hotspots involved in the signal propagation between the allosteric pocket and the active site identified during our analyses can guide mutagenesis experiments seeking to dissect the inhibition mechanism.[101] For instance, an S46 mutation might impair the propagation of signals originated from the allosteric pocket and heading to H174 upon Cpd66 binding, which could have an impact on the measured activity of FP-2.

Finally, we calculated the titration curves of acidic and His residues of free FP-2 and the FP-2:Cpd66 complex using CpHMD simulations in explicit solvent at different pHs. According to our predictions, Cpd66 is able to modify the pK_a values of various residues lying in different FP-2 regions. The occurrence of pK_a shifts provided further support to the occurrence of conformational changes triggered by the allosteric inhibitor.[142] Most of the residues undergoing pK_a shifts upon Cpd66 binding are located in loops, some of them flanking the active site, which suggests that loop motions might be somehow related to the FP-2 catalytic activity. Interestingly, the existence of motion-function relationships has been already established for other systems.[148, 149]

A particular residue undergoing a pK_a shift upon Cpd66 binding that is worth analyzing in detail is H174, since this change reveals the ability of the inhibitor to influence directly the functioning of the enzyme’s active site. Incidentally, the basal protonation state of the catalytic diad of cysteine proteases has been the subject of debate, and various models involving the occurrence of an ion-pair ($\text{Cys-S}^-/\text{HisH}^+$), fully protonated states ($\text{Cys-SH}/\text{HisH}^+$) and neutral species (Cys-SH/His) have been proposed so far.[150-153] As pointed out earlier, we were not able to titrate the C42 residue of FP-2 with a step-wise CpHMD protocol, which precluded the prediction of the catalytic diad protonation state. In spite of that, some consequences to the catalytic activity of FP-2 can be derived from the predicted pK_a shift of H174 caused by Cpd66 binding. For example, as the predicted pK_a values of H174 are close to the FP-2 functional pH (5.5), the pK_a shift can modify the basal ratio of protonated to deprotonated populations. Moreover, if one assumes that the catalytic diad of FP-2 operates in a similar fashion to that of cruzain (Cys-SH/His), in which the His residue abstracts a proton from the neighboring Cys upon substrate binding,[153] then, an enhanced His basicity due to the allosteric ligand binding would increase the fraction of $HisH^+$ at pH 5.5. This, in turn, would change the catalytic diad protonation state from its native form (Cys-SH/His) to a fully protonated one (Cys-SH/$HisH^+$). The catalytic activity could slow down because the abstraction of the C42 proton prior to the nucleophilic attack to the substrate carbonyl group becomes more difficult if H174 is already protonated ($HisH^+$). A more basic His residue can also disfavor the transfer of its proton attached to the Nδ atom to the peptide leaving group nitrogen, which is also detrimental for a catalytic mechanism involving an ion-pair state of the diad. Overall, the predicted pK_a shift of H174 could impair the proper functioning of the active site without necessarily modifying the latter's native shape.

5 ROLE OF WATER BRIDGES IN THE AFFINITY AND SELECTIVITY FOR FP-2 OF NITRILES BEARING PYRIDINE SUBSTITUENTS AT P2

5.1 Structural analyses

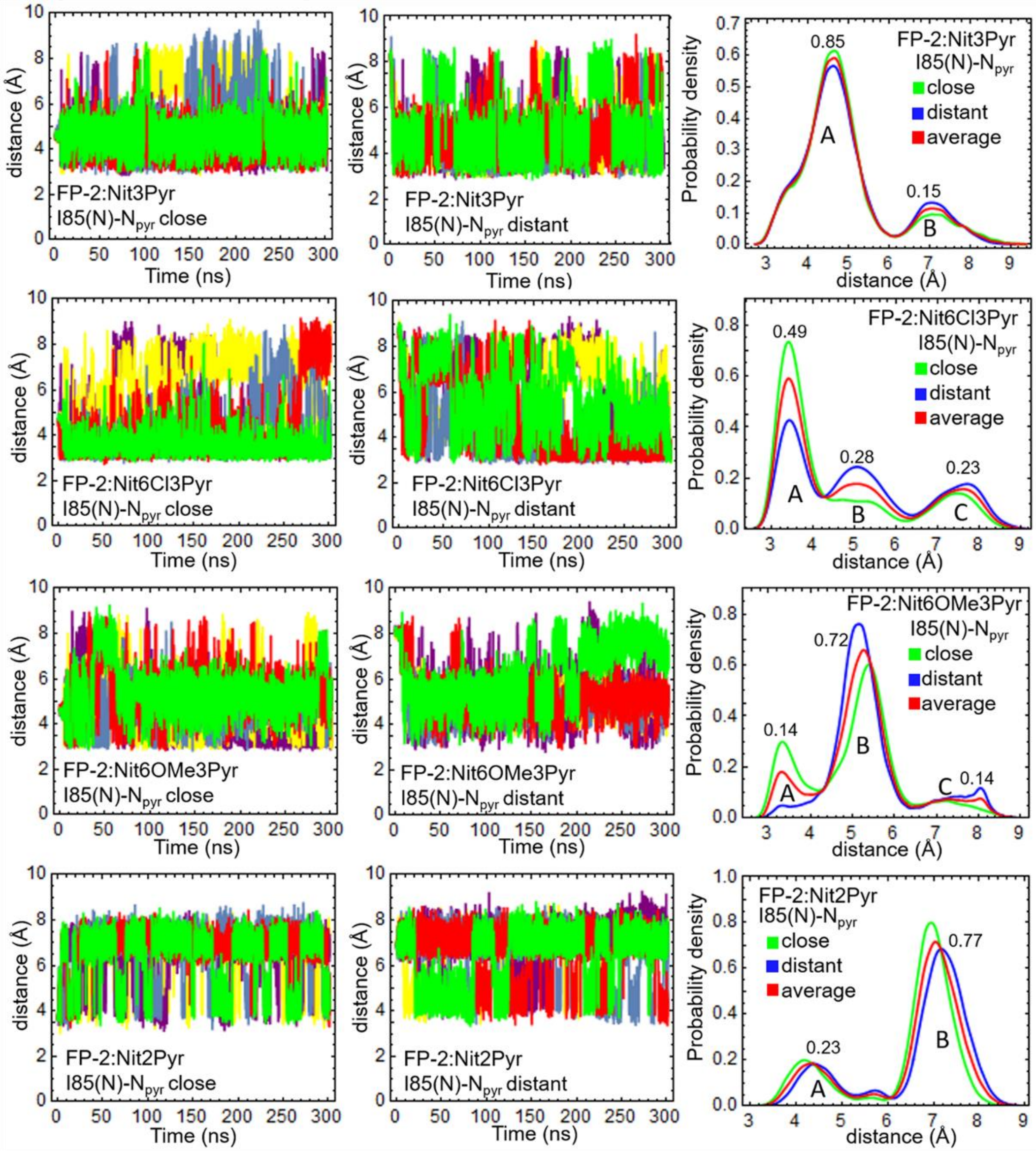
5.1.1 Analysis of H-bonds and water bridges between P2-Pyr substituents and FP-2 residues

Due to the weakly basic nature of pyridine (pK_a=5.0),[154] it is possible for the Pyr-containing nitriles to interact in the protonated form with FP-2. We addressed this issue by conducting a thorough energetic analysis of the putative FP-2:Nit3PyrH$^+$ complex, complemented with long MD simulations, which were also carried out for the FP-2:Nit2PyrH$^+$ and FP-2:Nit4PyrH$^+$ complex (Appendix D, Text D1 and figures cited therein). Our results show that FP-2 binds the neutral form of Nit3Pyr much more favorably than Nit3PyrH$^+$, with a remarkable affinity difference of roughly 6 kcal/mol (Appendix D, Table D1). For the other two compounds, it was inferred that the neutral forms must preferentially bind the enzyme as well (Appendix D, Text D1 and figures cited therein). Likewise, Nit6Cl3Pyr and Nit6OMe3Pyr are expected to interact with FP-2 in the neutral form, not only on the basis of the results obtained for Nit3Pyr, but also due to the lower basicity of their substituted pyridine moieties (Appendix D, Text D1).

Once ruled out the occurrence of protonated Pyr-containing nitriles in complex with FP-2, we studied the interaction of this enzyme with the deprotonated compounds. Through visual inspection of the MD trajectories and distance-based analyses, we determined that, at least, two different orientations of the asymmetric P2-Pyr rings at the interface of the studied FP-2:nitrile complexes were possible, namely, *i*) a close orientation in which the N_{pyr} atom is deeply buried and pointing towards I85 and *ii*) a distant orientation in which the N_{pyr} atom is more exposed to the solvent and pointing away from the S2-pocket base. Subsequently, frames corresponding to differently-oriented poses of the 2Pyr- and 3Pyr-containing nitriles bound to FP-2 were selected to conduct replicate MD simulations of each complex. The time profiles and distributions of the I85(N)-N_{pyr} distance calculated during these MD simulations are shown in Fig. 5.1. As can be seen, all the accessible positions of the P2-pyridine ring with respect to I85(N) were sampled during MD simulations started from either orientation. However, some discrepancies in the distributions derived from the MD simulations of the distant and close orientations of Nit6Cl3Pyr and

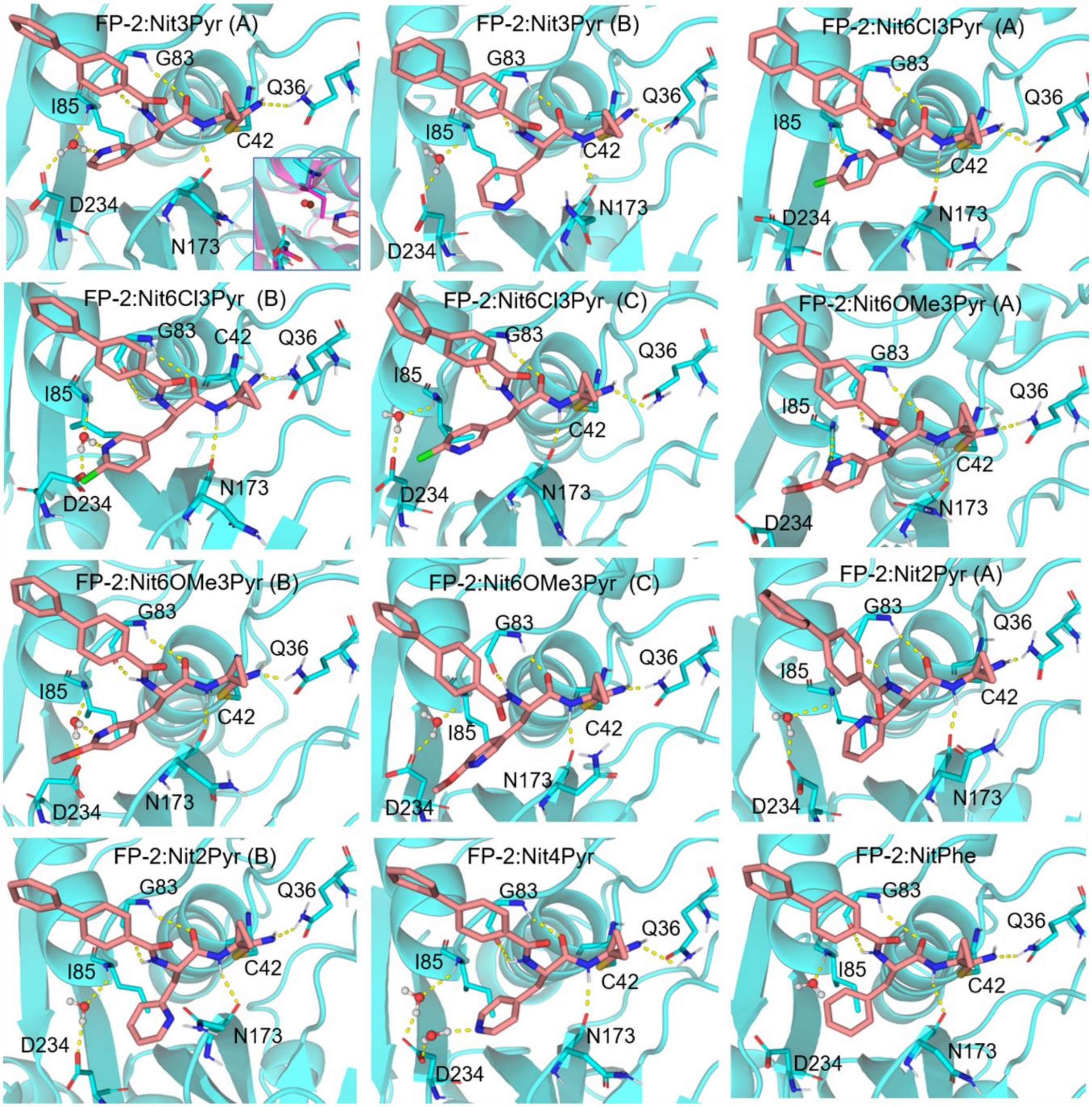
Nit6OMePyr (Fig. 5.1) indicate that longer simulation time and/or replicas are needed to reach full convergence.

Figure 5.1. Distance time profiles and distributions during MD simulations of FP-2 in complex with 3Pyr- and 2Pyr-containing nitriles. Distances between the indicated atoms were calculated for every complex during ten replicate MD simulations, half started from a pose in which the N_{pyr} atom is close to the reference atom I85(N), and the other half started from a distant orientation. Each replicate MD simulation is colored differently in the distance vs. time graphs. Distance distributions were calculated separately for the close and distant orientations by concatenating the values collected from the corresponding replicate MD simulations. An average distribution for each complex was also obtained by combining all results. The fraction of structures pertaining to each peak is indicated on top of the corresponding maximum. Position restraints were applied to complex heavy atoms during the first 3 ns of each distance time profile, corresponding to NVT and NPT equilibrations.



Source: Prepared by the author.

Figure 5.2. Structural representation of the studied FP-2:nitrile complexes. Central structures corresponding to each peak of the I85(N)-N_{pyr} distance distribution (Fig. 5.1) were obtained for every nitrile bearing an asymmetric P2-Pyr moiety, i.e., Nit3Pyr, Nit2Pyr, Nit6Cl3Pyr and Nit6Cl3Pyr, in complex with FP-2. The letter in parentheses indicates the peak in Fig. 5.1 to which the structure belongs. A single central structure was calculated for FP-2:Nit4Pyr and FP-2:NitPhe complexes. H-bonds and water bridges are represented as yellow dashed lines and residues participating in these interactions at the interfaces, plus the catalytic residue C42, are depicted as cyan sticks and labelled accordingly. In all cases, the protein (cartoon) and the ligand (stick) are colored in cyan and salmon, respectively. The inset in the FP-2:Nit3Pyr (A) panel corresponds to a structural alignment of the complex (cyan) and the crystal structure of FP-2 (magenta, PDB: 2OUL). The nearly-overlapping positions of the water molecule mediating the interaction between I85(N) and N_{pyr} (red sphere) and a crystallographic water (firebrick-colored sphere) is highlighted.



Source: Prepared by the author.

The N_{pyr} atoms of 3Pyr-containing nitriles remain mostly oriented towards I85(N), with distance values within the 3.0 to 6.2 Å range, thereby suggesting the occurrence of stabilizing polar interactions within the S2 subsite (Fig. 5.1). Smaller fractions of the respective ensembles contain the N_{pyr} atom in a distant orientation (peaks located at distances > 6.2 Å, Fig. 5.1). For the FP-2:Nit3Pyr complex, the central structure calculated for the largest distance-distribution peak (peak A) reveals that N_{pyr} can interact simultaneously with I85(N) and D234(OD1,2) via a water bridge (Fig. 5.2). A subsequent analysis of the interaction stability showed that the I85(N)-N_{pyr} water bridge occurs in nearly 50% of the frames collected during the FP-2:Nit3Pyr MD simulations (Table 5.1). The I85(N)-D234(OD1,2)-N_{pyr} water bridge is observed in 26% of all trajectory frames, while the D234(OD1,2)-N_{pyr} one, not necessarily involving I85(N), has a slightly higher occupancy (28%, Table 5.1). Therefore, only in 2% of the structures, the water molecule bridging D234(OD1,2) and N_{pyr} is not simultaneously interacting with I85(N). Moreover, distance distributions involving the water molecules closest to the N_{pyr} atom in the FP-2:Nit3Pyr complex indicate that I85(N) and D234(OD1,2) remain at distances ≤ 3.5 Å during 80 and 49% of the simulation time, respectively (Appendix D, Fig. D3). On the other hand, the I85(N)-N_{pyr} direct H-bond occurs very rarely (6%, Table 5.1), thus indicating that the water bridge is essential to mediate a polar interaction between both atoms. Interestingly, we observed a crystallographic water close to I85(N) in the PDB structure 2OUL, that nearly overlaps the one interacting with I85(N) and N_{pyr} in the FP-2:Nit3Pyr complex (Fig. 5.2). This finding reinforces our predictions, as it corroborates, on the basis of a crystal structure, that a water molecule lies in an optimal position to form the aforementioned interactions.

The remaining FP-2 complexes involving nitriles with substituted 3Pyr moieties, i.e., FP-2:Nit6Cl3Pyr and FP-2:Nit6OMe3Pyr, display more complex I85(N)-N_{pyr} distance profiles (Fig. 5.1). For these systems, two peaks, A and B, are observed at distances <6.2 Å, thus indicating the occurrence of two distinct groups of conformations within the aforementioned distance range. Peak A comprises structures in which the N_{pyr} atom of Nit6Cl3Pyr and Nit6OMe3Pyr can form a direct H-bond with I85(N) (Figs. 5.1 and 5.2). Water bridges between I85(N), D234(OD1,2) and N_{pyr} are observed in frames extracted from peak B of both complexes (Figs. 5.1 and 5.2). Finally, as in FP-2:Nit3Pyr, small fractions of the substituted 3Pyr rings occur in conformations in which the N_{pyr} atom is oriented away from I85(N), more exposed to the water bulk (state C, Figs. 5.1 and 5.2). Remarkably, we noticed that the -Cl group increases the likelihood of forming direct H-bonds with

I85(N), an interaction that is less prevalent in the FP-2:Nit6OMe3Pyr and FP-2:Nit3Pyr complexes, in that order (Table 5.1). Concomitantly, the occurrence of water bridges involving the N_{pyr} atom of Nit6Cl3Pyr sharply decreases with respect to the other 3Pyr-containing nitriles. On the contrary, the ability of N_{pyr} atom to form water bridges remains relatively unchanged in the FP-2:Nit6OMe3Pyr complex when compared to FP-2:Nit3Pyr (Table 5.1). Overall, we observed the occurrence of direct H-bonds and water bridges with variable stabilities at the S2 subsite of FP-2 in complex with 3Pyr containing nitriles.

Table 5.1. Fractional occupancies of water bridges and H-bonds involving the P2-Pyr moiety of the studied nitriles in complex with FP-2[a]

Complex	**Direct H-bond[b]**	**Wat. Bridge #1[c]**	**Wat. Bridge #2[d]**	**Wat. Bridge #3[e]**
FP-2:Nit3Pyr	0.06 (0.01)[f]	0.50 (0.03)	0.28 (0.02)	0.26 (0.02)
FP-2:Nit6Cl3Pyr	0.26 (0.05)	0.22 (0.03)	0.16 (0.03)	0.14 (0.03)
FP-2:Nit6OMe3Pyr	0.12 (0.02)	0.55 (0.06)	0.33 (0.07)	0.28 (0.06)
FP-2:Nit2Pyr	0	0	0	0
FP-2:Nit4Pyr	0	0	0.55 (0.05)	0

[a] The fraction of frames (= fractional occupancy) showing each interaction is provided
[b] H-bond between I85(N) and N_{pyr}.
[c] Water bridge between I85(N) and N_{pyr}.
[d] Water bridge between D234(OD1,2) and N_{pyr}.
[e] Simultaneous water bridge between I85(N), D234(OD1,2) and N_{pyr}.
[f] Mean values were obtained by averaging the fractional occupancies corresponding to each replicate MD simulation. *SEM*s are shown in parentheses.
Source: Prepared by the author.

Differently from Nit3Pyr and its derivatives, Nit2Pyr and Nit4Pyr are unable to form H-bonds or water bridges with I85(N) of FP-2 (Fig. 5.2), due to the geometric position of the N_{pyr} atom in these compounds. We did not observe any other polar interactions at the S2 subsite of FP-2 in complex with Nit2Pyr, and the N_{pyr} atom of the latter remains distantly oriented with respect to I85(N) most of the time, regardless of the initial pose chosen to start the MD simulations (Figs. 5.1 and 5.2 and Table 5.1). On the other hand, Nit4Pyr is able to interact with D234(OD1,2) through a water bridge in 55% of the trajectory frames, but the water molecule involved is more exposed to the bulk and does not simultaneously bridge I85(N), as in the FP-2:Nit3Pyr complex (Fig. 5.2, Table 5.1 and Appendix D, Fig. D3). Of note, the energy contribution of H-bonds formed in a medium of greater polarity, e.g., D234(OD1,2)-N_{pyr} water bridges in FP-2:Nit4Pyr, is expected to be smaller than that of H-bonds occurring in more nonpolar environments, e.g., I85(N)-N_{pyr} water bridges in FP-2:Nit3Pyr.[155] Furthermore, the free energy associated to multiple H-bonds involving the same groups or molecules, as in the I85(N)-D234(OD1,2)-N_{pyr} water bridge described for FP-2:Nit3Pyr, is enhanced because the entropic effects barely reduce the energy contribution of the extra H-bonds.[156, 157] Since Nit2Pyr and Nit4Pyr have low activity against FP-2, it is suggested from the results presented so far that the ability of N_{pyr} to form polar interactions with I85(N) is critical to increase the affinity for the enzyme.

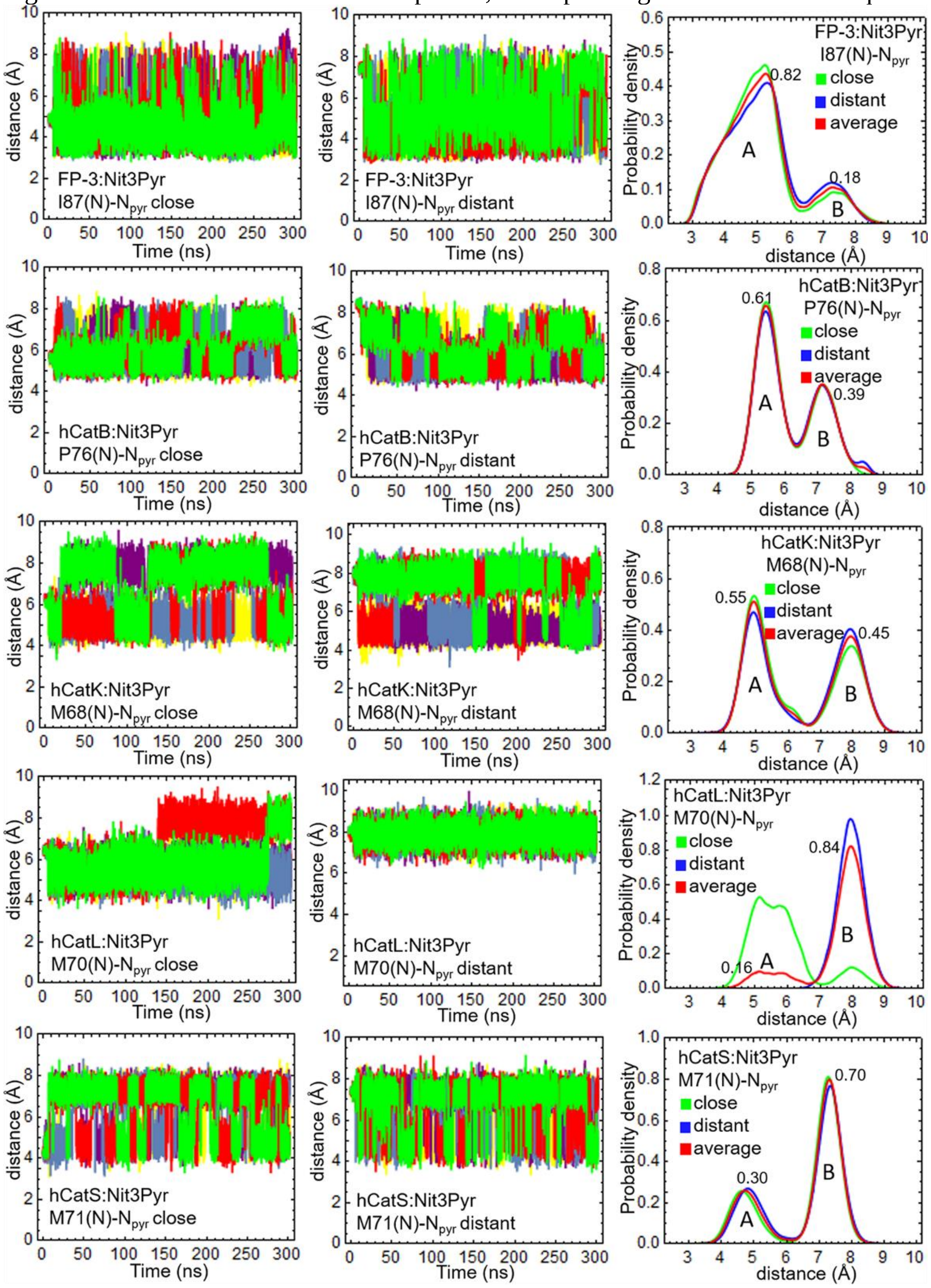
Finally, we show the binding mode of NitPhe to FP-2, obtained from RMSD clustering analysis of a single 300 ns MD simulation of the complex (Fig. 5.2). As can be seen, the crystal water molecule close to I85(N) remains interacting with the latter in the central structure of FP2:NitPhe (Fig. 5.2), but due to the hydrophobicity of the phenyl group, the compound is unable to form water-mediated interactions within the S2 subsite. This could explain the lower affinity of NitPhe for FP-2 when compared to that of 3Pyr-containing nitriles (Fig. 1.8).

5.1.2 Analysis of H-bonds and water bridges between the 3Pyr substituent and S2 residues of FP-3 and hCats

As done for the FP-2 complexes, we carried out several replicate MD simulations for Nit3Pyr in complex with FP-3 and the hCats of interest (hCatB, hCatK, hCatL and hCatS), started from close and distant orientations of the 3Pyr substituent within the respective S2 subsites. Moreover, distance time profiles and distributions equivalent to those shown in Section 5.1.1, were obtained for the systems analyzed here (Fig. 5.3). As can be observed in the graphs, excellent convergence of the distance distributions obtained from different initial 3Pyr orientations was achieved for all the complexes, except for hCatL:Nit3Pyr. The latter's average distribution was calculated by combining the distance distributions of the closely and distantly-oriented poses according to the probabilities associated to each, derived from TI $\Delta\Delta G$ calculations (see Appendix D, Text D2, Fig. D5 and Table D2).

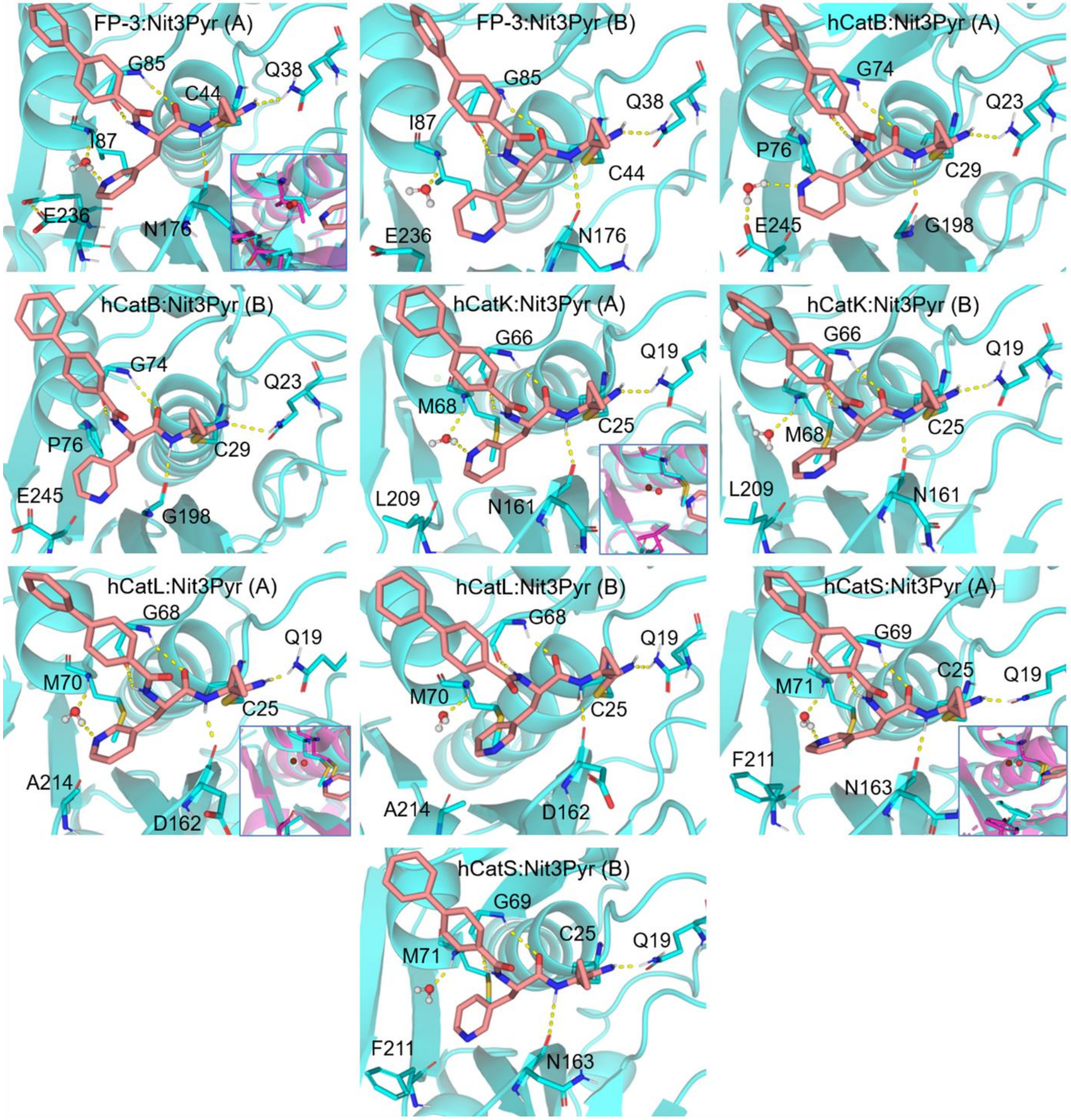
The distance distribution of the FP-3:Nit3Pyr complex shows that the N_{pyr} atom remains oriented towards I87(N) in 82% of the trajectory frames (peak A), whereas just 18% of the frames occur in the distant orientation (peak B, Figs. 5.3 and 5.4). A water bridge between N_{pyr} and I87(N) is observed in the central structure of peak A (Fig. 5.4). Remarkably, the water molecule mediating this interaction lies in a position equivalent to that of a crystallographic water of the FP-3 crystal structure 3BWK (inset in Fig. 5.4). Apart from the main I87(N)-N_{pyr} water bridge, formed during nearly half of the simulation time, the H-bond analysis revealed the occurrence of less prevalent

Figure 5.3. Distance time profiles and distributions during MD simulations of FP-3 and hCats in complex with Nit3Pyr. Distances between the indicated atoms were calculated for every complex during ten replicate MD simulations, half started from a closely-oriented pose, and the other half, from a distantly-oriented one. Replicate MD simulations are colored differently in the distance vs time graphs. Distance distributions were calculated separately for the MD simulations started from each 3Pyr orientation. Average distance distributions, except for hCatL, were obtained by combining the former two. The fraction of frames within each peak is indicated on top of its maximum. Position restraints were applied to complex heavy atoms during the first 3 ns of each distance time profile, corresponding to NVT and NPT equilibrations.



Source: Prepared by the author.

Figure 5.4. Structural representation of Nit3Pyr in complex with FP-3 and various hCats. Central structures corresponding to each peak of the distance distributions shown in Fig. 5.3 were obtained for all the analyzed complexes. The letter in parentheses indicates the peak to which the structure belongs. H-bonds and water bridges are represented as yellow dashed lines and residues participating in these interactions at the interfaces, plus the catalytic residue C42, are depicted as cyan sticks and labelled accordingly. In all cases, the protein (cartoon) and the ligand (stick) are colored in cyan and salmon, respectively. The insets show a structural alignment between the Nit3Pyr complexes (cyan) and the corresponding crystal structures of the enzymes (magenta, PDBs: 3BWK, 4DMY, 5MAE and 2FQ9). The relative positions of the water molecules involved in water bridges in the analyzed complexes (red sphere) and the closest crystallographic waters (firebrick-colored sphere) are highlighted. Due to the absence of a polar hydrogen, P76 of hCatB is unable to form a water bridge with N_{pyr}; therefore, this interaction, as well as the inset showing close water molecules, are missing from the panels corresponding to this enzyme.



Source: Prepared by the author.

E236(OE1,2)-N_{pyr} and I87(N)-E236(OE1,2)-N_{pyr} water bridges, with fractional occupancies of 0.14 and 0.13, respectively (Table 5.2). As shown by the FP-3:Nit3Pyr central structure, E236 bends away from the S2 subsite and forms an H-bond with Y238(OH), thereby adopting a distant position that prevents the former residue from establishing more stable water bridges with I87(N) and N_{pyr}. The I87(N)-N_{pyr} direct H-bond was observed in a small fraction of the complex structures (0.06, Table 5.2). These results demonstrate that the FP-3:Nit3Pyr and FP-2:Nit3Pyr complexes possess several common features, e.g., their respective distance distributions are similar, the stability of the I87/I85(N)-N_{pyr} direct H-bonds and water bridges are almost identical, and the acid residues E236/D234 can also interact with N_{pyr} via water bridge (Figs. 5.2 and 5.4). However, the latter interactions are two-fold more prevalent in FP-2:Nit3Pyr (Tables 5.1 and 5.2). This, in turn, indicates that the carboxyl oxygens have a more suitable position to form the water bridge in the shorter side-chain of D234, as corroborated by the distance distributions shown in Appendix D, Fig. D3. Further analyses carried out for the free enzymes confirmed that the I85(N)-D234(OD1,2) water bridge in FP-2 is much more stable than its equivalent, I87(N)-E236(OE1,2), in FP-3 (see Text D3, Fig. D6 and Table D3). Overall, the structural difference between FP-2 and FP-3 at position 234/236, may explain why the affinity of Nit3Pyr is higher for the former enzyme (Fig. 1.8).

On the other hand, the distance distributions of Nit3Pyr in complex with various hCats show an overall increase in the relative abundance of peaks B with respect to those of FP-3:Nit3Pyr and FP-2:Nit3Pyr (Figs. 5.1 and 5.3). Hence, the polar interactions favoring the 3Pyr close orientation within the S2 pocket of hCats must be weaker than those occurring in the analog FP complexes. Noticeably, hCatB possesses a Pro residue at position 76 that precludes the formation of H-bonds and water bridges with N_{pyr} of Nit3Pyr, due to the lack of a polar hydrogen attached to its N atom (Fig. 5.4). Instead, a solvent-exposed water bridge between E245(OE1,2) and N_{pyr}, with a fractional, occupancy of 0.20 is observed in hCatB:Nit3Pyr complex (Fig. 5.4 and Table 5.2). However, its contribution to the complex affinity is expected to be slight, considering the low stability of the water bridge and the more polar environment in which the interaction occurs.

