

**UNIVERSITY OF NAPLES FEDERICO II
SCHOOL OF MEDICINE
DEPARTMENT OF PHARMACY**



PhD programme in Pharmaceutical Sciences

Identification of new bioactive molecules by computational methods

**PhD student
OLEH TKACHUK**

**Tutor
Prof. CATERINA FATTORUSSO**

**Coordinator
Prof. ROSARIA MELI**

**Co-Tutor
Prof. MARCO PERSICO**

XXXVIII CYCLE (a.y. 2024/2025)

List of abbreviations

AFM	Atomic Force Microscopy
BA	Binding Approach
CAS	Computational Alanine Scanning
CCR	Catalytic Core Regulator
CDKN2A	Cyclin-Dependent Kinase Inhibitor 2A
C-PCM	Conductor-like Polarizable Continuum Model
CQ	Chloroquine
CRBN	Cereblon E3 Ligase
DFT	Density Functional Theory
DM1	7-Methyl-2-phenylimidazo[1,2-b]pyridazin-3-carboxylic acid
DM2	6-Methoxy-2-phenylimidazo[1,2-b]pyridazine-3-carboxylic acid
ExoN	Exoribonuclease
FRET	Fluorescence Resonance Energy Transfer
H2T4	Meso-tetrakis(4-N-methylpyridyl)-porphyrin
H2TPPS	Meso-tetrakis(4-sulphonatophenyl)-porphyrin
HBS	High Affinity Binding Site
HSP60	Heat Shock Protein 60
IDE	Insulin-Degrading Enzyme
LBS	Low Affinity Binding Site
LUMO	Lowest Unoccupied Molecular Orbital
MIF	Molecular Interaction Field
MM	Molecular Mechanics
Mpro (3CLpro)	Main Protease / Chymotrypsin-like Cysteine Protease
N7-MTase	N7-Methyltransferase
NBO	Natural Bond Orbital
NP	Natural Product
NSPs	Nonstructural Proteins
PfGAPDH	Plasmodium falciparum Glyceraldehyde 3-Phosphate Dehydrogenase
PH	Pleckstrin Homology Domain
PICS	Proteomic Identification of Cleavage-site Specificity
PIDs	Protein Interaction Domains
PIPs	Proteasome Interacting Proteins
PLpro	Papain-like Protease

PDB	Protein Data Bank
pLDH	Parasite Lactate Dehydrogenase
POI	Protein of Interest
PPIs	Protein–Protein Interactions
PROTAC	Proteolysis Targeting Chimera
PS	Phosphatidylserine
PTZ	Pentylene-tetrazol
RGS	Regulators of G Protein Signaling
RMSD	Root-Mean-Square Distance
RMSF	Root-Mean-Square Fluctuation
ROS	Reactive Oxygen Species
RP	Regulatory Particles
RTC	Replication–Transcription Complex
SAM	S-Adenosyl-L-Methionine
SAH	S-Adenosyl-L-Homocysteine
SAR	Structure–Activity Relationship
SH2/SH3	Src Homology Modules 2 and 3
SMD	Steered Molecular Dynamics
SLiMs	Short Linear Motifs
SMI	Small Molecule Inhibitor
SWDs	Spikes-and-Waves Discharges
t-BuOOH	Tert-Butyl Hydroperoxide
THiAs	1,2,4-Thiadiazolidin-3,5-diones
TMPC	5,10,15-Tri(N-methyl-4-pyridyl)corrole
TPS	Tetrasubstituted Pyrrole
TrisT4	5-(Phenyl)-10,15,20-(Tri-N-methyl-4-pyridyl)-porphyrin
UPP	Ubiquitin–Proteasome Pathway
VHL	Von Hippel–Lindau E3 Ligase
WT	Wild Type

Abstract

This thesis explores the “language” of biomolecules, interpreting their interactions as a grammatical system in which protein domains, short linear motifs, and hot spots represent the fundamental units of communication with functional ligands. The structural and conformational properties of bioactive small molecules—derived either from rational design or from natural products used as “privileged scaffolds”—can mimic those of endogenous ligands, thereby speaking the same “language”.

Building on this concept, various computational approaches were employed to investigate key protein interaction sites and their binding with both endogenous and exogenous functional ligands. The ultimate goal of these studies is to develop computational models as tools for identifying new lead compounds or optimizing existing ones.

The chapters of this thesis describe:

- i) the role of stereochemistry in the activity of nature-inspired isooxazoline derivatives;
- ii) the design of tetrasubstituted pyrrole derivatives that mimic α -helix-based protein–protein interaction motifs;
- iii) the allosteric activation of the human 20S proteasome by charged porphyrins that mimic interactions with 20S regulatory proteins;
- iv) the interaction of 2-phenylimidazo[1,2-b]pyridazine derivatives with calcium channels, highlighting their antiepileptic and neuroprotective properties;
- v) a computer-aided SAR analysis of 1,2,4-thiadiazolidine-3,5-dione derivatives as selective covalent inhibitors of coronavirus cysteine proteases;

vi) computational alanine-scanning mutagenesis of coronavirus non-structural proteins to identify novel druggable binding sites;

vii) the investigation of PROTAC ternary complexes using Molecular Interaction Fields, focusing on linker optimization for selective degradation of BET family bromodomains.

Collectively, these studies support the hypothesis that decoding the structural language of biomolecular actors is a powerful strategy in rational drug discovery and the design of innovative therapeutic agents.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
Chapter 1. Introduction.....	7
1.1. The “language” of protein-mediated interactions.....	7
1.2. Protein communication is mediated by protein interaction domains, short linear motifs and hot spots.....	9
1.2.1. Protein interaction domains.....	10
1.2.2. Short Linear Motifs.....	11
1.2.3. Hot Spots.....	13
1.3. Privileged scaffolds and natural compounds.....	16
1.4. Stereoselectivity and conformational enantiomers.....	18
1.5. References.....	19
Chapter 2. Role of Stereochemistry on the Biological Activity of Nature-Inspired 3-Br-Acivicin Isomers and Derivatives.....	23
2.1. Introduction.....	23
2.2. Results and Discussion.....	26
2.2.1. Effects of stereochemistry on in Vitro Inhibitory Activity towards PfGAPDH.....	26
2.2.2. Molecular Modelling Studies.....	29
2.2.3. Effect of stereochemistry on the antiparasitic activity..	42
2.3. Conclusions.....	45
2.4 References.....	46
Chapter 3. Tetrasubstituted Pyrrole Derivative Mimetics of Protein– Protein Interaction Hot-Spot Residues.....	48
3.1. Introduction.....	48
3.2. Results and Discussion.....	52

3.2.1. Design of the New TSP Derivatives.....	52
3.2.2. Solubility Assay.....	53
3.2.3. Biological evaluation.....	54
3.2.5. Computational studies.....	58
3.2.6. SAR Analysis.....	74
3.3. Conclusions.....	75
3.4. References.....	76
Chapter 4. Tetra-anionic porphyrin mimics protein-protein interactions between regulatory particles and the catalytic core, allosterically activating human 20S proteasome.....	78
4.1. Introduction.....	78
4.2 Results and Discussion.....	88
4.2.1. Effect of H2TPPS on the h20S proteolytic activity.....	88
4.2.2. AFM Imaging.....	93
4.2.3. Dynamic docking studies.....	95
4.2.4. Ligand analysis and generation of the starting complexes.	96
4.2.5. Docking calculations.....	100
4.3. Conclusions.....	112
4.5. References.....	114
Chapter 5. Molecular Modelling of 2-Phenylimidazo[1,2-b]pyridazine Derivatives Exhibiting In Vitro and In Vivo Pharmacological Activity	120
5.1 Introduction.....	120
5.2. Results and Discussion.....	122
5.2.1. Molecular Modelling Studies.....	122
5.2.2 Investigation of the Antiepileptic Potential of DM1 and DM2	

5.3 Conclusions.....	144
5.5. References.....	145
Chapter 6. 1,2,4-Thiadiazolidin-3,5-Diones as Inhibitors of Cysteine Proteases.....	148
6.1 Introduction.....	148
6.2. Results and Discussion.....	153
6.2.1. Inhibitory Activities of THIAs on Cysteine Proteases..	153
6.2.2. Molecular Modelling studies.....	163
6.3. Conclusions.....	208
6.5. References.....	210
Chapter 7. Mapping the Vulnerabilities of Coronaviruses: Druggable Hotspots in RNA Methylation Machinery.....	215
7.1. Introduction.....	215
7.2. Results and Discussion.....	219
7.2.1. Computational Alanine Scanning (CAS) mutagenesis experiments.....	220
7.2.2. SARS-CoV-2 Nsp10 Predicted Hot Spot Residues.....	222
7.2.3 Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp10 Hot Spots.....	230
7.2.4. SARS-CoV-2 Nsp14 Hot Spot Residue Prediction.....	232
7.2.5 Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp14 Hot Spots.....	242
7.2.6. SARS-CoV-2 Nsp16 Hot Spot Residues Prediction.....	244
7.2.7. Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp16 Hot Spots.....	253
7.2.8. Structural and Functional Comparison with Other Human Coronaviruses (HCoVs).....	255

7.3. Conclusions.....	264
7.5. References.....	265
Chapter 8. Rational PROTAC Design and Modelling.....	270
8.1. Introduction.....	270
8.1.1. Proteolysis Targeting Chimera (PROTACs).....	271
8.1.2. Molecular Interaction Fields.....	274
8.1.3. Rational Design of PROTACs using the RAINDROP procedure.....	277
8.2. Results and Discussion.....	280
8.2.1 Structure selection and docking procedure.....	281
8.2.3 Model generation and Correlation Analysis.....	282
8.3. Conclusions.....	288
8.5. References.....	289
Experimental Section.....	292
1. Pka calculations.....	292
2. Molecule preparation.....	293
2.1 Molecule Building.....	293
2.2 Forcefield.....	293
3. Conformational analysis.....	293
3.1 Conformational Sampling at MM level.....	293
3.2 MM-level conformer classification.....	295
3.3 DFT refinement of MM conformers.....	296
3.4 Pharmacophore-based geometric analysis.....	297
4. Protein structure selection, preparation and modelling.....	299
4.1. Selection of experimental structures.....	299
4.2. General preparation and protonation protocol.....	301

4.3. Homology modelling and rebuilding of missing regions. .	302
4.4. Structurally conserved regions and restraints in simulations	304
5. Docking and dynamic docking protocols.....	309
5.1. General considerations.....	309
5.2. Flexible docking (PfGAPDH and h20S proteasome).....	311
5.3. Grid-based docking: LigandFit and CDOCKER.....	312
5.4. Dynamic docking and simulated annealing refinement.....	314
5.5. Covalent docking of THIA derivatives.....	315
6. Molecular dynamics and steered molecular dynamics simulations	316
6.1. System setup and equilibration.....	317
6.2. Steered molecular dynamics protocol.....	318
6.3. Trajectory analysis and criteria for binding mode preservation 320	
7. Bioinformatics and structural analyses.....	321
7.1. Analysis of protein–protein interaction motifs and ligand pharmacophores.....	322
7.2. Surface mapping and interaction analysis on the 20S proteasome.....	323
7.3. Comparative analysis of Cav channels and cysteine proteases	325
7.4. Structural clustering, hotspot prediction and computational alanine scanning (nsp10, nsp14, nsp16).....	326

7.5. Cross-species sequence alignment of nsp10, nsp14 and nsp16	328
8. PROTAC ternary complex modelling (LigTBM and RAINDROP workflow).....	330
8.1. Binary complex generation with LigTBM.....	330
8.2. Ternary complex assembly with RAINDROP.....	331
8.3. Energetic filtering and clustering of ternary ensembles.....	332
8.4. Surface and interface analysis.....	333
References.....	335
Conclusions.....	337
Appendices.....	339

Chapter 1. Introduction

1.1. The “language” of protein-mediated interactions.

Proteins control most of the biological processes in living cells in concert with functional partners, forming a complex network of molecular signals. These inter-protein signals mutually regulate protein's function through specific interactions with functional ligands, which may be endogenous small-molecules or other proteins, in this latter case they are named protein-protein interactions (PPI) [1]. All the biological processes mediated by proteins and the ability to exert a specific functional effect depend on the interactions they can establish with their binding partners. The possibility to intervene on the whole set of molecular interactions present in a cell, the interactome gives opportunities for therapeutic strategies. However, each protein can have multiple partners and multiple conformations that enable the molecule to have yet another layer of complexity, thus making the deciphering of the basis of this network a major challenge [2].

The identification of protein interactions that mediate key biological functions may enable novel therapeutic approaches, either through their disruption or enhancement. A better understanding of how protein interactions change in pathological versus normal conditions will allow the discovery of novel drug targets for a large number of diseases [3].

Among the interactions proteins can establish with functional partners, PPIs have been considered “undruggable”, that is, which could not be targeted by a drug-like molecule. The situation changed in the last decades, and now PPI are among the most studied molecular interactions in the search of novel therapeutic tools.

PPIs can occur through binding surfaces or clefts in which are generally present smaller regions, called hot spots, that have the major contribution in terms of energy of the complex, a sort of energetical focal points [4-7]. The residues surrounding these hot spots probably acts as a “shield” impeding to the bulk solvent to access these regions. Interacting hot spots have a high degree of complementarity, in terms of shape and electronic properties, with hot spots packed in the PPI one against another. Moreover, these hot spots have a high degree of plasticity, for example proteins can bind to different partners with the same hot spots. Importantly these protein properties are reflected in their ligands, that is, similar substructures contained in different ligands bind to a similar hot spot in different proteins [8-10]. Accordingly, molecular communication in biological systems can be seen in the view of a “language”, and we can compare the structure of human grammar to that of biological molecules (proteins, RNA and DNA) [11-17]. With these premises, the link between phrases and discourses and the structure-function relationship in proteins would be an added value to the process of understanding protein communication (Figure 1).

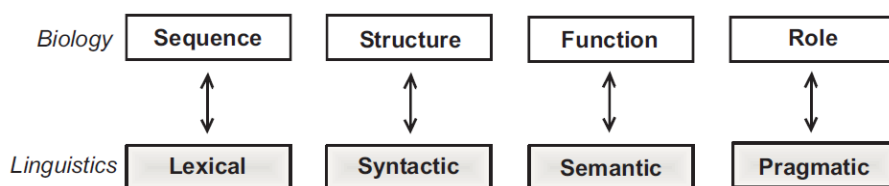


Figure 1. Similarities between the two hierarchies that apply to proteins (top row) and languages (bottom row). Reprinted by permission from Macmillan Publishers Ltd: Ref. [15], copyright 2006.

Decoding this “protein language” not only can help in the understanding of many biological processes, but also to design and/or identify small molecules that are able to modulate “wrong” protein interactions that can lead to pathological states, since the small molecule, an exogenous compound, can represent a portion of a “speech” (depending of the complexity of the molecule, it can be a letter, a word, a phrase) that can be recognized by our endogenous molecules, because they “speak” the same language. In this way a small molecule can change the meaning (semantic) of a message, and by doing so it can also change the role (pragmatic) of the targeted protein. This view can represent new opportunities in drug discovery.

1.2. Protein communication is mediated by protein interaction domains, short linear motifs and hot spots.

To understand an unknown language the first thing to do is to identify syntactic and semantic units. Proteins are characterized by protein interaction domains (PIDs) [18] and short linear motifs (SLiMs) [19] which enables the PPIs [20]. Thus, these PIDs and SLiMs might be the units of language in proteins. Some are exposed to the solvent, others become accessible only after a conformational change and/or aggregation state of a protein, depending strongly on the conformational state of a protein.

To understand the biological role of functional domains/motifs it is necessary to identify them and validate their individual function [15], keeping in mind that the full spectrum of a protein’s biological functions is the results of the concerted actions of these individual domains/motifs.

Hot spots are key regions within a biomolecular interface that contribute in a disproportionate way to the overall binding free energy. These regions often include a limited number of residues that play a dominant role in stabilizing the interaction through hydrogen bonds, van der Waals contacts, or electrostatic interactions [21]. Hot spots play a key role in PPI recognition, which enables the possibility to identify target sites for small molecules to act on that particular biological system.

1.2.1. Protein interaction domains.

PIDs can be seen as modular domains with a common core recognition ability and a high specificity, modulating in this way PPIs and the signal transduction system. Some examples of PIDs are the EF-hand-like domain or the pore domain of the Cytosolic C-terminal domain of human T-type voltage-dependent calcium channel (hCav_{3.1}), the NAD⁺-binding domain of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the grooves present on the surface of the α -ring of 20S proteasome (Figure 2).

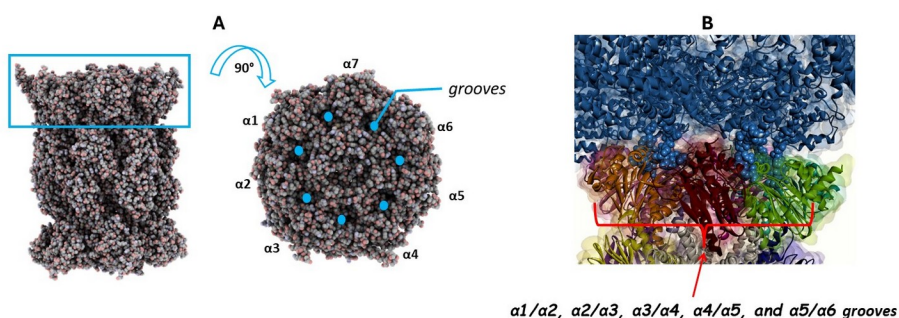


Figure 2. An example of PIDs in the human 20S proteasome. A: The grooves present on the surface of the α -ring of 20S involved in the interactions with the regulatory particles (i.e., 19S; PA28 and PA200) are evidenced. B: zoom on the interaction between human 20S and regulatory particle 19S (PDB ID: 5T0J).

These domains share common properties such as the ability to maintain their structure in isolation; in some cases the domain can be plugged into surface of proteins after a conformational change. It is worth noting that also intrinsically disorder regions of proteins can modulate protein functionality. Some PIDs can be seen in the same combination while others are seen in different arrangements. These are “promiscuous” domains, in the same way a word used frequently can have more meanings depending on the context. This versatility may have appeared, in evolutionary terms, because some domains had the potential to be useful in various contexts.

1.2.2. Short Linear Motifs.

Short Linear Motifs (SLiMs) are functional elements composed of a limited number of contiguous, specific amino acid residues within protein sequence, typically ranging from 3 to 11 residues in length [22].

The first example of interaction of a SLiM was described in 1990 with the pairing of SLiM KDEL (Lys-Asp-Glu-Leu) with the receptor ERD2, involved in endoplasmic-reticulum retrieval. Compared to PIDs it is difficult to detect SLiMs because of its intrinsically nature of short length sequences. The step for SLiMs identification usually involves the identification of a set of proteins with similar function and then with experimental techniques gradually going to identify a short common segment associated with that function. In other words, going back from the phrase to the word, but maintaining the context to understand the message that specific word can carry. Some examples of SLiMs are the HbYX activating motif present on regulatory proteins (RPs) of the proteasome

and RxFxV motif present in 14^{ARF} protein involved in its interaction with MDM2

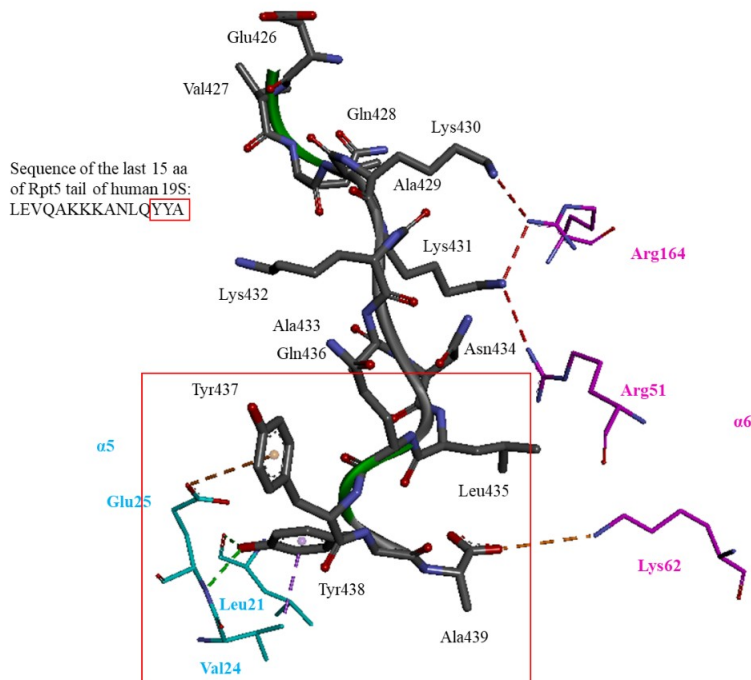


Figure 3. An example of SLiM. The C-terminal HbYX activating motif (dark grey stick) contained in the Rpt-5 subunit of 19S, a regulatory protein of human proteasome 20S (h20S). Key interactions with h20S (cyan and magenta) are displayed (PDB ID 5T0J).

They exhibit complex switching behaviour by cooperating with each other and/or by post-translational modifications, facilitating the switching between different functional states of a protein [23,24].

A key difference between SLiMs and PIDs is their affinity for their binding partner. Domains bind with each other with strong affinities, while the short length of short linear motifs means that they have rarely strong affinities as PIDs. Accordingly, SLiMs tend to mediate transient

interactions. Another difference between SLiMs and PIDs is the structural organization, that is, the secondary and tertiary structure. Indeed, SLiMs may need a specific bioactive conformation, anyway, limited to the secondary structure of a short sequence fragment, while PIDs are characterized by a specific tertiary structure (fold). In evolutionary terms, since the amino acid sequence is less conserved than the tertiary structure (fold), then, SLiMs are more prone to mutations, with just one mutation that can substantially modify the protein function. The consequence of this is that the entire motif is rarely conserved over long evolutionary distances, nevertheless, there are many consensus sequences identified in motifs of evolutionary distant proteins.

1.2.3. Hot Spots.

Hot spots are specific amino acid residues within protein–protein or protein–ligand interfaces that contribute significantly to the overall binding free energy, often with a $\Delta\Delta G \geq 2$ kcal/mol upon alanine substitution [25-27]. The discovery of hot spots has been instrumental in rational drug design, as targeting these key residues can enhance the specificity and efficacy of therapeutic inhibitors, and their identification has revealed druggable binding sites in several proteins [28]. These residues often correspond to the residues contained in SLiMs and/or PIDs that are conserved across homologous proteins and/or evolutionary distant species [29,30].

Despite being only a limited fraction of the interface residues, these sites are essential for stability and specificity of molecular recognition events. Hot spot residues are assumed to be clustered in specific regions involved in the binding with endogenous ligands and often form networks of

interactions involved in allosteric mechanisms. [31]. Accordingly, these residues are often found to be involved in hydrogen bonds, hydrophobic contacts, and electrostatic interactions. The discovery of hot spots has been instrumental in rational drug design, as targeting these key residues can enhance the specificity and efficacy of therapeutic inhibitors, and their identification has revealed druggable binding sites in several proteins [28]. Computational methods such as molecular dynamics simulation or computational alanine scanning are widely employed to predict and characterize hot spots [26,32,33].

The interfaces where hot spots are present are formed by proteins with good shape and electrostatic complementarity to their binding partners.

Protein–protein interfaces are typically hydrophobic, characterized by a large buried non-polar surface [34]. Hydrophobicity is therefore one of the main driving forces in protein–protein interactions [35,36], with hydrophobic residues engaging in van der Waals contacts and transitioning from a polar aqueous environment to a non-polar interface [37]. This transition releases ordered water molecules, resulting in a substantial increase in entropy that favors complex formation [38]. Despite the low enthalpic contribution of hydrophobic interactions, the entropic gain from desolvation significantly contributes to a favorable free energy change. This principle also applies when the functional ligand is not a protein.

In addition to hydrophobic forces, electrostatic interactions and hydrogen bonds play a major role in protein–protein recognition (Figure 4).

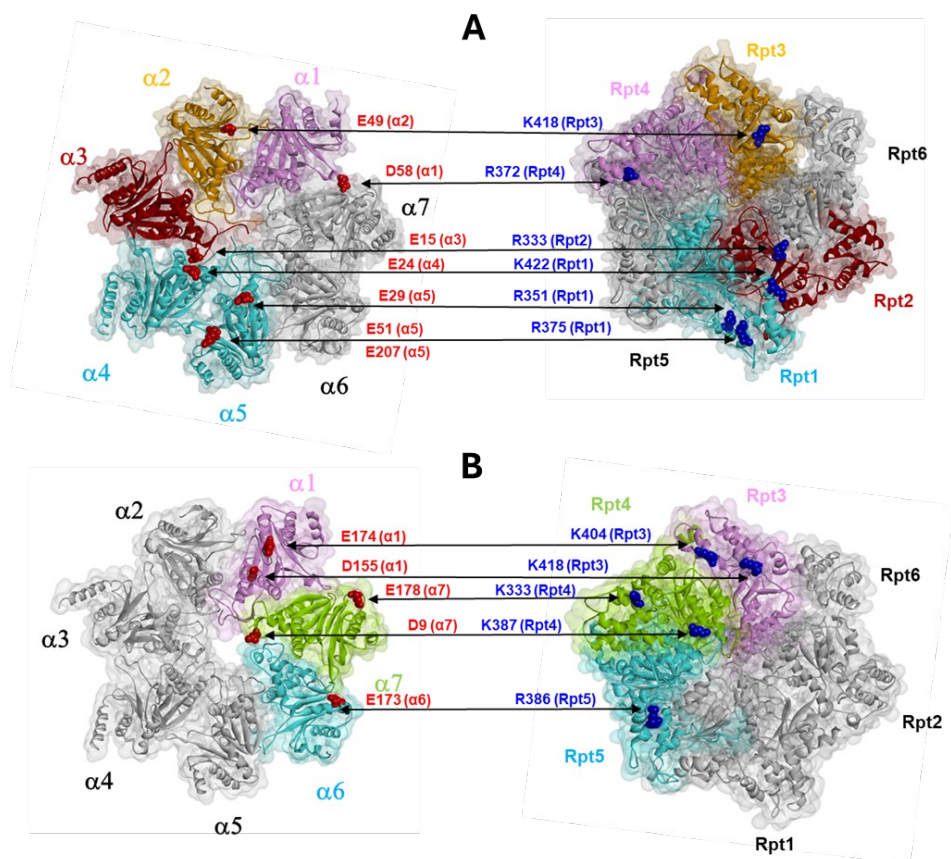


Figure 4. Ionic interactions occurring between human 20S proteasome and 19S RP complex: top view of the $\alpha 1$ - $\alpha 7$ ring of human 20S (left) is reported next to the bottom view of the Rpt1-6 subunits of 19S (right). (A) SA (PDB ID: 5T0G) and (B) SD (PDB ID: 5T0J) conformational/functional states. Proteins are displayed as solid ribbons and their van der Waals volume is displayed as transparent surface. The residues involved in ionic interactions between 20S (negative: red) and 19S (positive: blue) are displayed as CPK and labeled. The α subunits of 20S containing negatively charged interacting residue(s) are colored by the same color as the Rpt subunits of 19S containing the corresponding positively charged interacting residue(s). The rest of the 20S and 19S subunits are colored gray. Image modified from ref. Di Dato, A. et al. "Electrostatic map of proteasome α -rings encodes the design of allosteric porphyrin-based inhibitors able to affect 20S conformation by cooperative binding." *Scientific reports* 7.1 (2017): 17098.

While the entropic contribution from desolvation is less pronounced for polar or charged residues, the formation of strong electrostatic interactions and hydrogen bonds provides a substantial enthalpic contribution to the binding free energy. Electrostatic complementarity between binding partners promotes correct orientation and specificity, while hydrogen bonds stabilize the interface through directional and specific interactions, enhancing both affinity and association rate [39,40]. These interactions are equally relevant when the ligand is a non-protein molecule.

The same hot spot can be used to bind to several partners: in the same way a word can be used in different contexts. As an example, analyzing the molecular interactions of a hot spot at the interface of a protein heterodimer can provide valuable insights into potential interaction sites for other binding partners.

The same hot spot can be used to bind to several partners; in the same way a word can be used in different contexts. Analyzing the role of a hot spot within a specific heterodimer can also provide valuable insights into potential interaction sites for other binding partners. Moreover, these hot spot residues often overlap with structurally conserved residues [28,29].

1.3. Privileged scaffolds and natural compounds.

The discovery and development of drugs draw an inestimable source of inspiration from nature itself, particularly from natural products (NPs). The vast range of structurally diverse molecules and their biological activities is unmatched by any of the currently available synthetic screening libraries [41]. Indeed, NPs have evolved over millennia to acquire specific ligand-protein binding motifs and cover a biologically relevant chemical space that synthetic compounds often fail to efficiently

explore. This can account for the large number of NPs characterized by molecular core structures that, when shared by multiple compounds, confers bioactivity, thus providing new "privileged scaffolds"[41] for drug discovery [40-41].

However, natural products are also prone to pan-assay interference effects (PAINS), highlighting the need for rigorous evaluation in drug discovery. For instance, polyphenols and quinones often emerge as “false hits” in high-throughput screening assays. Rather than participating in selective binding, these compounds often interfere with assay measurements through non-specific mechanisms such as redox cycling, fluorescence quenching or colloidal aggregate formation. These kinds of drawbacks are rather common during the biological assessment of NPs. Thus, identifying and eliminating PAINS is a key step toward distinguish true privileged structures from assay artifacts, ensuring that the selected scaffolds possess genuine drug-like potential.

Although the structural complexity of natural products (NPs), their potential toxicity, unfavorable pharmacokinetics (PK), and pan-assay interference effects (PAINS) can limit their clinical applicability and large-scale production, the approach based on NPs and their privileged scaffolds is still considered today the most successful in identifying new drugs capable of interfering with protein interactions [41]. Current strategies emphasize the rational design of structural modifications to enhance pharmacological properties, coupled with the design of synthetic approaches for the development of libraries of structural analogues.

Indeed, the identification of privileged scaffolds matches with the evidence that binding sites in different proteins share common "motifs",

disclosing a mechanism that nature exploits to select biologically active molecules and offering new perspectives in drug discovery.

A perfect example of this evolutionary intelligence is the class of flavonoids. These natural products can bind to therapeutic targets that show a three-dimensional combination of secondary structure elements identical to that present in the binding site of their biosynthetic enzymes. This structural and topological correspondence, known as Protein Fold Topology (PFT), is crucial. Flavonoids are natural compounds known for their ability to inhibit protein kinases by occupying the same binding site as ATP [42]. The way quercetin binds to the therapeutic target PI3K is partially replicated when it binds to one of its own biosynthetic enzymes, anthocyanidin synthase [42]. This suggests that flavonoids use an evolutionarily conserved binding pattern to interact with different proteins.

1.4. Stereoselectivity and conformational enantiomers.

The role of stereoisomerism is central to drug design, as a molecule's spatial configuration profoundly influences its biological interactions, shaping its pharmacokinetics, metabolism, and target binding [43]. In this context, it is fundamental to distinguish between two main categories of stereoisomers: configurational enantiomers and conformational enantiomers.

Configurational enantiomers are non-superimposable mirror images characterized by permanent, non-interconvertible configurations around a

chiral center. Nature tends to biosynthesize natural products as a single, precise enantiomer.

In parallel, we encounter conformational enantiomers which are a pair of non-superimposable mirror-image isomers that can interconvert through the rotation of one or more single bonds. Among the atropoisomers are characterized by a rotational energy barrier sufficiently high (e.g., greater than 24-25 kcal/mol) to confer temporary stability to the chiral conformation [44].

1.5. References.

1. Che, Y., B. R. Brooks, and G. R. Marshall. "Development of small molecules designed to modulate protein-protein interactions." *Journal of Computer-Aided Molecular Design*, vol. 20, no. 2, 2006, pp. 109-130.
2. Akiva, E., G. Friedlander, Z. Itzhaki, and H. Margalit. "A Dynamic View of Domain-Motif Interactions." *PLoS Computational Biology*, vol. 8, no. 1, 2012, p. e1002341.
3. Deng, J. Z., D. J. Newman, and S. M. Hecht. "Use of COMPARE analysis to discover functional analogues of bleomycin." *Journal of Natural Products*, vol. 63, no. 9, 2000, pp. 1269-1272.
4. Bogan, A. A., and K. S. Thorn. "Anatomy of hot spots in protein interfaces." *Journal of Molecular Biology*, vol. 280, no. 1, 1998, pp. 1-9.
5. Arkin, M. R., and J. A. Wells. "Small-molecule inhibitors of protein-protein interactions: progressing towards the dream." *Nature Reviews Drug Discovery*, vol. 3, no. 4, 2004, pp. 301-317.
6. Moreira, I. S., P. A. Fernandes, and M. J. Ramos. "Hot spots: a review of the protein-protein interface determinant amino-acid residues." *Proteins: Structure, Function, and Bioinformatics*, vol. 68, no. 4, 2007, pp. 803-812.
7. Kozakova, D., et al. "Structural conservation of druggable hot spots in protein-protein interfaces." *Proceedings of the National Academy of Sciences*, vol. 108, no. 33, 2011, pp. 13528-13533.
8. Ferlini, Cristiano, et al. "Paclitaxel directly binds to Bcl-2 and functionally mimics activity of Nur77." *Cancer Research*, vol. 69, no. 17, 2009, pp. 6906-6914.
9. Andreoli, Mirko, et al. "Identification of the first inhibitor of the GBP1: PIM1 interaction. Implications for the development of a new class of

- anticancer agents against paclitaxel resistant cancer cells." *Journal of Medicinal Chemistry*, vol. 57, no. 19, 2014, pp. 7916-7932.
10. Persico, Marco, et al. "New anticancer agents mimicking protein recognition motifs." *Journal of Medicinal Chemistry*, vol. 56, no. 17, 2013, pp. 6666-6680.
 11. Baker, M. C. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. Basic Books, 2001.
 12. Pesole, G., M. Attimonelli, and C. Saccone. "Linguistic approaches to the analysis of sequence information." *Trends in Biotechnology*, vol. 12, no. 10, 1994, pp. 401-408.
 13. Searls, D. B. "Reading the book of life." *Bioinformatics*, vol. 17, no. 6, 2001, pp. 579-580.
 14. Popov, O., D. M. Segal, and E. N. Trifonov. "Linguistic complexity of protein sequences as compared to texts of human languages." *Biosystems*, vol. 38, no. 1, 1996, pp. 65-74.
 15. Gimona, M. "Protein linguistics - a grammar for modular protein assembly?" *Nature Reviews Molecular Cell Biology*, vol. 7, no. 1, 2006, pp. 68-73.
 16. Jensen, O. N. "Interpreting the protein language using proteomics." *Nature Reviews Molecular Cell Biology*, vol. 7, no. 5, 2006, pp. 391-403.
 17. Botstein, D., and J. M. Cherry. "Molecular linguistics: extracting information from gene and protein sequences." *Proceedings of the National Academy of Sciences*, vol. 94, no. 11, 1997, pp. 5506-5507.
 18. Cohen, G. B., R. Ren, and D. Baltimore. "Modular binding domains in signal transduction proteins." *Cell*, vol. 80, no. 2, 1995, pp. 237-248.
 19. Neduva, V., and R. B. Russell. "Linear motifs: evolutionary interaction switches." *FEBS Letters*, vol. 579, no. 15, 2005, pp. 3342-3345.
 20. Pang, E., and K. Lin. "Yeast protein-protein interaction binding sites: prediction from the motif-motif, motif-domain and domain-domain levels." *Molecular BioSystems*, vol. 6, no. 11, 2010, pp. 2164-2173.
 21. Bogan, Andrew A., and Kurt S. Thorn. "Anatomy of hot spots in protein interfaces." *Journal of Molecular Biology*, vol. 280, no. 1, 1998, pp. 1-9.
 22. Van Roey, K., et al. "The ELM resource: a compendium of conditional regulatory interaction interfaces." *Science Signaling*, vol. 6, no. 296, 2013, p. rs7.
 23. Van Roey, K., et al. "Capturing cooperative interactions with the PSI-MI format." *Database (Oxford)*, vol. 2013, 2013, p. bat066.
 24. Van Roey, K., T. J. Gibson, and N. E. Davey. "Motif switches: decision-making in cell regulation." *Current Opinion in Structural Biology*, vol. 22, no. 3, 2012, pp. 378-385.
 25. Lise, S., et al. "Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods." *BMC Bioinformatics*, vol. 10, no. 1, 2009, p. 365.
 26. DeLano, W. L. "Unraveling hot spots in binding interfaces: progress and challenges." *Current Opinion in Structural Biology*, vol. 12, no. 1, 2002, pp. 14-20.

27. Cunningham, B. C., and J. A. Wells. "Comparison of a structural and a functional epitope." *Journal of Molecular Biology*, vol. 234, no. 2, 1993, pp. 554-563.
28. Kozakov, Dima, et al. "Structural Conservation of Druggable Hot Spots in Protein-Protein Interfaces." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 33, 2011, pp. 13528-33.
29. Keskin, O., B. Ma, and R. Nussinov. "Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues." *Journal of Molecular Biology*, vol. 345, no. 5, 2005, pp. 1281-1294.
30. Li, W., et al. "Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions." *Biochemistry*, vol. 37, no. 32, 1998, pp. 11771-11779.
31. Wells, J. A. "Additivity of mutational effects in proteins." *Biochemistry*, vol. 29, no. 36, 1990, pp. 8509-8517.
32. Guharoy, G., and P. Chakrabarti. "Empirical energy-based analysis of protein-protein interfaces identifies strongly interacting residues." *Journal of Computer-Aided Molecular Design*, vol. 23, no. 11, 2009, pp. 895-902.
33. Smilova, Mihaela D., et al. "Fragment Hotspot Mapping to Identify Selectivity-Determining Regions between Related Proteins." *Journal of Chemical Information and Modeling*, vol. 62, no. 2, 2022, pp. 284-294.
34. Young, L., R. L. Jernigan, and D. G. Covell. "A role for surface hydrophobicity in protein-protein recognition." *Protein Science*, vol. 3, no. 5, 1994, pp. 717-729.
35. Korn, A. P., and R. M. Burnett. "Distribution and complementarity of hydrophobicity in multisubunit proteins." *Proteins: Structure, Function, and Bioinformatics*, vol. 9, no. 1, 1991, pp. 37-55.
36. Lijnzaad, P., and P. Argos. "Hydrophobic patches on protein subunit interfaces: characteristics and prediction." *Proteins: Structure, Function, and Bioinformatics*, vol. 28, no. 3, 1997, pp. 333-347.
37. Dill, K. A. "Dominant forces in protein folding." *Biochemistry*, vol. 29, no. 31, 1990, pp. 7133-7155.
38. Ross, P. D., & Subramanian, S. (1981). "Thermodynamics of protein association reactions: forces contributing to stability." *Biochemistry*, 20(11), 3096-3102.
39. Schreiber, G., & Fersht, A. R. (1996). "Rapid, electrostatically assisted association of proteins." *Nature Structural Biology*, 3(5), 427-431.
40. Jones, S., & Thornton, J. M. (1996). "Principles of protein-protein interactions." *Proceedings of the National Academy of Sciences*, 93(1), 13-20.
41. Davison, Emma K., and Margaret A. Brimble. "Natural product derived privileged scaffolds in drug discovery." *Current opinion in chemical biology* 52 (2019): 1-8.

42. Persico, Marco, et al. "From protein communication to drug discovery." *Current Topics in Medicinal Chemistry*, vol. 15, no. 20, 2015, pp. 2019-2031.
43. McConathy, Jonathan, and Michael J. Owens. "Stereochemistry in drug action." *Primary care companion to the Journal of clinical psychiatry* 5.2 (2003): 70
44. Brethome, Alexandre V., Stephen P. Fletcher, and Robert S. Paton. "Conformational effects on physical-organic descriptors: the case of sterimol steric parameters." *ACS Catalysis*, vol. 9, no. 3, 2019, pp. 2313-2323.

Chapter 2. Role of Stereochemistry on the Biological Activity of Nature-Inspired 3-Br-Acivicin Isomers and Derivatives.

2.1. Introduction.

Stereochemistry shapes the “syntax” by which ligands are “read” by the hot spots of the protein binding site; thus, enantiomeric and diastereomeric variants can deliver distinct “messages” by altering shape and electrostatic complementarity at recognition sites.

The unique chemical diversity of natural products (NPs) allows the expansion of chemical space investigated by medicinal chemists [1]. The majority of NPs are chiral and are biosynthesized in a precise enantiomer [2]. Stereochemistry has a strong impact on the action of a drug since it can affect the binding to metabolic enzymes as well as to the molecular target. Moreover, it also affects recognition by protein transport systems, resulting in a stereospecific uptake of molecules within the cell [3].

This chapter reports the investigation performed on the role of stereochemistry on the activity of nature-inspired isooxazoline derivatives.

Stereochemistry shapes the “syntax” by which ligands are read by PIDs/SLiMs and by interface hot spots; thus, enantiomeric and diastereomeric variants can deliver distinct “meanings” at PPIs by altering complementarity, desolvation, and electrostatics at recognition sites.

The natural compound (5S, α S) acivicin (Figure 1A) produced by *Streptomyces svinceus*, and its analogue 3-Br-acivicin (3-BA, **1a**, Figure 1A) have been described as L-glutamine analogues able to inhibit in an

irreversibly mode several glutamine-dependent amidotransferase, such as CTP synthetase, carbamoyl phosphate synthetase, XMP aminase[4-6] and γ -glutamyl transpeptidase[7].

The mechanism of action of acivicin and 3-BA involves the nucleophilic attack of an activated catalytic cysteine residue of the target enzyme to the Cl or Br-substituted C-3 of the 4,5-dihydroisoxazole ring (Figure 1C) [8,9].

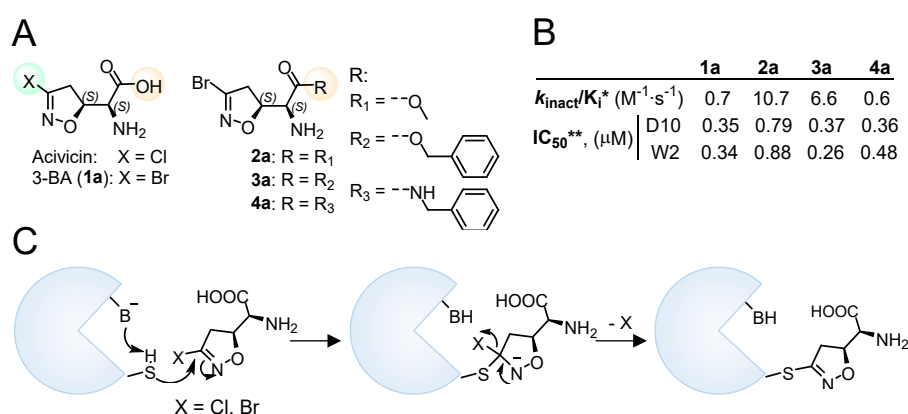


Figure 1. (A) Chemical structure of acivicin, 3-BA (**1a**) and 3-BA derivatives (**2a–4a**). (B) Biological data for compounds **1a–4a**. * Second-order rate constant used to characterize covalent binding of irreversible inhibitors to *Pf*GAPDH. ** IC_{50} towards *P. falciparum* D10 and W2 strains. (C) Proposed reaction mechanism for the 3-halo-4,5-dihydroisoxazole moiety.

Interestingly, towards *Trypanosoma brucei* CTPS, 3-BA showed a three-fold potency with respect to acivicin, demonstrating that the leaving group plays an important role in the irreversible mode of action. This difference in potency is translated into enhanced antitrypanosomal activity in vivo and in vitro [10]. The substitution effect, from chlorine to bromine, can be

seen as modulation of the “phonetics” of the ligand: tuning leaving-group propensity and solvent exposure (SASA) to better match the shape of the active site.

When tested on other therapeutically relevant targets, such as *Plasmodium falciparum* glyceraldehyde 3-phosphate dehydrogenase (*Pf*GAPDH), a key enzyme of the *Plasmodium* glycolytic pathway [11]. Acivicin was inactive, whereas 3-BA was able to irreversibly inhibit *Pf*GAPDH.

The aminoacidic portion of 3-BA (compound **1a**) was then modified obtaining a series of analogues but keeping the original stereochemistry unchanged. [11-14]. This series of derivatives, two ester analogues (**2a**, **3a**) and an amide analogue (**4a**) were investigated for their activity on *P. falciparum* and for the ability to covalently inhibit *Pf*GAPDH. All compounds showed antiparasitic activity against W2 (chloroquine resistant) and D10 (chloroquine sensitive) strain of *P. falciparum*, with values of IC₅₀ lower than 1 μ M (Figure 1B).

Thus, while chemical modifications of the amino acid portion were extensively explored, the impact of stereochemistry on antimalarial activity and protein binding has not yet been investigated.

Based on this, to expand structure–activity relationships (SARs), in the herein reported study stereochemical variants of 3-BA have been examined to assess whether this privileged natural scaffold retains its functional recognizability.

This study was executed in the context of a multidisciplinary collaboration involving the research groups of Prof. Paola Conti’s group (University of Milan “La Statale”) for compound synthesis and Prof. Stefano Bruno’s group (University of Parma) for biological assays. Research activities were carried out in the framework of the COST action network CA21111- “One

Health drugs against parasitic vector borne diseases in Europe and beyond” and are the object of the scientific publication reported in ref. [15].

I was directly involved in performing and analyzing computational studies as well as in the interpretation of the experimental results (particularly SAR analysis). I participated in the meetings of the research groups involved in the study and attended school and congresses related to this research topic such as the School of COST action “MedChem and structural biology: tools and strategies for hit and lead optimization in the One Health perspective” held at university of Siena.

The experimental procedure of synthetic and biological studies is not herein reported and can be found in the published manuscript [15].

2.2. Results and Discussion.

2.2.1. Effects of stereochemistry on in Vitro Inhibitory Activity towards PfGAPDH.

Compounds **1a-d**, **2a-d**, **3a-d**, **4a-d** were synthesized and tested for the ability to inhibit *PfGAPDH* activity. The incubation time was of 3 hours and the concentration at which the molecules were tested was 100 μ M (Figure 2). Compounds **2a-d** and **3a-d** were selected for a further investigation and tested at 50 μ M for 30 min.

Interestingly, it resulted that the effect of stereochemistry depends on the nature of the substituent (subclasses **1-4**). Indeed, 3-Br-acivicin (**1a**) inhibited *PfGAPDH* with a residual activity less than 50% while its diastereoisomers (**1b** and **1c**) and its enantiomer (**1d**) were inactive. The amide isomers displayed a similar trend comparing **4a** to **4b-d**. Instead, the

methyl (**2a–d**) and benzyl (**3a–d**) ester derivatives and their isomers were able to inhibit completely *Pf*GAPDH at 100 μ M, with a residual activity lower than 1% (Figure 2A).

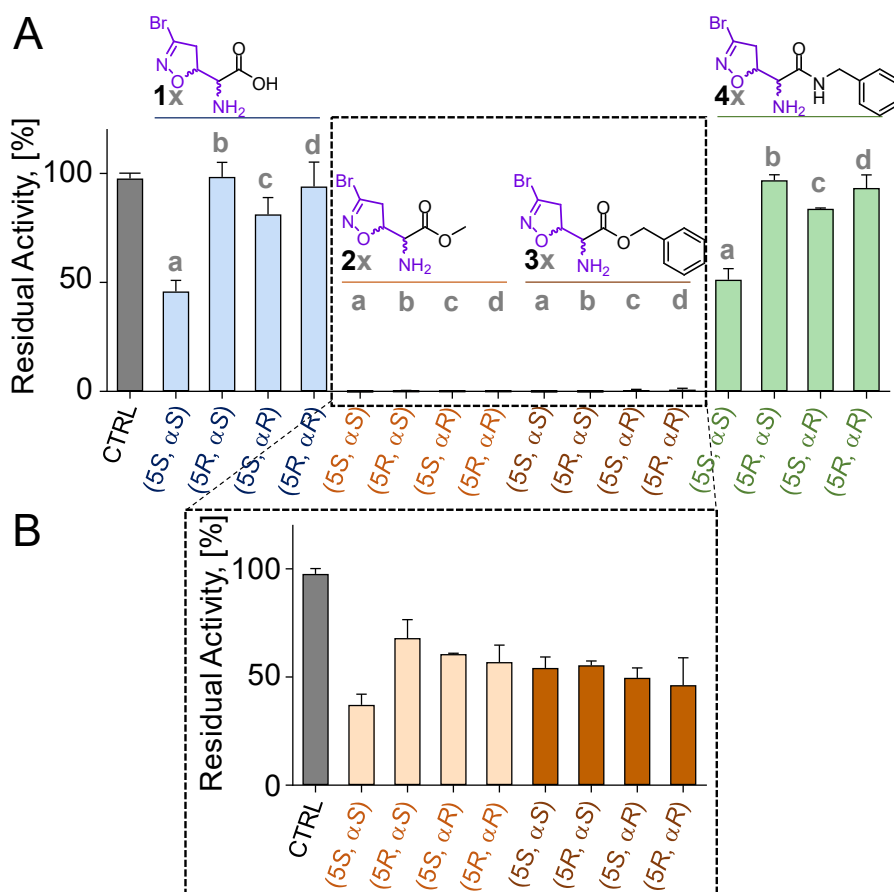


Figure 2. (A, B) Residual activity (%) of *Pf*GAPDH upon incubation with 100 μ M of compounds **1a–d**, **2a–d**, **3a–d**, and **4a–d** for 3 h (A) or 50 μ M of compounds **2a–d** and **3a–d** for 30 min (B). All experiments were performed at 25 °C in a buffered solution containing 10 mM of TEA, 5 mM of EDTA, and 10 mM of sodium arsenate, pH 7.6. *Pf*GAPDH was at 2 μ M concentration. CTRL: activity of *Pf*GAPDH maintained under the same conditions in the absence of inhibitors. Experiments were performed in independent triplicates. Data are shown as mean $\pm$ SD.

For this reason, these series of compounds (**2a-d** and **3a-d**) were further tested at 50 μ M for 30 minutes. The enzyme inhibition was higher than 50% for all the diastereoisomers, with **2a** the most active with 37% residual activity (Figure 2B).

The covalent mechanism of action of compounds **2a** and **3a** was characterized through mass spectrometry [11] and also the importance of the catalytic Cys153 with no involvement of other Cys residues was proved [11]. Moreover, the k_{inact}/K_i ratio used to characterize a mode of binding that is covalent was measured for compounds **1a-4a** [12]. Consequently, it is reasonable to say that all synthesized derivatives, possessing the same 3-halo-4,5-dihydroisoxazole warhead, share the potential to undergo the same specific nucleophilic attack mechanism observed for the natural (5S, α S) isomers.

However, the experimental data reveal that the impact of stereochemistry on the pre-reaction alignment is strictly dependent on the subclass (Figure 2). For the acid (**1a-d**) and amide (**4a-d**) derivatives, the lack of inhibitory activity observed for the non-natural stereoisomers confirms that the natural configuration is essential for placing the warhead in a position compatible with the covalent bond formation.

Conversely, the ester derivatives (**2a-d** and **3a-d**) exhibited a greater stereochemical tolerance. While the natural (5S, α S) isomers remained the most potent (Figure 2B), the non-natural stereoisomers retained the ability to inhibit the enzyme, leading to complete inactivation at 100 μ M (Figure 2A). This suggests that for the ester series, unlike the acid and amide analogues, the reactive moiety can still access and modify Cys153 despite the stereochemical inversion.

2.2.2. Molecular Modelling Studies.

The four series of compounds were subjected to conformational analysis and the resulting low energy conformers ($\Delta E_{\text{GM}} \leq 5$ kcal/mol) were placed in the *Pf*GAPDH catalytic site to perform flexible docking simulations as previously applied to **2a**, **3a** and **4a**. Obtained results (Appendix A: Table 1-38) were integrated with those previously obtained for **2a**, **3a** and **4a** [12].

The aim of the simulations was to investigate the binding of the compounds just before the nucleophilic attack by the catalytic cysteine residue (C153). Specifically, molecular docking was employed to characterize the geometry of the non-covalent complex, which is a prerequisite for the subsequent chemical reaction. It was proposed a flip-flop model for the catalytic reaction mechanism of GAPDH, where the substrate rotates around C153 and the phosphate group at the C-3 position undergoes a shift between two interaction sites, transitioning from the first site (PI) to the second (PII) [13,14]. Taking into account these information, two different binding approaches to C153 were considered (BA1 and BA2) for each compound which corresponded to the approach from the P_I or P_{II} site, respectively. These starting binding poses corresponded to the approach from the P_I and P_{II} sites (Figure 3 and 4, Appendix A: Tables 15-32). The docked ligands were compared to the conformers obtained by the conformational analysis in terms of geometry and conformational energy (Appendix A: Table 34). It was calculated the solvent-accessible surface area (SASA) of the leaving group (bromine atom) (Appendix A: Table 34), and the ligand-protein interactions were analyzed (Appendix A: Tables 35-38). Moreover, the complexes obtained were filtered by the distance (≤ 4

Å) between the electrophilic carbon C3 of the 3-bromo-4,5-dihydroisoxazole ring and the sulphur atom of C153 (See experimental section 2.4.3 Docking Studies).

It resulted that the active isomers **2a-2d**, could approach the catalytic cysteine C153 assuming either BA1 or BA2. In BA1 (Figures 3A and 4) the P_I site interacts with the protonated amine function and the ester function, while the bromine atoms are directed toward NAD⁺ thus, the bromine atom could leave the active site together with NAD⁺, in a similar way as the hydride ion of the substrate. In BA2 (Figures 3B and 5), the protonated amine and ester function are in the P_{II} site, while the bromine atom is in the P_I site pointing towards the active segment. This binding approach, taking in consideration the orientation of the leaving group, is favoured by a large solvent-accessible surface area (SASA) of the bromine atom (Appendix A: Table 34). In summary, both binding modes are compatible with the reaction with Cys135. In addition, the docked conformations of **2a-2d** were all within 5 kcal/mol from the global energy minimum (GM) (Appendix A: Table 33).

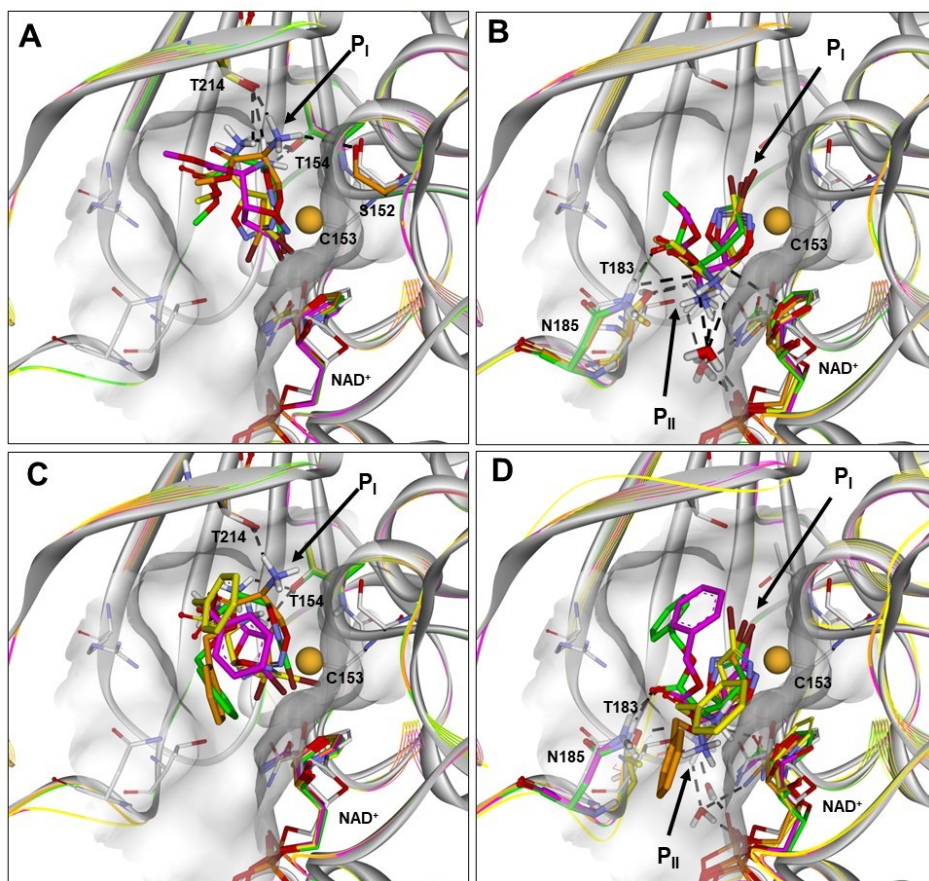


Figure 3. (A–D) Docked structures of *Pf*GAPDH in complex with: (i) **2a** (orange), **2b** (yellow), **2c** (green), and **2d** (magenta) approaching from the P_I site (BA1; **A**) or P_{II} site (BA2; **B**); (ii) **3a** (orange), **3b** (yellow), **3c** (green), and **3d** (magenta) approaching from the P_I site (BA1; **C**) or P_{II} site (BA2; **D**). All structures are superimposed on the starting *Pf*GAPDH conformation by fitting Ca atoms. The starting protein conformation (gray) is displayed as solid ribbons and the Connolly surface of the active site is shown. The backbone of the docked complexes is in line ribbons, key interaction residues are displayed and labeled, and heteroatoms are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The van der Waals volume of the sulphur atom of C153 is scaled by 50%. Hydrogen atoms are omitted for clarity, except those involved in ligand–protein hydrogen bond interactions (black dashed lines).

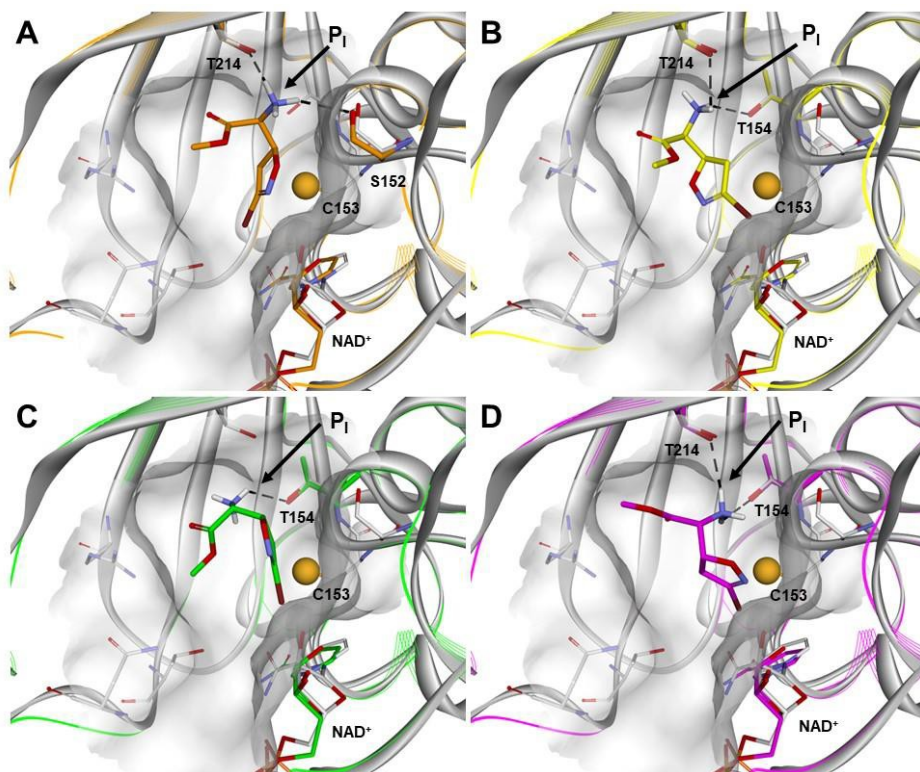


Figure 4. Docked complexes of **2a** (orange), **2b** (yellow), **2c** (green) and **2d** (magenta) bound to PI site of *Pf*GAPDH (BA1). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

The diastereoisomers **3a-3d** showed similar results, even if they had a benzyl moiety that resulted in a more hindered moiety respect to the methyl group of **2a-2d**.

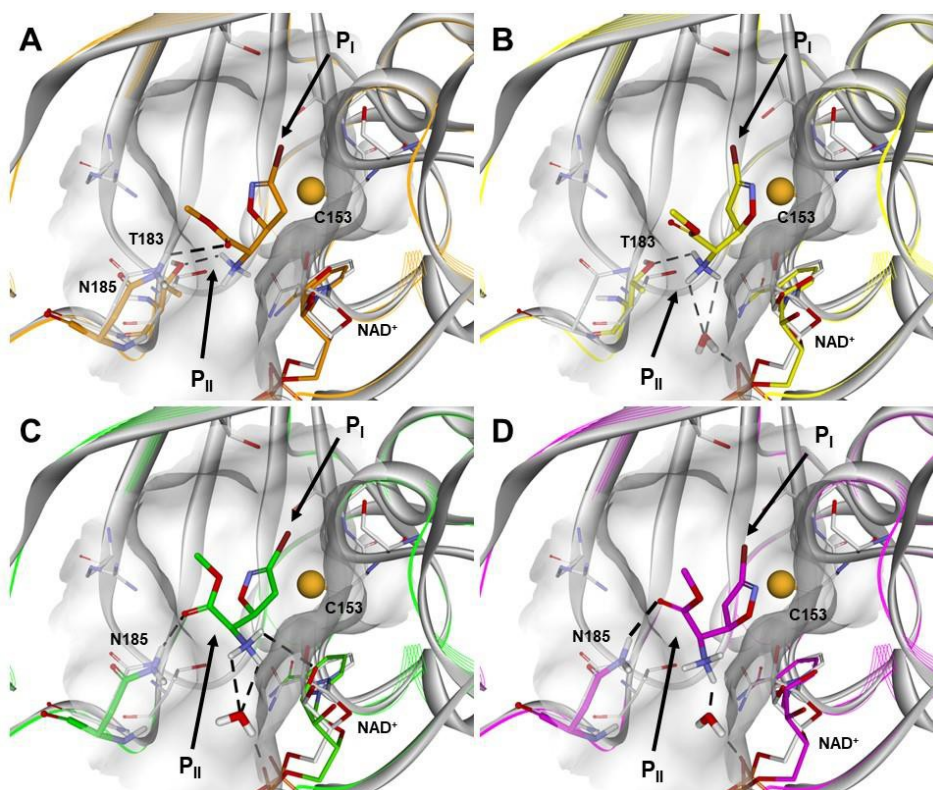


Figure 5. Docked complexes of **2a** (orange), **2b** (yellow), **2c** (green) and **2d** (magenta) bound to PII site of *P/GAPDH* (BA2). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

In binding mode BA1 (Figures 3 C and 6), the position of the phenyl ring is in a cleft formed between the S-loop (aa182-210) and the S7-S8 loop (aa122-130), with a little different orientation depending on the stereochemistry of C5 carbon (in P_{II} site for **3a** and **3c**, 5S diastereoisomers; Figure 3C). In binding mode BA2 instead the phenyl ring is oriented by the C α carbon that points towards the P_I site or the P_{II}

site in α R (**3c**,**3d**) and α S (**3a**, **3b**) diastereoisomers, respectively (Figures 3D and 7). In any case, there is enough space to accommodate the aromatic substituent adopting both binding modes, without affecting the approach to the catalytic cysteine.

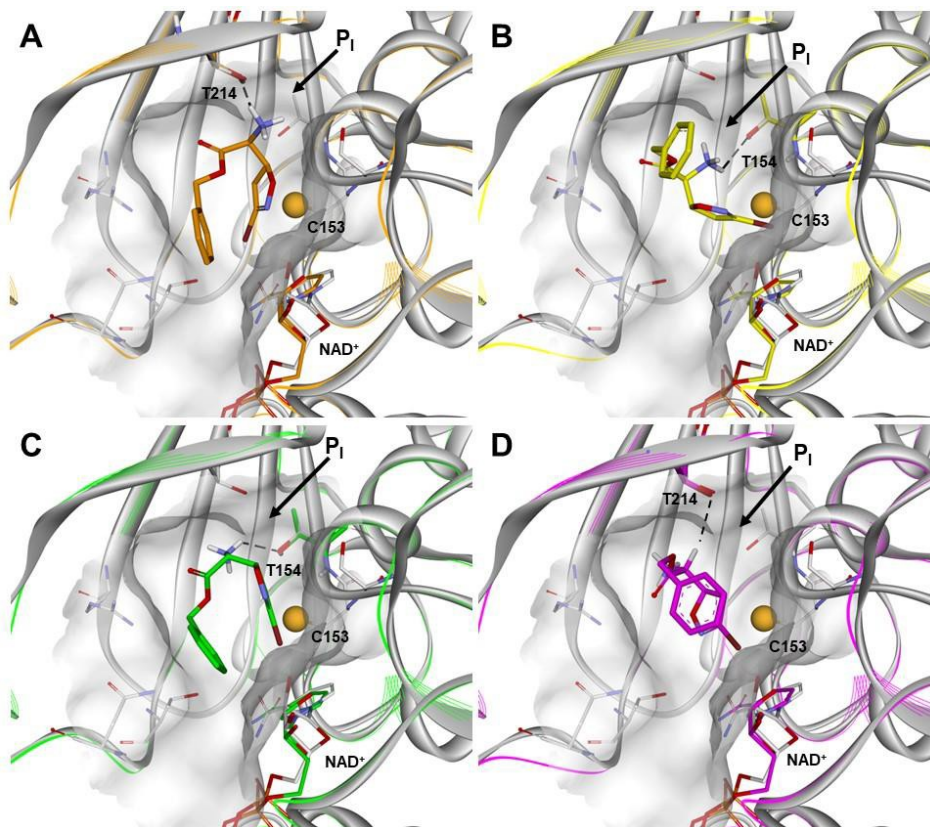


Figure 6. Docked complexes of **3a** (orange), **3b** (yellow), **3c** (green) and **3d** (magenta) bound to P_I site of P/GAPDH (BA1). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

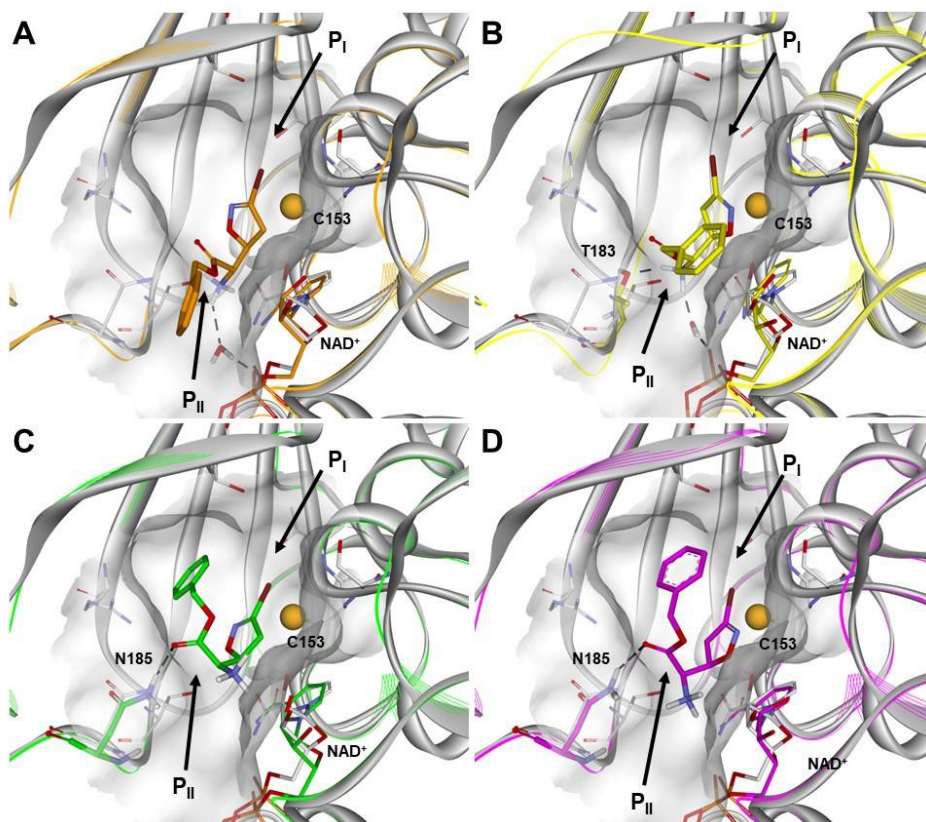


Figure 7. Docked complexes of **3a** (orange), **3b** (yellow), **3c** (green) and **3d** (magenta) bound to P_{II} site of *Pf*GAPDH (BA2). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

The results obtained for **1a-1d** indicate that a salt bridge between the negatively charged carboxylic group of the ligands and R237 of *Pf*GAPDH (engaged in phosphate binding and translocation) could be accountable for the significant reduction in the inhibitory activity of **1b-1d**.

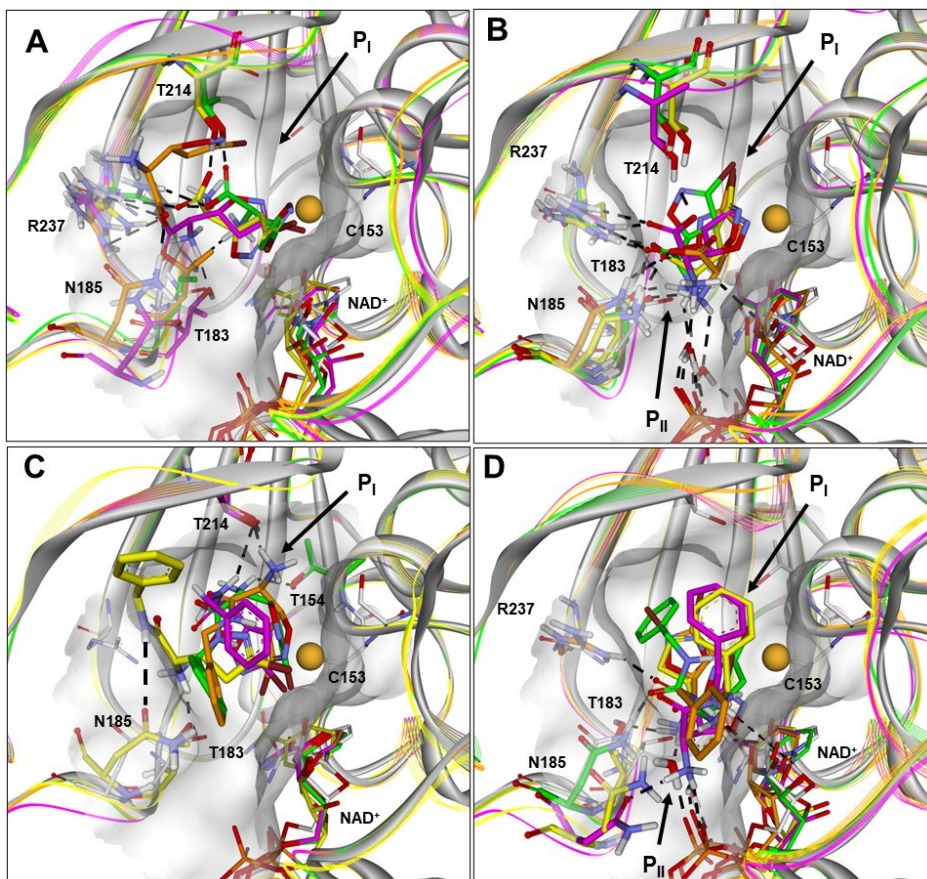


Figure 8. (A–D) Docked structures of *Pf*GAPDH in complex with: (i) **1a** (orange), **1b** (yellow), **1c** (green), and **1d** (magenta) approaching from the P_I site (BA1; **A**) or P_{II} site (BA2; **B**); (ii) **4a** (orange), **4b** (yellow), **4c** (green), and **4d** (magenta) approaching from the P_I site (BA1; **C**) or P_{II} site (BA2; **D**). All structures are superimposed on the starting *Pf*GAPDH conformation by fitting C α atoms. The starting protein conformation (gray) is displayed as solid ribbons and the Connolly surface of the active site is shown. The backbone of the docked complexes is in line ribbons, key interaction residues are displayed and labeled, and heteroatoms are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The van der Waals volume of the sulfur atom of C153 is scaled by 50%. Hydrogen atoms are omitted for clarity, except those involved in ligand–protein hydrogen bond interactions (black dashed lines).

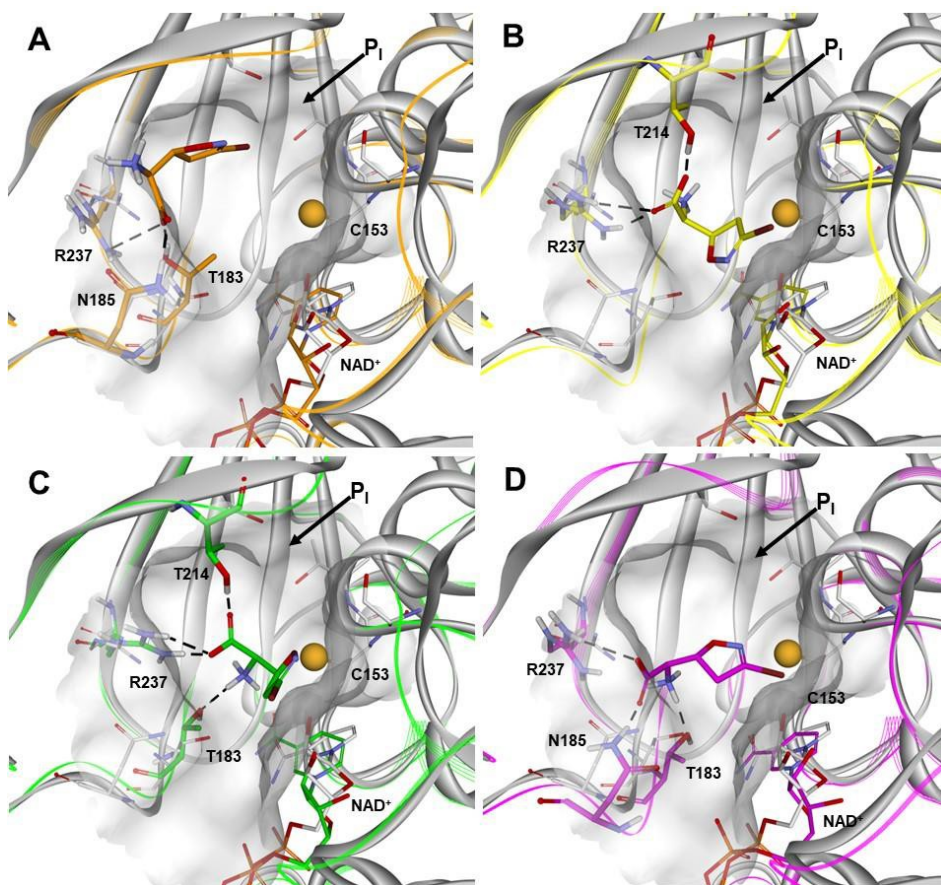


Figure 9. Docked complexes of **1a** (orange), **1b** (yellow), **1c** (green) and **1d** (magenta) bound to P_I site of *Pf*GAPDH (BA1). The starting protein (grey) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

When the carboxylic group is positioned at the P_I site (BA1, Figures 8A and 9), this interaction keeps the electrophilic carbon of 3-bromo-4,5-dihydroisoxazole ring (C3) away from the sulfur atom of C153 (distance >4 Å; **1a**, **1b** and **1c**), or, as in **1d**, places the 4,5-dihydroisoxazole nitrogen

atom between the C3 carbon and C153 sulfur atom. In both cases this interaction impairs the formation of the carbon sulfur bond.

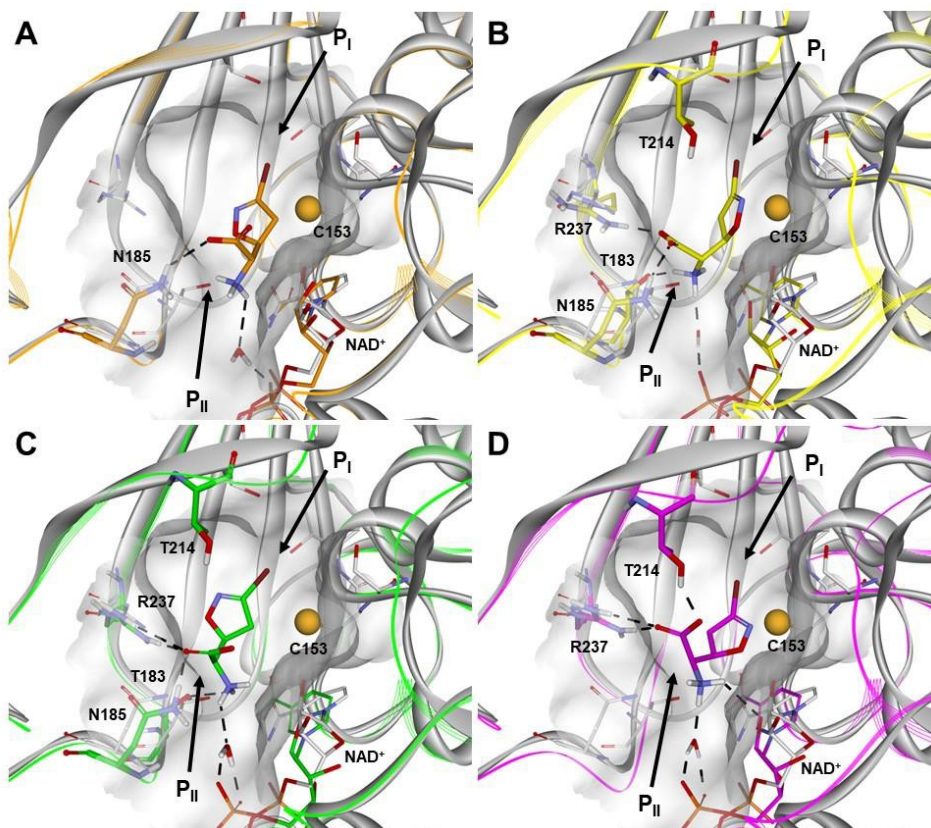


Figure 10. Docked complexes of **1a** (orange), **1b** (yellow), **1c** (green) and **1d** (magenta) bound to P_{II} site of *Pf*GAPDH (BA2). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

When the carboxylic acid of compounds **1b-1d** occupy the P_{II} site (BA2; Figures 8B and 10 B-D), it still establishes the salt-bridge with R237,

keeping C3 away from the sulfur atom of Cys153, along with a charge-assisted hydrogen bond with T214, reducing the SASA of the bromine atom in P_I site (Appendix A: Table 34). Of important notice, assuming BA2, the carboxylic acid of **1a** (the only active diastereoisomer) does not have interaction with R237 and the SASA of the bromine atom is not decreased (Figure 10A; Appendix A: Tables 34 and 35).

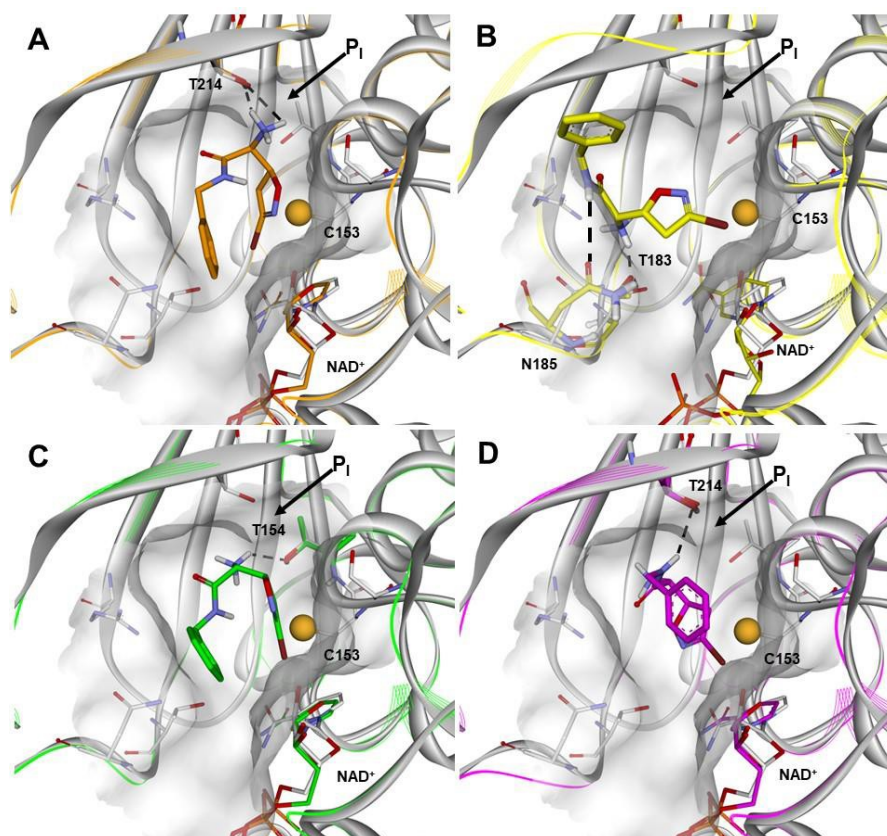


Figure 11. Docked complexes of **4a** (orange), **4b** (yellow), **4c** (green) and **4d** (magenta) bound to P_I site of PfGAPDH (BA1). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

In line with these results, amide derivative **4b** and **4d** with the amide group have additional interactions with the protein (particularly T183 and T214) caused by the presence of the NH, moving, also in this case, the ligand away from the catalytic cysteine C153 (Figures 8 C-D, 11 B,D , and 12 B,D; Appendix A: Tables 27-30; Figures). **4a** showed a similar behaviour only when bound to PII (BA2; Figure 11A vs Figure 12A).

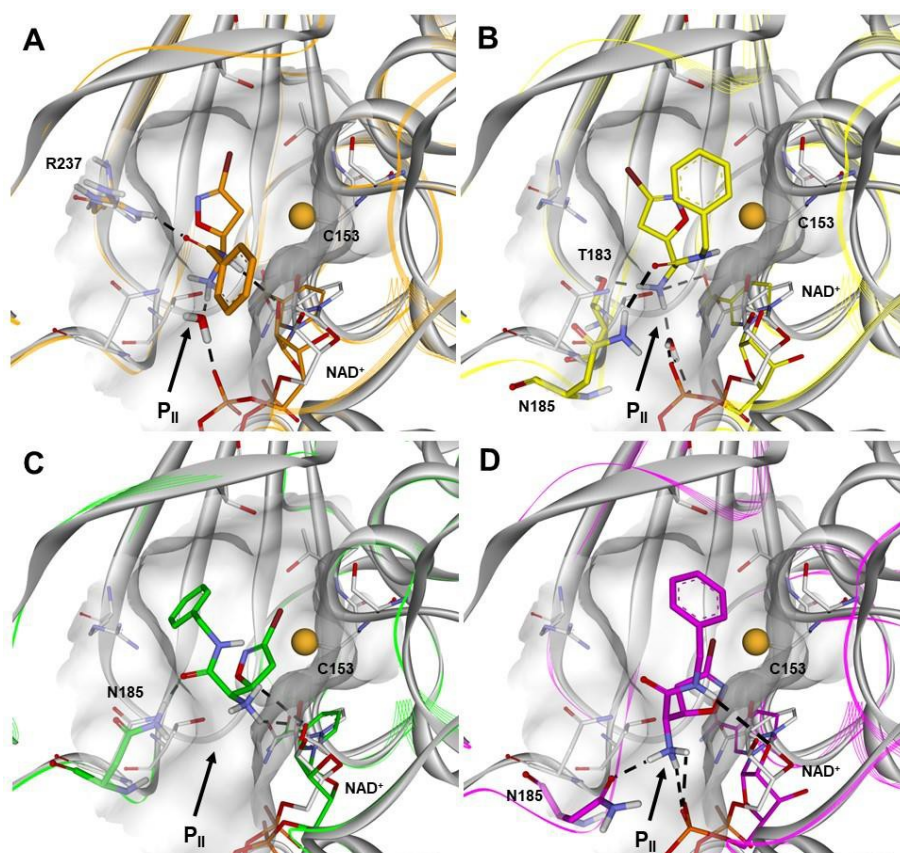


Figure 12. Docked complexes of **4a** (orange), **4b** (yellow), **4c** (green) and **4d** (magenta) bound to PII site of *Pf*GAPDH (BA2). The starting protein (gray) is displayed as Connolly surface and solid ribbons. The key interaction residues are displayed and labelled. The ligands, NAD⁺ and C153, as well as the residues involved in the interactions with ligands (stick) are colored by atom type (N, blue; O, red; S, yellow; Br, brown). The sulfur atom of C153 is displayed in CPK (scaled by 50%). Hydrogen atoms are omitted for clarity, except those involved in hydrogen bond interactions (black dashed lines).

On the other hand, **4c** has a highly distorted conformation in both binding modes (Figures 13 and 14), with an energy from the GM higher than 15 kcal/mol (Appendix A: Table 34).

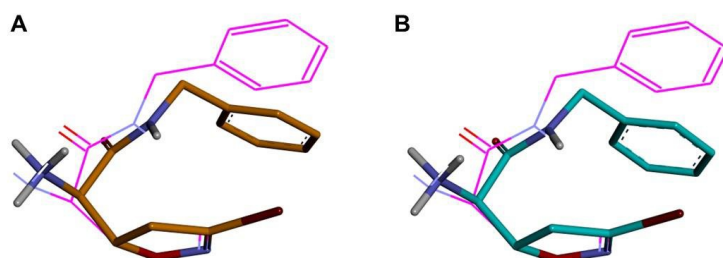


Figure 13. Superimposition of the two **4c** conformations obtained from docking studies (A: orange BA1; B: cyan BA2) on the corresponding local energy minimum conformer obtained from conformational analysis (magenta and lines). The structures are superimposed by 4,5-dihydroisoxazole ring heavy atoms and colored by atom type (N, blue; O, red; Br, brown).

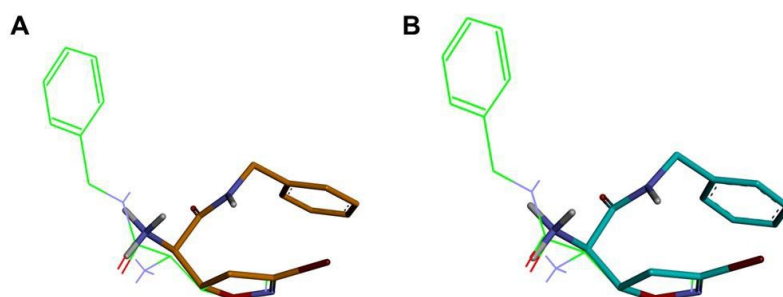


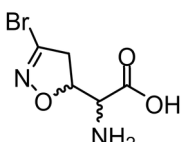
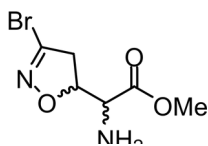
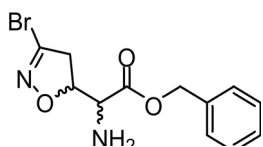
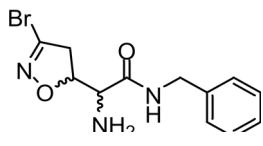
Figure 14. Superimposition of the two **4c** conformations obtained from docking studies (A: orange BA1; B: cyan BA2) on the global energy minimum conformer obtained from the conformational analysis (green and lines). The structures are superimposed by the 4,5-dihydroisoxazole ring heavy atoms and colored by atom type (N, blue; O, red; Br, brown).

These computational results are in agreement with the *Pf*GAPDH inhibitory activities, and the molecular models produced represent crucial tools for future rational structural modifications.

2.2.3. Effect of stereochemistry on the antiplasmodial activity.

Compounds **1a-d**, **2a-d**, **3a-d**, **4a-d** were tested for their in vitro antiplasmodial activity (Table 1).

Table 1. Chemical structures, clogD, and IC₅₀s against *P. falciparum* D10 and W2 strains.

Structure	Compound	clogD ^a	<i>P. falciparum</i> D10 IC ₅₀ (μM) ^b	<i>P. falciparum</i> W2 IC ₅₀ (μM) ^b
	(5 <i>S</i> , α <i>S</i>)- 1a	0.02	0.35 ± 0.08	0.34 ± 0.12
	(5 <i>R</i> , α <i>S</i>)- 1b		23.54 ± 0.31	24.75 ± 0.90
	(5 <i>S</i> , α <i>R</i>)- 1c		7.49 ± 1.48	8.47 ± 2.06
	(5 <i>R</i> , α <i>R</i>)- 1d		8.79 ± 1.12	10.18 ± 1.75
	(5 <i>S</i> , α <i>S</i>)- 2a	0.04	0.79 ± 0.21	0.88 ± 0.23
	(5 <i>R</i> , α <i>S</i>)- 2b		17.14 ± 5.91	17.18 ± 7.39
	(5 <i>S</i> , α <i>R</i>)- 2c		40.67 ± 16.20	48.18 ± 14.81
	(5 <i>R</i> , α <i>R</i>)- 2d		8.27 ± 1.05	7.42 ± 3.54
	(5 <i>S</i> , α <i>S</i>)- 3a	1.41	0.37 ± 0.12	0.26 ± 0.05
	(5 <i>R</i> , α <i>S</i>)- 3b		32.65 ± 18.10	34.85 ± 5.30
	(5 <i>S</i> , α <i>R</i>)- 3c		19.35 ± 4.34	16.51 ± 8.11
	(5 <i>R</i> , α <i>R</i>)- 3d		4.31 ± 0.51	4.50 ± 1.82
	(5 <i>S</i> , α <i>S</i>)- 4a	0.76	0.36 ± 0.11	0.48 ± 0.18
	(5 <i>R</i> , α <i>S</i>)- 4b		18.20 ± 1.15	18.20 ± 0.93
	(5 <i>S</i> , α <i>R</i>)- 4c		19.35 ± 2.15	19.35 ± 3.36
	(5 <i>R</i> , α <i>R</i>)- 4d		8.55 ± 1.95	8.91 ± 3.40

^a The clogD values (pH 7.4) were calculated by using ACD/pKa GALAS algorithm of ACD/Percepta software (ACD/Percepta, Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2017, <http://www.acdlabs.com> (accessed on 5 September 2022)). ^b pLDH method; data are the mean of three independent experiments ± SD; D10: CQ-susceptible *P. falciparum* strain; W2: CQ-resistant *P. falciparum* strain.

The assays were performed against W2 (chloroquine resistant) and D10 (chloroquine sensitive) strains of *P. falciparum*, using the parasite lactate dehydrogenase (pLDH) and chloroquine (CQ) as control.

Compounds of all four subclasses (**1–4**) that retained the natural configuration (5S, α S)—namely **1a**, **2a**, **3a**, and **4a**—were significantly more active than the other stereoisomers against both D10 and W2 *Plasmodium falciparum* strains ($IC_{50} < 1 \mu M$). Importantly, the equipotency observed against both the drug-sensitive (D10) and chloroquine-resistant (W2) strains suggests that these acivicin derivatives are not substrates for the common resistance mechanisms that affect chloroquine and other antimalarials. Their enantiomers (**1d**, **2d**, **3d**, and **4d**), bearing the (5R, α R) configuration, exhibited moderate antiplasmodial activity ($1 \mu M < IC_{50} < 10 \mu M$). Finally, the diastereomers with (5S, α R) and (5R, α S) configurations showed a marked reduction in antimalarial potency, with compounds displaying poor activity (see Table 1).

For the methyl ester derivatives, the enantiomer of **2a**, namely **2d**, with the configuration (5R, α R) was ten times less potent than the enantiomer with natural configuration with both D10 and W2 strains. The diastereoisomers **2b** and **2c** were inactive ($IC_{50} > 15 \mu M$). For the other compounds (**1a–d**, **3a–d**, **4a–d**) the same trend was observed, except for **1c**, with an $IC_{50} < 10 \mu M$ (Table 1).

In summary, the change in stereochemistry led to important differences in the antiplasmodial activity, with isomers with the natural configuration (5S, α S) being the most potent compounds.

Importantly, it can be excluded that the difference in the antiplasmodial potency of the different stereoisomers within the same subclass can be due

to a different passive diffusion across the cell membranes since the stereoisomers have the same calculated logD (cLogD; Table 1). Moreover, with the exception of **3a-d**, all compounds show cLogD values < 1 , not sufficient to cross cell membranes via passive diffusion. The natural compound acivicin (5S, α S) competes with amino acids for cellular uptake via transporters and it is likely that also our derivatives follow the same mechanism to get into the cell. Acivicin uptake is performed by the L-amino acids transport system which is in charge of the Na⁺-independent transport of neutral amino acids [15,16]. Two membranes control the microenvironment of the *P. falciparum*-infected human erythrocytes. The membrane of normal mature erythrocytes is impermeable to L-Gln and so also to acivicin while, when infected, the passage across the membrane is controlled by different transporter system with different substrate specificity [14,15]. Accordingly, our results suggest that the recognition of the compounds by the transport system might be influenced by the stereochemistry, not only for the 3-BA (**1a**), but also for its amide and ester derivatives (**2a-4a**). Indeed, the proliferation of *P. falciparum* was inhibited in a significant way only by the natural (5S, α S) isomers (**1a-4a**). Interestingly, when comparing the enzymatic and cellular data for the (5R, α R) enantiomers we can see that there is a discrepancy between enzymatic assays and cellular assays. This observation raises questions regarding their mechanism of action in the cellular context. One possibility is a potential lack of optical stability of the α -chiral center in the biological medium over the 72-hour assay, which could lead to partial epimerization to the active (S) configuration, although this was not experimentally assessed in this study. Alternatively, and perhaps more likely given the known promiscuity of the acivicin scaffold [4–7], these derivatives might

be acting on additional biological targets essential for parasite survival, thus compensating for the lack of PfGAPDH inhibition.

2.3. Conclusions.

The stereochemical variants **1b-d**, **2b-d**, **3b-d**, **4b-d** of 3-Br-acivicin (**1a**) and its derivatives **2a**, **3a**, and **4a**, were investigated to elucidate the role of stereochemistry in biological activity. While the natural (5S, α S) configuration proved essential for cellular uptake, this stereochemical requirement did not consistently correlate with PfGAPDH inhibition.

Docking studies provided an explanation for these results showing that the compounds can reach the PfGAPDH catalytic cysteine from opposite sides adopting two alternative binding modes, thereby reducing stereochemical requirements for enzyme inhibition. The observed variation in potency and stereoselectivity in inhibiting PfGAPDH depends on the nature of the substituent at C-5 and its ability to interact with specific arginine and threonine residues within the active site.

Conversely, the calculated low membrane permeability of the compounds suggests that cellular uptake may be mediated by the L-amino acid transport system, as previously demonstrated for acivicin. This mechanism could explain the stricter stereochemical requirements observed for antiparasmodial activity, where only the natural (5S, α S) isomers are active.

It has to be underlined that the antiparasmodial activity may not solely depend on PfGAPDH inhibition, pointing towards the involvement of additional biological targets. Future work will focus on investigating this hypothesis and on the structural optimization of the compounds, guided by the interaction models produced in this study.

2.4 References.

1. Nelson, A., and G. Karageorgis. "Natural Product-Informed Exploration of Chemical Space to Enable Bioactive Molecular Discovery." *RSC Medical Chemistry*, vol. 12, 2021, pp. 353-62.
2. A, N.L.B., et al. "Enantiomeric Mixtures in Natural Product Chemistry: Separation and Absolute Configuration Assignment." *Molecules*, vol. 23, no. 3, 2018, article 492.
3. Elder, F.C.T., et al. "The Role of Stereochemistry of Antibiotic Agents in the Development of Antibiotic Resistance in the Environment." *Environment International*, vol. 139, 2020, article 105681.
4. Miles, B.W., et al. "Inactivation of the Amidotransferase Activity of Carbamoyl Phosphate Synthetase by the Antibiotic Acivicin." *Journal of Biological Chemistry*, vol. 277, 2002, pp. 4368-73.
5. Denton, J.E., et al. "Rapid in Vivo Inactivation by Acivicin of CTP Synthetase, Carbamoyl-Phosphate Synthetase II, and Amidophosphoribosyltransferase in Hepatoma." *Life Sciences*, vol. 39, no. 9, 1986, pp. 817-24.
6. Earhart, R.H., and G.L. Neil. "Acivicin in 1985." *Advances in Enzyme Regulation*, vol. 24, 1985, pp. 179-205.
7. Williams, K., et al. "Crystal Structure of Acivicin-Inhibited Gamma-Glutamyltranspeptidase Reveals Critical Roles for its C-Terminus in Autoprocessing and Catalysis." *Biochemistry*, vol. 48, no. 12, 2009, pp. 2459-67.
8. Chittur, S.V., et al. "Mechanism for Acivicin Inactivation of Triad Glutamine Amidotransferases." *Biochemistry*, vol. 40, no. 4, 2001, pp. 876-87.
9. Galbiati, A., et al. "Covalent Inhibitors of GAPDH: From Unspecific Warheads to Selective Compounds." *European Journal of Medicinal Chemistry*, vol. 207, 2020, article 112740.
10. Conti, P., et al. "Synthesis and in Vitro/in Vivo Evaluation of the Antitrypanosomal Activity of 3-Bromoacivicin, a Potent CTP Synthetase Inhibitor." *ChemMedChem*, vol. 6, no. 2, 2011, pp. 329-33.
11. Bruno, S., et al. "Discovery of Covalent Inhibitors of Glyceraldehyde-3-Phosphate Dehydrogenase, a Target for the Treatment of Malaria." *Journal of Medicinal Chemistry*, vol. 57, no. 17, 2014, pp. 7465-71.
12. Cullia, G., et al. "Covalent Inhibitors of *Plasmodium falciparum* Glyceraldehyde 3-Phosphate Dehydrogenase with Antimalarial Activity in Vitro." *ACS Medicinal Chemistry Letters*, vol. 10, no. 4, 2019, pp. 590-95.
13. Moniot, S., et al. "Trapping of the Thioacylglyceraldehyde-3-Phosphate Dehydrogenase Intermediate from *Bacillus stearothermophilus*. Direct

Evidence for a Flip-Flop Mechanism." *Journal of Biological Chemistry*, vol. 283, 2008, pp. 21693-702.

14. Cook, W.J., et al. "An Unexpected Phosphate Binding Site in Glyceraldehyde 3-Phosphate Dehydrogenase: Crystal Structures of Apo, Holo and Ternary Complex of *Cryptosporidium parvum* Enzyme." *BMC Structural Biology*, vol. 9, 2009, article 9.
15. Huber, K.R., et al. "Uptake of Glutamine Antimetabolites 6-Diazo-5-Oxo-L-Norleucine (DON) and Acivicin in Sensitive and Resistant Tumor Cell Lines." *International Journal of Cancer*, vol. 41, no. 5, 1988, pp. 752-55.
16. Maeda, K., et al. "Effects of Acivicin on Growth, Mycotoxin Production and Virulence of Phytopathogenic Fungi." *Letters in Applied Microbiology*, vol. 59, no. 4, 2014, pp. 377-83.

Chapter 3. Tetrasubstituted Pyrrole Derivative Mimetics of Protein–Protein Interaction Hot-Spot Residues

3.1. Introduction.

One of the approaches of anticancer agents is to target PPIs engaged in the regulation of cell cycle [1]. In line with the conceptual framework outlined in Chapter 1, α -helical motifs can be considered as short “sentences”, in which hot spot residues function as key 'words' that define both meaning and biological function. The design of molecules capable of mimicking these residues represents an effort to create synthetic analogues that emulate the natural molecular interactions, which can be interpreted as part of a biochemical “grammar”.

This chapter reports the design, *in vitro* activity and SAR analysis of a second series of tetrasubstituted pyrrole (TSP) analogues able to mimic the spatial arrangement of three key hydrophobic residues ("hot spots") typical of an α -helix.

62% of PPIs in the Protein Data Bank (PDB) have an α -helix at the interface, making this structural element very important in protein-protein recognition [2,3]. This high prevalence of α -helices at PPI interfaces highlights their role as recurrent structural motifs. Important effort has been made in order to develop a non-peptide α -helix mimetics, able to mimic the key residue's spatial arrangement in α -helix-mediated PPIs. The identification of such ligands through conventional drug-discovery processes was difficult, nevertheless some successful examples have been reported [4-9] (Figure 1).

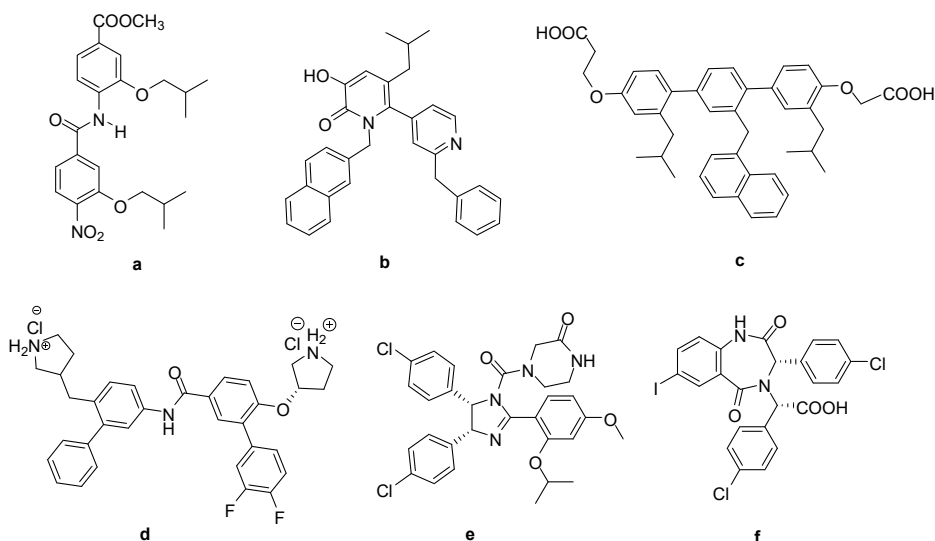


Figure 1. Chemical structure of representative anticancer PPIs mimics: **(a)** bis-benzamide derivative, AR/co-activator inhibitor; **(b)** pyridylpyridone derivative, ER/co-activator inhibitor; **(c)** terphenyl derivative, Bak/Bcl-xL inhibitor; **(d)** biphenyl carboxamide derivative, β -catenin/BCL9 complex inhibitor; **(e)** imidazoline derivative (nutlin-3a), p53/MDM2 inhibitor; **(f)** benzodiazepinedione derivative, p53/MDM2 inhibitor.

This α -helix-mediated PPIs shows substantial diversity, but hydrophobic residues are a majority of hot spot residues at the interface between helices. In all the compounds of Figure 1 there are common structural features: three aliphatic and/or aromatic groups positioned on a constrained chemical skeleton in such a way as to reproduce a specific tridimensional orientation of three hydrophobic residues on a α -helix. These aromatic moieties are associated with an increased affinity for the target sequence, probably due to the formation of π - π or π -alkyl interactions [10]. Molecules **a** and **b** were described as mimic of the i, i + 3 and i + 4 residues [4,5], whereas compounds **c** and **d** were described as mimic of i, i + 3 and i + 7 residues of helix-based PPI motifs [6,7]. The last two compounds, nutlin-3a (**e** in Figure 1) and **f** were shown to mimic the i, i + 4 and i + 7 residues of the MDM2 interacting motif on p53 protein $^{19}\text{FxxLWxxL}^{26}$.

The design of these mimetics relies on a scaffold hopping approach around various privileged scaffolds. By replicating the hydrophobic and stereochemical distribution of natural residues, these diverse structures are able to 'speak' the same molecular language used by proteins at their interaction interfaces. In this context, it was identified a tetrasubstituted pyrrole (TPS) derivative, namely 4-benzoyl-*N*-[(3-chlorophenyl)methyl]-5-methyl-1-[(4-methylphenyl)methyl]-1*H*-pyrrole-2-carboxamide (**1**, Figure 2), as the lead structure capable of mimic helix-based hydrophobic protein-recognition motifs [11].

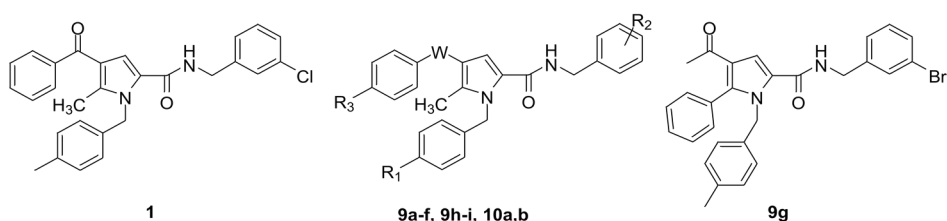


Figure 2. Representation of 4-benzoyl-*N*-(3-chlorobenzyl)-5-methyl-1-(4-methylbenzyl)-1*H*-pyrrole-2-carboxamide (**1**) and the new TSPs **9a-i** and **10a,b**.

These TSP compounds aim to mimic the spatial arrangement of three key hydrophobic residues ("hot spots") typical of an α -helix. The ability to replicate the hydrophobic and stereochemical distribution of these natural residues is what allows the TSP scaffold to "speak the same molecular language" used by proteins for interactions. TSPs, acting as synthetic privileged scaffolds, interfere with crucial pathological pathways.

Compound **1** displayed *in vitro* anticancer activity against M14 and A375 melanoma cell lines. Moreover, this compound showed a synergetic effect when used with nutlin-3a, suggesting that they do not compete for the same binding site but rather have effect on the same pathway, reducing p53-MDM2 complex formation [11,12]. In order to expand the structure-

activity relationship (SAR) of this class of peptidomimetics, a second series of TSP analogues (Compounds **9a-i**, **10a,b**; Figure 2) was developed among which it was identified **10a** as the most potent of the present and previous series of TSPs.

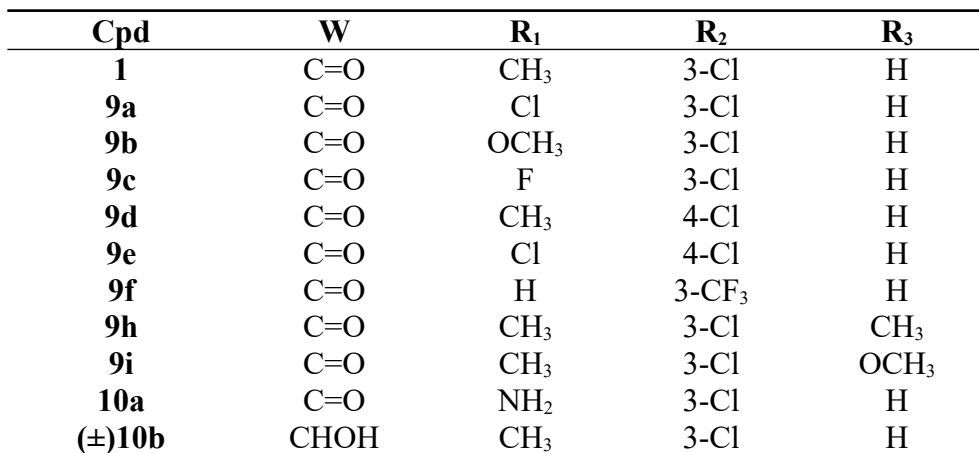
This study was performed in the context of a multidisciplinary collaboration involving the research groups of: i) Prof. Anna Ramunno (University of Salerno) for the synthesis, ii) Prof. Michela Varra (University of Naples “Federico II”) for the structural characterization and iii) Prof. Carlo Irace (University of Naples “Federico II”) for the biological assays. Moreover, this study was supported by a grant from Regione Campania-POR Campania FESR 2014/2020 “Combattere la resistenza tumorale: piattaforma integrata multidisciplinare per un approccio tecnologico innovativo alle oncoterapie-Campania Oncoterapie” (Project N. B61G18000470007) and is the object of the scientific publication reported in ref [13].

I was directly involved in performing and analyzing computational studies as well as in the interpretation of the experimental results (particularly SAR analysis). I participated in the meetings of the research groups involved in the study and I attended schools and congresses related to this research topic, in particular School of COST action “MedChem and structural biology: tools and strategies for hit and lead optimization in the One Health perspective” held at University of Siena, where this study was object of an oral presentation.

The synthesis and the procedure of experimental studies is not herein reported but can be found in the published manuscript [13]

3.2.1. Design of the New TSP Derivatives.

Table 1. Representation of 2D structures of compounds **1**, **9a-i**, and **10a, b**.



A loss of activity was observed when the Z ring was substituted with a methyl group as well as when the length of the methylene linker that connects the amide bond to the Y ring was modified or when the Y ring was substituted with a cyclohexane. Moreover, some more information

was obtained about the substitution of Y and X rings. In particular, compounds resulted inactive when the R₁ group was a bulk electron-withdrawing group such as NO₂ and CF₃. On the contrary, for the Y ring, the *meta*-substitution with a halogen resulted in active compounds.

Strating from this information, the exploration of the TSP skeleton was continued by changing the structure of the reference compound **1** (Table 1).

The exploration of the *para* position of the X ring was continued by changing the nature (electron-withdrawing/electron-donating) and the size of R₁ group (i.e Cl, OCH₃, F and NH₂; **9a-c** and **10a**; Table 1). Moreover, **1** was modified at both R1 and R2 to obtain additional combinations of *para*-substituted X ring and *meta*-substituted Y ring (i.e R₁: *p*-Cl and R2: *p*-Cl (**9e**); R1: H and R2: *m*-CF₃(**9f**)). The *para* position of the Y ring was investigated by shifting the meta chlorine atom of **1** to the *para* position (**9d**) while the Z ring of **1** was modified with two electron-donating groups at the *para* position (i.e CH₃ and OCH₃; **9h** and **9i** respectively.) The Z ring was also exchanged with a methyl group at position 5 (**9g**). Moreover, the reduction of the carbonyl group at position 4 of the pyrrole ring (W) to hydroxyl group ((±)**10b**) provided an additional hydrogen bond interaction with the target and modified the geometry of the carbon atom from planar to tetrahedral.

3.2.2. Solubility Assay.

The solubility of all compounds was tested in a PBS solution (pH = 7.4) by means of UV spectroscopy. The calculated solubility values of all the compounds are displayed in Table 2.

Table 2. Solubility (μM) in a PBS solution ($\text{pH} = 7.4$) of compounds **1**, **9a-i**, **10a,b**, and nutlin-3a.

Cpd	Solubility ^a (μM)
1	3.5 ± 0.1
9a	3.9 ± 0.1
9b	3.1 ± 0.2
9c	3.9 ± 0.2
9d	3.1 ± 0.2
9e	3.5 ± 0.3
9f	2.6 ± 0.1
9g	1.5 ± 0.2
9h	<1.0
9i	<1.0
10a	10.5 ± 0.1
($\pm$)10b	6.7 ± 0.6
nutlin-3a	2.4 ± 0.2

^a data are reported as the mean $\pm$ SEM.

The values of solubility for compounds **9a-i** were very close to the values of reference compound **1** ($3.5 \mu\text{M}$) and nutlin-3a ($2.4 \mu\text{M}$; Table 2). Instead, compound **10a** showed values of solubility 5.2 and 2.8-fold higher than nutlin-3a and **1**, respectively. **10a** was the most soluble among tested compounds, followed by the racemic mixture of **($\pm$)10b**. The remaining compounds, including nutlin-3a, were less soluble, with values ranging from 1.5 to $3.9 \mu\text{M}$.

3.2.3. Biological evaluation.

In order to evaluate the biological effects of compounds **1**, **9a-i** and **10a-b**, human malignant cells with high replicative potential *in vitro*, such as colorectal carcinoma HCT-116 cells and human melanoma p53wt A375 cells were used. Moreover, selective toxicity against healthy cell line L6 was also investigated. The ability of compounds **1**, **9a-i**, and **10a,b** to

inhibit cell growth was evaluated, comparing them with the reference compound nutlin-3a. It is important to note that for biological assays, compounds were dissolved in DMSO to ensure complete solubilization before dilution in the culture medium. Cells were incubated for 48 h (typically 1.5–25 μM , extended up to 100 μM for less active compounds to allow accurate IC_{50} determination), and IC_{50} values were calculated by concentration/effect curves based on the estimation of a "cell survival index", a parameter derived concurrently from the evaluation of the cellular metabolic activity and the live/dead cell ratio (Table 3).

Table 3. IC_{50} values (μM) and selectivity index (SI) of compounds **1**, **9a-i**, **10a,b**, and the positive control nutlin-3a against human cancer (A375 and HCT-116) and healthy (L6) cell lines following 48 h of incubation.

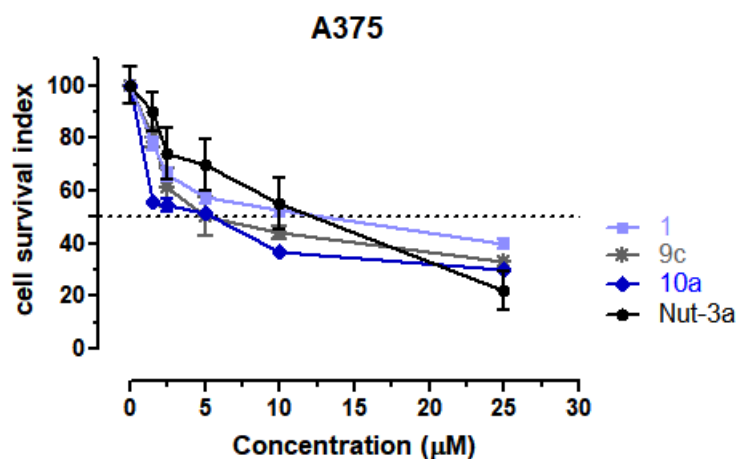
Cpd	A375 ^a IC_{50} (μM) ^b	L6 ^c IC_{50} (μM) ^b	SI ^d	HCT-116 ^e IC_{50} (μM) ^b
1	20 $\pm$ 4	42 $\pm$ 3	2.1	41.4 $\pm$ 6
9a	18 $\pm$ 4	48 $\pm$ 5	2.7	36.4 $\pm$ 8
9b	24 $\pm$ 14	40 $\pm$ 5	1.7	46 $\pm$ 18
9c	13 $\pm$ 4	35 $\pm$ 4	2.7	55.9 $\pm$ 7
9d	15 $\pm$ 5	25 $\pm$ 6	1.7	35.4 $\pm$ 6
9e	23 $\pm$ 6	38 $\pm$ 2	1.6	34.6 $\pm$ 6
9f	27 $\pm$ 8	35 $\pm$ 7	1.3	28.4 $\pm$ 7
9g	18 $\pm$ 5	71 $\pm$ 5	3.9	34 $\pm$ 6
9h	24 $\pm$ 6	42 $\pm$ 7	1.7	56.3 $\pm$ 15
9i	16 $\pm$ 6	66 $\pm$ 5	4.2	167.6 $\pm$ 9
10a	10 $\pm$ 4	55 $\pm$ 7	5.6	41.3 $\pm$ 6
($\pm$)10b	15 $\pm$ 5	54 $\pm$ 6	3.6	25 $\pm$ 5
nutlin-3a	15 $\pm$ 3	18 $\pm$ 4	1.2	12.5 $\pm$ 5

^a Human melanoma cell line. ^b IC_{50} values are expressed as mean values $\pm$ SEM ($n=30$). ^c

Rat myoblast cell line. ^d $\text{SI} = \text{IC}_{50} \text{ L6} / \text{IC}_{50} \text{ A375}$. ^e Human colon cancer cell line.

This “cell survival index” displays a concentration-dependent sigmoid trend, with IC_{50} values in the low micromolar range (Figure 1).

(A)



(B)

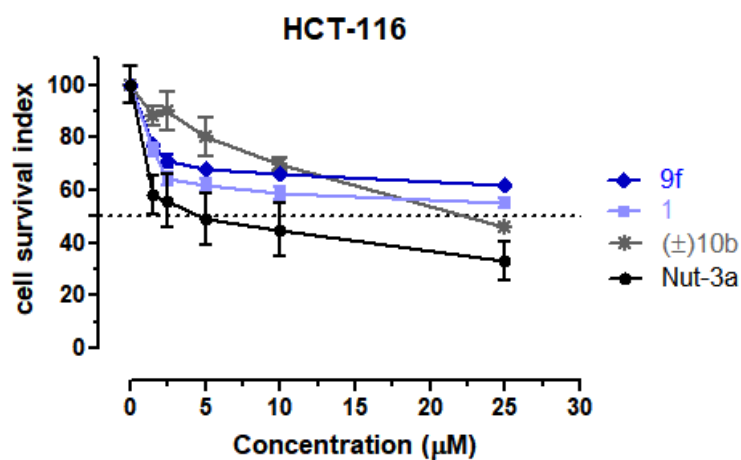


Figure 3. Cell survival index, evaluated by the MTT assay and the estimation of live/dead cell ratio, following 48 h of incubation with the indicated concentrations (the range 1→25 μM has been explored) of compounds **1**, **9c**, and **10a** for A375 cells (A), and of compounds **9f**, **1** and ($\pm$)**10b** for HCT-116 cells (B), respectively, as indicated in the legends. Nutlin-3a (**Nut-3a**) was used as positive control for cytotoxic and antiproliferative effects. Data are expressed as percentage of untreated control cells and are reported as mean of five independent experiments $\pm$ SEM ($n = 30$).

Overall, as shown in Table 3, all tested compounds exhibited higher activity against the melanoma A375 cell line compared to the colon cancer HCT-116 line. The value of IC_{50} for the compound tested against the melanoma A375 cell line resulted in the range from 10 to 27 μM , consistent with the value of nutlin-3a ($15 \pm 3 \mu M$; Table 3). When tested against HCT-116 cells, new TSP derivative **9i** was inactive ($IC_{50} > 100 \mu M$), while the rest of the compounds showed IC_{50} values ranging from 25 to 56 μM , higher than that of nutli-3a ($IC_{50} = 12.5 \pm 5 \mu M$; Table 3).

For the healthy L6 cell line, all compounds displayed IC_{50} values ranging from 35 to 72 μM . The only exception is **9d** ($IC_{50} = 25 \pm 6 \mu M$). The selectivity index respect to A375 was between 1.3 and 5.6, higher than that shown by nutlin-3a (i.e. 1.2; Table 3). The SI was specifically calculated for the A375 cell line because melanoma represents the primary therapeutic target for this class of compounds. This cell line is characterized by wild-type p53 (p53wt), making it the ideal model to evaluate the efficacy of molecules designed to interfere with the p53-MDM2 pathway, unlike HCT-116 which served as a comparative cancer cell model. In particular, there was no important difference between the newly synthesized pyrrole derivatives and lead **1** in terms of IC_{50} of the A375 cell line. The exception was i) **9c** ($IC_{50} = 13 \pm 4 \mu M$), with a strong electron-withdrawing and potential hydrogen bond acceptor substituent such as the fluorine atom, and ii) **10a** ($IC_{50} = 10 \pm 4 \mu M$), with an electron-donating and hydrogen bond donor/acceptor substituent (i.e the amino group) at the same position (Tables 1 and 3). This latter was the most active derivative among the tested molecules, slightly more active than nutlin-3a (Table 3).

These results indicate that the mechanism of action of these new TSP derivatives is similar to that of reference compound **1** and is based on the interference with a specific molecular pathway, that ultimately led to apoptosis and is not depending on unspecific cytotoxicity.

Moreover, apoptosis studies revealed that these compounds are able to kill the cancer cell by activating the apoptotic way of cell death (for more details see ref. [13])

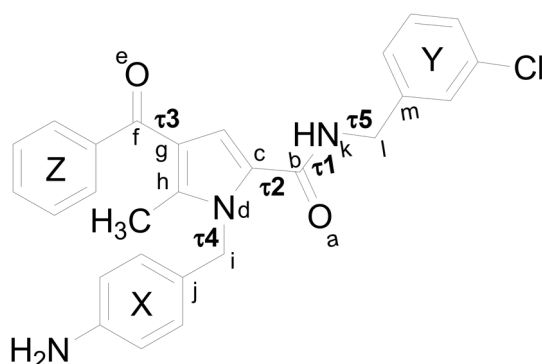
3.2.5. Computational studies.

3.2.5.1 Conformational Analysis.

The conformational properties of compounds **9a-i**, **10a** and ($\pm$)-**10b** were evaluated through force-field based and quantum mechanical computational methods. Compound **1** was included in the analysis of conformational properties in order to estimate the effects introduced by the structural modifications. Indeed, the first series of TSPs was analysed by applying in vacuum semi-empirical PM6 [14] calculations [11] while with the new series the method applied was the density functional theory (DFT) [15] calculations and, to mimic an aqueous environment, the conductor-like polarizable continuum model (C-PCM)[16] was used. This transition was driven by the need for higher accuracy in describing the electronic effects introduced by specific substituents (e.g., F, Cl, OMe).

The results of conformational analysis displayed two set of specular conformers for each compound, characterized by same conformational energy and the same absolute value of torsion angles but with opposite site (i.e conformational enantiomers, named E1 and E2; Table 4 and Tables 1-23 in Appendix B; Figure 4)

Table 4. Conformational energy, structural classification, and pharmacophore distances (Å) of calculated DFT conformers of **10a**.



Family ^a	Sub-Family	ΔE_{GM} (kcal/mol)	Distances (Å)			Torsional Angles (°) ^{b-f}				
			d1 (X-Y)	d2 (X-Z)	d3 (Y-Z)	$\tau 1$	$\tau 2$	$\tau 3$	$\tau 4$	$\tau 5$
TCC	I	0.00	6.4	8.7	9.3	-174	26	-21	77	-112
TCC	II	0.49	5.6	9.3	10.2	-174	28	22	75	-82
TCT	I	1.57	7.4	7.2	11.0	-174	27	-148	102	-111
TCT	III	1.65	9.7	7.2	10.7	-173	27	-148	101	123

^a The data reported refer to the enantiomer E1; the other enantiomer presents the same conformational energy, the same pharmacophore distances, and the same absolute torsion angle values but with the opposite sign. ^b $\tau 1$ torsional angle is calculated considering the amide bond atoms. ^c $\tau 2$: a, b, c, and d atoms. ^d $\tau 3$: e, f, g, and h atoms. ^e $\tau 4$: c, d, i, and j atoms. ^f $\tau 5$: b, k, l, and m atoms.

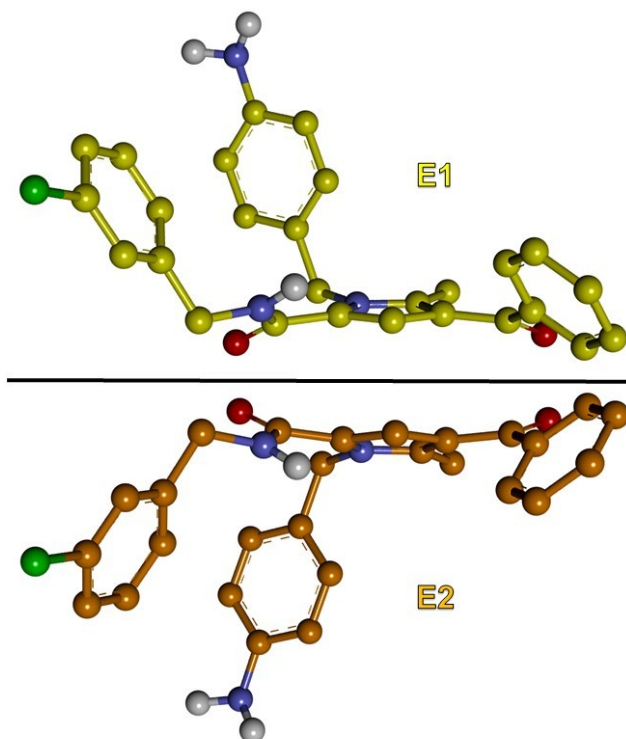


Figure 4. Comparison of the conformational enantiomers of the global minimum conformer of compound **10a**. Structures are displayed as ball and sticks. Heteroatoms are colored by atom type: O, red; N, blue and Cl, light green. Hydrogen atoms are omitted for clarity with the exception of those bonded to the amide and amine nitrogen.

In the case of **10b**, a racemic mixture, the specular conformers corresponded to those of the enantiomer with the opposite configuration.

The conformers of **1**, **9a-i** and **10a** with low energy ($\Delta E_{\text{GM}} < 2$ kcal/mol) were clustered in two families, named TCC and TCT, according to the antiperiplanar (T) or synperiplanar (C) conformation of the torsion angles τ_1 , τ_2 and τ_3 (Figure 5 and Table 4).