Water bridges involving N_{pyr} and M68(N)/M70(N)/M71(N) can be found at the interfaces of the hCatK:Nit3Pyr, hCatS:Nit3Pyr and hCatL:Nit3Pyr complexes (Fig. 5.4), with fractional occupancies of 0.35, 0.05 and 0.11, respectively (Table 5.2). Therefore, all these interactions are less stable than those observed in the FP-3:Nit3Pyr and FP-2:Nit3Pyr complexes (Tables 5.1 and

5.2). Moreover, no extra water bridges are possible in hCatK, hCatL and hCatS, as they bear hydrophobic residues, L209, A214 and F211, respectively, in lieu of the acidic ones found in FP-2 and FP-3 at equivalent positions (Figs. 5.2 and 5.4). Incidentally, the water molecules involved in the water bridges described for hCatK:Nit3Pyr, hCatL:Nit3Pyr and hCatS:Nit3Pyr complexes are close to crystallographic waters of the respective structures of these enzymes. However, for these systems, the water overlap is not as perfect as that observed in FP-2 and FP-3 bound to Nit3Pyr (Figs. 5.2 and 5.4), thus providing a structural hint about the lower water bridge stabilities of the analyzed hCat complexes (Table 5.2). Overall, the results presented here suggest that the water bridges between N_{pyr} of Nit3Pyr and S2 residues of the studied hCats contribute less to the binding affinity than those of the equivalent FP complexes.

Table 5.2. Fractional occupancies of water bridges and H-bonds involving the N_{pyr} atom of Nit3Pyr in complex with FP-3 and various hCats[a]

Complex	Direct H-bond	Wat. Bridge #1[c]	Wat. Bridge #2[d]	Wat. Bridge #3[e]
FP-3:Nit3Pyr	0.06 (0.01)[f]	0.49 (0.03)	0.14 (0.02)	0.13 (0.02)
hCatB:Nit3Pyr	-[g]	-	0.20 (0.02)	-
hCatK:Nit3Pyr	0	0.35 (0.07)	-	-
hCatL:Nit3Pyr	0	0.05 (0.01)[h]	-	-
hCatS:Nit3Pyr	0	0.20 (0.01)	-	-

[a] The fraction of frames (= fractional occupancy) displaying each interaction is provided.
[b] H-bonds between N_{pyr} and I87(N)/M68(N)/M70(N)/M71(N) of FP-3/hCatK/hCatL/hCatS.
[c] Water bridge between N_{pyr} and I87(N)/M68(N)/M70(N)/M71(N) of FP-3/hCatK/hCatL/hCatS.
[d] Water bridge between N_{pyr} and E236(OE1,2) of FP-3.
[e] Simultaneous water bridge between N_{pyr} and I87(N), E236(OE1,2) of FP-3.
[f] Mean values were obtained by averaging the fractional occupancies corresponding to each replicate MD simulation. SEMs are shown in parentheses.
[g] Nonexistent interaction due to the chemical nature of the residues that would be involved, P76 in hCatB and hydrophobic residues L209, A214 and F211 of hCatK, hCatL and hCatS, respectively (Figs. 5.4).
[h] The water bridge fractional occupancy of the hCatL:Nit3Pyr complex was calculated based on the MD simulations corresponding to the closely-oriented pose of Nit3Pyr, which was then multiplied by p_A, i.e., the probability finding Nit3Pyr in the close orientation (0.16, see Appendix D, Table D2). The *SEM* in this case was calculated by error propagation.
Source: Prepared by the author.

5.2 Energetic analyses

5.2.1 Prediction of relative affinities of the studied nitriles for FP-2

In Section 5.1.1 we have analyzed the polar interactions between the Pyr moieties of several nitriles and residues of the S2 subsite of FP-2. Here, we extend our study by conducting TI $\Delta\Delta G$ calculations that will allow us to compare the predicted relative affinities with the experimental values and to derive important conclusions regarding the energy contribution of the observed water bridges to the complex formation.

Table 5.3: $\Delta\Delta G$ values for several nitriles bound to FP-2 with respect to the FP-2:Nit3Pyr complex

Energy	**FP-2 complex**					
(kcal/mol)	**Nit6Cl3Pyr**	**Nit6OMe3Pyr**	**Nit2Pyr**	**Nit2Pyr$_{restr.}$**[a]	**Nit4Pyr**	**NitPhe**
$\Delta\Delta G_{calc}$ [b]	1.2 (0.2)[c]	1.8 (0.3)	2.8 (0.2)	2.6 (0.2)	3.1 (0.2)	1.6 (0.1)
$\Delta\Delta G_{calc}^{wb+}$ [d]	n.d.	n.d.	4.3 (0.2)	4.1 (0.2)	4.6 (0.2)	3.1 (0.1)
$\Delta\Delta G_{calc}^{wb-}$ [e]	n.d.[f]	n.d.	1.7 (0.2)	1.5 (0.2)	1.9 (0.1)	0.5 (0.1)
$\Delta\Delta G_{exp}$ [g]	1.5	1.4	>2.9	>2.9	>2.9	2.5
$\Delta\Delta G_{err}$ [h]	0.3	0.4	>0.1	>0.3	n.d.	0.9
$\Delta\Delta G_{err}^{wb+}$ [h]	n.d.	n.d.	n.d.	n.d.	n.d.	0.6
$\Delta\Delta G_{err}^{wb-}$ [h]	n.d.	n.d.	>1.2	>1.4	>1.0	2.0
$\Delta\Delta G_{hys}$ [i]	0.3	0.2	0.2	0.2	0.2	0.1
$\Delta\Delta G_{hys}^{wb+}$ [i]	n.d.	n.d.	0.2	0.2	0.3	0.2
$\Delta\Delta G_{hys}^{wb-}$ [i]	n.d.	n.d.	0.2	0.2	0.2	0.1

[a] Free energies for the FP-2:Nit2Pyr complex were calculated by subjecting it to distance restraints using a 10 kcal·mol^{-1}·Å^{-2} harmonic constant that kept the aromatic carbon atom at position 6 (next to N$_{pyr}$) and I85(N) separated by a distance > 6 Å during all λ MD simulations.
[b] $\Delta\Delta G$ values calculated as indicated in Fig. 2.3-B, with NitX$_2$ standing for the compounds shown in the table and NitX$_1$, for Nit3Pyr. See the corresponding $\partial V/\partial\lambda$ vs λ graphs in Appendix D, Fig. D5.
[c] Mean values obtained from six replicas, three of them performed in the forward direction and the other three, in the backward direction. The *SEM* in parentheses was calculated through the bootstrap method.[90]
[d] $\Delta\Delta G$ values calculated with respect to the FP-2:Nit3Pyr complex in which the I85(N)-N$_{pyr}$ water bridge was enhanced ($wb+$) by restraining the motion of the bridging water O atom with a harmonic constant of 10 kcal·mol^{-1}·Å^{-2} when its distance from I85(N) and/or N$_{pyr}$ increases above 3 Å.
[e] $\Delta\Delta G$ values calculated with respect to the FP-2:Nit3Pyr complex in which the formation of water bridges involving N$_{pyr}$, I85(N) and D234(OD1,2) were prevented (*wb*-) by keeping the former atom away from I85(N) at distances > 6 Å by means of a distance restraint of 10 kcal·mol^{-1}·Å^{-2}.
[f] Non-determined.
[g] Experimental $\Delta\Delta G$ values calculated through Eq. A16 for all complexes, except for FP-2:NitPhe, for which Eq. A17 was used (see Appendix A). The experimental IC_{50}^{app} values in the equations were taken from Fig. 1.5.
[h] Error of calculated $\Delta\Delta G$ values obtained as follows: $\Delta\Delta G_{err} = |\Delta\Delta G_{exp} - \Delta\Delta G_{calc}|$.
[i] $\Delta\Delta G$ hysteresis calculated as follows: $\Delta\Delta G_{hys} = \left|\Delta\Delta G_{fwd} - \Delta\Delta G_{bwd}\right|$, where $\Delta\Delta G_{fwd}$ and $\Delta\Delta G_{bwd}$ stand for the relative free energies calculated during forward and backward transformations, respectively. Mean values and their *SEM*s were obtained through bootstrapping.
Source: Prepared by the author.

Table 5.3 shows the $\Delta\Delta G_{calc}$ values for the same set of FP-2 complexes analyzed before with respect to FP-2:Nit3Pyr, the positive results indicating that the latter complex possesses higher affinity. The predictions for the FP-2:Nit6Cl3Pyr and FP-2:Nit6OMe3Pyr complexes are close to the experimental values, with errors below 0.5 kcal/mol, which are within the level of accuracy for these type of alchemical free energy calculations. In fact, it has been shown that the mean unsigned error (MUE) of TI $\Delta\Delta G$ performed in a large perturbation dataset using ff14SB force-field for the proteins and gaff1.8 for the ligands is 1.14 kcal/mol.[158] Note, however, that Nit6Cl3Pyr and Nit6OMe3Pyr are not correctly ranked according to their respective $\Delta\Delta G$ values, as their slight experimental relative affinity of 0.1 kcal/mol is very difficult to reproduce in alchemical free energy calculations considering the overall MUE associated to the latter.

The lower affinities of Nit2Pyr and Nit4Pyr for FP-2 with respect to that of Nit3Pyr were also reproduced by the free energy calculations, with $\Delta\Delta G_{calc}$ values of 2.8 and 3.1 kcal/mol, respectively (Table 5.3). Of note, the $\Delta\Delta G_{exp}$ values are expected to be > 2.9 kcal/mol, as the IC_{50}'s of these two compounds exceeded the maximal inhibitor concentration tested in the assays (Fig. 1A). Therefore, only the lower boundary of the experimental relative affinities is known and, consequently, the actual deviations of the theoretical values remain elusive. Nonetheless, it can be seen that the results obtained for FP-2:Nit2Pyr and FP-2:Nit4Pyr roughly fulfill the aforementioned experimental condition, with the $\Delta\Delta G_{calc}$ value for the former complex being just 0.1 kcal/mol below the 2.9 kcal/mol threshold ($\Delta\Delta G_{err}$ > 0.1 kcal/mol, Table 5.3). Likewise, the prediction for FP-2:NitPhe is in agreement with the experimental trend, although it underestimates the experimental relative affinity by 0.9 kcal/mol, an error slightly smaller than the MUE reported for this type of calculations.[158] Overall, the FP-2:nitrile structures proposed here lead to $\Delta\Delta G_{calc}$ values that are consistent with the experimental measurements within the level of accuracy of the employed computational techniques, thus reinforcing the validity of such structural models.

We have reasons to believe that the $\Delta\Delta G_{calc}$ values for FP-2:Nit2Pyr, FP2:Nit4Pyr and FP2:NitPhe complexes suffer from a systematic underestimation of the water bridge strength in the FP-2:Nit3Pyr reference complex. This may have arisen from limitations of the MD simulations, such as their failure to account for the polarizability of the atoms involved in the water bridge, especially the water molecule, which is treated exactly as the bulk solvent, despite being located in a deeply-buried and negatively-charged (D234) environment (Appendix D, Fig. D3). In fact, it has been shown that the polarizability of water molecules increases in the presence of charged species, which, in turn, strengthens the H-bonds mediated by them.[159]

A simple procedure was employed here to assess the impact of artificially strengthening the water bridge between I85(N) and N_{pyr} of Nit3Pyr on the $\Delta\Delta G_{calc}$ values by applying restraints that kept the bridging water close (up to 3 Å) to the previous atoms. Under such conditions, the calculated relative free energies ($\Delta\Delta G_{calc}^{wb+}$) are 1.5 kcal/mol greater than the corresponding $\Delta\Delta G_{calc}$ values obtained from unrestrained MD simulations, thus improving the agreement with the experimental measurements (Table 5.3). In fact, the affinities of Nit2Pyr and Nit4Pyr for FP-2 are now predicted to be 4.3 and 4.6 kcal/mol lower than that of Nit3Pyr, respectively. These values are well above the experimental lower boundary (> 2.9 kcal/mol). Remarkably, the $\Delta\Delta G_{calc}^{wb+}$ value for the FP-2:NitPhe complex is also closer to $\Delta\Delta G_{exp}$, the difference between both being reduced from 0.9 to 0.6 kcal/mol (Table 5.3). We did not evaluate here the impact of directly strengthening the D234(OD1,2)-N_{pyr} water bridge, but its formation and that involving I85(N)-N_{pyr} must be highly interdependent, as they depend on the same water molecule. Therefore, by increasing the stability of any these interactions, one probably ends up strengthening the other as well. Overall, the results so far confirm that a stronger water bridge between I85(N) and N_{pyr} of Nit3Pyr would improve the quality of our predictions.

Furthermore, it is worth noting that we did not apply restraints strengthening the I85(N)-N_{pyr} water bridge in the FP-2:Nit6Cl3Pyr and FP-2Nit6OMe3Pyr complexes (Table 5.3). The reason for this is that, despite the stability differences, all the 3Pyr-containing compounds are able to form water bridges and H-bonds with I85(N) (Table 5.1). Therefore, the underestimation of the energy contributions of polar interactions in the aforementioned complexes is expected to largely cancel during the $\Delta\Delta G$ calculations with respect to FP-2:Nit3Pyr. This error cancellation can explain, to some extent, why the $\Delta\Delta G_{err}$ values associated to the $\Delta\Delta G_{calc}$ values of the FP-2:Nit6Cl3Pyr and FP-2:Nit6OMe3Pyr complexes (0.3 and 0.4 kcal/mol, respectively) are smaller than that of FP-2:NitPhe (0.9 kcal/mol). Arguably, the water bridge observed in the FP-2:Nit4Pyr complex (Fig. 5.2, Table 5.1) can also be stronger than it is modelled using a classical non-polarizable force-field. However, the water molecules transiently involved in the previous interaction are more solvent-exposed and exchange faster with the bulk than those bridging D234(OD1,2), I85(N) and N_{pyr} in the FP-2:Nit3Pyr complex, with exchange rates of (10.3±0.2) ns^{-1} and (0.70±0.07) ns^{-1}, respectively. This is expected to lower the energy contribution of the D234(OD1,2)-N_{pyr} water bridge in FP-2:Nit4Pyr with respect to that of I85(N)-D234(OD1,2)-N_{pyr} in FP-2:Nit3Pyr, as already mentioned in Section 5.1.1. Moreover, notice that the unrestrained TI

$\Delta\Delta G$ calculations predict a large affinity difference between FP-2:Nit4Pyr and FP-2:Nit3Pyr (Table 5.3), thus, the process of fine-tuning the strength of the water bridges in both complexes is not likely to decrease their $\Delta\Delta G_{calc}$ value below the 2.9 kcal/mol experimental threshold.

Finally, we determined the effect of abolishing the water bridges involving the N_{pyr} atom of the FP-2:Nit3Pyr complex on the $\Delta\Delta G_{calc}$ values. This was done by applying distance restraints to the latter system in order to keep the N_{pyr} atom in a distant position (state B, Figs. 5.1 and 5.2). The $\Delta\Delta G_{calc}$ values predicted in the presence of such restraints ($\Delta\Delta G_{calc}^{wb-}$) are 1.1 kcal/mol lower than those obtained in their absence (Table 5.3), as the affinity of FP-2:Nit3Pyr complex decreases by the same amount upon water bridge disruption. Consequently, the accuracy of the results is undermined under such circumstances and all $\Delta\Delta G_{err}^{wb-}$ values become greater than 1.0 kcal/mol. Of note, even when no water bridge is formed, the affinity of Nit3Pyr for FP-2 is still higher than that of Nit2Pyr and Nit4Pyr (1.7 and 1.9 kcal/mol, respectively). Hence, other phenomena, presumably of entropic nature, favor the binding of the former compound over the other two. Furthermore, the FP-2:NitPhe and FP-2:Nit3Pyr affinities become similar when the water bridge is disrupted by the distance restraints ($\Delta\Delta G_{calc}^{wb-}$=0.9 kcal/mol, Table 5.3). This means that the substitution of the phenyl group by 3Pyr is advantageous in terms of FP-2 inhibition, to a large extent, because of the ability of the N_{pyr} atom to establish polar interactions within the S2 subsite residues I85 and D234.

As a negative control to the previous calculations, we applied a similar distance restraint to the FP-2:Nit2Pyr complex, but this time involving the carbon at position 6 of the 2Pyr ring, equivalent to that of N_{pyr} in Nit3Pyr, and I85(N). By doing so, the 2Pyr orientation was restricted to state B (Figs. 5.1 and 5.2) during the free energy calculations. The results shown in Table 5.3 demonstrate that the impact of such restraints on the $\Delta\Delta G_{calc}$ values is almost negligible (0.2 kcal/mol), since no polar interactions are broken in this case. Overall, we have shown in this section that the formation of the water bridge involving D234(OD1,2), I85(N) and N_{pyr} is an essential structural feature underlying the higher affinity of Nit3Pyr for FP-2 with respect to that of the related nitriles Nit2Pyr, Nit4Pyr and NitPhe.

5.2.2 Energy contributions of water bridges to the affinity of Nit3Pyr for the studied enzymes

So far, we have shown the occurrence of water bridges between the N_{pyr} atom of Nit3Pyr and S2 residues of FPs and hCats. The multiplicity and stability of such interactions depend on each system, being particularly more prevalent in the FP-2:Nit3Pyr and FP-3:Nit3Pyr complexes, in that order. Here, we conduct MM-GBSA free energy calculations in the presence and absence of the water molecules bridging the N_{pyr} atom of Nit3Pyr and key residues of the studied enzymes (Table 5.4).

Table 5.4. Energy contribution of water bridges of Nit3Pyr in complex with FPs and hCats calculated through MM-GBSA[a]

Complex	$\Delta G_{eff(w)}$ (kcal/mol)	$\Delta G_{eff(nw)}$ (kcal/mol)	$\Delta\Delta G_{eff(wb)}$ (kcal/mol)
FP-2:Nit3Pyr	-30.6 (0.2)[b]	-28.0 (0.2)	-2.56 (0.06)
FP-2:Nit3Pyr(wb+)[c]	-31.3 (0.2)	-28.0 (0.1)	-3.3 (0.2)
FP-3:Nit3Pyr	-34.3 (0.2)	-32.7 (0.1)	-1.63 (0.06)
hCatB:Nit3Pyr	-32.55 (0.05)	-31.72 (0.05)	-0.83 (0.01)
hCatK:Nit3Pyr	-33.0 (0.2)	-31.6 (0.1)	-1.4 (0.1)
hCatL:Nit3Pyr	-31.4 (0.1)	-29.66 (0.05)	-0.28 (0.04)[d]
hCatS:Nit3Pyr	-34.5 (0.3)	-33.5 (0.4)	-0.98 (0.05)

[a] The free energy calculations were conducted for the ten replicate 300 ns MD simulations of every complex; except for hCatL:Nit3Pyr, for which only the trajectories corresponding to the 3Pyr close orientation were taken, and for FP-2:Nit3Pyr(wb+), for which the trajectories of ten replicate 12 ns MD simulations were employed after discarding the first 2 ns of each.
[b] Mean values and *SEM*s in parentheses calculated based on the independent MD simulations.
[c] Distance restraints of 10 kcal·Å^{-2}·mol^{-1} were set on the O atom of the bridging water molecule in order to keep it close (up to 3 Å) to I85(N) and N_{pyr}.
[d] The $\Delta\Delta G_{eff(wb)}$ for this complex was predicted using the following equation: $\Delta\Delta G_{eff(wb)} = p_A\left(\Delta G_{eff(wb)} - \Delta G_{eff(nw)}\right)$, where p_A is the probability of Nit3Pyr to occur in the close orientation (0.16, see Appendix D, Table D2). The *SEM* in this case was calculated by error propagation.
Source: Prepared by the author.

The $\Delta\Delta G_{eff(wb)}$ values in Table 5.4 show that the analyzed water bridges possess by far the largest energy contribution to the affinity, in terms of absolute value, in the FP-2:Nit3Pyr complex (-2.56 kcal/mol). For FP-3 and hCats, the energy contributions of their respective water bridges range from -0.28 to -1.63 kcal/mol, with the former enzyme displaying the largest $\Delta\Delta G_{eff(wb)}$ value (Table 5.4).. These results are in agreement with the relative stabilities and multiplicity of water bridges in the Nit3Pyr complexes shown in previous sections (Tables 5.1 and 5.2). Interestingly, the experimental IC_{50} values indicate that Nit3Pyr binds the studied proteases with the following preference: FP-2 > FP-3 > hCatK > hCatS > hCatL > hCatB (Fig. 1.8), which roughly matches that of water bridge energy contributions in all systems, except for hCatL and hCatB, which are wrongly ranked with respect to each other. However, the nonpolar interactions between the 3Pyr ring and the S2 residues of these enzymes and, probably, entropic effects, also have an impact on the relative affinities of the complexes, as can be inferred from the experimental IC_{50} values of NitPhe (Fig. 1.8). The latter compound contains a similar but more hydrophobic aromatic ring at P2, and has the following preference: hCatK > hCatS > FP-2 > hCatL > hCatB (no data for FP-3 available). Remarkably, Nit3Pyr and NitPhe display the same order of selectivity for the hCats; hence, for complexes in which the water bridge energy contribution is expected to be relatively low (Table 5.4), the influence of other nonpolar and/or entropic effects on the complex affinity is likely to become dominant. In light of our findings, the striking difference in the selectivity of Nit3Pyr and NitPhe for FP-2 can only be attributed to the large stabilizing role of water bridges involving the N_{pyr} atom of the former nitrile, and I85(N) and D234(OD1,2) of the enzyme.

Our previous TI $\Delta\Delta G$ calculations indicated that stronger-than-predicted water bridges in the FP-2:Nit3Pyr complex would be required to improve the agreement with the experimental results (Table 5.3). Likewise, the $\Delta\Delta G_{eff(wb)}$ value for FP-2:Nit3Pyr must have been underestimated, since it was calculated from the classical MD simulations of the complex. The impact of neglecting the water polarizability in the remaining systems is expected to be less significant due to the weaker interaction of the bridging waters with a negatively-charged residue in FP-3 (Tables 5.1 and 5.2 and Appendix D, Fig. D3), and the non-existence of it in hCats. The $\Delta\Delta G_{eff(wb)}$ value obtained from MD simulations in which the bridging water was restrained to a position close to I85(N) of FP-2 and N_{pyr} of Nit3Pyr, as done in the TI $\Delta\Delta G$ calculations, yielded a considerably higher result of -3.3 kcal/mol. This value is more consistent with the large energy contribution of water bridges in the FP-2:Nit3Pyr complex, inferred from the available SAR data.

Finally, note that the $\Delta G_{eff(w)}$ values fail to predict the correct selectivity of Nit3Pyr for the target enzymes. When comparing different complexes, two approximations of the MM-GBSA method employed here are likely to lead to the previous incorrect predictions. Firstly, the contribution of configurational entropy to the complex formation was neglected. This term is usually calculated via Normal Mode Analysis, but usually it does not improve the quality of the results.[136] Secondly, it is intrinsically assumed that the energy contribution of the covalent bond formation is equal for all the complexes involving different cysteine proteases. These issues are less likely to undermine the accuracy of the predictions when referring to relative free energy changes in the same complex, as analyzed in this section. Overall, the results shown here indicate that robust free energy calculations, presumably including the explicit contribution of the covalent bond formation in the different enzymes, would be required to predict accurate binding affinities of Nit3Pyr for the cysteine proteases. This, however, is beyond the scope of the current work and not necessary to determine the role of water bridges in the studied complexes.

5.3 Discussion

Ehmke *et al.* first reported the good selectivity for FP-2 displayed by nitriles containing 3Pyr at P2.[43] They also proposed a binding mode of a 3Pyr-containing nitrile to FP-2 and observed that the N_{pyr} atom was close to I85(N). However, no information on how these two atoms interact and the potential role of the interaction in the selectivity was provided.[35] Further clarifications on this point have not been made ever since, but the properties of the 3Pyr substituent were subsequently exploited by Nizi *et al.* in a work pursuing the identification of selective FP-2 inhibitors.[40] The lack of structural knowledge on the interaction of the 3Pyr ring with FP-2 highlights the importance of addressing this topic.

In this work, we have thoroughly studied the interaction with hCats and FPs of several nitriles containing Pyr isomers and derivatives, and phenyl groups at P2.[40] Firstly, we ruled out, through free energy calculations and structural analyses, the binding of nitriles bearing protonated Pyr rings to FP-2, and, by inference, to the remaining cysteine proteases included here. Hence, all subsequent analyses were conducted with the neutral forms of the compounds. The preferential binding of the neutral forms is reinforced by experimental data showing that nitriles containing substituted 3Pyr rings with low basicity, e.g., Nit6Cl3Pyr and Nit6OMe3Pyr, still display high affinity for FP-2.[40] Therefore, the protonation of N_{pyr} must not be the underlying cause leading to the greater affinity of Nit3Pyr and its derivatives (Nit6Cl3Pyr and Nit6OMe3Pyr) for FP-2 with respect to that of its isomers or analogs (Nit4Pyr, Nit2Pyr and NitPhe).[40] This result was rather surprising, as, in theory, the protonated 3Pyr ring can establish a H-bond with D234(OD1,2) of FP-2. This is probably the reason that led the authors reporting the SAR data of these compounds, to speculate about the role of the Pyr ring basicity in the selectivity of Nit3Pyr for FP-2.[40] Furthermore, the experimental studies showed that the activity profiles of the nitriles against FP-2 were not improved by introducing positively-charged lysine-like moieties at P2.[40] Altogether, the theoretical and experimental findings mentioned above discourage the design of FP-2 inhibitors bearing positively-charged groups at P2 capable, in principle, of interacting with D234(OD1,2), as has been previously recommended.[8]

We demonstrate in this work that residues I85 and D234 at the bottom of the S2 subsite of FP-2 are key structural determinants for the interaction of FP-2 and Nit3Pyr. These residues are capable of forming either individual or simultaneous water bridges with the N_{pyr} atom of Nit3Pyr during the MD simulations, involving a water molecule that is present in the crystal structure 2OUL. A detrimental impact on the formation of these interactions was observed when placing the N_{pyr} atom in other positions of the Pyr ring. Further structural and energetic analyses conducted here showed that equivalent water bridges in FP-3 and various hCats are less prevalent and energetically-relevant, which, in turn, explains the increased selectivity of Nit3Pyr for FP-2. The main reason underlying the differential pattern of water bridge formation in the analyzed complexes is the nature of non-conserved residues occupying positions equivalent to I85 and D234 of FP-2 in the other cysteine proteases. Remarkably, the structural alignment comprising FP-2, FP-3 and eight human cysteine cathepsins (B, C, F, K, L, S, V and X) shows that none of the enzymes in the latter group possesses a combination of non-proline and acidic residues aligned to I85 and D234 of FP-2, respectively (positions 107 and 358 of the alignment shown in Appendix B, Fig. B1). Therefore, Nit3Pyr is expected to display lower affinity for other hCats not included in the experiments by Nizi *et al.*,[40] which highlights its suitability to reduce the potential unwanted effects of cross-inhibition in human cells.

Based on the previous results, we propose to address the design of selective FP-2 inhibitors by taking into account the formation of water bridges with residues I85 and D234. This can be achieved by keeping the crystal water close to I85(N) in the PDB structure 2OUL as part of the receptor in SBVSs,[160-162] or by taking for such purposes the FP-2 central structure including the water molecule bridging I85(N) and D234(OD1,2) obtained from MD simulations (Appendix D, Fig. D6-A). Rescoring methods, such as Nwat-MMGBSA, can be employed to further enrich the compound lists obtained from SBVS with potential hits interacting with the key water molecule.[85] Redesign strategies of the P2 substituents, in which the bridging water molecule is replaced by groups mimicking its interactions with I85(N) and D234(OD1,2) constitutes another strategy to increase the affinity and selectivity of FP-2 inhibitors. However, according to the results obtained by Garcia-Sosa, the success of this approach is not always guaranteed, and better results may be obtained by keeping the tightly-bound bridging water molecule.[163] Therefore, the use of rigorous free energy calculations, such as TI or free energy perturbation (FEP), would be required to account for the impact of introducing such substituents.[90, 158, 164, 165]

Interestingly, Shah *et al.* employed a computational approach based on solvent thermodynamics, referred to as WaterMap, to predict the role of water molecules at the active sites of FP-2 and FP-3 in the affinity of these enzymes for several inhibitors.[166] The authors predicted that more favorable binding free energies could be achieved by displacing the water molecules near I85/87 of FP-2/3 (labeled W3 and W9, respectively, in the cited work) by complementary chemical groups of the ligands.[166] However, the impact of such water displacement on the selectivity of redesigned inhibitors for FP-2 and/or FP-3 with respect to hCats constitutes another issue that must be carefully assessed in a case-by-case fashion.

Overall, our study has revealed differences in the formation of key water bridges at the bottom of the S2 subsite of FP2, FP3 and several hCats, which involve a water molecule close to I85(N) of FP2 and its equivalents in the remaining proteases. The major role played by such interactions in determining the activity profile of Nit3Pyr against the previous enzymes opens the gate toward the design of more potent and selective inhibitors against FP-2.

6 CONCLUDING REMARKS

In this work we have explored different *in silico* approaches aiming at the identification of novel FP-2 inhibitors with improved selectivity. Two promising compounds, HTS07940 and HTS08262, were identified through SBVSs directed towards the active sites of FP-2, FP-3 and hCatK and subsequent experimental validation. The compounds display inhibitory activity against FP-2 in the 10^{-5}-10^{-6} M range and no inhibition of hCatK was detected. The higher selectivity for FP-2 was attributed to differences in the residue composition at the bottom of the S2 subsite among the analyzed cysteine proteases. In addition, HTS07940 and HTS08262 possess micromolar inhibition of *P. falciparum* growth *in vitro* and suitable selectivity indices when tested against HeLa cells. Both compounds belong to a new scaffold of FP-2 inhibitor, which increases the chemical repertoire of antimalarial drug candidates. We also probed the potential of site 6 of FP-2 as a druggable allosteric site through SBVS. Two compounds, ZINC03225317 and ZINC72290660, were identified as non-competitive FP-2 inhibitors after the experimental validation. Despite their weak inhibitory activity against FP-2, K_i values within the 10^{-4}-10^{-3} M range, our results show the suitability of site 6 as a druggable cavity in FP-2.

Moreover, the search for other potential allosteric cavities of FP-2 was expanded by predicting the binding mode of a reported non-competitive inhibitor, Cpd66, with available experimental information on the likely location of its binding site. The formation of a transient pocket at the site 3 region of FP-2, which remains mostly occluded by K34 side-chain, was revealed. The predicted binding mode of Cpd66 within the identified pocket fulfills all the experimental findings, thus reinforcing the potential of site 3 as an allosteric pocket. In accordance with the established classical non-competitive mechanism of Cpd66, the absence of conformational changes at the active site upon the compound binding was corroborated *in silico*. It was shown that a combination of subtle structural aspects, e.g., variations in loop motions, community and signaling pathways rearrangements, and pK_a shifts, constitute the likely molecular fingerprint of an impaired catalytic function of FP-2 when bound to Cpd66.

Finally, we conducted a computational study to decipher the bases of the selectivity for FP-2 of nitriles containing Pyr derivatives at P2. Our results revealed that water bridges involving residues I85 and D234 of FP-2, and the deprotonated nitrogen of the 3Pyr moiety at P2, which are either less prevalent or nonexistent in the analog complexes involving hCats, explain the SAR data.

The presence of water molecules in the crystal structures of the studied enzymes close to the bridging water positions reinforces the likelihood of such interactions occurring. The presence of the nitrogen atom in positions 2 and 4 of the pyridine ring, also precludes the formation of the stabilizing water bridges within the S2 subsite of FP-2. Our results show that selective FP-2 inhibitors can be designed by promoting the formation of the above-mentioned water bridge and/or by introducing substituents that displace the bridging water.

REFERENCES

(1) Sachs, J.; Malaney, P., The Economic and Social Burden of Malaria. . *Nature* **2002**, 415, 680–685.

(2) World Malaria Report 2017 http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/.

(3) Teixeira, C.; Gomes, J. R.; Gomes, P., Falcipains, *Plasmodium falciparum* Cysteine Proteases as Key Drug Targets against Malaria. *Curr Med Chem* **2011**, 18, 1555-1572.

(4) Rosenthal, P. J., Falcipains and Other Cysteine Proteases of Malaria Parasites. *Adv Exp Med Biol* **2011**, 712, 30-48.

(5) Rowe, J. A.; Claessens, A.; Corrigan, R. A.; Arman, M., Adhesion of *Plasmodium falciparum*-Infected Erythrocytes to Human Cells: Molecular Mechanisms and Therapeutic Implications. *Expert Rev Mol Med* **2009**, 11, e16.

(6) Zheng, J.; Pan, H.; Gu, Y.; Zuo, X.; Ran, N.; Yuan, Y.; Zhang, C.; Wang, F., Prospects for Malaria Vaccines: Pre-Erythrocytic Stages, Blood Stages, and Transmission-Blocking Stages. *Biomed Res Int* **2019**, 2019, 9751471.

(7) Bekono, B. D.; Ntie-Kang, F.; Owono Owono, L. C.; Megnassan, E., Targeting Cysteine Proteases from *Plasmodium falciparum*: A General Overview, Rational Drug Design and Computational Approaches for Drug Discovery. *Curr Drug Targets* **2018**, 501-526.

(8) Ettari, R.; Bova, F.; Zappala, M.; Grasso, S.; Micale, N., Falcipain-2 Inhibitors. *Med Res Rev* **2010**, 30, 136-167.

(9) Shenai, B. R.; Sijwali, P. S.; Singh, A.; Rosenthal, P. J., Characterization of Native and Recombinant Falcipain-2, a Principal Trophozoite Cysteine Protease and Essential Hemoglobinase of *Plasmodium falciparum*. *J Biol Chem* **2000**, 275, 29000-29010.

(10) Sijwali, P. S.; Shenai, B. R.; Gut, J.; Singh, A.; Rosenthal, P. J., Expression and Characterization of the *Plasmodium falciparum* Haemoglobinase Falcipain-3. *Biochem J* **2001**, 360, 481-489.

(11) Roy, K. K., Targeting the Active Sites of Malarial Proteases for Antimalarial Drug Discovery: Approaches, Progress and Challenges. *Int J Antimicrob Agents* **2017**, 50, 287-302.

(12) Sijwali, P. S.; Koo, J.; Singh, N.; Rosenthal, P. J., Gene Disruptions Demonstrate Independent Roles for the Four Falcipain Cysteine Proteases of *Plasmodium falciparum*. *Mol Biochem Parasitol* **2006**, 150, 96-106.

(13) Kerr, I. D.; Lee, J. H.; Pandey, K. C.; Harrison, A.; Sajid, M.; Rosenthal, P. J.; Brinen, L. S., Structures of Falcipain-2 and Falcipain-3 Bound to Small Molecule Inhibitors: Implications for Substrate Specificity. *J Med Chem* **2009**, 52, 852-857.

(14) Pandey, K. C.; Wang, S. X.; Sijwali, P. S.; Lau, A. L.; McKerrow, J. H.; Rosenthal, P. J., The *Plasmodium falciparum* Cysteine Protease Falcipain-2 Captures Its Substrate, Hemoglobin, Via a Unique Motif. *Proc Natl Acad Sci U S A* **2005**, 102, 9138-9143.

(15) Sijwali, P. S.; Shenai, B. R.; Rosenthal, P. J., Folding of the *Plasmodium falciparum* Cysteine Protease Falcipain-2 Is Mediated by a Chaperone-Like Peptide and Not the Prodomain. *J Biol Chem* **2002**, 277, 14910-14915.

(16) Cotrin, S. S.; Gouvea, I. E.; Melo, P. M.; Bagnaresi, P.; Assis, D. M.; Araujo, M. S.; Juliano, M. A.; Gazarini, M. L.; Rosenthal, P. J.; Juliano, L.; Carmona, A. K., Substrate Specificity Studies of the Cysteine Peptidases Falcipain-2 and Falcipain-3 from *Plasmodium falciparum* and Demonstration of Their Kininogenase Activity. *Mol Biochem Parasitol* **2013**, 187, 111-116.

(17) Teixeira, C.; Gomes, J. R.; Couesnon, T.; Gomes, P., Molecular Docking and 3d-Quantitative Structure Activity Relationship Analyses of Peptidyl Vinyl Sulfones: *Plasmodium falciparum* Cysteine Proteases Inhibitors. *J Comput Aided Mol Des* **2011**, 25, 763-775.