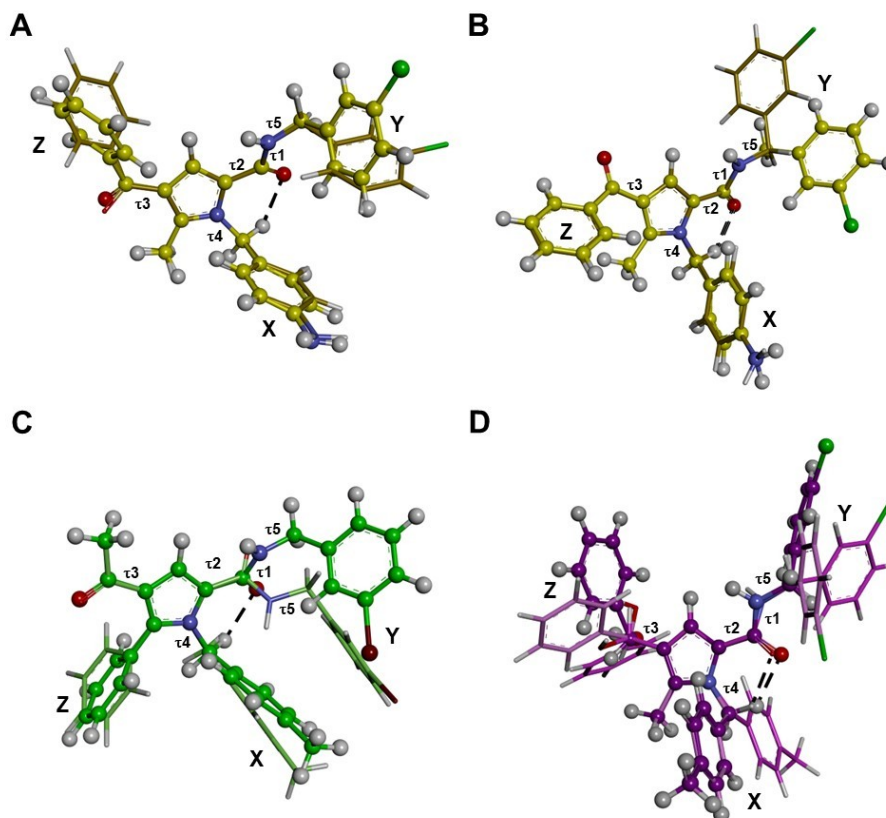


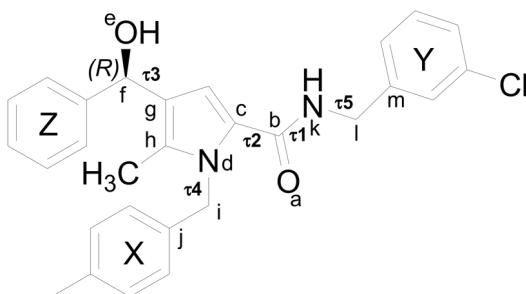
Figure 5. TSP conformers: TCC I (yellow) and TCC II (dark yellow) (**10a**; A); TCT I (yellow) and TCT III (dark yellow) (**10a**; B); TCC I (green) and TTC IV (light green) (**9g**; C); TCG⁻ (violet), TCA⁻ (pink) and TCG⁺ (magenta) (**R-10b**; D). The lowest energy conformers of the family are displayed in ball&stick and the others (if any) in stick. The τ_1 - τ_5 torsion angles and the pharmacophoric rings are labelled. Heteroatoms are colored by atom type: O, red; N, blue; Cl, green; Br, brown. The intra-molecular hydrogen-bond interactions are evidenced with a black dashed line.

The TCC family is favored with respect to the TCT family in terms of both conformer population (~28% vs ~25%) and conformational energy.

In compound ($\pm$)-**10b**, the hybridization of the τ_3 torsion angle is changed from sp^2 to sp^3 since the carbonyl group at C4 is reduced to a hydroxyl function. This compound has the same conformational preferences as for

1, **9a-i** and **10a** for the torsion angles τ_1 and τ_2 (family TC) but with three different orientations of the Z ring (i.e. three values corresponding to the G-, A-, and G+; Table 5, Figure 6D and Figure 7).

Table 5. Conformational energy, structural classification, and pharmacophore distances (Å) of calculated DFT conformers of **(R)-10b**.



Famil y ^a	Sub- family	ΔE_{GM} (kcal/mol)	Distances (Å)			Torsional Angles (°) ^{b-f}				
			d1 (X-Y)	d2 (X-Z)	d3 (Y-Z)	τ_1	τ_2	τ_3	τ_4	τ_5
TCG ⁻	III	0.00	9.6	8.3	7.9	-173	23	-85	95	12 1
TCA ⁻	I	0.60	7.0	7.1	9.5	-172	23	-145	95	-1 03
TCG ⁺	III	0.89	7.3	7.0	10.3	174	-22	64	-94	11 2

^aThe data reported are referred to the enantiomer R, the enantiomer S presents the same conformational energy, the same pharmacophore distances, and the same absolute torsion angle values but with opposite sign. ^b τ_1 torsional angle is calculated considering the amide bond atoms. ^c τ_2 : a, b, c, and d atoms. ^d τ_3 : e, f, g, and h atoms. ^e τ_4 : c, d, i, and j atoms. ^f τ_5 : b, k, l, and m atoms.

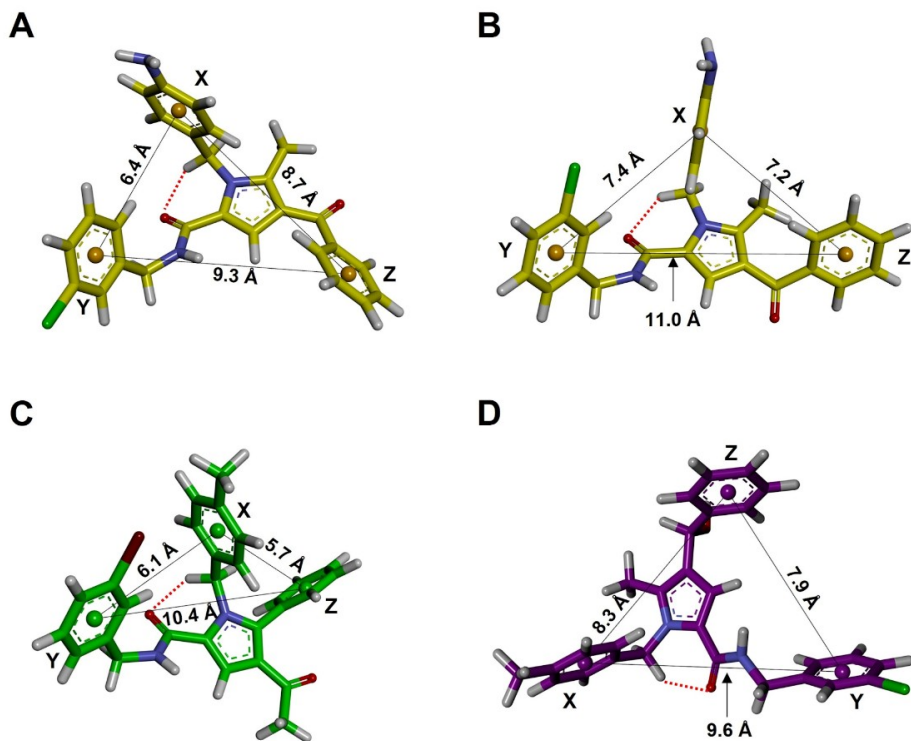


Figure 6. DFT lowest energy conformers of TSP conformational families: TCC I (**10a**; (A)), TCT I (**10a**; (B)), TCC I (**9g**; (C)), and TCG-III (R-**10b**; (D)). The intra-molecular distances between the X, Y, and Z rings are reported.

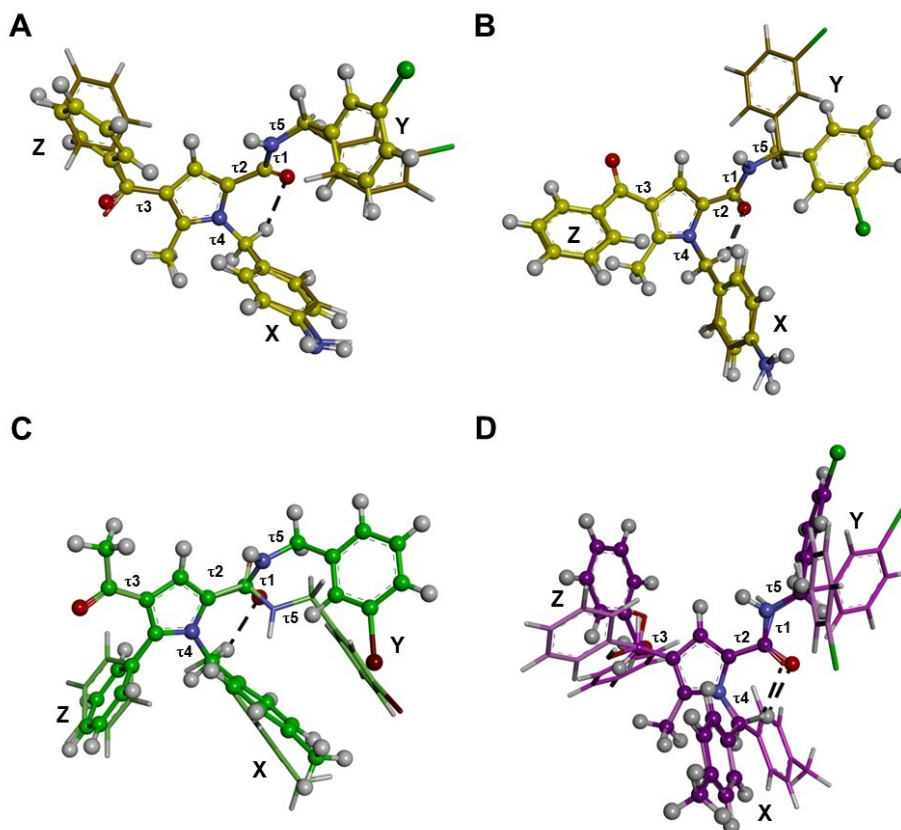
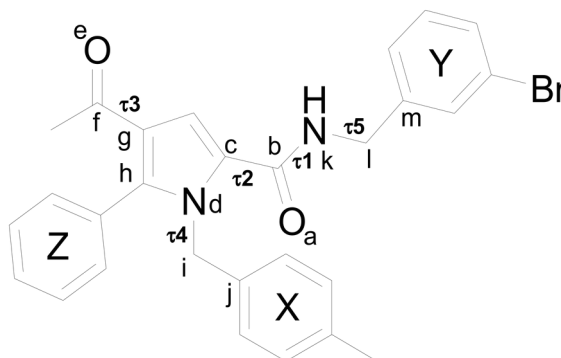


Figure 7. TSP conformers: TCC I (yellow) and TCC II (dark yellow) (**10a**; A); TCT I (yellow) and TCT III (dark yellow) (**10a**; B); TCC I (green) and TTC IV (light green) (**9g**; C); TCG⁻ (violet), TCA⁻ (pink) and TCG⁺ (magenta) (**R-10b**; D). The lowest energy conformers of the family are displayed in ball&stick and the others (if any) in stick. The τ_1 - τ_5 torsion angles and the pharmacophoric rings are labelled. Heteroatoms are colored by atom type: O, red; N, blue; Cl, green; Br, brown. The intra-molecular hydrogen-bond interactions are evidenced with a black dashed line.

For the low energy conformers of **9g**, the values of torsion angles τ_1 and τ_2 were corresponding to the families TC (most favored also in this case) and TT. In this compound, the Z ring is moved from the carbonyl group at C4 (whose orientation is defined by τ_3) to C5. In this case, the τ_3 torsion angle

has always a synperiplanar (C) conformation respect to C5. (Table 6, Figures 6C and 7).

Table 6. Conformational energy, structural classification, and pharmacophore distances (Å) of calculated DFT conformers of **9g**.



Family ^a	Sub-family	ΔE_{GM} (kcal/mol)	Distances (Å)			Torsional Angles (°) ^{b-f}				
			d1 (X-Z)	d2 (X-Z)	d3 (Y-Z)	$\tau 1$	$\tau 2$	$\tau 3$	$\tau 4$	$\tau 5$
TCC	I	0.00	6.1	5.7	10.4	-17 3	27	-2	71	-107
TTC	IV	1.91	4.5	5.4	9.2	165	154	3	90	-176

^a The data reported are referred to the enantiomer E1, the other enantiomer presents the same conformational energy, the same pharmacophore distances, and the same absolute torsion angle values but with opposite sign ^b $\tau 1$ torsional angle is calculated considering the amide bond atoms. ^c $\tau 2$: a, b, c, and d atoms. ^d $\tau 3$: e, f, g, and h atoms. ^e $\tau 4$: c, d, i, and j atoms. ^f $\tau 5$: b, k, l, and m atoms.

Importantly, all compounds showed not only the $\tau 1$ and $\tau 2$ torsion angles (TC family) in the same conformational preference, but also very similar values of $\tau 4$ (~90°), likely due to the formation of intra-molecular hydrogen bond between the X ring and the carbonyl group of the amide function (Figures 6-7). On the other hand, for the torsion angle $\tau 5$ four conformations were found: i) A⁻ (-90° to -150°; subfamily I), ii) G⁻ (30° to

90°; subfamily II, iii) A⁺ (90° to 150°; subfamily III), and T (±150° to 180°; subfamily IV) (Figure 7; Table 4, 5, 6 and Tables 13-21 in Appendix B).

Taken together, these results evidenced that the new TSP's low energy conformers as well as those of **1** can be grouped into two families, which are defined by similar values of τ_1 , τ_2 and τ_4 , but differing for τ_3 and τ_5 .

Comparing the previously obtained results [11] some important differences can be noted: i) the PM6 low-energy conformers previously calculated comprised the *cis* conformation of the amide bond (τ_1) and ii) the distance between ring X and Y resulted decreased in the conformations belonging to the TCC and TCT families compared to DFT results (Figure 8), probably due to overestimation of intra-molecular π - π interaction and the absence of solvent.

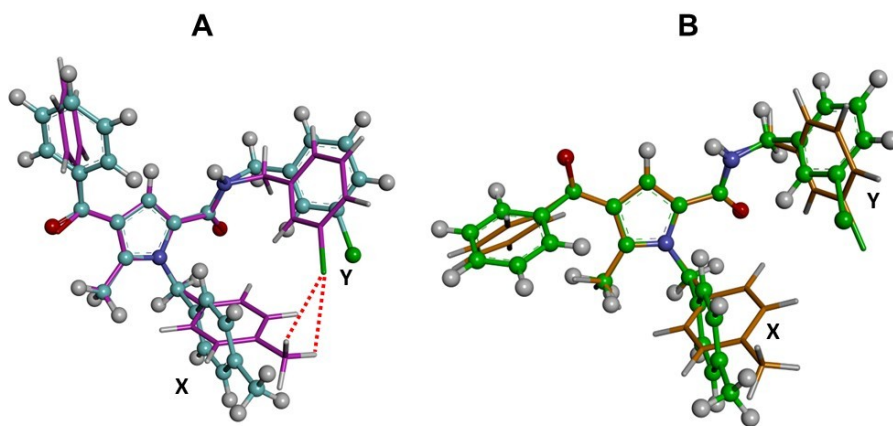


Figure 8. A: Comparison between the PM6 (stick and carbon atoms in magenta) and the DFT (ball&stick and carbon atoms in light blue) TCC I conformer of **1**. B: Comparison between the PM6 (stick and carbon atoms in orange) and the DFT (ball&stick and carbon atoms in green) TCT I conformer of **1**. Heteroatoms are colored by atom type: O, red; N, blue; Cl, light green. The intra-molecular hydrogen-bond interactions between methyl group and chlorine atom are evidenced with red dashed lines.

3.2.5.2. Investigation of the Peptidomimetic Mimicry Ability of the New TSPs.

The peptidomimetic ability of compounds **1**, **9a-i**, **10a**, and ($\pm$)-**10b** was studied by comparing the low-energy DFT conformers with experimentally determined structures of hydrophobic PPI motifs (LxxLL and coiled coil). This was done in order to identify the consensus structure of the helix-based hydrophobic motif(s) (i.e the spatial arrangement of the hot spot residues) that can be mimicked by TSPs.

The TSP conformers were clustered by their intra-molecular distances (Table 7), the PPI hot spot residues were identified, and a comparison was made between the distances of the centroids of the X, Y, and Z rings (Table 4 and Tables 13-21 in Appendix B).

Table 7. Distances between the centroids of the side chains of the hot-spot residues of the PPI motifs considered.

Motifs	PDB ID	Distances (Å)					
		$i, i+3$	$i, i+4$	$i+3, i+4$	$i, i+7$	$i+4, i+7$	$i+3, i+7$
²³ VxxLxxxV ³⁰ (Parallel)	2ZTA	6.6	-	-	10.1	-	7.7
¹⁴ LxxxLxxL ²¹ (Antiparallel)	1GRJ	-	7.6	-	10.2	6.8	-
³⁴⁵ LxxLL ³⁴⁹	2J7X	6.9	5.2	7.2	-	-	-

When the distance similarity criteria between the conformers and the PPI motifs matched, a superimposition was made on the PPI motifs by fitting the centroids of the side chains of the hot spot residues and the centroids of the aromatic rings X, Y and Z (Table 8).

The conformers with a calculated root-mean-square distance (RMSD) value < 2.5 Å were selected as mimetic of the PPI motif.

Table 8. Calculated RMSD (Å) by fitting the centroids of the aromatic rings X, Y, and Z of the DFT conformers of TSPs on the centroids of the side chains of the considered PPI motif.

Pharmacophore	Compound	Conformer	RMSD (Å) ^a			
			<i>i i+3 i+7</i> ²³ V _{xx} L _{xxx} V ₃₀		<i>i i+4 i+7</i> ¹⁴ L _{xxx} L _{xx} L ₂₁	
			Parallel coiled coil PDB ID: 2ZTA		Antiparallel coiled coil PDB ID: 1GRJ	
1	10a	TCC I	1.94	2.25	2.18	1.95
2	10a	TCT I	1.80	2.08	1.65	2.12
3	9g	TCC I	2.15	2.42	2.35	2.40
4	10b	TCG ⁻	1.99	1.82	1.74	2.25

^aThe RMSD values calculated for the conformational enantiomer E2 or for the configurational enantiomer S of **10b** are evidenced in bold

These studies led to the identification of four different pharmacophores (Figure 6) that are able to reproduce the intra-molecular distances between hydrophobic hot spot residues of hydrophobic PPI engaged in the formation of parallel and antiparallel coiled coil structures (Table 9).

Notably, the spatial orientation of the pharmacophoric moieties of TSP (i.e., the X, Y, and Z rings), along with the presence of conformational enantiomers, appears to mirror specific structural features of PPI motifs involved in the formation of both parallel and antiparallel coiled-coil structures. In these arrangements, the *i*+3 and *i*+4 hot spot residues are symmetrically positioned relative to the helix axis and to other key residues (*i* and *i*+7), thereby enabling the selective recognition of distinct

faces of the α -helix in the two PPI types (parallel and anti-parallel coiled coil structures; Figure 9).

Table 9. Correspondence between the aromatic rings of TSP derivatives and the interacting residues of the considered PPI motifs.

Pharmacophore	TSP Ring	PPI Motif			
		²³ VxxLxxxV ³⁰		¹⁴ LxxxLxxL ²¹	
		Hot-Spot Residues ^a			
1	X	i + 3	i + 3	i + 4	i + 4
	Y	i	i + 7	i	i + 7
	Z	i + 7	i	i + 7	i
2	X	i + 3	i + 3	i + 4	i + 4
	Y	i	i + 7	i	i + 7
	Z	i + 7	i	i + 7	i
3	X	i + 3	i + 3	i + 4	i + 4
	Y	i	i + 7	i	i + 7
	Z	i + 7	i	i + 7	i
4	X	i	i + 7	i	i + 7
	Y	i + 7	i	i + 7	i
	Z	i + 3	i + 3	i + 4	i + 4

^a The PPI residues mimicked by the conformational enantiomer E2 or by the configurational enantiomer S of **10b** are evidenced in bold.

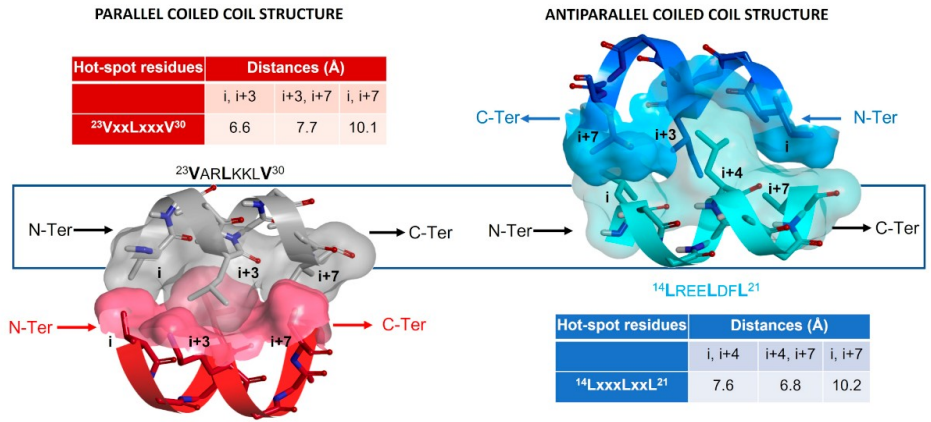


Figure 9. Parallel (left) and antiparallel (right) coiled coil structures (PDB: 2ZTA and 1GRJ, respectively). The two interacting coiled coil motifs are represented as ribbons, the heavy atoms and the polar hydrogens of the hot-spot residues are displayed as sticks and

colored by atom type, the residue position is labeled, and the solvent accessible surface is shown. A dark blue frame evidences the coiled coil motifs mimicked by the TSPs; the sequence of the motif is reported together with the distances between the hot-spot residues (red and blue tables).

The conformational/configurational enantiomers of TSP derivatives (Figure 10) have a similar structural relationship, and this can contribute to their peptidomimetic ability.

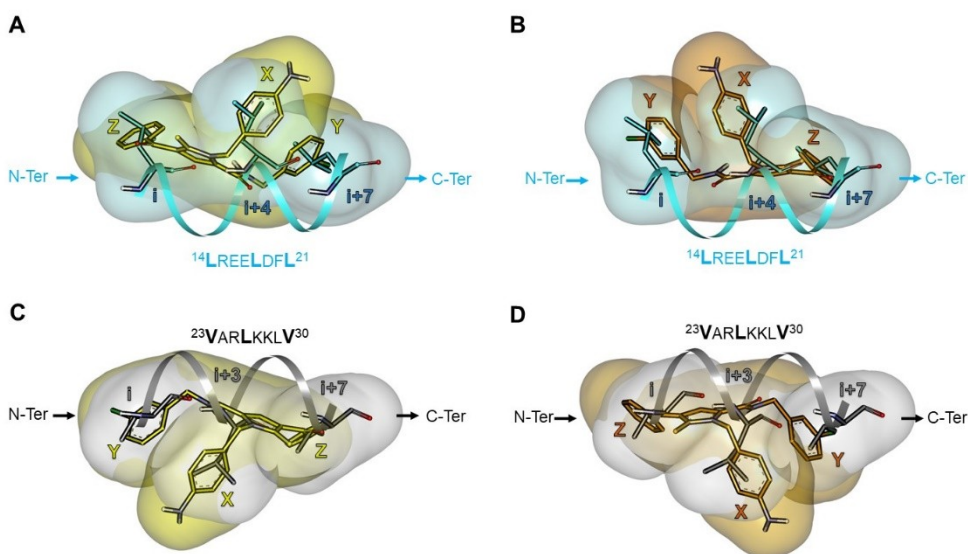


Figure 10. TCC I conformer of **10a** (i.e., pharmacophore 1) superimposed on the hot-spot residues of coiled coil motifs involved in the formation of parallel and antiparallel PPI. E1 conformational enantiomer (yellow) superimposed on the “antiparallel” (cyan; (A)) and parallel (white; (C)) coiled coil motif; E2 conformational enantiomer (orange) superimposed on the “antiparallel” (cyan; (B)) and parallel (white; (D)) coiled coil motif. The centroids of the X, Y, and Z rings are superimposed on the centroids of the side chains of the hot-spot residues. The heavy atoms and the polar hydrogens of **10a** and the hot-spot residues are displayed as sticks and colored by atom type. Molecular volumes are showed as soft surface.

The central hot spot residue of the motif (i.e. $i + 3$ or the $i + 4$ in parallel or antiparallel coiled coil structure, respectively) was mimicked by the X

ring, while Y and Z had the longest intra-molecular distance (~ 10 Å; Figure 6A-C and Table 7). The hot spot residues i or $i + 7$ can be mimicked by the Z and Y ring alternatively depending on: i) the enantiomer when comparing the two enantiomer to the same PPI motif (Figure 10A vs. Figure 10B and Figure 10C vs. Figure 10D) or ii) on the PPI motif when comparing the two motifs to the same enantiomer (Figure 10A vs. Figure 10C and Figure 10B vs. Figure 10D). The distance between the X and Y ring is the longest only in the case of one pharmacophore of ($\pm$)-**10b** (n.4 in table 9). For this reason, hot spot residue i or $i + 7$ are alternately mimicked by the X and Y ring, while the central hot spot residue of the motif is mimicked by the Z ring.

This mechanism of action of compound **1** is confirmed by the pharmacological studies, since they evidenced that the anticancer activity of TSPs is connected to the activation of apoptotic cell-death pathways [12] and in line with the computational results, the same mechanism of action is proposed for the new TSPs. The melanoma viability was impaired by **1** in the previous pharmacological investigation on melanoma cells (M14 [11] and A375 [12]) by its ability to reduce p53-MDM2 interaction and to restore p53 transcriptional activities and functions. Despite p53 is not frequently mutated in melanoma, the inactivation of p53 can be reached by the deletion of CDKN2A (Cyclin-Dependent Kinase Inhibitor 2A) locus that encodes for p14^{ARF}, that in turn leads to a stable activation of MDM2 [17].

p53 is degraded with the formation of p53-MDM2 complex which in turn is regulated by p14^{ARF}, a tumor suppressor. The 37 N-terminal amino acid portions of p14^{ARF} binds and sequesters MDM2 within nucleoli causing cell cycle arrest, but in a different region from that necessary for the

binding of p53 [18]. To validate the hypothesis that p14^{ARF} could be mimicked by the TSP derivatives in its interaction with MDM2, a bioinformatic and structural analysis was performed on the two first helices of p14^{ARF} (α H1 (4-14) and α H2 (20-29)) by using the x-ray structure of the mouse p19^{ARF} tumor suppressor protein (PDB ID: 1HN3) (Figure 11).

			α -helix H1	α -helix H2		
p19 ^{ARF} mouse	1	MGR	RFLVTIVRIQRA	RP	Q	ERVFLVKFVRSRRPRTASCALAFVNMLLRLEILRRGPHRNP-GPGDDDGQ 69
p14 ^{ARF} human	1	MVR	RFLVTILRIRRA	SP	PR	VRVFVVIHPIRTGEWAAPGAPAAVALVLMLLRSQRLGQQPLPRRPGHDDGQ 70
Consensus aa:		Ms	RRFLVTIRIP	RAt.P.p.	RVFIV+hSR.p...	hAstA.AhVshIL.LbR.bR.G.p...PGcDDGQ
p19 ^{ARF} mouse	70	RS	SSSSAQLRCR	FELRGPHYLL	PPGARRSAGRL	FGHAGGAARVSGSAGCARCLGSPAARLGPRAGTSRH 139
p14 ^{ARF} human	71	RP	SGGAAAPRRGA	QLRRPRHSHPT	RARRCPGGL	FGHAGGAAPGRGAAGRARCLGPSARGPG----- 132
Consensus aa:		R	sp	ttttA..R..h	pLR.P+g.hPs.ARR	tsG.LFGHAGGAA.sRGtAG.ARCLGssA...G.....
p19 ^{ARF} mouse	140	RA	IFAFRWVLFV	FRWVVFVIR	WERRPDRRA	169
p14 ^{ARF} human		-----				
Consensus aa:						

Figure 11. Sequence alignment of the human p14^{ARF} and the mouse p19^{ARF} tumor suppressor protein. The first two N-terminal helices of p14^{ARF} (H1 (4-15) and H2 (20-30)) involved in the interaction with MDM2 are evidenced with a red box.

It was found that a coiled-coil motif is present on the α H2 helix of p14^{ARF} characterized by the presence of hydrophobic residues at position i , $i + 3$, $i + 4$ and $i + 7$ (i.e. ²⁰VRVFVVIH²⁷). Interestingly, this motif is on α H2 helix, a region of p14^{ARF} involved in the interaction with MDM2, acting as a negative modulator for the binding of p53 to MDM2 [18]. Thus, a hypothesis can be made, in which TSP can bind to MDM2 by reproducing some hot spot residues of p14^{ARF}, leading to a decrease in the formation of the p53-MDM2 complex (Figure 12; Table 10).

Since there is no structural data for the p14^{ARF}/MDM2 complex the modelling of the interaction of TSPs with MDM2 deserves future studies.

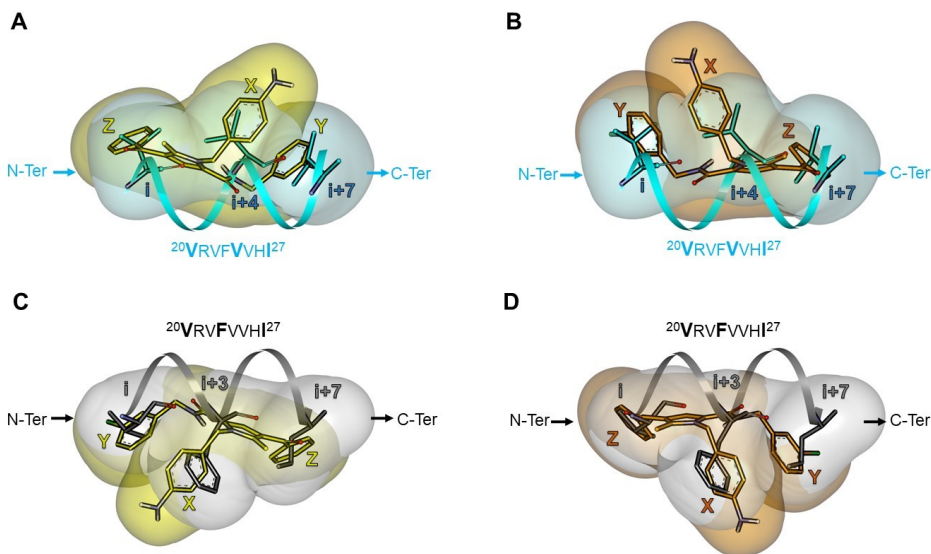


Figure 12. TCC I conformer of **10a** (i.e., pharmacophore 1) superimposed on the coiled coil motif present on α H2 of p14^{ARF}. E1 conformational enantiomer (yellow) superimposed on the i, i + 4, and i + 7 residues (cyan; (A)) and the i, i + 3, and i + 7 residues (white; (C)); E2 conformational enantiomer (orange) superimposed on the i, i + 4, and i + 7 residues (cyan; (B)) and the i, i + 3, and i + 7 residues (white; (D)). The centroids of the X, Y, and Z rings are superimposed on the centroids of the side chains of the hot-spot residues. The heavy atoms and the polar hydrogens are displayed as sticks and colored by atom type. Molecular volumes are shown as soft surface.

Table 10. Calculated RMSD (Å) by fitting the centroids of the aromatic rings X, Y, and Z of the selected DFT conformers on the centroids of the side chains of the considered PPI motif on human p14^{ARF}.

Pharmacophore	Compound	Sub-family	RMSD (Å)			
			20VRVFVVHI ²⁷		20VRVFVVHI ²⁷	
			i, i+3, i+7		i, i+4, i+7	
1	10a	TCC I	2.27	1.90	2.38	2.55
2	10a	TCT I	1.62	1.47	1.99	2.34
3	9g	TCC I	2.51	2.11	2.08	2.03
4	10b	TCG ⁻	1.63	1.93	2.45	2.06

^aThe RMSD values calculated for the conformational enantiomer E2 or for the configurational enantiomer S of **10b** are evidenced in bold.

3.2.6. SAR Analysis.

Considering that all the compounds exhibit the ability to reproduce (with their X, Y, and Z rings) the spatial arrangement of the side chains corresponding to the i, i+3/4, and i+7 residues within coiled-coil PPI motifs, we aimed to rationalize the observed structure–activity relationships (SARs) by distinguishing the respective contributions of pharmacokinetic and pharmacodynamic properties to the compounds' biological activity.

According to the results of solubility assays, the values of cLogD calculated for nutlin-3a and the new TSPs derivatives **9a-i**, **10a,b** resulted about 5 or higher (Table 11), indicating a high lipophilicity of the compounds.

Considering the obtained results, some consideration can be made: i) a high number of compounds show the same solubility but different activity, and ii) although a strict linear correlation was not found between solubility and activity ($R^2 = 0.2963$), it is noteworthy that ($\pm$)-**10b** and **10a** are the most active and also show higher solubility. In particular, introducing NH_2 group as R_1 (**10a**) and reducing the carbonyl group (($\pm$)-**10b**) led to an increase in both activity and solubility. In contrast, substituting CH_3 group with fluorine atom in the *para* position of the X ring (**9c**), shifting the chlorine atom from the *meta* to the *para* position on Y ring (**9d**) and introducing a methoxy group in the *para* position of the Z ring (**9i**) led to an increase in activity unrelated to solubility, also in the case of (**9g**) when the Z ring is substituted by a methyl group at C5 (Figure 13).

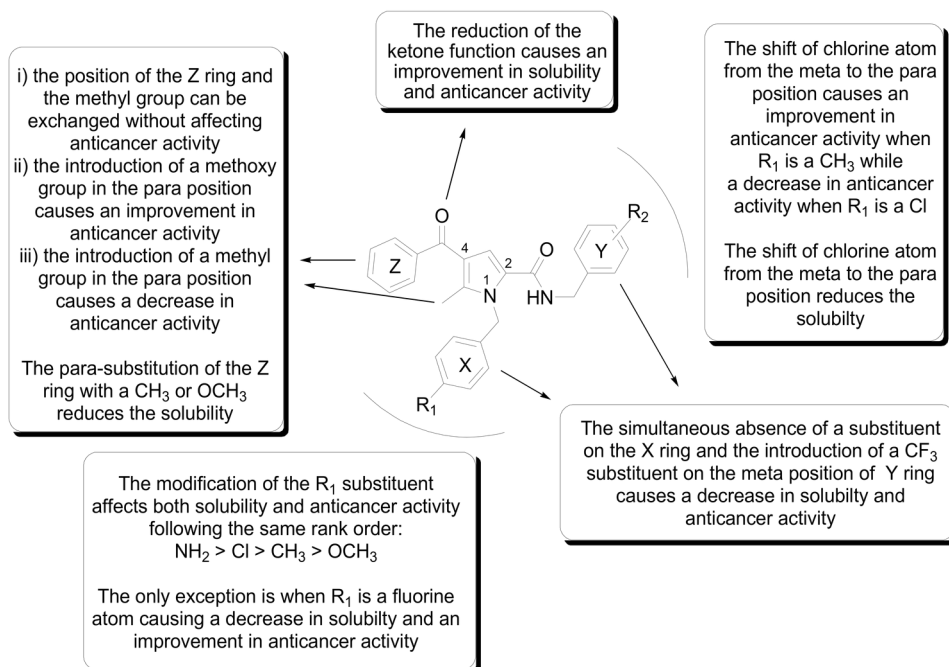


Figure 13. Schematic representation of the SARs of new TSP derivatives.

3.3. Conclusions.

A new small and rationally designed series of anticancer TSP derivatives, starting from the previously identified hit candidate **1**, was developed leading to the identification of a new hit (**10a**) with the highest activity and improved solubility among the present and previous series of TSPs.

The combination of computational studies with experimental studies allowed to identify the role of the different groups and substituents on the pharmacokinetics and pharmacodynamics of these compounds leading to the refinement of the 3D pharmacophore model.

The obtained results support the hypothesis that the anticancer ability of these compounds is due to their ability to mimic structural motifs involved in the PPIs activating apoptotic cell-death pathway. In particular, these

TSP derivatives share some peculiar structural features with the hot spot residues of protein motifs involved in formation of antiparallel and parallel coiled coil structures, such as that present on the N-terminal helix α H2 of p14^{ARF}.

The results of these studies can be a starting point for the rational design of new anticancer agents against melanoma (p53 WT, CDKN2A mut.).

3.4. References.

1. Ivanov, A.A., et al. "Targeting Protein-Protein Interactions as an Anticancer Strategy." *Trends in Pharmacological Sciences*, vol. 34, 2013, pp. 393-400.
2. Bullock, B.N., et al. "Assessing Helical Protein Interfaces for Inhibitor Design." *Journal of the American Chemical Society*, vol. 133, 2011, pp. 14220-23.
3. Algar, S., et al. "Evolution in Non-Peptide α -Helix Mimetics on the Road to Effective Protein-Protein Interaction Modulators." *European Journal of Medicinal Chemistry*, vol. 211, 2021, article 113015.
4. Ravindranathan, P., et al. "Peptidomimetic Targeting of Critical Androgen Receptor–Coregulator Interactions in Prostate Cancer." *Nature Communications*, vol. 4, 2013, article 1923.
5. Becerril, J., and A.D. Hamilton. "Helix Mimetics as Inhibitors of the Interaction of the Estrogen Receptor with Coactivator Peptides." *Angewandte Chemie International Edition*, vol. 46, 2007, pp. 4471-73.
6. Kutzki, O., et al. "Development of a Potent Bcl-xL Antagonist Based on α -Helix Mimicry." *Journal of the American Chemical Society*, vol. 124, 2002, pp. 11838-39.
7. Hoggard, L.R., et al. "Rational Design of Selective Small-Molecule Inhibitors for β -Catenin/B-Cell Lymphoma 9 Protein–Protein Interactions." *Journal of the American Chemical Society*, vol. 137, 2015, pp. 12249-60.
8. Vassilev, L.T., et al. "In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2." *Science*, vol. 303, 2004, pp. 844-48.

9. Grasberger, B.L., et al. "Discovery and Cocrystal Structure of Benzodiazepinedione HDM2 Antagonists that Activate p53 in Cells." *Journal of Medicinal Chemistry*, vol. 48, 2005, pp. 909-12.
10. Ribas, J., et al. "Theoretical Study of Alkyl- π and Aryl- π Interactions. Reconciling Theory and Experiment." *The Journal of Organic Chemistry*, vol. 67, 2002, pp. 7057-65.
11. Persico, M., et al. "New Anticancer Agents Mimicking Protein Recognition Motifs." *Journal of Medicinal Chemistry*, vol. 56, 2013, pp. 6666-80.
12. Proto, M.C., et al. "Tetrasubstituted Pyrrole Derivatives Act as Potent Activators of p53 in Melanoma Cells." *Investigational New Drugs*, vol. 38, 2020, pp. 634-49.
13. Persico, Marco, et al. "Tetrasubstituted Pyrrole Derivative Mimetics of Protein-Protein Interaction Hot-Spot Residues: A Promising Class of Anticancer Agents Targeting Melanoma Cells." *Molecules*, vol. 28, no. 10, 2023, article 4161.
14. Stewart, J.J. "Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements." *Journal of Molecular Modeling*, vol. 13, 2007, pp. 1173-213.
15. Budyka, M.F., and I.V. Oshkin. "Comparative Semiempirical and DFT Study of Styrylnaphthalenes and Styrylquinolines and Their Photocyclization Products." *International Journal of Quantum Chemistry*, vol. 111, 2011, pp. 3673-80.
16. Cossi, M., et al. "Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model." *Journal of Computational Chemistry*, vol. 24, 2003, pp. 669-81.
17. Hodis, E., et al. "A Landscape of Driver Mutations in Melanoma." *Cell*, vol. 150, 2012, pp. 251-63.
18. DiGiammarino, E.L., et al. "Solution Structure of the p53 Regulatory Domain of the p19Arf Tumor Suppressor Protein." *Biochemistry*, vol. 40, 2001, pp. 2379-86.

Chapter 4. Tetra-anionic porphyrin mimics protein-protein interactions between regulatory particles and the catalytic core, allosterically activating human 20S proteasome.

4.1. Introduction.

Consistent with the conceptual framework outlined in Section 1.3 of Chapter 1, nature-inspired privileged scaffolds can be considered as evolutionary refined molecular structures that recognize homologous structural elements in the protein target(s). These scaffolds frequently occupy regions of chemical space that are poorly represented in conventional synthetic libraries. Based on these premises, the charged porphyrins were selected for the development of novel proteasome modulators, aiming to mimic protein-protein interactions between the catalytic core of proteasome and the regulatory particles (RPs).

In order to expand the structure-activity relationship (SAR) of this class of inhibitors/activators and to investigate the effect of changing the electrostatic charges on the porphyrin scaffold from positive to negative, this chapter reports the computational and experimental studies performed on a tetra-anionic porphyrin able to mimic PPI between the catalytic core and RPs, allosterically activating human 20S proteasome.

Proteasome is a multi-subunit cytosolic protease that constitutes the key element of the ubiquitin-proteasome pathway (UPP), the major protagonist of regulated protein catabolism [1-4]. Proteasome malfunctioning is directly linked to the pathogenesis of a large number of diseases, from cancer to neurodegeneration [1, 5-7].

The UPP system is the principal mechanism of intracellular protein degradation in eukaryotic organisms and has an essential role in the equilibrium of cellular homeostasis by regulating quality, quantity and functionality of proteins [8-9]. The process is based on the covalent binding of ubiquitin, a little polypeptide of 76 amino acids highly conserved, to the proteins that have to be degraded.

Ubiquitination happens through an enzymatic cascade that involves three classes of enzymes: E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) and E3 (ubiquitin ligase). The latter give specificity to the process, by catalyzing an isopeptidic bond between ubiquitin and a lysine residue of the target protein. [9,10]. By adding more units of ubiquitin there is the formation of polyubiquitin chains that are recognized by proteasome 26S (Figure 1) [11,12].

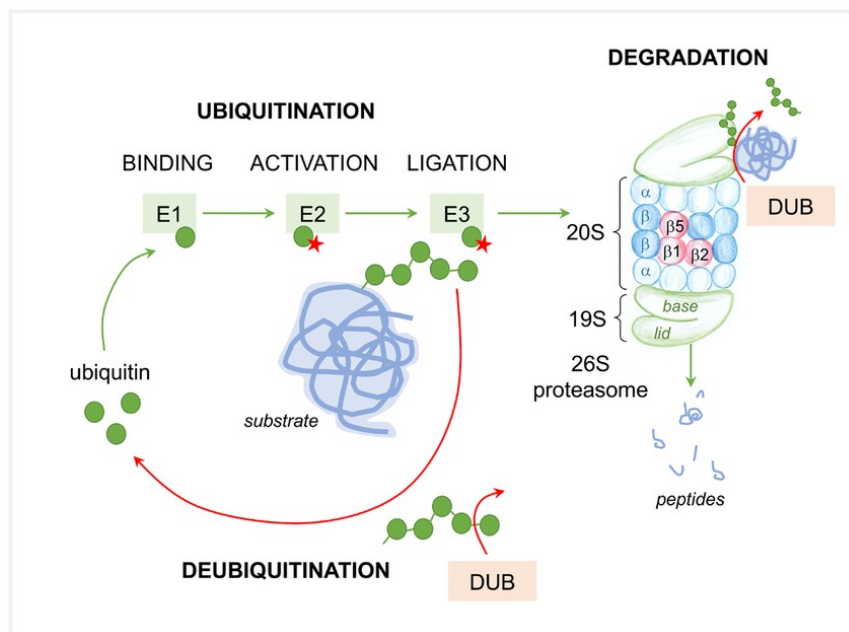


Figure 1. Schematic representation of the UPP system. [Meyer-Schwesinger, Lukas Heintz Catherine. "The intertwining of autophagy and the ubiquitin proteasome system in podocyte (patho) physiology." *Cell Physiol Biochem* 55.S4 (2021): 68-95.]

The mature proteasome is composed of a core particle in a barrel-shape (20S CP; human 20S proteasome (h20S)) which contains the catalytic subunits (Figure 2 A and B), and different RPs (namely 19S, PA28 and PA200 for h20S) that can bind to the core (Figure 3) [13-15].

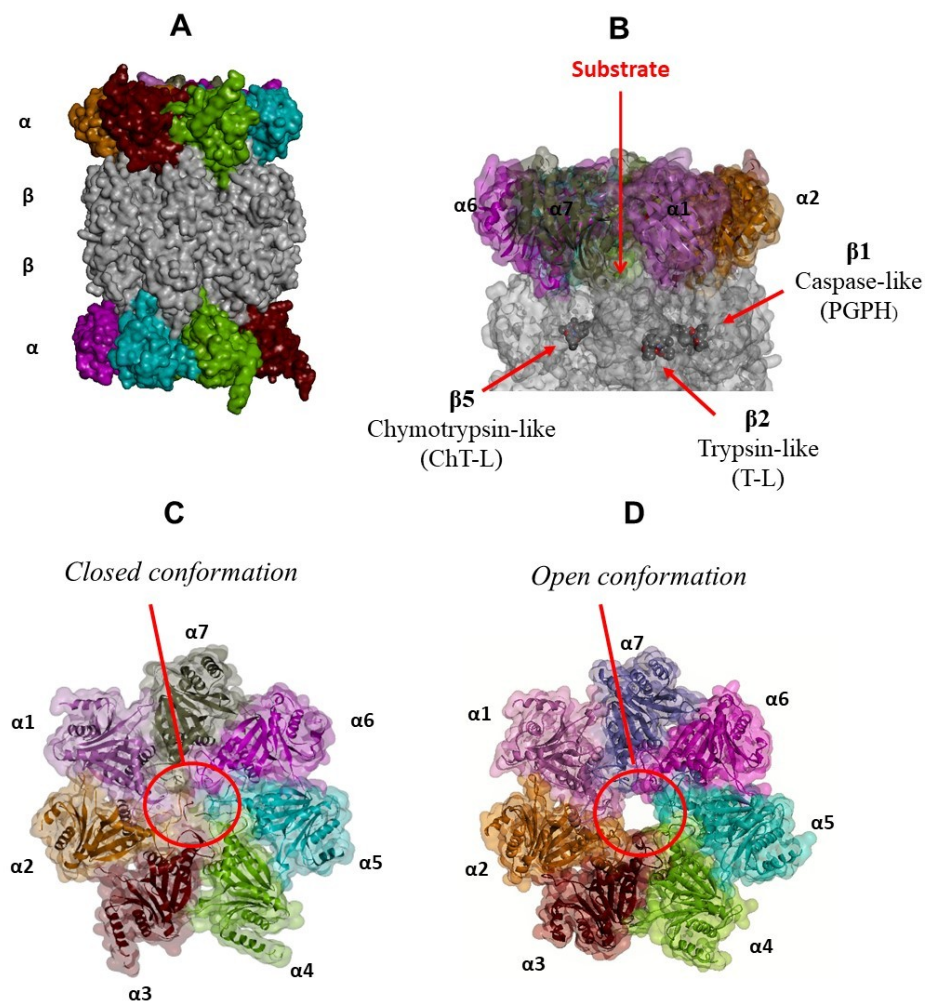


Figure 2. A: Overall structure of h20S (PDB ID: 4R3O). B: Transversal view of the h20S proteasome in its open conformation with evidenced the three catalytic subunits and the gate. C: Top view of the h20S proteasome in its closed conformation (PDB ID: 5T0C). D: Top view of the h20S proteasome in its open conformation (PDB ID: 5T0J). The backbone is displayed as solid ribbons and colored in pink ($\alpha 1$), orange ($\alpha 2$), brown ($\alpha 3$), light green ($\alpha 4$), cyan ($\alpha 5$), magenta ($\alpha 6$), and grey ($\alpha 7$ and all β subunits).

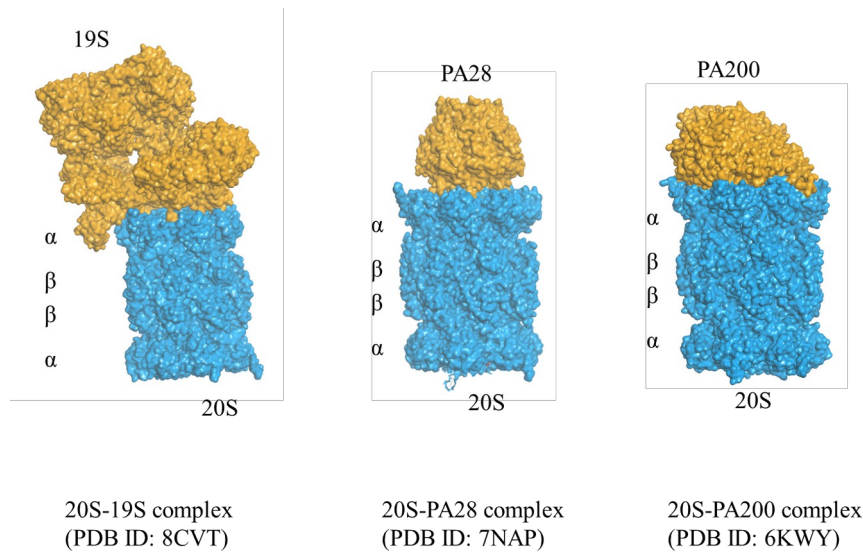


Figure 3. Human 20S proteasome in complex with the regulatory particles 19S, PA28 and PA200.

There is also the presence of other modulators (e.g. Ecm29 [16], heat shock protein 60 (HSP60) [17], insulin-degrading enzyme (IDE) [18], Catalytic Core Regulators (CCRs) [19] also often called as proteasome interacting proteins (PIPs) [20,21]. The 20S CP is made by four heptameric stacked rings, two external α rings (α 1- α 7) and two internal β rings (β 1- β 7) [22,23]. Each of the β rings contains the catalytic subunits β 1, β 2 and β 5 that shows postglutamyl peptide hydrolising (PGPH), trypsin-like (T-L), and chymotrypsin-like (ChT-L) activity, respectively [24]. The access to the catalytic chamber by the substrate happens through the gate placed in the centre of each of the two α rings (Figure 2B). When there is no substrate and/or activator (e.g RPs), the gate is closed thanks to the N-terminal tails of α 2, α 3 and α 4 subunits. The activation of the proteasome is tightly controlled through complex conformational equilibria with a

dynamic crosstalk between α and β subunits, which results in a fine tuning of the substrate access to catalytic sites inside 20S CP barrel [25-30].

Within the framework of proteasomal regulatory particles (RPs), I conducted an in-depth analysis of the role of IDE, with a particular focus on exploring its structure–function relationship. The details of this investigation are thoroughly reported in the following review “Tundo GR, Grasso G, Persico M, Tkachuk O, Bellia F, Bocedi A, Marini S, Parravano M, Graziani G, Fattorusso C, Sbardella D. The Insulin-Degrading Enzyme from Structure to Allosteric Modulation: New Perspectives for Drug Design. *Biomolecules*. 2023 Oct 7;13(10):1492. doi: 10.3390/biom13101492.” [31].

IDE, also named insulysin, is a 110-kDa neutral zinc and thiol-dependant metallopeptidase and member of the invercinzin family of metallopeptidases, characterized by an inversion of the active site motif “HExxH” to “HxxEH” [32,33]. IDE is formed by four structurally homologous $\alpha\beta$ domains (domain 1, residues 1-285; domain 2, residues 286-515; domain 3, residues 542-768; and domain 4, residues 769-1019). The N-terminal portion of the IDE (i.e IDE-N) that includes domains 1 and 2 (colored white and blue, respectively; Figure 4 and the C-terminal portion (i.e. IDE-C) that includes domains 3 and 4 (colored pink and magenta, respectively; Figure 4) are connected by a 28-residue hinge loop (colored red; Figure 4).

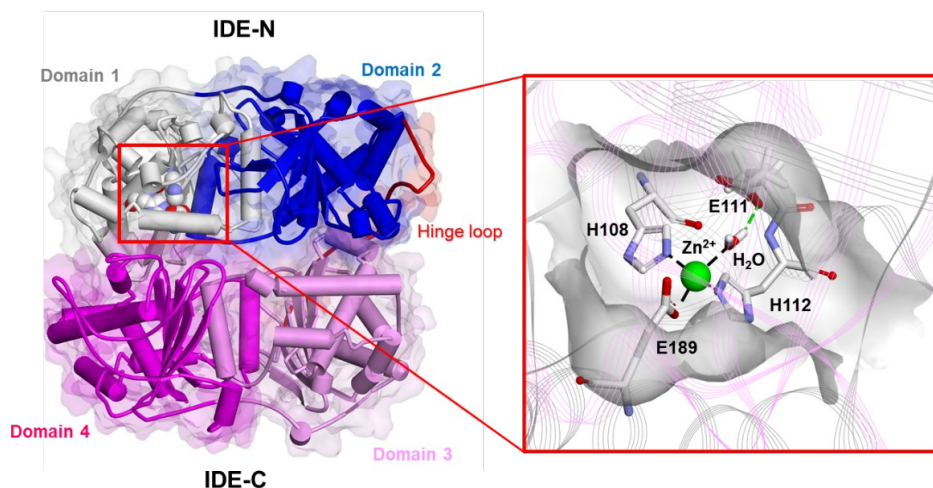


Figure 4. Experimentally determined structure of the human IDE in the monomeric apo form (PDB ID: 2JG4). Left: overall structure of the human IDE. The secondary structure is displayed as wide cylinders (α -helices), arrows (beta-sheets), and tubes (coils and turns), and is colored by protein domains: white (domain 1: aa 43–285), blue (domain 2: aa 286–515), pink (domain 3: aa 542–768), magenta (domain 4: aa 769–1016), and red (hinge loop: aa 516–541). The protein Connolly surface is displayed (transparency 90%). The catalytic residues are displayed in CPK. Right: close-up view of the catalytic site. The protein is displayed as line ribbons. The catalytic residues and water molecules are displayed in stick and colored by atom type (C: white; N: blue; O: red). The zinc ion is displayed in CPK (scale: 0.5) and colored green. Only the hydrogen atoms of the catalytic water molecules are displayed for the sake of clarity.

The activities and properties through which IDE is expected to contribute to protein homeostasis are: i) “dead-end” chaperone; ii) Heat Shock Protein (HSP); and iii) interaction with components of the UPP [32]. In this latter case, IDE has been proposed to: i) interact and modulate the activity of the 20S CP and ii) control the equilibrium between the proteasome particle, 20S, 26S and 30S. So, IDE could bind 20S CP and induce a conformational change that modulates its catalytic activity and prevents the binding of 19S, increasing the pool of uncapped 20S.

The proteasome emerged as an attractive pharmacological target due to its cell's physiology [7,34]. There are already several drugs used in anti-cancer field of action as competitive inhibitors. Another way of action can be boosting proteasome activity in neurodegenerative disorders [35,36], that is associated in UPP decline. Proteotoxicity *in cells* were reduced by proteasome activation [37-40] and also symptoms of neurodegeneration were attenuated in animal models of Alzheimer's disease [41]. Progressive decline in UPP effectiveness contributes to deleterious effects of senescence [42] and this is in agreement with the observation of increased proteasome activity in the exceptionally long-lived naked male rat [43,44] and long-lived humans (centenarians) [45]. Based on this information, different strategies have been developed to increase UPP activity in the fight against neurodegenerative diseases, with several compounds identified as potential proteasome activators [39]. These activators can be grouped into i) direct and ii) indirect; the first are molecules that bind directly the 20S and enhance its proteolytic activity, the second group include compounds that boost proteasome activity through regulation of deubiquitinase, transcription factors and upstream kinases [46-48].

The compounds, either synthetic or natural, that act by relaxing the narrow 20S gate and consequently promote the entry into the catalytic chamber of 20S are often referred to as gate openers. Some of them are derived from RPs, in particular the conserved HbYX (Hb = hydrophobic; Y = Tyr; X = any residue) activating motif present at the 20S-binding C-terminals of 19S Rpt (regulatory particle ATPase) subunits 2, 3, and 5, and PA200 [49-54]. Among these there is a HbYX-like dipeptide mimetic Z-Tyr-Ala [37]. Several small-molecule ligands have been reported as direct activators as well, among them several natural compounds, such as ursolic, betulinic,

oleanolic acids [55], lithocholic acid [56], cycloastragenol [57], polyphenols [58,59], aflatoxins [60], sylibins [61], oleuropein [62], citicoline [63] and syrotingopine [64]. Also synthetic small molecules can act as direct activators, such as fluspirilene analogs, phenothiazines, dihydroquinazolines [65], pyrazolones [66], imidazolines [67], MK886 and its derivatives [68], N-arachidonoylphenolamine (AM404) [69], carboranes [70] and a cytosine derivative [71].

The cationic porphyrins were recently identified as a new class of proteasome modulators [72-75], which is able to interact with charged residues in the grooves between the α subunits. Notably, these residues are also engaged in ionic interactions with RPs. This interaction may represent an “electrostatic key code” that is able fine-tune the ability of charged ligands to affect with proteasome gate movements. Porphyrins have a constrained and symmetric planar structure, so they represent a valuable probe to assess the validity of this hypothesis. Moreover, SAR studies are possible thanks to the easy modification of the side groups. This in turn results in a rational design of the modifications of the porphyrin scaffold in order to generate compounds that have the desired effect on proteasome enzymatic abilities.

In agreement with the “electrostatic key code” hypothesis (Figure 5, Figure 4 in Chapter 1), it was reported [75] the diverse effect of changing the number of charged groups and, in the meantime, of introducing a hydrophobic group in the porphyrin structure by putting a phenyl ring in place of one of the protonated N-methyl pyridine of meso-tetrakis(4-N-methylpyridyl)-porphyrin (H2T4) (Figure 6).

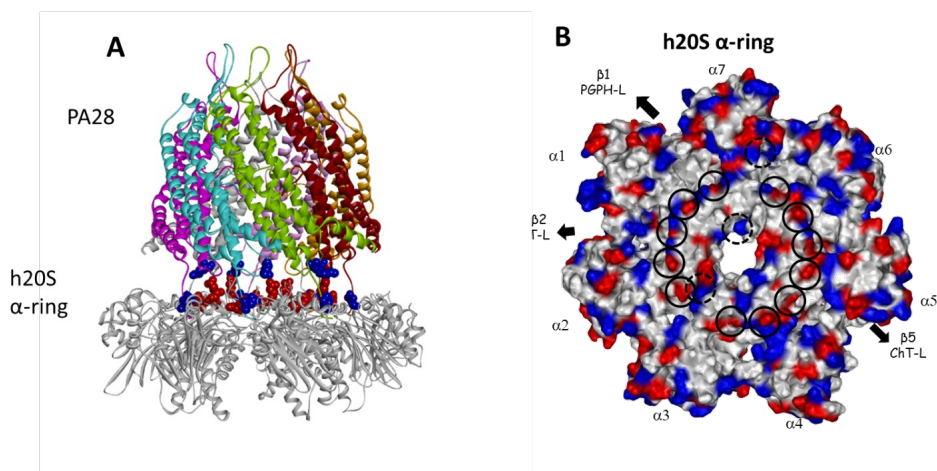


Figure 5. A: 20S negatively charged residues interacting with positively charged residues of PA28; B: charged residues on the h20S α -ring (colored blue = positive and red = negative), those involved in the interaction with PA28 are circled.

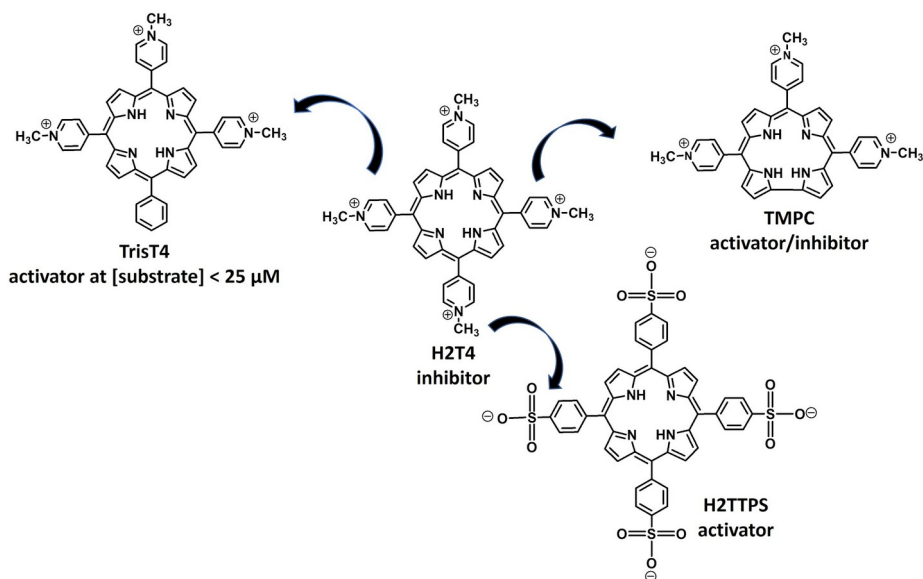


Figure 6. 2D structures of cationic and anionic porphyrins identified as 20S inhibitors or activators.

The derived compound, 5-(phenyl)-10,15,20-(tri N-methyl-4-pyridyl)-porphyrin (TrisT4), showed an activating effect on h20S due to the ability

of the phenyl ring to mimic a key interaction of the HbYX activating motif at the C-terminals of PA200 and 19S [75]. Nevertheless, the enzymatic activity was diminished with the increase of substrate concentrations, and this effect was no longer detectable at substrate concentration $> 25 \mu\text{M}$. 5,10,15-Tri(N-methyl-4-piridyl) corrole (TMPC; Figure 4) has also showed to improve ChT-L activity of h20S, but the activation was followed by inhibition upon increase of ligand concentration [73].

Starting by the map of ionic interactions that occur between charged residues present on the surface of h20S and the docking regions of RPs [72], in order to expand the structure-activity relationship (SAR) of this class of inhibitors/activators and to investigate the effect of changing the electrostatic charges on the porphyrin scaffold from positive to negative, this chapter reports: i) the effect of a tetra-anionic porphyrin derivative, meso-tetrakis (4-sulphonatophenyl)-porphyrin (H2TPPS; Figure 4) on the ChT-L activity of h20S; ii) the monitoring of the conformational equilibria of the h20S gate area (closed, intermediate, and open state) upon exposure to this porphyrin by using single-molecule imaging by atomic force microscopy (AFM) and iii) the dynamic docking studies to investigate the binding of H2TPPS to the different h20S conformations.

This study was performed in the context of a multidisciplinary collaboration involving the research groups of: i) Prof. Massimo Coletta (Fondazione Bietti) for the biochemical assays, ii) Prof. Pawel Osmulski and Prof. Maria Gaczynska (University of Texas; USA) for AFM imaging experiments and iii) Prof. Purrello (University of Catania), Dr. Danilo Milardi and Dr. Anna Maria Santoro (National Research Council of Catania) for proteasome activity assays. Moreover, this research activity was performed in the framework of two projects funded by the Italian

Ministry of University and Research (MUR): “PRIN 2022 PNRR Prot. P2022YRPHS (entitled “Multifaced-peptidomimetics as proteasome modulators”) and “PRIN2022 Prot. 2022PAAYZE (entitled “Modulating conformational Equilibria of prion protein and proteasome to tune proteostasis network (RESET)”). The results reported in this chapter are object of the scientific publication reported in ref [76].

I was directly involved in performing and analyzing computational studies as well as in the interpretation of the experimental results (particularly SAR analysis). I participated in the meetings of the research groups involved in the study and attended school and congresses related to this research topic; in particular I attended XXVIII EFMC International Symposium on Medicinal Chemistry held in Rome, where I presented a poster about this study.

The experimental procedure of biochemical assays and AFM imaging is not reported here but can be found in the publication "Tetra-anionic porphyrin mimics protein-protein interactions between regulatory particles and the catalytic core, allosterically activating human 20S proteasome." *Journal of Enzyme Inhibition and Medicinal Chemistry* 40.1 (2025): 2482892[76].

4.2 Results and Discussion.

4.2.1. Effect of H2TPPS on the h20S proteolytic activity.

In Figure 7 is reported the double-reciprocal Lineweaver-Burk plot of data of ChT-L enzymatic activity of h20S as function of fluorogenic substrate Suc-LLVY-AMC, selective for the ChT-L activity, in presence and absence of H2TPPS with concentration ranging from 0 to 10 μ M.

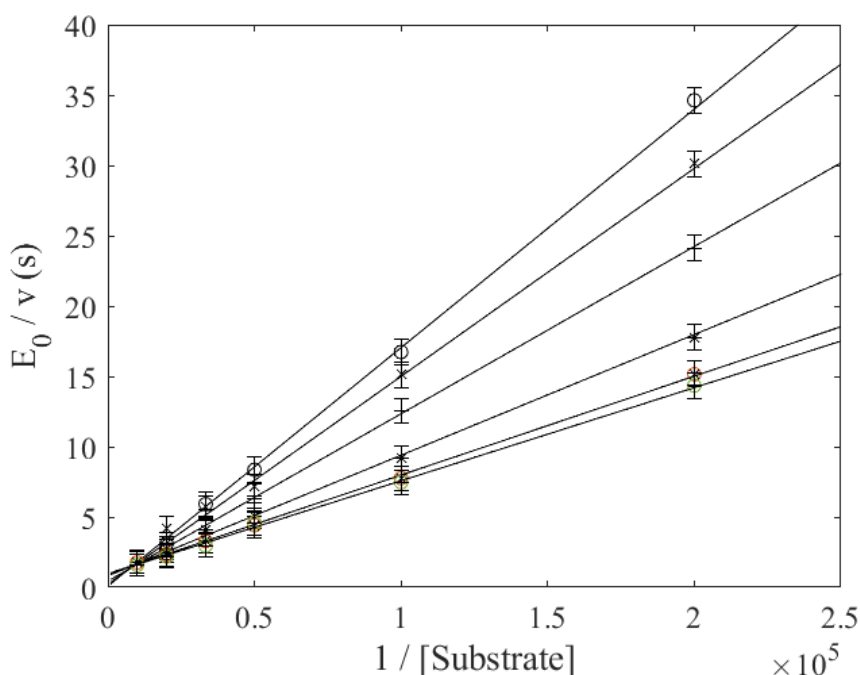


Figure 7. Lineweaver-Burk plot of the ChT-L activity of human 20S at pH 8.0 and 37°C for different concentrations of H2TPPS porphyrin, namely no (o), 0.1 μM (x), 0.2 μM (+), 0.5 μM (*), 1 μM ($\otimes$) and 10 μM ($\oplus$). For the sake of simplicity data at 0.05 μM and at 5 μM H2TPPS were omitted, since they were closely similar to those with no and 10 μM H2TPPS, respectively. Continuous lines are linear least-squares fitting of data according to Eq. (1).

As evidenced by the decrease in the slope, the tetra-anionic porphyrin H2TPPS activates the h20S, over the concentration range used, rising up the catalytic efficiency. The overall effect can be divided into several contributions, namely the H2TPPS-linked effect on the velocity of the catalysis rate-limiting step (i.e k_{cat} ; Figure 8A) and on the substrate affinity (i.e. K_{m} ; Figure 8B). The ratio between these two values (i.e. $k_{\text{cat}}'/K_{\text{m}}'$; Figure 8C) represents the catalytic efficiency of the enzyme.

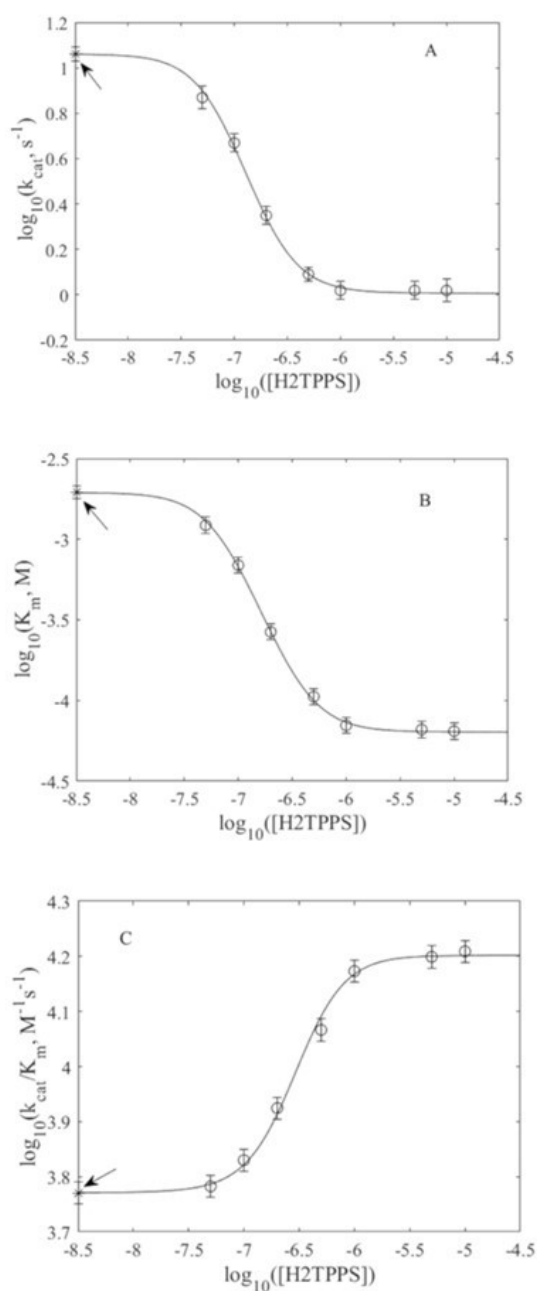


Figure 8. Dependence on H2TPPS concentration of ChT-L catalytic parameters of 20S proteasome, namely of k_{cat} (panel A), K_m (panel B) and k_{cat}/K_m (panel C). Continuous lines represent the non-linear least-squares fitting of data according to Eqs (2a-c), employing the common set of parameters reported in Table 1. Asterisk, pointed by an arrow, indicates the parameter value in the absence of H2TPPS.

These catalytic parameters show a cooperative dependence, highlighting an allosteric mechanism of this anionic porphyrin, whose binding induces a change of the 20S proteasome from a state with lower substrate affinity (state A) to one with higher substrate affinity (state B).

Tables 1 and 2 show the kinetic parameters obtained by experimental data [75].

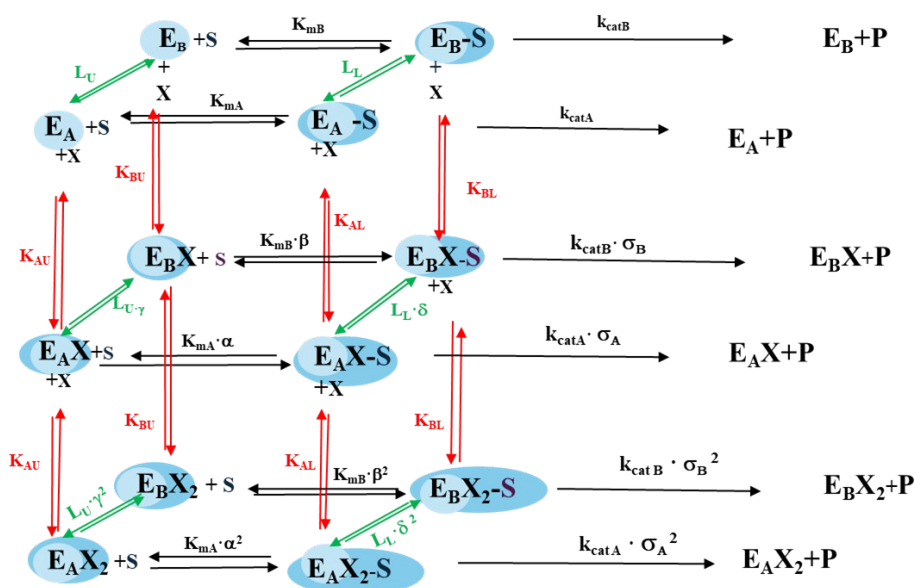
Table 1. Kinetic parameters of human 20S [74].

h20S	
k_{catA} (s ⁻¹)	11.5 ± 1.3
k_{catB} (s ⁻¹)	0.32 ± 0.05
K_{mA} (M)	$1.9 (\pm 0.4) \times 10^{-3}$
K_{mB} (M)	$1.5 (\pm 0.3) \times 10^{-4}$
L_U	$3.0 (\pm 0.5) \times 10^{-6}$
L_L	$3.9 (\pm 0.6) \times 10^{-5}$

Table 2. Association constants of H2TPPS for the A and B states, unbound to substrate (K_{AU} and K_{BU}) and substrate-bound (K_{AL} and K_{BL}), number of H2TPPS molecules for full h20S activation (n), and interaction parameters describing the effect on k_{cat} of H2TPPS binding (σ).

h20S +H2TPPS	
K_{AU} (M ⁻¹)	$2.0 (\pm 0.4) \times 10^4$
K_{BU} (M ⁻¹)	$1.5 (\pm 0.3) \times 10^9$
K_{AL} (M ⁻¹)	$1.2 (\pm 0.3) \times 10^5$
K_{BL} (M ⁻¹)	$2.3 (\pm 0.4) \times 10^9$
n	2
σ_A	1.0 ± 0.2
σ_B	1.8 ± 0.3

As can be seen from Table 1, in the absence of H2TPPS, h20S in the A state has a slightly higher catalytic efficiency (k_{cat}/K_m) than in the B state ($6052 \text{ M}^{-1}\text{s}^{-1}$ vs $2133 \text{ M}^{-1}\text{s}^{-1}$). On the other hand, the EB-X2 and EB-X2-S states are predominant in the doubly liganded fully activated h20S (as from $L_U \times \gamma_2 = 1.7 \times 10^4$ and $L_L \times \delta_2 = 1.4 \times 10^4$; Scheme 1) due to the higher affinity of H2TTPS for the state B compared to the state A.



Scheme 1. Thermodynamic description of equilibria involved in the modulation of enzymatic activity of 20S proteasome (E) on the substrate (S) by the anionic porphyrin H2TPPS (X). Arrows indicate equilibrium constants (green), kinetic parameters (black), ligand affinity constants (red). In the upper part is the scheme for the first H2TPPS binding step, while in the lower part is the scheme for the second H2TPPS binding step. Parameters concerning the enzymatic activity are depicted in black, those regarding porphyrin binding are depicted in red and conformational equilibria are depicted in green.

By consequence, the corresponding enzymatic efficiency is defined by $k_{catB} \times \sigma_B^2 / (K_{mB} \times \beta^2) = 1.6 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$, and resulted, in any case, increased with respect to that of the “initial” A or B states in the absence of H2TPPS.

In summary: i) H2TPPS activates h20S through the concerted action of at least two molecules, bringing a change from the E_A to the $E_B\text{-X}_2\text{-S}$ conformational/functional state, ii) the affinity of H2TPPS is very high for the state B, 75.000-fold higher than A in free enzyme and 19.000-fold higher than in substrate-bound enzyme and iii) this affinity is much higher than of the other porphyrins [73-75], suggesting that the site is positively charged or has higher affinity for negative effectors.

4.2.2. AFM Imaging.

The possible conformational changes in the proteasome gate area induced by H2TPPS was investigated by means of AFM imaging. The images of the α face can be classified as “open-gate”, “intermediate” and “closed-gate”, depending on their topography [77,78] and all the particles can assume any of the conformations when scanned and imaged every 2-3 minutes. Nonetheless, the partition of conformers depends on the presence of a specific ligand. In Figure 7 we can see the partitions of closed, intermediate and open h20S particles in presence or absence of the peptide substrate for the ChT-L activity and/or H2TPPS at two different concentrations (1 μM and 2 μM) (Figure 9).

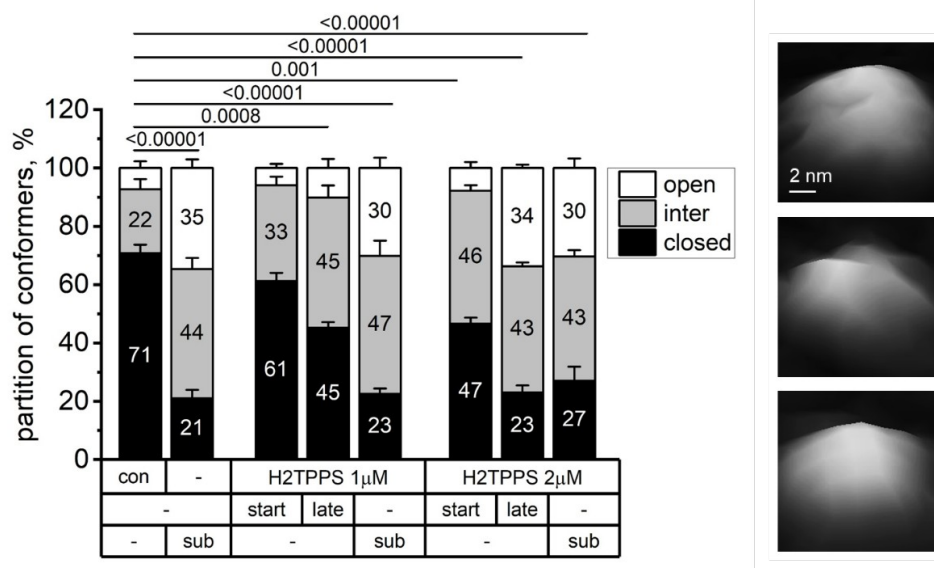


Figure 9. Relative (%) partitions of the three conformational forms of h2OS (from the top: open, intermediate and closed; AFM images at the right). Data are reported in the absence (control; “con”) and in the presence of 1 μ M (central set of columns) or 2 μ M (rightmost set of columns) H2TPPS within and over 15 minutes after its addition (“start” and “late” columns, respectively). The effect of adding the substrate (Suc LLVY-AMC 100 μ M) at both H2TPPS concentrations is also reported (“+Sub” columns). All values are the average of partitions calculated for at least $n=3$ fields with at least 150 proteasome particles. The partitions between control and substrate or porphyrin treated h2OS were statistically different with p values at the top (Chi-Square test). The AFM images are pseudo-3D renderings of zoomed in h2OS particles.

Control “idle” proteasome existed in a conformational equilibrium as predominantly in closed-gate forms ($71\% \pm 3\%$), with a fraction of intermediate-gate ($22\% \pm 3\%$) and a smaller part as open-gate ($7\% \pm 2\%$) particles (Figure 9, left). With the addition of the peptide substrate, the contribution partition of closed gate was drastically reduced to $21\% \pm 3\%$, with an increase in intermediate partition ($44\% \pm 4\%$) and open partition ($35\% \pm 3\%$). This kind of partition was clear shortly after the addition of

the substrate (first scan) and remained stable through repeated scans. H2TPPS was added in the absence of the substrate and induced a concentration-dependent reduction in closed gate conformers. This reduction was slow, with stabilization after 15 min (5 scans) from the addition of the porphyrin (Figure 9, center and right, early and late effects). Treatment with 2 μ M porphyrin led to the “late” partition of conformers not significantly different from that with the substrate alone (Figure 9, left and right). Moreover, addition of the substrate to proteasomes treated with H2TPPS did not alter the conformations partition when 2 μ M and shifted the conformation partition to “substrate-alone”, like when the concentration used of the porphyrin was of 1 μ M (Figure 9).

Previous structural studies based on cryogenic electron microscopy (cryo-EM) and Nuclear Magnetic Resonance (NMR) techniques showed that the conformational transition of 20S is characterized by several intermediate states with some of them corresponding grossly to the semi-open conformation of the gate [29, 79-80]. Our AFM data confirmed the presence of semi-open intermediates of the h20S and showed that the treatment with H2TPPS shifts the conformational equilibrium toward the opening of the gate, consequently marked by lowered participation of the closed-gate conformers and elevated partition of the intermediate and open-gate states.

4.2.3. Dynamic docking studies.

The tetra-anionic ligand H2TPPS was docked to different conformational states of h20S to analyze the molecular bases of its ability to affect the 20S conformational equilibrium. The aim was to explore possible binding sites for the tetra-anionic ligand and to identify the one with the best binding

energy. Only one ligand molecule for each h20S conformational form was docked, so no information was provided about the binding of additional ligand molecules. The analysis was guided by the working hypothesis of an electrostatic code on the α -surface mirroring the charged motifs of endogenous regulatory particles (RPs). Accordingly, α -grooves enriched in solvent-exposed Lys/Arg residues were prioritized as putative anchoring sites for H2TPPS.

4.2.4. Ligand analysis and generation of the starting complexes.

As a first step, the molecular model of H2TPPS was built in tetra-anionic ionization state, and its conformational space was sampled. This enabled the identification of the global minimum energy conformer of H2TPPS and the successive calculation of conformational energy in the docked complexes as well as the generation of a pharmacophore model necessary for the starting binding hypothesis (Figure 10).

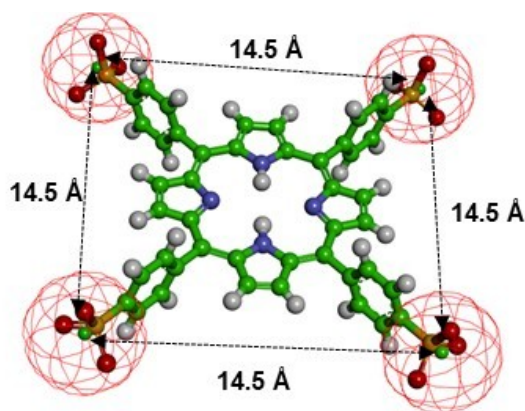


Figure 10. H2TPPS pharmacophore and related inter-atomic distances: i) the negative ionizable sulfonate groups are evidenced as mesh spheres and ii) the centroid of the oxygen atoms of each sulfonate group is displayed as a green ball. The global minimum conformer of H2TPPS is displayed in ball & stick and colored by atom type (C: green; N: blue; S: yellow; O: red and H: white).

The docking procedure requires a reasonable starting complex even if during the calculation the ligand is rotated and translated with respect to the protein and its conformation is changed. Then, all the positively charged residues of h20S (closed, open and semi-open conformation) engaged in ionic interactions with RPs (i.e. 19S, PA200, PA28; Figure 11; Tables 1-6 Appendix C), were mapped and fitted with the H2TPPS pharmacophore.

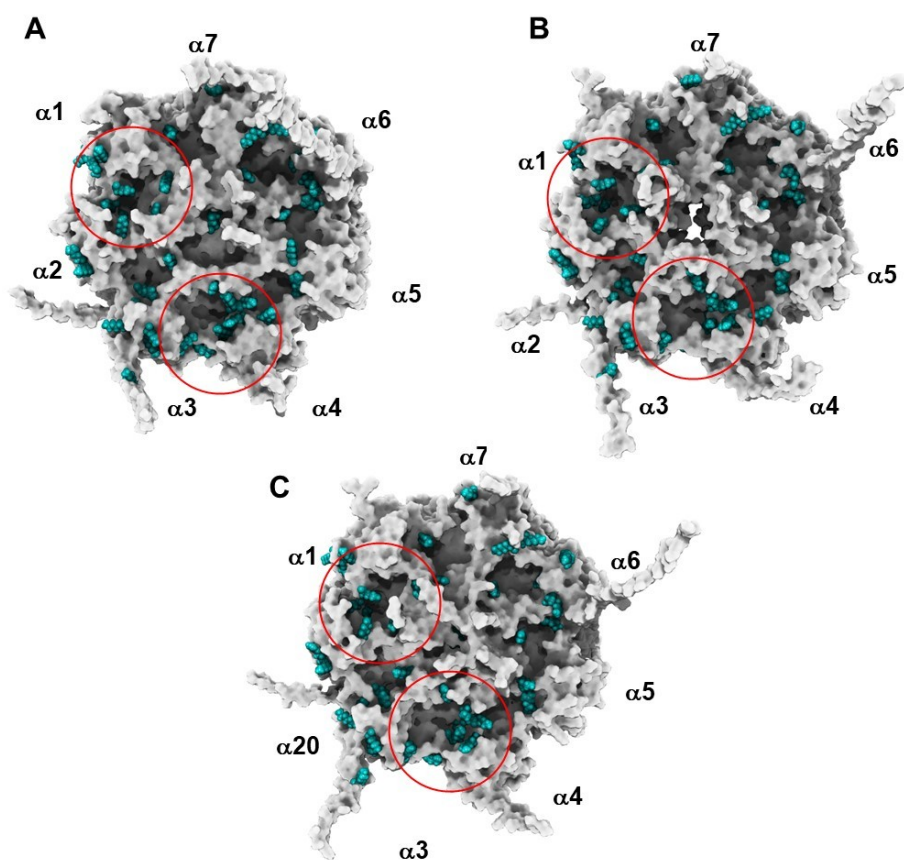


Figure 11. Top view of the human 20S proteasome structure in the closed (A), open (B), and semi-open (C) state. The protein is displayed as Connolly surface and colored in grey; the positively charged residues involved in ionic interactions with RPs (i.e., 19S, PA28 α and PA200) are displayed as CPK and colored in cyan. The putative binding sites identified for H2TPPS are evidenced with red circles.

From this analysis, two clusters of four solvent exposed positively charged amino acids were identified on the α ring of h20S ($\alpha 1$ - $\alpha 2$ and $\alpha 3$ - $\alpha 4$ grooves) which could match the planar, rigid and negatively charged pharmacophore of H2TPPS (Figure 11; Tables 3-4).

Table 3. Identified clusters of positively charged amino acids on the α -ring surface of human 20S proteasome (closed, semi-open and open conformation) able to interact with H2TPPS and related inter-residue distances (C α).

	Closed state			
Groove	D1 (Å)	D2(Å)	D3(Å)	D4(Å)
$\alpha 1$ - $\alpha 2$	25.4 K30($\alpha 1$)- K181($\alpha 1$)	18.7 K181($\alpha 1$)- K51($\alpha 2$)	13.1 K51($\alpha 2$)- K165($\alpha 2$)	26.8 K30($\alpha 1$)- K165($\alpha 2$)
$\alpha 3$ - $\alpha 4$	33.0 R17($\alpha 3$)- K199($\alpha 3$)	14.5 K199($\alpha 3$)- K176($\alpha 3$)	26.3 K176($\alpha 3$)- R163($\alpha 4$)	15.1 R17($\alpha 3$)-R163 ($\alpha 4$)
	Open state			
$\alpha 1$ - $\alpha 2$	25.9 K30($\alpha 1$)- K181($\alpha 1$)	18.9 K181($\alpha 1$)- K51($\alpha 2$)	13.7 K51($\alpha 2$)- K165($\alpha 2$)	27.1 K30($\alpha 1$)- K165($\alpha 2$)
$\alpha 3$ - $\alpha 4$	35.1 R17($\alpha 3$)- K199($\alpha 3$)	14.9 K199($\alpha 3$)- K176($\alpha 3$)	29.9 K176($\alpha 3$)- R163($\alpha 4$)	14.9 R17($\alpha 3$)-R163 ($\alpha 4$)
	Semi-open state			
$\alpha 1$ - $\alpha 2$	27.5 K30($\alpha 1$)- K181($\alpha 1$)	18.1 K181($\alpha 1$)- K51($\alpha 2$)	14.1 K51($\alpha 2$)- K165($\alpha 2$)	29.2 K30($\alpha 1$)- K165($\alpha 2$)
$\alpha 3$ - $\alpha 4$	35.4 R17($\alpha 3$)- K199($\alpha 3$)	16.2 K199($\alpha 3$)- K176($\alpha 3$)	29.2 K176($\alpha 3$)- R163($\alpha 4$)	18.2 R17($\alpha 3$)-R163 ($\alpha 4$)

Table 4. Percentage of the sidechain solvent accessibility of the identified clusters of positively charged amino acids on the α -ring surface of human 20S proteasome able to interact with H2TPPS.

Conformational state	Groove	Residue	Percent Sidechain Solvent Accessibility
Closed	$\alpha 1$ - $\alpha 2$	K30($\alpha 1$)	75.31
		K181($\alpha 1$)	100
		K51($\alpha 2$)	45.54
		K165($\alpha 2$)	75.08
	$\alpha 3$ - $\alpha 4$	R17($\alpha 3$)	41.63
		K176($\alpha 3$)	61.69
		K199($\alpha 3$)	89.21
		R163($\alpha 4$)	93.24
Open	$\alpha 1$ - $\alpha 2$	K30($\alpha 1$)	45.48
		K181($\alpha 1$)	100
		K51($\alpha 2$)	34.24
		K165($\alpha 2$)	68.78
	$\alpha 3$ - $\alpha 4$	R17($\alpha 3$)	65.91
		K176($\alpha 3$)	46.90
		K199($\alpha 3$)	89.97
		R163($\alpha 4$)	79.67
Semiopen	$\alpha 1$ - $\alpha 2$	K30($\alpha 1$)	28.92
		K181($\alpha 1$)	100
		K51($\alpha 2$)	23.94
		K165($\alpha 2$)	77.43
	$\alpha 3$ - $\alpha 4$	R17($\alpha 3$)	33.12
		K176($\alpha 3$)	63.87
		K199($\alpha 3$)	68.02
		R163($\alpha 4$)	70.67

Accordingly, two ligand starting poses were produced and since H2TPPS was docked to the closed, intermediate and open conformation, a total of six starting complexes were subjected to docking simulation.

4.2.5. Docking calculations.

In the docking procedure, the ligand was considered flexible and the whole h20S structure was defined as search space. Only, the backbone atoms of conserved secondary structures were restrained and scaled during the molecular dynamic (MD) run. The resulting complexes were subjected to a final round of unconstrained energy minimization. Each run produced a maximum of 20 docked complexes. The ligand explored the whole h20S structure and several putative binding sites were identified as well as ligand-induced protein conformational changes were observed in the generated complexes. An example of dynamic docking results obtained for H2TPPS is reported in figure 12.

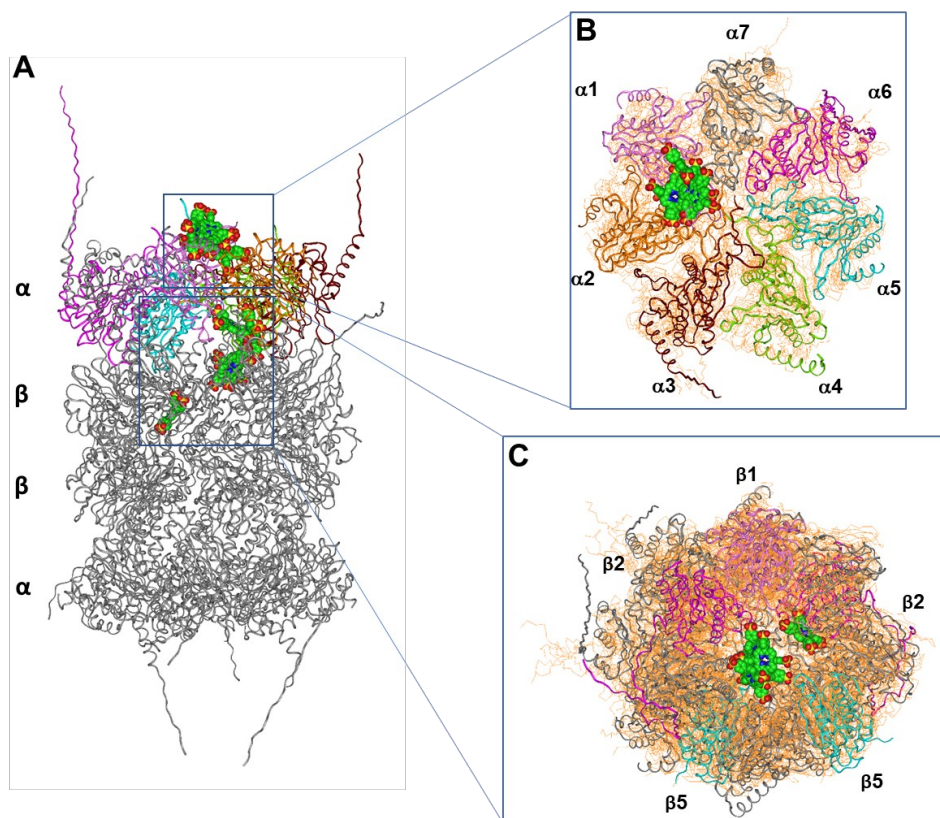


Figure 12. A) Dynamic docking results obtained for H2TPPS using as starting structure the $\alpha 1$ - $\alpha 2$ groove of the closed conformation of human 20S proteasome. For clarity only the starting protein is shown. B) Top view of dynamic docking results at the first α -ring: the first five generated complexes show H2TPPS bound at the $\alpha 1$ - $\alpha 2$ groove (although presenting a different binding mode with respect to the starting position). C) Top view of dynamic docking results without the first ring of α subunits. The backbone of the starting complex is displayed as solid ribbons and colored in pink ($\alpha 1$), orange ($\alpha 2$), brown ($\alpha 3$), light green ($\alpha 4$), cyan ($\alpha 5$), magenta ($\alpha 6$), and gray ($\alpha 7$, α subunits of the second ring and all β subunits). The backbone of the calculated complexes is displayed as line ribbons and colored in orange (B and C). The porphyrin ligands are colored by atom type (C: green; N: blue; S: yellow; O: red) and displayed as CPK. In B α subunits and the catalytic β subunits are labeled. In C the catalytic β subunits are labeled and colored in pink ($\beta 1$), violet ($\beta 2$), and cyan ($\beta 5$).

A filter was applied to the generated complexes by the conformational energy, and these complexes were ranked by the ligand-protein interaction energy, calculated for each protein residue (Tables 7-10 Appendix C).

For each run of docking calculations, the complex with the most favorable ligand-protein non-bond interaction energy was chosen (Table 5).

Table 5. Interaction energies (kcal/mol) of the best h20S-H2TPPS complexes obtained by docking calculations.

Groove	Starting h20S conformation	Ligand-Protein interaction energy (kcal/mol)		
		Tot	vdW	Electrostatic
$\alpha 1-\alpha 2$	Closed	-130.57	-80.97	-49.60
$\alpha 3-\alpha 4$	Closed	-41.07	-21.58	-19.49
$\alpha 1-\alpha 2$	Open	-96.29	-71.77	-24.52
$\alpha 3-\alpha 4$	Open	-132.15	-79.17	-52.98
$\alpha 1-\alpha 2$	Semi-open	-113.81	-68.98	-44.83
$\alpha 3-\alpha 4$	Semi-open	-113.26	-72.69	-40.56

The selected complexes were evaluated for protein structural quality by using Molprobit structure evaluator [81] and compared with the reference PDB structure (Table 6).

In agreement with the binding hypothesis, the higher ligand-protein interaction energy value was in the complexes that displayed H2TPPS bound to the $\alpha 1-\alpha 2$ and $\alpha 3-\alpha 4$ grooves (Table 5). In particular, to the $\alpha 3-\alpha 4$ groove when docking H2TPPS to the h20S open conformation and to the $\alpha 1-\alpha 2$ groove when docking H2TPPS to the h20S closed conformation. The two selected complexes showed a similar binding energy, with a ΔE of about 1.5 kcal/mol in favor of the binding to the $\alpha 3-\alpha 4$ groove in the open conformation.

Table 6. Summary of Molprobit results obtained for the experimentally determined structures of human 20S proteasome, their atomic models, and the docked H2TTPS/20S complexes.

Structure	Conformational State	Residues favored regions	Residues allowed regions	Residues outliers	Poor rotamers
X-ray (4R3O) ^a	Closed	94.9%	4.3%	0.8%	4.3%
Cryo-EM (6MSK) ^b	Open	97.0%	2.8%	0.2%	0.2%
Cryo-EM (5T0I)	Semi-open	95.5%	4.3%	0.2%	1.2%
Molecular model ^a	Closed	94.8%	4.3%	0.9%	4.1%
Molecular model ^b	Open	92.6%	6.9%	0.5%	0.9%
Molecular model	Semi-open	90.9%	8.3%	0.8%	1.5%
20S-H2TTPS (groove $\alpha 1$ - $\alpha 2$)	Closed	78.8%	18.2%	3.0%	2.4%
20S-H2TTPS (groove $\alpha 3$ - $\alpha 4$)	Closed	78.8%	18.2%	3.0%	1.8%
20S-H2TTPS (groove $\alpha 1$ - $\alpha 2$)	Open	78.0%	18.6%	3.4%	2.4%
20S-H2TTPS (groove $\alpha 3$ - $\alpha 4$)	Open	79.1%	17.2%	3.7%	2.3%
20S-H2TTPS (groove $\alpha 1$ - $\alpha 2$)	Semi-open	77.4%	19.5%	3.1%	1.9%
20S-H2TTPS (groove $\alpha 3$ - $\alpha 4$)	Semi-open	79.0%	18.2%	2.8%	2.3%

^aData from reference [71]. ^bData from reference [72].

On the contrary, in the case of binding to the semi-open conformation the best docked complexes displayed H2TTPS bound at $\alpha 1$ - $\alpha 2$ or $\alpha 3$ - $\alpha 4$ grooves (Figures 13-14) with a very similar interaction energy ($\Delta E \sim 0.5$ kcal/mol), but less favorable compared to the one calculated for the binding with the open ($\alpha 3$ - $\alpha 4$) and closed ($\alpha 1$ - $\alpha 2$) h20S conformations ($\Delta E \sim 18$ kcal/mol; Table 6).

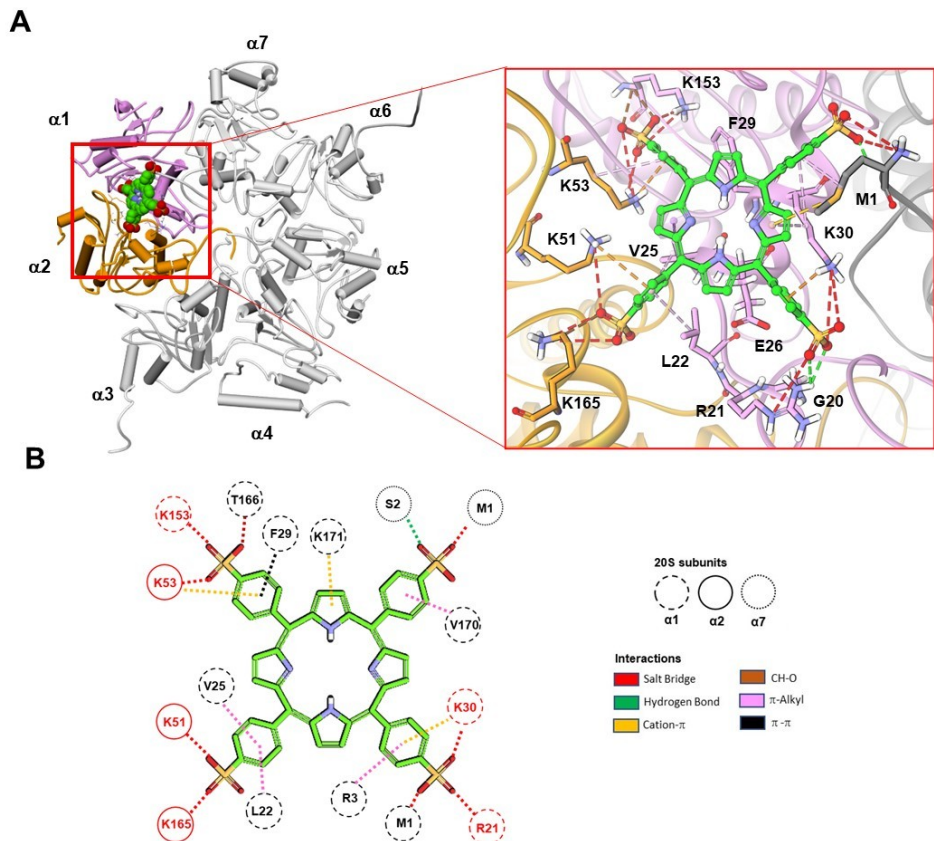


Figure 13. Selected docked complex of H2TPPS bound to h20S semi-open conformation at the $\alpha 1$ - $\alpha 2$ groove. A: top-view of the α -ring (left) and close-up view of the binding site (right). The interacting residues involved in ionic interactions or engaged in the binding of h20S to endogenous RPs and the small activator ZYA are displayed in stick. H2TPPS is displayed in CPK (left) and ball&stick (right) and colored by atom type (C = green, O = red, N = blue, S = orange, H = white). Only polar hydrogens are displayed for clarity. Interactions are evidenced with dashed lines. H20S α -ring is displayed as wide cylinders (α -helices), arrows (beta-sheets), and tubes (coils and turns) (left) or solid ribbons (right) and colored in grey with the exception of the $\alpha 1$ subunit colored in pink and the $\alpha 2$ subunit colored in orange. B: 2D representation of the interactions between h20S (semi-open state) and H2TPPS bound at the $\alpha 1$ - $\alpha 2$ groove.

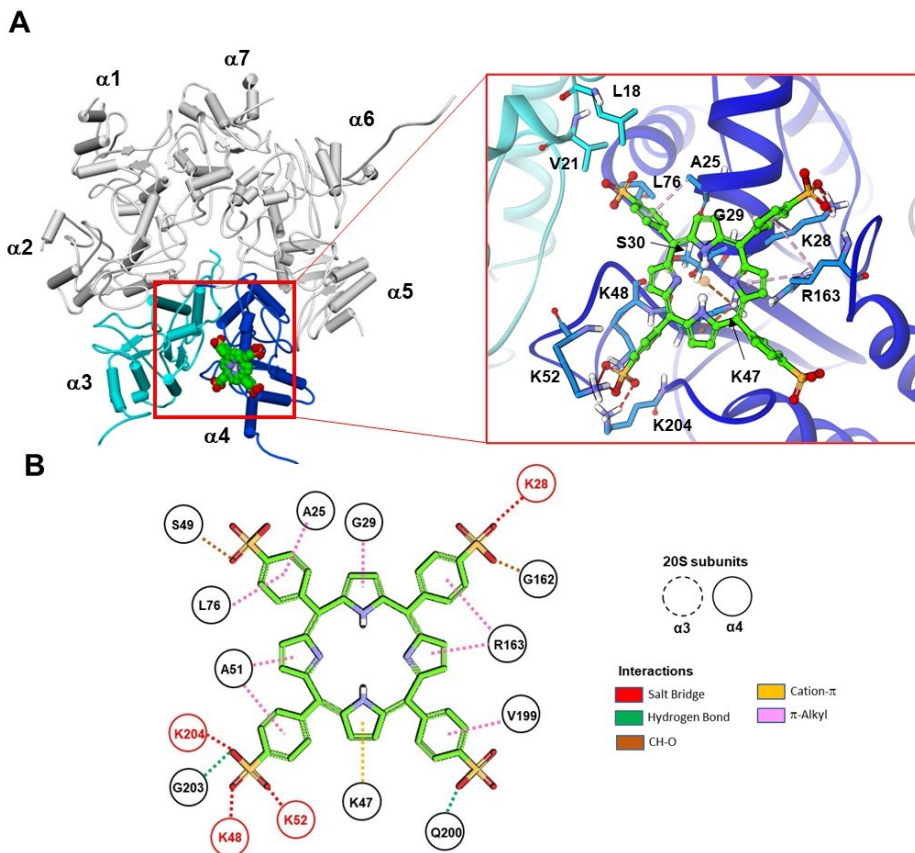


Figure 14. Selected docked complex of H2TPPS bound to h20S semi-open conformation at the $\alpha 3$ - $\alpha 4$ groove. A: top-view of the α -ring (left) and close-up view of the binding site (right). The interacting residues involved in ionic interactions or engaged in the binding of h20S to endogenous RPs and the small activator ZYA are displayed in stick. H2TPPS is displayed in CPK (left) and ball&stick (right) and colored by atom type (C = green, O = red, N = blue, S = orange, H = white). Only polar hydrogens are displayed for clarity. Interactions are evidenced with dashed lines. H20S α -ring is displayed as wide cylinders (α -helices), arrows (beta-sheets), and tubes (coils and turns) (left) or solid ribbons (right) and colored in grey with the exception of the $\alpha 3$ subunit colored in cyan and the $\alpha 4$ subunit colored in blue. B: 2D representation of the interactions between h20S (semi-open state) and H2TPPS bound at the $\alpha 3$ - $\alpha 4$ groove.

In Figure 15 we can see that, when H2TPPS was bound to the $\alpha 1$ - $\alpha 2$ groove of closed h20S there was the displacement of the N-terminal tail of $\alpha 2$, $\alpha 3$ and $\alpha 4$ from the center of the channel, opening partially the central gate. The N-terminal region of $\alpha 2$ undergoes the most pronounced conformational rearrangement, shifting from $\alpha 7$ (Tyr119) toward $\alpha 3$ (Tyr121), whereas the N-termini of $\alpha 3$ and $\alpha 4$ lose key interactions with the $\alpha 5$ - and $\alpha 6$ -loops of the α -anulus, respectively (R3 $\alpha 3$ -Y123 $\alpha 6$ and E126 $\alpha 5$ -R5 $\alpha 4$; Table 7).

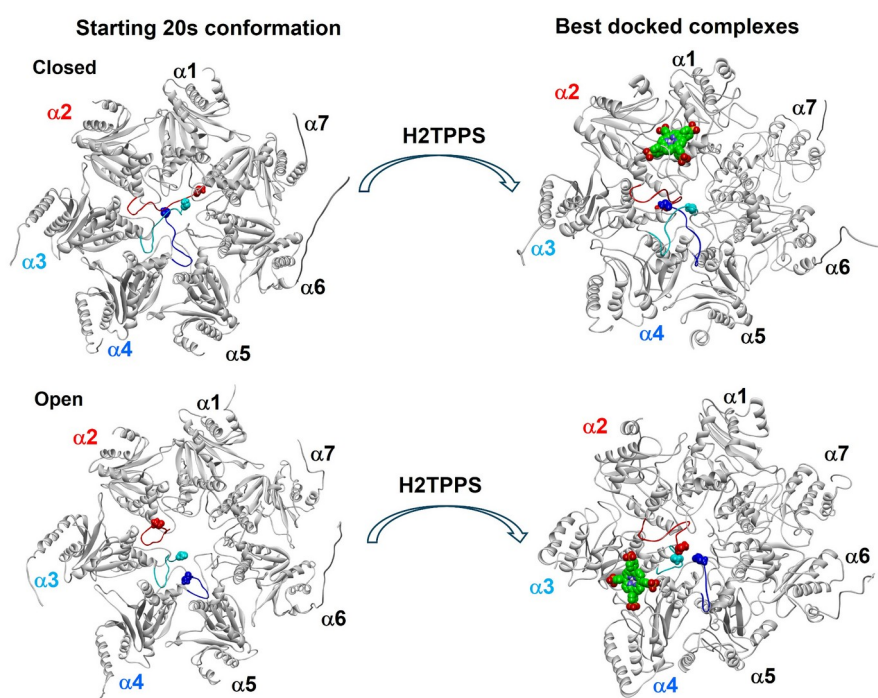


Figure 15. Results of dynamic docking of H2TPPS to h20S: starting protein conformation and final docked complex (top view; only the α -ring is displayed for sake of clarity). The N-terminal tails of the $\alpha 2$, $\alpha 3$, and $\alpha 4$ subunits of h20S are colored in red, cyan, and blue, respectively, and their first residue displayed as CPK. H2TPPS is displayed as CPK and colored by atom type, h20S secondary structure is displayed as white ribbon.

Table 7. Molecular interactions involving the N-terminals, $\alpha 5$ -loop, and $\alpha 6$ -loop present in the experimentally determined structures of human 20S proteasome wt closed (PDB ID: 4R3O), wt open (6MSK) and wt semi-open (PDB ID: 5T0I) and in the selected docked complex.

Structure	Interactions
<i>Human 20S closed</i>	R3 α 3-Y123 α 6 E125 α 5-R3 α 3 E126 α 5-R5 α 4 D127 α 5- R122 α 6 D127 α 5- R125 α 6
<i>Human 20S semi-open</i>	D127 α 5- R4 α 2 D127 α 5- R122 α 6 D127 α 5- R125 α 6
<i>Human 20S open</i>	E125 α 5- R125 α 6 E125 α 5- R126 α 6
<i>Docked complex of H2TPPS bound to h20S closed conformation</i>	R4 α 2 – E126 α 5 E125 α 5-R3 α 3 D127 α 5- R125 α 6
<i>Docked complex of H2TPPS bound to h20S open conformation</i>	R4 α 2 – D6 α 3 D127 α 5- R122 α 6 E125 α 5-R125 α 6
<i>Docked complex of H2TPPS bound to h20S semi- open conformation ($\alpha 1$-$\alpha 2$)</i>	E126 α 5- R125 α 6 D127 α 5- R4 α 2
<i>Docked complex of H2TPPS bound to h20S semi- open conformation ($\alpha 3$-$\alpha 4$)</i>	E125 α 5- R126 α 6 E126 α 5- R125 α 6 D127 α 5-R125 α 6

On the other hand, considering the h20S open conformation, H2TPPS preferentially binds to the $\alpha 3$ - $\alpha 4$ groove forming a stable complex with the gate open.

Then, the comparison of the molecular interactions formed by H2TPPS with h20S with those formed by PA28, PA200, 19S and a small

peptidomimetic ligand named ZYA that mimic the activating HbYX-COOH motif [36] was performed (Appendix C: Tables 7-10). It resulted that when H2TPPS is bound at the $\alpha 3$ - $\alpha 4$ groove (starting conformation: open), it interacts with seven positively charged residue that are also involved in the binding of h20S to endogenous RPs and ZYA (i.e. R17 ($\alpha 3$), K27 ($\alpha 4$), K28 ($\alpha 4$), K47 ($\alpha 4$), K48 ($\alpha 4$), K52 ($\alpha 4$), and R163 ($\alpha 4$) (Figure 16) (Appendix C: Table 8)).

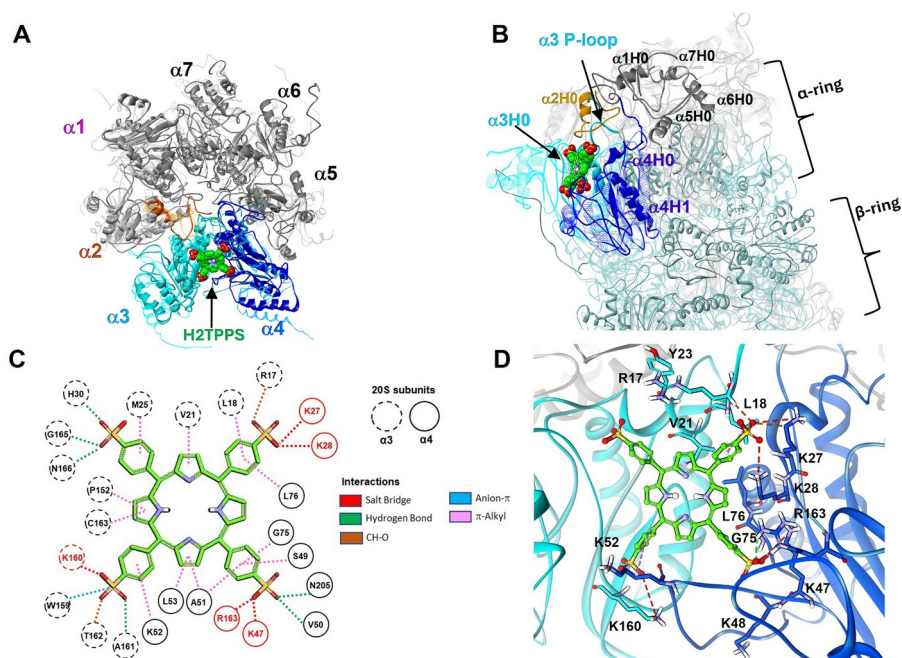


Figure 16. Selected complex of H2TPPS docked to h20S open conformation. A: top-view of the α -ring of the resulting complex (solid ribbons) superimposed on the starting h20S structure (line ribbons), H2TPPS is displayed in CPK. B: lateral view of the H2TPPS-h20S complex. C: schematic representation of the interactions between H2TPPS and h20S residues. D: close up view of H2TPPS (ball&stick; carbons: green) bound to h20S $\alpha 3$ - $\alpha 4$ groove; only protein residues establishing ionic interactions and/or also involved in the binding of h20S to endogenous RPs or ZYA are displayed and labeled. Only polar hydrogens are displayed for clarity.

In a similar way, when H2TPPS is bound to the $\alpha 1$ - $\alpha 2$ (starting conformation: closed), it interacts with five positively charged residues which are also involved in the binding of h20S to endogenous RPs and ZYA(R21 ($\alpha 1$), K30($\alpha 1$), K51 ($\alpha 2$), K53 ($\alpha 2$), and K64 ($\alpha 2$)) (Figure 17, Table 7 Appendix C).

These findings support the hypothesis [72] that charged porphyrins employ a sort of electrostatic code located on the surface of the α ring, which is necessary to the anchoring of RPs and takes part to the allosteric network leading to gate opening [73,74]. Specifically, R21 ($\alpha 1$) and R17 ($\alpha 3$) are part of the inverse Pro-loop located at the N-terminal of the α -subunits (P-loop), while K20 ($\alpha 1$) and K64 ($\alpha 2$) were reported as crucial for the activation of 20S by the C-terminal tails of 19S and PA200 [14, 54]. The interaction of the P-loop with an invariant PA28 motif known as the activation loop located between helices 2 and 3, motif that contains a conserved glutamate residue, is responsible for the opening of the gate [79]. In a similar way the conformational change of the P-loop induced by the ligand in the subunits $\alpha 2$ -, $\alpha 3$ - and $\alpha 4$ -subunits causes the displacement of the corresponding N-terminal tails from the center of the gate when the h20S closed conformation is docked by H2TPPS (Figures 15 and 17A-B).

If we consider also non-charged residues, H2TPPS is engaged in interaction with other residues involved in the binding of h20S to endogenous RPs (Figures 16-17; Tables 7-8 Appendix C), including the highly conserved hydrophobic residues at the edge of the h20S $\alpha H1$ helix (M79 ($\alpha 2$); L76 ($\alpha 4$)), which is proved to be a key element in the allosteric network that lead to gate opening [80] and to be involved in the binding with the activating motif HbYX [28,26].

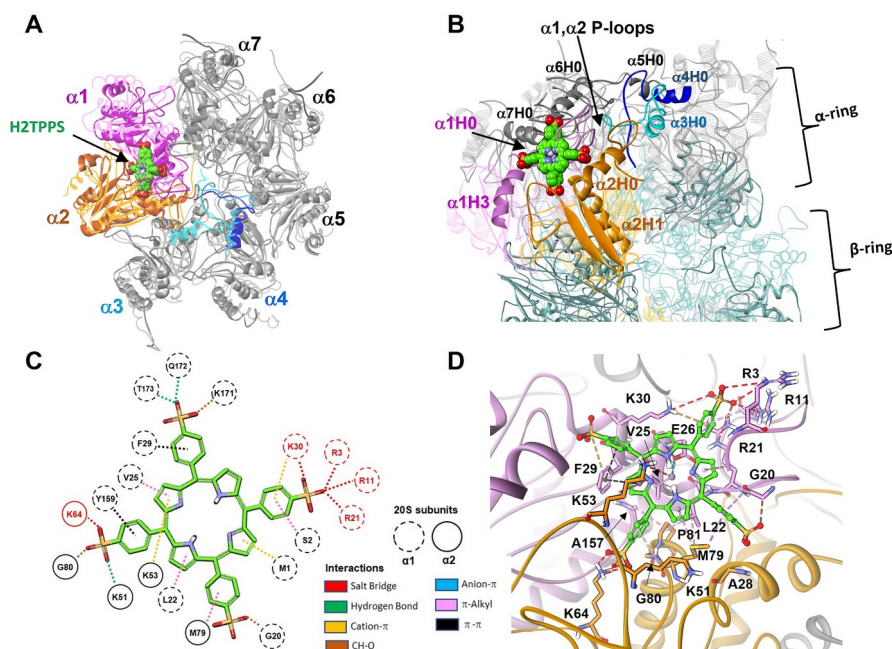


Figure 17. Selected complex of H2TPPS docked to h20S closed conformation. A: top-view of the α -ring of the resulting complex (solid ribbons) superimposed on the starting h20S structure (line ribbons), H2TPPS is displayed in CPK. B: lateral view of the H2TPPS-h20S complex. C: schematic representation of the interactions between H2TPPS and h20S residues. D: close up view of H2TPPS (ball&stick; carbons: green) bound to h20S $\alpha 1$ - $\alpha 2$ groove; only protein residues establishing ionic interactions and/or also involved in the binding of h20S to endogenous RPs or ZYA are displayed and labeled. Only polar hydrogens are displayed for clarity.

Noteworthy, also TrisT4 and TMPC, the previously studied porphyrins, reproduced this crucial interaction, but they were not able to interact with the P-loop and consequently and, accordingly, did not induce the displacement of the $\alpha 2$ -, $\alpha 3$ -, and $\alpha 4$ -N-terminals from the center of the gate, as was observed when H2TPPS is docked to the closed conformation (Figures 15 and 17A-B).

Previous molecular docking studies on porphyrins (i.e., TrisT4 and TMPC) [72-73] or other small-ligand allosteric modulators [38, 52, 65] evidenced that these molecules act as proteasome activators interacting with different α -grooves of proteasome. The computational studies reported in this chapter confirmed these results evidencing a scenario in which H2TPPS could bind to both the closed and open conformation of the h20S with a comparable binding energy but preferring a different α -groove according to the h20S conformational state. By binding to the $\alpha 1$ - $\alpha 2$ groove of closed h20S, H2TPPS promotes h20S gate opening exploiting the RP interaction sites (“induced fit activation mechanism”). At the same time, the binding of H2TPPS to the $\alpha 3$ - $\alpha 4$ groove of h20S could stabilize the open gate conformation (“conformational selection activation mechanism”). Both phenomena could contribute to the observed shift of the equilibrium toward the opening of the h20S gate in the presence of the ligand and the consequent augmentation of the catalytic activity. Thus, the computational data are in full agreement with the experimental results obtained with AFM and enzymatic activity studies. Indeed, experimental results indicate that treatments with the high concentration ($\geq 2 \mu\text{M}$) of H2TPPS lead to the lowest achievable participation of closed-gate forms and the highest participation of intermediate and open-gate conformers. Interestingly, such partition is observed also for the substrate-treated proteasomes fully engaged in catalytic activity.

When comparing the binding properties of H2TPPS with other porphyrin-based activators, such as Tris-T4, it is noteworthy that H2TPPS exhibits a significantly higher affinity (approximately tenfold) for the substrate-bound h20S proteasome compared to Tris-T4 [75]. This enhanced interaction is attributed to the presence of negatively charged groups,

which confer a distinct binding mode to the α -ring of h20S, differing from that of previously studied activating porphyrins.

Docking studies suggest that H2TPPS not only reproduces key interactions of the C-terminal HbYX motif found in regulatory particles (RPs), but also, when bound to the $\alpha 1$ – $\alpha 2$ groove of the closed h20S conformation, engages with the P-loop and promotes the displacement of the N-terminal tails of $\alpha 2$, $\alpha 3$, and $\alpha 4$ subunits from the central gate. This mechanism closely resembles the activation loop of PA28. Such behavior may explain why only two H2TPPS molecules are sufficient to induce the observed enhancement in proteasome activity, whereas three molecules of Tris-T4 are required to achieve full activation [75].

Furthermore, since only two H2TPPS molecules are accommodated within the grooves, a less crowded occupancy of the α -ring is expected compared to Tris-T4. This reduced steric hindrance could account for the absence of inhibitory effects at high substrate concentrations, a phenomenon observed with Tris-T4.

4.3. Conclusions.

A complementary set of experimental and computational approaches has consistently demonstrated that H2TPPS acts as an allosteric modulator of the h20S proteasome, promoting the activation of its chymotrypsin-like (ChT-L) activity at low micromolar concentrations. The molecular modelling results were found to be in excellent agreement with experimental observations, indicating that H2TPPS binding shifts the conformational equilibrium of the 20S proteasome toward the open state of the substrate gate. The molecular fit between the tetra-anionic pharmacophore of H2TPPS and the positively charged surface residues

involved in electrostatic interactions with regulatory particles (RPs) confirms that H2TPPS is capable of mimicking the key structural determinants responsible for RP anchoring to the 20S core. Remarkably, unlike previously investigated RP-derived activators (including other charged porphyrins) H2TPPS reproduces not only the interactions mediated by the C-terminal HbYX motif of 19S and PA200, but also those associated with the activation loop of PA28, which contains conserved negatively charged glutamate residues. As a consequence, H2TPPS displays a higher activation efficiency compared to other porphyrin-based modulators such as TrisT4.

Taken together, the present findings support the rational discovery of small ligands mimicking key structural elements of 20S RPs as a feasible approach for the rational discovery of new chemical scaffolds acting as proteasome activators. Given its structural and functional properties, H2TPPS may serve as a novel lead compound for drug-discovery programs aimed at activating proteasomal degradation in the context of neurodegenerative disease therapy.

4.5. References.

1. Ciechanover, A. "Proteolysis: From the Lysosome to Ubiquitin and the Proteasome." *Nature Reviews Molecular Cell Biology*, vol. 6, no. 1, 2005, pp. 79-87.
2. Glickman, M.H., and A. Ciechanover. "The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction." *Physiological Reviews*, vol. 82, no. 2, 2002, pp. 373-428.
3. Livneh, I., et al. "The Life Cycle of the 26S Proteasome: From Birth, Through Regulation and Function, and Onto its Death." *Cell Research*, vol. 26, no. 8, 2016, pp. 869-85.
4. Tomko, R.J., and M. Hochstrasser. "Molecular Architecture and Assembly of the Eukaryotic Proteasome." *Annual Review of Biochemistry*, vol. 82, no. 1, 2013, pp. 415-45.
5. Klaips, C.L., et al. "Pathways of Cellular Proteostasis in Aging and Disease." *The Journal of Cell Biology*, vol. 217, no. 1, 2018, pp. 51-63.
6. Thibaudau, T.A., and D.M. Smith. "A Practical Review of Proteasome Pharmacology." *Pharmacological Reviews*, vol. 71, no. 2, 2019, pp. 170-97.
7. Tundo, G.R., et al. "The Proteasome as a Druggable Target with Multiple Therapeutic Potentialities: Cutting and Non-Cutting Edges." *Pharmacology & Therapeutics*, vol. 213, 2020, article 107579.
8. Ciechanover, A., and A.L. Schwartz. "The Ubiquitin-Proteasome Pathway: The Complexity and Myriad Functions of Proteins Death." *Proceedings of the National Academy of Sciences*, vol. 95, no. 6, 1998, pp. 2727-30.
9. Hershko, A., and A. Ciechanover. "The Ubiquitin System." *Annual Review of Biochemistry*, vol. 67, 1998, pp. 425-79.
10. Komander, D., and M. Rape. "The Ubiquitin Code." *Annual Review of Biochemistry*, vol. 81, 2012, pp. 203-29.
11. Finley, D. "Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome." *Annual Review of Biochemistry*, vol. 78, 2009, pp. 477-513.
12. Collins, G.A., and A.L. Goldberg. "The Logic of the 26S Proteasome." *Cell*, vol. 169, no. 5, 2017, pp. 792-806.
13. Schweitzer, A., et al. "Structure of the Human 26S Proteasome at a Resolution of 3.9 Å." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 28, 2016, pp. 7816-21.
14. Toste Rêgo, A., and P.C.A. da Fonseca. "Characterization of Fully Recombinant Human 20S and 20S-PA200 Proteasome Complexes." *Molecular Cell*, vol. 76, no. 1, 2019, pp. 138-147.e5.
15. Zhao, J., et al. "Structural Insights into the Human PA28-20S Proteasome Enabled by Efficient Tagging and Purification of Endogenous Proteins." *Proceedings of the National Academy of Sciences*

- of the United States of America*, vol. 119, no. 33, 2022, article e220720011.
16. Kalderon, B., et al. "Cytosolic HSP60 Can Modulate Proteasome Activity in Yeast." *Journal of Biological Chemistry*, vol. 290, no. 6, 2015, pp. 3542-51.
 17. Sbardella, D., et al. "The Insulin-Degrading Enzyme is an Allosteric Modulator of..." *Note: Title incomplete in source.*
 18. Deshmukh, F.K., et al. "Allosteric Regulation of the 20S Proteasome by the Catalytic Core Regulators (CCRS) Family." *Nature Communications*, vol. 14, no. 1, 2023, article 3126.
 19. Gaczynska, M., and P.A. Osmulski. "Targeting Protein-Protein Interactions in the Ubiquitin-Proteasome Pathway." *Advances in Protein Chemistry and Structural Biology*, vol. 110, 2018, pp. 123-65.
 20. Guerrero, C., et al. "Characterization of the Proteasome Interaction Network Using a QTAX-Based Tag-Team Strategy and Protein Interaction Network Analysis." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 36, 2008, pp. 13333-38.
 21. Groll, M., et al. "Structure of 20S Proteasome from Yeast at 2.4 Å Resolution." *Nature*, vol. 386, no. 6624, 1997, pp. 463-71.
 22. Harshbarger, W., et al. "Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib." *Structure*, vol. 23, no. 2, 2015, pp. 418-24.
 23. Groll, M., et al. "Structure of 20S Proteasome from Yeast at 2.4 Å Resolution." *Nature*, vol. 386, no. 6624, 1997, pp. 463-71.
 24. Harshbarger, W., et al. "Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib." *Structure*, vol. 23, no. 2, 2015, pp. 418-24.
 25. Groll, M., et al. "The Catalytic Sites of 20S Proteasomes and Their Role in Subunit Maturation: A Mutational and Crystallographic Study." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 20, 1999, pp. 10976-83.
 26. Chen, S., et al. "Structural Basis for Dynamic Regulation of the Human 26S Proteasome." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 46, 2016, pp. 12991-96.
 27. Chuah, J.J., et al. "High Resolution Structures Define Divergent and Convergent Mechanisms of Archaeal Proteasome Activation." *Communications Biology*, vol. 6, no. 1, 2023, article 733.
 28. Collins, G.A., and A.L. Goldberg. "The Logic of the 26S Proteasome." *Cell*, vol. 169, no. 5, 2017, pp. 792-806.
 29. Dong, Y., et al. "Cryo-EM Structures and Dynamics of Substrate-Engaged Human 26S Proteasome." *Nature*, vol. 565, no. 7737, 2019, pp. 49-55.
 30. Whitby, F.G., et al. "Structural Basis for the Activation of 20S Proteasomes by 11S Regulators." *Nature*, vol. 408, no. 6808, 2000, pp. 115-20.