(18) Fernandez, L.; Bustos, R. H.; Zapata, C.; Garcia, J.; Jauregui, E.; Ashraf, G. M., Immunogenicity in Protein and Peptide Based-Therapeutics: An Overview. *Curr Protein Pept Sci* **2018**, 19, 958-971.

(19) Cianni, L.; Feldmann, C. W.; Gilberg, E.; Gutschow, M.; Juliano, L.; Leitao, A.; Bajorath, J.; Montanari, C. A., Can Cysteine Protease Cross-Class Inhibitors Achieve Selectivity? *J Med Chem* **2019**, 10497-10525.

(20) Lu, S.; Li, S.; Zhang, J., Harnessing Allostery: A Novel Approach to Drug Discovery. *Med Res Rev* **2014**, 34, 1242-1285.

(21) Liu, J.; Nussinov, R., Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput Biol* **2016**, 12, e1004966.

(22) Nussinov, R.; Tsai, C. J., Allostery in Disease and in Drug Discovery. *Cell* **2013**, 153, 293-305.

(23) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E., Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem Rev* **2016**, 116, 6370-6390.

(24) Li, Z.; Kienetz, M.; Cherney, M. M.; James, M. N.; Bromme, D., The Crystal and Molecular Structures of a Cathepsin K:Chondroitin Sulfate Complex. *J Mol Biol* **2008**, 383, 78-91.

(25) Novinec, M.; Korenc, M.; Caflisch, A.; Ranganathan, R.; Lenarcic, B.; Baici, A., A Novel Allosteric Mechanism in the Cysteine Peptidase Cathepsin K Discovered by Computational Methods. *Nat Commun* **2014**, 5, 3287.

(26) Novinec, M.; Lenarcic, B.; Baici, A., Probing the Activity Modification Space of the Cysteine Peptidase Cathepsin K with Novel Allosteric Modifiers. *PLoS One* **2014**, 9, e106642.

(27) Novinec, M.; Rebernik, M.; Lenarcic, B., An Allosteric Site Enables Fine-Tuning of Cathepsin K by Diverse Effectors. *FEBS Lett* **2016**, 590, 4507-4518.

(28) Li, Z.; Yasuda, Y.; Li, W.; Bogyo, M.; Katz, N.; Gordon, R. E.; Fields, G. B.; Bromme, D., Regulation of Collagenase Activities of Human Cathepsins by Glycosaminoglycans. *J Biol Chem* **2004**, 279, 5470-5479.

(29) Almeida, P. C., et al, Cathepsin B Activity Regulation. Heparin-Like Glycosaminogylcans Protect Human Cathepsin B from Alkaline pH-Induced Inactivation. *J Biol Chem* **2001**, 276, 944-951.

(30) Judice, W. A., et al, Heparin Modulates the Endopeptidase Activity of *Leishmania mexicana* Cysteine Protease Cathepsin L-Like Rcpb2.8. *PLoS One* **2013**, 8, e80153.

(31) Lima, A. P., et al, Heparan Sulfate Modulates Kinin Release by *Trypanosoma cruzi* through the Activity of Cruzipain. *J Biol Chem* **2002**, 277, 5875-5881.

(32) Costa, T. F.; dos Reis, F. C.; Lima, A. P., Substrate Inhibition and Allosteric Regulation by Heparan Sulfate of *Trypanosoma brucei* Cathepsin L. *Biochim Biophys Acta* **2012** 1824, 493–501.

(33) Almeida, P. C., et al Cysteine Proteinase Activity Regulation. A Possible Role of Heparin and Heparin-Like Glycosaminoglycans. *J Biol Chem* **1999**, 274, 30433-30438.

(34) Marques, A. F.; Esser, D.; Rosenthal, P. J.; Kassack, M. U.; Lima, L. M., Falcipain-2 Inhibition by Suramin and Suramin Analogues. *Bioorg Med Chem* **2013**, 21, 3667-3673.

(35) Marques, A. F.; Gomes, P. S.; Oliveira, P. L.; Rosenthal, P. J.; Pascutti, P. G.; Lima, L. M., Allosteric Regulation of the *Plasmodium falciparum* Cysteine Protease Falcipain-2 by Heme. *Arch Biochem Biophys* **2015**, 573, 92-99.

(36) Bertoldo, J. B.; Chiaradia-Delatorre, L. D.; Mascarello, A.; Leal, P. C.; Cordeiro, M. N.; Nunes, R. J.; Sarduy, E. S.; Rosenthal, P. J.; Terenzi, H., Synthetic Compounds from an in House Library as Inhibitors of Falcipain-2 from *Plasmodium falciparum*. *J Enzyme Inhib Med Chem* **2015**, 30, 299-307.

(37) Pant, A.; Kumar, R.; Wani, N. A.; Verma, S.; Sharma, R.; Pande, V.; Saxena, A. K.; Dixit, R.; Rai, R.; Pandey, K. C., Allosteric Site Inhibitor Disrupting Auto-Processing of Malarial Cysteine Proteases. *Sci Rep* **2018**, 8, 16193.

(38) Alberca, L. N.; Chuguransky, S. R.; Alvarez, C. L.; Talevi, A.; Salas-Sarduy, E., *In Silico* Guided Drug Repurposing: Discovery of New Competitive and Non-Competitive Inhibitors of Falcipain-2. *Front Chem* **2019**, 7, 534.

(39) Machin, J. M.; Kantsadi, A. L.; Vakonakis, I., The Complex of *Plasmodium falciparum* Falcipain-2 Protease with an (E)-Chalcone-Based Inhibitor Highlights a Novel, Small, Molecule-Binding Site. *Malar J* **2019**, 18, 388.

(40) Nizi, E.; Sferrazza, A.; Fabbrini, D.; Nardi, V.; Andreini, M.; Graziani, R.; Gennari, N.; Bresciani, A.; Paonessa, G.; Harper, S., Peptidomimetic Nitrile Inhibitors of Malarial Protease Falcipain-2 with High Selectivity against Human Cathepsins. *Bioorg Med Chem Lett* **2018**, 28, 1540-1544.

(41) Verissimo, E.; Berry, N.; Gibbons, P.; Cristiano, M. L.; Rosenthal, P. J.; Gut, J.; Ward, S. A.; O'Neill, P. M., Design and Synthesis of Novel 2-Pyridone Peptidomimetic Falcipain 2/3 Inhibitors. *Bioorg Med Chem Lett* **2008**, 18, 4210-4214.

(42) Lee, B. J.; Singh, A.; Chiang, P.; Kemp, S. J.; Goldman, E. A.; Weinhouse, M. I.; Vlasuk, G. P.; Rosenthal, P. J., Antimalarial Activities of Novel Synthetic Cysteine Protease Inhibitors. *Antimicrob Agents Chemother* **2003**, 47, 3810-3814.

(43) Ehmke, V.; Kilchmann, F.; Heindl, C.; Cui, K.; Huang, J.; Schirmeister, T.; Diederich, F., Peptidomimetic Nitriles as Selective Inhibitors for the Malarial Cysteine Protease Falcipain-2. *MedChemComm* **2011**, 2, 800-804.

(44) Webb, B.; Sali, A., Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinformatics* **2016**, 5 6 1-5 6 37.

(45) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res* **2007**, 35, W522-525.

(46) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A., Comparative Protein Structure Modeling of Genes and Genomes. *Annu Rev Biophys Biomol Struct* **2000**, 29, 291-325.

(47) Sali, A.; Blundell, T. L., Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol* **1993**, 234, 779-815.

(48) Rognan, D., Docking Methods for Virtual Screening: Principles and Recent Advances. In *Virtual Screening*. Sotriffer C., Ed.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, 2011; Chapter 6, pp 153-176.

(49) Trott, O.; Olson, A. J., Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **2010**, 31, 455-461.

(50) Hernandez-Gonzalez, J. E.; Salas-Sarduy, E.; Hernandez Ramirez, L. F.; Pascual, M. J.; Alvarez, D. E.; Pabon, A.; Leite, V. B. P.; Pascutti, P. G.; Valiente, P. A., Identification of (4-(9H-Fluoren-9-yl) Piperazin-1-yl) Methanone Derivatives as Falcipain 2 Inhibitors Active against *Plasmodium falciparum* Cultures. *Biochim Biophys Acta Gen Subj* **2018**, 1862, 2911-2923.

(51) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., Autodock4 and Autodocktools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* **2009**, 30, 2785-2791.

(52) Hernandez Gonzalez, J. E.; Hernandez Alvarez, L.; Pascutti, P. G.; Valiente, P. A., Predicting Binding Modes of Reversible Peptide-Based Inhibitors of Falcipain-2 Consistent with Structure-Activity Relationships. *Proteins* **2017**, 85, 1666-1683.

(53) Hernandez Alvarez, L.; Barreto Gomes, D. E.; Hernandez Gonzalez, J. E.; Pascutti, P. G., Dissecting a Novel Allosteric Mechanism of Cruzain: A Computer-Aided Approach. *PLoS One* **2019**, 14, e0211227.

(54) Nguyen, H.; Roe, D. R.; Simmerling, C., Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J Chem Theory Comput* **2013**, 9, 2020-2034.

(55) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; T E Cheatham, I.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; C Sagui; Salomon-Ferrer, R.; Seabra, G.;

Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *Amber14* University of California, San Francisco, 2014.

(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

(57) Molecular Operating Environment (MOE), 2013.08;Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016.

(58) Hernandez Gonzalez, J. E.; Hernandez Alvarez, L.; Pascutti, P. G.; Leite, V. B. P., Prediction of Noncompetitive Inhibitor Binding Mode Reveals Promising Site for Allosteric Modulation of Falcipain-2. *J Phys Chem B* **2019**, 123, 7327-7342.

(59) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M., Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J Comput Chem* **2010**, 31, 797-810.

(60) Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *J Comput Chem* **2011**, 32, 866-877.

(61) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J Cheminform* **2012**, 4, 17.

(62) 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.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *Amber 2016*, University of California, San Francisco, 2016.

(63) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails,

J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*, University of California, San Francisco, 2018.

(64) 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.

(65) Norberg, J.; Nilsson, L., Advances in Biomolecular Simulations: Methodology and Recent Applications. *Q. Rev. Biophys.* **2003**, 36, 257-306.

(66) MacKerell, A. D. Atomistic Models and Force Fields. In *Computational Biochemistry and Biophysics*, Becker, O. M.; MacKerell, A. D.; Roux, B.; Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, NY, 2001; Chapter 2, pp 7-38.

(67) Hug, S., Classical Molecular Dynamics in a Nutshell. *Methods Mol Biol* **2013**, 924, 127-152.

(68) Lindahl, E. R. Molecular Dynamics Simulations. In *Methods in Molecular Biology (Methods and Protocols)*, Kukol, A., Ed.; Humana Press: New York, NY, 2015; Vol. 1215, pp 3-26.

(69) 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.

(70) Reif, M. M.; Hunenberger, P. H.; Oostenbrink, C., New Interaction Parameters for Charged Amino Acid Side Chains in the Gromos Force Field. *J Chem Theory Comput* **2012**, 8, 3705-3723.

(71) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, 1, 19-25.

(72) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., Lincs: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* **1997**, 18, 1463-1472.

(73) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, 81, 3684-3690.

(74) Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling through Velocity Rescaling. *J Chem Phys* **2007**, 126, 014101.

(75) Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1981**, 52, 7182-7190.

(76) Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, 120, 11919-11929.

(77) Wereszczynski, J.; McCammon, J. A., Accelerated Molecular Dynamics in Computational Drug Design. *Methods Mol Biol* **2012**, 819, 515-524.

(78) Swails, J. M.; York, D. M.; Roitberg, A. E., Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J Chem Theory Comput* **2014**, 10, 1341-1352.

(79) Socher, E.; Sticht, H., Mimicking Titration Experiments with Md Simulations: A Protocol for the Investigation of pH-Dependent Effects on Proteins. *Sci Rep* **2016**, 6, 22523.

(80) Mathematica, Version 7.0; Wolfram Research. Inc: Champaign, Illinois, 2008.

(81) Hayes, J. M.; Archontis, G. MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies. In *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*, Wang, L., Ed.; InTech: Rijeka, Croatia, 2012; Chapter 9, pp 171-190.

(82) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, 112, 6127–6129.

(83) Miller, B. R., 3rd; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* **2012**, 8, 3314-3321.

(84) Maffucci, I.; Contini, A., Improved Computation of Protein-Protein Relative Binding Energies with the Nwat-MMGBSA Method. *J Chem Inf Model* **2016**, 56, 1692-1704.

(85) Maffucci, I.; Hu, X.; Fumagalli, V.; Contini, A., An Efficient Implementation of the Nwat-MMGBSA Method to Rescore Docking Results in Medium-Throughput Virtual Screenings. *Front Chem* **2018**, 6, 43.

(86) Martins, S. A.; Perez, M. A.; Moreira, I. S.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A., Computational Alanine Scanning Mutagenesis: MM-PBSA vs TI. *J Chem Theory Comput* **2013**, 9, 1311-1319.

(87) Kuhn, B.; Tichy, M.; Wang, L.; Robinson, S.; Martin, R. E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; Diederich, F.; Hert, J., Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *J Med Chem* **2017**, 60, 2485-2497.

(88) Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H., Calculation of Binding Free Energies. *Methods Mol Biol* **2015**, 1215, 173-209.

(89) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L., Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J Chem Theory Comput* **2018**, 14, 6346-6358.

(90) Bhati, A. P.; Wan, S.; Wright, D. W.; Coveney, P. V., Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *J Chem Theory Comput* **2017**, 13, 210-222.

(91) Graf, M. M.; Maurer, M.; Oostenbrink, C., Free-Energy Calculations of Residue Mutations in a Tripeptide Using Various Methods to Overcome Inefficient Sampling. *J Comput Chem* **2016**, 37, 2597-2605.

(92) Kästner, J., Umbrella Sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1, 932-942.

(93) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J Chem Theory Comput* **2010**, 6, 3713-3720.

(94) Doudou, S.; Burton, N. A.; Henchman, R. H., Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J Chem Theory Comput* **2009**, 5, 909-918.

(95) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins* **2010**, 78, 1950-1958.

(96) Wagner, J. R.; Sorensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E., POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J Chem Theory Comput* **2017**, 13, 4584-4592.

(97) Bowerman, S.; Wereszczynski, J., Detecting Allosteric Networks Using Molecular Dynamics Simulation. *Methods Enzymol* **2016**, 578, 429-447.

(98) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z., Dynamical Networks in Trna:Protein Complexes. *Proc Natl Acad Sci U S A* **2009**, 106, 6620-6625.

(99) Girvan, M.; Newman, M. E., Community Structure in Social and Biological Networks. *Proc Natl Acad Sci U S A* **2002**, 99, 7821-7826.

(100) Yao, X. Q.; Malik, R. U.; Griggs, N. W.; Skjaerven, L.; Traynor, J. R.; Sivaramakrishnan, S.; Grant, B. J., Dynamic Coupling and Allosteric Networks in the Alpha Subunit of Heterotrimeric G Proteins. *J Biol Chem* **2016**, 291, 4742-4753.

(101) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E., Weighted Implementation of Suboptimal Paths (Wisp): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J Chem Theory Comput* **2014**, 10, 511-517.

(102) Lange, O. F.; Grubmuller, H., Generalized Correlation for Biomolecular Dynamics. *Proteins* **2006**, 62, 1053-1061.

(103) Skjaerven, L.; Yao, X. Q.; Scarabelli, G.; Grant, B. J., Integrating Protein Structural Dynamics and Evolutionary Analysis with Bio3d. *BMC Bioinformatics* **2014**, 15, 399.

(104) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C., Essential Dynamics of Proteins *Proteins* **1993**, 17, 412-425.

(105) DeLano, W. L. *Pymol*, 2.1.0; DeLano Scientific: San Carlos, CA, 700, 2002.

(106) Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J Mol Graph* **1996**, 14, 33-38, 27-38.

(107) Lerbret, A.; Mason, P. E.; Venable, R. M.; Cesaro, A.; Saboungi, M. L.; Pastor, R. W.; Brady, J. W., Molecular Dynamics Studies of the Conformation of Sorbitol. *Carbohydr Res* **2009**, 344, 2229-2235.

(108) Sarduy, E. S.; Munoz, A. C.; Trejo, S. A.; de los, A. C. P. M., High-Level Expression of Falcipain-2 in Escherichia Coli by Codon Optimization and Auto-Induction. *Protein Expr Purif* **2012**, 83, 59-69.

(109) Linnevers, C. J.; McGrath, M. E.; Armstrong, R.; Mistry, F. R.; Barnes, M. G.; Klaus, J. L.; Palmer, J. T.; Katz, B. A.; Bromme, D., Expression of Human Cathepsin K in Pichia Pastoris and Preliminary Crystallographic Studies of an Inhibitor Complex. *Protein Sci* **1997**, 6, 919-921.

(110) GraphPad Software, La Jolla California USA, www.graphpad.com.

(111) Trager, W.; Jensen, J. B., Human Malaria Parasites in Continuous Culture. *Science* **1976**, 193, 673-675.

(112) Pabon, A.; Ramirez, O.; Rios, A.; Lopez, E.; de Las Salas, B.; Cardona, F.; Blair, S., Antiplasmodial and Cytotoxic Activity of Raw Plant Extracts as Reported by Knowledgeable Indigenous People of the Amazon Region (Vaupes Medio in Colombia). *Planta Med* **2016**, 82, 717-722.

(113) Bianco, A. E.; Favaloro, J. M.; Burkot, T. R.; Culvenor, J. G.; Crewther, P. E.; Brown, G. V.; Anders, R. F.; Coppel, R. L.; Kemp, D. J., A Repetitive Antigen of *Plasmodium falciparum* That Is Homologous to Heat Shock Protein 70 of Drosophila Melanogaster. *Proc Natl Acad Sci U S A* **1986**, 83, 8713-8717.

(114) Pascual, M. J.; Merwaiss, F.; Leal, E.; Quintana, M. E.; Capozzo, A. V.; Cavasotto, C. N.; Bollini, M.; Alvarez, D. E., Structure-Based Drug Design for Envelope Protein E2 Uncovers a New Class of Bovine Viral Diarrhea Inhibitors That Block Virus Entry. *Antiviral Res* **2018**, 149, 179-190.

(115) Copeland, R. A. Assay Considerations for Compound Library Screening In *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*; John Wiley & Sons: 2013; Chapter 4, pp 123-168.

(116) Selwyn, M. J., A Simple Test for Inactivation of an Enzyme During Assay. *Biochim Biophys Acta* **1965**, 105, 193-195.

(117) In *Assay Guidance Manual*, Sittampalam, G. S.; Grossman, A.; Brimacombe, K.; Arkin, M.; Auld, D.; Austin, C. P.; Baell, J.; Bejcek, B.; Caaveiro, J. M. M.; Chung, T. D. Y.; Coussens, N. P.; Dahlin, J. L.; Devanaryan, V.; Foley, T. L.; Glicksman, M.; Hall, M. D.; Haas, J. V.; Hoare, S. R. J.; Inglese, J.; Iversen, P. W.; Kahl, S. D.; Kales, S. C.; Kirshner, S.; Lal-Nag, M.; Li, Z.; McGee, J.; McManus, O.; Riss, T.; Saradjian, P.; Trask, O. J., Jr.; Weidner, J. R.; Wildey, M. J.; Xia, M.; Xu, X., Eds.; Bethesda (MD), 2004.

(118) Ettari, R.; Zappala, M.; Micale, N.; Grazioso, G.; Giofre, S.; Schirmeister, T.; Grasso, S., Peptidomimetics Containing a Vinyl Ketone Warhead as Falcipain-2 Inhibitors. *Eur J Med Chem* **2011**, 46, 2058-2065.

(119) Alves, M. F.; Puzer, L.; Cotrin, S. S.; Juliano, M. A.; Juliano, L.; Bromme, D.; Carmona, A. K., S3 to S3' Subsite Specificity of Recombinant Human Cathepsin K and Development of Selective Internally Quenched Fluorescent Substrates. *Biochem J* **2003**, 373, 981-986.

(120) Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D., Cysteine Cathepsins: From Structure, Function and Regulation to New Frontiers. *Biochim Biophys Acta* **2012**, 1824, 68-88.

(121) Cotrin, S. S.; Puzer, L.; de Souza Judice, W. A.; Juliano, L.; Carmona, A. K.; Juliano, M. A., Positional-Scanning Combinatorial Libraries of Fluorescence Resonance Energy Transfer Peptides to Define Substrate Specificity of Carboxydipeptidases: Assays with Human Cathepsin B. *Anal Biochem* **2004**, 335, 244-252.

(122) Puzer, L.; Cotrin, S. S.; Alves, M. F.; Egborge, T.; Araujo, M. S.; Juliano, M. A.; Juliano, L.; Bromme, D.; Carmona, A. K., Comparative Substrate Specificity Analysis of Recombinant Human Cathepsin V and Cathepsin L. *Arch Biochem Biophys* **2004**, 430, 274-283.

(123) Bromme, D.; Bonneau, P. R.; Lachance, P.; Storer, A. C., Engineering the S2 Subsite Specificity of Human Cathepsin S to a Cathepsin L- and Cathepsin B-Like Specificity. *J Biol Chem* **1994**, 269, 30238-30242.

(124) Baell, J. B.; Holloway, G. A., New Substructure Filters for Removal of Pan Assay Interference Compounds (Pains) from Screening Libraries and for Their Exclusion in Bioassays. *J Med Chem* **2010**, 53, 2719-2740.

(125) Kuo, M. R.; Morbidoni, H. R.; Alland, D.; Sneddon, S. F.; Gourlie, B. B.; Staveski, M. M.; Leonard, M.; Gregory, J. S.; Janjigian, A. D.; Yee, C.; Musser, J. M.; Kreiswirth, B.; Iwamoto, H.; Perozzo, R.; Jacobs, W. R., Jr.; Sacchettini, J. C.; Fidock, D. A., Targeting Tuberculosis and Malaria through Inhibition of Enoyl Reductase: Compound Activity and Structural Data. *J Biol Chem* **2003**, 278, 20851-20859.

(126) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E., Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem Rev* **2002**, 102, 4639-4750.

(127) Storer, A. C.; Ménard, R., Recent Insights into Cysteine Protease Specificity: Lessons for Drug Design. *Perpect Drug Discovery Des* **1996**, 6, 33-46.

(128) Shah, F.; Mukherjee, P.; Gut, J.; Legac, J.; Rosenthal, P. J.; Tekwani, B. L.; Avery, M. A., Identification of Novel Malarial Cysteine Protease Inhibitors Using Structure-Based Virtual Screening of a Focused Cysteine Protease Inhibitor Library. *J Chem Inf Model* **2011**, 51, 852-864.

(129) Changeux, J. P.; Edelstein, S., Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *F1000 Biol Rep* **2011**, 3, 19.

(130) Hernandez Alvarez, L.; Naranjo Feliciano, D.; Hernandez Gonzalez, J. E.; Soares, R. O.; Barreto Gomes, D. E.; Pascutti, P. G., Insights into the Interactions of Fasciola Hepatica Cathepsin L3 with a Substrate and Potential Novel Inhibitors through *In Silico* Approaches. *PLoS Negl Trop Dis* **2015**, 9, e0003759.

(131) Durrant, J. D.; Keranen, H.; Wilson, B. A.; McCammon, J. A., Computational Identification of Uncharacterized Cruzain Binding Sites. *PLoS Negl Trop Dis* **2010**, 4, e676.

(132) Brady, C. P.; Brinkworth, R. I.; Dalton, J. P.; Dowd, A. J.; Verity, C. K.; Brindley, P. J., Molecular Modeling and Substrate Specificity of Discrete Cruzipain-Like and Cathepsin L-Like Cysteine Proteinases of the Human Blood Fluke Schistosoma Mansoni. *Arch Biochem Biophys* **2000**, 380, 46-55.

(133) Lecaille, F.; Choe, Y.; Brandt, W.; Li, Z.; Craik, C. S.; Bromme, D., Selective Inhibition of the Collagenolytic Activity of Human Cathepsin K by Altering Its S2 Subsite Specificity. *Biochemistry* **2002**, 41, 8447-8454.

(134) Lecaille, F.; Chowdhury, S.; Purisima, E.; Bromme, D.; Lalmanach, G., The S2 Subsites of Cathepsins K and L and Their Contribution to Collagen Degradation. *Protein Sci* **2007**, 16, 662-670.

(135) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T., Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J Phys Chem B* **2013**, 117, 8408-8421.

(136) Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin Drug Discov* **2015**, 10, 449-461.

(137) Copeland, R. A. Reversible Inhibitors. In *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*; Wiley-VCH, Inc.: New York, 2000; Chapter 8, pp 266-304.

(138) Waller, R. F.; Keeling, P. J.; Donald, R. G.; Striepen, B.; Handman, E.; Lang-Unnasch, N.; Cowman, A. F.; Besra, G. S.; Roos, D. S.; McFadden, G. I., Nuclear-Encoded Proteins Target to the Plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **1998**, 95, 12352-12357.

(139) Nicoll-Griffith, D. A., Use of Cysteine-Reactive Small Molecules in Drug Discovery for Trypanosomal Disease. *Expert Opin Drug Discov* **2012**, 7, 353-366.

(140) Verespy, S., 3rd; Mehta, A. Y.; Afosah, D.; Al-Horani, R. A.; Desai, U. R., Allosteric Partial Inhibition of Monomeric Proteases. Sulfated Coumarins Induce Regulation, Not Just Inhibition, of Thrombin. *Sci Rep* **2016**, 6, 24043.

(141) Ghisaidoobe, A. B.; Chung, S. J., Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Forster Resonance Energy Transfer Techniques. *Int J Mol Sci* **2014**, 15, 22518-22538.

(142) Lang, E. J.; Heyes, L. C.; Jameson, G. B.; Parker, E. J., Calculated pKa Variations Expose Dynamic Allosteric Communication Networks. *J Am Chem Soc* **2016**, 138, 2036-2045.

(143) Bowman, G. R.; Geissler, P. L., Equilibrium Fluctuations of a Single Folded Protein Reveal a Multitude of Potential Cryptic Allosteric Sites. *Proc Natl Acad Sci U S A* **2012**, 109, 11681-11686.

(144) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The Ensemble Nature of Allostery. *Nature* **2014**, 508, 331-339.

(145) Tsai, C. J.; del Sol, A.; Nussinov, R., Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *J Mol Biol* **2008**, 378, 1-11.

(146) Cooper, A.; Dryden, D. T., Allostery without Conformational Change. A Plausible Model. *Eur Biophys J* **1984**, 11, 103-109.

(147) Nussinov, R.; Tsai, C. J., Allostery without a Conformational Change? Revisiting the Paradigm. *Curr Opin Struct Biol* **2015**, 30, 17-24.

(148) Ma, W.; Tang, C.; Lai, L., Specificity of Trypsin and Chymotrypsin: Loop-Motion-Controlled Dynamic Correlation as a Determinant. *Biophys J* **2005**, 89, 1183-1193.

(149) Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D., Intrinsic Dynamics of an Enzyme Underlies Catalysis. *Nature* **2005**, 438, 117-121.

(150) Storer, A. C.; Menard, R., Catalytic Mechanism in Papain Family of Cysteine Peptidases. *Methods Enzymol* **1994**, 244, 486-500.

(151) Otto, H. H.; Schirmeister, T., Cysteine Proteases and Their Inhibitors. *Chem Rev* **1997**, 97, 133-172.

(152) Shokhen, M.; Khazanov, N.; Albeck, A., Challenging a Paradigm: Theoretical Calculations of the Protonation State of the Cys25-His159 Catalytic Diad in Free Papain. *Proteins* **2009**, 77, 916-926.

(153) Zhai, X.; Meek, T. D., Catalytic Mechanism of Cruzain from *Trypanosoma cruzi* as Determined from Solvent Kinetic Isotope Effects of Steady-State and Pre-Steady-State Kinetics. *Biochemistry* **2018**, 57, 3176-3190.

(154) Haynes, W. M., *Crc Handbook of Chemistry and Physics*. CRC press: 2014.

(155) Herschlag, D.; Pinney, M. M., Hydrogen Bonds: Simple after All? *Biochemistry* **2018**, 57, 3338-3352.

(156) Jencks, W. P., On the Attribution and Additivity of Binding Energies. *Proc Natl Acad Sci U S A* **1981**, 78, 4046-4050.

(157) Ward, W. H.; Holdgate, G. A., Isothermal Titration Calorimetry in Drug Discovery. *Prog Med Chem* **2001**, 38, 309-376.

(158) Song, L. F.; Lee, T. S.; Zhu, C.; York, D. M.; Merz, K. M., Jr., Using Amber18 for Relative Free Energy Calculations. *J Chem Inf Model* **2019**, 59, 3128-3135.

(159) Xu, H.; Stern, H. A.; Berne, B. J., Can Water Polarizability Be Ignored in Hydrogen Bond Kinetics? *J Phys Chem B* **2002**, 106, 2054-2060.

(160) Lu, J.; Hou, X.; Wang, C.; Zhang, Y., Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J Chem Inf Model* **2019**, 59, 4540-4549.

(161) Huang, N.; Shoichet, B. K., Exploiting Ordered Waters in Molecular Docking. *J Med Chem* **2008**, 51, 4862-4865.

(162) Amaro, R. E.; Li, W. W., Emerging Methods for Ensemble-Based Virtual Screening. *Curr Top Med Chem* **2010**, 10, 3-13.

(163) Garcia-Sosa, A. T., Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies. *J Chem Inf Model* **2013**, 53, 1388-1405.

(164) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., 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.

(165) Cournia, Z.; Allen, B.; Sherman, W., Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J Chem Inf Model* **2017**, 57, 2911-2937.

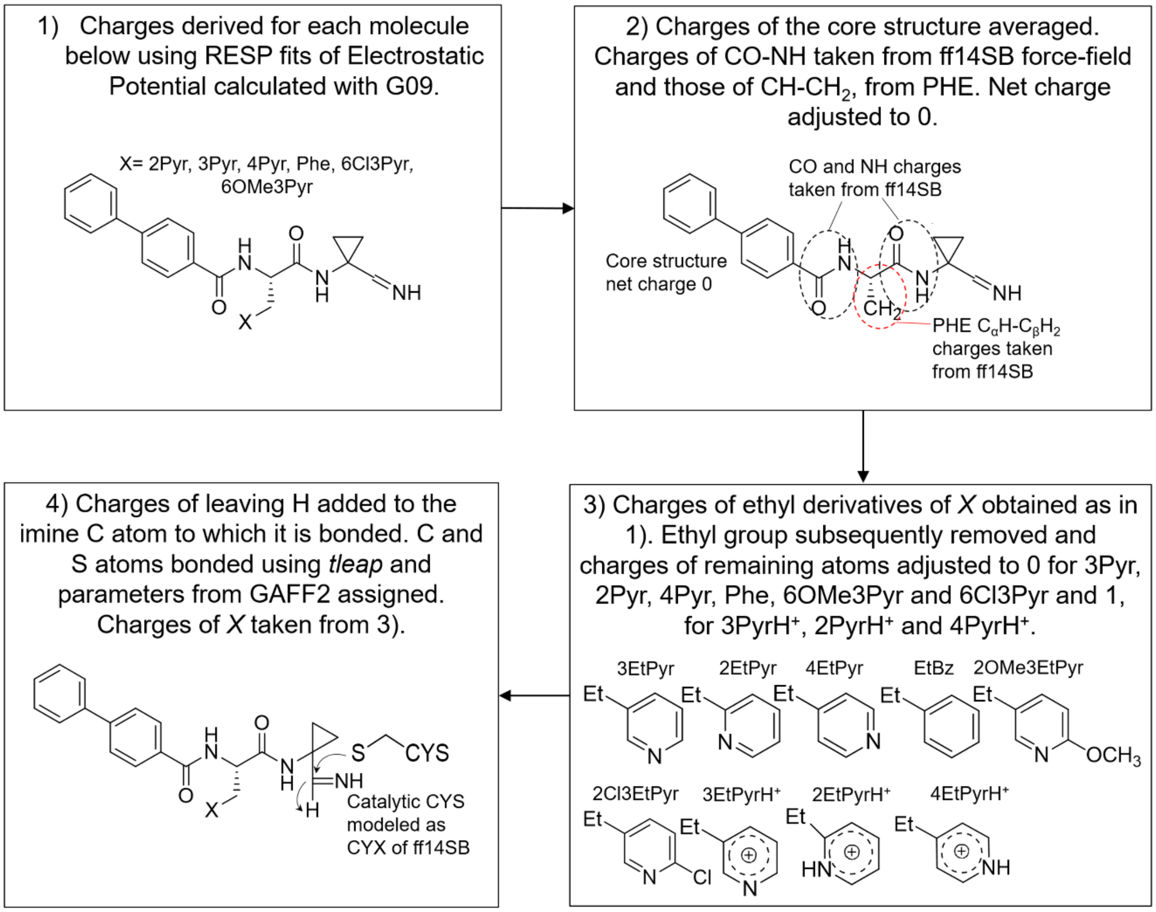
(166) Shah, F.; Gut, J.; Legac, J.; Shivakumar, D.; Sherman, W.; Rosenthal, P. J.; Avery, M. A., Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure-Activity Relationships, Hydration Site Thermodynamics, and Reactivity Analysis. *J Chem Inf Model* **2012**, 52, 696-710.

APPENDIX A – Additional information to methods

Text A1. Compound parametrization

The charges of the atoms belonging to the core structure, i.e., the common structural region present in all the studied compounds, were obtained by subjecting each inhibitor in neutral state to quantum mechanical calculations (see step 1 in Fig. A1). These charges were then averaged and those of the atoms involved in peptide bonds (-CO-NH-) were manually adjusted to their values in ff14SB force-field (see step 2 in Fig. A1).[1] Similarly, the charges of the aliphatic group >CH-CH2-*ϕ*, where *ϕ* is the P2 substituent, were replaced by those of the equivalent atoms in PHE (see step 2 in Fig. A1). The core net charge was neutralized by uniformly distributing the oppositely-signed excess charge in all of the atoms contained in this region, except for those whose charges were taken directly from ff14SB force-field (-CONH- and >CH-CH2-*ϕ*) (see step 2 in Fig. A1).

Figure A1. Workflow employed to parametrize the studied covalent inhibitors. Four main steps described in detail in Text A1 were conducted to derive charges, bonded and nonbonded parameters for the studied molecules.



1) Charges derived for each molecule below using RESP fits of Electrostatic Potential calculated with G09.