31. Tundo, Grazia Raffaella, et al. "The insulin-degrading enzyme from structure to allosteric modulation: new perspectives for drug design." *Biomolecules* 13.10 (2023): 1492.
32. Zhang, Z., et al. "Identification of an Activation Region in the Proteasome Activator REGα." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, 1998, pp. 2807-11.
33. Sbardella, D., et al. "Human Matrix Metalloproteinases: An Ubiquitarian Class of Enzymes Involved in Several Pathological Processes." *Molecular Aspects of Medicine*, vol. 33, 2012, pp. 119-208.
34. Tundo, G.R., et al. "Multiple Functions of Insulin-Degrading Enzyme: A Metabolic Crosslight?" *Critical Reviews in Biochemistry and Molecular Biology*, vol. 52, 2017, pp. 554-82.
35. Cromm, P.M., and C.M. Crews. "The Proteasome in Modern Drug Discovery: Second Life of a Highly Valuable Drug Target." *ACS Central Science*, vol. 3, no. 8, 2017, pp. 830-38.
36. Fernández-Cruz, I., and E. Reynaud. "Proteasome Subunits Involved in Neurodegenerative Diseases." *Archives of Medical Research*, vol. 52, no. 1, 2021, pp. 1-14.
37. Thibaudau, T.A., et al. "A Common Mechanism of Proteasome Impairment by Neurodegenerative Disease-Associated Oligomers." *Nature Communications*, vol. 9, no. 1, 2018, article 1097.
38. Chuah, J.J., et al. "Minimal Mechanistic Component of HbYX-Dependent Proteasome Activation that Reverses Impairment by Neurodegenerative-Associated Oligomers." *Communications Biology*, vol. 6, no. 1, 2023, article 725.
39. Fiolek, T.J., et al. "Fluspirilene Analogs Activate the 20S Proteasome and Overcome Proteasome Impairment by Intrinsically Disordered Protein Oligomers." *ACS Chemical Neuroscience*, vol. 12, no. 8, 2021, pp. 1438-48.
40. Njomen, E., and J.J. Tepe. "Proteasome Activation as a New Therapeutic Approach to Target Proteotoxic Disorders." *Journal of Medicinal Chemistry*, vol. 62, no. 14, 2019, pp. 6469-81.
41. Wang, X., and H. Wang. "Priming the Proteasome to Protect Against Proteotoxicity." *Trends in Molecular Medicine*, vol. 26, no. 7, 2020, pp. 639-48.
42. Chocron, E.S., et al. "Genetic and Pharmacologic Proteasome Augmentation Ameliorates Alzheimer's-Like Pathology in Mouse and Fly APP Overexpression Models." *Science Advances*, vol. 8, no. 23, 2022, article eabk2252.
43. Hipp, M.S., et al. "The Proteostasis Network and its Decline in Ageing." *Nature Reviews Molecular Cell Biology*, vol. 20, no. 7, 2019, pp. 421-35.
44. Du, Z., et al. "Probing the Unfolded Protein Response in Long-Lived Naked Mole-Rats." *Biochemical and Biophysical Research Communications*, vol. 529, no. 4, 2020, pp. 1151-57.

45. Chondrogianni, N., et al. "Fibroblast Cultures from Healthy Centenarians Have an Active Proteasome." *Experimental Gerontology*, vol. 35, no. 6-7, 2000, pp. 721-28.
46. Rodriguez, K.A., et al. "Altered Composition of Liver Proteasome Assemblies Contributes to Enhanced Proteasome Activity in the Exceptionally Long-Lived Naked Mole-Rat." *PLoS One*, vol. 7, no. 5, 2012, article e35890.
47. Dagar, G., et al. "Ubiquitination and Deubiquitination: Implications on Cancer Therapy." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1866, no. 4, 2023, article 194979.
48. Mathien, S., et al. "Regulation of Mitogen-Activated Protein Kinase Signaling Pathways by the Ubiquitin-Proteasome System and Its Pharmacological Potential." *Pharmacological Reviews*, vol. 73, no. 4, 2021, pp. 263-96.
49. Zheng, L-L., et al. "Recent Advances in the Development of Deubiquitinases Inhibitors as Antitumor Agents." *European Journal of Medicinal Chemistry*, vol. 266, 2024, article 116161.
50. Cekała, K., et al. "Peptidomimetics Based on C-Terminus of BLM10 Stimulate Human 20S Proteasome Activity and Promote Degradation of Proteins." *Biomolecules*, vol. 12, no. 6, 2022, article 777.
51. Gillette, T.G., et al. "Differential Roles of the COOH Termini of AAA Subunits of PA700 (19 S Regulator) in Asymmetric Assembly and Activation of the 26 S Proteasome." *Journal of Biological Chemistry*, vol. 283, no. 46, 2008, pp. 31813-22.
52. Jones, C.L., and J.J. Tepe. "Proteasome Activation to Combat Proteotoxicity." *Molecules*, vol. 24, no. 15, 2019, article 2841.
53. Opoku-Nsiah, K.A., et al. "The YΦ Motif Defines the Structure-Activity Relationships of Human 20S Proteasome Activators." *Nature Communications*, vol. 13, no. 1, 2022, article 1226.
54. Smith, D.M., et al. "Docking of the Proteasomal ATPases' Carboxyl Termini in the 20S Proteasome's α Ring Opens the Gate for Substrate Entry." *Molecular Cell*, vol. 27, no. 5, 2007, pp. 731-44.
55. Farzan, M., et al. "Neuroprotective Properties of Betulin, Betulinic Acid, and Ursolic Acid as Triterpenoids Derivatives: A Comprehensive Review of Mechanistic Studies." *Nutritional Neuroscience*, vol. 27, no. 3, 2024, pp. 223-40.
56. Dang, Z., et al. "Synthesis of Lithocholic Acid Derivatives as Proteasome Regulators." *ACS Medicinal Chemistry Letters*, vol. 3, no. 11, 2012, pp. 925-30.
57. Yilmaz, S., et al. "The Role of Cycloastragenol at the Intersection of Nrf2/ARE, Telomerase, and Proteasome Activity." *Free Radical Biology and Medicine*, vol. 188, 2022, pp. 105-16.
58. Carmona, V., et al. "Several Targets Involved in Alzheimer's Disease Amyloidogenesis are Affected by Morin and Isoquercitrin." *Nutritional Neuroscience*, vol. 23, no. 8, 2020, pp. 575-90.

59. Khan, T.K., et al. "Modulation of Proteasome Activity by Curcumin and Didemethylcurcumin." *Journal of Biomolecular Structure and Dynamics*, vol. 40, no. 18, 2022, pp. 8332-39.
60. Amici, M., et al. "Binding of Aflatoxins to the 20S Proteasome: Effects on Enzyme Functionality and Implications for Oxidative Stress and Apoptosis." *Biological Chemistry*, vol. 388, no. 1, 2007, pp. 107-17.
61. Persico, M., et al. "Silybins are Stereospecific Regulators of the 20S Proteasome." *Bioorganic & Medicinal Chemistry*, vol. 66, 2022, article 116813.
62. Chondrogianni, N., et al. "Anti-Aging Properties of the Olive Constituent Oleuropein in Human Cells." *Olives and Olive Oil in Health and Disease Prevention*, Academic Press, 2010, pp. 1335-43.
63. Sbardella, D., et al. "Structural and Functional Evidence for Citicoline Binding and Modulation of 20S Proteasome Activity: Novel Insights into its Pro-Proteostatic Effect." *Biochemical Pharmacology*, vol. 177, 2020, article 113977.
64. Sadahiro, Y., et al. "Syrotingopine Enhances 20S Proteasome Activity and Degradation of α -Synuclein." *Journal of Natural Products*, vol. 87, no. 3, 2024, pp. 554-59.
65. George, D.E., and J.J. Tepe. "Advances in Proteasome Enhancement by Small Molecules." *Biomolecules*, vol. 11, no. 12, 2021, article 1789.
66. Santoro, A.M., et al. "Pyrazolones Activate the Proteasome by Gating Mechanisms and Protect Neuronal Cells from β -Amyloid Toxicity." *ChemMedChem*, vol. 15, no. 3, 2020, pp. 302-16.
67. Njomen, E., et al. "Small Molecule Modulation of Proteasome Assembly." *Biochemistry*, vol. 57, no. 28, 2018, pp. 4214-24.
68. Liao, E.E., et al. "Proteasomal Stimulation by MK-886 and Its Derivatives Can Rescue Tau-Induced Neurite Pathology." *Molecular Neurobiology*, vol. 60, no. 10, 2023, pp. 6133-44.
69. Halder, S., et al. "Oleic Amide Derivatives as Small Molecule Stimulators of the Human Proteasome's Core Particle." *RSC Medicinal Chemistry*, vol. 13, no. 9, 2022, pp. 1077-81.
70. Ban, H.S., et al. "Discovery of Carboranes as Inducers of 20S Proteasome Activity." *ChemMedChem*, vol. 5, no. 8, 2010, pp. 1236-41.
71. Coleman, R.A., and D.J. Trader. "Development and Application of a Sensitive Peptide Reporter to Discover 20S Proteasome Stimulators." *ACS Combinatorial Science*, vol. 20, no. 5, 2018, pp. 269-76.
72. Di Dato, A., et al. "Electrostatic Map of Proteasome α -Rings Encodes the Design of Allosteric Porphyrin-Based Inhibitors Able to Affect 20S Conformation by Cooperative Binding." *Scientific Reports*, vol. 7, no. 1, 2017, article 17098.
73. Persico, M., et al. "Modulation of the 20S Proteasome Activity by Porphyrin Derivatives is Steered Through Their Charge Distribution." *Biomolecules*, vol. 12, no. 6, 2022, article 741.
74. Santoro, A.M., et al. "Cationic Porphyrins are Tunable Gatekeepers of the 20S Proteasome." *Chemical Science*, vol. 7, no. 2, 2016, pp. 1286-97.

75. Santoro, A.M., et al. "Cooperative Binding of the Cationic Porphyrin tris-T4 Enhances Catalytic Activity of 20S Proteasome Unveiling a Complex Distribution of Functional States." *International Journal of Molecular Sciences*, vol. 21, no. 19, 2020, article 7190.
76. Santoro, A.M., et al. "Tetra-Anionic Porphyrin Mimics Protein-Protein Interactions Between Regulatory Particles and the Catalytic Core, Allosterically Activating Human 20S Proteasome." *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 40, no. 1, 2025, article 2482892.
77. Giżyńska, M., et al. "Proline- and Arginine-Rich Peptides as Flexible Allosteric Modulators of Human Proteasome Activity." *Journal of Medicinal Chemistry*, vol. 62, no. 1, 2019, pp. 359-70.
78. Giletto, M.B., et al. "Pipelicolic Esters as Minimized Templates for Proteasome Inhibition." *Organic & Biomolecular Chemistry*, vol. 17, no. 10, 2019, pp. 2734-46.
79. Lesne, J., et al. "Conformational Maps of Human 20S Proteasomes Reveal PA28- and Immuno-Dependent Inter-Ring Crosstalks." *Nature Communications*, vol. 11, no. 1, 2020, article 6140.
80. Ruschak, A.M., and L.E. Kay. "Proteasome Allostery as a Population Shift Between Interchanging Conformers." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 50, 2012, pp. E3454-62.
81. Vincent B. Chen, W. Bryan Arendall III, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson and David C. Richardson (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D* 66: 12-21.

Chapter 5. Molecular Modelling of 2-Phenylimidazo[1,2-b]pyridazine Derivatives Exhibiting In Vitro and In Vivo Pharmacological Activity

5.1 Introduction

The imidazo[1,2-b]pyridazine system is considered a privileged scaffold in medicinal chemistry due to its broad clinical potential in both therapeutic and diagnostic applications [1-8], showing promise in the development of anticancer, antineuropathic, antipsychotic, antiparasitic, antibacterial, antiviral, and anti-inflammatory agents [9]. Extensive research efforts have been devoted to the design, synthesis, and biological evaluation of compounds containing related heterocyclic moieties [10–13].

In this context, two derivatives bearing a 2-phenylimidazo[1,2-b]pyridazine core—namely 7-methyl-2-phenylimidazo[1,2-b]pyridazin-3-carboxylic acid (DM1) and 6-methoxy-2-phenylimidazo[1,2-b]pyridazin-3-carboxylic acid (DM2)—have shown in vivo anti-absence and antinociceptive activity [14,15]. In particular, DM1 and DM2 demonstrated potent anti-absence effects by suppressing spike-and-wave discharges (SWDs) in WAG/Rij rats, a validated model of absence epilepsy [14].

The mechanism of action of these compounds is related to their ability to bind and block human T-type voltage-dependent calcium channels (hCav3.1). Specifically, when tested on Ba²⁺ currents elicited by step depolarization in HEK-293 cells stably expressing hCav3.1 channels, DM1 and DM2 showed significant inhibitory activity at the highest

concentration tested (300 μ M). DM1 reduced Ba²⁺ currents to 32.3% $\pm$ 2.7% of control values, while DM2 reduced them to 39.9% $\pm$ 4.15%. [15].

In the present chapter, molecular modelling studies on DM1 and DM2, such as ligand conformational analysis, DFT calculations, and docking simulations, are reported. These investigations were conducted to explore the interaction of DM1 and DM2 with hCAV_{3.1} channels, laying the groundwork for future structural optimization of these promising lead compounds. Indeed, the new *in vitro* and *in vivo* pharmacological investigations confirmed their antiepileptic potential. In particular, DM1 and DM2 were tested in CD-1 ICR mice after pentylenetetrazol (PTZ) administration and in DBA/2 mice with seizure induced by audiogenic stimuli. Since the seizures activity generates reactive oxygen species (ROS) [16-19] that contributes to neuronal damage and cell death [20], their protective effects against oxidative stress in C6 rat brain glioma cells were also evaluated.

This research was performed in a multidisciplinary framework, with the group of Prof. Maria Grazia Rimoli and the group of Prof. Rosaria Meli from the University of Naples for the synthesis and the pharmacological assays, respectively. The results were published in a scientific paper [21].

I was directly involved in performing conformational analysis of the ligands and conducting docking simulations, as well as in the interpretation of computational results.

The optimization of imidazo[1,2-b]pyridazine scaffold synthesis and the experimental procedure of pharmacological studies are not reported herein but can be found in the publication “Eco-friendly Synthesis and Molecular Modelling of 2-Phenylimidazo[1,2-b]pyridazine Derivatives: *In*

Vitro and *In Vivo* Studies for Lead Optimization” *ChemMedChem* 20.5 (2025): e202400721 [21].

5.2. Results and Discussion.

5.2.1. Molecular Modelling Studies.

Previously it was demonstrated that DM1 and DM2 can block the hCav_{3.1} isoform of T-type VGCCs, thus giving the anti-absence effect *in vivo* [14]. To further elucidate the molecular bases of this blocking activity and to gather information for structural optimization of these compounds, the interactions of DM1 and DM2 with the hCav_{3.1} channel were modelled.

5.2.1.1 Ligand Analysis

DM1 and DM1 were built in the anionic form (pKa values in Table 1) and subjected to conformational analysis by using a stochastic conformational search procedure.

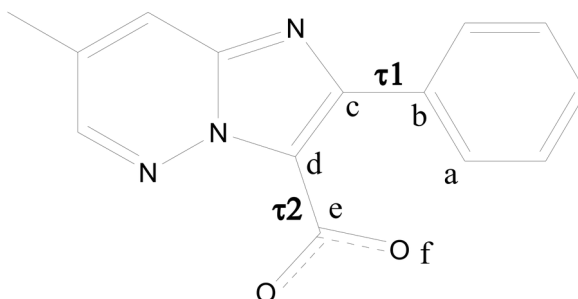
Table 1. Percentage of ionic forms and calculated pKa values of compounds DM1 and DM2.

Compd	Calculated pKa	Prevalent Ionic Form ^a (%)	
		pH 7.4	pH 7.2
DM1	pKa ₁ 4.7 ± 0.9	A (100)	A (100)
	pKa ₂ : 2.1 ± 1.0		
DM2	pKa ₁ 4.7 ± 0.9	A (100)	A (100)
	pKa ₂ : 2.1 ± 1.0		

^a A = anionic. ACD/pKa GALAS algorithm of ACD/Percepta software

The resulting conformations were then subjected to Molecular Mechanics (MM) calculations. Then they were ranked according to their potential energy values (difference from the global energy minimum, ΔE_{GM}) and grouped into conformational families (Table 2 and 3).

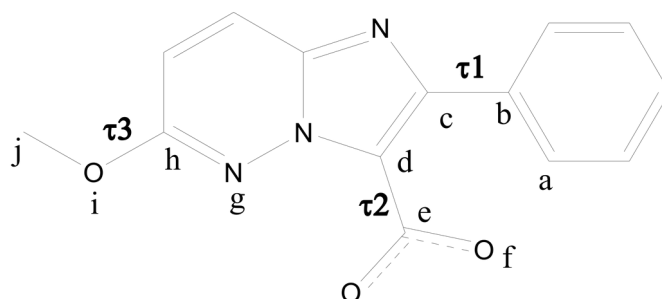
Table 2. ΔE_{GM} values (kcal/mol) and torsional angle values (degrees) of MM and DFT conformers of DM1.



Conformer	Starting conformer (MM)			DFT conformer		
	ΔE (kcal/mol)	$\tau 1^a$	$\tau 2^a$	ΔE (kcal/mol)	$\tau 1^a$	$\tau 2^a$
1A	0.00	-78	-1	0.00	-20	-87
1B	0.00	78	1	0.00	20	87

^a $\tau 1$: abcd; $\tau 2$: cdef.

Table 3. ΔE_{GM} values (kcal/mol) and torsional angle values (degrees) of MM and DFT conformers of DM2.



Conformer	Starting conformer (MM)				DFT conformer			
	ΔE (kcal/ mol)	$\tau 1^a$	$\tau 2^a$	$\tau 3^a$	ΔE (kcal/ mol)	$\tau 1^a$	$\tau 2^a$	$\tau 3^a$
1A				0		-2	-8	0
	0.00	-78	-1		0.00	1	2	
1B	0.00	78	1	0	0.00	21	82	0
2A				180		-2	-8	18
	1.76	-78	-1		3.18	0	8	0
2B				180				18
	1.76	78	1		3.18	20	88	0

^a $\tau 1$: abcd; $\tau 2$: cdef; $\tau 3$: ghij

Representative conformers of each family were selected and further optimized using DFT full minimization calculations with the conductor-like polarizable continuum model (C-PCM) to mimic an aqueous environment [22].

The DFT minimized structures were different from the starting MM conformers, in particular in the orientation of the phenyl ring ($\tau 1$) and the carboxylate function ($\tau 2$), highlighting that quantum-mechanical methods

are necessary in modelling highly conjugated structures (Figure 1; Tables 2 and 3).

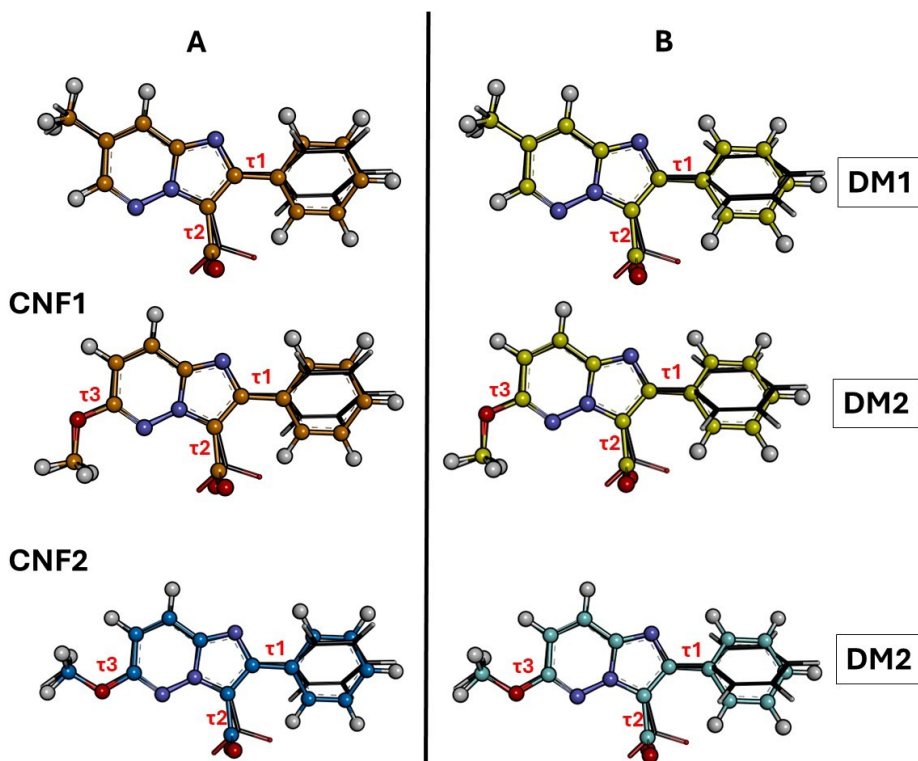


Figure 1. Structural comparison between the MM (stick and black) and DFT (ball&stick) conformers of DM1 and DM2. The conformational enantiomers A ($\tau_2 = -20^\circ$) are showed in A while the conformational enantiomers B ($\tau_2 = 20^\circ$) in B. Heteroatoms are colored by atom type: O, red; N, blue; H, white.

The conformers of DM1 and DM2 were limited in number, mostly for the formation of a non-conventional hydrogen bond between a carboxylate oxygen atom and the hydrogen atom at the *ortho* positions of the phenyl ring (Donor-H-Acceptor, $\alpha \sim 145^\circ$; H-Acceptor, $d \sim 2.6 \text{ \AA}$; Figure 2A and B) [23,24]

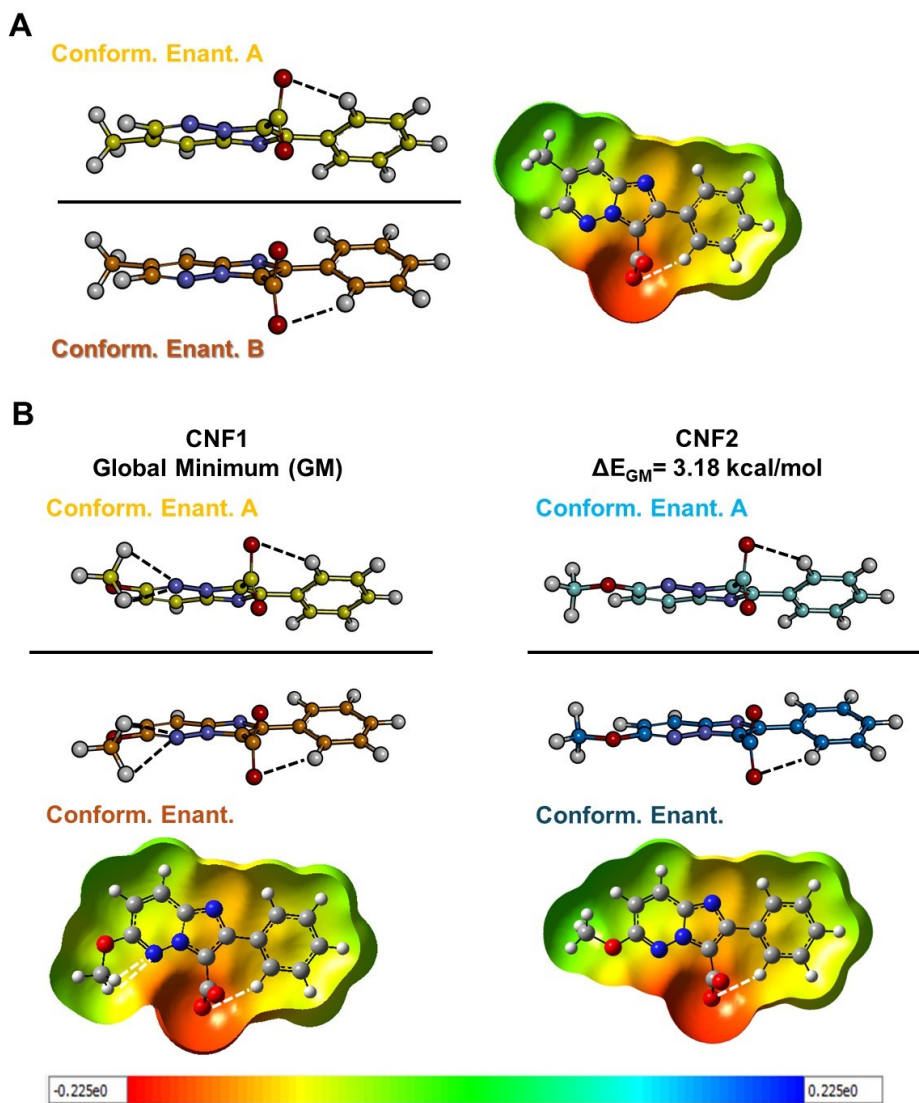


Figure 2. (A) DFT conformers of DM1: conformational enantiomer A ($\tau_2 = -20^\circ$) and B ($\tau_2 = 20^\circ$). Structures are displayed as ball and sticks; heteroatoms are coloured by atom type: O, red; N, blue; hydrogen bonds are highlighted with dashed lines. Right: molecular electrostatic potential surface (isovalue 0.0004 e-/a.u.³); the bond order is reported according to bond length; the frontmost portions of the surface has been removed to allow views into the interior of the molecular display (z-clip option). **(B)** DFT conformers of

DM2: two couple of conformational enantiomers characterized by the different rotation of the methoxyl substituent. Structures are displayed as ball and sticks; heteroatoms are coloured by atom type: O, red; N, blue; hydrogen bond interactions are highlighted with dashed lines. Bottom: molecular electrostatic potential surfaces (isovalue 0.0004 e-/a.u.³); the bond order is reported according to bond length; the frontmost portions of the surfaces has been removed to allow views into the interior of the molecular display (z-clip option).

Notably, DM1 showed two specular and isoenergetic conformers, i.e. conformational enantiomers, that are characterized by i) the opposite rotation of the phenyl ring about τ_1 ($\pm 20^\circ$) and ii) same orientation of the carboxylate group ($\tau_2 \sim 90^\circ$) (Figure 2A and Table 2). On the other hand, DM2 showed similar values for τ_1 (phenyl ring) and τ_2 (carboxylate group) (Figure 2B and Table 3). In addition, the rotation of the methoxy substituent (τ_3) in DM2 resulted in two pairs of conformational enantiomers. One pair includes the minimum energy conformational enantiomer ($\tau_3 \sim 0^\circ$) favoured by 3.18 kcal/mol over the other pair ($\tau_3 \sim 180^\circ$). The attraction of the hydrogen atoms of the methoxy group and the pyridazine nitrogen atom N5, that forms a non-conventional hydrogen bond interaction, is probably the reason there is this strong conformational preference.

Conformational enantiomers are present in achiral molecules (i.e., molecules without a permanent stereocenter) that can adopt chiral conformations (non-superimposable mirror images) due to restricted rotation or specific spatial arrangements. They are most commonly observed in atropisomeric compounds, where rotation around a single bond is hindered (e.g., biaryl systems) and in cyclic systems with rigid structures that prevent free rotation [25,26]. Although we did not calculate the rotation barrier of DM1 and DM2, the presence of a biaryl system with an ortho substituent strongly suggest that they are atropisomeric

compounds. Accordingly, the docking simulations were set up to treat each conformer separately, allowing the evaluation of potential stereospecific binding preferences.

5.2.1.2 Definition of The Protein Search Area

The definition of the region of hCa_v3.1 channels to be explored during docking studies was guided by a structural and bioinformatic analysis on the structures reported in Table 4.

Table 4. Experimentally determined structures of mammalian Ca_v channels used in the structural and bioinformatics analysis.

PDB		
ID	Organism	Structure
6KZO	human	Cryo-EM structure of apo human Ca _v 3.1
6KZP	human	Cryo-EM structure of human Ca _v 3.1-Z944 Complex
	rabbit	Cryo-EM structure of Rabbit Ca _v 1.1-Verapamil
6JPA		Complex
	rabbit	Cryo-EM structure of Rabbit Ca _v 1.1-Diltiazem
6JPB		Complex

hCa_v channels consist of a single peptide chain that folds into four homologous domains (repeat I–IV), each containing six transmembrane helices (S1_{I–IV}–S6_{I–IV}). The central ion conducting pore domain is enclosed by the S5_{I–IV} and S6_{I–IV} helices, the selectivity filter (SF_{I–IV}) and the pore helices P1_{I–IV} and P2_{I–IV}. The shifting of the S5 and S6 helices of each domain opens the channel.

Dihydropyridines, verapamil and diltiazem are known Cav_{1.1} (type L) blockers, along with the Cav_{3.1} (type T) blocker Z944, bind to the central

cavity of the channel at S5 and S6 helices level, regardless of the channel domain (repeat I-IV) to which they bind (Figure 3).

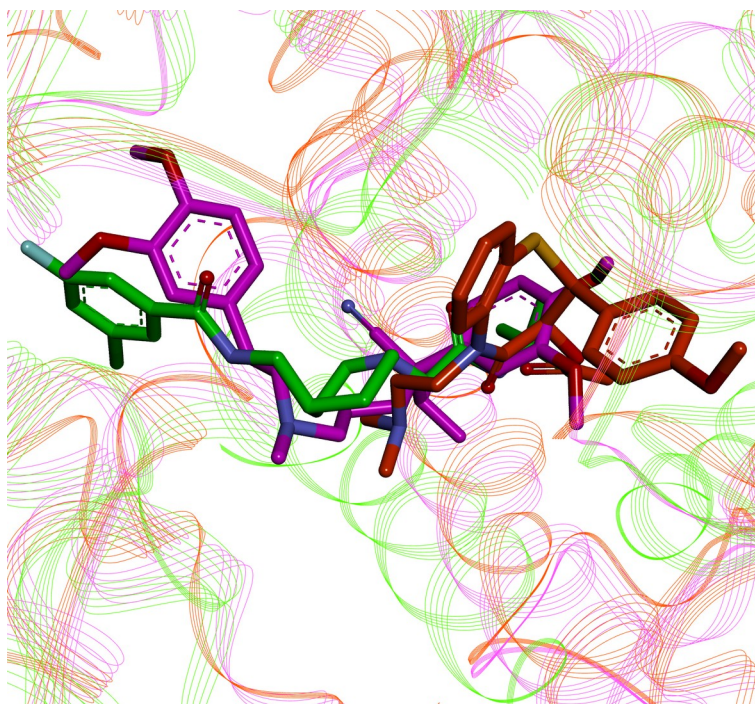


Figure 3. Superimposition of Cryo-EM structure of human $\text{Ca}_v3.1$ -Z944 Complex (green ribbon and green stick), Cryo-EM structure of Rabbit $\text{Ca}_v1.1$ -Verapamil Complex (purple ribbon and purple stick) and Cryo-EM structure of Rabbit $\text{Ca}_v1.1$ -Diltiazem Complex (red ribbon and red stick).

The calculated ligand-protein interactions of the complexes under investigation are reported in Figures 4-7.

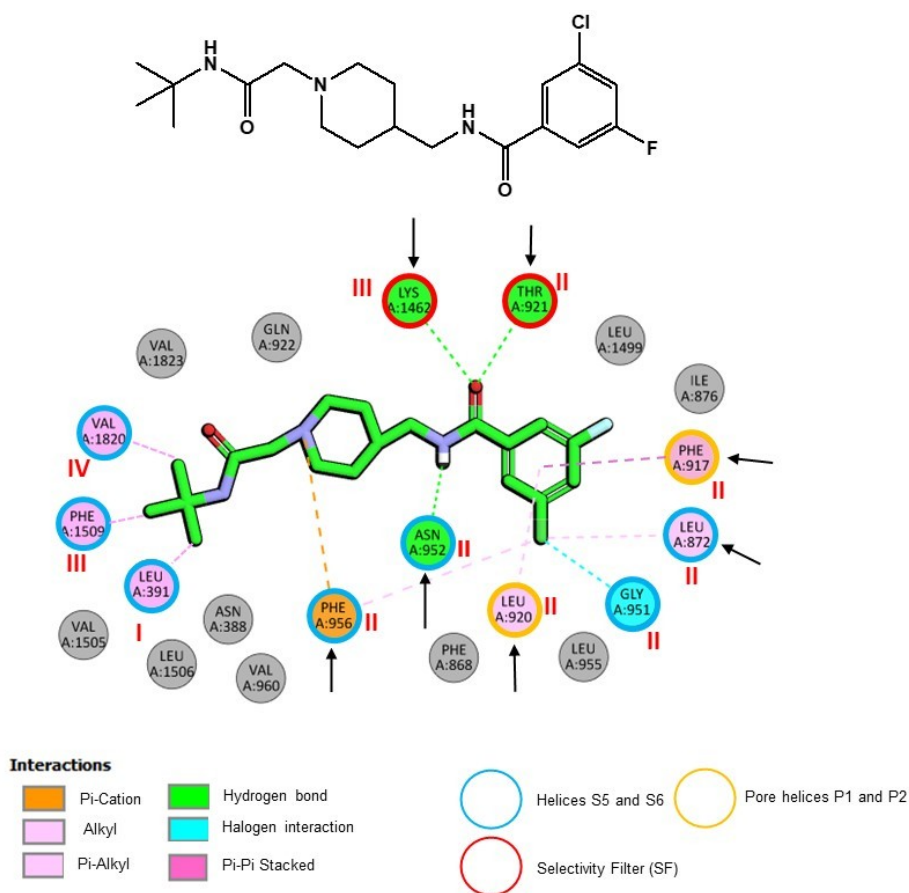


Figure 4. Schematic representation of the interactions between Z944 and hCa_v3.1 (PDB ID: 6KZP). All residues within 5 Å from any ligand atom are displayed. The inhibitor is colored by atomic type (C= green, O = red, N = blue, F = cyan, Cl = light green, H = white). The residues involved in the interaction with DM1 and DM2 are indicated with a black arrow.

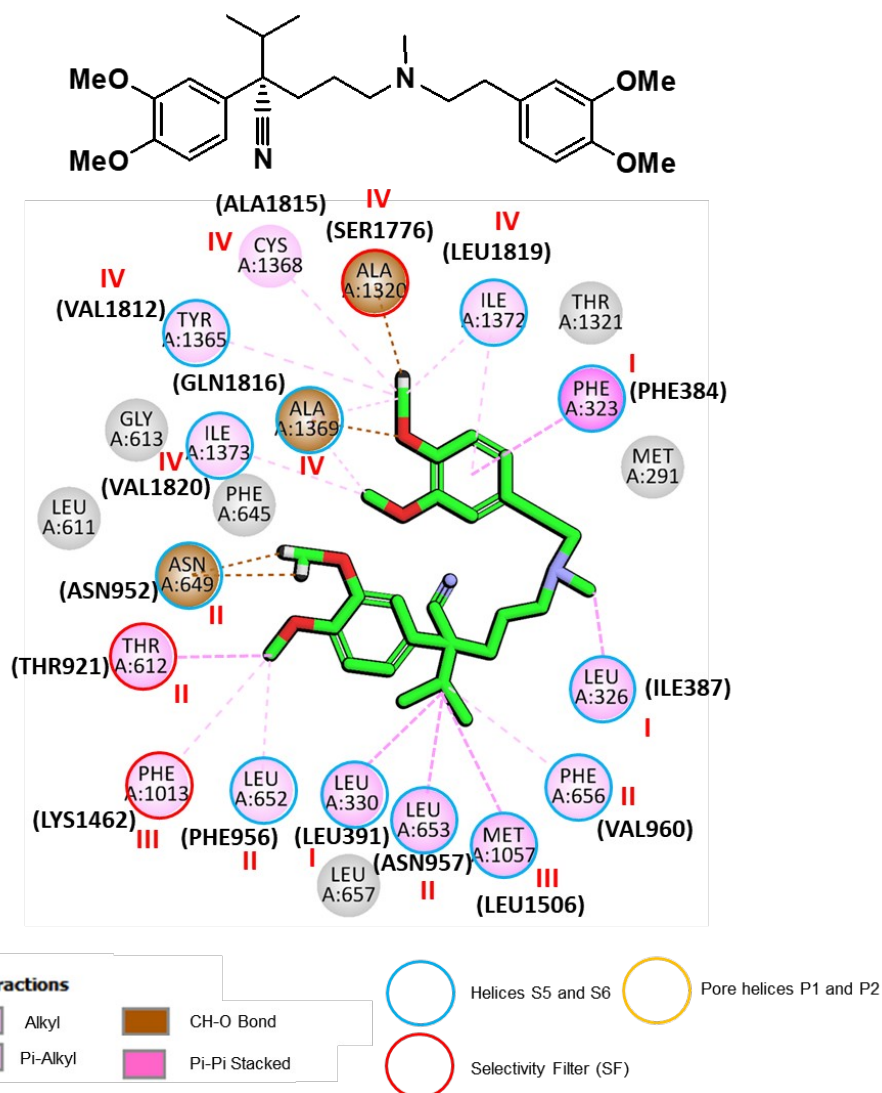


Figure 6. Schematic representation of the interactions between Verapamil and rabbit Ca_v1.1 (Binding mode II; PDB ID: 6JPA). All residues within 5 Å from any ligand atom are displayed. In brackets the corresponding residue of human Ca_v3.1. The inhibitor is colored by atomic type (C= green, O = red, N = blue, H = white).

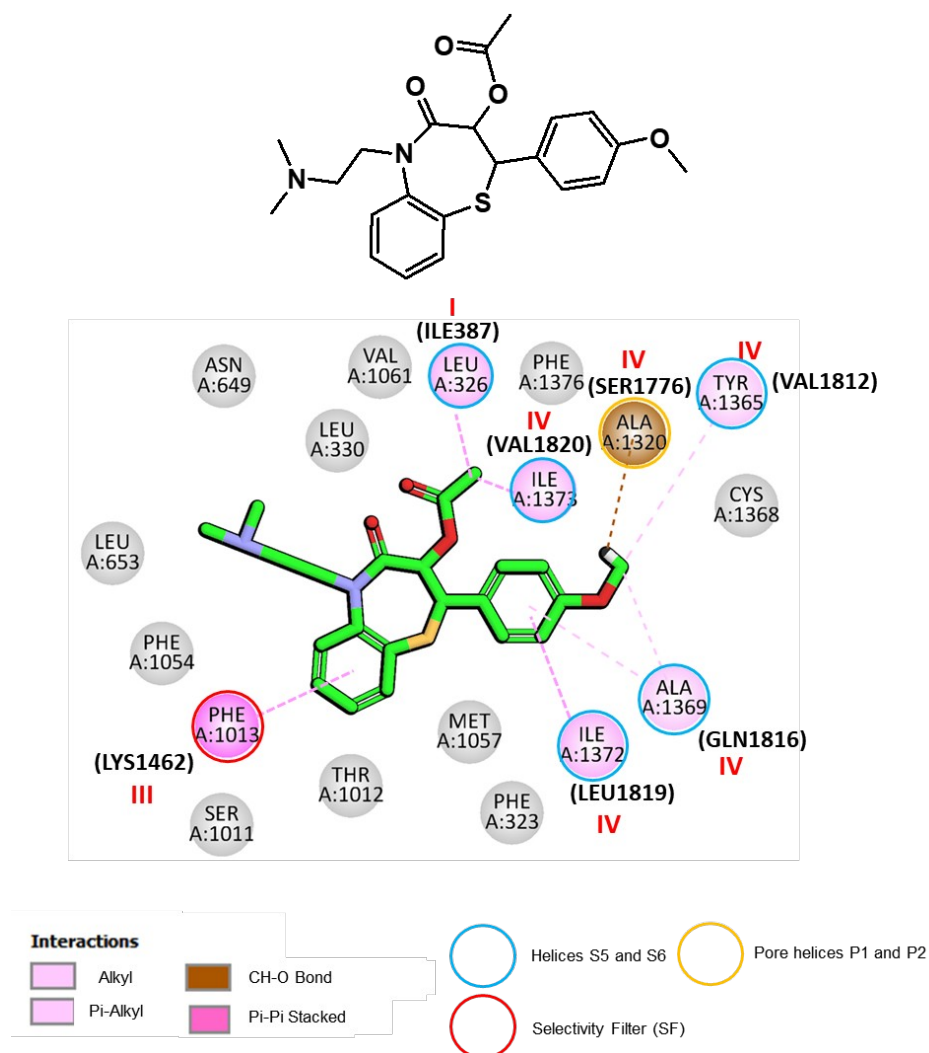


Figure 7. Schematic representation of the interactions between Diltiazem and rabbit $\text{Ca}_v1.1$ (PDB ID: 6JPB). All residues within 5 Å from any ligand atom are displayed. In brackets the corresponding residue of human $\text{Ca}_v3.1$. The inhibitor is colored by atomic type (C= green, O = red, N = blue, S = yellow, H = white).

Based on this analysis, two different search domains area were defined on the structure of $\text{hCa}_v3.1$ calcium channel in inactivated state (i.e. closed intracellular gate and depolarized “up” voltage sensing domains (VSDs)) [27]: i) a binding site sphere of 20 Å around the mass centre of the ligands

(named “BS1”) or ii) a binding site sphere created from the cavity present in the centre of the channel structure (Radius 21.7 Å; named “BS2”) (Figure 8). The experimentally determined structure of hCav_{3.1}-Δ8b in complex with the selective blocker Z944 (PDB ID: 6KZP) was used as starting protein conformation and prepared for docking simulation.

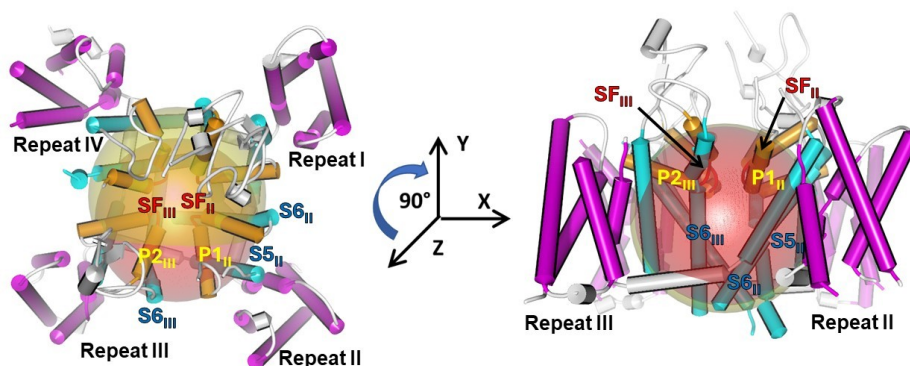


Figure 8. The cryo-electron microscopy structure of the human Ca_v3.1 channel used as protein starting structure in docking studies (PDB ID: 6KZP). The two search areas defined for ligand docking are represented as transparent spheres (yellow:BS1; red:BS2). The S1–S4 helices of each repeat forming the peripheral (VSD) are coloured in magenta; the S1_{I-IV}–S6_{I-IV} helices are coloured in cyan, the pore helices P1_{I-IV}–P2_{I-IV} in orange, and the loops constituting the selectivity filter SF_{I-IV} in red. The four repeats and the substructures of repeats II and III are labelled in yellow.

5.3.1.3 Docking Simulations

Different starting ligand-protein complexes were generated by using the DFT conformers of DM1 (two conformers) and DM2 (four conformers), ensuring in such way that all possible ligand conformers were considered while preserving the DFT-calculated structure during the docking (force-field based) simulations. Torsion angles τ_1 and τ_2 were restrained in order to tether the conformation of these ligands during the docking simulations [27,28]. This enables the assessment of a potential binding selectivity

between conformational enantiomers. Considering the two search domain area, a total of twelve docking run was performed.

The resulting poses (Figure 9) were then minimized with unconstrained energy (i.e. the ligand were left free to reorient in the active site and residues included in the binding domain area were left free to move) and grouped into families based on structure similarity.

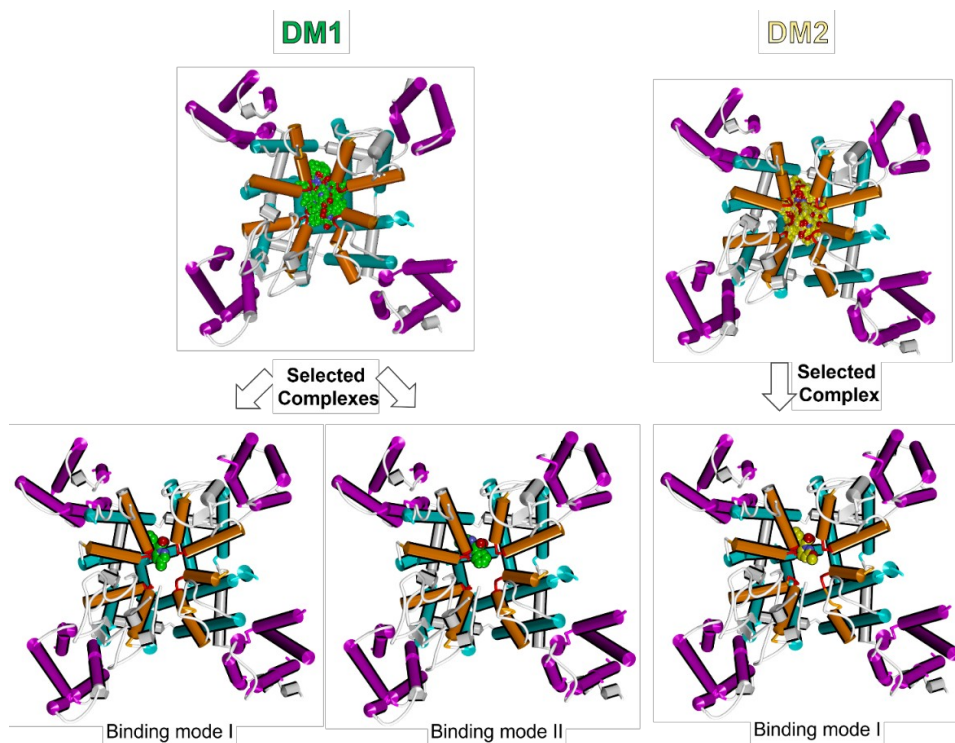


Figure 9. DM1 and DM2 binding poses resulting from docking studies (up) and selected best docked complexes (down). The ligand is shown as CPK and coloured by atom type (carbon =green); $\text{Ca}_{v3.1}$ secondary structure is coloured as follows: S1-I-IV -S4-I-IV helices are coloured in magenta, S5-I-IV-S6-I-IV helices are coloured in cyan, P1-P2-I-IV helices are coloured in orange, and the selectivity filter loops SF-I-IV in red.

The complexes with the most favourable ligand/protein interaction energy were selected as the best-docked complexes and their structural quality was assessed using the structure evaluator software Procheck (Table 5).

Table 5. Summary of Procheck results obtained for the selected docked complexes of DM1/hCa_v3.1 and DM2/hCa_v3.1.

Complex	Pos e	Residu es favored regions (%)	Residu es allowed regions (%)	Residues generous ly allowed regions (%)	Residues disallow ed regions (%)	Poor rotame r (%)
DM1/ hCa _v 3.1	119	91.6	7.8	0.5	0.1	2.1
DM1/ hCa _v 3.1	101	91.7	7.7	0.5	0.1	2.1
DM2/ hCa _v 3.1	32	91.9	7.4	0.5	0.1	2.1

Two complexes with the same binding energy ($\Delta E = 0.03$ kcal/mol) were selected for DM1 and one complex was selected for DM2 (Table 6; Figures 9 and 10).

Table 6. Ligand-protein nonbonded interaction energies (kcal/mol) of the best docked complexes.

Complex	Ligand Conf.	Binding mode	Ligand-Protein interaction energies (kcal/mol)
hCa _v 3.1-DM1	B	I	-45.89
hCa _v 3.1-DM1	B	II	-45.86
hCa _v 3.1-DM2	2B	I	-47.54

The contribution of each protein residue to the binding energy was evaluated (Appendix F: Table 2; Tables 7,8), and the number and types of ligand-protein interactions were calculated (Figures 10 and 11) and

compared with those present in experimentally determined structures of VGCC in complex to know ligands (Figures 4-7).

Table 7. Ligand-residue nonbonded interaction energies (kcal/mol) of hCav3.1/DM1 best docked complex (Binding Mode II; residues within 5 Å from any ligand atom).

Residue	Interaction Energy (kcal/mol)	VDW Interaction Energy (kcal/mol)	Electrostatic Interaction Energy (kcal/mol)
F868	-0.184636	-0.139097	-0.045539
L872	-1.484386	-1.510254	0.025868
F875	-0.368664	-0.343244	-0.025420
I876	-0.353827	-0.367214	0.013387
F917	-1.781752	-1.893987	0.112236
L920	-4.814382	-4.817216	0.002834
T921	-6.598563	-4.794982	-1.803580
Q922	-2.149928	-2.067512	-0.082417
G951	-1.450503	-1.429146	-0.021357
N952	-2.896945	-2.601230	-0.295715
L955	-0.377873	-0.423971	0.046097
F956	-5.460884	-5.497628	0.036744
S1461	-0.056905	-0.155670	0.098765
K1462	-13.746023	-2.802122	-10.943901
L1498	-0.521261	-0.718333	0.197072

L1499	-0.472323	-0.592656	0.120333
A1502	-0.970089	-0.678588	-0.291501

Table 8. Ligand-residue nonbonded interaction energies (kcal/mol) of hCav3.1/DM2 best docked complex (residues within 5 Å from any ligand atom).

Residue	Interaction	VDW	Electrostatic
	Energy (kcal/mol)	Interaction Energy (kcal/mol)	Interaction Energy (kcal/mol)
L872	-1.243542	-1.259184	0.015642
F875	-0.192107	-0.150239	-0.041869
I876	-0.676733	-0.666056	-0.010677
F917	-1.659222	-1.993220	0.333998
L920	-4.397705	-4.265539	-0.132166
T921	-7.553552	-5.700613	-1.852939
Q922	-3.311208	-2.974990	-0.336219
E923	-0.015719	-0.200802	0.185084
G951	-0.651939	-0.578987	-0.072952
N952	-2.368714	-2.151309	-0.217404
L955	-0.157085	-0.196318	0.039234
F956	-3.760215	-3.808562	0.048347
S1461	-0.252648	-0.293554	0.040906
K1462	-15.869071	-3.736326	-12.132745

L1498	-0.668147	-0.842299	0.174152
L1499	-1.672995	-1.691714	0.018720
A1502	-0.998697	-0.633829	-0.364867

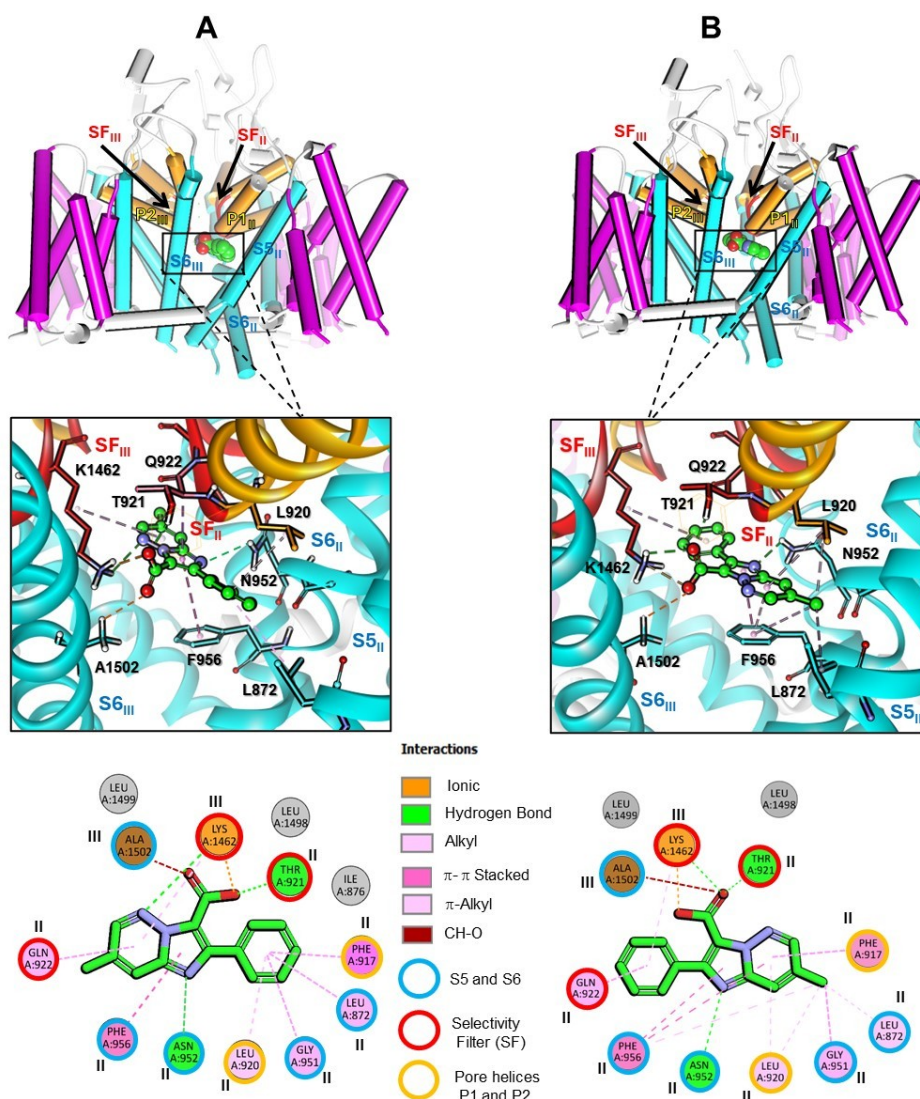


Figure 10. Ca_v3.1-DM1 best-docked complexes: Binding mode I (A) and Binding mode II (B). Up: overall lateral view of the complex; middle: ligand binding site; bottom: 2D

representation of Ligand-protein interactions. Protein secondary structures are displayed as cylinder (up) or ribbons (bottom), ligand is displayed as CPK (up) or balls&sticks (middle) and coloured in green. hCa_v3.1 substructures are labelled in yellow (3D-structures) or black (2D-structures). Protein residues showing an interaction energy with the ligand > 1 kcal/mol are displayed and labelled in black except for F917 (displayed as lines for clarity of presentation). Ca_v3.1 substructures are coloured as follows: S1_{I-IV} -S4_{I-IV} helices are coloured in magenta, S5_{I-IV}-S6_{I-IV} helices are coloured in cyan, P1-P2_{I-IV} helices are coloured in orange, and the selectivity filter loops SF_{I-IV} in red. Hydrogens are omitted for clarity except those involved in hydrogen bonds.

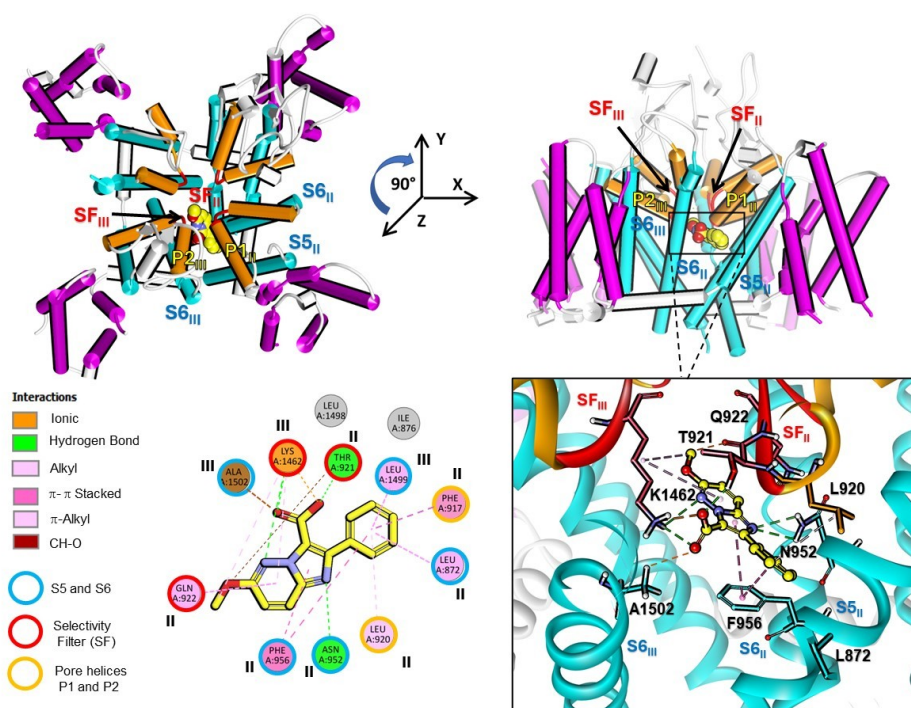


Figure 11. Ca_v3.1-DM2 best-docked complex. Up: overall view of the complex; bottom: 2D representation of ligand-protein interactions and ligand binding site. Protein secondary structures are displayed as cylinder (up) or ribbons (bottom), ligand is displayed as CPK (up) or balls&sticks (bottom) and coloured in yellow. hCa_v3.1 substructures are labelled. Protein residues showing an interaction energy with the ligand > 1 kcal/mol are displayed and labelled in black except for F917 (for clarity of presentation). Ca_v3.1 substructures are coloured as follows: S1_{I-IV} -S4_{I-IV} helices are coloured in magenta, S5_{I-IV}-S6_{I-IV} helices are coloured in cyan, P1-P2_{I-IV} helices are

coloured in orange, and the selectivity filter loops SF_{I-IV} in red. Hydrogens are omitted for clarity except those involved in hydrogen bonds.

In the selected best-docked complexes, DM1 and DM2 showed the same conformation of τ_1 and τ_2 (i.e., conformational enantiomers B; Figure 2) and were positioned into the fenestration between repeats II and III. The convergence of the best solutions on form B suggests a conformational enantiomer preference at the II–III fenestration site. Notably, it has been reported that binding to this site may highlight the state dependence for T-type inhibition [27].

In particular, the binding site for DM1 and DM2 is enclosed by the interfaces of S5_{II}, S6_{II} and S6_{III} helices and the central cavity of the channel (i.e., selectivity pore: P1_{II}, SF_{II} and SF_{III}) (Figures 10 and 11).

A salt bridge is formed between the carboxylate group and K1462 (numbering according to PDB entry 6KZP; uniport sequence: O43497-9) on the selectivity loop SF_{III} and a hydrogen bond is established between the same carboxylate and T921 on the selectivity filter loop SF_{II}. At the same time, the imidazole ring interacted with the S6_{II} helix trough a π - π contact with F956 and a hydrogen bond with the side chain of N952.

These interactions were present in the best-docked solutions for both DM1 and DM2, with the highest contribution to ligand-protein binding energy (~14 kcal/mol; (Appendix F: Table 1 and 2)). The selectivity filter presents a unique to Cav₃ (T-type) VGCC channel residue (K1462), replaced by a Phe in Cav₁ and Gly in Cav₂ channels, respectively. F956 is another unique residue in Cav₃ (leucine in Cav₁ and Cav₂ channels) involved in π - π interactions with DM1 and DM2 and located on S6_{II} helix. These interactions suggest that DM1 and DM2 possesses a potential selectivity

for hCav_{3.1}, hypothesis supported also by their *in vivo* anti-absence activity [14].

As can be seen from Table 5 and Figure 9, DM1 could adopt two equally favoured binding modes by “switching” the phenyl and pyridazine ring.

In the first binding mode (Mode I; Figure 10A), the ligand positioned its phenyl ring within the cavity formed by the S5_{II} (L872), S6_{II} (F956), and P1_{II} (F917, L920) helices, whereas the pyridazine ring was oriented toward the channel centre. In this configuration, the N1 nitrogen of the pyridazine established an additional hydrogen bond with the side chain of K1462 (SF_{III}).

In the second binding mode (Mode II; Figure 10B), the 7-methylpyridazine ring was instead surrounded by the S5_{II} (L872), S6_{II} (F956), and P1_{II} (F917, L920) helices, while the phenyl ring extended toward the selectivity pore, forming π -alkyl interactions with the aliphatic chain of K1462.

Conversely, the best-docked conformation of DM2 adopted binding mode I, engaging the same key residues involved in the interaction with DM1. Additionally, the phenyl ring formed a hydrophobic contact with L1499 on helix S6_{III}, while the methyl group of the methoxy substituent established a non-conventional (carbon) hydrogen bond with T921 and alkyl interactions with the aliphatic chain of K1462 (Figure 10; (Appendix F: Table 1 and 2)). Notably, the oxygen atom of the methoxy group did not directly interact with the protein as it lacked the intra-molecular hydrogen bond between the methoxy substituent and the pyridazine nitrogen atom N5 (Figure 2B). Instead, this nitrogen is involved in interaction with K1462 in the bioactive conformation (Figure 11).

These findings explain the comparable in vitro efficacy of DM1 and DM2 in inhibiting hCav_{3.1} channels and are consistent with the slightly higher potency observed for DM1 relative to DM2 [14]. Notably, the best-docked conformations of both ligands reproduced most of the key interactions characteristic of the Cav_{3.1}-selective blocker Z944, including contacts with the unique hCav₃ residues F956 and K1462. Figure 12 report the structural comparison of our docking results with the cryo-EM structure of hCav_{3.1}-Z944 complex.