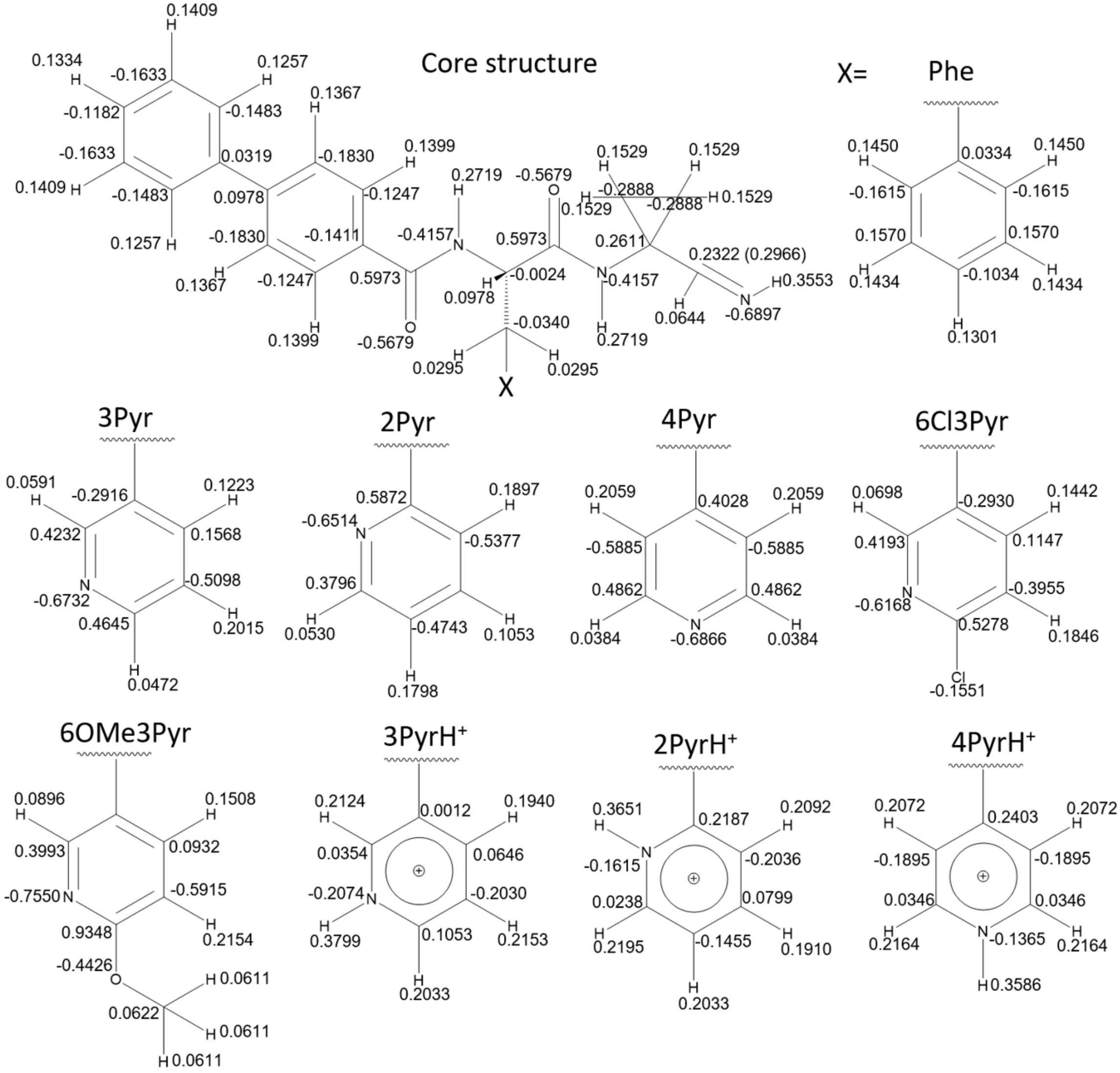
2) Charges of the core structure averaged. Charges of CO-NH taken from ff14SB force-field and those of CH-CH_2, from PHE. Net charge adjusted to 0.

3) Charges of ethyl derivatives of *X* obtained as in 1). Ethyl group subsequently removed and charges of remaining atoms adjusted to 0 for 3Pyr, 2Pyr, 4Pyr, Phe, 6OMe3Pyr and 6Cl3Pyr and 1, for 3PyrH$^+$, 2PyrH$^+$ and 4PyrH$^+$.

4) Charges of leaving H added to the imine C atom to which it is bonded. C and S atoms bonded using *tleap* and parameters from GAFF2 assigned. Charges of *X* taken from 3).

Source: Prepared by the author.

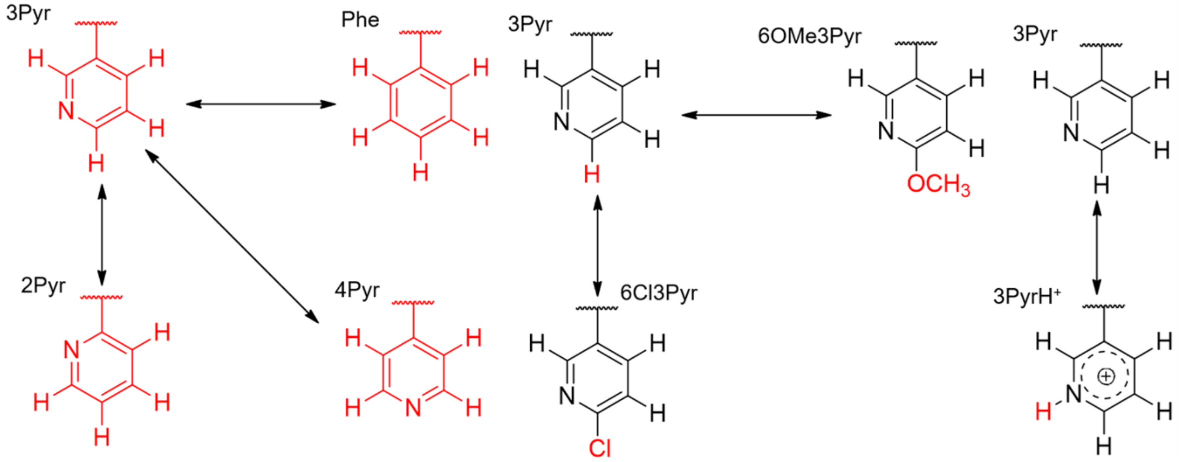
Figure A2. Partial charges assigned to the atoms of the studied compounds. The charge of the imine C atom in parentheses is the one used to simulate the covalent complexes after the removal of the H atom.



Source: Prepared by the author.

On the other hand, the P2-aromatic rings bearing ethyl groups in lieu of the core structure were subjected to independent quantum mechanical calculations and subsequent RESP fits to derive partial atomic charges with *antechamber* of Amber18.[1] Then, the ethyl groups were removed and the total charge of the P2 substituent was adjusted to either 0 or 1, depending on the protonation state, through uniform distribution of the oppositely-signed excess charge in the remaining atoms (see step 3 in Fig. A1). This approach was found to yield more suitable and consistent charges for the atoms undergoing alchemical transformations than those obtained from the parametrization of the whole compounds. Of note, it is essential that the atoms subjected to alchemical transformation have zero net charge in order to keep the charge constant during the TI free energy calculations. This condition is satisfied by the above-mentioned parametrization scheme. When dealing with charged ligands, such as the protonated forms of the studied compounds, the conservation of the net charge is guaranteed by another strategy, as explained in Section 2.1.9.[2]

Figure A3. Atoms defining the TI regions during the alchemical transformations. Atoms depicted in red were treated with soft-core potentials as they appear/disappear during the van der Waals transformations, the remaining atoms were linearly transformed. When performing the uncharging and recharging steps, all atoms shown in the figure were linearly transformed. Arrows indicate that alchemical transformations were performed in both directions.



Source: Prepared by the author.

Covalent adducts formed by each compound and the catalytic Cys were created by removing the H bonded to the imine carbon and then adding its charge to the latter atom in order to keep the net charge constant (see step 4 in Fig. A1). Instead of parametrizing the whole adducts, we took the parameters of the covalently-modified Cys, including its partial charges, directly from the CYX residue type of ff14SB force-field, which greatly simplified the generation of covalent complexes. We are assuming here that the impact of the neglected charge variations on the overall protein-ligand dynamics and on the relative free energy calculations is slight, especially when the affected atoms are distal to the mutated groups. All bonded and nonbonded parameters describing the interaction of the imine carbon and the Cys S atom were set to those of the c2 and ss atom types of gaff2, respectively.[1]

Finally, it is worth noting that the charges of the PHE phenyl ring were not directly taken from ff14SB force-field, as we deemed that it was more consistent to derive the charges of all atoms undergoing alchemical transformations by following the same exact approach. By doing so, one can guarantee higher error cancelation when calculating the $\Delta\Delta G$ values. Figure A2 shows the partial charges assigned to the atoms of the studied inhibitors.

Text A2. Thermodynamic cycle employed for relative free energy calculations

Complexes involving cysteine proteases and reversible covalent inhibitors form through the following mechanism:[3]

$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \bullet I \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E - I$$

The enzyme (E) and the inhibitor first bind to form a noncovalent complex ($E \bullet I$), which subsequently leads to the formation of the adduct ($E–I$). Note that k_1, k_{-1}, k_2 and k_{-2} are kinetic rate constants corresponding to the indicated processes. The equilibrium constants associated to each reversible process of the mechanism (K_1 and K_2) are defined by the equations below:[3]

$$K_1 = \frac{k_{-1}}{k_1} = \frac{[E][I]}{[E \bullet I]} \tag{A1a}$$

$$K_2 = \frac{k_{-2}}{k_2} = \frac{[E \bullet I]}{[E - I]} \tag{A1b}$$

The inhibition equilibrium constant (K_i) of the overall process calculated from experimental kinetic assays can be written as follows:[3]

$$K_i = \frac{[E][I]}{[E \bullet I] + [E - I]} \tag{A2}$$

By combining all previous equations, we can write the K_i as follows:

$$K_i = \frac{K_1 K_2}{1 + K_2} = \frac{K_3}{1 + K_2} \tag{A3}$$

where K_3 is the dissociation equilibrium constant of the overall process: $E + I \rightleftharpoons E–I$.

If one is interested in calculating the $\Delta\Delta G$ value between two reversible covalent inhibitors I_1 and I_2 targeting the same enzyme E, one can use the expression below that relates the free energies to the K_i's:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = RTln\left[\frac{K_i^{(2)}}{K_i^{(1)}}\right] \tag{A4}$$

where ΔG_1 and ΔG_2 are the binding free energies of the complexes and $K_i^{(1)}$and $K_i^{(2)}$, their respective inhibition constants. By substituting Eq. A3 into A4, one obtains the following expression:

$$\Delta\Delta G = RTln\left[\frac{K_3^{(2)}}{K_3^{(1)}}\right] + RTln\left[\frac{1 + K_2^{(1)}}{1 + K_2^{(2)}}\right] \quad \text{(A5)}$$

The second term on the right side of Eq. A5 can be neglected when *i*) $K_2^{(1)} = K_2^{(2)}$ and/or *ii*) $K_2^{(1)} \ll 1$ and $K_2^{(2)} \ll 1$. Condition *i*) is satisfied when the structural differences between inhibitors I_1 and I_2, belonging to the same compound series and having identical warheads, do not affect the covalent bond formation equilibrium. Condition *ii*) is fulfilled when most of the complexed ligand is covalently bonded to the enzyme (*E–I*), i.e., [*E–I*]≫[*E•I*]. When any of the above conditions is met, Eq. A5 can be approximated to:

$$\Delta\Delta G \approx RTln\left[\frac{K_3^{(2)}}{K_3^{(1)}}\right] \quad \text{(A6)}$$

Eq. A6 indicates that the calculation of ΔΔ*G* values for covalent inhibitors fulfilling conditions *i*) and/or *ii*) can be conducted through a thermodynamic cycle that involves the alchemical transformations of the inhibitor free in solution and covalently bonded to the enzyme (Fig. 2.3), as has been already done by others.[4] Therefore, it is not necessary to take into account the free energy contribution of the alchemical process transforming one ligand into the other when they are bound in a noncovalent fashion to the enzyme (*E•I*).

We have assumed that the inhibitors studied in the current work fulfill, at least, one of the abovementioned conditions. Firstly, the chemical changes take place in a region distal to the cyano group; hence, no significant modifications of the electrophilicity of the warhead and of its accommodation within the enzyme S1 subsite are expected to occur.[4] Besides, the crystal structures of cysteine proteases complexed with similar nitrile inhibitors show that the covalent form is invariantly preferred over the noncovalent one, at least, under the crystallization conditions (e.g., PDBs: 2FQ9, 2FRA, 1MS6, 4AXM, 4QH6, 3I06, 1GMY, 1U9V, 1U9X, 1U9W). This, in turn, suggests that K_2 must be lower than 1. As a consequence of all the previous statements, the thermodynamic cycle shown in Fig. 2.3-A can be employed as a reasonable starting approximation to conduct ΔΔ*G* calculations for the studied systems. The comparison with the experimental values will ultimately show whether these assumptions can yield results consistent with the experiments.

Moreover, we performed the alchemical transformations of the P2 moieties in solution with the compounds bearing an imino instead of a cyano group, even though the latter is the functional group actually present in the free molecules (Fig. 2.3-B). This allowed us to employ the imines already parametrized (see Text A1), which constitute the noncovalent analogs of the bound state, thus eliminating the unnecessary step of parametrizing the nitriles. The previous simplification relies on the assumption that free energies associated to the alchemical transformations of the compounds in solution are not affected by the specific nature of the warhead (either a cyano or an imino group), given the latter's distal position with respect to the P2 moiety and the small chemical change undergone.

Text A3. Calculation of experimental ΔΔ*G* from IC_{50} values and pH corrections

To transform the IC_{50} value of a reversible competitive inhibitor into its K_i, as is the case for the inhibitors studied here, one can employ the expression below, upon the assumption that inhibitors obey the Michaelis-Menten equation:[5]

$$K_i = \frac{IC_{50} - \frac{1}{2}[E]}{1 + \frac{[S]}{K_M}} \tag{A7}$$

where $[E]$ and $[S]$ are the concentrations of the enzyme and the substrate in the assay, and K_M is the substrate concentration required to attain half-maximal velocity. When $IC_{50} \gg [E]$, the previous equation can be simplified to the one below:[6]

$$K_i \approx \frac{IC_{50}}{1 + \frac{[S]}{K_M}} \tag{A8}$$

Assuming that $[S]$ and K_M remain constant when performing inhibition assays for different inhibitors (I_1 and I_2) binding the same enzyme under the same conditions, one can express their relative free energy in terms of the IC_{50} values by substituting Eq. A8 into Eq. A4:

$$\Delta\Delta G \approx RTln\left[\frac{IC_{50}^{(2)}}{IC_{50}^{(1)}}\right] \tag{A9}$$

When the inhibitors undergo ionization processes in solution, some corrections depending on the pH and the pK_a of the ionizable groups must be taken into account.

Let us consider that the covalent inhibitor I coexists as two species IH^+ and I in aqueous solution, as indicated by the equilibrium shown below:

$$IH^+ + H_2O \rightleftharpoons I + H_3O^+$$

The equilibrium constant for this process (K_a) is defined by the following equation:

$$K_a = \frac{[H_3O^+][I]}{[IH^+]} \tag{A10}$$

Due to the underlying protonation/deprotonation equilibrium, an apparent K_i (K_i^{app}) instead of the real one is obtained from the experiments. K_i^{app} can be defined as follows:

$$K_i^{app} = \frac{[E][I]'}{[E-I]+[E \bullet I]} \tag{A11}$$

where $[I]'$ is the total concentration of the inhibitor in solution, regardless of its protonation state, which is equal to the following sum:

$$[I]' = [I] + [IH^+] \tag{A12}$$

Notice that Eq. A11 is valid when the occurrence of $E{\bullet}IH^+$ and $E\text{-}IH^+$ species is negligible; therefore, they can be eliminated from the denominator on the equation's right side. If one expresses $[IH^+]$ in terms of $[I]$ using Eq. A10, inserts the resulting expression into Eq. A12, then, after some algebra, one obtains the equation below:

$$\frac{[I]}{[I]'} = \frac{K_a}{[H_3O^+]+K_a} \tag{A13}$$

From eqs. A2, A8, A11 and A13, one arrives to the following relations:

$$\frac{IC_{50}}{IC_{50}^{app}} \approx \frac{K_i}{K_i^{app}} = \frac{[I]}{[I]'} = \frac{K_a}{[H_3O^+]+K_a} \tag{A14}$$

Therefore, the IC_{50} of an ionizable inhibitor is related to its IC_{50}^{app} by the equation below:

$$IC_{50} = IC_{50}^{app}\frac{K_a}{[H_3O^+]+K_a} = IC_{50}^{app}\frac{10^{-pK_a}}{10^{-pH}+10^{-pK_a}} = IC_{50}^{app}\frac{10^{pH}}{10^{pH}+10^{pK_a}} \tag{A15}$$

The $\Delta\Delta G$ value between two ionizable inhibitors I_1 and I_2 that bind the same enzyme can be calculated by inserting Eq. A15 into A9:

$$\Delta\Delta G \approx RTln\left[\frac{IC_{50}^{(app,2)}}{IC_{50}^{(app,1)}}\right] + RTln\left[\frac{10^{pH}+10^{pK_a^{(1)}}}{10^{pH}+10^{pK_a^{(2)}}}\right] \tag{A16}$$

The second term on the right side of the above equation represents the pH correction due to the occurrence of ionization. If only one of the two inhibitors, e.g., I_2, is susceptible to ionization, eq. A16 can be simplified as follows:

$$\Delta\Delta G \approx RTln\left[\frac{IC_{50}^{(app,2)}}{IC_{50}^{(1)}}\right] + RTln\left[\frac{10^{pH}}{10^{pH}+10^{pK_a^{(2)}}}\right] \tag{A17}$$

The last two equations were employed to calculate $\Delta\Delta G$ values for the studied inhibitors based on the IC_{50} and IC_{50}^{app} values reported by Nizi *et al.*[7] We assumed that the pH of the assays was 5.5, as it is customary when working with lysosomal cysteine proteases,[8, 9] because the actual value was not reported by the authors. pK_a's for the P2-Pyr substituents were approximated to experimental values reported in literature for Pyr derivatives (see Appendix D, Text D1). As the concentration of enzyme in the assays was not reported either, we had to use the simplified expression A8, which neglects the $[E]$ term. Of note, Eq. A8 is more likely to yield somewhat inaccurate results for tight-binding inhibitors, since the condition $IC_{50} \gg [E]$ is typically not fulfilled. However, in the absence of more experimental details, we believe that eqs. A9, A16 and A17 are good approximations that allow us to compare the predicted $\Delta\Delta G$ values for the studied systems with those estimated from experimental measurements within a certain margin of error.

Text A4. Umbrella sampling protocol

After careful orientation using gmx editconf of GROMACS v5.1.4, the FP-2:Cpd66 and FP-2:peptide:Cpd66 complexes were embedded in 7x7x10 nm boxes. Of note, identical orientation of both complexes inside the box was guaranteed by structural alignment prior to solvation with TIP3P water molecules. A neutralizing amount of NaCl was added to each, so that the final concentration was 0.1 M. The systems were then subjected to 50000 steps of steepest descents energy minimization, followed by 100 ps of NVT and NPT equilibrations in order to reach a final temperature and pressure of 298 K and 1 bar, respectively. The heating process during the NVT equilibration was conducted using a linear temperature gradient, from 50 to 298 K, the target temperature being achieved at 90 ps. The Berendsen thermostat and barostat were used for temperature and pressure control, respectively, during the equilibrations.[10] Pulling simulations were carried out during 25 ns at a pulling rate of 0.192 nm/ns on the *z* axis. The reference atoms for pulling and US simulations were the carbonyl carbon atom of residue I146 and the alkenyl carbon atom of Cpd66 bonded to the furanyl ring (Fig. 1.6). Harmonic restraints on the orthogonal axes of the distance vector between the reference atoms were set during all pulling and US simulations. The orthogonal restraint constant was 8000 $kJ \cdot mol^{-1} \cdot Å^{-2}$ during pulling simulations, and was lowered to 4000 $kJ \cdot mol^{-1} \cdot Å^{-2}$ during US simulations. The equilibrium values for the *xy* distance components were calculated by averaging the results from five 25 ns replicate MD simulations of the complexes carried out with no restraints, after discarding the first 5 ns of each.

In addition, the protein's six external degrees of freedom were restrained to avoid rotation and translation with respect to the center of mass without affecting the conformational equilibrium in both, pulling and US simulations.[11] This was achieved by attaching the protein to three dummy atoms, which were kept fixed in the simulation cell by means of position restraints of 21000 kJ·mol^{-1}·Å^{-2}. The six restraints involving the three dummy atoms (D1, D2 and D3) and three protein atoms (carbonyl carbons of I57, S147 and A88) corresponded to D1-I57 distance, D2-D1-I57 and D1-I57-S147 angles, and D3-D2-D1-I57, D2-D1-I57-S147 and D1-I57-S147-A88 dihedrals. A harmonic constant of 21000 kJ·mol^{-1}·Å^{-2} was employed for the D1-I57 distance restraint, and 418.4 kJ·mol^{-1}rad^{-2}, for the angular and dihedral restraints. Frames were collected from the pulling simulations every 0.1 nm intervals along the *z* axis until reaching a distance of 3 nm between the ligand and the protein reference atoms. Beyond that distance, frames were collected every 0.2 nm. In the 0-1.20 nm interval, a restraint constant of 6000 kJ·mol^{-1}·Å^{-2} was set for every umbrella window, whereas a value of 1000 kJ·mol^{-1}·Å^{-2} was employed for the remaining ones. The stiffer constant allowed us to keep the ligand reference atom close to the equilibrium position while still partially buried in the binding pocket. The starting frame of each window was equilibrated in the NPT ensemble during 200 ps at temperature 298 K and pressure 1 bar. The production runs were conducted in the NPT ensemble at 298 K using the velocity rescaling and Parrinello-Rahman algorithms for temperature and pressure control, respectively.[12, 13] Up to 1.20 nm, all windows were simulated during 25 ns during the production runs, the remaining ones being simulated over 15 ns. Data collection started after discarding the first 5 ns. The *gmx wham* program of GROMACS v5.1.4 was then used to calculate the PMF along the reaction coordinate.[14] Standard errors of the mean were estimated by conducting 1000 bootstrapping simulations taking into account the autocorrelation time for each window.

In all simulations, a distance cut-off of 1.0 nm was set for electrostatic and van der Waals non-bonded interactions. Particle Mesh Ewald (PME) was used to handle long-range electrostatics. PBCs were employed in all cases. The equation of motion was solved using the leap frog algorithm with a Δ*t* of 2 fs,[15] while constraining the bond lengths with LINCS.[16]

References

(1) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand,

S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A., *AMBER2018*; University of California, San Francisco, 2018.

(2) Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H., Calculation of Binding Free Energies. *Methods Mol Biol* **2015**, 1215, 173-209.

(3) Chatterjee, P.; Botello-Smith, W. M.; Zhang, H.; Qian, L.; Alsamarah, A.; Kent, D.; Lacroix, J. J.; Baudry, M.; Luo, Y., Can Relative Binding Free Energy Predict Selectivity of Reversible Covalent Inhibitors? *J Am Chem Soc* **2017**, 139, 17945-17952.

(4) Kuhn, B.; Tichy, M.; Wang, L.; Robinson, S.; Martin, R. E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; Diederich, F.; Hert, J., Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *J Med Chem* **2017**, 60, 2485-2497.

(5) Copeland, R. A., Tight binding inhibitors. In *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. Wiley-VCH, Inc.:New York, 2000; Chapter 9, pp 305-317.

(6) Copeland, R. A., Reversible Inhibitors. In *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. Wiley-VCH, Inc.: New York, 2000; Chapter 8, pp 266-304.

(7) Nizi, E.; Sferrazza, A.; Fabbrini, D.; Nardi, V.; Andreini, M.; Graziani, R.; Gennari, N.; Bresciani, A.; Paonessa, G.; Harper, S., Peptidomimetic Nitrile Inhibitors of Malarial Protease Falcipain-2 with High Selectivity against Human Cathepsins. *Bioorg Med Chem Lett* **2018**, 28, 1540-1544.

(8) Novinec, M.; Korenc, M.; Caflisch, A.; Ranganathan, R.; Lenarcic, B.; Baici, A., A Novel Allosteric Mechanism in the Cysteine Peptidase Cathepsin K Discovered by Computational Methods. *Nat Commun* **2014**, 5, 3287.

(8) Hernandez-Gonzalez, J. E.; Salas-Sarduy, E.; Hernandez Ramirez, L. F.; Pascual, M. J.; Alvarez, D. E.; Pabon, A.; Leite, V. B. P.; Pascutti, P. G.; Valiente, P. A., Identification of (4-(9H-Fluoren-9-yl) Piperazin-1-yl) Methanone Derivatives as Falcipain 2 inhibitors Active against *Plasmodium falciparum* Cultures. *Biochim Biophys Acta Gen Subj* **2018**, 1862, 2911-2923.

(10) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, 81, 3684-3690.

(11) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K., Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *J Chem Theory Comput* **2017**, 13, 3260-3275.

(12) Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling through Velocity Rescaling. *J Chem Phys* **2007**, 126, 014101.

(13) Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1981**, 52, 7182-7190.

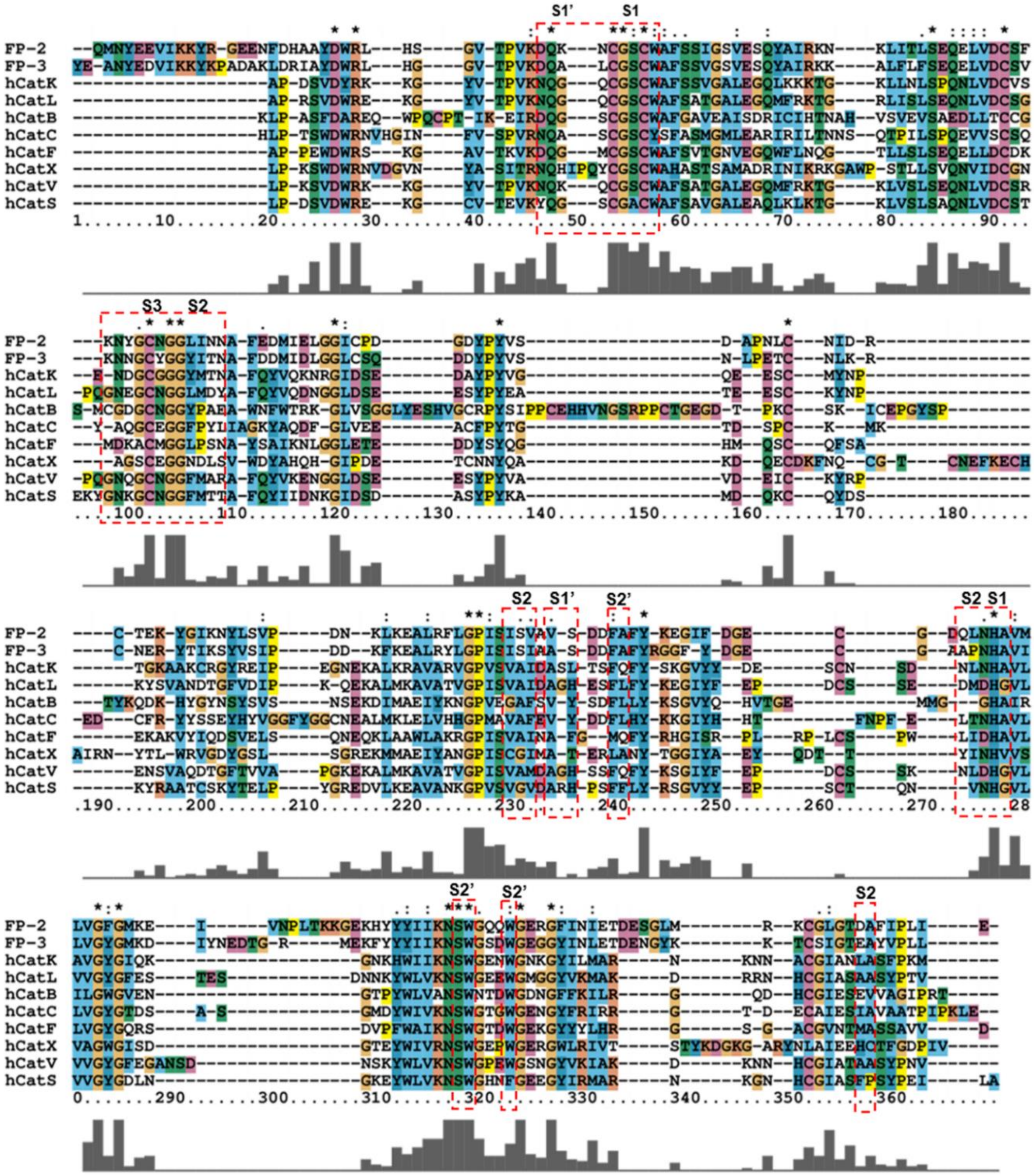
(14) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J Chem Theory Comput* **2010**, 6, 3713-3720.

(15) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *Softwarex* **2015**, 1, 19-25.

(16) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* **1997**, 18, 1463-1472.

APPENDIX B – Additional information to identification of FP-2 inhibitors through an integrated *in silico* and experimental approach

Figure B1. Structural alignment of FPs and hCats, shown at the level of their primary sequences. Red dashed rectangles enclose active site residues of the aligned proteases. On top of the rectangles, the subsites to which the enclosed residues belong are indicated. Asteriks (*) indicate conserved positions, while colons (:) and periods (.) indicate conservation between groups of amino acids with strongly and weakly similar properties, respectively. Grey bars show the %ID for each position. The ClustalX2 default coloring scheme is used.[1] The PDBs used for alignment with Modeller 9v9[2] were: 2OUL, 3BWK, 1AYV, 5MAE, 1GMY, 1M6D, 1EF7, 1FH0 and 3KWN.



Source: Prepared by the author.

Table B1: %IDs for hCats with respect to FP-2 and FP-3 in the active site region

	hCatK	**hCatL**	**hCatB**	**hCatC**	**hCatF**	**hCatX**	**hCatV**	**hCatS**
FP-2	53.85%	48.72%	48.72%	46.15%	46.15%	35.71%	48.72%	43.59%
FP-3	61.54%	48.72%	46.15%	46.15%	48.72%	40.48%	46.15%	43.59%

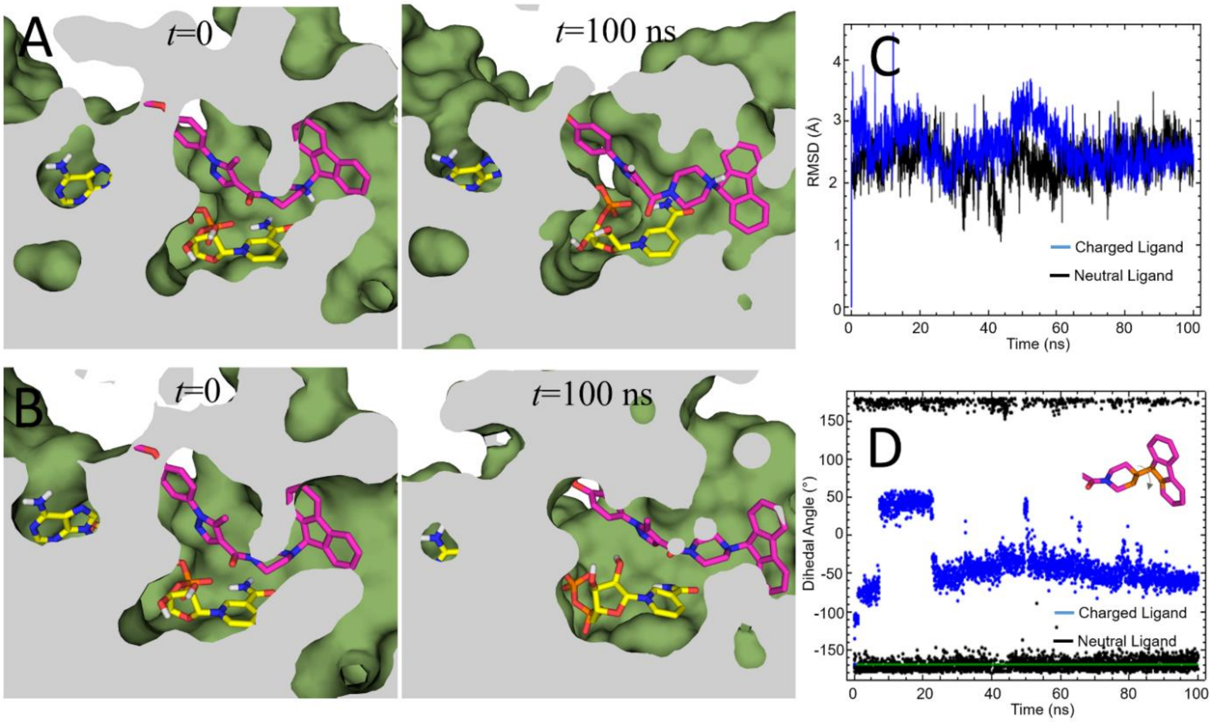
The %IDs were calculated based on the alignment shown in Fig. B1, and only considering the residues enclosed by the dashed red rectangles.
Source: Prepared by the author.

Text B1. *In silico* analysis of the interaction of HTS07940 with PfENR

We built a model of the ternary complex PfENR:HTS07940:NAD$^+$ by superimposing the PfNER molecule (PDB: 2OL4) onto InhA (PDB: 1P44). The coordinates of Genz10850 present in the latter structure were used as a starting point to build the structure of HTS07940 with *Avogadro*[3] by replacing the indol group by the methylpyrazolyl and methoxyphenyl rings characteristic of the latter compound. The coordinates of NAD$^+$ were taken directly from PDB 2OL4 after superposition. NAD$^+$ was parametrized following the same general procedures outlined in Section 2.1.4. Differently from FP-2, which acts at pH=5.5, PfENR is an enzyme that works at neutral pH.[4] Of note, a previous study has measured pK_a values ranging from 8.06 to 9.73 for various piperazine derivates, at room and physiological temperatures (298 and 313 K, respectively).[5] Therefore, the chances of HTS07940 to become deprotonated in solution increase, as the pH gets closer to the pK_a vaule of the amine group. Additionally, the hydrophobic environment of the deeply buried PfENR cavity can influence the actual pK_a value of the amine group, favoring the deprotonation process.[6] Taking into account these facts, we parametrized a deprotonated form of HTS07940 and performed subsequent 100 ns MD simulations of the ternary complex with both variants of the compound (Fig. B2).

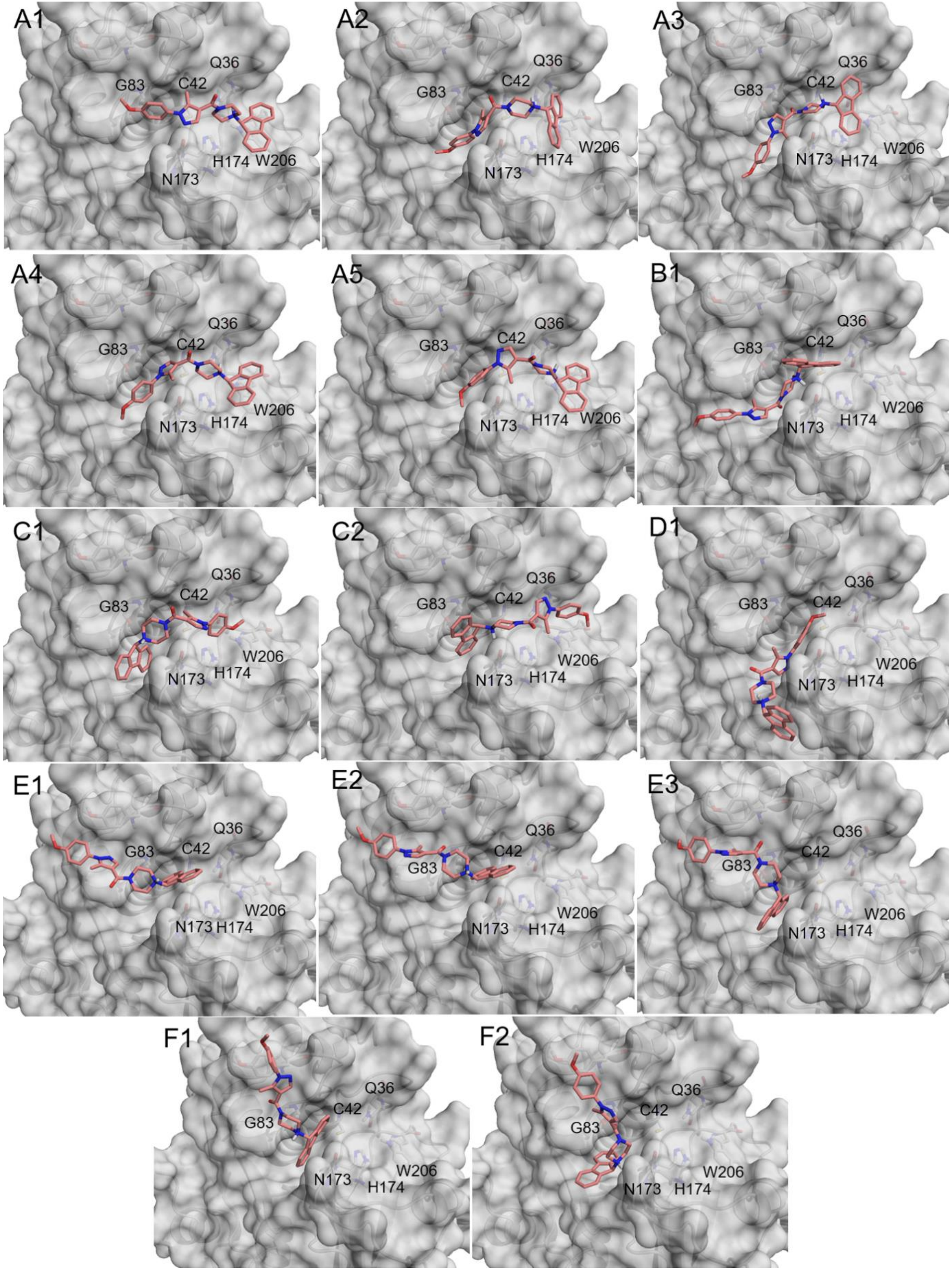
The RMSD analysis for the ligand atoms during both MD simulations shows that the compound remained bound into the protein cavity in two stable binding modes, depending on the protonation state of its amine group (Figs. B2A, B2B and B2C). However, in both cases, the compound deviates ~2.5 Å from the starting structure; thus indicating the occurrence of differences between the putative binding mode of HTS07940 to PfENR and that of the analog Genz10850 to InhA, observed in the crystal structure. Despite the deviation of the binding modes of both HTS07940 protonation states from the initial structure, the relative orientation of the fluorenyl substituent with respect to the piperazine ring of HTS07940 is closer to that observed for Genz10850 in the PDB 1P44 only during the MD simulation of neutral form (Fig. B2D). Hence, the correct protonation state for HTS07940 in a putative interaction with PfENR at physiological pH could be the neutral one. Overall, the relative stability of HTS07940 (and, by inference, of HTS08262) into the PfENR cavity suggests that the compound may be capable of inhibiting this enzyme in addition to FP-2. However, the starting pose of HTS07940 is so deeply buried into the PfENR cavity, that the likelihood of observing a dissociation event of the compound during a 100 ns MD simulation is rather small. Interestingly, the PfENR active pocket changed its shape appreciably during the MD simulation in order to accommodate HTS07940 (Fig. B2). Hence, the enzyme pocket seems to be quite flexible, being able to adapt to ligands of different sizes and shapes, e.g., Genz10850 and larger compounds such as HTS07940 and HTS08262. Further *in vitro* experiments are required to corroborate our predictions about a putative interaction of HTS07940 and HTS08262 with PfENR.

Figure B2. MD simulations of two different protonation states of HTS07940 in complex with PfENR bound to NAD$^+$. A) Starting and final structures of the MD simulation of PfENR:NAD$^+$ in complex with protonated HTS07940. **B)** Starting and final structures of the MD simulation of PfENR:NAD$^+$ in complex with neutral HTS07940. **C)** RMSD values for HTS07940 heavy atoms calculated with respect to the structure at t=0. The sudden initial jump of RMSD values is a consequence of the exclusion of the equilibration time from the graph. **D)** Time profile of the dihedral angle indicated in the figure. The green line corresponds to the dihedral angle in the starting pose of HTS07940.



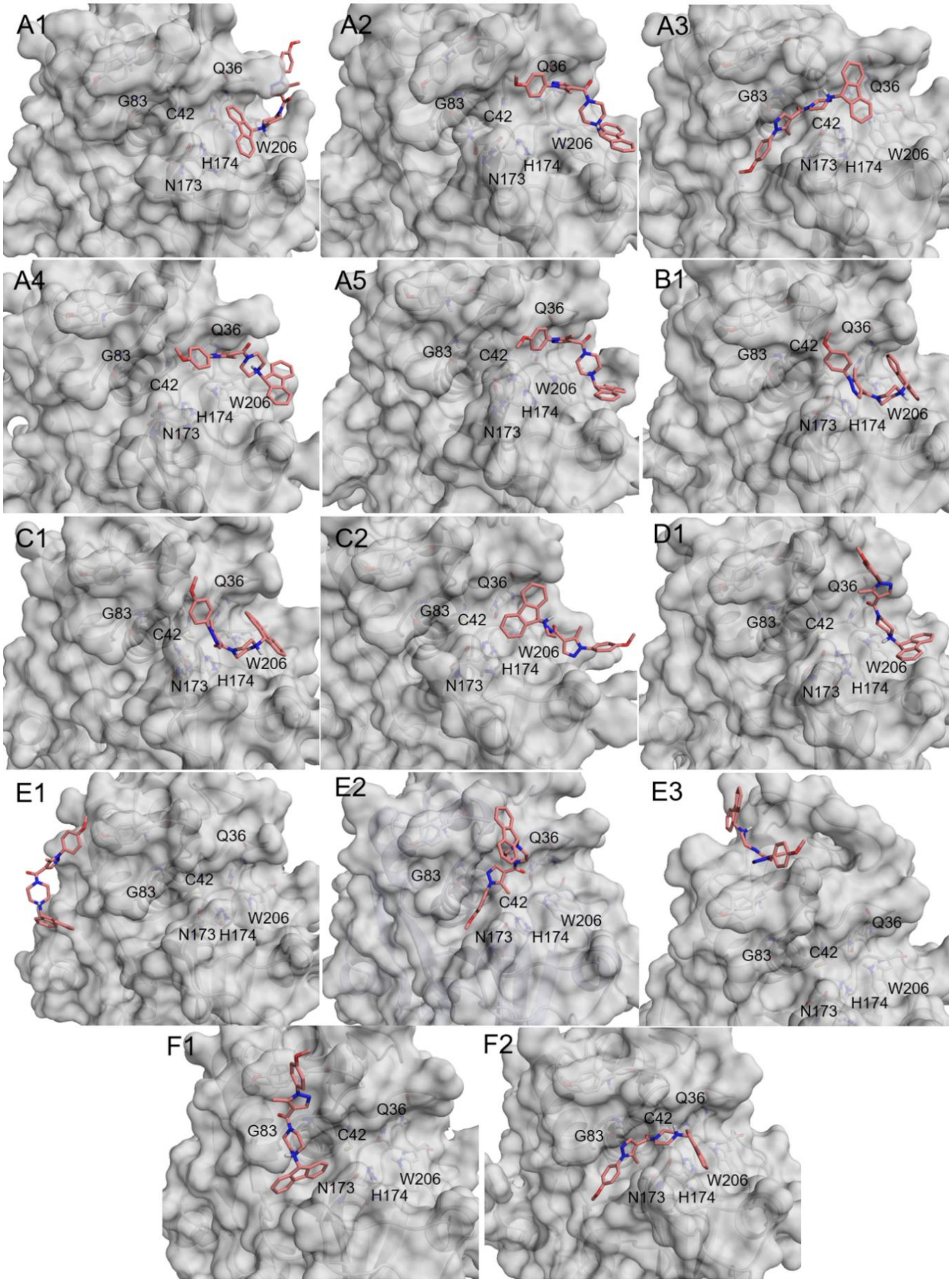
Source: Prepared by the author.

Figure B3. Selected docking poses of compound HTS07940 into the FP-2 binding site. Poses were clustered into groups identified by letters based on their structural similarities.



Source: Prepared by the author.

Figure B4. Final conformations after 100 ns MD simulations of the docking poses of compound HTS07940 into the FP-2 binding site. Final conformations are labeled according to Fig. B3.



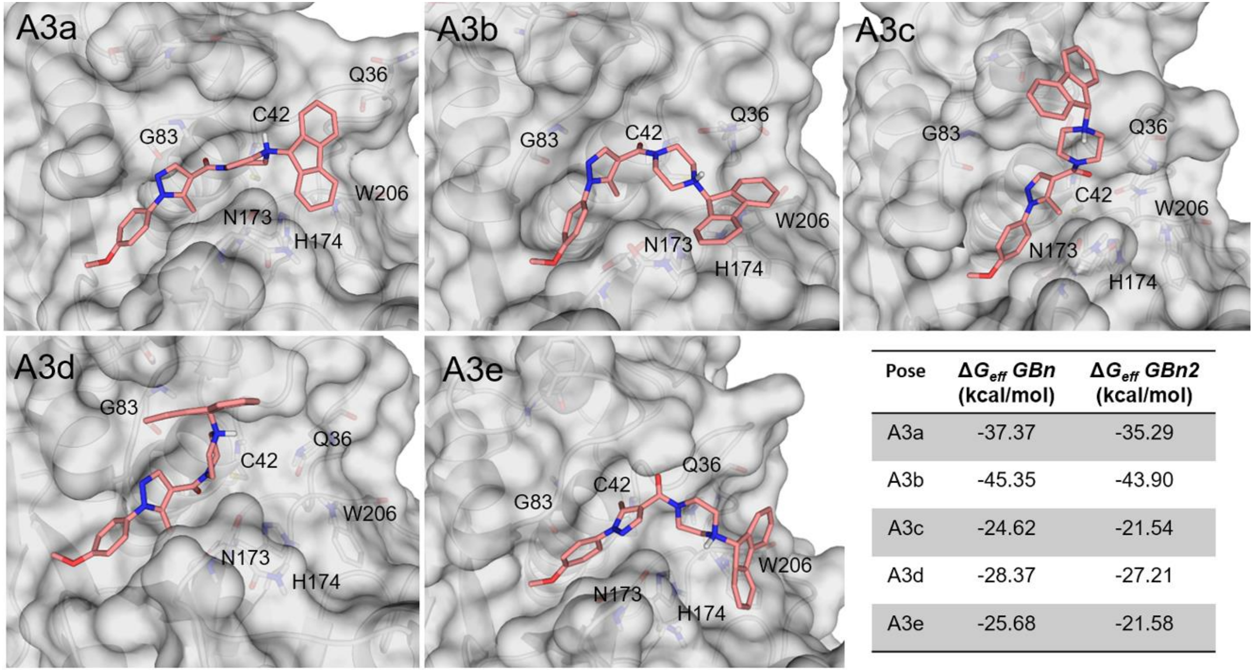
Source: Prepared by the author.

Table B2. Δ*G*$_{eff}$-values for the last 3 ns of the MD simulations of the docking poses of compound HTS07940 in complex with FP-2

Poses	Δ*G*$_{eff}$-GB*n* (kcal/mol)	Δ*G*$_{eff}$-GB*n2* (kcal/mol)
A1	-21.4(0.2)	-20.8(0.2)
A2	-23.5(0.3)	-22.1(0.2)
A3	-28.7(0.2)	-28.6(0.3)
A4	-18.3(0.3)	-17.9(0.3)
A5	-20.6(0.3)	-19.1(0.3)
B1	-18.0(0.3)	-17.7(0.3)
C1	-19.7(0.2)	-18.7(0.2)
C2	-20.5(0.2)	-19.3(0.2)
D1	-22.1(0.3)	-20.6(0.3)
E1	-11.8(0.2)	-11.4(0.2)
E2	-23.5(0.2)	-21.8(0.2)
E3	-22.2(0.2)	-19.7(0.2)
F1	-23.5(0.2)	-22.5(0.2)
F2	-23.3(0.2)	-21.7(0.2)

Source: Prepared by the author.

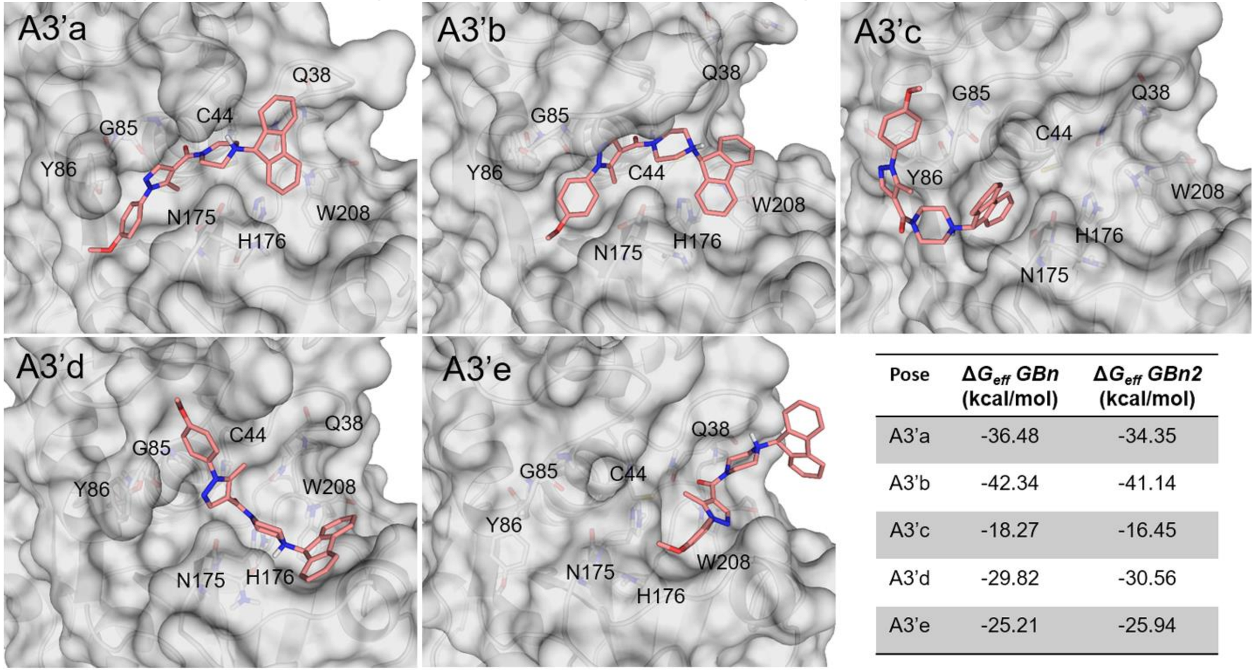
Figure B5. Central structures obtained from clustering analysis of multiple aMD simulations of HTS07940 in complex with FP-2 and their respective Δ*G*$_{eff}$ values. Eleven 100ns aMD simulations were performed for different conformations of the complex generated during a previous MD simulation started with pose A3. Five clusters were obtained for each trajectory, and the resulting 55 single structures were scored with GB*n* and GB*n2* models. For brevity's sake, the figure only shows the five most divergent structures among the 55 central structures, containing the top-scoring ones.



Pose	Δ*G*$_{eff}$ *GBn* (kcal/mol)	Δ*G*$_{eff}$ *GBn2* (kcal/mol)
A3a	-37.37	-35.29
A3b	-45.35	-43.90
A3c	-24.62	-21.54
A3d	-28.37	-27.21
A3e	-25.68	-21.58

Source: Prepared by the author.

Figure B6. Central structures obtained from clustering analysis of multiple aMD simulations of HTS07940 in complex with FP-3 and their respective Δ*G*_{*eff*} values. Eleven 100ns aMD simulations were performed for different conformations of the complex generated during a previous MD simulation started with pose A3’ (similar to pose A3’a). Five clusters were obtained for each trajectory, and the resulting 55 single structures were scored with GB*n* and GB*n2* models. For brevity’s sake, the figure only shows the five most divergent structures among the 55 central structures, containing the top-scoring ones.



Pose	ΔG_{eff} GB*n* (kcal/mol)	ΔG_{eff} GB*n2* (kcal/mol)
A3’a	-36.48	-34.35
A3’b	-42.34	-41.14
A3’c	-18.27	-16.45
A3’d	-29.82	-30.56
A3’e	-25.21	-25.94

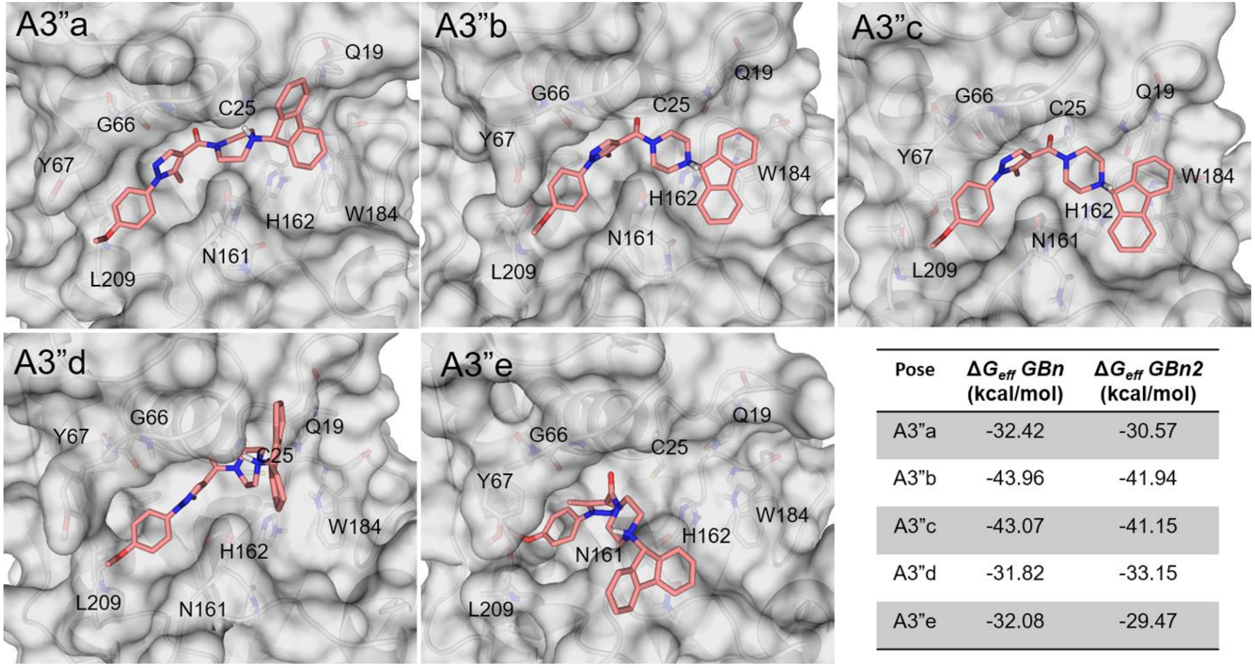
Source: Prepared by the author.

Text B3. Prediction of the binding mode of HTS07940 to hCatK

Following the same strategy employed for FP-3 (see Section 3.1.4), we also predicted the lowest-energy binding mode of compound HTS07940 to hCatK (Fig. B7). In this case, two nearly-isoenergetic top-scoring poses, mainly differing in the conformation of residue Y67, were identified (Fig. B7). The structural features of the central structures of both poses calculated from 100 ns MD simulations are presented in more details in Figs. 3.4 and B8-A. From these figures, it becomes apparent that Y67 side chain protrudes into the solvent in pose A3”b, thereby leading to an ‘up’ conformation. Conversely, the side chain bends towards the protein in pose A3”c, hereinafter referred to as ‘down’ conformation. We also predicted that pose A3”b is, by a narrow margin, more favorable than pose A3”c, based on mean Δ*G*_{*eff*} values calculated from 100 ns MD simulations of both complexes (Table B3). Therefore, further analyses were necessary to definitely propose the lowest energy binding mode of HTS07940 to hCatK.

The ‘down’ conformation of Y67 in pose A3”c is similar to that observed in the crystal structure of human pro-cathepsin K (Figs. B8-B and C). By visual inspection of the latter structure, we concluded that Y67 is pushed down by the propeptide residue P-18 to avoid steric clashes. Consequently, once the propeptide is removed, the Y67 side chain adopts an ‘up’ conformation in the mature form of the enzyme (Fig. B9-A), as well as in all the structures of hCatK bound to organic ligands reported at PDB (data not shown). Of note, the ‘down’ conformation of Y67 was sampled during the MD simulations of free hCatK (Fig. B8-D), but the probability associated with this event was very low, i.e., $3{\cdot}10^{-4}$. This result reinforces the intrinsic instability of the ‘down’ conformation of Y67 in the mature form of free hCatK, as previously inferred from the analysis of the available crystal structures.

Figure B7. Central structures obtained from clustering analysis of multiple aMD simulations of HTS07940 in complex with hCatK and their respective Δ*G_{eff}* values. Eleven 100ns aMD simulations were performed for different conformations of the complex generated during a previous MD simulation started with pose A3” (similar to pose A3”a) Five clusters were obtained for each trajectory, and the resulting 55 single structures were scored with GB*n* and GB*n2* models. For brevity’s sake, the figure only shows the five most divergent structures among the 55 central structures, containing the top-scoring ones.

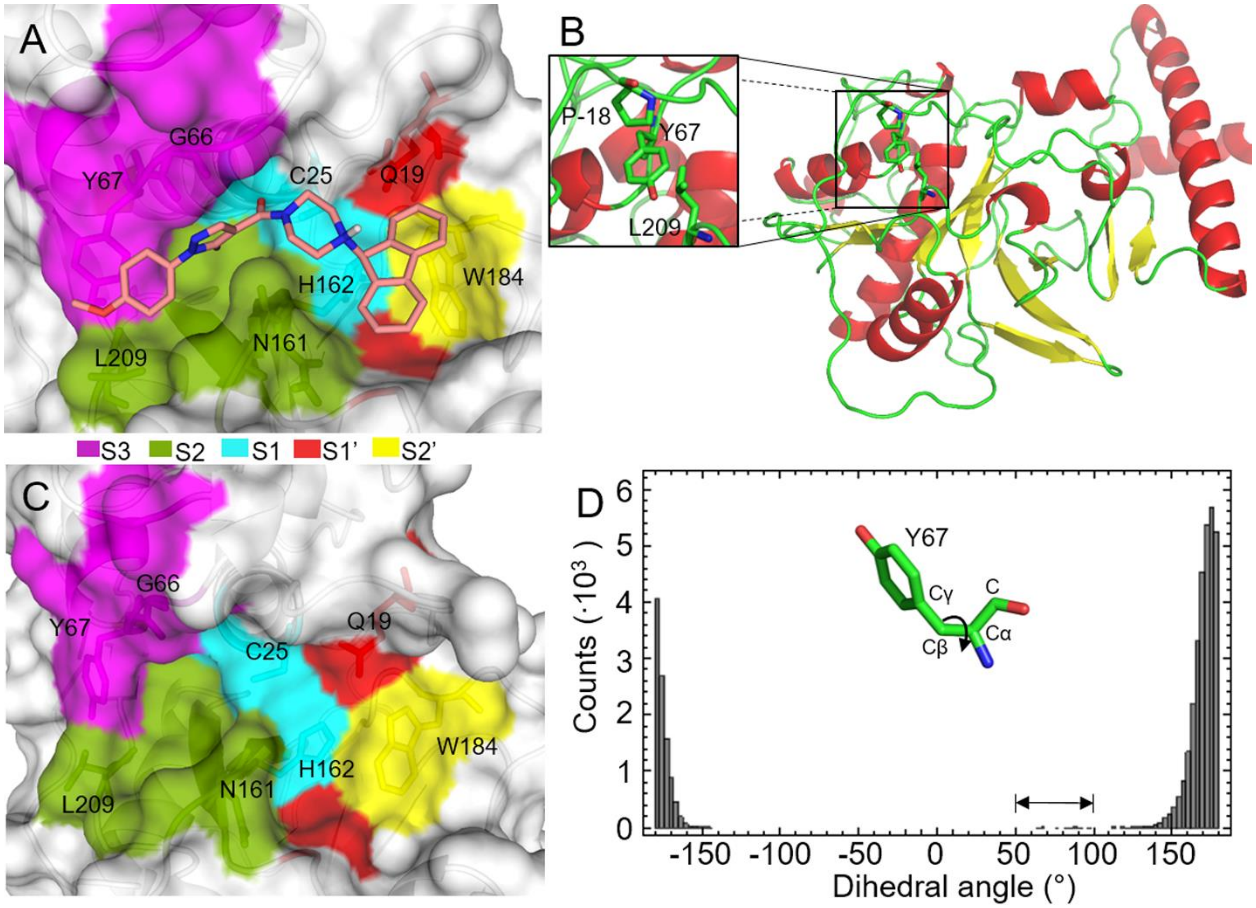


Pose	ΔG_{eff} *GBn* (kcal/mol)	ΔG_{eff} *GBn2* (kcal/mol)
A3”a	-32.42	-30.57
A3”b	-43.96	-41.94
A3”c	-43.07	-41.15
A3”d	-31.82	-33.15
A3”e	-32.08	-29.47

Source: Prepared by the author.

On the other hand, it can be noticed from Fig.3.4-B that the *p*-methoxyphenyl moiety of HTS07940 is partially inserted between residues Y67 and L209 of hCatK. Remarkably, the side-chains of these two residues are in close contact in the structure of free hCatK (PDB: 5TUN, Fig. B9-A),

Figure B8. Alternative ‘down’ conformation of Y67 in hCatK. A) Predicted central structure of hCatK:HTS07940 from a 100 ns MD simulation started with pose A3”c (Fig. B7). Subsites are colored according to legend. **B)** Cartoon representation of human pro-cathepsin K highlighting the Y67-L209 pair and the propeptide residue P-18 in sticks. The zoomed-in view of the three residues Y67, L209 and P-18, shows how the latter pushes down the former’s sidechain to avoid steric clashes between them. The human procathepsin K structure is colored by secondary structure (helices in red, β strands in yellow and loops in green). **C)** Crystal structure of human pro-cathepsin K (PDB: 1BY8), the N-terminal propeptide was removed to expose the active site. **D)** Distribution of the C-Cα-Cβ-Cγ dihedral angle of Y67 during a 0.8 µs MD simulation of apo hCatK. Dihedral angles within the range 50-100° (see double-headed arrow) correspond to a ‘down’ conformation of Y67 in apo hCatK. The probability of finding this ‘down’ conformation in the whole ensemble of structures is ~$3 \cdot 10^{-4}$.



Source: Prepared by the author.

thereby forming a wall that occludes the bottom of the S2 subsite. Thus, the insertion of the *p*-methoxyphenyl moiety between L209 and Y67 side-chains can only occur if both residues become separated from each other. In this sense, we noticed that the histogram of Y67-L209 distance obtained from the MD simulations of free hCatK overlaps, to some extent, that obtained from pose A3”b (Fig. B10). Therefore, Y67-L209 distances typical of the complex are sampled during the MD simulations of free hCatK. By visual inspection, we determined that the *p*-methoxyphenyl moiety fits well into the S2 bottom at Y67-L209 distances larger than 8 Å, a condition fulfilled by nearly 1.2% of the structures in the free hCatK ensemble (probability=1.2 10^{-2}). Hence, it is roughly 40 times more likely to find free hCatK in a conformation capable of accommodating the P2 moiety of compound HTS07940 as in pose A3"b than as in A3"c. When incorporating the relative abundancies of both conformations into the $\Delta\Delta G$ calculation, it turns out that pose A3"b is the most stable one by a $\Delta\Delta G$ of ~3 kcal/mol (Table B3). Therefore, we selected pose A3"b as the most stable binding mode of compound HTS07940 to hCatK.

Table B3. Free energy values for the two top-scoring poses of HTS07940 in complex with hCatK during a 100 ns MD simulation.

Free energies (kcal/mol)	**Poses**	
	A3"b	**A3"c**
ΔG_{eff} (GB*n*)	-34.06	-33.59
ΔG_{eff} (GB*n2*)	-33.78	-33.09
$\Delta\Delta G$ (GB*n*)[a]	-	0.47
$\Delta\Delta G$ (GB*n2*)[a]	-	0.69
$\Delta G_{penalty}$[b]	2.62	4.85
$\Delta G'$ (GB*n*)[c]	-31.44	-28.74
$\Delta G'$ (GB*n2*)[c]	-31.16	-28.24
$\Delta\Delta G'$ (GB*n*)[d]	-	2.70
$\Delta\Delta G'$ (GB*n2*)[d]	-	2.92

[a] Relative free energies were calculated as follows: $\Delta\Delta G=\Delta G_{eff}$(A3"c)-$\Delta G_{eff}$(A3"b), where both ΔG_{eff} values correspond to the same GB model (either GB*n* or GB*n2*).

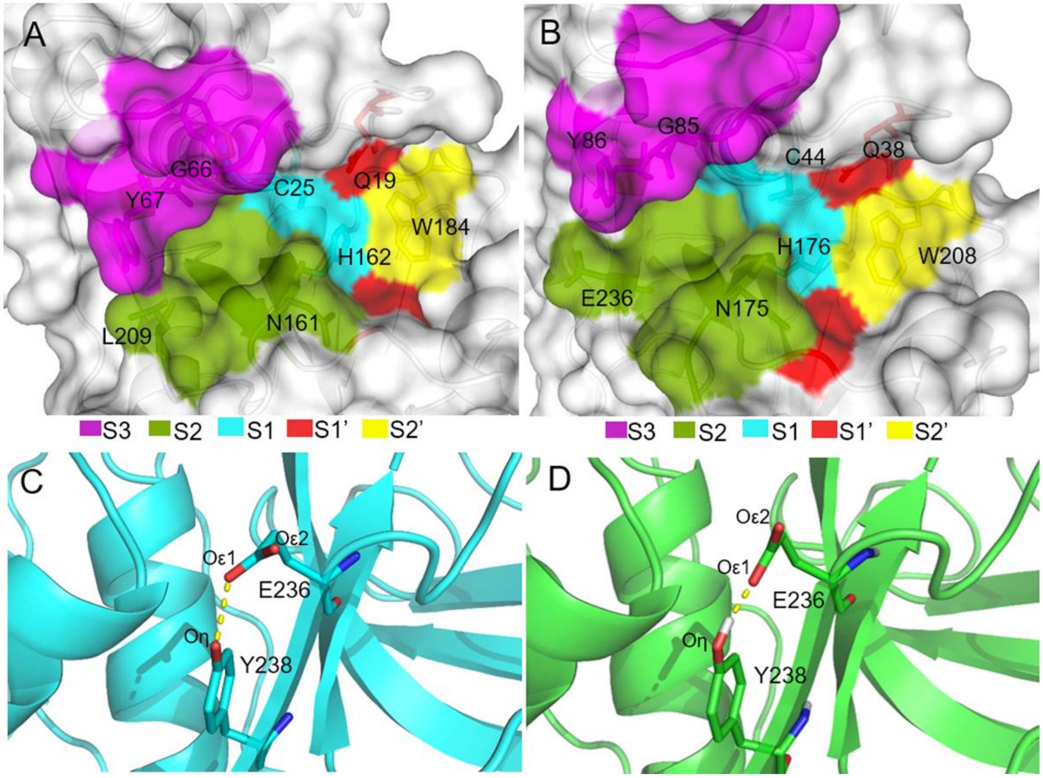
[b] The free energy penalty ($\Delta G_{penalty}$) for binding to conformations i of the protein was calculated as follows: $\Delta G_{penalty}=-RTln(p_i)$, where p_i is the probability of finding conformations i in the whole ensemble of apo structures.

[c] Corrected free energies ($\Delta G'$) were calculated as follows: $\Delta G'=\Delta G_{eff}+\Delta G_{penalty}$.

[d] Corrected relative free energies ($\Delta\Delta G'$) were calculated as follows: $\Delta\Delta G'=\Delta G'$(A3"c)-$\Delta G'$(A3"b), where both $\Delta G'$ values correspond to the same GB model (either GB*n* or GB*n2*).

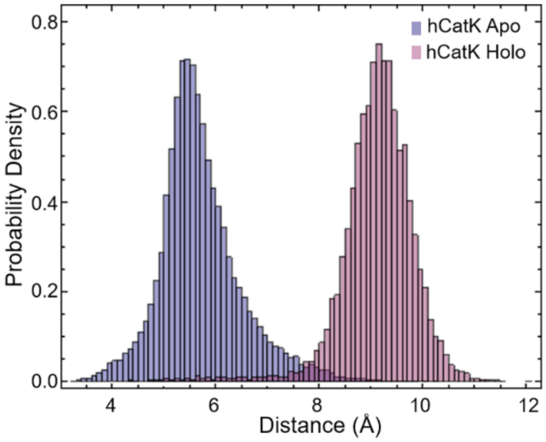
Source: Prepared by the author.

Figure B9. Active sites of hCatK and FP-3. A) Surface representation of the active site of free hCatK (PDB: 5TUN), after reversing the C25S mutation. **B)** Surface representation of the active site of FP-3 after removing the ligand (PDB: 3BWK). Subsites are colored according to the legends. **C)** H-bond between residues E236 and Y238 in the crystal structure of FP-3 (PDB: 3BWK). **D)** H-bond between E236 and Y238 in the central structure of the FP-3:HTS07940 complex. The ligand was removed for clarity's sake. The H-bond mediated by either Oε1 or Oε2 of E236 with Oη of Y238 has a total occupancy of 72% throught the MD simulation of the complex.



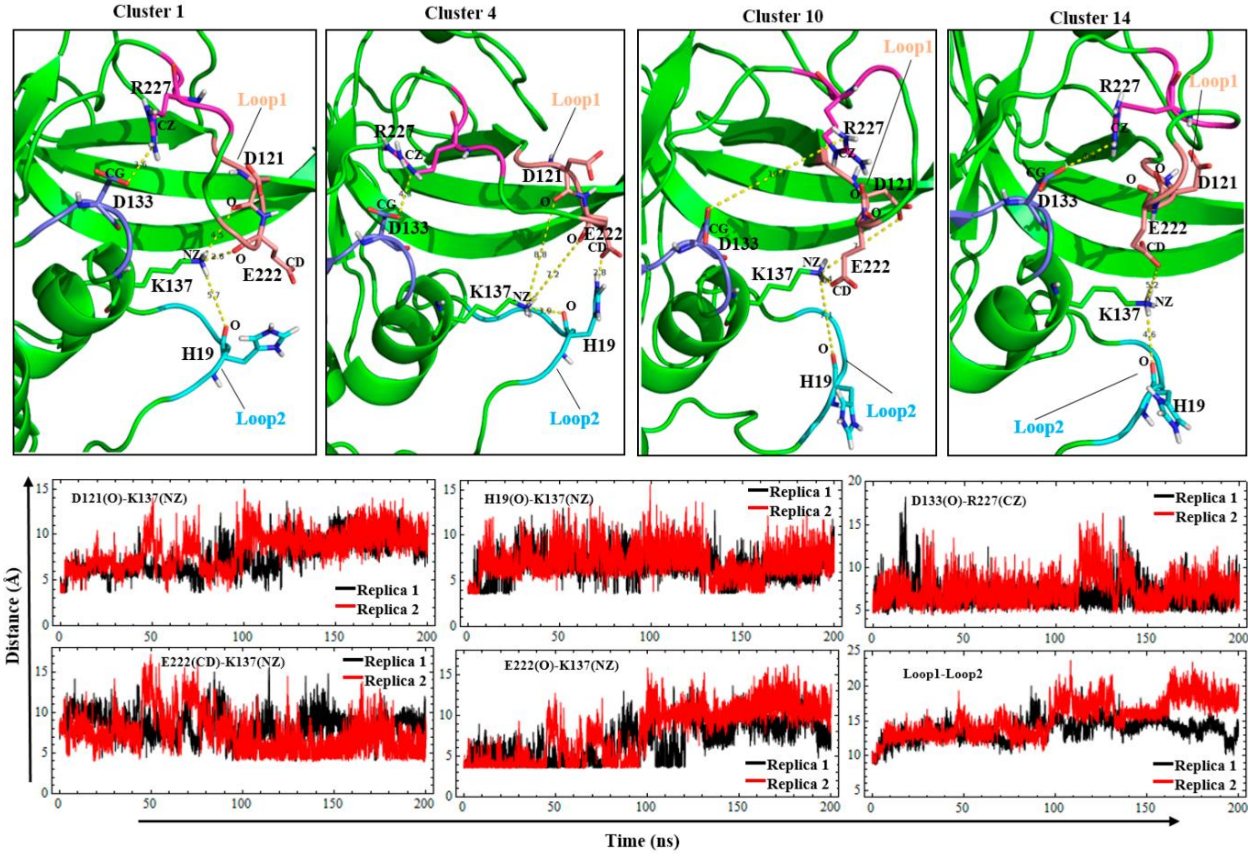
Source: Prepared by the author.