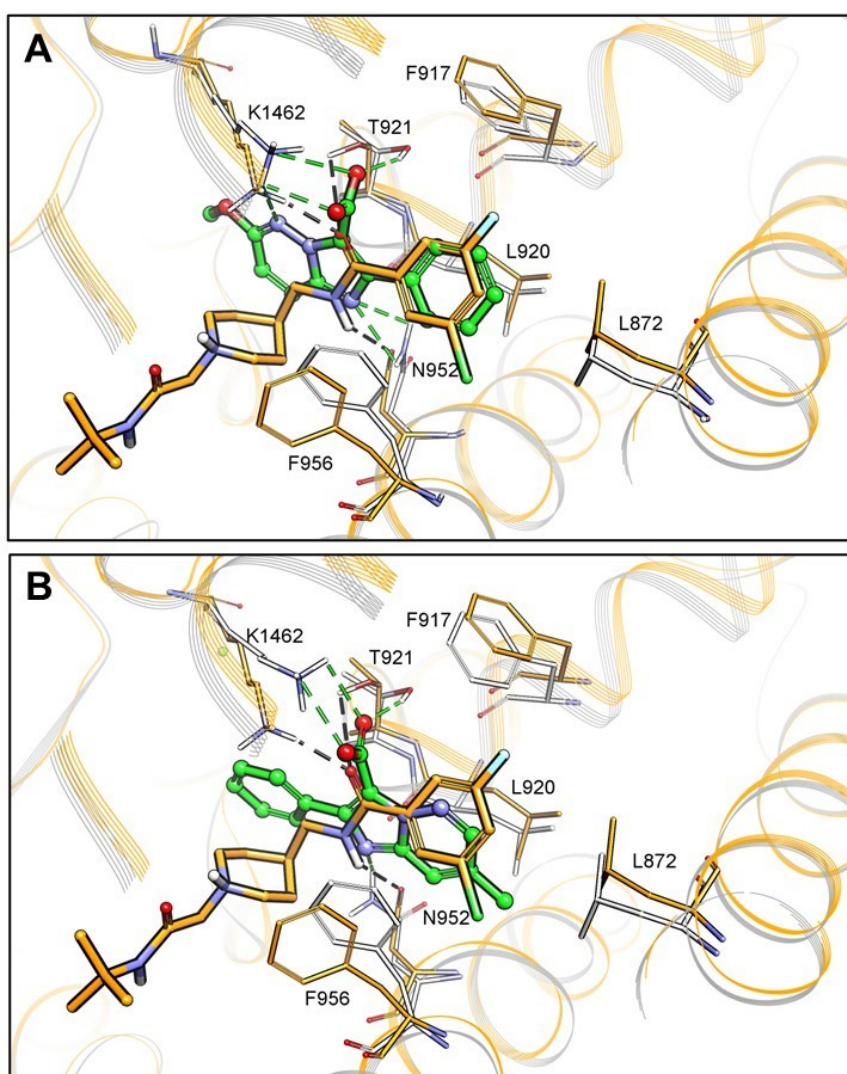


Figure 12. A: structural overlap ($\text{C}\alpha$ pairs with a $\text{RMSD} < 1 \text{ \AA}$) of $\text{Ca}_{\text{v}3.1}$ -DM2 best-docked complex (grey) on the cryo-EM structure of human $\text{Ca}_{\text{v}3.1}$ -Z944 complex (orange; PDB ID: 6KZP). B: structural overlap ($\text{C}\alpha$ pairs with a $\text{RMSD} < 1 \text{ \AA}$) of $\text{Ca}_{\text{v}3.1}$ -DM1 best-docked complex (binding mode II; grey) on the cryo-EM structure of human $\text{Ca}_{\text{v}3.1}$ -Z944 complex (orange; PDB ID: 6KZP). The protein backbone is displayed as line ribbon. DM1 and DM2 are displayed in ball&stick. The common interacting residues for DM1, DM2 and Z944 are displayed in stick and labelled. The ligands and the interacting residues are colored by atomic type (O=red, N=blue, Cl=light green, H=white). Hydrogen bonds are highlighted by green (docked complexes) and black (cryo-EM) dashed lines.

5.2.2 Investigation of the Antiepileptic Potential of DM1 and DM2

For more information about the Antiepileptic Potential of DM1 and DM2 the reader can refer to ref. [21]

5.3 Conclusions

Both DM1 and DM2 showed anti-seizure activity in *in-vivo* animal models and demonstrated neuroprotective effects in *in-vitro* studies. Computational studies indicate a putative binding site and binding mode for DM1 and DM2 within the hCav3.1 calcium channel in its depolarized conformation. Overall, conformational enantiomerism and the preferential binding of one conformer at Cav3.1 emerge as key structural drivers, pointing to optimization strategies that raise the rotational barrier/atropisomerism to “freeze” the bioactive form. The restricted conformational flexibility of the compounds, the existence of conformational enantiomers, the preferential affinity of one enantiomer toward the target, and the structural comparison between the docked hCav3.1/DM1 and hCav3.1/DM2 complexes and experimentally resolved VGCC-blocker complexes collectively provide a robust basis for the rational design and further optimization of this promising class of molecules.

5.5. References.

1. Akkurt, B. "Recent Advances on the Synthesis of Imidazo[1,2-b]pyridazines." *Journal of the Turkish Chemical Society, Section A*, vol. 8, 2021, pp. 1217-50.
2. Zeng, F., et al. "Discovery of Novel Potent and Selective Imidazo[1,2-b]pyridazine Antagonists of the Corticotropin-Releasing Factor-1 (CRF1) Receptor." *ACS Medicinal Chemistry Letters*, vol. 1, no. 3, 2010, pp. 80-84.
3. Gehlert, D. R., et al. "A Novel CRF1 Receptor Antagonist, NBI 77860, Blocks Behavioral and Endocrine Effects of Emotional Stress." *The Journal of Neuroscience*, vol. 27, no. 10, 2007, pp. 2718-26.
4. Krall, J., et al. "Design, Synthesis, and Pharmacological Characterization of Imidazo[1,2-b]pyridazines as Novel Positive Allosteric Modulators of the $\alpha 1$ -Subunit-Containing GABAA Receptors." *Journal of Medicinal Chemistry*, vol. 62, no. 6, 2019, pp. 2798-2813.
5. Sabnis, R. W. "Tirzepatide: A New Era of Dual GIP/GLP-1 Receptor Agonists for Diabetes and Obesity." *ACS Medicinal Chemistry Letters*, vol. 12, no. 10, 2021, pp. 1526-27.
6. Anderson, M., et al. "The discovery of a novel series of potent and selective Aurora kinase inhibitors." *Bioorganic & Medicinal Chemistry Letters*, vol. 13, no. 18, 2003, pp. 3021-26.

7. Byth, K. F., et al. "Imidazo[1,2-b]pyridazines: a potent and selective class of STAT3 inhibitors." *Bioorganic & Medicinal Chemistry Letters*, vol. 14, no. 9, 2004, pp. 2245-48.
8. Byth, K. F., et al. "Inhibitors of STAT3: the first selective, small molecule inhibitors of STAT3 protein-protein interaction." *Bioorganic & Medicinal Chemistry Letters*, vol. 14, no. 9, 2004, pp. 2249-52.
9. Garrido, A., et al. "Imidazo[1,2-b]pyridazine: A privileged scaffold in medicinal chemistry." *European Journal of Medicinal Chemistry*, vol. 226, 2021, article 113867.
10. Abignente, E., et al. "Derivatives of imidazo[1,2-b]pyridazine with antiinflammatory and analgesic activities." *Research Communications in Chemical Pathology and Pharmacology*, vol. 67, no. 1, 1990, pp. 43-54.
11. Abignente, E., et al. "Synthesis and pharmacological screening of new imidazo[1,2-b]pyridazine derivatives." *Il Farmaco*, vol. 45, no. 10, 1990, pp. 1075-87.
12. Luraschi, E., et al. "Imidazo[1,2-b]pyridazine derivatives. VI--Synthesis and pharmacological activity of new 2-aryl-6-chloro-7-methyl-derivatives." *Il Farmaco*, vol. 52, no. 4, 1997, pp. 213-17.
13. Frecentese, F., et al. "New Imidazo[1,2-b]pyridazine Derivatives with Anticonvulsant Activity: Synthesis, In Vitro and In Vivo Evaluation." *International Journal of Molecular Sciences*, vol. 24, no. 13, 2023, article 10722.
14. Rimoli, M. G., et al. "Anticonvulsant activity of a new imidazo[1,2-b]pyridazine derivative in rodent models of epilepsy." *Neuropharmacology*, vol. 56, no. 3, 2009, pp. 637-46.
15. Palazzo, E., et al. "The imidazo[1,2-b]pyridazine derivative, Vg-3, reduces neuropathic pain by inhibiting spinal cord glial activation and PARP-1 activity." *Neuropharmacology*, vol. 58, no. 3, 2010, pp. 660-67.
16. Frantseva, M. V., et al. "Specific inhibition of mitochondrial permeability transition pore opening by a novel imidazo[1,2-b]pyridazine derivative." *European Journal of Neuroscience*, vol. 12, no. 4, 2000, pp. 1431-39.
17. Malinska, D., et al. "The imidazo[1,2-b]pyridazine derivative, CGP37157, is a specific inhibitor of the mitochondrial Na⁺/Ca²⁺ exchanger." *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1797, no. 6-7, 2010, pp. 1163-70.
18. Pestana, R. R. F., et al. "Inhibition of mitochondrial Na⁺/Ca²⁺ exchanger by CGP-37157 protects against neuronal damage induced by kainic acid." *Neuroscience Letters*, vol. 484, no. 3, 2010, pp. 187-91.
19. Waldbaum, S., and M. Patel. "Mitochondrial dysfunction and oxidative stress: a contributing link to epilepsy." *Journal of Bioenergetics and Biomembranes*, vol. 42, no. 6, 2010, pp. 449-55.
20. Kovac, S., et al. "Mitochondrial permeability transition in epilepsy: modulation by ketogenic diet." *Cell Death & Disease*, vol. 4, 2013, article e693.

21. Schiano, Marica Erminia, et al. "Eco-friendly Synthesis and Molecular Modelling of 2-Phenylimidazo [1, 2-b] pyridazine Derivatives: In Vitro and In Vivo Studies for Lead Optimization." *ChemMedChem*, vol. 20, no. 5, 2025, article e202400721.
22. Cossi, M., et al. "Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model." *Journal of Computational Chemistry*, vol. 24, no. 6, 2003, pp. 669-81.
23. Imran, M., et al. "Crystal structure of 2-phenyl-6-(4-chlorophenyl)imidazo[1,2-b]pyridazine." *Zeitschrift für Kristallographie - Crystalline Materials*, vol. 225, no. 1, 2010, pp. 36-40.
24. Desiraju, G. R., and T. Steiner. *The Weak Hydrogen Bond: In Structural Chemistry and Biology*. Oxford University Press, 1999.
25. Deng, Chenghua, et al. "Homochiral BINOL-Based Metal–Organic Frameworks for Luminescence Sensing of Hydrobenzoin Enantiomers." *Inorganic Chemistry* 64.5 (2025): 2583-2589.
26. Khan, Ruhima, Anjana Ningombam, and Moirangthem Dhaneshwar Singh. "Atropisomerism about the C (1)-N single bond in N, N-disubstituted-1-aminoanthracenes: isolation of conformational diastereomers." *Journal of Chemical Sciences* 130.6 (2018): 61.De Sarro, G., et al. "Epilepsy and Antiepileptic Drugs: A Review." *Epilepsy & Behavior*, vol. 71, Part B, 2017, pp. 165-73.
27. Zhao, Y., et al. "Structural basis for the channel gating mechanism of the human voltage-gated sodium channel Nav1.7." *Cell*, vol. 177, no. 6, 2019, pp. 1495-1506.e13.
28. Zhao, Y., et al. "Molecular basis for the binding of lidocaine to human Nav1.7." *Nature*, vol. 576, 2019, pp. 492-97.

Chapter 6. 1,2,4-Thiadiazolidin-3,5-Diones as Inhibitors of Cysteine Proteases

6.1 Introduction

In the context of viral infections, such as SARS-CoV-2, cysteine proteases have emerged as critical therapeutic targets. Since these enzymes play an essential role in immune evasion and viral replication, they have been of interest in the development of selective covalent inhibitors.

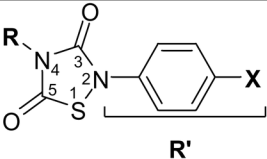
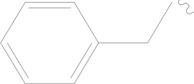
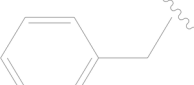
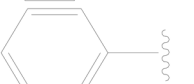
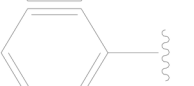
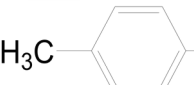
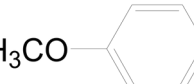
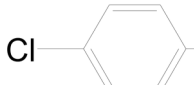
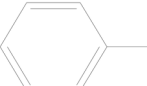
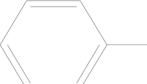
On the other hand, 1,2,4-Thiadiazolidin-3,5-diones (THIAs) are a versatile class of heterocycles with diverse pharmacological applications. They have been describing as inhibitors of glycogen synthase kinase 3 β (GSK3 β) [1], regulators of G protein signaling (RGS) proteins [2], and hydrogen sulfide (H₂S) donors [3]. The reaction of the sulfur atom of THIAs with the sulfhydryl groups (SH) contained in compounds that release H₂S [3] or with the catalytic cysteine residue in enzymatic reactions [2] are the basis of the mechanism of the activities of these compounds. Given this reactivity, inhibition of cysteine proteases is an expected activity for THIA derivatives. Nonetheless, 4-(4-fluorobenzyl)-2-p-tolyl-1,2,4-thiadiazolidine-3,5-dione (CCG-50014) did not inhibit Papain [4]. This observation indicates that covalent inhibition depends not only on the reaction between the THIA ring and the catalytic cysteine, but also on the proper accommodation of the compound within the enzyme's active site, as originally described for Papain by Schechter and Berger [5] and

investigated in other cysteine proteases using substrate libraries and systematic series of synthetic peptide analogues [6-10].

Based on these considerations, THIA derivatives were selected as potential inhibitors of cysteine proteases. In particular, this chapter examines the inhibitory activity of THIA derivatives and elucidates the molecular determinants underlying their selectivity through an integrated experimental and computational approach, keeping in mind that the presence of conformational enantiomers, are likely to play a role in their ability to fit into the cysteine protease active site. Stereochemistry plays a key role in the mechanism of action of a drug since it can affect the binding to the molecular target [11].

Going into details, a selected library of ten THIA derivatives (**THIA-1-10**) (Table 1) was chosen based on the previous considerations, to evaluate their inhibitory activity against cysteine proteases with distinct substrate specificities.

Table 1. The 1,2,4-thiadiazolidine-3,5-diones: THIA 1-10.

		
Compounds	R	X
THIA-1		-H
THIA-2		-OCH ₃
THIA-3		-H
THIA-4		-OCH ₃
THIA-5		-OCH ₃
THIA-6		-OCH ₃
THIA-7		-OCH ₃
THIA-8		-OCH ₃
THIA-9		-CH ₃
THIA-10		-F

Apart from the reference proteases Papain and Cathepsin L, two proteases of SARS-CoV-2 were selected: the papain-like protease (PLpro), which exhibits deubiquitinating and deISGylating activities and 3CLpro, essential for viral polyprotein processing. The aim of this study is double: i) to evaluate the potential of **THIA-1-10** compounds as selective covalent inhibitors of cysteine proteases, and ii) to elucidate the molecular basis underlying their activity through computational analyses and thus enhance future structure-based optimization efforts.

Papain and Cathepsin L show similar crystal structures [12,13], with well-defined S1 to S1' subsites that can contribute to binding specificity and that can accommodate substrate residues [14-16].

3CLpro is active only in the dimeric form [17]. Each homodimer is organized into three domains: domains I and II are composed of antiparallel β -barrels forming a characteristic chymotrypsin-like fold, whereas domain III is responsible for mediating dimerization. A Cys¹⁴⁵-His⁴¹ dyad forms the catalytic center. The inhibitors in crystal structure of 3CLpro-inhibitor complexes have hydrophobic residues from P₃ to P₁' that fit into the corresponding protease subsites S₃ to S₁' [18-22].

PLpro is a cysteine protease with a canonical catalytic triad composed of Cys¹¹¹, His²⁷², and Asp²⁸⁹. The deubiquitinating and deISGylating activities of the enzyme are permitted by the accommodation of the Gly-Gly motif in the narrow channel in which are located the S₂ and S₁ subsites [23-27].

Initially it was evaluated the stability of **THIA-1-10** in organic solvents and buffer solutions. Subsequently, the inhibition mechanism of each protease was determined using **THIA-3** as the reference compound, after

which IC₅₀ values were calculated for all other THIA derivatives. Although the number of compounds examined in this study was limited, they demonstrated remarkable *in vitro* potency and selectivity toward therapeutically relevant cysteine proteases. Consequently, a series of computational investigations, including conformational analysis, density functional theory (DFT) calculations, molecular docking simulations, and structural and bioinformatics analysis, were conducted to elucidate the molecular basis of the observed activity, thereby providing a foundation for the future optimization of these promising hits.

This study was performed in a multidisciplinary framework, involving the research group of: i) Prof. Giuseppe Caliendo and Prof. Vincenzo Santagada for the synthesis and ii) Prof. Luiz Juliano (University of Sao Paulo; Brazil) for the biological assays. The results obtained in this study are the object of the scientific publication reported in ref. [28].

I was directly involved in performing computational studies, such as conformational search and docking studies, as well as in the interpretation of the experimental results.

The experimental procedure for the synthesis and biochemical assays is not here reported but can be found in the published manuscript [28].

6.2. Results and Discussion.

6.2.1. Inhibitory Activities of THIAs on Cysteine Proteases.

6.2.1.1 Kinetic Mechanism of Inhibition.

The type of inhibition in each protease was identified using **THIA-3** as reference compound for all the tested THIA derivatives. As it can be seen from Figure 1, for both Papain and Catepsin L there was a progressive decrease in activity after 30 min pre-incubation time, suggesting irreversible inhibition.

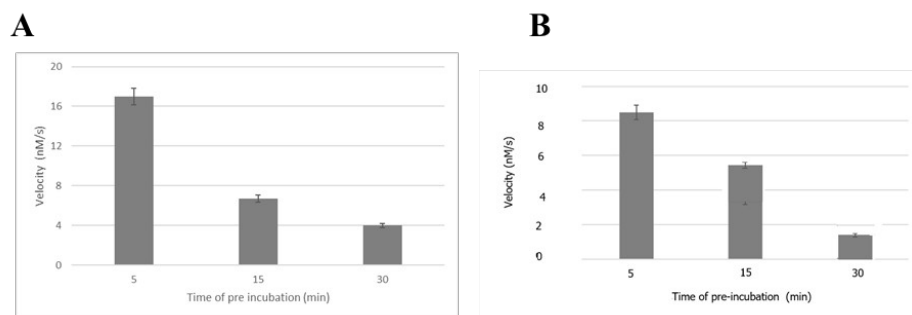


Figure 1. Irreversible inhibition of papain (A) and Catepsin L (B) by **THIA-3**.

In particular, a covalent irreversible inhibition was observed on Papain and 3CLpro by **THIA-3**, reverted by reducing agents (Figure 2), and reversible inhibition on PLpro (Figure 3).

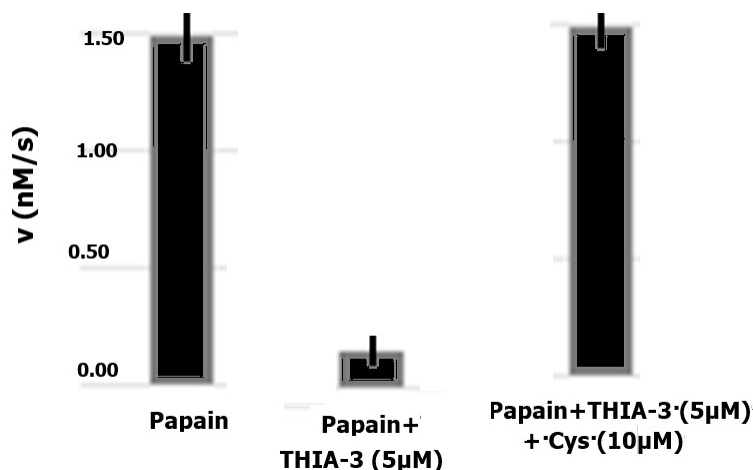


Figure 2. Inhibition of Papain by THIA-3 and its reversibility by cysteine (Cys) as a reducing agent. Z-FR-MCA as substrate, Papain activity was evaluated before and after pre-incubation with 5μM THIA-3 for 5 min and then assayed after Cys addition (10 μM).

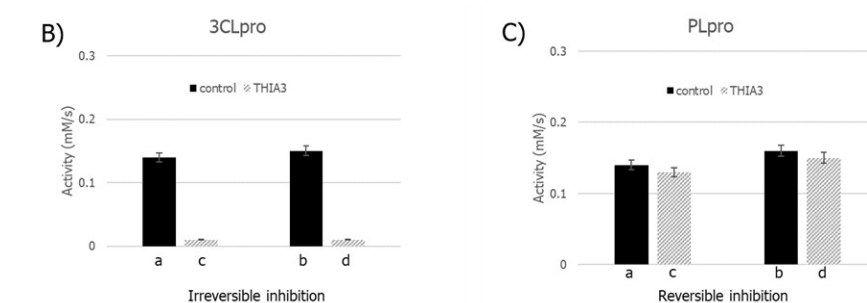
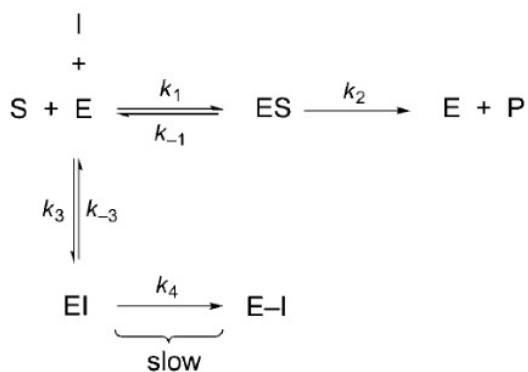


Figure 3. Irreversible inhibition of 3CLpro and reversible inhibition of PLpro by THIA-3.

Due to the susceptibility of THIA derivatives to reducing agents such as cysteine (Cys) or dithiothreitol (DTT) (Figure 2), the proteases Papain, Cathepsin L, 3CLpro, and PLpro were individually pre-activated with 5 mM DTT for 30 minutes. The reducing agent was subsequently removed by gel filtration, yielding stock solutions of the activated enzymes, which were then stored at -20°C under a nitrogen atmosphere for future use.

The reactions of **THIA-3** and **THIA-10** with Papain, and **THIA-3** with 3CLpro are displayed in Figure 4, and they are compatible with a two-step covalent enzyme inactivation mechanism (Scheme 1)



Scheme 1. Scheme of a two-step covalent enzyme inactivation mechanism.

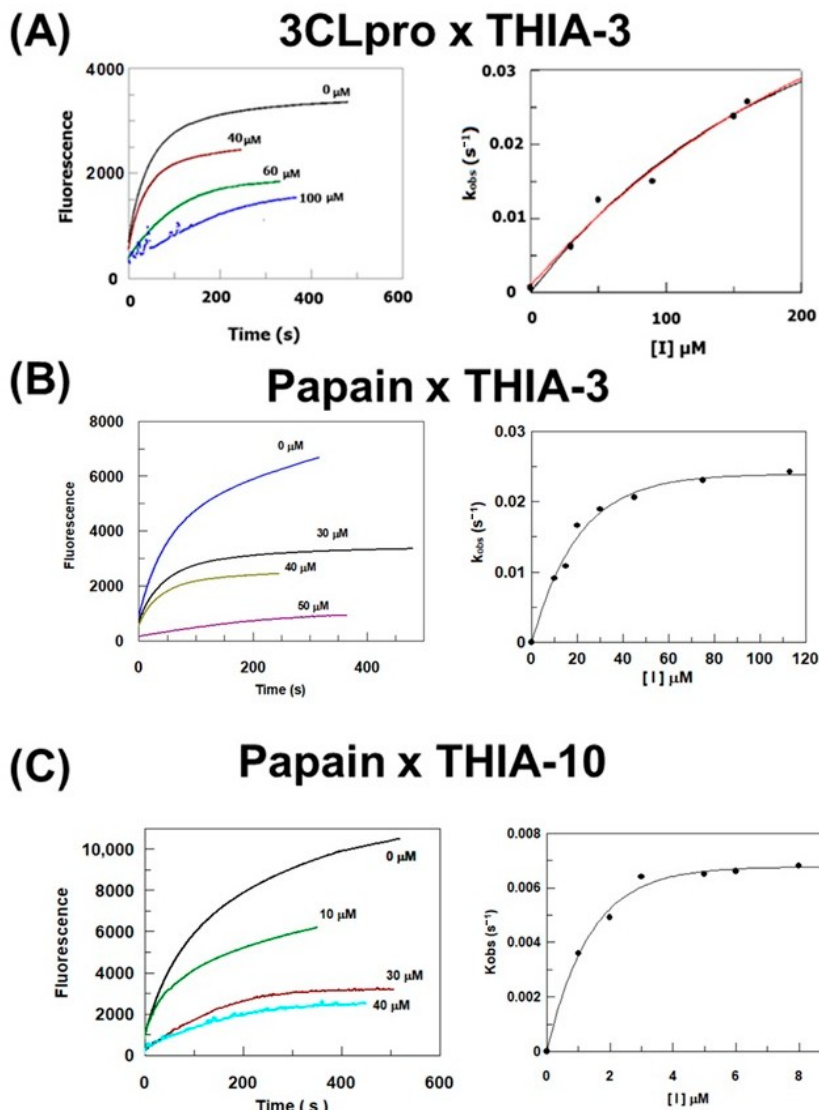


Figure 4. (A) On the left graphic: Time course of the hydrolysis of Abz-SAVLQSGFRK(Dnp)NH₂ by 3CLpro in the absence and the presence of THIA-3 (concentration indicated close to the curves) – On the right graphic: inactivation rate constants as a function of the concentration of THIA-3. (B) and (C)- On the left graphics: Time course of the hydrolysis of Z-FR-MCA by Papain in the absence and the presence of THIA-3 or THIA-10 (concentration indicated close to the curves)- On the right graphics: inactivation rate constants as a function of the concentration of THIA-3 or THIA-10.

The kinetic constant values for the reactions of the ten THIA compounds with Papain, corresponding to the reaction pathways illustrated above, are reported in Table 1.

Table 1. Values of the kinetic constants for Papain inactivation kinetics analyses by all THIA compounds

Inhibitors	k_3 ($\mu\text{M.s}$)⁻¹	k_4 (s^{-1})	K_i (μM)
THIA-1	19.9	0.013	0.26
THIA-2	24.6	0.012	0.32
THIA-3	19.7	0.008	0.17
THIA-4	4.9	0.015	0.08
THIA-5	12.3	0.021	0.27
THIA-6	14.5	0.019	0.27
THIA-7	17.3	0.009	0.16
THIA-8	109.9	0.016	0.68
THIA-9	14.5	0.024	0.34
THIA-10	13.7	0.022	0.31

Conditions: Papain 48nM, substrate Z-FR-MCA 1.5 μM . Buffer: 100 mM sodium phosphate, 1 mM EDTA, pH 6.0. The kinetic constants for the reaction are indicated in the scheme below.

These results confirm a two-step, irreversible reaction: inhibitor binding and the formation of a covalent inhibitory complex.

The IC₅₀ values were obtained for all four-cysteine protease inhibited by THIA derivatives with pre-incubation time at 10 min. These values are summarized in Table 2.

Table 2. IC₅₀ for Inhibition of Papain, 3CLpro, PLpro, and Cathepsin L by THIA compounds.

Inhibitors	Papain (μM)	Cathepsin L (μM)	3CLpro (μM)	PLpro (μM)
THIA-1	2.16 ± 0.09	14±1	0.13 ± 0.01	25 ± 3
THIA-2	1.28 ± 0.07	21± 2	0.12 ± 0.01	8.3 ± 0.8
THIA-3	0.14±0.02	0.12± 0.01	0.11 ± 0.01	8.5 ± 0.2
THIA-4	0.071±0.006	0.40±0.05	0.13 ± 0.01	7.6 ± 0.9
THIA-5	0.102 ± 0.003	0.15±0.01	0.09 ± 0.01	3.9 ± 0.7
THIA-6	0.010±0.001	0.35±0.02	0.03 ± 0.01	12.5 ± 0.9
THIA-7	0.13± 0.01	0.42 ± 0.05	0.01± 0.01	14.4 ± 0.7
THIA-8	2.4± 0.1	27±2	0.19± 0.01	4.2 ± 0.1
THIA-9	0.11± 0.01	0.21 ± 0.01	0.16± 0.08	10.4 ± 0.4
THIA-10	0.011± 0.001	0.33 ± 0.03	0.02± 0.01	10 ± 1

Conditions: Acetonitrile was the solvent in all stock solutions. Temperature of the reactions: 30 °C. Pre-incubation of 10 min with each inhibitor before addition of the substrates. Buffers and substrates:

Papain: 100 mM sodium phosphate, 1 mM EDTA, pH 6.0, Substrate: Z Phe-Arg-MCA.

Cathepsin L: 100 mM sodium acetate, 1 mM EDTA, pH 5.0, Substrate: Z Phe-Arg-MCA.

3CLpro: 50 mM Tris, 1 mM EDTA, pH 7.5, Substrate: Abz-SAVLQSGFRK(Dnp)-NH₂.

PLpro: 50 mM Tris, 1 mM EDTA, pH 7.5, Substrate: Abz-TLKGGAPIKQ-EDDnp

6.2.1.2. IC₅₀ Values with Papain and Cathepsin L.

As showed in Table 2, the inhibitory activity trend of **THIA-1-10** is very parallel for cathepsin L and papain, with the majority of the compounds more active on papain. Among the worst inhibitor for both proteases, we have **THIA-1**, **THIA-2** and **THIA-8** with IC₅₀ values in the μM range. Instead, the other THIAs showed values in the nM range. A common feature is present in these compounds: an alkyl group as R substituent. In the other compounds R is a phenyl ring with or without *para*-substitution.

Taking in account the papain family, the subsites have some particular features: S₁ subsite is strictly conserved (Table 3) with a preference for positively charged residues (i.e. arginine and lysine); S₂ subsite, on the other hand, show differences in shape and size and represent the factor in defining substrate or inhibitor specificity.[7,12,13,15,29]; some differences are present in S₃ subsite (Table 3) between papain and cathepsin L due to their broader interaction specificity [8].

Table 3. List of amino acids located in the subsites S₁, S₂ and S₃ of 3CLpro, PLpro, Cathepsin L, Papain, and Bromelain.

Subsite	3CLpro	PLpro	Cathepsin L	Papain	Bromelain
S ₁	Phe ¹⁴⁰	Trp ¹⁰⁶	Gln ¹⁹	Gln ¹⁹	Gln ²⁰
	Asn ¹⁴²	Asn ¹⁰⁹	Gly ²³	Gly ²³	Gly ²⁴
	Ser ¹⁴⁴	Cys ¹¹¹	Cys ²⁵	Cys ²⁵	Cys ²⁶
	Cys ¹⁴⁵	Tyr ¹¹²	Asp ¹⁶²	Asp ¹⁵⁸	Asn ¹⁵⁷
	His ¹⁶³	His ²⁷²	His ¹⁶³	His ¹⁵⁹	His ¹⁵⁸
	His ¹⁷²		Trp ¹⁸⁹	Trp ¹⁷⁷	Trp ¹⁸¹
S ₂	His ⁴¹	Leu ¹⁶²	Met ⁷⁰	Pro ⁶⁸	Glu ⁶⁸
	Cys ⁴⁴	Gly ¹⁶³	Ala ¹³⁵	Val ¹³³	Ala ¹³³
	Met ⁴⁹	Tyr ²⁶⁴	Met ¹⁶¹	Val ¹⁵⁷	Leu ¹⁵⁶
	Pro ⁵²	Gly ²⁷¹	Gly ¹⁶⁴	Ala ¹⁶⁰	Ala ¹⁵⁹
	Tyr ⁵⁴	Tyr ²⁷³	Ala ²¹⁴	Ser ²⁰⁵	Asp ²⁰⁹
	Met ¹⁶⁵				
	Asp ¹⁸⁷				
	Gln ¹⁸⁹				
S ₃	Glu ¹⁶⁶	Asp ¹⁶⁴	Gly ⁶¹	Arg ⁵⁹	Lys ⁵⁹
	Pro ¹⁶⁸	Tyr ²⁶⁸	Asn ⁶²	Ser ⁶⁰	Gly ⁶⁰
		Gln ²⁶⁹	Glu ⁶³	Tyr ⁶¹	Tyr ⁶¹
			Gly ⁶⁷	Gly ⁶⁵	Gly ⁶⁵
			Gly ⁶⁸	Gly ⁶⁶	Gly ⁶⁶
			Leu ⁶⁹	Tyr ⁶⁷	Trp ⁶⁷
			Tyr ⁷²	Ser ⁷⁰	Arg ⁷⁰

THIA-1 and **THIA-2**, with a benzyl group, and **THIA-8**, with a cyclohexyl group, are the least active inhibitors on both enzymes, indicating that the effect exhibited by the presence of an alkyl carbon at N4

on the reactivity of the THIA ring may be involved in the drop of inhibitory activity. Notably, compound CCG-50014, not active on papain [4], has a benzyl group in R position, similar to **THIA-1** and **THIA-2**, but with the difference of the presence of a fluorine atom in *para* position. It is possible that the properties of CCG-50014 are altered by the fluorine atom in *para* position of the benzyl group [30-32], impairing the reaction with papain. On the other hand, **THIA-7** (R = 4-chlorophenyl) inhibited papain and Cathepsin L with IC₅₀ values in nanomolar range, similar to **THIA-4**, unsubstituted at the R position. Since the halogen at the *para* position of the phenyl ring did not affect the inhibition process, the cause of the drop in activity is probably related to the benzyl group.

One of the most active inhibitors of Papain is **THIA-6** (4-methoxyphenyl at N2 and N4). Interestingly, **THIA-6** IC₅₀ value (0.01 µM) is lower than that of **THIA-4** (0.07 µM), presenting one 4-methoxyphenyl at N2 (R' substituent). In turn, **THIA-4** is lower than **THIA-3**, with two unsubstituted phenyl rings as substituents. This indicates that the presence of 4-methoxyphenyl substituents may have favourable interactions with Papain.

Finally, among the best inhibitors there is **THIA-10**, with IC₅₀ (0.01 µM) lower than that of **THIA-3** (0.14 µM). The only difference between these compounds is the presence of fluorine atom in *para* position of the phenyl ring. This fluorine atom is supposed to accelerate the nucleophilic attack of the papain catalytic Cys on the sulphur atom (S1) of the THIA ring, thanks to its electron-negative inductive effect, evident in the release of H₂S in the reaction between THIA-10 and Cys [3]. For the Cathepsin L inhibition, **THIA-10** showed a IC₅₀ similar to that of **THIA-3**, while for **THIA-1**,

THIA-2 and **THIA-8**, the values of IC_{50} were ten times higher than that of Papain.

Bromelain is unique among cysteine protease, since it requires a substrate containing a positively charged residue at P_2 position because of the presence of Glu₆₈ in S_2 subsite. For this reason, it was evaluated the role of S_2 subsite in selective inhibition against bromelain [33-34]. Inhibition was observed only in the case of **THIA-7** with an IC_{50} of $2.0 \pm 0.1 \mu M$. This compound presents a *para*-Cl-phenyl group at R position, with the chlorine atom that induces dipole charge resonance structures in the phenyl group, forming a favourable interaction with Glu⁶⁸ carboxylate group in S_2 subsite (Figure 5).

6.2.1.3. IC_{50} Values with SARS-CoV-2 Cysteine Proteases.

The ten THIA derivatives showed inhibition of 3CLpro in the nanomolar range (Table 2), thus exhibiting good performance as protease inhibitors. Among the best inhibitors we have **THIA-6**, **THIA-7** and **THIA-10**, with IC_{50} values ten times lower of the other tested THIAs, except for **THIA-5**. 3CLpro and PLpro have 12 and 11 residues of cysteine, respectively, differently from Papain and Cathepsin L (Figure 5).

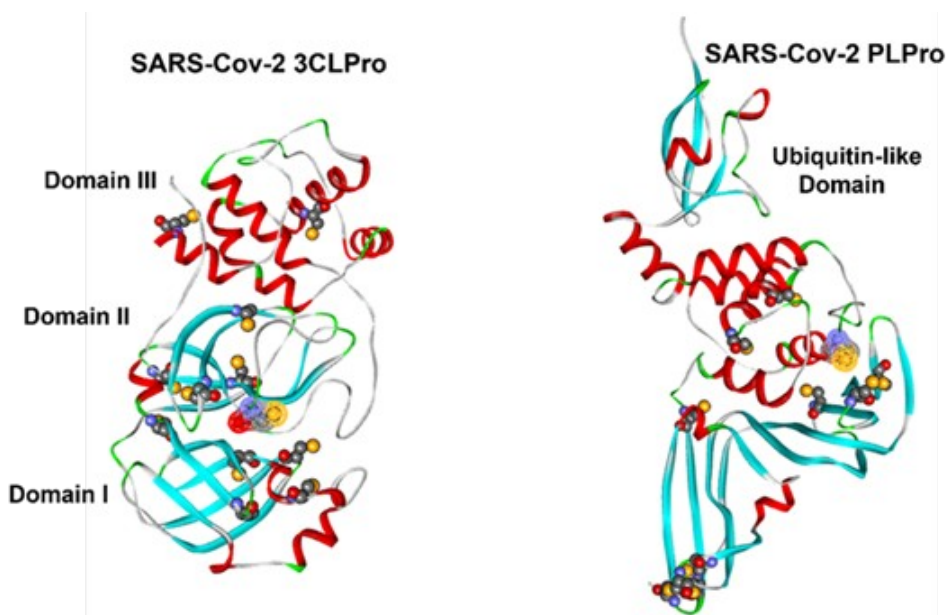


Figure 5. Free Cys residues in SARS-CoV-2 3CLpro (PDB ID: 7JKV) and PLpro (PDB ID: 6WX4). The cysteine residues are displayed as CPK and colored by atom type, with the solvent-accessible surface of the catalytic cysteine highlighted. The protein backbone is shown as ribbons and colored by secondary structure (white = loop; green = turn; red= α -helix; cyan= β -strand).

Since the room-temperature crystal of 3CLpro showed that only Cys¹⁴⁵ is oxidized and only Cys¹⁴⁵ and Cys¹⁵⁶ react with the alkylating agents N-ethylmaleimide [21], it can be deduced that the irreversible inhibition of 3CLpro by THIA derivatives happens with the reaction with Cys¹⁴⁵ in the catalytic site [22].

The IC₅₀ values of THIA compounds with PLpro are all in the μ M range, with about half of them exceeding 10 μ M (Table 2). The type of inhibition is reversible (Figure 3C) due to inaccessibility of S₁ and S₂ subsites to molecules that are larger than glycine-glycine, resulting in a weaker inhibition of these compounds respect to covalent inhibitors that resulted

from the nucleophilic reaction with Cys¹¹¹ in the catalytic site [35]. This kind of inhibition also suggests that the other ten Cys residues of PLpro (Figure 5) are not reactive towards THIA compounds under the conditions of the assays. The noncovalent deubiquitinase inhibitor 5-amino-2-methyl-N-[(1R)-1-(1-naphthalenyl)ethyl]benzamide (GRL0617) inhibited PLpro ($IC_{50} = 0.6 \mu M$) through an alternative mechanism, binding to the S₄-S₃ subsites and inducing a conformational change in the loop that closes the access to the catalytic site [36]. **THIA-5** and **THIA-8** have the lowest IC_{50} values (3.9 μM and 4.2 μM , respectively) and have some components such as GRL0617. The superimposition between GRL0617 and **THIA-5** in complex with PLpro reinforce the hypothesis that these compounds bind to the same allosteric site.

6.2.2. Molecular Modelling studies.

A comprehensive set of computational studies was performed in order to investigate the molecular basis of the inhibitory activity of THIA derivatives. Firstly, the conformational and electronic properties of the compound listed in Table 1 were evaluated by using a combination of Molecular Mechanics (MM) and Density Functional Theory (DFT) calculations. Then, molecular docking simulations were performed using representative subset of derivatives, spanning from the least active (**THIA-8**) to the most active (**THIA-7**). These simulations were designed to study the adaptation of the inhibitor in the active site of the enzyme by positioning the reactive warhead in proximity to the catalytic cysteine, facilitating in this way the nucleophilic attack necessary for covalent bond formation. Moreover, a bioinformatic and structural analysis was performed to better understand the observed selectivity with respect to other cysteine proteases and Steered Molecular Dynamics (SMD) was performed in order to investigate the dynamic nature of ligand-protein

interaction and also to validate docking results. At least covalent docking was performed to elucidate the potential adaptation of the inhibitor in the active site of the enzyme after the formation of a covalent bond.

6.2.2.1. Ligand Analysis.

The conformational space of THIA derivatives was explored by performing a stochastic conformational search [37] and then by DFT full optimization of the MM conformers mimicking an aqueous environment using the conductor-like polarizable continuum model (C-PCM) [38]. The comparison of MM conformers with the resulting DFT ones evidences significant structural changes and different MM conformers converge to the same DFT conformer (Appendix D: Tables 1-10; Figure 6). These results highlight the necessity to use DFT methods to model highly conjugated structures, such as THIA compounds.

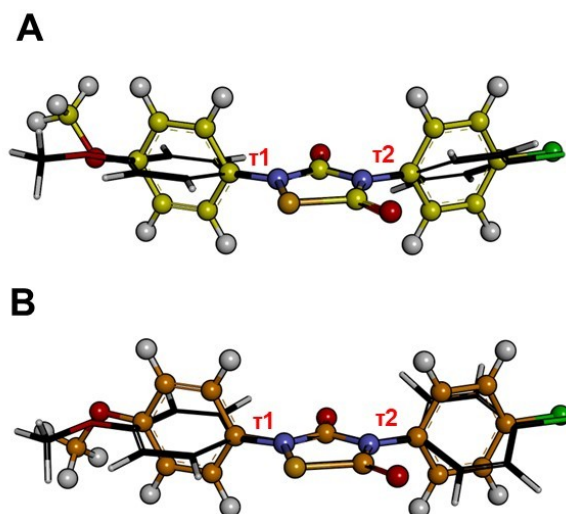


Figure 6. Structural comparison between the MM (stick; black) and DFT (ball& stick) conformers of THIA-7. The conformational enantiomer A is shown in (A) while the conformational enantiomer B in (B). The ligands are coloured by atom type: O=red; N=blue; S=yellow; Cl=green; H=white.

Going into details, the planes of the substituted phenyl rings at N2 (τ_1) and N4 (τ_2) assume always a perpendicular orientation with respect to that of 1,2,4-thiadiazolindine-3,5-dione ring (τ_1 and $\tau_2 \sim 90^\circ$). These findings indicate that the conjugation of N2 and N4 with the phenyl substituent is unfavoured respect to the conjugation with the adjacent carbonyl groups as confirmed by the nitrogen-carbonyl bond length and order (partial double) compared to those of the nitrogen-phenyl carbon bond (single). Therefore, oxygen atoms of the carbonyl functions, particularly O3, harbours the most negative values of the molecular electrostatic potential (Figures 7 and 8).

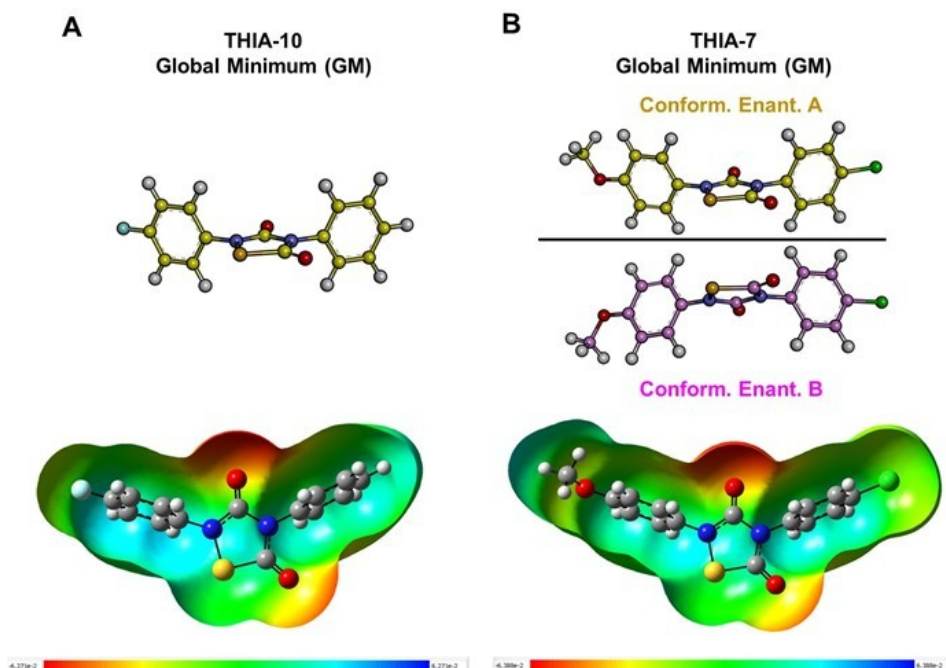


Figure 7. DFT conformers of (A) THIA-10; (B) THIA-7, with conformational enantiomers A and B. Structures are displayed as balls and sticks; heteroatoms are colored by atom type: O=red; N=blue; S=yellow; Cl=green. Bottom: molecular electrostatic potential surface (isovalue $0.0004 \text{ e}^-/\text{a.u.}^3$); the bond order is reported according to bond length; the frontmost portions of the surface have been removed to allow views into the interior of the molecular display (z-clip option).

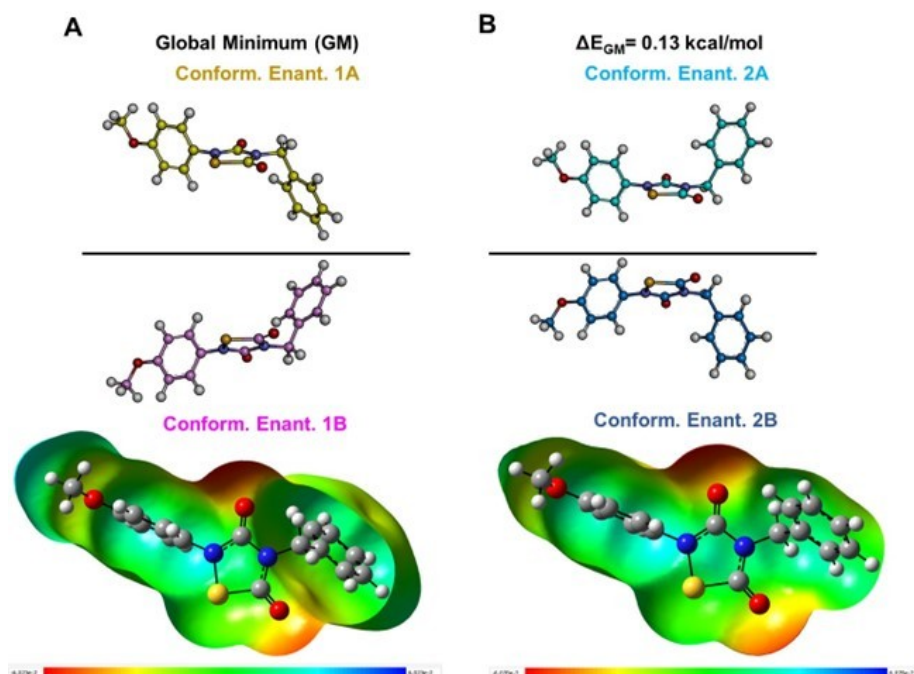


Figure 8. DFT conformers of THIA-2: the conformational enantiomers 1A and 1B are shown in (A), while the conformational enantiomers 2A and 2B are shown in (B). Structures are displayed as balls and sticks; heteroatoms are colored by atom type: O=red; N=blue; S=yellow. Bottom: molecular electrostatic potential surfaces (isovalue 0.0004 e⁻/a.u.³); the bond order is reported according to bond length; the frontmost portions of the surfaces have been removed to allow views into the interior of the molecular display (z-clip option).

The results of conformational analysis showed that compounds with two rotatable bonds (**THIA-3**, **THIA-9** and **THIA-10**) have just one conformer (Appendix D: Tables 3, 9 and 10; Figure 7A), while compounds such as **THIA-1**, **THIA-4**, **THIA-5** and **THIA-7**, with three rotatable bonds, present conformational enantiomers (couple of specular, not superimposable and isoenergetic, namely A and B in Figure 7B; Appendix D: Tables 1,3,4 and 6). Finally, **THIA-2** and **THIA-6**, with four rotatable bonds, and **THIA-8**, whose cyclohexyl ring adopts two almost

isoenergetic orientations, have two couples of conformational enantiomers (namely, 1A-B and 2A-B; $\tau_2 = 180^\circ$ or 0° , respectively) (Figure 8, Appendix D: Tables 2,6 and 9.)

The ability to fit in into the active site of cysteine protease by the THIA compounds are likely linked to the presence of conformational enantiomers and their limited energetically accessible conformational space. In line with the frontier orbital theory, the energy of the lowest unoccupied molecular orbital (LUMO) is responsible for the propensity of a molecule to acquire electrons. Thus, the LUMO energy of the DFT conformers was calculated (Table 4) in order to investigate the tendency of THIA compounds to undergo nucleophilic attack by the sulfur atom of the catalytic cysteine [39].

Table 4. Lowest unoccupied molecular orbital (LUMO) energy of the DFT conformers of THIA1-10.

Cmp	DFT Conformer	E _{LUMO} (kcal/mol)
THIA-7	1 A/B	-23.02
THIA-10	1 A	-22.95
THIA-3	1 A	-22.71
THIA-9	1 A	-22.53
THIA-4	1 A/B	-22.40
THIA-5	1 A/B	-22.02
THIA-6	1 A/B	-21.83
	2 A/B	-21.82
THIA-8	1 A/B	-20.41
	2 A/B	-20.94
THIA-1	1 A/B	-20.68
THIA-2	1 A/B	-19.99
	2 A/B	-19.82

These results evidenced that compounds presenting a benzyl (**THIA-1** and **THIA-2**) or a cyclohexyl (**THIA-8**) ring have less propensity to acquire electrons compared to the ones having a (substituted)-phenyl ring (**THIA-3**, **THIA-4**, **THIA-5**, **THIA-6**, **THIA-7**, **THIA-9** and **THIA-10**). In addition, among the substituted-phenyl ring bearing compounds there is a variation of the LUMO energies (~ 1 kcal/mol) depending on the nature of the *para*-substituent on the two phenyl rings. Electron-donating substituents, such as *p*-Ome or *p*-Me (**THIA-4**, **THIA-5**, **THIA-6** and **THIA-9**) have as effect an increase in the LUMO energy, while electron-withdrawing substituents, such as *p*-Cl (**THIA-7**) and *p*-F (**THIA-10**), determine a decrease in LUMO energy (Table 4). By comparing Table 4 with Table 3, it can be seen that **THIA-1**, **THIA-2** and **THIA-8** ability to act as covalent inhibitor is lower than that of the other THIA derivatives in the case of Papain and Cathepsin L, but the inhibitory activity on 3CLpro is hardly affected. This observation suggests that not only the compound reactivity, but also the role of R and R' substituents in the ability of the compound to interact with the 3CLpro active site by correctly orienting the reaction partner with the catalytic cysteine is crucial in the inhibitory activity toward this enzyme.

6.2.2.2. Docking Calculations.

Docking simulations were performed to investigate the accommodation of THIA derivatives in the active site of 3CLpro immediately before the nucleophilic attack by the catalytic cysteine residue. It was taken in account the activation of Cys¹⁴⁵ and the interaction with His⁴¹. The results were filtered based on the orientation of the reacting atoms. To reach this goal, it was applied a combination of shape-based grid docking (LigandFit) and MM energy minimization (CFF forcefield; Dassault Systèmes BIOVIA, San Diego, 2023) [40].

All conformers of **THIA-2** (four), **THIA-4** (two), **THIA-7** (two), **THIA-8** (four) and **THIA-10** (one) were used to generate distinct ligand-protein starting complexes for a total of thirteen docking run, during which the ligand conformation was tethered. This was done in order to consider all possible ligand conformers and preserve the DFT-calculated ligand conformation during the docking (force-field based) simulations.

A structural and bioinformatic analysis was performed to define the area of the protein to be explored during the docking run (the ligand binding domain). 3CLpro is formed by three domains: Domain I and II connected by a long loop region (LH-Loop) to Domain III (Figure 9A).

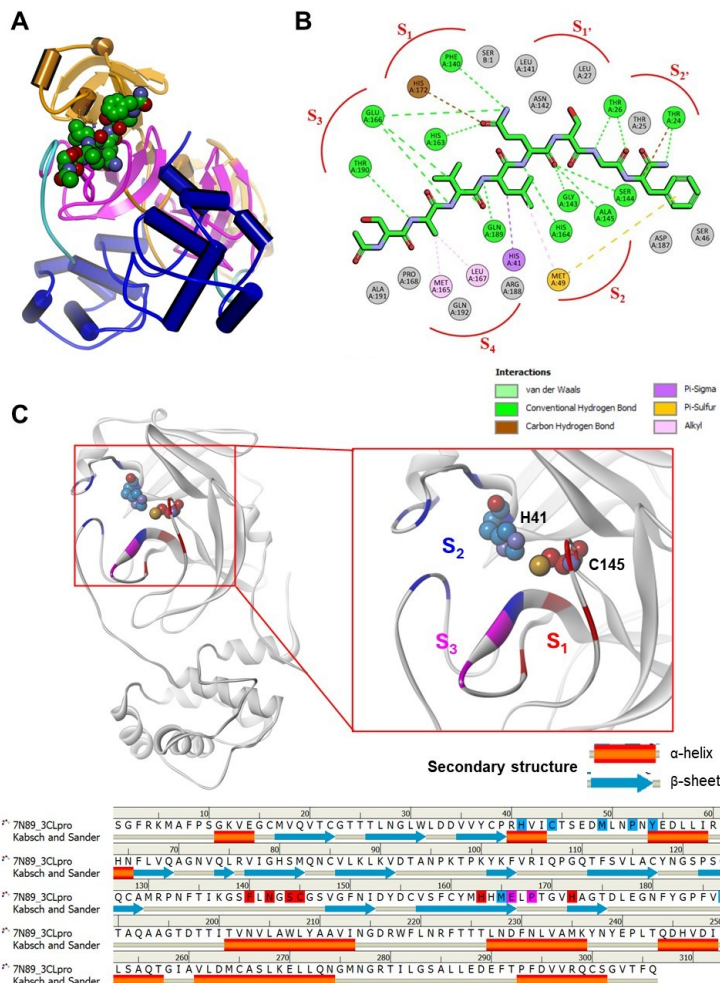


Figure 9. A: X-ray structure of SARS-CoV-2 3CLpro C145A mutant in complex with substrate Ac-SAVLQSGF-CONH₂ (PDB ID: 7N89). 3CLpro is coloured by protein domains: Domain I is coloured in orange; Domain II in magenta, LH-loop in cyan, and Domain III in blue. The secondary structure is displayed as wide cylinders (α -helices), arrows (beta-sheets), and tubes (coils and turns). The substrate is displayed in CPK and coloured by atom type (C= green, O = red, N = blue). B: schematic representation of the interactions between substrate and 3CLpro. All residues within 5 Å of any ligand atom are displayed. C: 3CLpro is displayed in solid ribbon and colored in grey. The ribbon of residues of S₁, S₂, and S₃ subsites is coloured in red, blue, and magenta, respectively. The catalytic residues Cys and His are displayed in CPK and labelled. The sequence and the secondary structure assignments (Kabsch and Sander) are reported.

The catalytic dyad (i.e. Cys¹⁴⁵ and His⁴¹) is buried by the substrate-binding site present in a cleft between Domains I and II. Upon the reaction of the sulfur atom of Cys¹⁴⁵, the hydrogen is transferred as a proton to the nitrogen atom of the scissile peptide bond via protonation of His⁴¹ that forms the oxyanion hole along with the backbone amides of Gly¹⁴³, Ser¹⁴⁴ and Cys¹⁴⁵ which assist the stabilization of the thiohemiketal intermediate [41]. The catalytic site is divided into subsites, namely, S₁-S₅ and S_{1'}-S_{2'} [42,43] (Figure 10B). To fully cover the subsites, a binding site sphere of 12 Å around the mass centre of the co-crystallized inhibitor of the 3CLpro-GRL-2420 complex (PDB ID: 7JKV) (Volume 522.5 Å³; Figure 10) defined the ligand binding domain area.

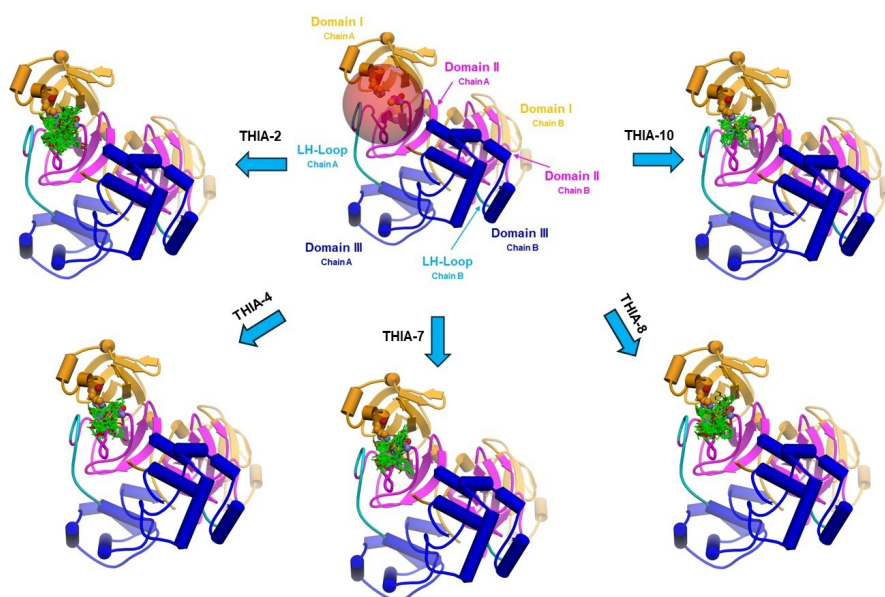


Figure 10. X-ray structure of SARS-CoV-2 3CLpro used as protein starting structure in docking studies (PDB ID: 7JKV). The search area defined for ligand docking is represented as a transparent red sphere. The secondary structure is displayed as wide cylinders (α -helices), arrows (β -sheets), and tubes (coils and turns), and is coloured by protein domains: Domain I of each monomer is coloured in orange; Domain II in magenta, LH-loop in cyan, and Domain III in blue. THIA-2, THIA-4, THIA-7, THIA-8, and THIA-10 binding poses resulting from docking studies are shown. The ligands are displayed in stick and coloured by atom type (N=blue; O=red; S=yellow; Cl=green; F=cyan).

A tethering restrain was applied to the hydrogen bond between Cys¹⁴⁵ and His⁴¹ during docking simulations to preserve the activation of the catalytic cysteine.

Each docking run produced up to 20 complexes (Figure 10) that were then subjected to energy minimization. The resulting complexes were filtered based on the maximum distance between the enzyme reactive atoms required for nucleophilic attack and the ligand (i.e. distance between sulfur atom S₁ of THIA compound and the sulfur atom of Cys¹⁴⁵ ≤ 4 Å) [44,45] and also the activation of Cys¹⁴⁵ by His⁴¹ was taken in account. Then, the ligand-protein interaction energy was calculated and the resulting structure were ranked and analyzed.

For **THIA-2/3CL**pro docked complexes, it can be seen that the ligand can approach the catalytic Cys¹⁴⁵ by placing the R and R' substituents in different subsites resulting in four possible binding modes (BM-I, BM-II, BM-III and BM-IV; Figure 11; Appendix D: Tables 11-15).

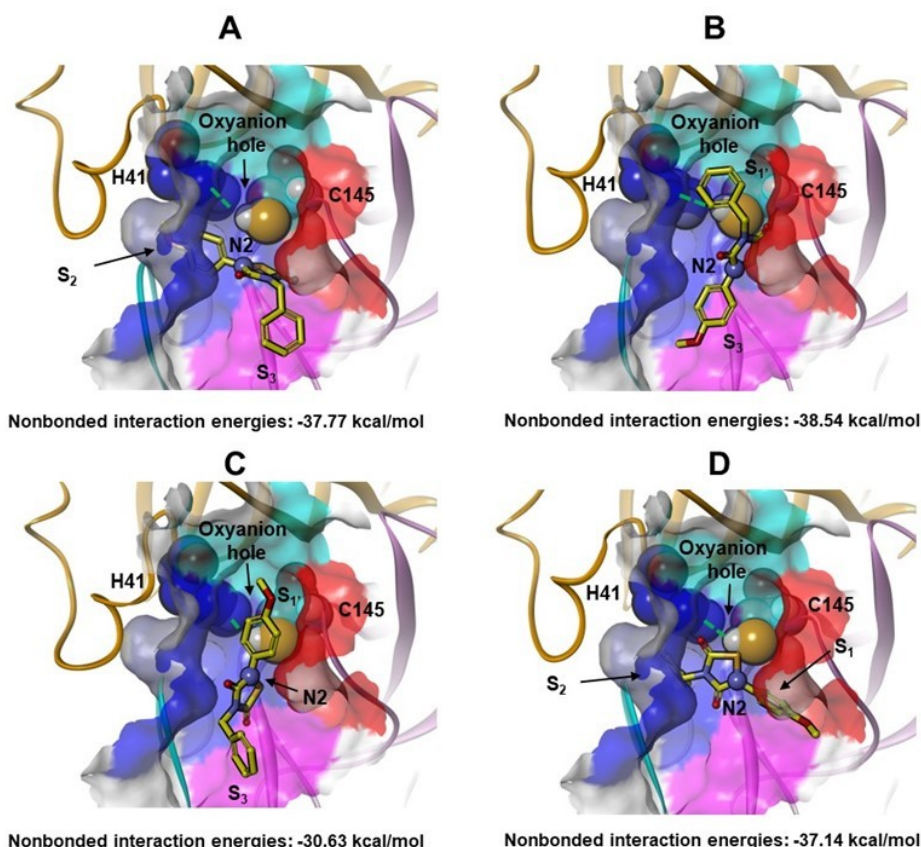


Figure 11. Docking calculations obtained four binding modes of THIA-2 to 3CLpro: BM-I (A), BM-II (B), BM-III (C), and BM-IV (D). The protein backbone is displayed as ribbons and colored by protein domains: orange (Domain I: aa 1-101), pink (Domain II: aa 102-184), and cyan (LH-Loop: aa 185-200). The solvent-accessible surface of S₁' (cyan), S₁ (red), S₂ (blue), and S₃ (magenta) subsites are displayed in solid (transparency 50%). The catalytic residues His⁴¹ and Cys¹⁴⁵ are displayed in ball&stick. The hydrogen bond between the catalytic dyad, His⁴¹, and Cys¹⁴⁵, is shown as a green dashed line. THIA-2 (yellow) is displayed on stick. Atoms are colored by type (N: blue; O: red; S: orange). THIA-2 N2 nitrogen atom is evidenced in a ball.

Some considerations can be made: i) the reaction mechanism of THIA derivatives with thiols includes the cleavage of the S1-N2 bond, together with the protonation of N2 [2], in a similar way to the protonation of the

scissile peptide bond nitrogen [41]; ii) in 3CLpro this proton exchange occurs in the oxyanion hole, where proton shifts from Cys¹⁴⁵ to the substrate nitrogen via His⁴¹. Accordingly, for THIA N2 to be protonated, it must be positioned in the oxyanion hole, a possibility achievable only in binding mode BM-I, with N2 in S2 subsite and N4 in S3 subsite (Figure 12A). On this basis, the BM-I complex with the best ligand-protein binding energy was selected as the best-docked structure (Appendix D: Table 5 and Tables 11-16; Figures 12-14).