Figure B10. Distance distributions of residues Y67-L209 in free hCatK and in hCatK complex with HTS07940. The distance of the Y67-L209 pair was defined as the minimum distance between Cγ of Y67 and either Cδ1 or Cδ2 of L209. hCatK Holo in the legend refers to hCatK in complex with HTS07940. Note that at a distance larger than 8 Å, the histograms of apo and holo forms of hCatK overlap; thus, open conformations of the Y67-L209 gate are sampled in the simulation of apo hCatK.



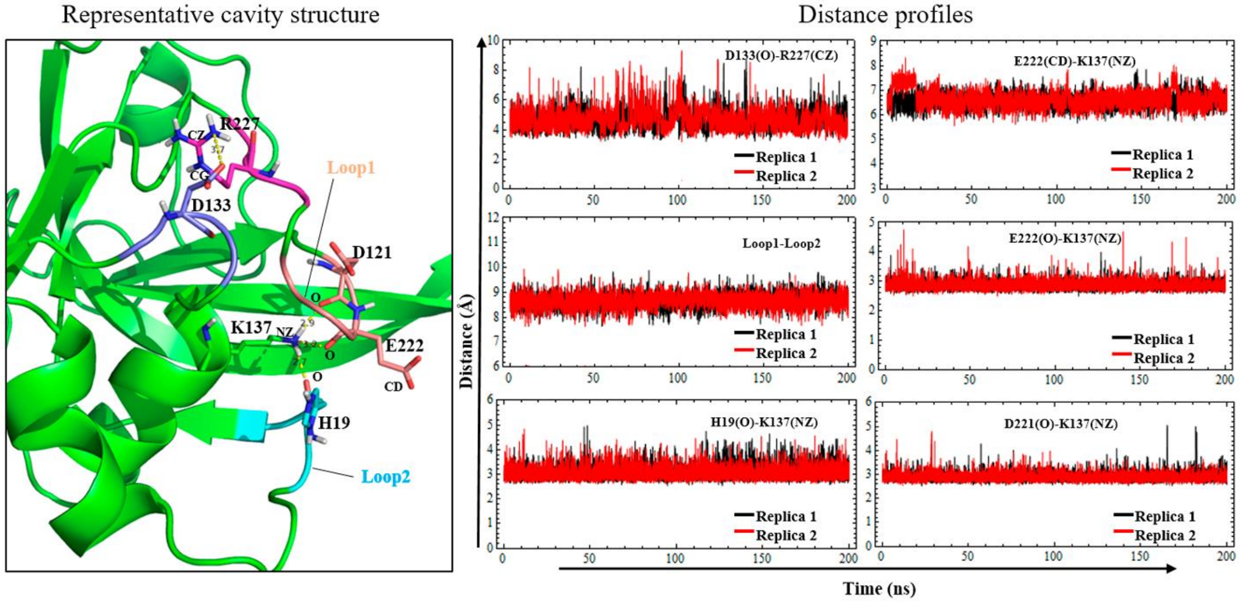
Source: Prepared by the author.

Figure B11: Distance time profiles of interacting oppositely-charged residues of FP-2 site 6 during the simulations conducted with GROMOS 54a8 force-field. The interacting residues are shown for each central structure of site 6. The distance between these loops was calculated with respect to their center of masses.



Source: Prepared by the author.

Figure B12: Distance time profiles of interacting oppositely-charged residues of FP-2 site 6 during the simulations conducted with Amber14SB force-field. The interacting residues are shown for each central structure of site 6. The distance between these loops was calculated with respect to their center of masses.



Source: Prepared by the author.

References

(1) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, 23, 2947-2948.

(2) Madhusudhan, M. S.; Webb, B. M.; Marti-Renom, M. A.; Eswar, N.; Sali, A., Alignment of Multiple Protein Structures Based on Sequence and Structure Features. *Protein Eng Des Sel* **2009**, 22, 569-574.

(3) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J Cheminform* **2012**, 4, 17.

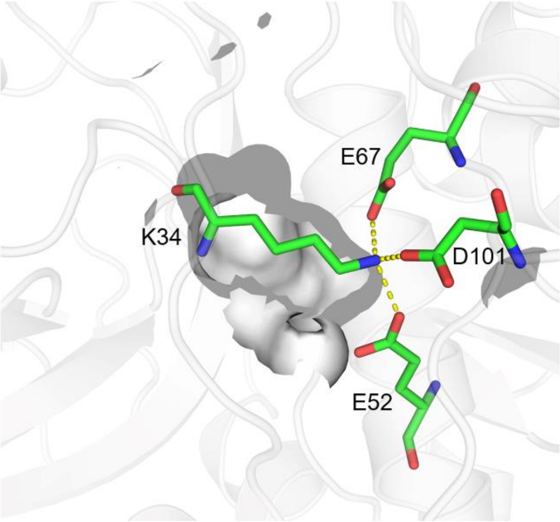
(4) Perozzo, R.; Kuo, M.; Sidhu, A.; Valiyaveettil, J. T.; Bittman, R.; Jacobs, W. R., Jr.; Fidock, D. A.; Sacchettini, J. C., Structural Elucidation of the Specificity of the Antibacterial Agent Triclosan for Malarial Enoyl Acyl Carrier Protein Reductase. *J Biol Chem* **2002**, 277, 13106-13114.

(5) Khalili, F.; Henni, A.; East, A. L., pKa values of some Piperazines at (298, 303, 313, and 323) K. *J. Chem. Eng. Data* **2009**, 54, 2914-2917.

(6) Harms, M. J.; Castaneda, C. A.; Schlessman, J. L.; Sue, G. R.; Isom, D. G.; Cannon, B. R.; Garcia-Moreno, E. B., The pK(a) Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors. *J Mol Biol* **2009**, 389, 34-47.

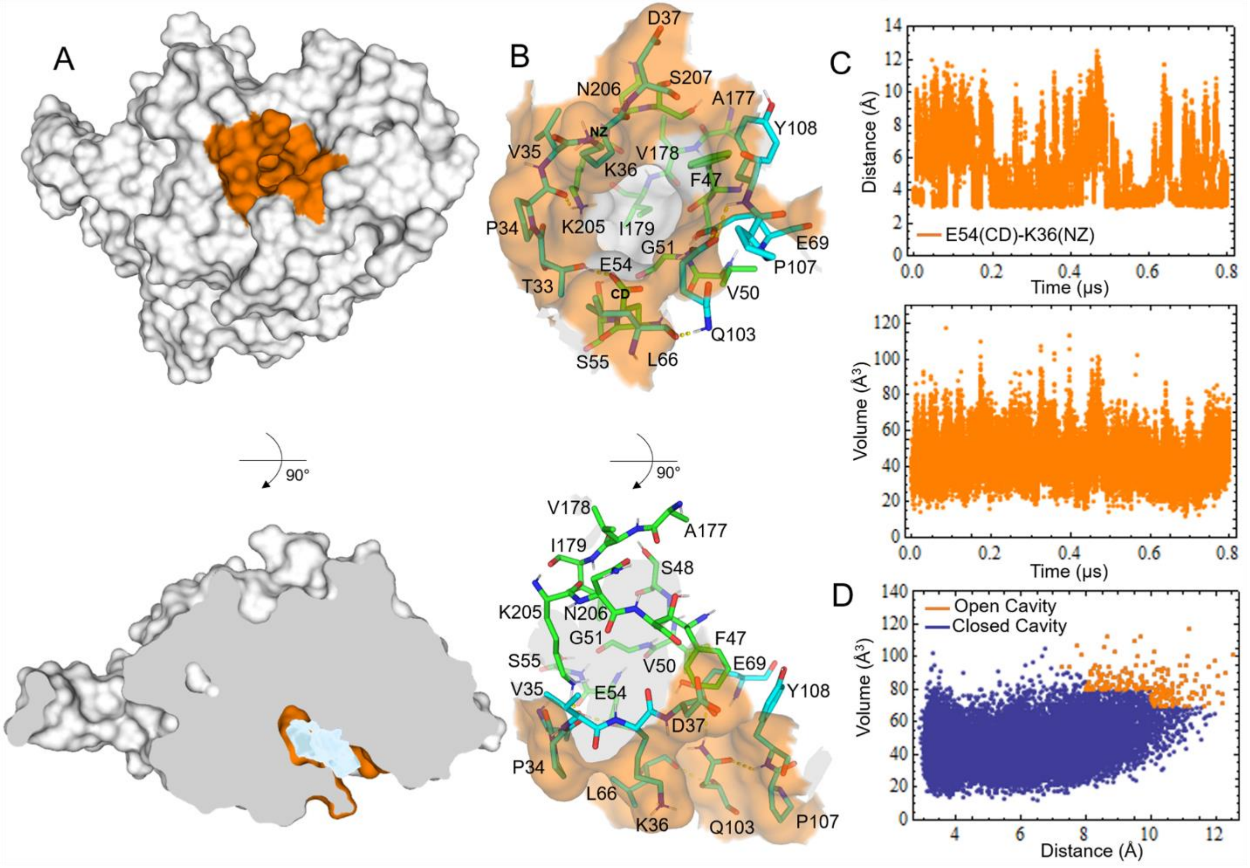
APPENDIX C – Additional information to the prediction of the binding mode of compound 66 to FP-2 and to the molecular mechanism of the exerted non-competitive inhibition

Figure C1. Occlusion by residue K34 of site 3 internal cavity in FP-2 crystal structure 2OUL. The internal cavity walls are shown in gray. Residue K34, which acts as a lid, and the acidic residues stabilizing its side-chain conformation are depicted as sticks.



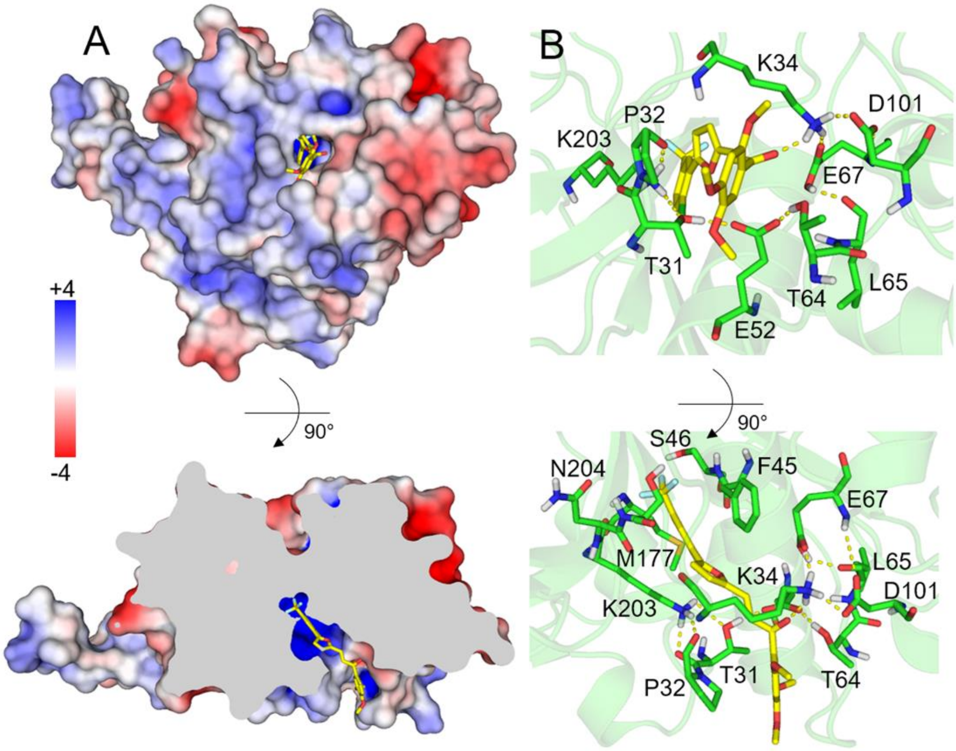
Source: Prepared by the author.

Figure C2. Formation of a transient pocket in the site 3 region of FP-3. A) Surface representation of a central structure of FP-3 having an open pocket conformation. The surface region corresponding to the transient pocket entrance is depicted in orange. A protein slice after 90° rotation shows the inner space of the pocket. The pocket internal volumetric density is represented in light blue. **B)** Detailed structural representation of the open pocket conformation. Residues lying at the pocket entrance are depicted as cyan sticks surrounded by a transparent orange surface. Residues forming the pocket internal wall are shown as green sticks. The internal surface of the pocket is colored in gray. H-bonds occurring at the pocket entrance are indicated with yellow dashed lines. **C)** Graphs showing the E54(CD)-K36(NZ) distance and site 3 internal volume time profiles during the concatenated MD simulations of FP-3. **D)** Volume versus E54(CD)-K36(NZ) distance graph. Dots depicted in orange correspond to frames bearing an open pocket. For FP-3, the conditions established to define an open cavity were 10 Å >E54(CD)-K36(NZ) distance ≥ 8 Å and a pocket inner volume ≥ 80 Å^3, or E54(CD)-K36(NZ) distance ≥ 10 Å and a pocket inner volume ≥ 70 Å^3.



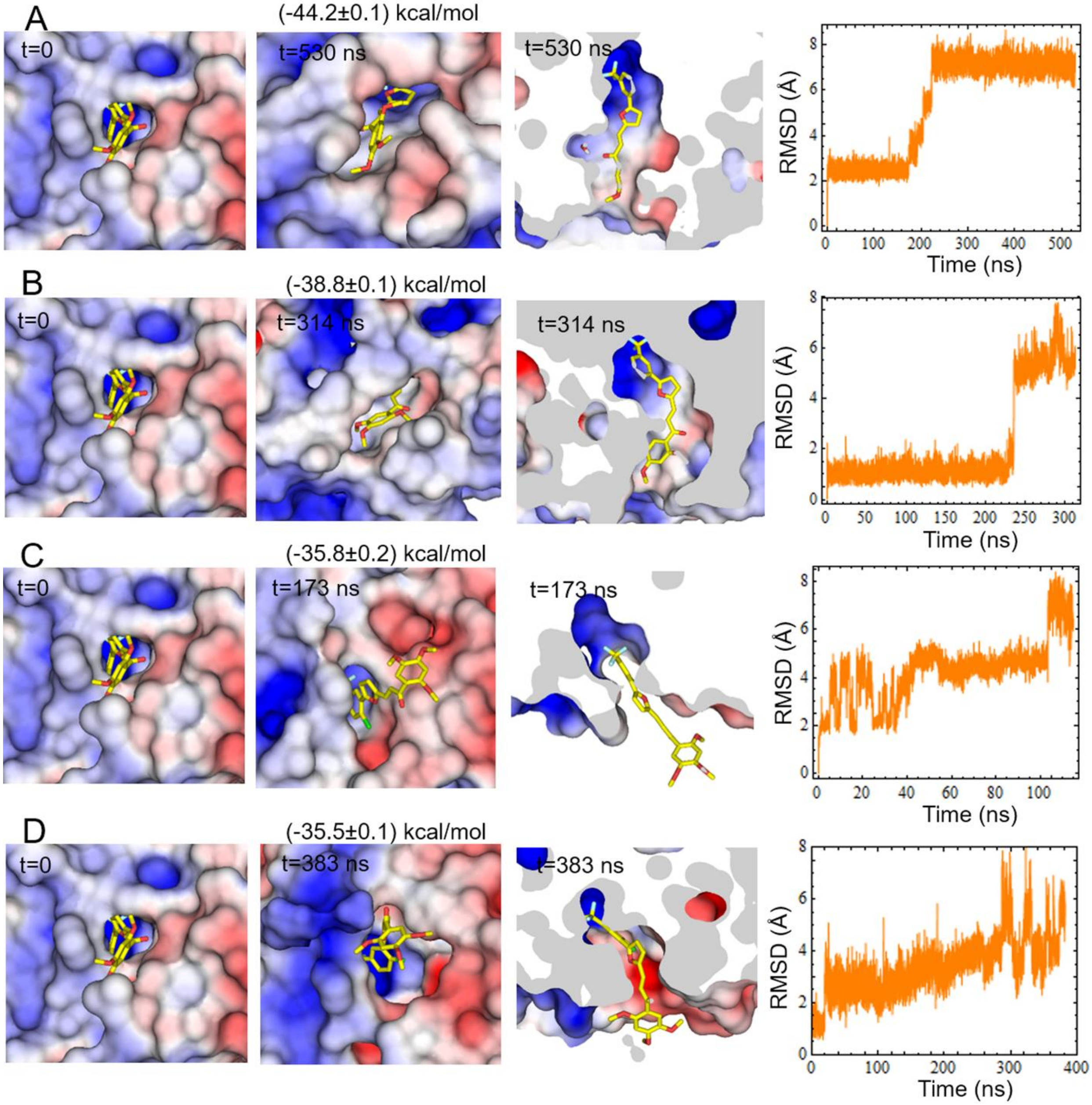
Source: Prepared by the author.

Figure C3. Best docking pose of Cpd66 into site 3. A) Two views of the best docking pose showing the electrostatic surface of FP-2 calculated with the APBS plugin of pymol 2.1.0. The lower image corresponds to a protein slice after performing a 90° rotation with respect to the upper representation. The surface is colored according to the gradient of electrostatic potential expressed in k_bT/e units, where k_b, T and e stand for the Boltzmann's constant, the temperature (298.15 K) and the electron charge, respectively. **B)** Detailed views of the complex interface. Residues interacting with Cpd66 are shown as sticks. Yellow dashed lines indicate H-bonds occurring at the interface. The inhibitor is depicted as yellow sticks in all cases.



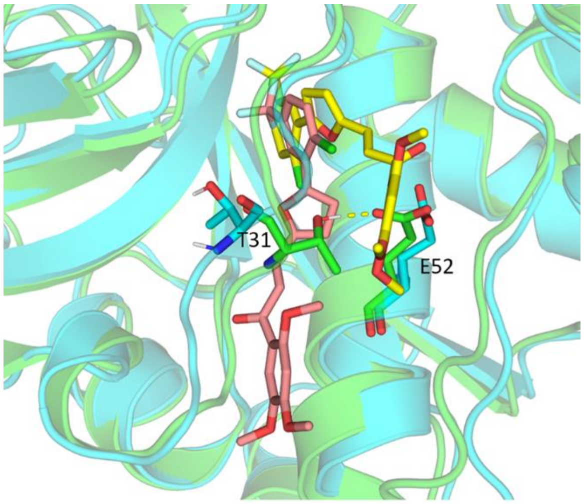
Source: Prepared by the author.

Figure C4. Time evolution of the FP-2:Cpd66 complex determined by docking during four independent MD simulations. A), **B)**, **C)** and **D)** represent independent MD simulations started from the same structure (*t*=0) but using different randomly-generated velocities. The electrostatic surface of the protein was colored according to the gradient shown in Fig. C3. Two views of last frame obtained from each simulation are depicted and labeled accordingly. On top of the first view, the ΔG_{eff} mean value corresponding to the last 50 ns of each trajectory is provided. Graphs displaying the time profiles of RMSD values calculated for the Cpd66 heavy atoms after fitting each trajectory to the corresponding starting frames and with respect to the protein's backbone atoms, are shown in all cases. Simulations were halted at different times, when sufficient data were obtained to definitely propose the most stable binding mode.



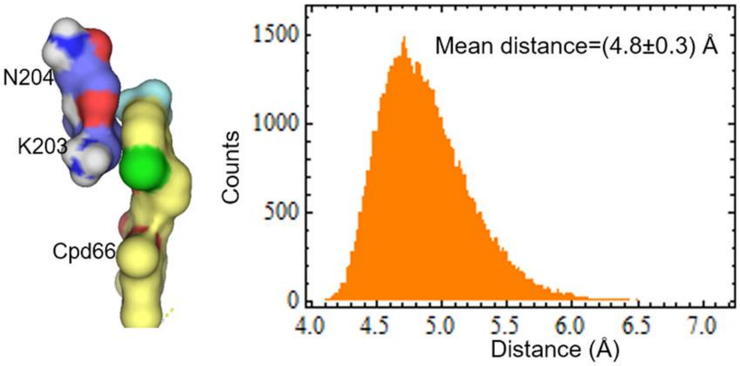
Source: Prepared by the author.

Figure C5. Comparison of the docking and MD-generated conformations of Cpd66 in complex with FP-2. The FP-2:Cpd66 complex used as starting structure for the MD simulations is represented as green cartoon (protein) and yellow sticks (ligand). The complex structure obtained from the MD simulation leading to the most stable binding mode is shown as cyan cartoon (protein) and salmon sticks (ligand). Note that the H-bond (yellow dashed line) mediated by T31 and E52 breaks when Cpd66 adopts the final conformation.



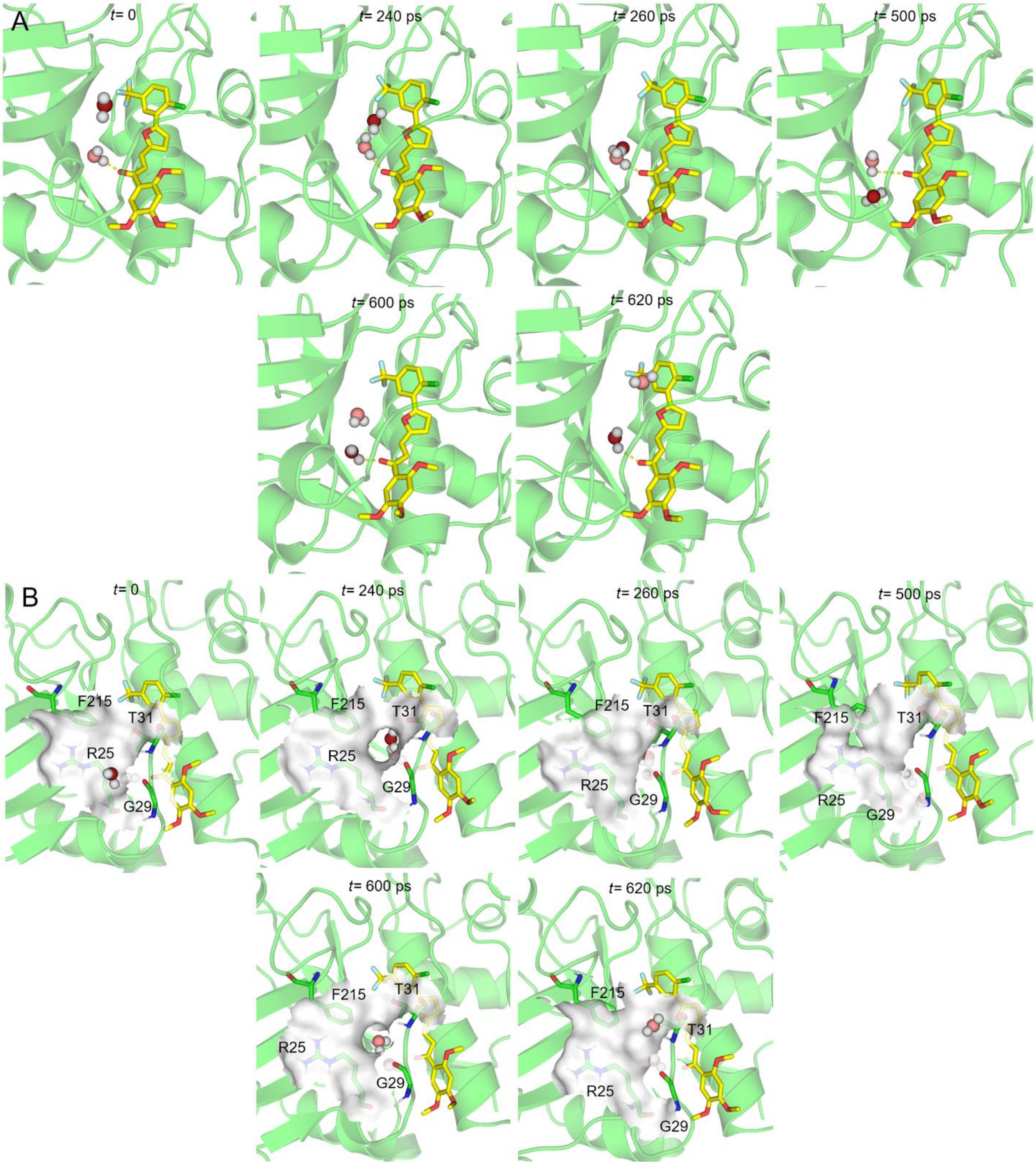
Source: Prepared by the author.

Figure C6. Contact between residue K203 and the six-membered aryl ring of Cpd66 in the FP-2:Cpd66 complex. Residues K203 and N204, whose peptide bond is susceptible to trypsin cleavage, and Cpd66 are shown in surface representation. The distribution of distances between the heavy atoms of K203 and those of the six-membered aryl ring of Cpd66 is shown on the right. The distances between centers-of-mass of the previously-mentioned atom groups were calculated during the 1.2 μs concatenated MD simulations of the FP-2:Cpd66 complex.



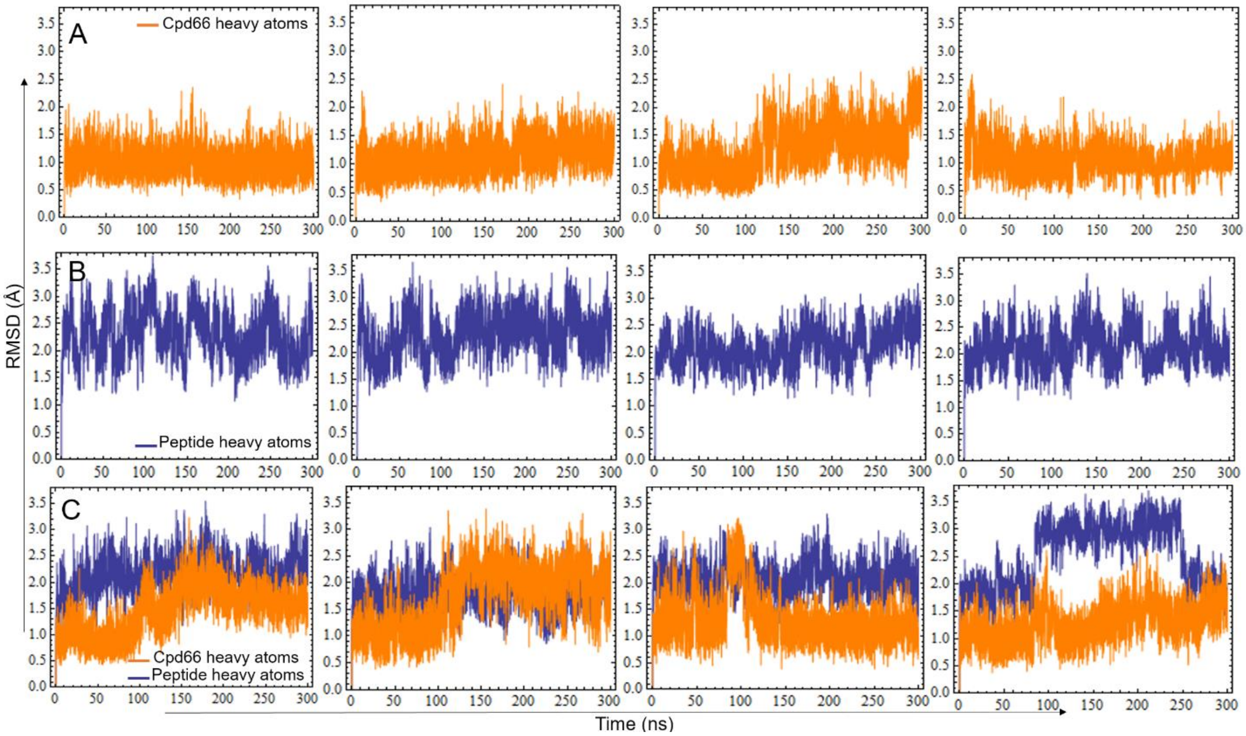
Source: Prepared by the author.

Figure C7. Exchange of water molecules mediating the H-bond between Cpd66 and G29. A) Frames collected from an MD simulation of the FP-2:Cpd66 complex showing the water exchange during an interval of 620 ps are sequentially represented. Note that *t*=0 does not indicate the first frame of the trajectory, but the one taken as a reference. The water molecules are colored differently to allow their identification throughout the frames. **B)** Same as A) but showing the protein surface (gray) created by residues R25, G29, T31, and F215. A hole on this surface is transiently created, thus allowing the water exchange.



Source: Prepared by the author.

Figure C8. RMSD time profiles for the Cpd66 and peptide during the replicate MD simulations of the studied complexes. RMSD values calculated during the MD simulations of **A)** the FP-2:Cpd66 complex, **B)** the FP-2:peptide complex and **C)** the FP-2:peptide:Cpd66 complex. Time profiles of RMSD values were calculated for the Cpd66 and the peptide heavy atoms after fitting each trajectory to the corresponding starting frames and with respect to the protein's backbone atoms.

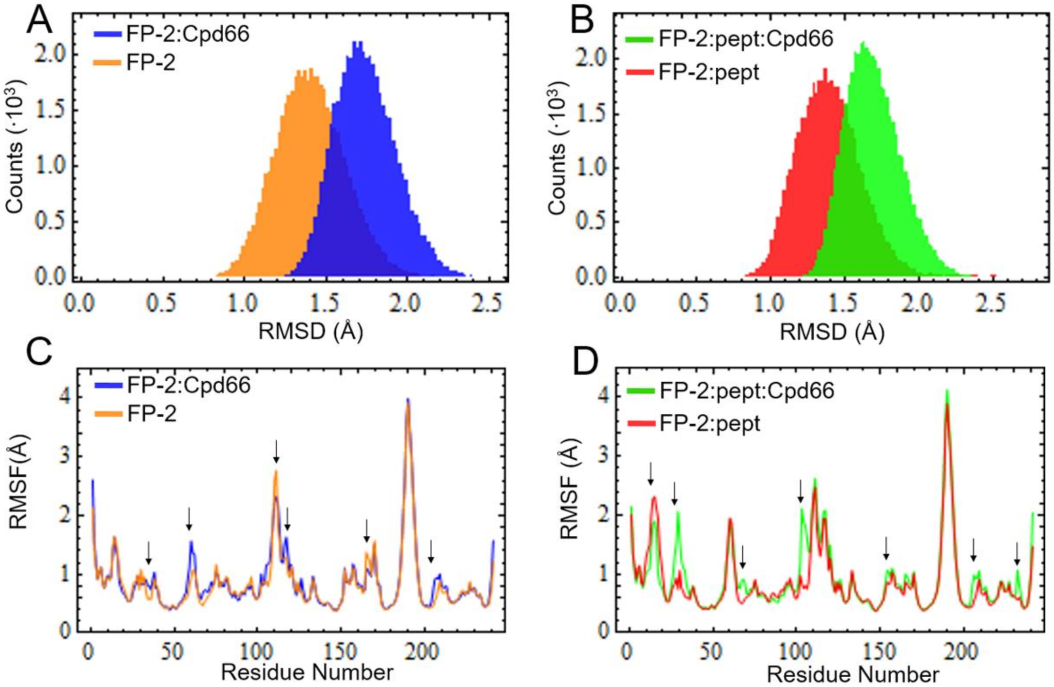


Source: Prepared by the author.

Text C1. Cpd66 binding induces subtle perturbations in the FP-2 conformational space

The allosteric inhibition exerted by Cpd66 on FP-2 must lead to some changes in the protein, ultimately responsible for slowing down the enzymatic activity. We started the characterization of such changes by means of coarse measurements, such as backbone RMSD and RMSF calculations throughout the MD simulations (Fig. C9). Interestingly, the Cpd66 binding causes a shift in the backbone RMSD histograms calculated with respect to the FP-2 crystal structure (PDB: 2OUL) toward the right, when compared to the systems lacking the inhibitor (Figs. C9-A and B). Therefore, when Cpd66 is bound, the enzyme occurs in conformations that are, on average, more divergent with respect to the crystal structure than in the absence of this ligand. However, this effect is largely caused by the reorganization of the loops surrounding Cpd66 and do not reflect necessarily distal changes. In addition, the calculation of RMSD histogram widths, i.e., the standard deviations, reveals that the global diversity of FP-2 conformations slightly narrows in the presence of Cpd66 (Figs. C9A and C9B). In other words, the protein conformations sampled during the MD simulations of the FP-2:Cpd66 and FP-2:peptide:Cpd66 complexes tend to be less diverse in terms of backbone RMSD values if compared to those generated for free FP-2 and the FP-2:peptide complex.

Figure C9. Comparison of RMSD distributions and per-residue RMSF values for the studied systems. A) and **B)** backbone RMSD distributions for the indicated systems with respect to the FP-2 crystal structure 2OUL. The standard deviations of the RMSD distributions for the FP-2, FP-2:Cpd66, FP-2:peptide and FP-2:peptide:Cpd66 systems are 0.214, 0.189, 0.227 and 0.187 Å, respectively; and the mean values, 1.40, 1.74, 1.40 and 1.69 Å, respectively. **C)** and **D)** Per-residue RMSF values calculate for the protein's backbone with respect to the average position of each residue during the concatenated 1.2 μs MD simulations conducted for each system. The arrows indicate regions that appreciably changed their flexibility in the compared systems.



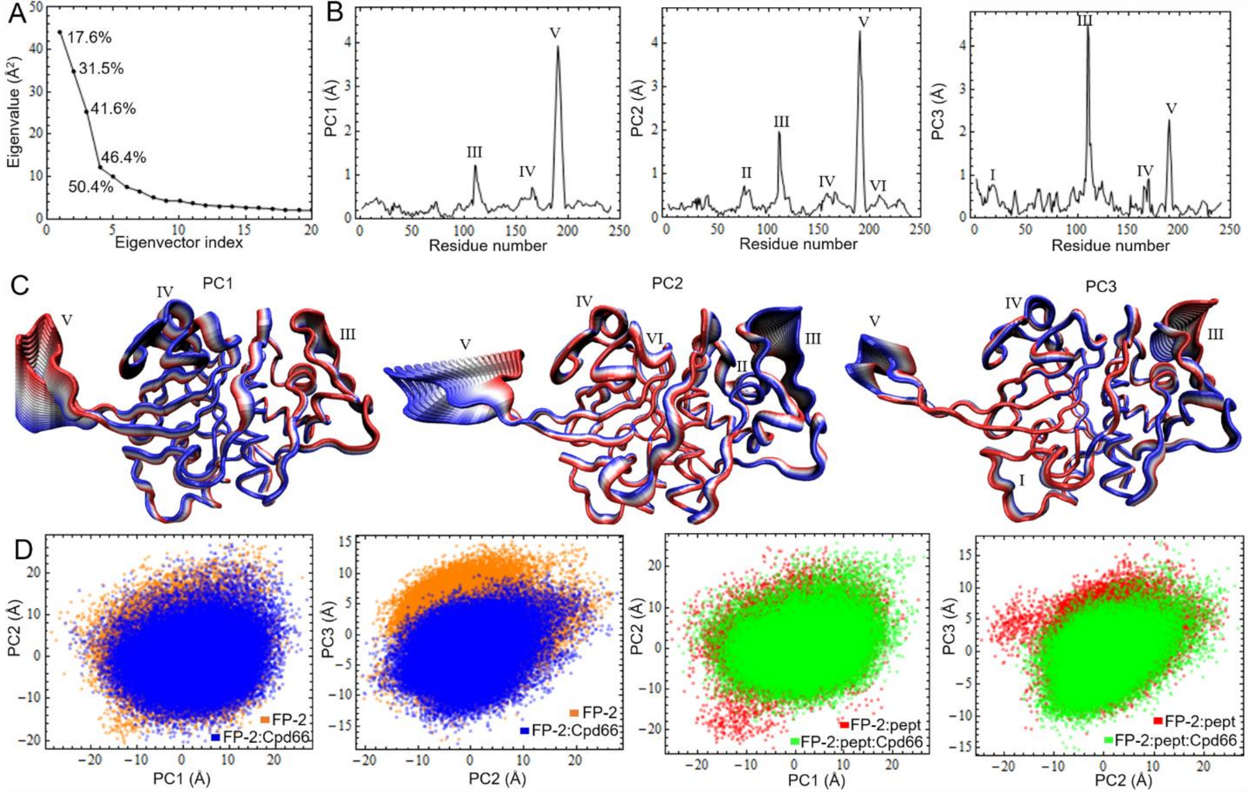
Source: Prepared by the author.

The analysis of per-residue backbone RMSF values allowed us to identify protein regions that changed their local flexibility in the presence of the inhibitor (see arrows, Figs C9C and C9D). The binding of Cpd66 to FP-2 promotes the most appreciable RMSF variations between the apo and holo systems in the following residues: 33-39, 58-63, 106-109, 111-114, 117-120, 166-168 and 206-210 (Fig. C9-C), all of them belonging to loops. Remarkably, the flexibility change in the latter loop, which contains residues W206 and W210, might also explain, to some extent, the variation in Trp intrinsic fluorescence emission between free FP-2 and the FP-2:Cpd66 complex,[1] as microenvironment modifications are expected to occur due to an increase in loop motion upon ligand binding. Residue W24, besides interacting directly with Cpd66, contacts some residues belonging to the segments 33-39 and 58-63, whose RMSF profiles varied because of the ligand binding (Fig. C9-C). This reinforces our conclusion that changes in the microenvironment of W24 are likely to be critical to explain the shift of Trp emission maximum observed in the fluorescence experiments.[1]

For the FP-2:peptide and FP-2:peptide:Cpd66 complexes, the main RMSF variations are found in the positions: 10-21, 27-36, 62-70, 100-107, 154-157, 205-211 and 233-235, all lying in loop regions as well (Fig. C9-D). Some of the previously-mentioned loop segments either contain residues directly interacting with the inhibitor in both FP-2:Cpd66 and FP-2:peptide:Cpd66 complexes, e.g., those involving positions from 27 to 39, or lie close to the interface, e.g., those comprising the 58-70 and 205-211 positions. However, the remaining loops are located farther from the allosteric pocket, thus indicating that the perturbations triggered by the inhibitor can propagate toward distal regions of the protein. Moreover, the binding of Cpd66 tends to increase the fluctuations of various loops (Figs. C9-C and D), especially those in vicinity of the compound. Of note, the previous result is not in contradiction with the previous RMSD analysis, which showed a decreased conformational diversity in the presence of the inhibitor. The RMSF measures the average fluctuation of each atom (or set of atoms) during the MD simulation with respect to a reference position, whereas the RMSD quantifies the overall divergence of each conformation of the ensemble with respect to the reference structure. Therefore, flexible but small regions do not necessarily lead to large RMSD variations. Conversely, less flexible but larger regions will have higher weights in the calculations and, consequently, their contribution to the final RMSD value is likely to be more significant.

In addition, we performed a PCA to compare the motions of the studied systems along the main PCs.[2] As shown in Fig. C10-A, the eigenvalues associated with the PCs of free FP-2 rapidly decrease up to the fourth PC (Fig. C10-A). This indicates that the first three PCs are associated with the largest motions of the protein and account for nearly 42% of the cumulative fluctuation (Fig. C10-A). Of note, PC1, PC2 and PC3 are associated with the collective motions of various enzyme loops that also display large RMSF values (Figs. C9-C and C10-B and C). The 2D projections of the FP-2 and FP-2:Cpd66 complex trajectories onto the PC1, PC2 and PC2, PC3 subsets of the free enzyme show a significant overlap; thereby indicating that protein motions along the selected PCs are rather similar in both systems. However, FP-2 samples, in general, motions of larger amplitude in the absence of the inhibitor, especially along PC3 (Fig. C10-D). Similar conclusions can be drawn from the 2D projections of the FP-2:peptide and FP-2:peptide:Cpd66 trajectories onto the same PC subsets employed before (Fig. C10-D). Therefore, the Cpd66 binding tends to slightly narrow the concerted movements of the protein along the first three PCs regardless of the presence of the peptide in the active site. Nonetheless, the peptide seems to counter this effect to some extent, as greater overlaps of the 2D projections are observed for the FP-2:peptide and FP-2:peptide:Cpd66 complexes (Fig. C10-D).

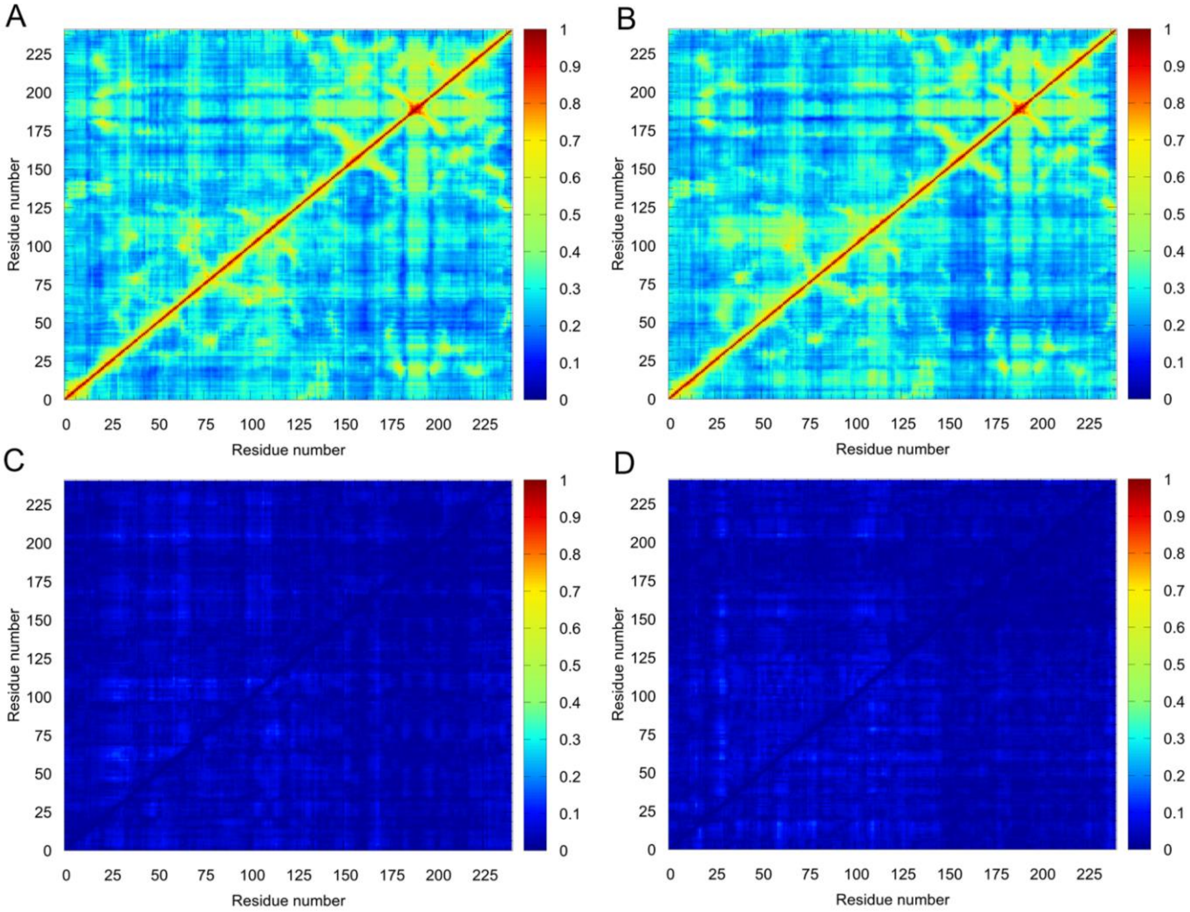
Figure C10. PCA results for the studied systems. A) Eigenvalues of the first twenty eigenvectors corresponding to the 1.2 μs concatenated MD simulations of free FP-2. The cumulative fluctuations of the first five eigenvectors are shown beside their corresponding points in the graph. **B)** Per-residue components of PC1, PC2 and PC3 of free FP-2. Peaks displaying the largest vector components are labeled using Roman numbers. **C)** Structural representation of the collective motions in free FP-2 along PC1, PC2 and PC3. Regions corresponding to the peaks shown in B) are labeled accordingly. **D)** 2D projections of the protein motions along the PC1, PC2 and PC2, PC3 subsets of free FP-2 PCs for the indicated systems. The PCA was conducted for the Cα atoms of the protein.



Source: Prepared by the author.

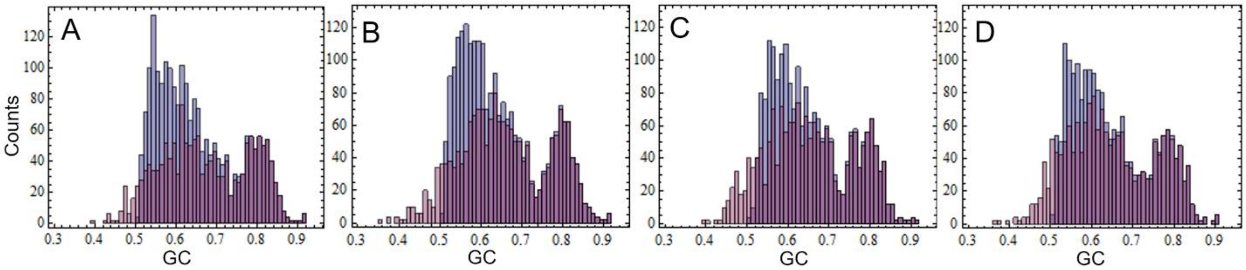
Figure C11. Generalized correlations for the studied systems. *GC* matirices for **A)** free FP-2 (upper triangle) and the FP-2:Cpd66 complex (lower triangle), and **B)** the FP-2:peptide (upper triangle) and the FP-2:peptide:Cpd66 complexes (lower triangle). RMSD of *GC* matrices for **C)** free FP-2 (upper triangle) and the FP-2:Cpd66 complex (lower triangle), and **D)** for the FP-2:peptide (upper triangle) and the FP-2:peptide:Cpd66 complexes (lower triangle). GC_{ij} RMSD values were calculated from the four replicate MD simulation of each system.

Source: Prepared by the author.



Source: Prepared by the author.

Figure C12. Distribution of raw and filtered generalized correlations for the four studied systems. A) Free FP-2, **B)** FP-2:Cpd66 complex, **C)** FP-2:peptide complex and **D)** FP-2:peptide:Cpd66 complex. The bars in blue correspond to distributions of raw $GC \geq 0.5$. The bars colored in light purple correspond to distributions of $GC \geq 0.5$ in at least one replicate MD simulation of each system and filtered using residue-residue contact maps, which zero all correlations of distant residues. The overlap between both distributions is highlighted in dark purple in each graph. Two residues are considered to be in contact if their heavy atoms are closer than 5.0 Å during, at least. 75% of the simulation time. Note that for $GC \geq 0.6$, both distributions largely match, thus indicating that long-range correlations have been greatly ruled out when using this higher *GC* cut-off.



Source: Prepared by the author.

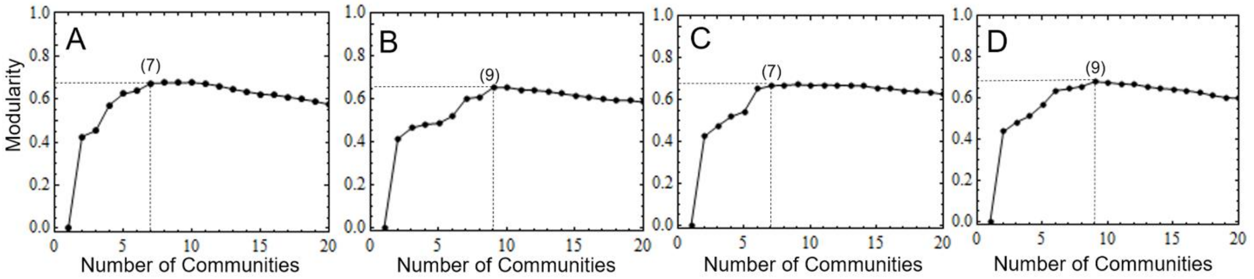
Table C1. Residue composition of each community of the studied systems using a 5.0 Å distance cut-off for residue-residue contact definition

Community No.	FP-2	FP-2:Cpd66	FP-2:peptide	FP-2:peptide:Cpd66
1	1:15, 126:144, 236:241.	1:15.	1:16.	1:16.
2	16:33, 179:202, 208:226.	16:33, 145, **178:204**, 215:226.	17:35, 156:164, 179:203, 205:226.	17:23, 179:202, 217:226.
3	34:35, 145:150, **174:178**, **203:207**[a], 231:232, 234:235.	34, 63:74, 96:124.	**36:63**, 238, 240:241.	24:33.
4	**36:41**, 76:96.	35, 153:167, 205:214.	64:74, 97:126.	34, 63:73, 95, 97:123.
5	**42:63**.	**36:41**, 75:95.	75:96.	35, 154:164, **203:216**.
6	64:75, 97:125.	**42:62**.	127:130, 145:155, **165:178**, **204**, 227:237, 239.	**36:47**, 74:94, 96.
7	151:173, 227:230, 233.	125:127, 238:241.	131:144.	48:62, 124:126, 238, 240:241.
8	-	128:130, 146:152, **168:177**, 227:237.	-	127:144, 237, 239.
9	-	131:144.	-	145:153, **165:178**, 227:236.

[a] Residue segments containing key residues for catalysis, Q36, C42, H174 and N204, are highlighted in bold.

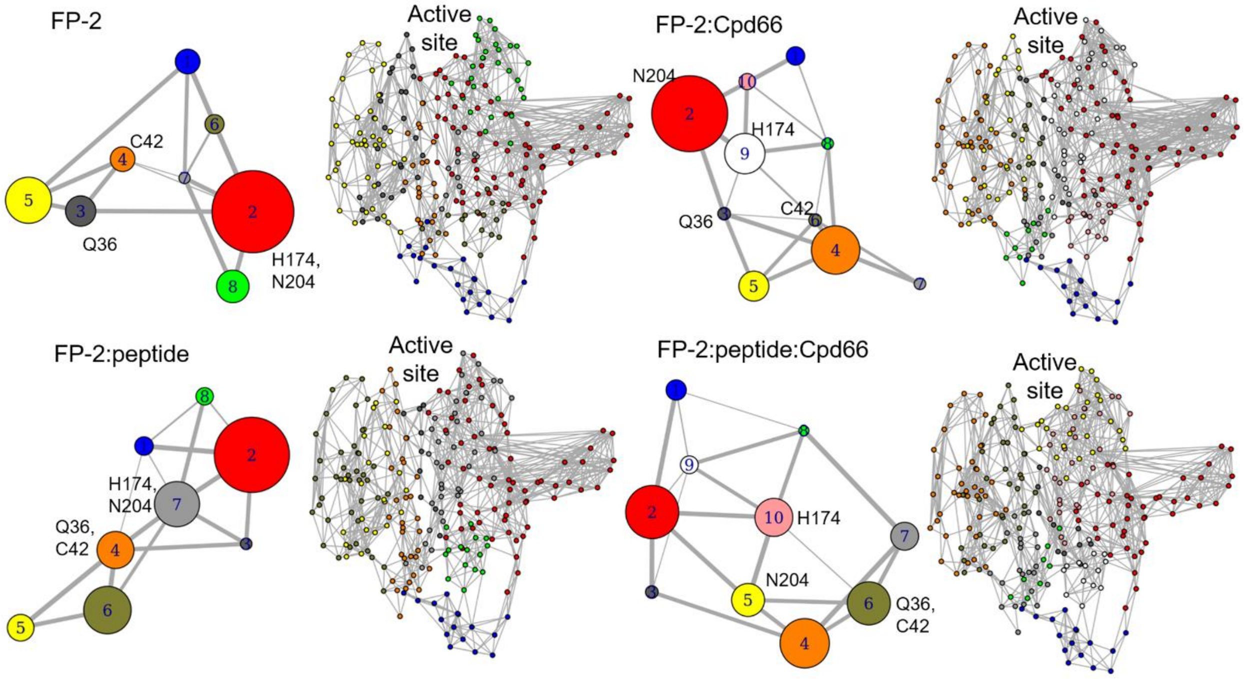
Source: Prepared by the author.

Figure C13. Dependency of modularity values on the number of communities for the analyzed systems. A) Free FP-2, **B)** FP-2:Cpd66 complex, **C)** FP-2:peptide complex and **D)** FP-2:peptide:Cpd66 complex. The optimal number of communities for each system is the smallest one yielding a nearly maximal modularity (a 0.01 tolerance from the modularity absolute maximum was chosen in each case). The calculated numbers of communities are indicated between parentheses in each graph.



Source: Prepared by the author.

Figure C14. Community analysis for the studied systems using a distance cut-off of 4.5 Å to define residue-residue contacts. 2D diagrams showing the organization of communities for each system are depicted on the left. The communities are represented by circles whose diameters are proportional to the number of residues contained within each. The lines account for intercommunity correlations, which are proportional to the line width. For convenience, communities are colored and numbered differently. The labels of residues Q36, C42, H174 and N204, are shown beside their corresponding communities. On the right, the community organization is represented as a diagram mimicking the 3D-structure of FP-2, with nodes colored according to their community membership. The dots and gray lines stand for the network nodes (Cα atoms) and their pairwise *GCs*, respectively.



Source: Prepared by the author.

Table C2. Residue composition of each community of the studied systems using a 4.5 Å distance cut-off for residue-residue contact definition

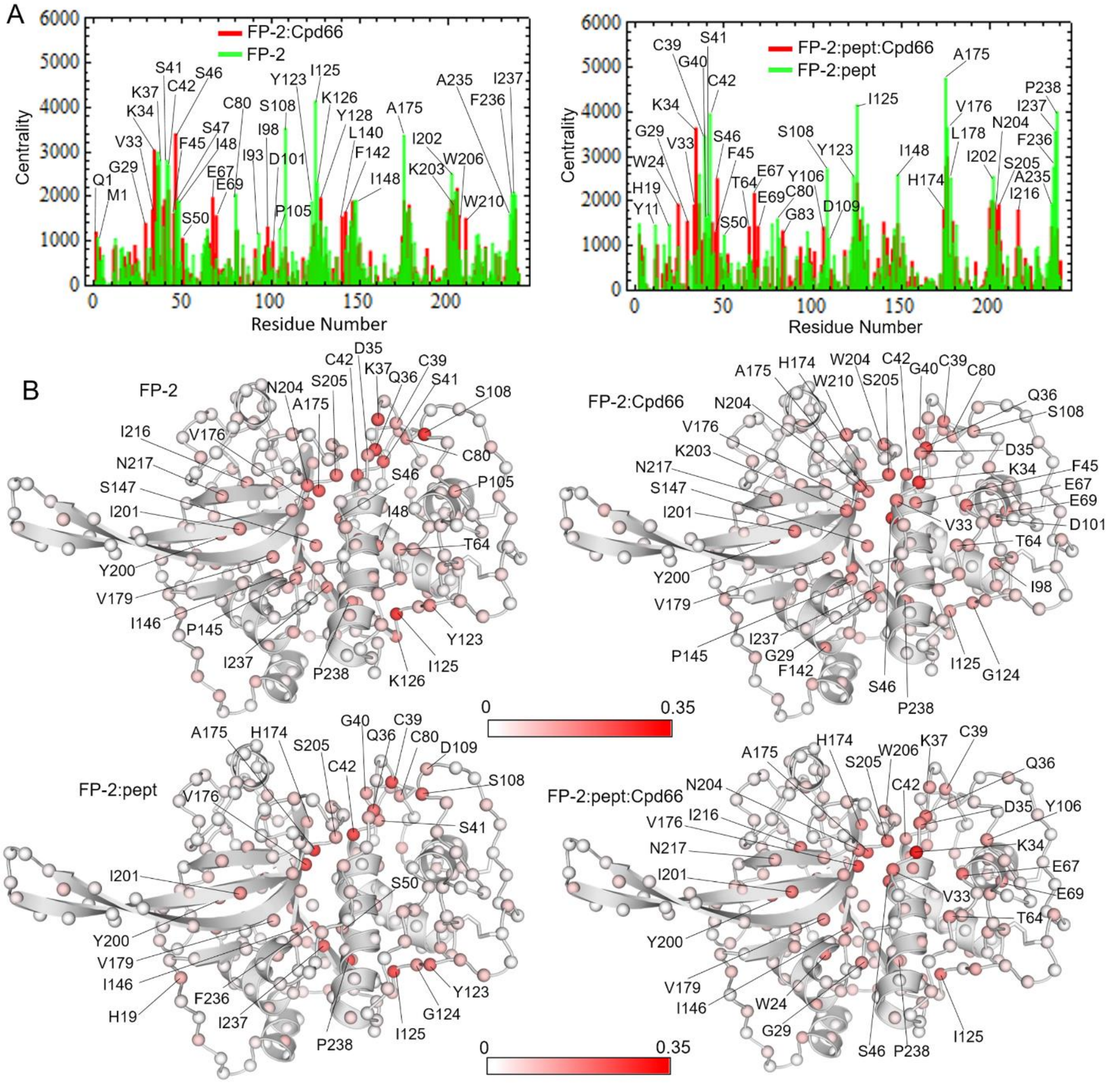
Community No.	FP-2	FP-2:Cpd66	FP-2:peptide	FP-2:peptide:Cpd66
1	1:16, 126:127, 238:241.	1:16.	1:16.	1:16.
2	17:35, 145, **174:226**[a].	17:25, 157:164, **179:204**, 206:228.	17:28, 156:164, 179:203, 208:226.	17:23, 178:202, 217:226.
3	**36:41**, 76:96.	26:33, **35:36**, 205.	29:35, 205:207.	24:33.
4	**42:63**.	34, 52, 63:74, 97:124.	**36:39**, **41:63**, 127, 238:241.	34, 64:74, 97:123.
5	64:75, 97:125.	37:40, 75:96	40, 75:96	35, 154:164, **203:216**
6	128:144.	**41:51**.	64:74, 97:126.	**36:47**, 75:96.
7	146:149, 232:237.	53:62.	128:130, 146:155, **165:178**, **204**, 227:237.	48:63, 124:126, 238, 240:241.
8	150:173, 227:231.	125:129, 237:241.	131:145.	127:130, 235:237, 239.
9	-	130, 145:156, **165:178**, 229:236.	-	131:144.
10	-	131:144.	-	145:153, **165:177**, 227:234.

[a] Residue segments containing key residues for catalysis, Q36, C42, H174 and N204, are highlighted in bold.
Source: Prepared by the author.

Text C2: Betweenness centrality analysis

Our results show that the centrality profiles of the analyzed systems are perturbed by the presence of the inhibitor (Fig. C15). For free FP-2 and the FP-2:Cpd66 complex, the residues undergoing the largest variations are: M2, K37, S41, S47, I48, I93, P105, S108, Y123, I125, K126, I148, A175 and F236, which decreased their centralities upon the inhibitor binding; and Q1, G29, V33, K34, G40, F45, S46, S50, E67, E69, L140, F142, K203 and W206, which possess higher values in the FP-2:Cpd66 complex (Fig. C15-A). Some of the previous residues, e.g., G29, V33, K34, F45, S46, E67 and K203, lie at the allosteric pocket (Fig. 4.2-B) and, interestingly, all belong to the group of residues whose centralities augmented appreciably when the inhibitor is bound. On the other hand, most of the residues displaying large centrality variations are not in direct contact with Cpd66. In fact, the perturbations caused by the inhibitor are not limited to the allosteric pocket vicinity, but are spread throughout the protein structure (compare the distribution of per-residue centralities in free FP-2 and the FP-2:Cpd66 complex, Fig. C15-B). Remarkably, some of the residues undergoing large centrality variations are located within the active site, e.g., G40, S41, A175 and W206, thus providing further clues about the impact of Cpd66 on this region of the protein.

Figure C15. Betweenness centralities for the FP-2 residues in four different states. A) Graphs show the per-residue centrality values for the indicated systems. Residues displaying large centrality variations between the compared systems are labeled. **B)** Distribution of per-residue centralities on the FP-2 structure in the different systems (see labels). The Cα atoms of FP-2 are depicted as spheres and colored according to their normalized centralities (see color gradient bars). Cpd66 (not shown) binds along the FP-2 central α-helix.

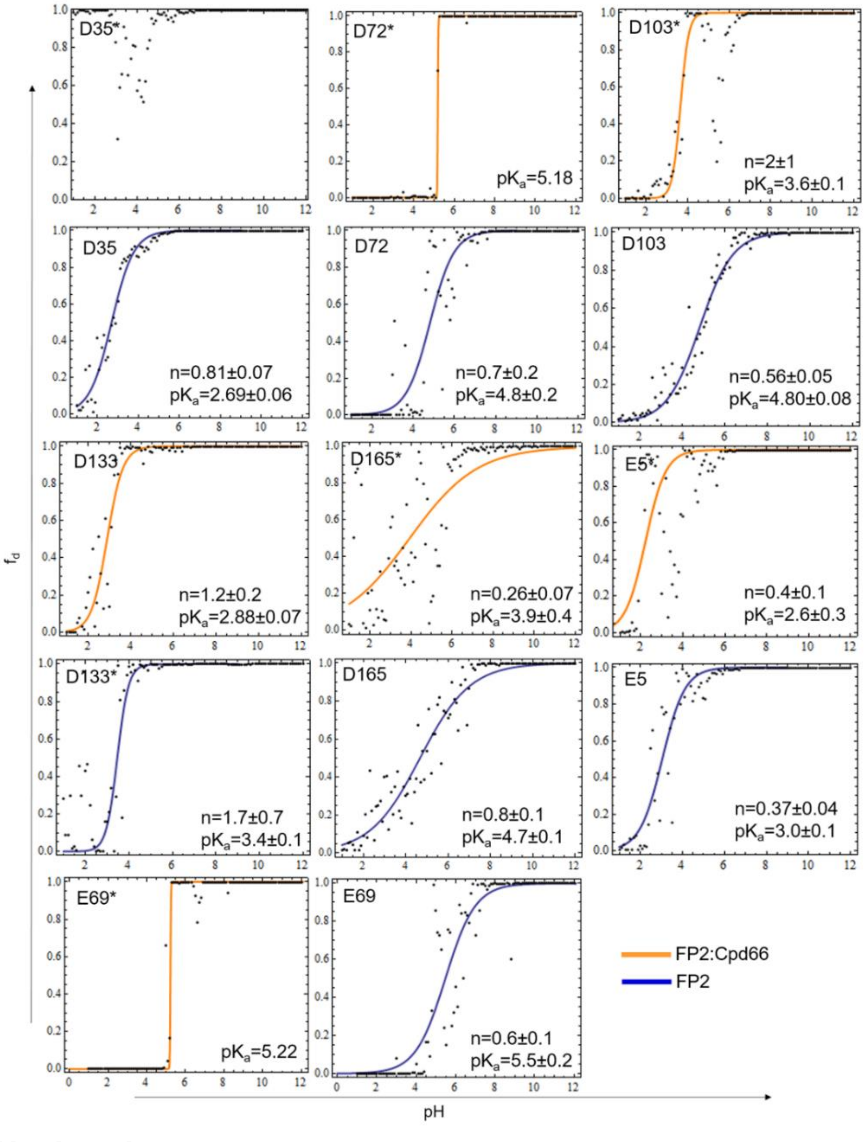


Source: Prepared by the author.

The comparison of per-residue centralities for the FP-2:peptide and FP-2:peptide:Cpd66 complexes leads to conclusions that are roughly similar to those obtained in the absence of the peptide. However, it is worth saying that the centrality differences between the peptide-bound complexes become more apparent (compare both graphs in Fig. C15-A). The following residues decreased significantly their centralities upon Cpd66 binding to the FP-2:peptide complex: C39,

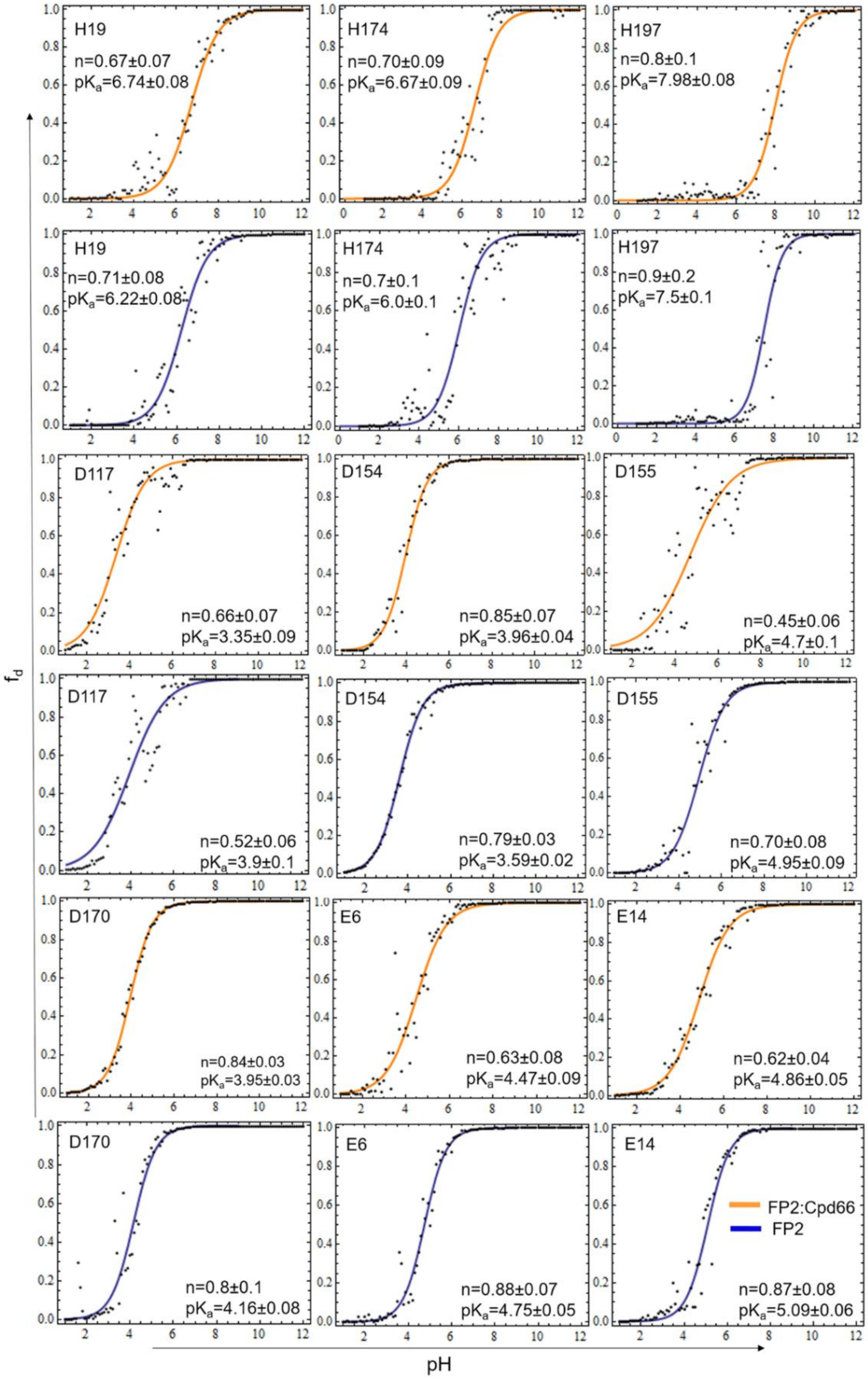
G40, S41, C42, S50, C80, G97, D109, S108, Y123, G124, I148, A175, L178, F236, I237 and P238. On the contrary, other residues, comprising W24, G29, V33, K34, K37, F45, S46, E67, E69, G83 and N204, notably increased their centralities in the presence of the inhibitor (Fig. C15-A). Many of these residues display a similar behavior in FP-2 and the FP-2:Cpd66 complex.

Figure C16. Titration curves that were reliably estimated in only one system. In every graph, the fraction of deprotonation (f_d) was plotted versus the pH. Calculated pK_a's and Hill's cooperativity indices (n) are indicated. 95% confidence intervals for the estimated values were calculated by error bootstrapping. Residues marked with asterisks were not accurately titrated.



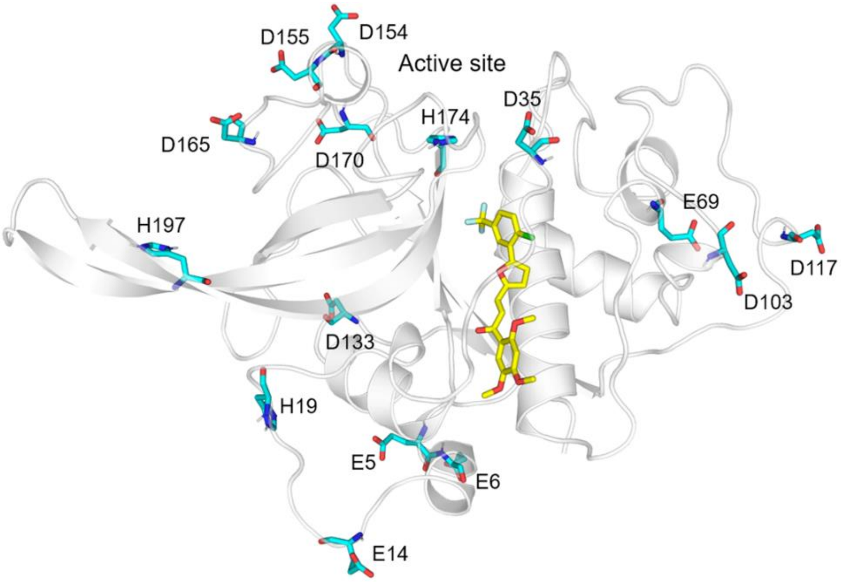
Source: Prepared by the author.

Figure C17. Titration curves displaying the largest shifts upon Cpd66 binding to FP-2. In every graph, the fraction of deprotonation (*f_d*) was plotted versus the pH. Calculated *pK_a*'s and cooperativity indices (*n*) are indicated. 95% confidence intervals for the estimated values were calculated by error bootstrapping.



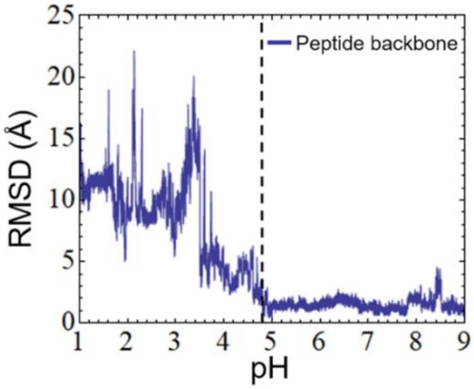
Source: Prepared by the author.

Figure C18. Residues displaying significant *pK_a* shifts upon the binding of Cpd66 to FP-2. The residues undergoing such variations are depicted as cyan sticks. The inhibitor is shown in yellow and the protein backbone structure, as white cartoon.



Source: Prepared by the author.

Figure C19. RMSD values for the peptide at different pHs. The RMSDs were calculated with respect to the peptide starting conformation in the FP-2:peptide complex at pH=5.5 after fitting all frames to FP-2 initial structure. As the titration was performed by changing the pH in discrete steps of 0.1, the graph was built by assigning evenly distributed pH values between *i* and *i+0.1* to the frames sequentially collected at pH *i.* At pH<4.8 (see dashed line) the peptide loses affinity for the enzyme's active site.



Source: Prepared by the author.

References

(1) Bertoldo, J. B.; Chiaradia-Delatorre, L. D.; Mascarello, A.; Leal, P. C.; Cordeiro, M. N.; Nunes, R. J.; Sarduy, E. S.; Rosenthal, P. J.; Terenzi, H., Synthetic compounds from an in house library as inhibitors of falcipain-2 from Plasmodium falciparum. *J Enzyme Inhib Med Chem* **2015**, 30, 299-307.