Table 5. THIA-3CLPro best docked complexes. The ligand conformation and LUMO energy, the ligand-protein non-bond interaction energy, and the distance between the sulfur atoms of THIA and Cys¹⁴⁵ are reported.

Complex	Nonbonded interaction energies (kcal/mol)	Dist S1-S _{Cys}	DFT conformer	E _{LUMO} (kcal/mol)
THIA-2/3CLpro	-37.77	3.43	1B	-19.82
THIA-8/3CLpro	-33.28	3.19	2B	-20.94
THIA-7/3CLpro	-31.07	3.34	1B	-23.02
THIA-4/3CLpro	-30.06	3.39	1B	-22.40
THIA-10/3CLpro	-25.50	3.05	1A	-22.95

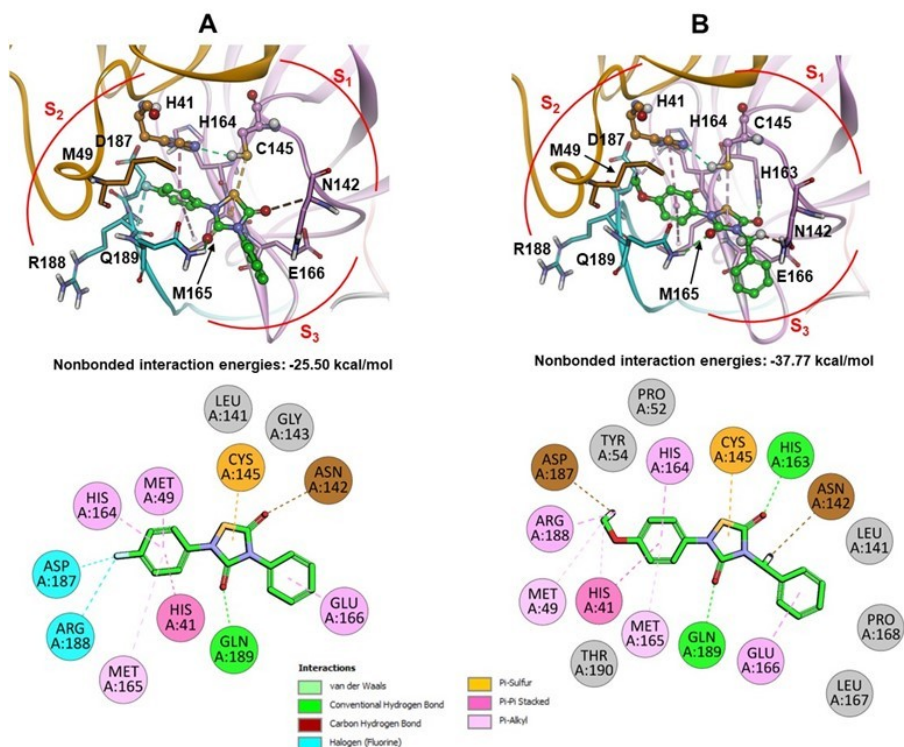


Figure 12. 3CLpro-THIA-10 best-docked complex (A) and 3CLpro-THIA-2 best-docked complex (B). Up: ligand binding site; bottom: 2-D representation of ligand-protein interactions. Protein is displayed as a ribbon; ligand is displayed as a ball and stick (up) or stick (bottom) and colored in green. Protein residues showing interaction energy with the ligand > 1 kcal/mol are displayed and labeled in black. 3CLpro substructures are colored as follows: orange (Domain I: aa 1-101), pink (Domain II: aa 102-184), and cyan (LH-Loop: aa 185-200). The residues and the ligand are colored by atomic type (C=green, O=red, N=blue, F=cyan, S=yellow, H=white).

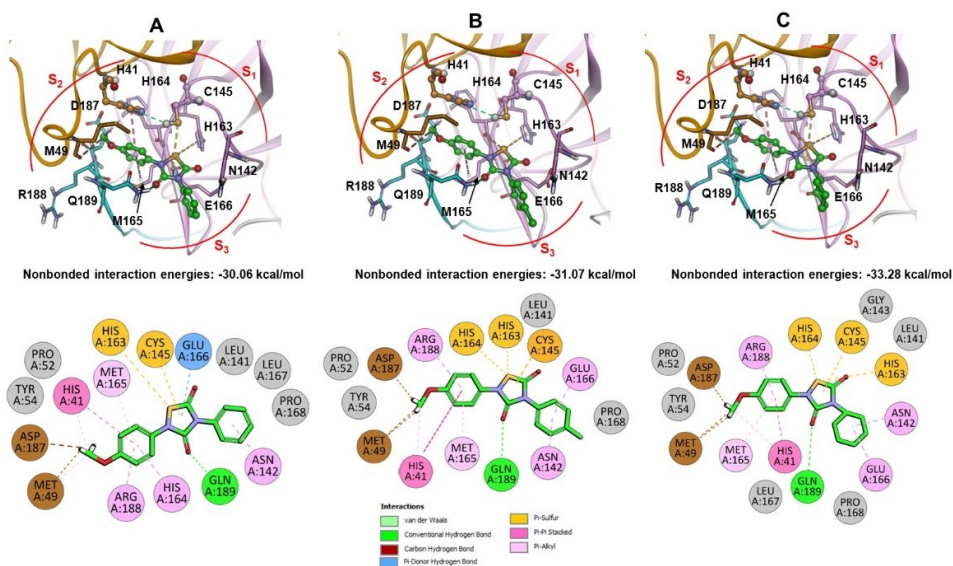


Figure 13. 3CLpro-THIA-4 best-docked complex (A), 3CLpro-THIA-7 best-docked complex (B), and 3CLpro-THIA-8 best-docked complex (C). Up: ligand binding site; bottom: 2-D representation of ligand-protein interactions. Protein is displayed as ribbons; ligand is displayed as ball&stick (up) or stick (bottom) and colored in green. Protein residues showing interaction energy with the ligand > 1 kcal/mol are displayed and labeled in black. 3CLpro substructures are colored as follows: orange (Domain I: aa 1-101), pink (Domain II: aa 102-184), and cyan (LH-Loop: aa 185-200). The residues and the ligand are colored by atomic type (C=green, O=red, N=blue, Cl=light green, S=yellow, H=white).

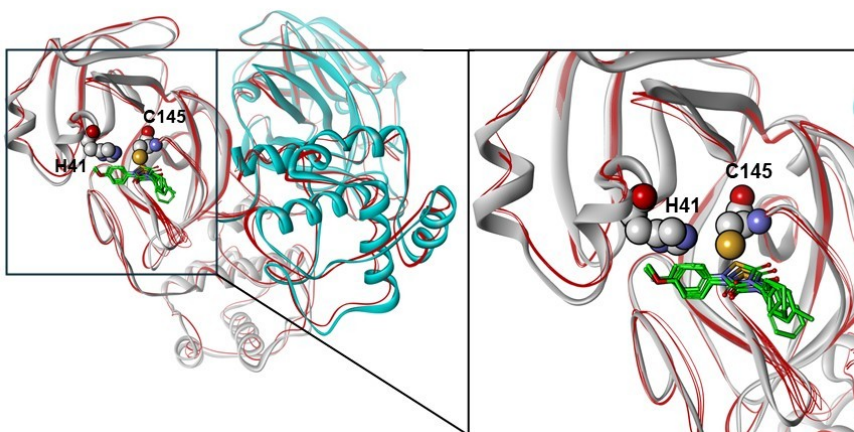


Figure 14. 3CLPro-THIA best docked complexes superimposed on the starting structure (PDB ID: 7JKV; Ca pairs with a RMSD < 1Å). The backbone of the starting complex is displayed as solid ribbon and coloured in grey monomer A and cyan monomer B. The backbone of the calculated complexes is displayed as line ribbons and coloured in red. The ligands are coloured by atom type (C: green; N: blue; S: yellow; O: red) and displayed in stick. The two catalytic residues are shown in CPK and labelled.

The contribution of each protein residue to the binding energy was evaluated (Appendix D: Tables 16-20), and by using a geometry-based method [42] it was estimated the number and types of ligand-protein interactions. All THIA derivatives interact with several residues also involved in the interaction with the substrate, such as Gln¹⁸⁹, His⁴¹, Met⁴⁹, His¹⁶³, His¹⁶⁴, Met¹⁶⁵ and Glu¹⁶⁶ (Figures 10B, 12 and 13) with the strongest contribution to the binding energy (~ -4 kcal/mol) represented by the hydrogen bond between O3 and the side chain of Gln¹⁸⁹ (S_2 subsite; Figures 12 and 13; Appendix D: Tables 16-20). In particular, the Ome group at R' of **THIA-2**, **THIA-4**, **THIA-7** and **THIA-8** contributes to the ligand binding energy by forming favourable interactions with the S_2 subsite (especially with Asp187, Met49 and Arg188; Figures 13B and 14; Appendix D: Tables 16-19) and always presents the same conformation

(conformational enantiomer B). The other carbonyl group (C5=O5) is oriented toward S₁, interacting with His¹⁶³ and partially occupying this subsite depending on the considered THIA derivative (Figures 12 and 13). **THIA-2** displays the best accommodation in the catalytic site (best binding energy), with a hydrogen bond between O5 and the side chain of His¹⁶³ (Figure 11B) mimicking in this way the interaction between the side chain of the P1 residue (Gln) of the substrate (Figure 11B).

The most favourable ligand-protein binding energy is displayed by **THIA-2** and **THIA-8** (Table 5). Notably, **THIA-1**, **THIA-2** and **THIA-8** have a sp³ carbon attached to N4 and show the highest LUMO energy (lower propensity to acquire electrons) in this series of THIA derivatives (Table 4). Docking results indicate that the lower reactivity is compensated by their ability to fit into 3CLpro active site by forming strong interactions with the protein (Figures 12B and 13C; Appendix D: Tables 16-19). **THIA-2** and **THIA-8** (but also **THIA-1**) are 3CLpro covalent inhibitors at sub- micromolar concentrations, while preserving potency against other proteases. Therefore, the presence of a sp³ carbon at N4 of the THIA ring decreases electron affinity but increases the right positioning of the R substituents into 3CLpro S₃ subsite. In this way we have compounds (**THIA-1**, **THIA-2** and **THIA-8**) with comparable 3CLpro inhibitory potency with respect to the phenyl-bearing compounds (**THIA-3** and **THIA-4**) but with improved selectivity.

THIA-10/3CLpro complex shows the highest ligand-protein binding energy among all THIA derivatives (still indicating a favourable interaction; Table 5). Notably **THIA-10** have only one binding mode (BM-I)(Appendix D: Table 15) while the other THIA compounds also have “unproductive” binding modes (BM-II to BM-IV), often with more

favourable interaction energies than BM-I (Appendix D: Tables 11-14). This selective adoption of the productive binding mode increases the probability of a successful interaction with the protein and can contribute to entropic advantages. Moreover, this complex showed the shortest distance between the sulfur atom of the catalytic residue Cys¹⁴⁵ and the sulfur atom S1 of the ligand (Table 5). Compared to other docked derivatives, in this complex the proximity of the two sulfur atoms is probably favoured by an interaction between the fluorine atom and the backbone of Asp¹⁸⁷ and Arg¹⁸⁸, which result in a shift in the ligand toward the S₂ subsite bringing the reactive centres together. Finally, the electron-withdrawing effect exerted by the para-fluorine substituent, which increase the electrophilicity of the THIA core and facilitates the nucleophilic attack (Tables 4 and 5), may contribute to the high inhibitory potency of this molecule toward 3CLpro (IC₅₀ = 0.02 μM; Table 2).

Thus, THIA ability to act as 3CLpro covalent inhibitor is probably a result of balance between the ability to interact properly with the catalytic site, orienting the reacting partners, and the chemical reactivity. In agreement with this, **THIA-7**, with the best compromise between ligand-protein interaction energy and the LUMO energy (Table 5) showed the lowest value of IC₅₀ (0.01 μM) towards 3CLpro among the tested THIAs (Table 2). However, **THIA-4**, with a ligand-protein binding energy value lower than **THIA-10** but higher than the other THIAs and a fairly low LUMO energy (lower than **THIA-2** and **THIA-8**, higher than **THIA-7** and **THIA-10**), displayed an inhibitory potency against 3CLpro (IC₅₀ = 0.13) comparable to that of **THIA-2** and **THIA-8**. The electron affinity is increased by the introduction of a para-halogen substituted phenyl ring either at N2 (R'; S₂ subsite; **THIA-10**) or N4 (R; S₃ subsite; **THIA-7**), thus with the effect of speeding up the reaction with the catalytic cysteine.

Consequently, **THIA-7** and **THIA-10** showed increased potency against 3CLpro but decreased selectivity, being also very potent against Papain and Cathepsin L.

The introduction of a *para*-methoxy (*p*-Ome) substituent on the phenyl ring at N4 (R substituent) has a positive effect on 3CLpro inhibitory activity (**THIA-6** vs. **THIA-4**; Table 2) that was elucidated by fitting **THIA-6** into the 3CLpro active site by pharmacophore fitting procedure into docking results obtained for **THIA-4**. The resulting **THIA-6**/3CLpro ligand-protein interaction were calculated and showed that the *p*-Ome group of the R substituent may establish favourable hydrophobic contacts with Pro¹⁶⁸ in the S₃ subsite (Figure 15).

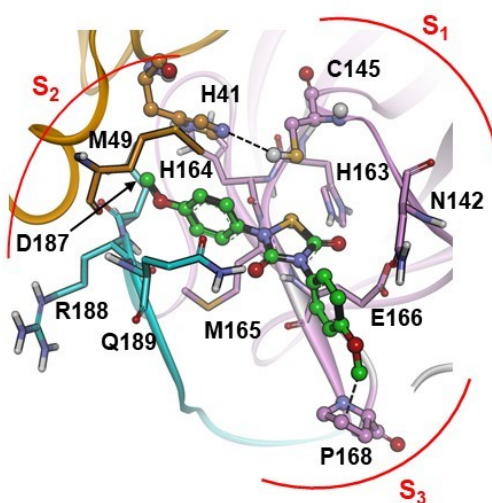


Figure 15. 3CLpro/THIA-6 molecular interaction model obtained by the superimposition of THIA-6 (ball&stick; green) on THIA-4 (stick; black) in the best docked complex with 3CLpro. The protein is displayed as solid ribbons and coloured by protein domains: orange (Domain I: aa 1-101), pink (Domain II: aa 102-184), and cyan (LH-Loop: aa 185-200). The residues involved in the interactions with the ligand are displayed in sticks; the catalytic residues His⁴¹ and Cys¹⁴⁵ as well as Pro¹⁶⁸ are displayed in ball&stick and the putative hydrophobic interaction of THIA-6 with Pro¹⁶⁸ is evidenced by black dashed lines.

However, this substituent can be also accommodated within the active sites of Cathepsin L and Papain, thus not enhancing compound selectivity.

6.2.2.3. Steered Molecular Dynamics Simulations.

To validate docking results and to investigate the dynamic nature of ligand-protein interactions, Steered Molecular Dynamics (SMD) simulations were performed using the final docked complexes as starting conformations. SMD simulations of ligand unbinding from catalytic sites yield key information on molecular recognition and indirectly clarify the binding mechanism [43]. In this case we are mimicking the reverse of the binding event rather than the unbinding process itself, by guiding the ligand away from the catalytic site along a plausible dissociation pathway [44].

The best pathway was the one directed through the S₃ subsite, the widest channel leading out of the catalytic pocket. In all the compounds the interaction energy with the protein diminished from the starting (docked) to the final (unbound) conformation (Figure 16).

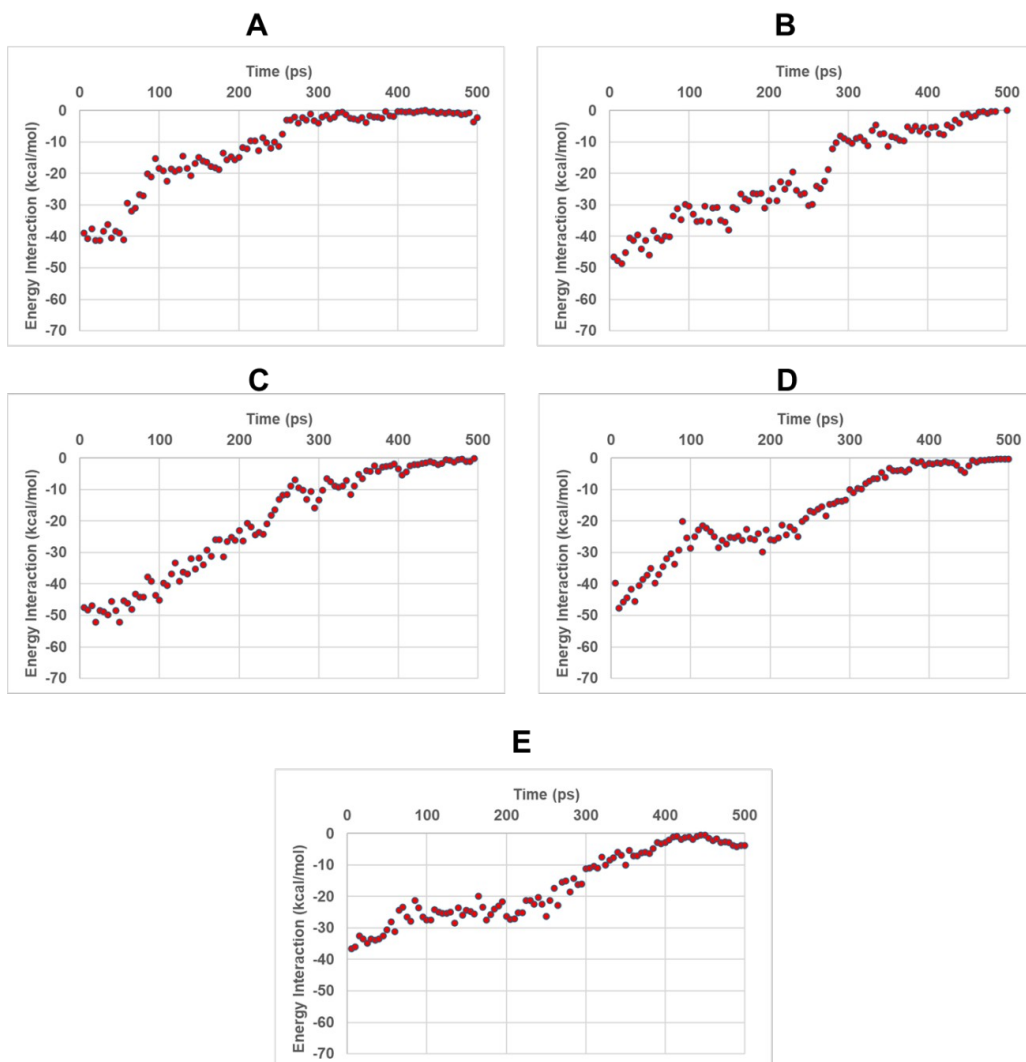


Figure 16. Graphs showing the relationship between time (ps) and interaction energy (kcal/mol) calculated considering all the resulting SMD structures of THIA-10 (A), THIA-2 (B), THIA-4 (C), THIA-7 (D) and THIA-8 (E). Red dots represent interaction energy values calculated every 5 ps.

All compounds exceed the 20 Å distance between S1 and the sulfur atom of Cys145 (Figure 17) at the end of the simulations. This distance is the

threshold beyond that a ligand can be considered unbound from the protein site [45]. Notably, BM-I binding mode complexes resulted the ones with the best interaction energy, supporting the result of docking studies, as reported in Figure 17.

SMD showed that the most effective compound in maintaining the binding mode BM-I is **THIA-2** with interactions energies spanning from -48.66 to -38.07 kcal/mol and the longest retention time (65ps) of this binding mode among all THIA compounds (Figure 16). This good fit with the catalytic site, in line with the docking results, can compensate for the low propensity to acquire electrons of **THIA-2** (highest LUMO energy among THIA derivatives).

Finally, conformational changes and local motions induced by THIA binding were assessed through the estimation of Root Mean Square Fluctuation (RMSF), by aligning each trajectory frame to the starting structure using all C α atoms. A notable flexibility was appointed to the residues that showed value of RMSF greater than 2 Å.

Results (Figures 17-21) showed that **THIA-4** exhibited the highest degree of fluctuation in the upper loop region (residues 44–54), which plays a crucial role in stabilizing both the active conformation of the catalytic residues and ligand binding [26].

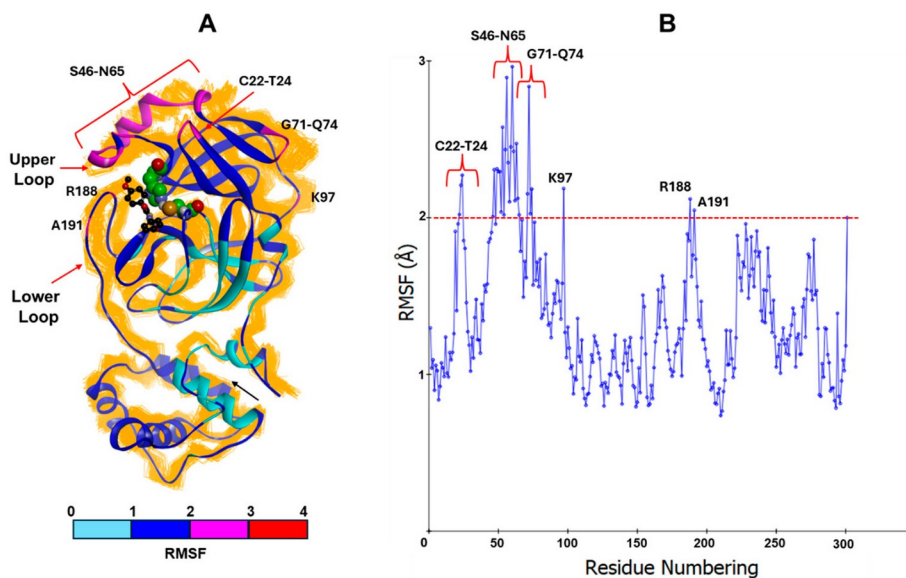


Figure 17. (A) Steered Molecular Dynamics (SMD) results obtained for **THIA-4**. The resulting complexes were superimposed on the starting structure by using $C\alpha$. The starting complex is displayed as solid ribbons and colored according to the calculated RMSF values. The $C\alpha$ of the resulting complexes is displayed as line ribbons and colored in orange. The ligand (black; ball-and-stick model) and the catalytic residues (CPK and green) are colored by atom type (O: red; S: orange; N: blue). **(B)** Mean Square Fluctuation (RMSF) values were calculated considering all the resulting SMD structures. The red dashed line represents the 2 Å RMSF threshold, which is used to highlight the residues with greater flexibility. The residues with an RMSF value > 2 Å are labelled.

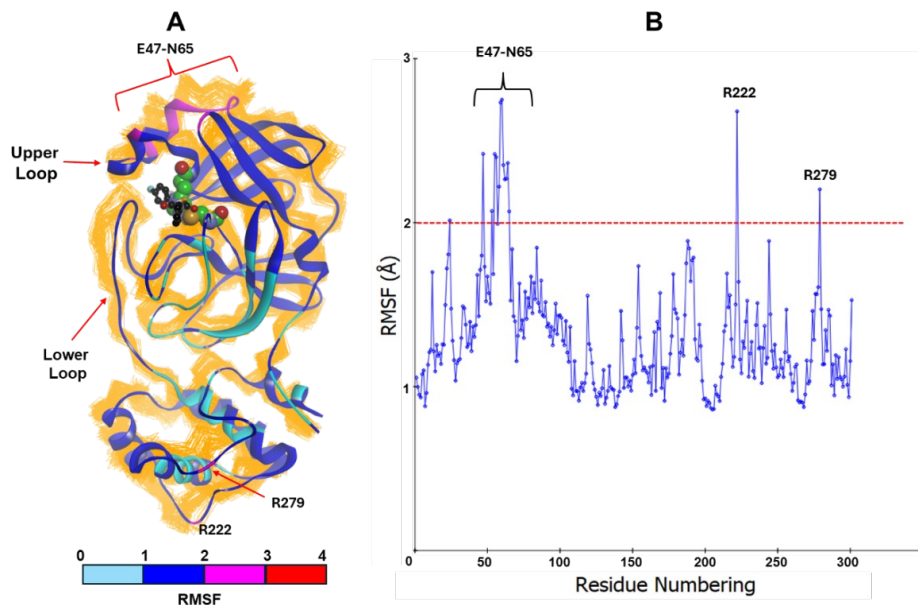


Figure 18. A) Steered Molecular Dynamics (SMD) results obtained for **THIA-10**. The resulting complexes were superimposed on the starting structure by using C α . The starting complex is displayed as solid ribbons and colored according to the calculated RMSF values. The C α of the resulting complexes are displayed as line ribbons and colored in orange. The ligand (black; ball&stick) and the catalytic residues (CPK and green) are colored by atom type (O: red; S: orange; N: blue; F: light blue). B) Mean Square Fluctuation (RMSF) values calculated considering all the resulting SMD structures. The residues with a RMSF value > 2 Å are labelled.

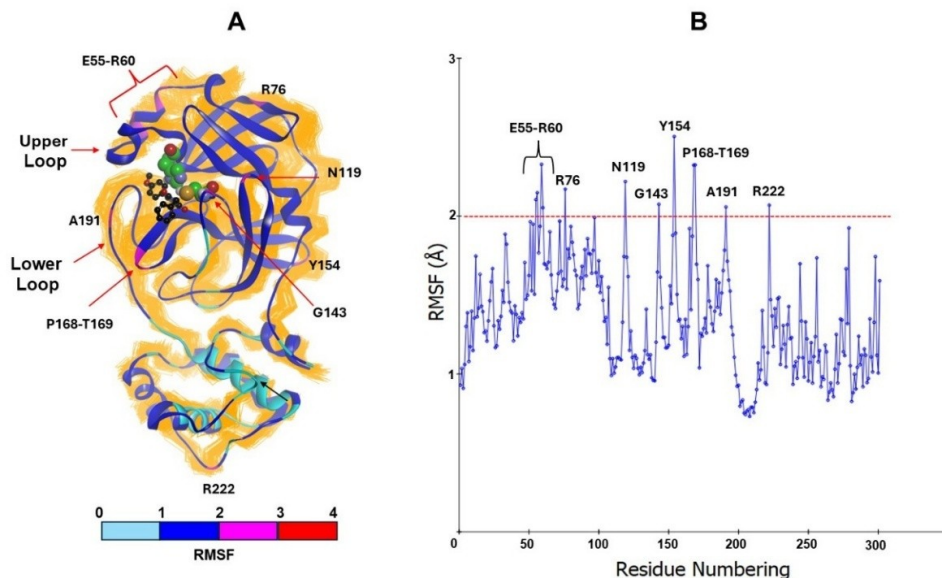


Figure 19. A) Steered Molecular Dynamics (SMD) results obtained for **THIA-2**. The resulting complexes were superimposed on the starting structure by using $C\alpha$. The starting complex is displayed as solid ribbons and colored according to the calculated RMSF values. The $C\alpha$ of the resulting complexes are displayed as line ribbons and colored in orange. The ligand (black; ball&stick) and the catalytic residues (CPK and green) are colored by atom type (O: red; S: orange; N: blue). B) Mean Square Fluctuation (RMSF) values calculated considering all the resulting SMD structures. The residues with a RMSF value $> 2 \text{ \AA}$ are labelled.

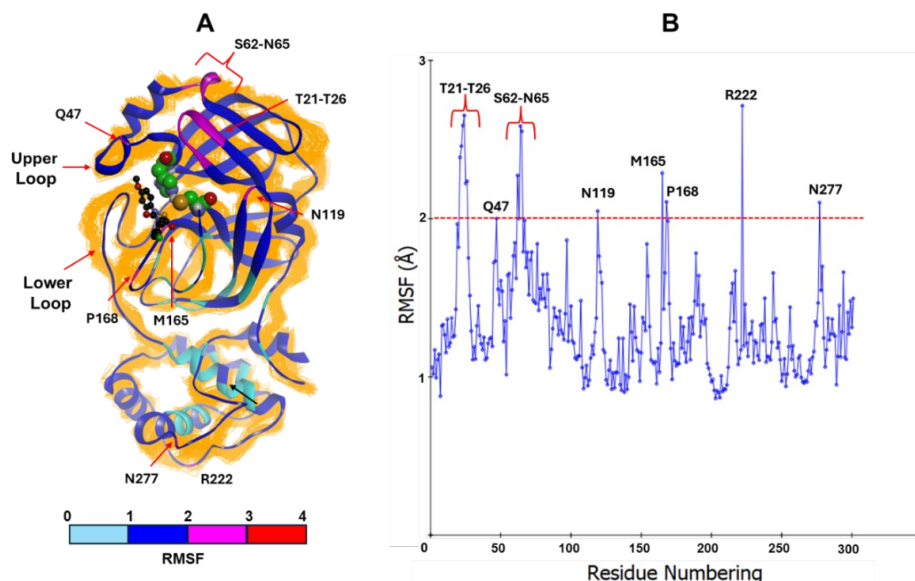


Figure 20. A) Steered Molecular Dynamics (SMD) results obtained for **THIA-7**. The resulting complexes were superimposed on the starting structure by using C α . The starting complex is displayed as solid ribbons and colored according to the calculated RMSF values. The C α of the resulting complexes are displayed as line ribbons and colored in orange. The ligand (black; ball&stick) and the catalytic residues (CPK and green) are colored by atom type (O: red; S: orange; N: blue; Cl: green). B) Mean Square Fluctuation (RMSF) values calculated considering all the resulting SMD structures. The residues with a RMSF value > 2 Å are labelled.

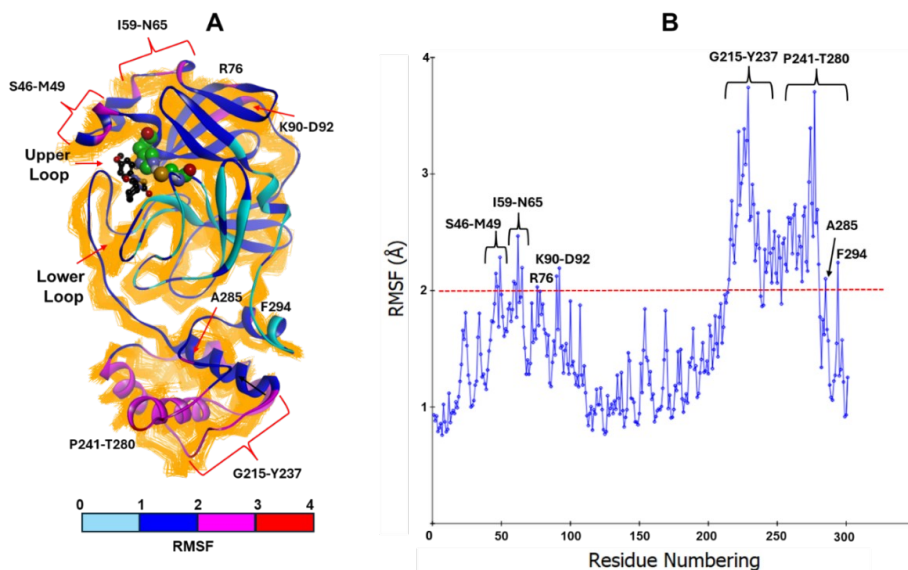


Figure 21. A) Steered Molecular Dynamics (SMD) results obtained for **THIA-8**. The resulting complexes were superimposed on the starting structure by using $\text{C}\alpha$. The starting complex is displayed as solid ribbons and colored according to the calculated RMSF values. The $\text{C}\alpha$ of the resulting complexes are displayed as line ribbons and colored in orange. The ligand (black; ball&stick) and the catalytic residues (CPK and green) are colored by atom type (O: red; S: orange; N: blue). B) Mean Square Fluctuation (RMSF) values calculated considering all the resulting SMD structures. The residues with a RMSF value $> 2 \text{ \AA}$ are labelled.

This result may explain the low inhibitory potency of **THIA-4** against 3CLpro, comparable to that of **THIA-2** and **THIA-8** (Table 2), despite its higher electron-accepting propensity (lower LUMO energy) relative to **THIA-2** and **THIA-8**.

6.2.2.4. Investigation of the Molecular Bases of THIA Selectivity.

To rationalize the selectivity profiles of the THIA derivatives toward the tested cysteine proteases (Table 2), we performed a comparative analysis

between our computational results and the experimentally determined structures of the corresponding enzymes. Initially, we examined the size, shape, and spatial arrangement of the catalytic subsites in 3CLpro, PLpro, Papain, Cathepsin L, and Bromelain (Table 6).

Table 6. Experimentally determined structures of 3CLpro, PLpro, Papain, Cathepsin L, and Bromelain we employed in the structural and bioinformatics analysis.

PDB ID	Organism	Structure	Resolution (Å)
7JKV	SarsCoV 2	Crystal Structure of SARS-CoV-2 main protease in complex with the inhibitor GRL-2420	1.25
7N89	SarsCoV 2	X-ray structure of SARS-CoV-2 main protease C145A mutant in complex with substrate Ac-SAVLQSGF-CONH2	2.00
6WX 4	SarsCoV 2	Crystal structure of the SARS CoV-2 Papain-like protease in complex with peptide inhibitor VIR251	1.66
7JIW	SarsCoV 2	The crystal structure of Papain-Like Protease of SARS CoV-2 in complex with PLP_Snyder530 inhibitor	2.30
3OF8	H.Sapiens	Structural Basis for Reversible and Irreversible Inhibition of Human Cathepsin L by their Respective Dipeptidyl Glyoxal and Diazomethylketone Inhibitors	2.20
6TCX	Papaya	Papain bound to a natural cysteine protease inhibitor from Streptomyces mobaraensis	1.65
6YCG	Ananas comosus	Structure of the bromelain protease from Ananas comosus in complex with the TLCK inhibitor	1.45

Subsequently, the THIA scaffold was positioned within the active sites of PLpro (PDB ID: 6WX4), Papain (PDB ID: 6TCX), Cathepsin L (PDB ID: 3OF8), and Bromelain (PDB ID: 6YCG) by superimposing it onto the co-crystallized peptide ligands through pharmacophore fitting.

Papain, Cathepsin L, and Bromelain share a conserved fold topology and display approximately 48% sequence similarity, which enables structural superimposition based on sequence alignment (Figure 21).

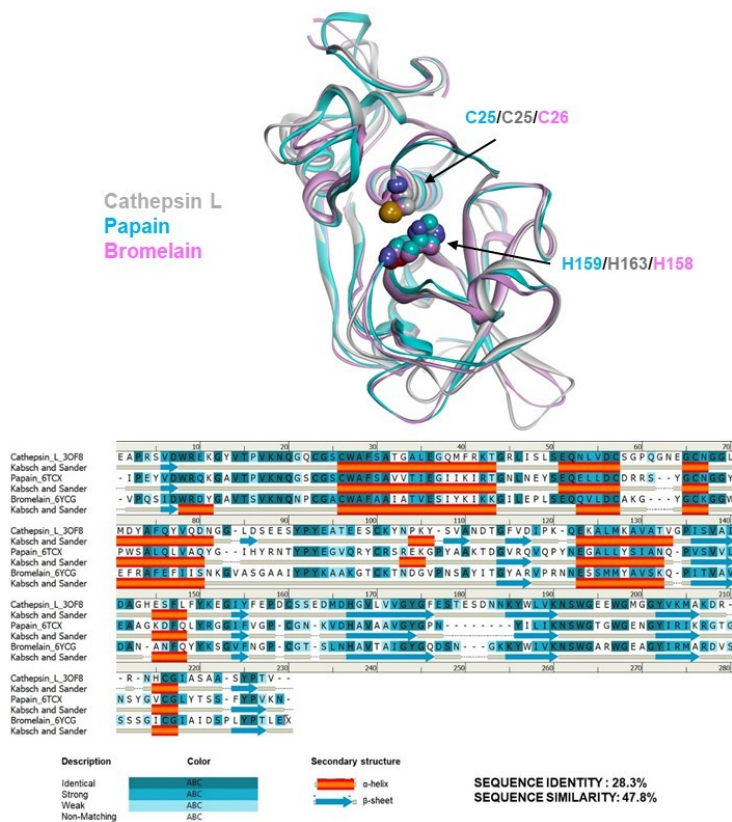


Figure 21. Up: superimposition among x-ray structure of Papain (cyan; PDB ID: 6TCX), x-ray structure of Cathepsin L (grey; PDB ID: 3OF8), and x-ray structure of Bromelain (pink; PDB ID: 6YCG). The proteins are displayed in solid ribbon. The catalytic residues Cys and His are shown in CPK and labelled. Bottom: secondary structure sequence alignment between Papain, Cathepsin L, and Bromelain. The sequence and the secondary structure assignments (Kabsch and Sander) are reported.

In contrast, PLpro and 3CLpro share less than 30% sequence similarity with these proteases and therefore cannot be superimposed based on sequence alignment. However, PLpro preserves a comparable active-site structure and catalytic residue orientation (Figure 22), allowing its inclusion in the structural superimposition by aligning its catalytic triad (Cys¹¹¹, His²⁷², Asp²⁸⁶) with that of Cathepsin L (Cys²⁶, His¹⁶⁴, Asn¹⁸⁸). 3CLpro, which differs significantly in both fold topology and catalytic residue orientation (Cys¹⁴⁵, His⁴¹), was superimposed by aligning the backbone atoms of its co-crystallized peptide ligand (PDB ID: 7JKV) with those of the peptide ligand bound to PLpro (PDB ID: 6WX4).

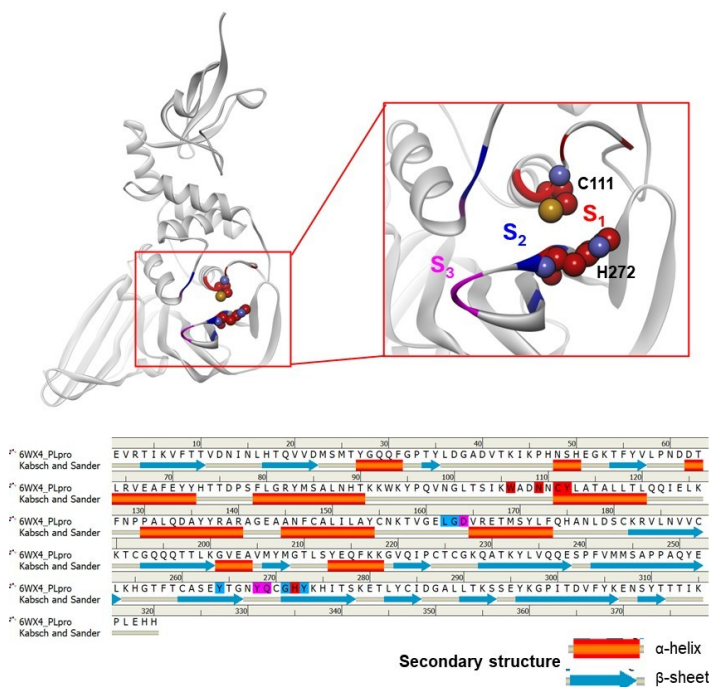


Figure 22. A: X-ray structure of SARS-CoV-2 PLpro (PDB ID: 6WX4). PLpro is displayed in solid ribbon and colored in grey. The residues of S₁, S₂, and S₃ subsites are coloured in red, blue, and magenta, respectively. The catalytic residues Cys and His are displayed in CPK and labelled. The sequence and the secondary structure assignments (Kabsch and Sander) are reported.

This ligand-based structural alignment allowed a comparative assessment of the positioning of the subsites relative to the catalytic cysteine and the scissile peptide bond. While the S₁ and S₂ subsites were approximately superimposable across all proteases, the S₃ subsite displayed a distinct spatial orientation. A schematic representation of these differences is shown in Figure 24.

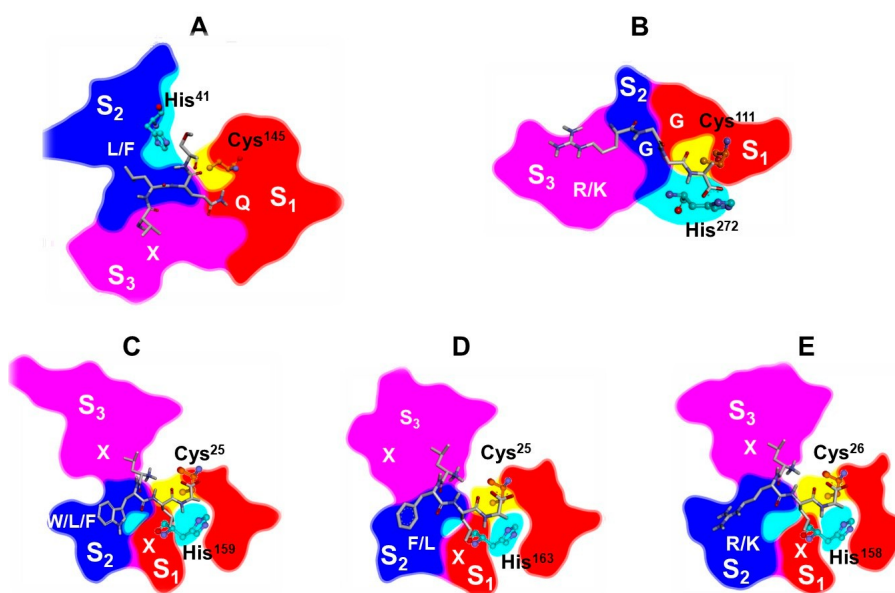


Figure 23. Schematic representations of the catalytic subsites S₁ (red), S₂ (blue), and S₃ (magenta) of five proteases, 3CLpro (A), PLpro (B), Papain (C), Cathepsin L (D), and Bromelain (E), generated from their Connolly surfaces. The spatial arrangement of the subsites was obtained using the superimposition method described in the Materials and Methods section. Representations were created by posterizing the corresponding 3D molecular structures using Inkscape (version 1.3). Catalytic residues Cys and His are shown in ball-and-stick models, labeled and colored in orange and cyan, respectively. The enzyme substrate is shown in stick (C: white; N: blue; O: red) and the consensus substrate sequence is labelled (X = any residue).

Notably, the oxyanion holes of Papain, Cathepsin L, Bromelain, and PLpro exhibit orientations that differ from those observed in 3CLpro. In 3CLpro, the oxyanion hole is located between the S₁ and S₂ subsites, whereas in the other proteases it lies within the S₁ subsite, on the opposite side relative to the catalytic cysteine (Figures 23, 24A–C, and 25A,B).

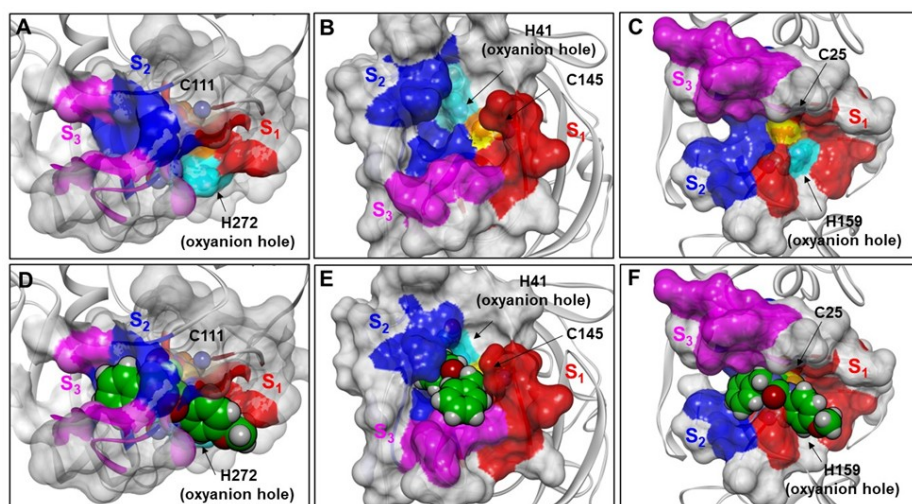


Figure 24. A: x-ray structure of SARS-CoV-2 PLpro (PDB ID: 6WX4). B: x-ray structure of SARSCoV- 2 3CLpro (PDB ID: 7JKV). C: x-ray structure of Papain (PDB ID: 6TCX). D: fitting of THIA-2 in the active site of PLpro (PDB ID: 6WX4). E: 3CLpro-THIA-2 best-docked complex. F: fitting of THIA-2 in the active site of Papain (PDB ID: 6TCX). The proteins are displayed as grey ribbons. The catalytic residues Cys and His are displayed in CPK, labeled and colored in orange and cyan, respectively. The Connolly surfaces of the proteases are displayed in solid. The residues of S₁, S₂, and S₃ subsites are colored red, blue, and magenta, respectively. The solvent-accessible surface (SASA) of the sulfur atom of the catalytic cysteine is displayed and colored in yellow. THIA-2 is displayed in CPK and colored by atom type (C = green, O = red, N = blue, S = yellow).

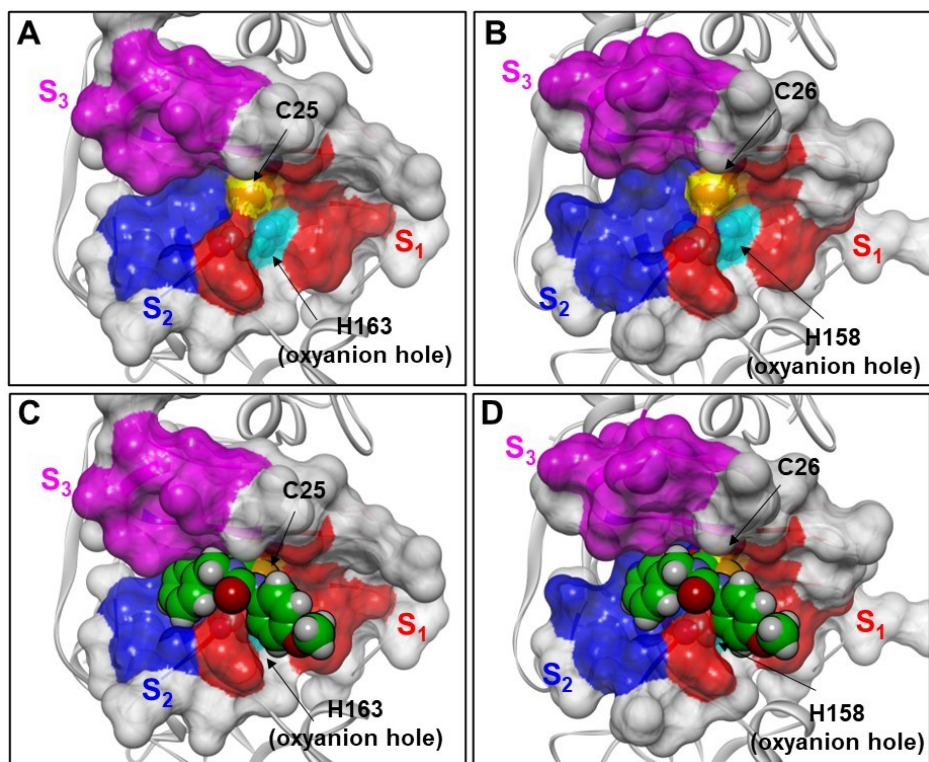


Figure 25. A: x-ray structure of Cathepsin L (PDB ID: 3OF8). B: X-ray structure of Bromelain (PDB ID: 6YCG). C: fitting of THIA-2 in the active site of Cathepsin L (PDB ID: 3OF8). D: fitting of THIA-2 in the active site of Bromelain (PDB ID: 6YCG). The proteins are displayed as grey ribbons. The catalytic residues Cys and His are shown in CPK, labelled and coloured in orange and cyan, respectively. The Connolly surfaces of the proteases are displayed in solid. The residues of S₁, S₂, and S₃ subsites are coloured in red, blue, and magenta, respectively. The solvent accessible surface (SASA) of the sulphur atom of the catalytic cysteine is displayed and coloured in yellow. THIA-2 is shown in CPK and coloured by atom type (C = green, O = red, N = blue, S = yellow).

This implies a distinct spatial arrangement of the ligand-reacting atoms, as confirmed by the analysis of the X-ray structures. Accordingly, when the THIA scaffold is fitted into the active sites of Papain (6TCX), Cathepsin L (3OF8), Bromelain (6YCG), and PLpro (6WX4), both the sulfur (S1) and

nitrogen (N2) atoms occupy the S₁ subsite, while in 3CLpro, the N2 atom is instead positioned within the S₂ subsite (Figures 25D–F and 26C,D).

In PLpro, the catalytic residues reside at the interface between the thumb (aa 62–178) and palm (aa 241–315) domains. The S₁ and S₂ subsites are very narrow (Figure 14A), in agreement with the substrate specificity of this enzyme at positions P1 and P2 [35,46,47]. Accordingly, the THIA derivatives are unable to fit within the PLpro active site (Figure 25D). However, the THIA derivatives could occupy the PLpro S₃ and S₄ subsites, thereby blocking access to the catalytic site without forming a covalent bond with Cys¹¹¹. This behaviour resembles that of GRL0617 (IC₅₀: 2.3 µM) and its analogues, a class of potent SARS-CoV-2 PLpro allosteric inhibitors sharing a structural similarity with the THIA scaffold (Figure 26) [26,48,49].

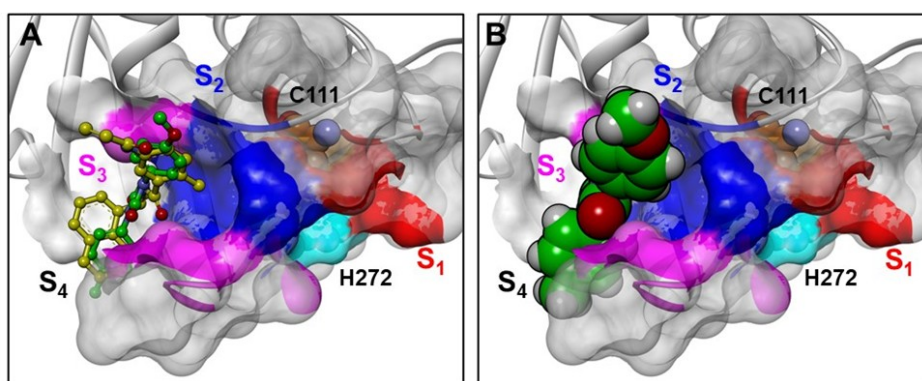


Figure 27. A: structural superimposition of THIA-5 (ball&stick; green) on a GRL0617 derivative in complex with PLpro (yellow; PDB ID: 7JIW); B: van der Waals volume of THIA-5 is displayed. The protein is displayed as grey ribbons. The catalytic residues Cys and His are displayed in CPK, labeled and colored in orange and cyan, respectively. The Connolly surface of the protease is displayed in solid. The residues of S₁, S₂, and S₃ subsites are colored red, blue, and magenta, respectively. THIA-5 is colored by atom type (C = green, O = red, N = blue, S = yellow).

Figure 12A shows the structural superimposition of **THIA-5** with a GRL0617 derivative bound to PLpro (PDB ID: 7JIW). Notably, **THIA-5** appears capable of interacting with most of the residues involved in the binding of the GRL0617 derivative (Figure 27).

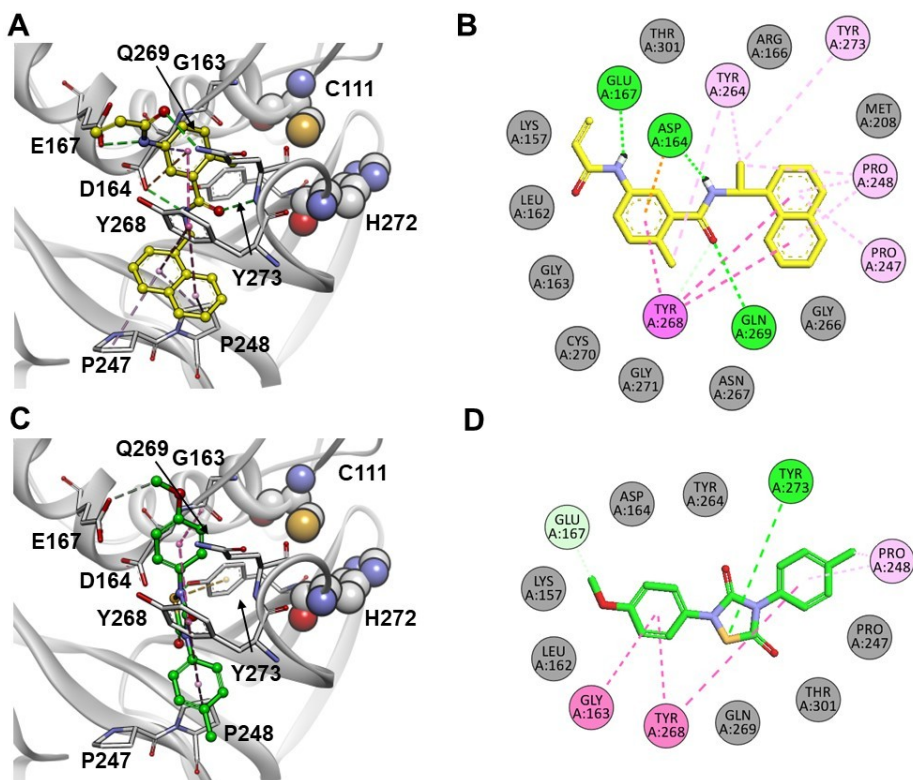


Figure 27. A: x-ray structure of SARS-CoV-2 PLpro in complex with a GRL0617 derivative (yellow; PDB ID: 7JIW). C: fitting of THIA-5 (green) in the allosteric site of PLpro (PDB ID: 7JIW). The two catalytic residues (CPK) and the residues involved in the interactions (stick) are displayed and labelled. Ligands are displayed in ball&stick and coloured by atom type (O = red; N = blue; S = yellow). B and D: 2D representation of ligand-protein interactions (hydrogen bond: green; π - π interaction: magenta; alkyl interaction: pink; CH-O bond: light green).

This accounts for the observed reversible Inhibition of PLpro by THIA compounds (Figure 3C).

On the contrary, the active sites of Papain and Cathepsin L are less constrained (Figures 26C and 27A), allowing accommodation of bulkier substrates, consistent with the broader substrate specificity of these proteases [8,50]. In particular, the catalytic cysteine is more solvent-exposed and accessible than in the other proteases, as evidenced by the solvent-accessible surface area (SASA) analysis (Table 7).

Table 7. Solvent accessible surface (SASA) of the sulphur atom of the catalytic cysteine.

PDB ID	Structure	Solvent accessible surface (SASA) (Å ²)
7JKV	3CLpro	9.668
6WX4	PLpro	6.732
6TCX	Papain	14.055
3OF8	Cathepsin L	14.135
6YCG	Bromelain	15.548

Accordingly, the THIA derivatives, including those bearing a benzyl substituent, such as **THIA-2**, can be accommodated within the active sites of Papain and Cathepsin L (Figures 25F and 26C). As previously discussed, the differences in their inhibitory activity toward these proteases appear to arise primarily from their distinct reactivity (E_{LUMO} ; Table 4). The overall lack of activity of the THIA derivatives against Bromelain is likely attributable to the peculiar amino acid composition of its S_2 subsite. Indeed, while the S_1 subsite is strictly conserved, the S_2 subsite shows pronounced differences among Papain, Cathepsin L, and Bromelain (Figure 28 and Table 3).

Bromelain is polar and negatively charged, due to the presence of Asp²⁰⁹ and Glu⁶⁸. This feature aligns with the enzyme's ability to recognize substrates bearing positively charged residues at the P₂ position [8,50]. The unique ability of **THIA-7** to inhibit Bromelain ($IC_{50} = 2.0 \pm 0.1 \mu M$) may arise from the capacity of the chlorine atom at the para position of the N4-phenyl ring (R substituent) to form a halogen bond with the carboxylate group of Glu⁶⁸ located in the S₂ subsite of the enzyme (Figure 30A), as indicated by the nonbonded geometric interaction analysis [42].

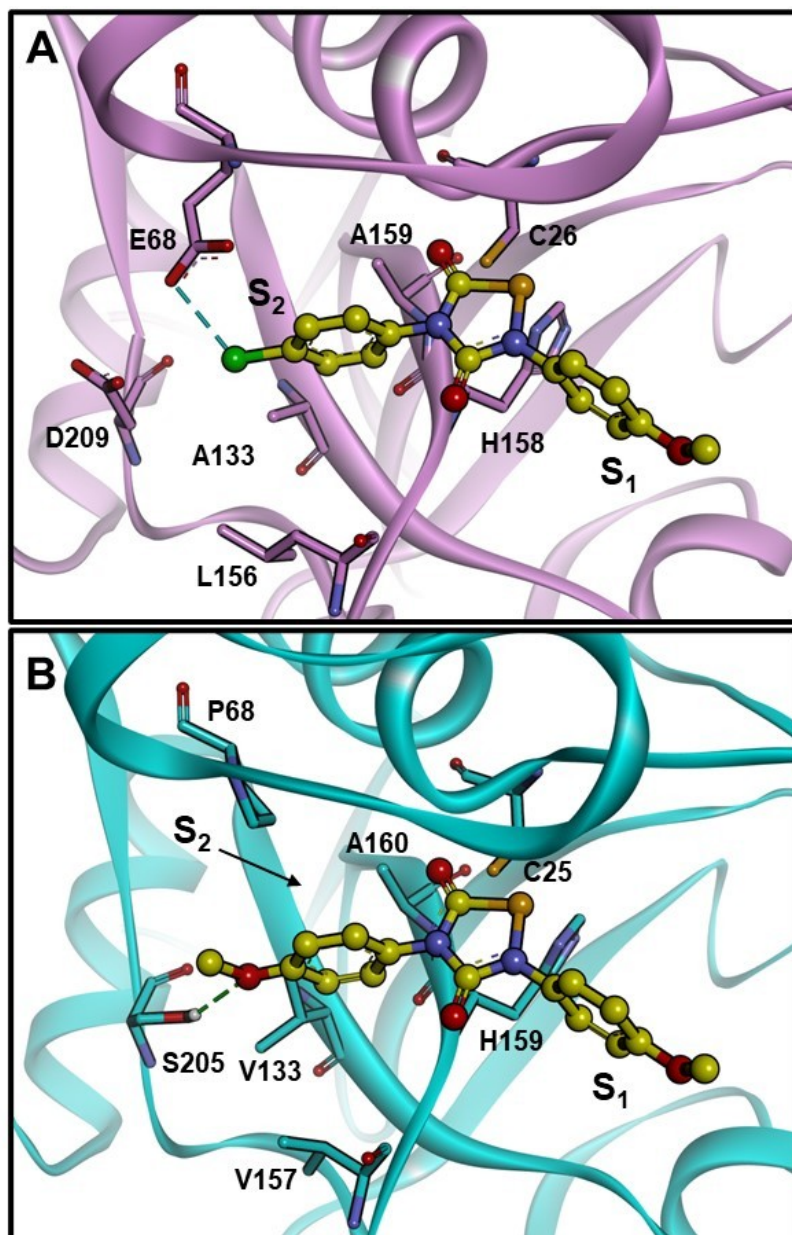


Figure 29. A: fitting of THIA-7 (yellow) in the active site of Bromelain (PDB ID:6YCG; pink). The hydrogen bond is displayed as a green dashed line. B: fitting of THIA-6 (yellow) in the active site of Papain (cyan). The halogen bond is displayed as a black dashed line. Papain and Bromelain are displayed as ribbons. Ligands are displayed in ball&stick and coloured by atom type (C= yellow; O=red; N=blue; S=orange; Cl=green). The catalytic residues and the residues of the S₂ subsite are displayed in stick and labelled.

The binding hypothesis presented in this study may also explain the generally higher inhibitory activity of all THIA derivatives toward Papain compared with Cathepsin L (Table 2). Indeed, the S₂ subsite of Papain contains less bulky residues than that of Cathepsin L (Papain: Pro⁶⁸, Val¹⁵⁷; Cathepsin L: Met⁷⁰, Met¹⁶¹; Table 3), allowing a better accommodation of the R substituent (Figure 28). In particular, the methoxy group of **THIA-6** may form a hydrogen bond with Ser²⁰⁵, a residue unique to Papain located in the S₂ subsite (Figures 28 and 29B; Table 3), consistent with the highest inhibitory activity of **THIA-6** toward Papain among the tested compounds (Table 2).

6.2.2.5. Covalent Docking Simulation.

To obtain molecular models of the covalent complexes of 3CLpro, Papain, and Cathepsin L with **THIA-3**, **-4**, **-6**, and **-8**, we explored the accommodation of the catalytic cysteine–THIA adduct within the catalytic pockets of the different proteases through covalent docking simulations performed with AutoDock4 [51]. No specific ligand–protein complex was used as a starting point; instead, the adduct was randomly generated to identify the most favourable complex formed after covalent bond formation. In addition, the conformation of the adduct generated upon THIA ring opening was allowed to adjust in order to achieve the best structural complementarity with the enzyme binding site. Since **THIA-6**, **THIA-7**, and **THIA-10** display stronger inhibitory effects (IC₅₀ < 0.05 μM) compared to the other THIA derivatives with similar IC₅₀ values, the analysis focused on comparing the covalent binding of 3CLpro with **THIA-6** and **THIA-4**, which differ only by the presence of a p-methoxyphenyl group in **THIA-6** instead of a phenyl group in **THIA-4** at

the N4 substituent (R). Covalent docking revealed that, upon reaction with the catalytic Cys¹⁴⁵, both compounds occupy the S₁–S₂–S₄ subsites and form a hydrogen bond with the N^δ atom of His⁴¹. In this binding mode, the N4 substituent (R) fits within the S₁ subsite (Phe¹⁴⁰, Leu¹⁴¹, Asn¹⁴², Met¹⁶⁵, Glu¹⁶⁶), whereas the N2 substituent (R'), carrying a p-methoxyphenyl group, extends deeply into the hydrophobic S₂ pocket (Met⁴⁹, Asp¹⁸⁷, Arg¹⁸⁸). The p-methoxy group further establishes additional interactions toward the S₄ subsite (Met¹⁶⁵, Glu¹⁶⁶, Gln¹⁸⁹, and Gln¹⁹²) (Figure 30A).