(2) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C., Essential Dynamics of Proteins *Proteins* **1993**, 17, 412-425.

APPENDIX D – Additional information to the role of water bridges in the affinity and selectivity for FP-2 of nitriles bearing pyridine substituents at P2

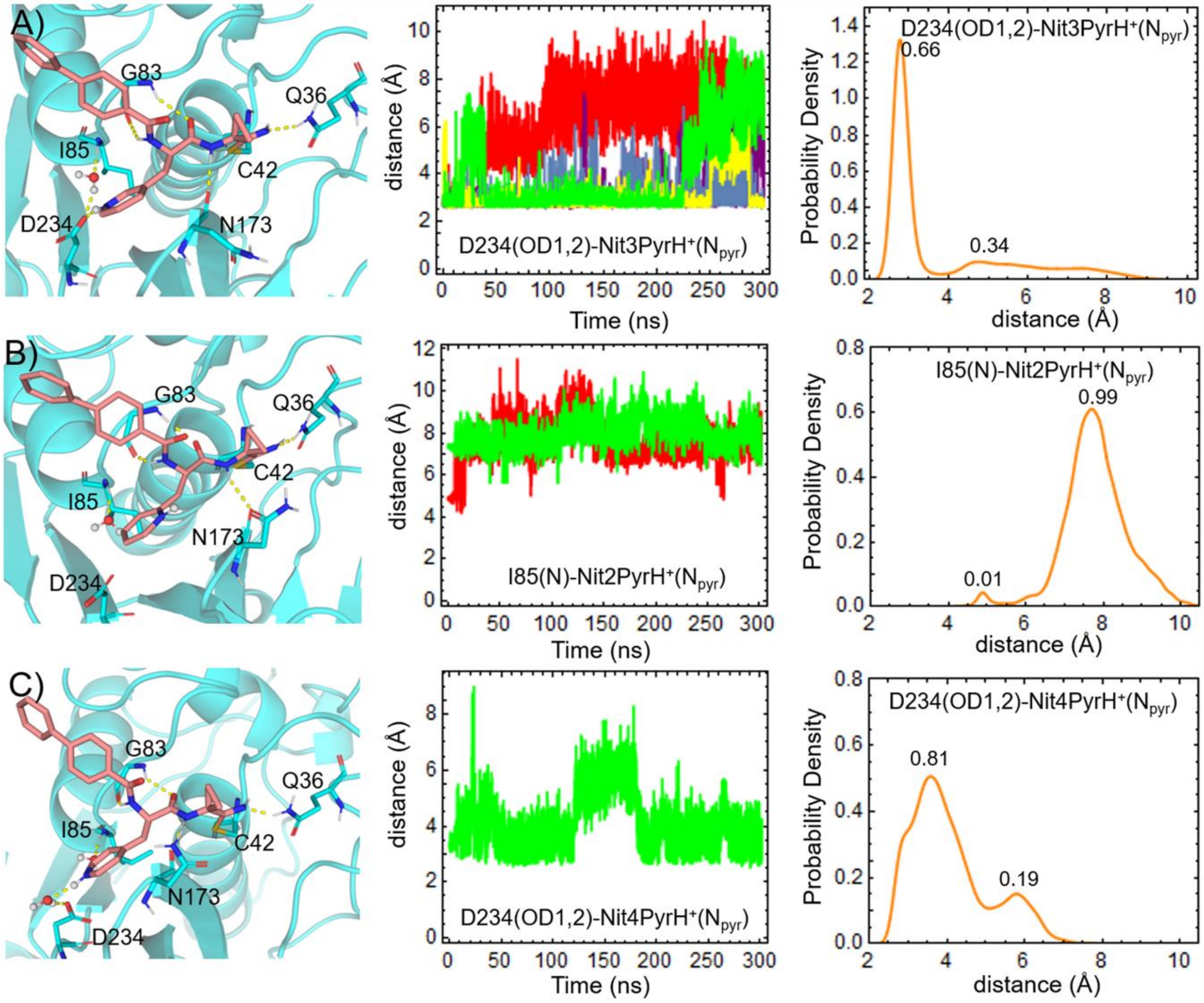
Text D1. Unfavorable interaction of the protonated studied compounds with FP-2

The pK_a of Pyr at 25 °C is 5.23, and increases to 6.00, 5.70 and 5.99 upon methylation at positions 2, 3 and 4, respectively [1]. Assuming that the Pyr rings of Nit2Pyr, Nit3Pyr and Nit4Pyr have pK_a values similar to the analog methylpyridines, roughly 60 to 75% of the molecules will be protonated at pH=5.5, the typical pH of lysosomes.[2] Under these conditions, protonated and deprotonated forms of the compounds coexist in considerable amounts in solution. Hence, it cannot be straightforwardly inferred that, upon binding to FP-2, the protonation state of the studied inhibitors will be that of their most abundant free species in solution.

Firstly, we analyze the most likely protonation state of Nit3Pyr in complex with FP-2. When protonated, the N_{pyr} atom of this compound lies in a position that, in principle, allows the formation of H-bonds with either OD1 or OD2 atoms of D234 (Fig. D1-A). In fact, this interaction was observed in nearly 66% of the frames collected throughout five 300 ns replicate MD simulations of the FP-2:Nit3PyrH$^+$ complex (Fig. D1-A). However, we noted that in some replicas, the N_{pyr} and OD1,2 atoms remained distant from each other during long periods of time, thus indicating that the H-bond between them might not as stable as first thought. Subsequently, it was confirmed through TI $\Delta\Delta G$ calculations that the FP-2:Nit3PyrH$^+$ complex is ~6 kcal/mol less stable than FP-2:Nit3Pyr, regardless of the ionic strength (Table D1). Of note, free energy calculations at two different NaCl concentrations were conducted as it may affect the affinity of some charged ligands for the target proteins.[3] The large $\Delta\Delta G$ associated to the binding of Nit3PyrH$^+$ and Nit3Pyr to FP-2 indicates that the occurrence of complexes involving the charged ligand ($E \bullet H^+$ and $E\text{-}H^+$) is negligible, as assumed in Appendix A, eq. A11.

On the other hand, the N_{pyr} atoms of Nit2PyrH$^+$ and Nit4PyrH$^+$ do not lie in optimal positions to establish stable direct H-bonds with D234(OD1,OD2) (Figs. D1-B and C). The protonated N_{pyr} atom of Nit2PyrH$^+$ remains exposed to the solvent due to the lack of polar interactions within the S2 subsite during two replicate MD simulations started from structures with different orientations of the Pyr ring. Remarkably, the N_{pyr} atom of Nit4PyrH$^+$ was found to form either a direct H-bond or a water bridge with D234(OD1,2) during a single 300 ns MD simulation.

Figure D1. Interaction of Nit3PyrH$^+$, Nit2PyrH$^+$ and Nit4PyrH$^+$ with FP-2. A) Interface of the FP-2:Nit3PyrH$^+$ complex and D234(OD1,2)-N$_{pyr}$ distance time profiles and distribution obtained from five 300 ns replicate MD simulations. As OD1 and OD2 of D234 are chemically equivalent, the minimal distance between any of them and N$_{pyr}$ for every trajectory frame is shown in the graphs. **B)** Interface of the FP-2:Nit2PyrH$^+$ complex and I85(N)-N$_{pyr}$ distance time profile and distribution obtained from two replicate MD simulations started using different initial orientations of the Pyr ring. **C)** Interface of the FP-2:Nit4PyrH$^+$ complex and D234(OD1,2)-N$_{pyr}$ distance time profile and distribution calculated as in A). Only a single 300 ns MD simulation of this complex was conducted. In all cases, direct H-bonds and water bridges are represented using yellow dashed lines. Water molecules lying close to I85(N) are shown as sphere and sticks. FP-2 residues (cyan sticks) forming H-bonds with the ligand (salmon stick) are labeled. The fraction of frames contained within each peak of the probability density distributions is shown on top of the corresponding maximum.



Source: Prepared by the author.

However, the stability of the direct H-bond (occupancy of 12%) was appreciably lower than that of Nit3PyrH$^+$ (direct H-bond occupancy of 66%). The water bridge was also less prevalent than one formed by the deprotonated compound (occupancies of 36 and 55% for Nit4PyrH$^+$ and Nit4Pyr, respectively, see Table 5.1). Moreover, it was observed that the neutral compounds are

Table D1: Relative free energies between FP-2:Nit3PyrH$^+$ and FP-2:Nit3Pyr complexes[a]

[NaCl] (M)	$\Delta\Delta G_{chg}$[b] (kcal/mol)	$\Delta\Delta G_{vdw}$[b] (kcal/mol)	$\Delta\Delta G$[b] (kcal/mol)	$\lvert\Delta\Delta G_{fwd}-\Delta\Delta G_{bwd}\rvert$[b,c] (kcal/mol)
0	5.9 (0.4)	0.05 (0.01)	6.0 (0.4)	0.4
0.15	6.2 (0.4)	0.05 (0.1)	6.2 (0.4)	0.3

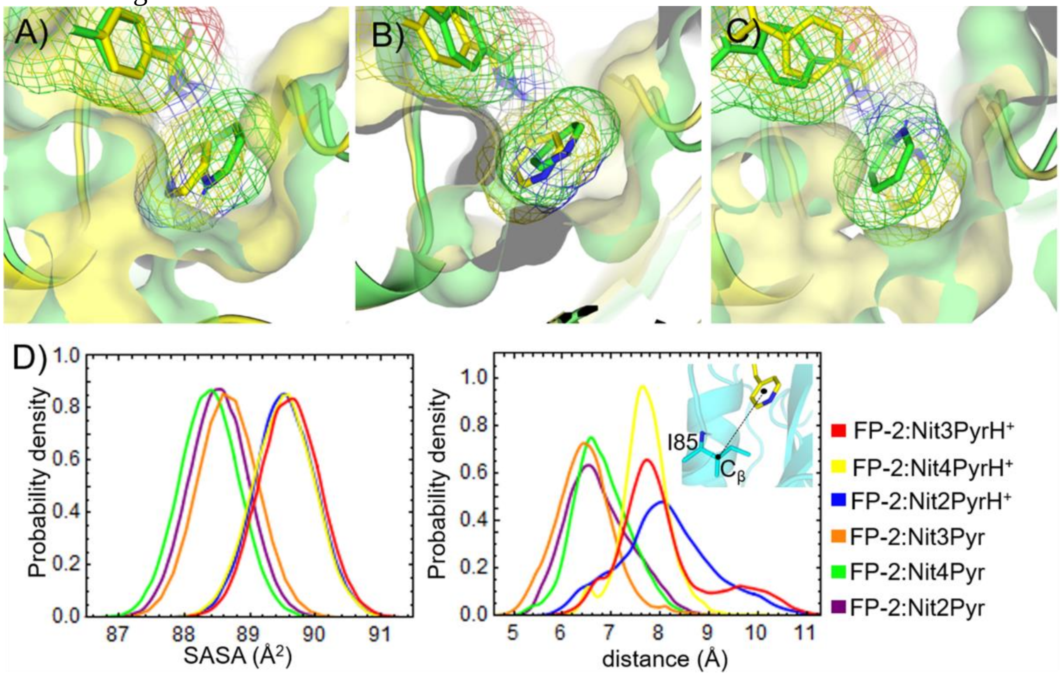
[a] The values were determined using the dual system/single box approach at two salt concentrations.

[b] $\Delta\Delta G_X=\Delta G_X(\text{Nit3PyrH}^+)-\Delta G_X(\text{Nit3Pyr})$, where *X* stands for *chg* (charge-related transformations), *vdw* (van der Waals transformations), *fwd* (forward transformation) or *bwd* (backward transformation). $\Delta\Delta G=\Delta\Delta G_{chg}+\Delta\Delta G_{vdw}$

[c] Hysteresis.

Source: Prepared by the author.

Figure D2. Insertion of P2-Pyr moieties at different protonation states into the S2 subsite of FP-2. A) FP-2:Nit3Pyr (yellow) and FP-2:Nit3PyrH$^+$ (green) complexes. **B)** FP-2:Nit4Pyr (yellow) and FP-2:Nit4PyrH$^+$ (green) complexes. **C)** FP-2:Nit2Pyr (yellow) and FP-2:Nit2PyrH$^+$ (green) complexes. In all cases, the ligands are depicted as sticks surrounded by a mesh. The proteins are shown in cartoon and surface representations. Central structures calculated from the respective MD simulations were selected for structural representation after structural alignment. **D)** Distributions of SASA (left graph) and distance (right graph) values calculated for the MD trajectories of the complexes indicated in the legend on the right. SASA values were calculated for the heavy atoms of the pyridine ring of each ligand. The distance between I85(Cβ) at the base of the S2 subsite and the center of mass of the pyridine ring of each compound was also taken as measurement of ring insertion.



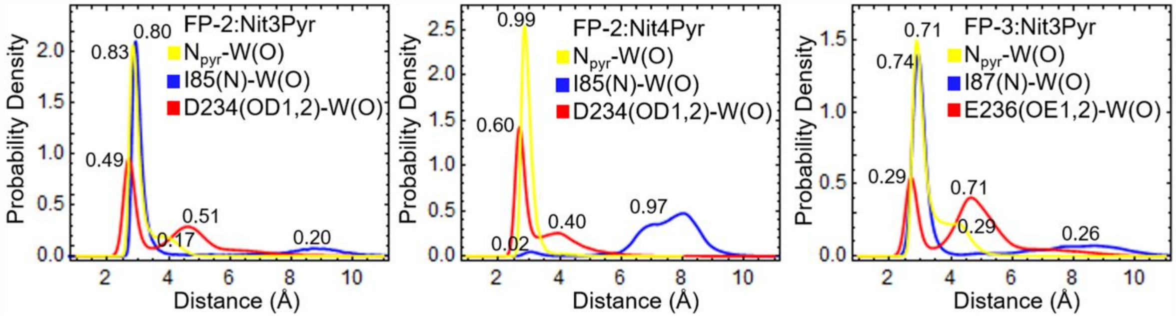
Source: Prepared by the author.

able to insert the P2-Pyr rings deeper into the S2 subsite than the respective protonated forms (Fig. D2). This indicates, in turn, that it is more favorable for the neutral forms to undergo desolvation and to accommodate into the hydrophobic S2 pocket than it is for protonated Pyr rings bearing a delocalized positive charge. On the basis of the previous results and on the large unfavorable $\Delta\Delta G$

value for Nit3PyrH$^+$/Nit3Pyr (Table D1), the protonated isomer with the highest capacity to form H-bonds with D234(OD1,2), we concluded that Nit2Pyr and Nit4Pyr also occur preferentially in deprotonated states when bound to FP-2. Hence, no time consuming $\Delta\Delta G$ calculations were deemed necessary for these two complexes.

Finally, note that the substituted 3Pyr moieties of other compounds studied in this work, e.g., Nit6Cl3Pyr and Nit6OMe3Pyr, are less basic than that of Nit3Pyr, with pK_a's of roughly 1.0 and 3.8, respectively. These values were estimated by taking as reference the pK_a's of 2-chloropyridine (0.49) and 2-methoxypyridine (3.28),[1] and by adding to each the pK_a shift (0.5) caused by a methyl group at position 4. This shift was obtained by subtracting the pK_a's of 3-methylpyridine (5.70) and pyridine (5.23), which mimics the effect of methylene linker of the P2 moiety (Fig. 1.5). Due to the smaller pK_a values of Nit6Cl3Pyr and Nit6OMePyr, it is even less likely for their protonated forms to bind FP-2 than it is for Nit3PyrH$^+$. Therefore, they were taken as neutral to conduct the MD simulations of their respective complexes.

Figure D3. Distance distributions involving the water molecules closest to N_{pyr} in three complexes. The water molecules closest to N_{pyr} in each complex were tracked during all the corresponding MD simulations. Distance profiles for the indicated atoms were calculated and distributions were subsequently obtained. The minimal distance of N_{pyr} with respect to OD1,2 and OE1,2 is plotted. The relative abundance of each peak is shown close to each distribution maximum.

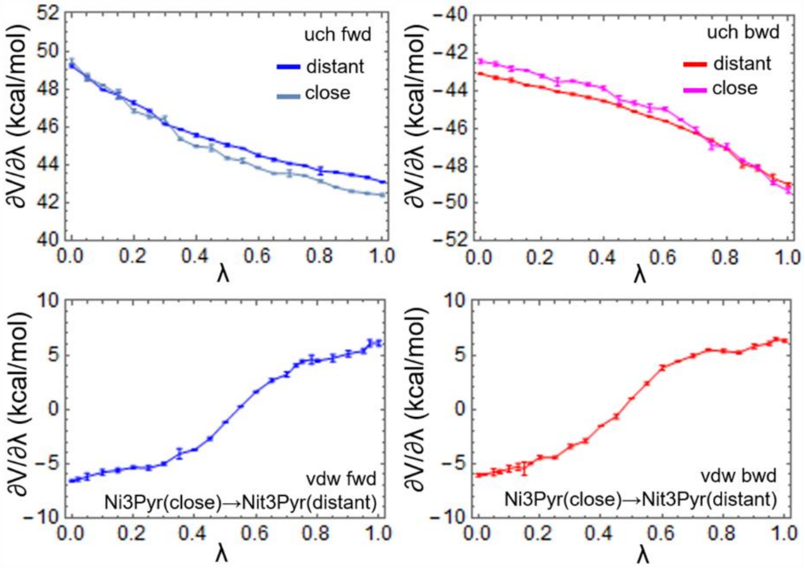


Source: Prepared by the author.

Text D2. Relative abundance of close and distant orientations of Nit3Pyr in complex with hCatL obtained form TI $\Delta\Delta G$ calculations.

The distance profiles of the hCatL:Nit3Pyr complex show that only two transitions, both in the close-to-distant direction, were observed in all the replicate MD simulations (Fig. 5.3). This indicates that the interconversion of both orientations is a slow process, probably due to a tight fit of the 3Pyr ring into the S2 subsite of hCatL. The small number of transitions observed in the MD simulations does not allow to determine the relative abundance of peaks A and B (Fig. 5.3). The more natural solution, i.e., increasing the simulation time, seems to be impractical given the very slow interconversion rate observed for this system during its ten replicate 300 ns trajectories, totalizing 3 μs of simulation time. Therefore, we decided to employ TI $\Delta\Delta G$ calculations to calculate the relative stability of each orientation, which, in turns, gives us the required information to combine peaks A and B in the correct proportion.

Figure D4. $\partial V/\partial\lambda$ vs λ plots for the alchemical transformation of close and distant orientations of Nit3Pyr in complex with hCatL. Error bars are added to each point. The abbreviation uch and vdw stand for the uncharging and van der Waals steps, respectively. Forward (fwd) and backward (bwd) transformations are labeled accordingly. The legend for each color is shown in the figure.



Source: Prepared by the author.

The $\Delta\Delta G$ calculation of pose interconversion can be conducted through the thermodynamic cycle shown Fig. 2.3-B after some considerations. Firstly, note that, since the ligand is not chemically modified during the alchemical transformations, the free energy associated to the left vertical leg of the cycle is zero by definition. Consequently, the free energy of right vertical leg can yield the desired $\Delta\Delta G$ value if NitX_1 and NitX_2 correspond, say, to the close and the distant orientations of the same ligand, respectively, during all the λ MD simulations. It is important to highlight that, during the alchemical transformations, one must ensure that no pose interconversion events occur, which can be achieved by employing harmonic restraints. Nonetheless, the latter were not necessary during the 10 ns λ MD simulations conducted during the uncharging steps of the 3Pyr ring in both orientations, as the slow transition rate observed in the hCatL:Nit3Pyr complex (Fig. 5.3) arises from the volume of the P2 substituent, which is kept constant during charge-related transformation. In fact, we checked that no transitions occurred by analyzing the distance profile for every λ MD simulation. On the other hand, restraints were required to preserve the desired orientation during the van der Waals alchemical transformations. Distance restraints similar to those applied to the FP-2:Nit3Pyr complex during the calculation of $\Delta\Delta G_{calc}^{wb-}$ values in Table 5.3 were employed, the reference atoms being, in this case, M70(N) and $\mathrm{N_{pyr}}$ of hCatL and Nit3Pyr, respectively. The restraint minimum was set to 6.8 Å, since it is a suitable threshold to define close (<6.8 Å) and distant (>6.8 Å) orientations, according to the distance distributions for hCatL:Nit3Pyr shown in Fig. 5.3.

Table D2: Relative free energies between the distant and close orientations of Nit3Pyr in complex with hCatL

$\Delta G_{uch(A)}^{bound}$ [a] (kcal/mol)	$\Delta G_{uch(B)}^{bound}$ [a] (kcal/mol)	$\Delta G_{vdw(A\to B)}^{bound}$ [a] (kcal/mol)	$\Delta\Delta G_{calc}$ (kcal/mol)	$\Delta\Delta G_{hys}$ [b] (kcal/mol)	p_A [c]	p_B [c]
45.00 (0.05)	45.41 (0.03)	-0.60 (0.08)	-1.0 (0.1)	0.2	0.16 (0.02)	0.84 (0.02)

[a] Values of the uncharging and van der Waals free energies for Nit3Pyr close (A) and distant (B) orientations, which were substituted into Eqs. D12 and D13 to calculate $\Delta\Delta G_{calc}$ and p_A, respectively. The close (distant) orientation were preserved during van der Waals transformations by applying distance restraints when the M70(N)-$\mathrm{N_{py}}$ distance was >6.8 Å (<6.8 Å). A harmonic constant of 10 $\mathrm{kcal\cdot mol^{-1}\cdot Å^{-2}}$ was set. No restraints were used during the uncharging process.

[b] Hysteresis

[c] Probabilities of finding the 3Pyr ring in close (A) or distant (B) orientations in the hCatL:Nit3Pyr complex.

Source: Prepared by the author.

The $\partial V/\partial\lambda$ vs λ plots for the uncharging steps conducted three times in both directions for each orientation are shown in Fig. D4. The $\Delta\Delta G$ value between the close and the distant orientations of the hCatL:Nit3Pyr complex was calculated as follows:

$$\Delta\Delta G_{calc} = \Delta G_{hCatL:Nit3Pyr(B)}^{bind} - \Delta G_{hCatL:Nit3Pyr(A)}^{bind} = \Delta\Delta G_{Nit3Pyr(A\to B)}^{bound}$$

$$\Delta\Delta G_{calc} = \Delta G_{uch(A)}^{bound} - \Delta G_{uch(B)}^{bound} + \Delta\Delta G_{vdw(A\to B)}^{bound} \tag{D12}$$

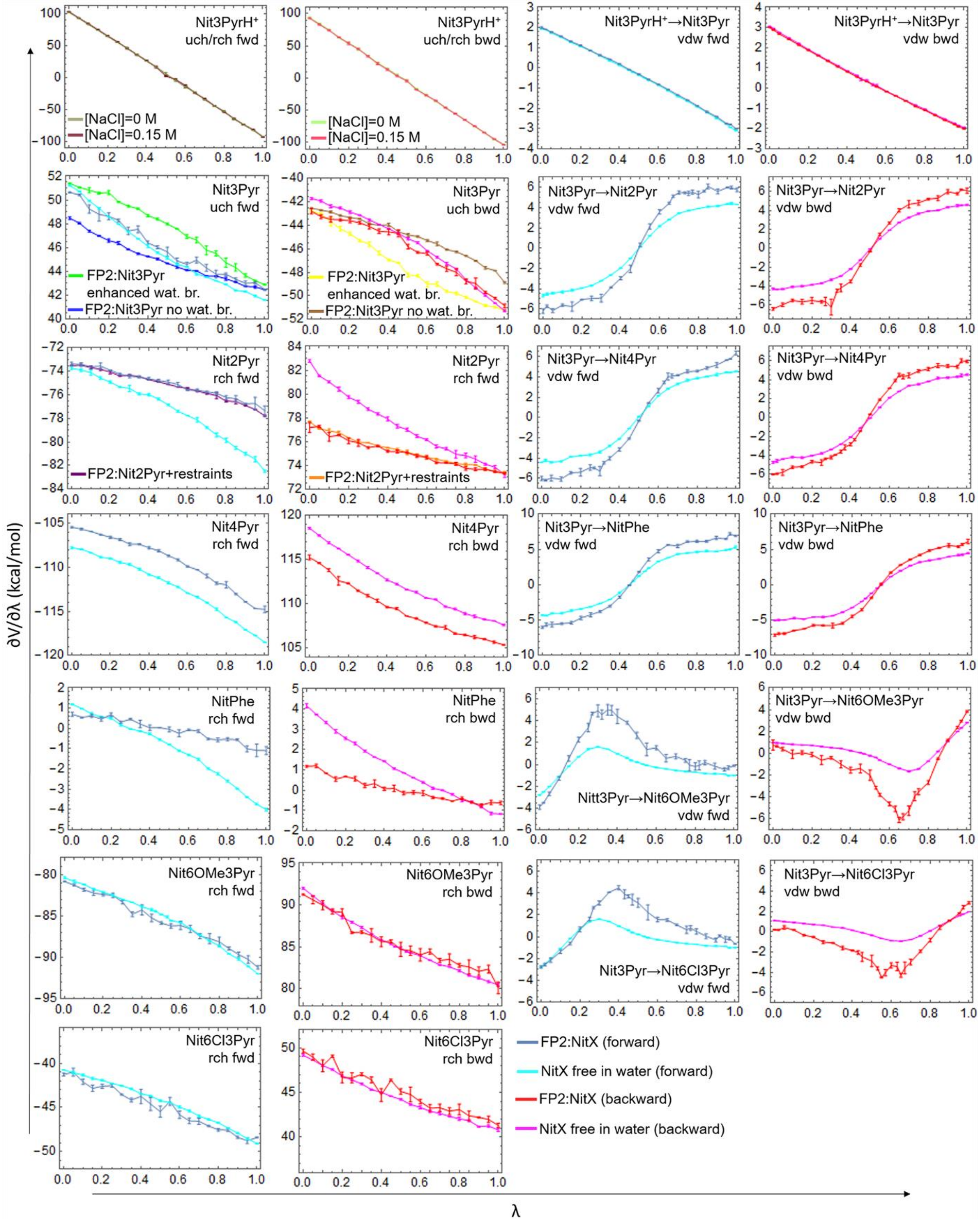
where A and B correspond to the close and distant orientations from peaks A and B, respectively (Fig. 5.3). Table D2 summarized the results of these free energy calculations. In addition, we show in the table the probabilities associated to each orientation, which were obtained by calculating the Boltzmann weights of each state as follows:

$$\frac{p_B}{p_A} = \frac{1-p_A}{p_A} = e^{-\frac{\Delta\Delta G}{RT}}$$

$$\therefore\ p_A = \frac{e^{-\frac{\Delta\Delta G}{RT}}}{1+e^{\frac{-\Delta\Delta G}{RT}}} \tag{D13}$$

Notice that we set $p_A + p_B = 1$, which assumes that the whole phase space is sampled from MD simulations carried out corresponding to the close and distant orientations. Finally, the histograms of each orientation were combined to yield the average distribution, so that the fraction of frames in each peak matched their respective probabilities, p_A and p_B. The latter was achieved by evenly eliminating the exceeding frames from peak A. It is worth noting that the distant orientation is more favorable than the close one, which explains why the only two transitions observed during the MD simulations were in the close-to-distant direction (Fig. 5.3).

Figure D5. $\partial V/\partial\lambda$ vs λ plots for all the alchemical transformation conducted during the $\Delta\Delta G$ calculations. Error bars are added to each point. The abbreviations uch, vdw and rch stand for the uncharging, van der Waals and recharging steps, respectively. For the charged ligand, the uncharging and recharging steps were carried out simultaneously (uch/rch) in the same simulation box. Forward (fwd) and backward transformations are labelled accordingly. The legend for each color is shown in the figure.



Source: Prepared by the author.

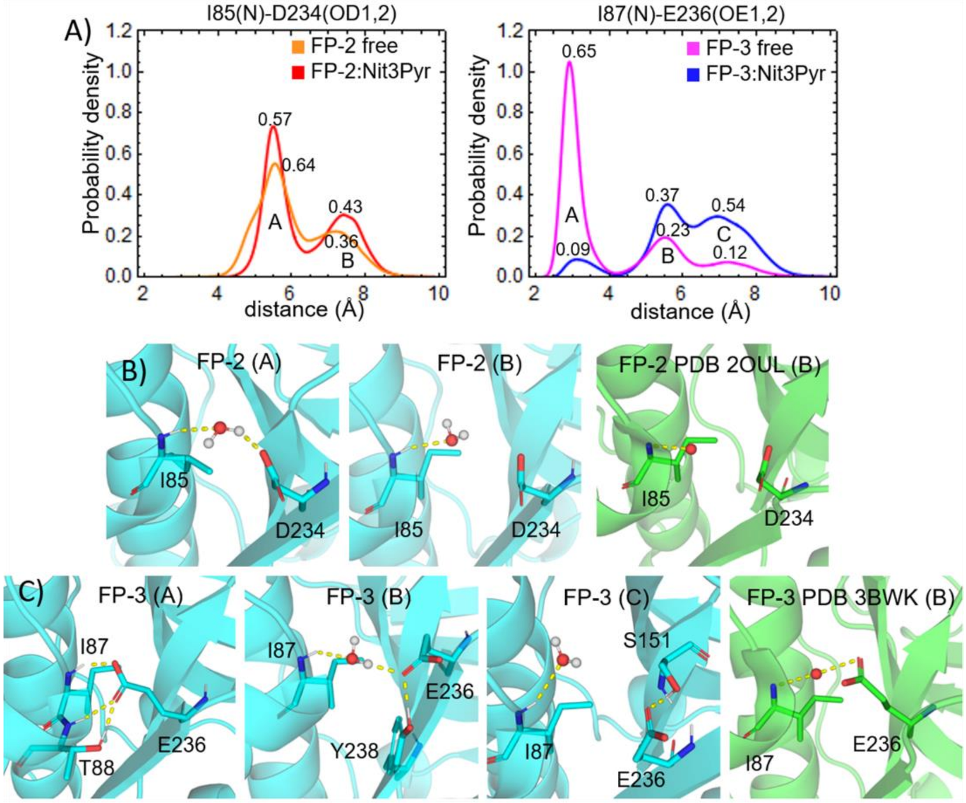
Text D3. Analysis of H-bonds and water bridges between residues I85 and D234 in free FP-2 and in FP-2:Nit3Pyr, and between residues I87 and E236 in free FP-3 and in FP-3:Nit3Pyr

We conducted H-bond and water bridge analyses for free FP-2 and FP-3 using four replicate 300 ns MD simulations of each enzyme that had been conducted in previous work.[4] The distributions of I85(N)-D234(OD1,2) distance in free FP-2 show the occurrence of two peaks, A and B, with relative abundances of 0.64 and 0.36, respectively (Fig. D6-A). 75% of the frames pertaining to peak A, which comprises I85(N)-D234(OD1,2) distances within 4 to 6.3 Å, can form I85(N)-D234(OD1,2) water bridges, which occur with an overall fractional occupancy of 0.44 (Figs. D6-A and B, and Table D3). Peak B corresponds to structures in which D234(OD1,2) and I85(N) are too distant from each other to interact (> 6.3 Å, Figs. D6-A and B). Remarkably, the crystal structure of FP-2 (PDB: 2OUL) lies within this peak (Fig. D6-B). Upon Nit3Pyr binding, the overall shape of the distribution does not change appreciably (Fig. D6-A). The same peaks, A and B, are still observed in the complex, the former peak being again more abundant (fraction of 0.57, Fig. D6-A). The fractional occupancy of the I85(N)-D234(OD1,2) water bridge in the FP-2:Nit3Pyr complex is 0.36, slightly smaller than that of free FP-2 (Table D3).

The I87(N)-E236(OE1,OE2) distance distribution in free FP-3 is significantly different from that of FP-2 (Fig. D6-A). In this case, three peaks, A, B and C, were detected, and the largest one (peak A) corresponds to structures in which the E236(OE1,2) and I85(N) atoms are close to each other (2 to 4 Å range) and can form direct H-bonds (fractional occupancy of 0.64, Figs. D6-A and C, and Table D3). The conformation of E236 side-chain observed in the central structure of peak A is further stabilized by the formation of additional H-bonds with T88(N) and T88(OG1) (Fig. D6-C). Peak B, occurring with a relative abundance of 0.37, comprises structures able to form the I87(N)-E236(OE1,OE2) water bridge with a fractional occupancy of 0.26, which is lower than that of FP-2 (Table D3). On the other hand, the conformation adopted by E236 side-chain in the less representative peak (C) precludes the formation of H-bonds and water bridges with I87 (Fig. D6-C). The FP-3 crystal structure occurs in a conformation typical of peak B, as it involves the formation of the analyzed water bridge (Fig. D6-C). The Nit3Pyr binding dramatically changes the I87(N)-E236(OE1,OE2) distance distribution (Fig. D6-A). Peak A becomes almost nonexistent

(9% of the frames), as both residues must lie farther from each other to let the 3Pyr moiety fit within the S2 subsite. Peak B is also less abundant under these circumstances (Fig. D6-A), but the I87(N)-E236(OE1,2) water bridge fractional occupancy remains relatively unchanged with respect to that of free FP-3 (Table D3). In the presence of the ligand, the largest peak (C), relative abundance of 0.54, corresponds to structures in which I87(N) and E236(OE1,2) are distant from each other (> 6.3 Å) and, consequently, do not interact (Fig. D6-A and C).

Figure D6. Distance distributions and central structures of free FP-2 and FP-3 corresponding to every distribution peak. A) Distance distributions for the indicated atoms and systems. The relative abundance of each peak is shown close to each distribution maximum **B)** Central structures of free FP-2, colored in cyan, corresponding to the peaks shown in parentheses. The crystal structure of FP-2, PDB 2OUL, is colored in green and the peak to which it belongs is shown in parentheses. **C)** Same as B) but for FP-3. The crystal structure of the enzyme is colored in green. In all cases, residues and water molecules involved in the formation of H-bonds and/or water bridges are depicted as stick and spheres, respectively, and residues were labelled.



Source: Prepared by the author.

The analyses performed here indicate that I85 and D234 are more prone to form water bridges in free FP-2 than the equivalent residues, I87 and E236, in free FP-3. This differential pattern of water bridge formation persists upon Nit3Pyr binding to each enzyme, thus highlighting that the greater relevance of these interactions in the FP-2:Nit3Pyr complex stems from an inherent property of the interacting residues in FP-2.

Table D3. Fractional occupancies of water bridges and H-bonds involving I85 and D234 of free FP-2 and FP-2:Nit3Pyr, and I87 and E236 of FP-3 and FP-3:Nit3Pyr

System	Direct H-bond[a]	Wat. Bridge[b]
FP-2	0	0.44 (0.05)[c]
FP-2:Nit3Pyr	0	0.36 (0.04)
FP-3	0.64 (0.09)	0.26 (0.06)
FP-3:Nit3Pyr	0.03 (0.01)	0.23 (0.02)

[a,b] I85(N)-D234(OD1,2) in FP-2 systems and I87(N)-E236(OE1,2) in the FP-3 ones.

[c] Mean value and standard error of the mean in parentheses.

Source: Prepared by the author.

References

(1) Haynes, W. M., *CRC handbook of chemistry and physics*. CRC press: 2014.

(2) Johnson, D. E.; Ostrowski, P.; Jaumouille, V.; Grinstein, S., The Position of Lysosomes within the Cell Determines their Luminal pH. *J Cell Biol* **2016**, 212, 677-692.

(3) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L., Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J Chem Theory Comput* **2018**, 14, 6346-6358.

(4) Hernandez Gonzalez, J. E.; Hernandez Alvarez, L.; Pascutti, P. G.; Leite, V. B. P., Prediction of Noncompetitive Inhibitor Binding Mode Reveals Promising Site for Allosteric Modulation of Falcipain-2. *J Phys Chem B* **2019**, 123, 7327-7342.