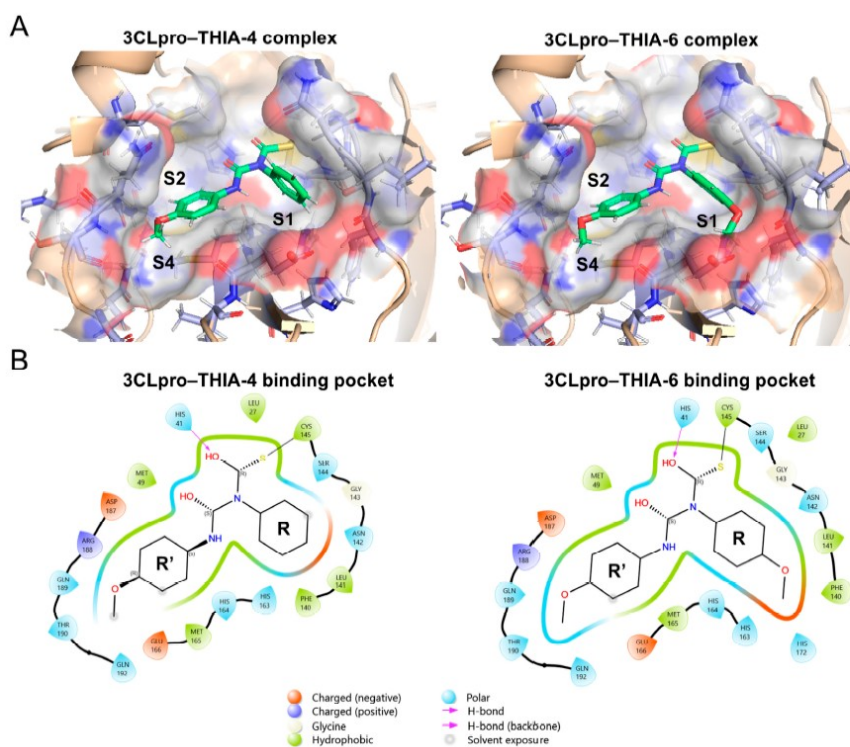


Figure 30. (A) 3D molecular visualization of the 3CLpro–THIA-4 (Lowest binding energy of –6.48 kcal/mol) and 3CLpro–THIA-6 (Lowest binding energy of –6.43 kcal/mol) complexes. Ligands are shown in green within the active site of 3CLpro. Subsites S1, S2, and S4 are highlighted, illustrating the spatial arrangement of the binding pocket and indicating specific interactions with each ligand. (B) 2D interaction diagrams of the 3CLpro binding pockets for THIA-4 and THIA-6.

Although both compounds adopt a similar orientation within the 3CLpro binding pocket, the *p*-methoxy substituent at the N4 position of **THIA-6** allows additional interactions within the S₁ subsite, including contacts with His¹⁷². This deeper interaction with the S₁ subsite enabled a better overall fit across the S₂–S₄ subsites and resulted in slightly greater solvent accessibility compared to **THIA-4** (Figure 30B). It is noteworthy that the covalent docking results differ from the non-covalent complex shown in Figure 14 in the positioning of the N4 substituents (R groups), which occupy the S₃ subsite in all covalent complexes. Next, we evaluated **THIA-4**, **THIA-6**, and **THIA-8** through covalent docking simulations with the Papain adducts. These compounds were selected because, despite differing only in the N4 substituent, **THIA-6** shows a sevenfold higher inhibitory potency than **THIA-4** and a 240-fold increase compared to **THIA-8**. Covalent docking simulations further revealed that the N4 substituent (R) generally occupies the S₂ subsite, while the N2 substituent (R') is positioned within the S₃ subsite, except for the **THIA-8** complex, in which this orientation is reversed (Figure 31A).

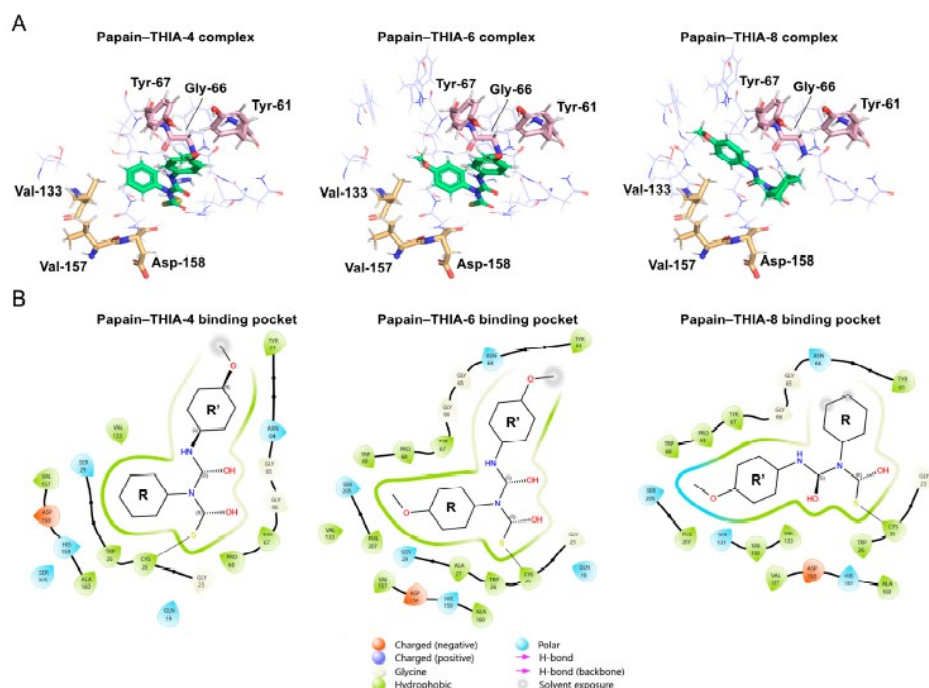


Figure 31. (A) Representation of 3D molecular structure of Papain bound to **THIA-4** (Lowest binding energy of -6.28 kcal/mol), **THIA-6** (Lowest binding energy of -6.19 kcal/mol), and **THIA-8** (Lowest binding energy of -5.94 kcal/mol). Key residues involved in ligand binding corresponding to S2-S3 subsites are labelled. (B) 2D interaction diagrams of Papain–THIA complexes. Each diagram illustrates the binding pocket and interaction profile for **THIA-4**, **THIA-6**, and **THIA-8**. Hydrogen bonds, hydrophobic contacts, and solvent exposure are indicated using standardized symbols. Amino acid residues and ligand structures are annotated to show the molecular basis of binding specificity.

The binding site of Papain should not be regarded as consisting of distinct S₂ (Val¹³³, Val¹⁵⁷, Asp¹⁵⁸) and S₃ (Gly⁶⁶, Tyr⁶⁷) subsites, but rather as a single large hydrophobic pocket encompassing both regions [50]. It was observed that **THIA-4** and **THIA-6** exhibited similar interactions involving the *p*-methoxyphenyl group of the N2 substituent within the S₃ subsite and demonstrated enhanced occupancy of the S₂ subsite core,

owing to the positioning of the N4 phenyl substituent near the S₂ residues. **THIA-6**, in particular, exhibited a greater number of interactions within the S₂ subsite compared to **THIA-4**, due to the presence of the O-methyl group, which enables additional contacts with Phe²⁹⁷ and Trp⁶⁹. In contrast, the *p*-methoxyphenyl group at the N2 position of **THIA-8** was deeply inserted into the S₂ subsite. In contrast, the N4 cyclohexyl substituent extended toward the S₃ region, remaining distant from Gly⁶⁶ and Tyr⁶⁷, and instead exposed to the solvent (Figure 32B). Considering the structural features of both the receptor and the ligands, it can be hypothesized that the Papain binding pocket is not deep enough to accommodate the bulky cyclohexyl group of **THIA-8**. This steric limitation likely leads to an unfavourable geometric alignment of the warhead within the active site, redirecting the overall binding orientation and forcing the N2 substituent into the S₂ subsite while leaving the warhead poorly positioned in the S₃ pocket.

Cathepsin L features a deep and narrow hydrophobic S₂ subsite formed by the side chains of Leu⁶⁹, Met⁷⁰, Ala¹³⁵, and Met¹⁶¹, along with the backbone atoms of Met¹⁶¹, Asp¹⁶², His¹⁶³, and Gly¹⁶⁴, underscoring their crucial role in substrate specificity [52], as exemplified by the CatL–E64d complex (PDB ID: 7ZXA). Moreover, the confined architecture of the S₂ pocket is further reinforced by Asp¹⁶⁰, Met¹⁶¹, Asp¹⁶², and Ala²¹⁴, which together create a continuous steric barrier that limits the extension of the S₂ subsite [53]. In this context, the bulky cyclohexyl moiety may fit into the S₂ subsite, similarly to the triazine ligand, which targets the hydrophobic S₂ pocket [54] (PDB ID: 4AXM) (Figure 32A).

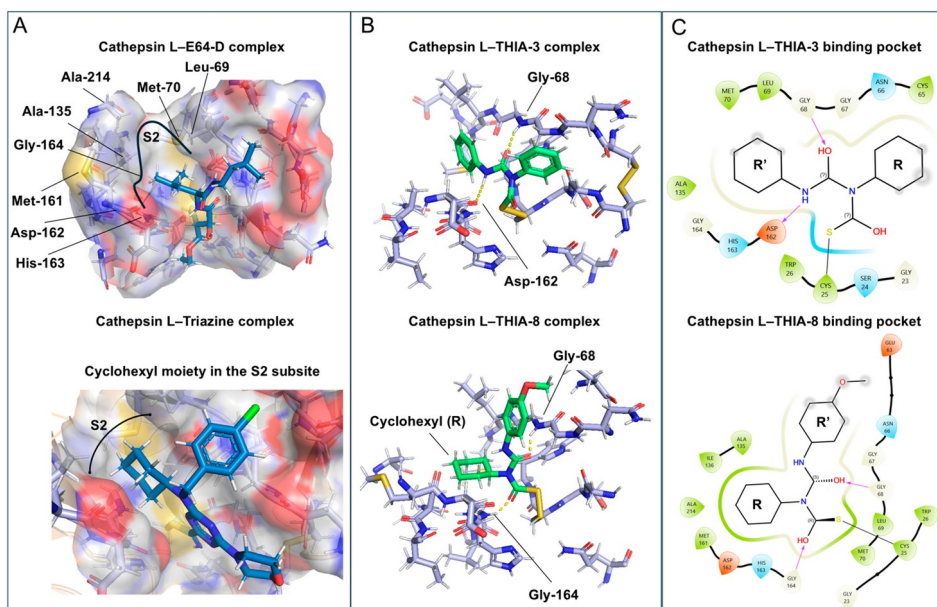


Figure 32. (A) Structural representation of the Cathepsin L–E64–D complex (PDB ID: 7zxa), highlighting the residues of the S₂ subsite and The Cyclohexyl moiety of Triazine occupying this pocket (PDB ID: 4axm). (B) Structural representation of Cathepsin L–**THIA-3** complex (Lowest binding energy of –5.09 kcal/mol) and Cathepsin L–**THIA-8** complex (Lowest binding energy of –6.18 kcal/mol), showing R and R' positioning and respective H-bonds (lines yellow), in which the Cyclohexyl group from **THIA-8** was positioned within the S₂ subsite, similar to Triazine. (C) 2D interaction diagrams of Cathepsin L with **THIA-3** and **THIA-8**. Hydrogen bonds, hydrophobic contacts, and other molecular interactions are illustrated. Ligand structures and binding residues are annotated to highlight key interactions within the binding pocket.

The distinct hydrophobic character of the S₂ and S₃ subsites (Gly⁶¹, Glu⁶³, Gly⁶⁷, Gly⁶⁸) favours the accommodation of hydrophobic side chains from different ligands, while key hydrogen-bonding hotspots across the Cathepsin L binding pocket include Asp¹⁶² and Gly⁶⁸ [55]. The calculated Cathepsin L adducts with **THIA-3** and **THIA-8** revealed that, in the **THIA-3**/Cathepsin L complex, the N₄ phenyl substituent is positioned within the S₁ subsite, whereas the N₂ phenyl group is oriented toward the

entrance of the S₂ subsite, forming hydrogen bonds with Gly⁶⁸ and Asp¹⁶². In the **THIA-8**/Cathepsin L complex, the N4 substituent (cyclohexyl; R) is oriented toward the S₂ subsite, forming hydrogen bonds with Gly⁶⁸ and Gly¹⁶⁴, while the N2 substituent remains solvent-exposed (Figure 33B). Although the cyclohexyl moiety fits well into the S₂ pocket of Cathepsin L, the p-methoxyphenyl group at the N2 position of **THIA-8** displays suboptimal accommodation within the S₃ subsite. This may be due to the relatively bulky volume of the moiety combined with the narrow geometry of the Cathepsin L S₃ pocket, leading to a steric clash with the pocket wall, particularly with Glu63, and resulting in partial solvent exposure. In contrast, the smaller substituent composition of THIA-3 allows for more extensive interactions within the S1 subsite and provides a better fit within the S₃ subsite (Figure 32C).

It is noteworthy that the bulky side chains of Met⁷⁰ and Met¹⁶¹ within the S₂ subsite (Table 3) restrict its occupancy, which may explain the generally weaker inhibitory activity of THIA derivatives toward Papain compared with 3CLpro and Cathepsin L (Table 2). This effect is particularly evident for **THIA-1**, **THIA-2**, and **THIA-8**, which bear bulkier N4 substituents compared with the other THIA.

6.3. Conclusions.

These studies contributed to clarifying the molecular basis of the viral inhibition mechanism of 1,2,4-thiadiazolidine-3,5-dione (THIA) derivatives on the 3CLpro enzyme and their selectivity towards other proteases. Computational studies supported the experimental findings and provided a rationale for the observed selectivity profiles. The inhibition mechanism proceeds through a two-step irreversible bimolecular process,

which becomes reversible in the presence of reducing agents. The inhibitory activity appears to depend not only on the electrophilic reactivity of the THIA core, which enables covalent binding to the catalytic cysteine, but also on the conformational properties of the scaffold and on the capacity of the R and R' substituents to orient the compound toward specific protease subsites, thereby promoting favourable hydrophobic and electrostatic interactions. The most potent inhibitors (**THIA-6**, **-7**, and **-10**) exhibited nanomolar IC₅₀ values against 3CLpro, whereas **THIA-1**, **-2**, and **-8** displayed enhanced selectivity. The presence of a benzyl or cyclohexyl group at the N4 nitrogen favoured selectivity toward 3CLpro, whereas para-halogenated phenyl substituents enhanced the overall inhibitory potency. In the case of Cathepsin L, the reduced inhibitory activity can be attributed to steric hindrance within the S₂ subsite. The observed reversible inhibition of PLpro is likely attributable to the inaccessibility of the S₁ and S₂ subsites. The strong structural overlap between the non-covalent inhibitor GRL0617 and **THIA-5** in complex with PLpro supports the hypothesis that both compounds bind to the same allosteric site. The dual functionality of THIA derivatives, as both H₂S-releasing agents and covalent inhibitors, supports their potential as multifunctional therapeutic candidates targeting cysteine proteases implicated in viral replication and inflammatory pathways.

6.5. References.

1. Martinez, A., et al. "SAR and 3D-QSAR Studies on Thiadiazolidinone Derivatives: Exploration of Structural Requirements for Glycogen Synthase Kinase 3 Inhibitors." *Journal of Medicinal Chemistry*, vol. 48, 2005, pp. 7103-12.
2. Turner, E.M., et al. "Small Molecule Inhibitors of Regulator of G Protein Signalling (RGS) Proteins." *ACS Medicinal Chemistry Letters*, vol. 3, no. 2, 2012, pp. 146-50.
3. Severino, B., et al. "1,2,4-Thiadiazolidin-3,5-diones as Novel Hydrogen Sulfide Donors." *European Journal of Medicinal Chemistry*, vol. 143, 2018, pp. 1677-86.
4. Blazer, L.L., et al. "A Nanomolar-Potency Small Molecule Inhibitor of Regulator of G-Protein Signaling Proteins." *Biochemistry*, vol. 50, no. 15, 2011, pp. 3181-92.
5. Schechter, I., and I.A. Berger. "On the Size of the Active Site in Proteases. I. Papain." *Biochemical and Biophysical Research Communications*, vol. 425, no. 3, 2012, pp. 497-502.
6. St Hilaire, P.M., et al. "Fluorescence-Quenched Solid Phase Combinatorial Libraries in the Characterization of Cysteine Protease Substrate Specificity." *Journal of Combinatorial Chemistry*, vol. 1, no. 6, 1999, pp. 509-23.
7. Turk, D., et al. "Revised Definition of Substrate Binding Sites of Papain-Like Cysteine Proteases." *Biological Chemistry*, vol. 379, no. 2, 1998, pp. 137-47.
8. Choe, Y., et al. "Substrate Profiling of Cysteine Proteases Using a Combinatorial Peptide Library Identifies Functionally Unique Specificities." *Journal of Biological Chemistry*, vol. 281, no. 18, 2006, pp. 12824-32.
9. Gosalia, D.N., et al. "High Throughput Substrate Specificity Profiling of Serine and Cysteine Proteases Using Solution-Phase Fluorogenic Peptide Microarrays." *Molecular & Cellular Proteomics*, vol. 4, no. 5, 2005, pp. 626-36.
10. Alves, L.C., et al. "S1 Subsite Specificity of a Recombinant Cysteine Proteinase, CPB, of *Leishmania mexicana* Compared with Cruzain, Human Cathepsin L and Papain Using Substrates Containing Non-Natural Basic Amino Acids." *European Journal of Biochemistry*, vol. 268, no. 5, 2001, pp. 1206-12.
11. McConathy, Jonathan, and Michael J. Owens. "Stereochemistry in drug action." *Primary care companion to the Journal of clinical psychiatry* 5.2 (2003): 70. Fujishima, A., et al. "The Crystal Structure of

- Human Cathepsin L Complexed with E-64." *FEBS Letters*, vol. 407, no. 1, 1997, pp. 47-50.
12. Varughese, K.I., et al. "Crystal Structure of a Papain-E-64 Complex." *Biochemistry*, vol. 28, no. 3, 1989, pp. 1330-32.
 13. Tsuge, H., et al. "Inhibition Mechanism of Cathepsin L-Specific Inhibitors Based on the Crystal Structure of Papain-CLIK148 Complex." *Biochemical and Biophysical Research Communications*, vol. 266, no. 2, 1999, pp. 411-16.
 14. Gour-Salin, B.J., et al. "E64 [trans-Epoxy succinyl-L-leucylamido-(4-guanidino)butane] Analogues as Inhibitors of Cysteine Proteinases: Investigation of S2 Subsite Interactions." *Biochemical Journal*, vol. 299, no. 2, 1994, pp. 389-92.
 15. Coulombe, R., et al. "Structure of Human Procathepsin L Reveals the Molecular Basis of Inhibition by the Prosegment." *The EMBO Journal*, vol. 15, no. 20, 1996, pp. 5492-503.
 16. Okamoto, D.N., et al. "Increase of SARS-CoV 3CL Peptidase Activity Due to Macromolecular Crowding Effects in the Milieu Composition." *Biological Chemistry*, vol. 391, no. 12, 2010, pp. 1461-68.
 17. Zhang, L., et al. "Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α -Ketoamide Inhibitors." *Science*, vol. 368, no. 6489, 2020, pp. 409-12.
 18. Konno, S., et al. "3CL Protease Inhibitors with an Electrophilic Arylketone Moiety as Anti-SARS-CoV-2 Agents." *Journal of Medicinal Chemistry*, vol. 65, no. 4, 2022, pp. 2926-39.
 19. Kitamura, N., et al. "Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors." *Journal of Medicinal Chemistry*, vol. 65, no. 4, 2022, pp. 2848-65.
 20. Kneller, D.W., et al. "Room-Temperature X-Ray Crystallography Reveals the Oxidation and Reactivity of Cysteine Residues in SARS-CoV-2 3CL Mpro: Insights into Enzyme Mechanism and Drug Design." *IUCrJ*, vol. 7, no. 6, 2020, pp. 1028-35.
 21. Kneller, D.W., et al. "Structural Plasticity of SARS-CoV-2 3CL Mpro Active Site Cavity Revealed by Room Temperature X-Ray Crystallography." *Nature Communications*, vol. 11, 2020, article 3202.
 22. Fuzo, C. A., et al. "Celastrol: A Lead Compound that Inhibits SARS-CoV-2 Replication, the Activity of Viral and Human Cysteine Proteases, and Virus-Induced IL-6 Secretion." *Drug Development Research*, vol. 83, no. 8, 2022, pp. 1823-40, doi:10.1002/ddr.21982.
 23. Ratia, K., et al. "A Noncovalent Class of Papain-Like Protease/Deubiquitinase Inhibitors Blocks SARS Virus Replication."

Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 42, 2008, pp. 16119-24.

24. Chou, C.-Y., et al. "Structural Basis for Catalysis and Ubiquitin Recognition by the Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease." *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 2, 2014, pp. 572-81.
25. Rut, W., et al. "Activity Profiling and Crystal Structures of Inhibitor-Bound SARS-CoV-2 Papain-Like Protease: A Framework for Anti-COVID-19 Drug Design." *Science Advances*, vol. 6, no. 42, 2020, article eabd4596.
26. Baez-Santos, Y.M., et al. "The SARS-Coronavirus Papain-Like Protease: Structure, Function and Inhibition by Designed Antiviral Compounds." *Antiviral Research*, vol. 115, 2015, pp. 21-38.
27. Lv, Z., et al. "Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development." *Frontiers in Chemistry*, vol. 9, 2022, article 819165.
28. Juliano, Maria Aparecida, et al. "1,2,4-Thiadiazolidin-3,5-Diones as Inhibitors of Cysteine Proteases." *Molecules*, vol. 30, no. 19, 2025, article 3896.
29. Portaro, F.C., et al. "Probing the Specificity of Cysteine Proteinases at Subsites Remote from the Active Site: Analysis of P4, P3, P2' and P3' Variations in Extended Substrates." *Biochemical Journal*, vol. 347, no. 1, 2000, pp. 123-29.
30. Abeles, R.H., and T.A. Alston. "Enzyme Inhibition by Fluoro Compounds." *Journal of Biological Chemistry*, vol. 265, no. 28, 1990, pp. 16705-08.
31. Bohm, H.J., et al. "Fluorine in Medicinal Chemistry." *ChemBioChem*, vol. 5, no. 5, 2004, pp. 637-43.
32. Olsen, J.A., et al. "A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an Enzyme Active Site: Evidence for C-F...C=O Interactions." *Angewandte Chemie International Edition*, vol. 42, no. 22, 2003, pp. 2507-11.
33. Azarkan, M., et al. "Structures of the Free and Inhibitors-Bound Forms of Bromelain and Ananain from *Ananas comosus* Stem and in Vitro Study of Their Cytotoxicity." *Scientific Reports*, vol. 10, 2020, article 19570.
34. Barrett, A.J., et al. "L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its Analogues as Inhibitors of Cysteine Proteinases Including Cathepsins B, H and L." *Biochemical Journal*, vol. 201, no. 1, 1982, pp. 189-98.

35. Osipiuk, J., et al. "Structure of Papain-Like Protease from SARS-CoV-2 and its Complexes with Noncovalent Inhibitors." *Nature Communications*, vol. 12, 2021, article 743.
36. Ratia, K., et al. "A Noncovalent Class of Papain-Like Protease/Deubiquitinase Inhibitors Blocks SARS Virus Replication." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 42, 2008, pp. 16119-24.
37. Nicholls, A., and B.A. Honig. "A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation." *Journal of Computational Chemistry*, vol. 12, no. 4, 1991, pp. 435-45.
38. Cossi, Maurizio, et al. "Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model." *Journal of computational chemistry* 24.6 (2003): 669-681. Ren, P., et al. "Discovery and Mechanism Study of SARS-CoV-2 3C-like Protease Inhibitors with a New Reactive Group." *Journal of Medicinal Chemistry*, vol. 66, no. 17, 2023, pp. 12266-83.
39. Ren P.; Li, H.; Nie, T.; Jian, X.; Yu C; Jian Li, J.; Haixia Su, H.; Zhang, X; Li, S.; Yang, X.; Peng, C.; Yin, Y.; Zhang, L; Xu, Y; Liu, H.; Fang Bai, F. Discovery and Mechanism Study of SARS-CoV-2 3C-like Protease Inhibitors with a New Reactive Group. *J Med Chem.* **2023**, 66 (17), 12266-12283. doi: 10.1021/acs.jmedchem.3c00818
40. Venkatachalam, C.M., et al. "LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites." *Journal of Molecular Graphics and Modelling*, vol. 21, no. 4, 2003, pp. 289-307.
41. Swiderek, K., and V. Moliner. "Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 Mpro by QM/MM Computational Methods." *Chemical Science*, vol. 11, no. 39, 2020, pp. 10626-30.
42. Bissantz, C., et al. "A Medicinal Chemist's Guide to Molecular Interactions." *Journal of Medicinal Chemistry*, vol. 53, no. 14, 2010, pp. 5061-84.
43. Okimoto, N., et al. "Evaluation of Protein-Ligand Affinity Prediction Using Steered Molecular Dynamics Simulations." *Journal of Biomolecular Structure and Dynamics*, vol. 35, no. 15, 2017, pp. 3221-31.
44. Hu, X., et al. "Steered Molecular Dynamics for Studying Ligand Unbinding of Ecdysone Receptor." *Journal of Biomolecular Structure and Dynamics*, vol. 36, no. 14, 2018, pp. 3819-28.

45. Patel, J.S., et al. "Steered Molecular Dynamics Simulations for Studying Protein-Ligand Interaction in Cyclin-Dependent Kinase 5." *Journal of Chemical Information and Modeling*, vol. 54, no. 2, 2014, pp. 470-80.
46. Ullrich, S., and C. Nitsche. "SARS-CoV-2 Papain-Like Protease: Structure, Function and Inhibition." *ChemBioChem*, vol. 23, no. 19, 2022, article e202200327.
47. Fu, Z., et al. "The Complex Structure of GRL0617 and SARS-CoV-2 PLpro Reveals a Hot Spot for Antiviral Drug Discovery." *Nature Communications*, vol. 12, 2021, article 488.
48. Shao, Q., et al. "Unraveling the Catalytic Mechanism of SARS-CoV-2 Papain-Like Protease with Allosteric Modulation of C270." *Journal of Physical Chemistry B*, vol. 125, no. 48, 2021, pp. 13206–15.
49. Biniossek, M.L., et al. "Proteomic Identification of Protease Cleavage Sites Characterizes Prime and Non-Prime Specificity of Cysteine Cathepsins B, L, and S." *Journal of Proteome Research*, vol. 10, no. 12, 2011, pp. 5363-73.
50. Kim, M.J., et al. "Crystal Structure of Papain-E64-c Complex. Binding Diversity of E64-c to Papain S2 and S3 Subsites." *Biochemical Journal*, vol. 287, no. 3, 1992, pp. 797-803.
51. Morris, G.M., et al. "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility." *Journal of Computational Chemistry*, vol. 30, no. 16, 2009, pp. 2785-91.
52. Xie, X., et al. "Structure-Based Design of Pan-Coronavirus Inhibitors Targeting Host Cathepsin L and Calpain-1." *Signal Transduction and Targeted Therapy*, vol. 9, 2024, article 54.
53. Li, Q., et al. "An Approach Combining Deep Learning and Molecule Docking for Drug Discovery of Cathepsin L." *Expert Opinion on Drug Discovery*, vol. 18, no. 4, 2023, pp. 347-56.
54. Ehmke, V., et al. "Optimization of Triazine Nitriles as Rhodesain Inhibitors: Structure–Activity Relationships, Bioisosteric Imidazopyridine Nitriles, and X-Ray Crystal Structure Analysis with Human Cathepsin L." *ChemMedChem*, vol. 8, no. 6, 2013, pp. 967-75.
55. Falke, S., et al. "Structural Elucidation and Antiviral Activity of Covalent Cathepsin L Inhibitors." *Journal of Medicinal Chemistry*, vol. 67, no. 9, 2024, pp. 7048-67.

Chapter 7. Mapping the Vulnerabilities of Coronaviruses: Druggable Hotspots in RNA Methylation Machinery

7.1. Introduction.

Hot spots are specific amino acid residues located at protein–protein or protein–ligand interfaces that contribute disproportionately to the overall binding free energy, typically exhibiting a $\Delta\Delta G \geq 2$ kcal/mol upon alanine substitution [1-3]. The identification of hot spots has been instrumental in rational drug design, since targeting these key residues can enhance both the specificity and efficacy of therapeutic inhibitors, and has led to the discovery of druggable binding sites in numerous proteins [4]. Hot spot residues are typically clustered within specific regions that interact with endogenous ligands and often form networks of interactions implicated in allosteric regulation [5]. Computational methods, such as computational alanine scanning (CAS), have refined the identification of hot spot regions by quantifying the energetic contribution of individual residues to binding [1,6,7].

Coronaviruses are positive-sense single-stranded RNA viruses possessing the largest genomes among RNA viruses. This genomic complexity contributes to their high rates of mutation and recombination, particularly within the receptor-binding domain (RBD) of the spike protein [8]. All known human coronaviruses, including those responsible for common colds, originate from zoonotic sources [8] and have caused several spillover events over the past two decades, including SARS-CoV-1 (2002), MERS-CoV (2012), and SARS-CoV-2 (2019). In this context, the

identification of novel druggable molecular targets in coronaviruses represents a major priority for the development of antiviral agents aimed at counteracting both current and emerging coronavirus-related diseases. A key hallmark of coronavirus replication is the RNA capping process, which consists of adding a cap structure to the 5' end of the viral mRNA. This modification is crucial for generating a functional 5' cap that enhances RNA stability and enables the viral genome to mimic host mRNA, thereby facilitating immune evasion [9]. In addition to its roles in translation and immune evasion [10], RNA methylation also modulates local RNA secondary structures [11].

Coronaviruses are capable of autonomously performing RNA capping through their own enzymatic machinery. In SARS-related coronaviruses, this process proceeds through a four-step sequential reaction catalyzed by viral nonstructural proteins (NSPs) [12]. While the early steps of the RNA capping pathway are still under investigation, with two alternative mechanisms proposed in the second step implicating either nsp13 [13] or nsp12 [14], on the contrary, the final methylation steps are well-characterized. (Figure 1).

These are carried out by nsp10 in complex with nsp14, an N7-methyltransferase (N7-MTase), and nsp10 in complex with nsp16, a 2'-O-ribose methyltransferase (2-O-MTase). Nsp10 lacks intrinsic enzymatic activity but serves as a central cofactor that recruits and activates both nsp14 and nsp16 [15] by providing structural support and by regulating the methylation process [16-18]. Both nsp14 and nsp16 require S-adenosyl-L-methionine (SAM) as a methyl donor and nsp10 as a cofactor.

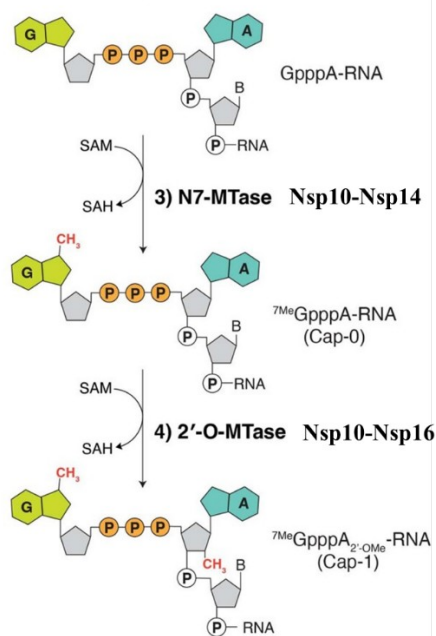


Figure 1. Scheme of the proposed RNA capping mechanism in SARS-CoV-2, with a focus on the last two steps.

The methyltransferase (MTase) activity of the nsp10/nsp16 complex appears to depend on the prior formation of the N7-methylguanosine cap catalyzed by the nsp10/nsp14 complex [19].

Nsp10/nsp14 and nsp10/nsp16 heterodimers are catalytically active *in vitro*, with nsp10 acting as a stimulatory cofactor for both complexes (e.g. nsp14 ExoN/MTase activity) [17,20]. In the cellular context, they are recruited to the replication–transcription complex (RTC), a multi-protein machinery essential for viral genome replication and transcription [21,22]. Although the full structure and composition of the replication–transcription complex (RTC) have not yet been completely elucidated, nsp7, nsp8, and nsp12 are regarded as its core components [23]. Other non-structural proteins (NSPs)—including nsp9, nsp10, nsp13, nsp14, and nsp16—are also essential elements that can be dynamically recruited during viral RNA synthesis [21,22].

Due to their central function as methylation enzymes within the coronavirus RNA capping machinery, nsp10, nsp14, and nsp16 have emerged as attractive targets for the rational design of new antiviral therapeutics [24,25].

This chapter reports the hotspot residue prediction carried out using computational alanine scanning (CAS) mutagenesis on the experimentally determined structures of SARS-CoV-2 nsp10, nsp14, and nsp16 listed in Table 1: Appendix E.

The structures were clustered according to their functional ligand(s), reflecting distinct conformational and functional states. For each structure, the effect of single-point mutations on binding affinity was computed as the mutation energy ($\Delta\Delta G_{\text{mut}}$). The conservation of the identified hotspot residues among coronaviruses was then assessed using experimentally determined structures of homologous NSPs sharing at least 70% sequence identity.

The hot regions predicted through our CAS mutagenesis analysis were mapped onto the surface of each protein structure. The root-mean-square fluctuation (RMSF) values of the predicted hotspot residues were calculated across all protein conformational and functional states. Overall, this study provides a comprehensive overview of conserved, druggable hotspots in key non-structural proteins of the coronavirus family, offering a valuable framework for the rational identification and design of antiviral agents targeting coronaviruses.

I was directly involved in conducting computational studies, including structure retrieval and preparation, CAS and RMSF calculations, and subsequent data analysis. I presented the results during multidisciplinary group meetings related to the financed project in which I am involved

UNmasking Viral RNA: targeting viral RNA capping machinery to tackle COVID-19 and future CoV emergencies (UNVIR-19) PRIN2022 (Prot. 2022MBK24T) which included, in addition to my supervisor Prof. Caterina Fattorusso's group, the research teams of i) Prof. Gaetano Malgieri (Università degli Studi della Campania Luigi Vanvitelli) for NMR studies, and (ii) Dr. Laura Zaccaro (Institute of Biostructures and Bioimaging – IBB, National Research Council – CNR) for peptide and protein production.

The results of this study, as presented here, were also showcased by me in poster format at the 14th European Conference on Computational and Theoretical Chemistry (EuChemS CompChem 2025, Naples, Italy).

7.2. Results and Discussion.

The coronavirus RNA-capping machinery plays a pivotal role for viral survival. It involves cascades of biological processes and a series of NSPs. Three of these are directly involved in RNA methylation: nsp10, the N7-methyltransferases (N7-MTase) nsp14, and 2'-O-MTase nsp16, representing, therefore, attractive drug targets.

Although nsp10 lacks intrinsic catalytic activity, its physical contact with the two cognate proteins ultimately determines the virus survival in the host cell. Therefore, in the quest for inhibitors of the formation of the nsp10-nsp14 and/or nsp10-nsp16 complexes is a promising approach to develop antivirals against SARS-CoV-2 and related coronaviruses. Advantages of nsp10 as suitable drug target are: its stable and soluble nature at room temperature and the availability of experimentally determined structures showing the interactions with sp14 and nsp16 [26]. For similar reasons, nsp14 and nsp16 are also considered favorable targets.

Moreover, they feature additional druggable pockets associated with their enzymatic activities [27]. Importantly, a wealth of computational and experimental data evidenced that nsp10, nsp14 and nsp16 have relatively low mutation rate, which indicates that these proteins remained relatively stable during evolution [28-29].

A large set of experimental structures was considered for the calculation of the binding mutation energy. Importantly, our analysis was not limited to protein partners but included the binding to all functional ligands found in the analyzed structures, such as RNA, SAM and SAH, inhibitors, and sulphate ions mimicking the RNA phosphate groups.

7.2.1. Computational Alanine Scanning (CAS) mutagenesis experiments.

The SARS-CoV-2 nsp10, nsp14, and nsp16 PDB structures listed in Appendix F: Table 1 were clustered according to their conformational and functional states, which correspond to binding with one or more active ligands (Table 1 and 2). Accordingly, Tables 1 and 2 summarize the structural clusters identified for the nsp10–nsp14 and nsp10–nsp16 complexes, respectively.

Table 1. Functional clusters of the nsp10-nsp14 complexes used in CAS studies.

Functional state	PDB ID	Cluster
SARS-CoV-2 nsp10 bound to nsp14	7DIY, 7MC5, 7MC6	1
SARS-CoV-2 nsp10-nsp14-RNA complex	7N0B, 7N0C,7N0D	2
Monomeric form of the SARS-CoV-2 replication-transcription complex	7EIZ	3

Dimeric form of the SARS-CoV-2 replication-transcription complex	7EGQ	4
--	------	---

Table 2. Functional clusters of the nsp10-nsp16 complexes used in CAS studies.

Functional state	PDB ID	Cluster
SARS-CoV-2 nsp10 bound to nsp16	7ULT	1
SARS-CoV-2 nsp10 bound to nsp16 and the cofactor SAM	7BQ7, 6W61, 7C2J, 7C2I, 6W4H, 6W75, 6XKM	2
SARS-CoV-2 nsp10 bound to nsp16 and the inhibitors analogues of SAM/SAH	8BZV, 8BSD, 8C5M, 8OT0, 8OTO, 8OV2, 8OV3, 8OV4, 8A23, 6YZ1, 6WKQ	3
SARS-CoV-2 nsp10 bound to nsp16 and allosteric inhibitors	8F4S, 8F4Y	4
SARS-CoV-2 nsp10 bound to nsp16, SAM and Cap-0	7KOA, 7JYY, 6WKS, 7JPE, 6WVN, 7R1U	5
SARS-CoV-2 nsp10 bound to nsp16, SAH and Cap-0	6WQ3, 6WRZ	6
SARS-CoV-2 nsp10 bound to nsp16, SAH and Cap-1	7JZ0, 7L6R, 7LW3, 7L6T, 7JHE	7
SARS-CoV-2 nsp10 bound to nsp16 and SAH	6WJT, 7LW4	8

Four distinct conformational and functional states were obtained for the nsp10–nsp14 complexes, and eight for the nsp10–nsp16 complexes.

The effect of each mutation on binding affinity (mutation energy, $\Delta\Delta G_{\text{mut}}$) was calculated as the difference between the binding free energy of the mutant and the corresponding wild-type structure: $\Delta\Delta G_{\text{mut}} = \Delta G_{\text{bind}}(\text{mutant}) - \Delta G_{\text{bind}}(\text{wild type})$. All residues were then ranked and classified according to their $\Delta\Delta G_{\text{mut}}$ values as follows: (i) ≥ 0.5 and < 1.0 kcal/mol = low, (ii) ≥ 1.0 and < 2.0 kcal/mol = medium, and (iii) ≥ 2.0 kcal/mol = high. To account for potential long-range allosteric effects,

residues with very low mutation energies (> 0.1 and < 0.5 kcal/mol) were also mapped, although they were not considered hotspot residues.

7.2.2. SARS-CoV-2 Nsp10 Predicted Hot Spot Residues.

The hot spot residues predicted by our CAS mutagenesis study were mapped on the 3D-structure of nsp10 (Figure 2).

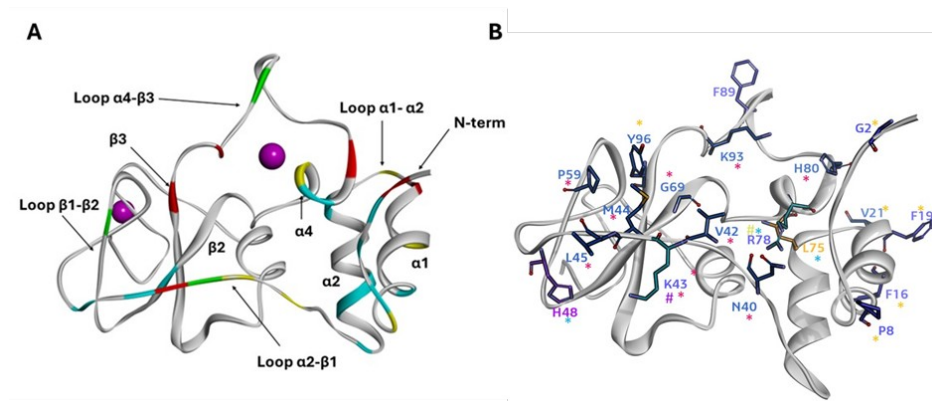


Figure 2. Nsp10 from the cryo-EM structure of SARS-CoV-2 nsp10-nsp14 (WT)-RNA complex (only Nsp10 is shown, PDB ID: 7N0B). Nsp10 is displayed in solid ribbon and colored in white. A) Protein backbone is colored according to $\Delta\Delta G_{mut}$ values (cyan=very low; green=low; yellow=medium; red=high) (B) Residues that are hot spot are showed in stick and colored accordingly with the binding partner: blue with the other proteins (Nsp16 or Nsp14), yellow with allosteric inhibitors, purple with the RTC macrocomplex (Nsp9). The colored asterisk represents the predicted hot spot residue with the partner protein: blue for Nsp16, yellow for Nsp14 and magenta for both the proteins. Colored hash represents residues that are also at the interface with: allosteric inhibitors for R78 (yellow hash) and SO_4^{2-} for K43 (purple hash).

In Figure 2 protein backbone is colored by the $\Delta\Delta G_{mut}$ values (Figure 2A) and highlighted according to the functional ligand involved in the interaction (Figure 2B). Nsp10 tertiary structure comprises six α -helices and an antiparallel β -sheet. The $\alpha 1$ helix (Pro8-Phe19) plays a key role in

binding nsp14 but it is also expendable for nsp16-engagement. From our study this region contains nsp10 hotspot residues Phe8, Thr12, Phe16, and Phe19 in nsp10-nsp14 complex structures, of which Phe16 and Phe19 are predicted in all structures.

Figure 3 shows the nsp10 solvent accessible surface colored according to the calculated $\Delta\Delta G_{\text{mut}}$ values, in A it is reported the nsp10-nsp14 complex and in B the nsp10-nsp16 complex.

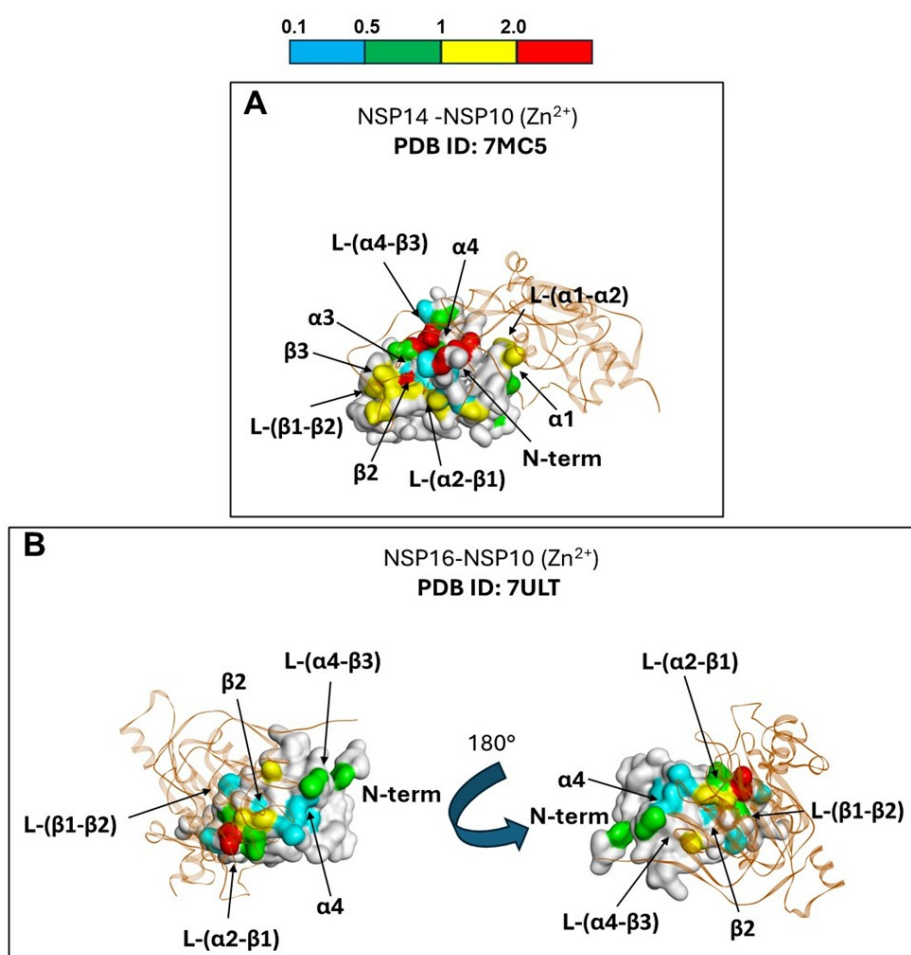


Figure 3. A) nsp10-nsp14 complex. B) nsp16-nsp10 complex. Nsp14 (A) and nsp16 (B) are displayed as line ribbons and colored in brown. The solvent-accessible surface of Nsp10 is displayed and colored according to the residue mutation energy ($\Delta\Delta G_{\text{mut}}$): red

($\Delta\Delta G_{\text{mut}} \geq 2$ kcal/mol, high), yellow ($1 \leq \Delta\Delta G_{\text{mut}} < 2$ kcal/mol, medium), and green ($0.5 \leq \Delta\Delta G_{\text{mut}} < 1$ kcal/mol, low).

Shown below (Figure 4) are the bar graphs reporting the predicted hotspots of SARS-CoV-2 nsp10 in complex with nsp14 (A) and nsp16 (B), respectively. X-axis = hotspot residues, Y-axis = Bars are color-coded according to the mutation energy, arrows to the binding partners.

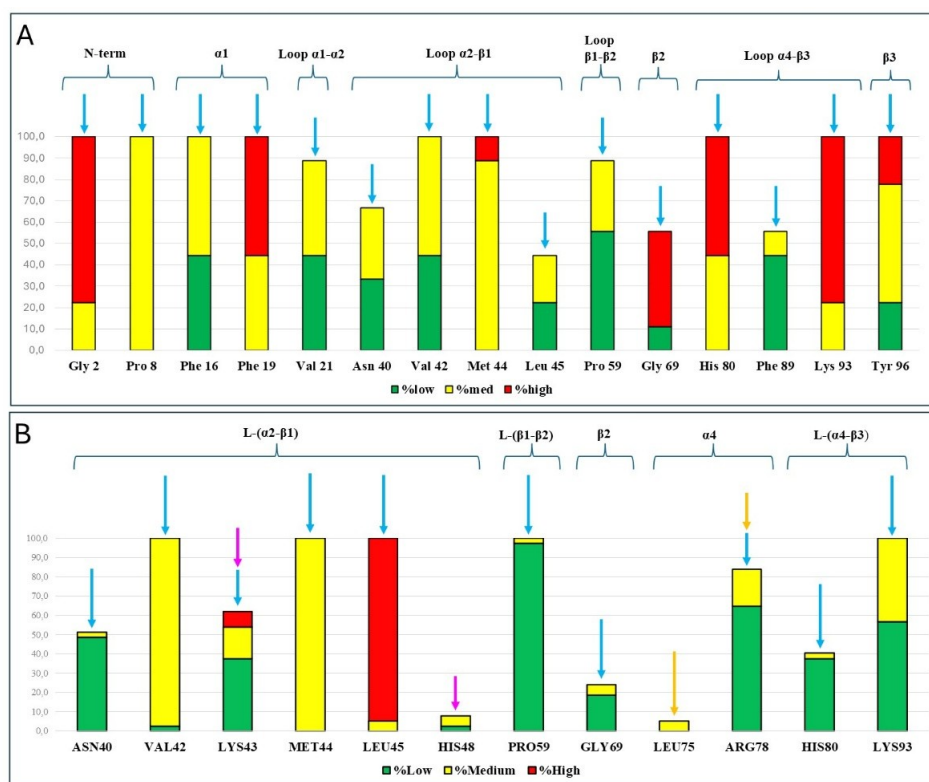


Figure 4 Bar graphs showing the hotspot residues (X-axis) and the percentage of complexes in which they are present (Y-axis) for (A) nsp10/nsp14 and (B) nsp10/nsp16 complexes. Bars are color-coded according to the mutation energy ($\Delta\Delta G_{\text{mut}}$): red ($\Delta\Delta G_{\text{mut}} \geq 2$ kcal/mol, high), yellow ($1 \leq \Delta\Delta G_{\text{mut}} < 2$ kcal/mol, medium), and green ($0.5 \leq \Delta\Delta G_{\text{mut}} < 1$ kcal/mol, low). Arrows indicate the functional ligand involved in the interaction: blue = nsp14 (A) or nsp16 (B); magenta = sulfate groups (mimicking RNA phosphate positions); yellow = inhibitors.

There are several nsp10 residues identified as hotspots for the binding to both nsp14 and nsp16. The long $\alpha 2$ – $\beta 1$ loop (Gly34–Ala54) of nsp10 acts as an interface with either nsp14 or nsp16 and, accordingly, contains several shared hotspots, including Asn40, Val42, Met44, and Leu45. In particular, Met44 presents a $\Delta\Delta G_{\text{mut}}$ value in between 1 and 2 kcal/mol in all the nsp10-nsp14 and nsp10-nsp16 complexes (Figure 4). Met44 is, indeed, involved in similar interactions in the binding of both proteins. Conversely, Leu45 displays a high mutation energy exclusively in the complexes with Nsp16, where it forms a hydrogen bond with Gln87 and engages in alkyl–alkyl interactions with Ile40, Leu244, and Met247. In the nsp10–nsp14 complexes, Leu45 interacts with nsp14 only through hydrophobic contacts (with Leu38 or Val40; Appendix E: Table 9 and 10).

Lys93 is another nsp10 residue identified as a common hotspot for both nsp14 and nsp16 binding. In particular, in the nsp10-nsp16 complexes, this residue is involved in hydrogen bond interactions with Ala107 and Ser105, while, when Nsp10 is bound to Nsp14, Lys93 establishes an ionic interaction with Asp126 and a hydrogen bond with Thr127. Accordingly, Lys93 shows a high predicted mutation energy in the complexes with Nsp14 (Figure 4; Appendix E: Tables 11 and 12). Mutants of this residue in nsp10 of different viruses, exhibited a drastic loss of interaction with nsp14 and nsp16, thereby abolished the Exon activity, and 2-O-methyltransferase activity [30-32].

Pro59 is also predicted as a hot spot residue in both nsp10-nsp14 and nsp10-nsp16 complexes, present in almost all the analyzed structures. Pro59 is located in the loop region from Val57 to Asp64 flanking the short beta-sheet-1 (Ile55-Thr56) and beta-sheet-2 (Gln65-Gly69).

Finally, Gly69 is involved in the binding with both nsp14 and nsp16 of nsp10, with a high value of $\Delta\Delta G_{mut}$ when in complex with nsp14 (Figure 4). It is located in the second beta sheet and is extended by another Gly70 residue to the third helix.

To refine our analysis, the identified hot spot residues were analyzed considering the functional states of the proteins. Figure 5 reports the results obtained for the nsp10-nsp14(A) and nsp10-nsp16 (B) complexes, respectively.

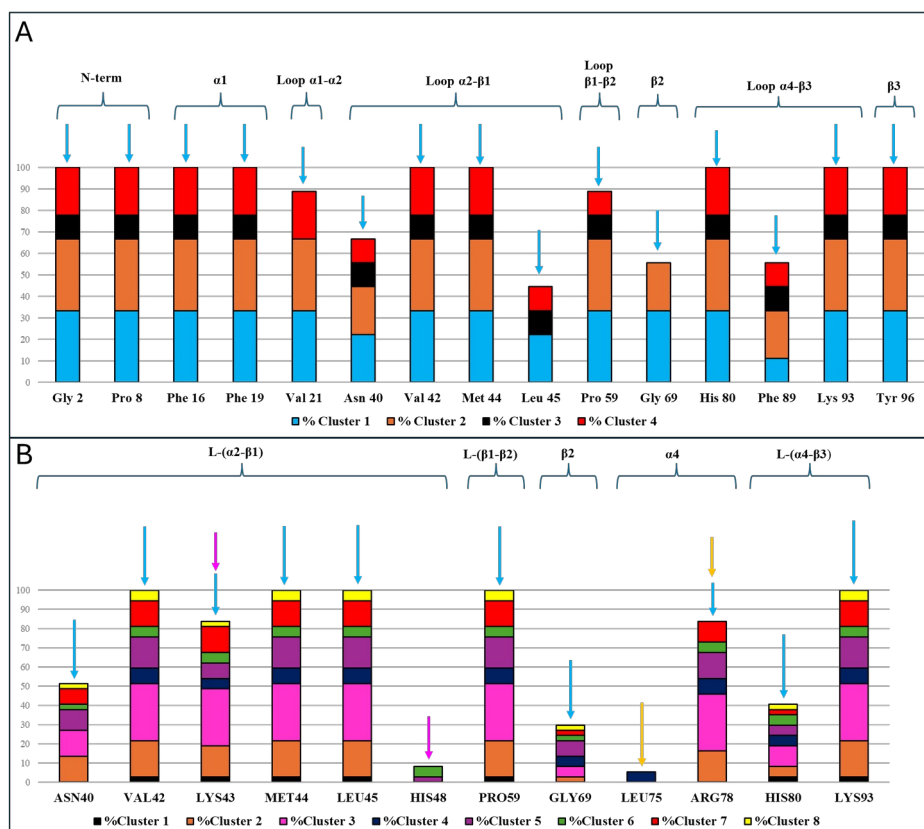


Figure 5. A: percentage of SARS CoV-2 nsp10 predicted hotspot residues in nsp10-nsp14 complexes according to the protein functional states. The clusters are defined in Table 1. Blue arrows indicate residues at the interface with nsp14. B:percentage of SARS CoV-2 nsp10 predicted hotspot residues in nsp10-nsp16 complexes according to the protein functional states. The clusters are defined in table 2. Blue arrows indicate residues binding

nsp16; magenta arrows indicate residues binding SO_4^{2-} ; orange arrows indicate residues binding inhibitors.

As shown in Figure 4A, 5A, and 6, in the complexes with nsp14, the effects of mutations on nsp10 binding affinity ($\Delta\Delta G_{\text{mut}}$) are exclusively related to its interaction with nsp14, since no other molecule is directly bound to nsp10 (except the structural zinc ions, excluded from this analysis). The identified residues are distributed across several surface regions, including the N-terminal tail (Gly2, Pro8), α -helix $\alpha 1$ (Phe16, Phe19), loop $\alpha 1$ – $\alpha 2$ (Val21), loop $\alpha 2$ – $\beta 1$ (Asn40, Val42, Met44, Leu45), loop $\beta 1$ – $\beta 2$ (Pro59), β -sheet $\beta 2$ (Gly69), loop $\alpha 4$ – $\beta 3$ (His80, Phe89, Lys93), and β -sheet $\beta 3$ (Tyr96) (Figures 2, 3, and 6).

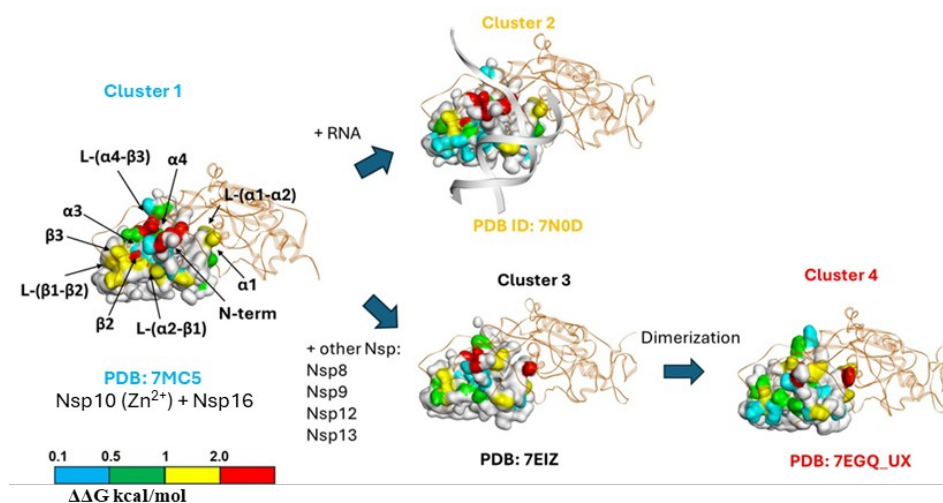


Figure 6. The best resolution structure of the functional clusters of Nsp10 in complex with Nsp14 is displayed. The solvent-accessible surface of Nsp10 is displayed and colored on the bases of the following $\Delta\Delta G_{\text{mut}}$ range: cyan for residue with $\Delta\Delta G_{\text{mut}}$ between 0.1 kcal/mol and 0.5 kcal/mol; green for residues with $\Delta\Delta G_{\text{mut}}$ between 0.5 kcal/mol and 1.0 kcal/mol; yellow for residues with $\Delta\Delta G_{\text{mut}}$ between 1.0 kcal/mol and 2.0 kcal/mol; red for residues with $\Delta\Delta G_{\text{mut}}$ above 2.0 kcal/mol. Nsp14 is displayed as line ribbons and colored in brown.

Conversely, when in complex with Nsp16, Nsp10 also participates in the binding of additional ligands such as sulphate ions or the allosteric inhibitor pyrimidin-2-ol (Figures 4B, 5B, and 7). Interestingly, one sulphate ion is positioned at the interface with nsp16, mimicking the location of the RNA phosphate group when bound to its low-affinity site [21].

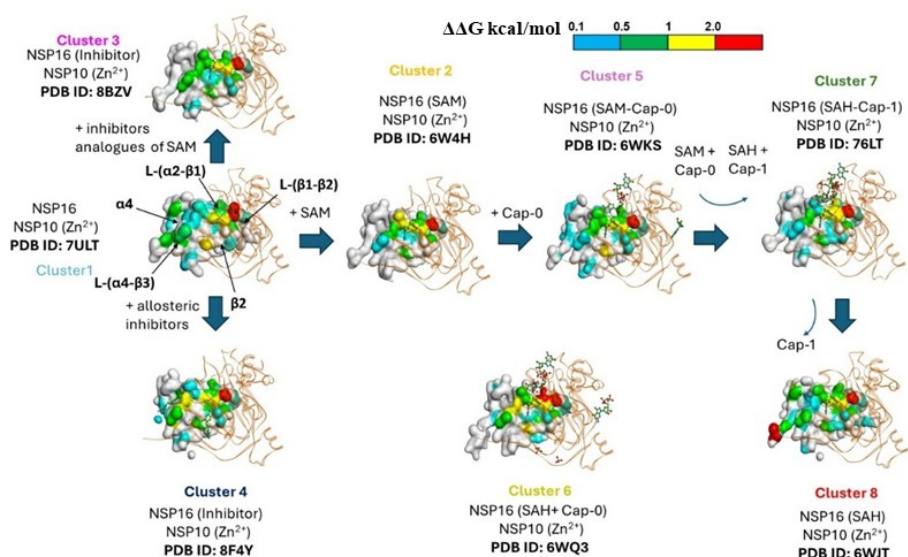


Figure 7. The best resolution structure for each functional cluster of Nsp10 in complex with Nsp16 is displayed. The solvent-accessible surface of Nsp10 is displayed in solid and colored on the bases of the following $\Delta\Delta G_{\text{mut}}$ range: cyan for residue with $\Delta\Delta G_{\text{mut}}$ between 0.1 kcal/mol and 0.5 kcal/mol; green for residues with $\Delta\Delta G_{\text{mut}}$ between 0.5 kcal/mol and 1.0 kcal/mol; yellow for residues with $\Delta\Delta G_{\text{mut}}$ between 1.0 kcal/mol and 2.0 kcal/mol; red for residues with $\Delta\Delta G_{\text{mut}}$ above 2.0 kcal/mol. Nsp16 is displayed as line ribbons and colored in brown.

Consistently with prior structural/biochemical work, CAS highlights a compact set of nsp10 hot spot residues. The Asn40–Leu45 loop acts as a

conserved region used across nsp14-, nsp16- and oligomeric contexts, aligning with mutagenesis data that reported severe losses of binding/activity for K43A, M44A, V42A in SARS/SARS-CoV-2 backgrounds [21,33,34]. The $\alpha 1$ helix (Pro8–Phe19) contributes asymmetrically, being critical for nsp14 engagement yet more expendable for nsp16, again in accord with legacy studies. Together, these observations delineate a PPI hotspot region suitable for small-molecule or peptidomimetics targeting.

The N-terminal loop Ala1–Pro8 includes residues Gly2 and Pro8, both exhibiting $\Delta\Delta G_{\text{mut}}$ values greater than 1 kcal/mol exclusively in the Nsp10–Nsp14 complexes. Similarly, Phe16 and Phe19, located within the first α -helix, were identified as hotspots in all Nsp10–Nsp14 structures, displaying energetically significant mutation energies. Conversely, Val21, positioned in the $\alpha 1$ – $\alpha 2$ loop, was predicted as a hotspot only in the nsp10–nsp14 complexes, except for the SARS-CoV-2 replication–transcription complex (monomeric form; PDB ID: 7EIZ). In this specific structure, Val21 loses its interactions with Lys200 (alkyl–alkyl interaction) and Phe217 (π –alkyl interaction) (Appendix E: Table 13). It can be observed that, beyond the direct “contacting residues”, both the Nsp14–Nsp10 and Nsp16–nsp10 complexes also involve “nearby” residues, likely reflecting allosteric effects. Pro59, at the $\beta 1$ loop (Val57–Asp64), emerges from CAS across nsp10–nsp14 and nsp10–nsp16 and is conserved in SARS-CoV and MERS. Gly69 ($\beta 2$) is a recurrent hotspot in nsp10–nsp14; prior mapping showed that nsp10 residues 65–107 is enough for nsp16 binding, whereas 42–120 are required for catalytic stimulation of 2'-O-MTase; G70A/S72A mutations impair Nsp16 engagement/activity [34,35], which mirrors hotspot distribution around G69–G70–S72. It should be noted, however, that due to the absence of molecular dynamics calculations on

the experimentally determined structures, allosteric effects may be underestimated. Accordingly, the information provided by our analysis warrants further investigation.

7.2.3 Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp10 Hot Spots

Although no molecular dynamics simulations were performed, a preliminary assessment of residue flexibility was conducted through root-mean-square fluctuation (RMSF) analysis. Specifically, RMSF values were calculated for both the analyzed complexes and their corresponding apo forms.

This analysis revealed the presence of three regions in Nsp10 (i.e., $\alpha 1/\alpha 1-\alpha 2$ loop (aa10–22), the $\alpha 4-\beta 3$ loop (aa80–95), and the C-terminal region (aa125–130); Table 3) all characterized by a certain degree of flexibility (RMSF value > 1 Å; Figure 8; Appendix E : Table 2).

Table 3. Substructures of nsp10.

Residues	Secondary structure^a
10-19	$\alpha 1$
23-33	$\alpha 2$
55-56	$\beta 1$
65-69	$\beta 2$
70-73	$\alpha 3$
75-79	$\alpha 4$
96-100	$\beta 3$
101-103	$\alpha 5$
107-113	$\alpha 6$
115-116	$\beta 4$
123-124	$\beta 5$

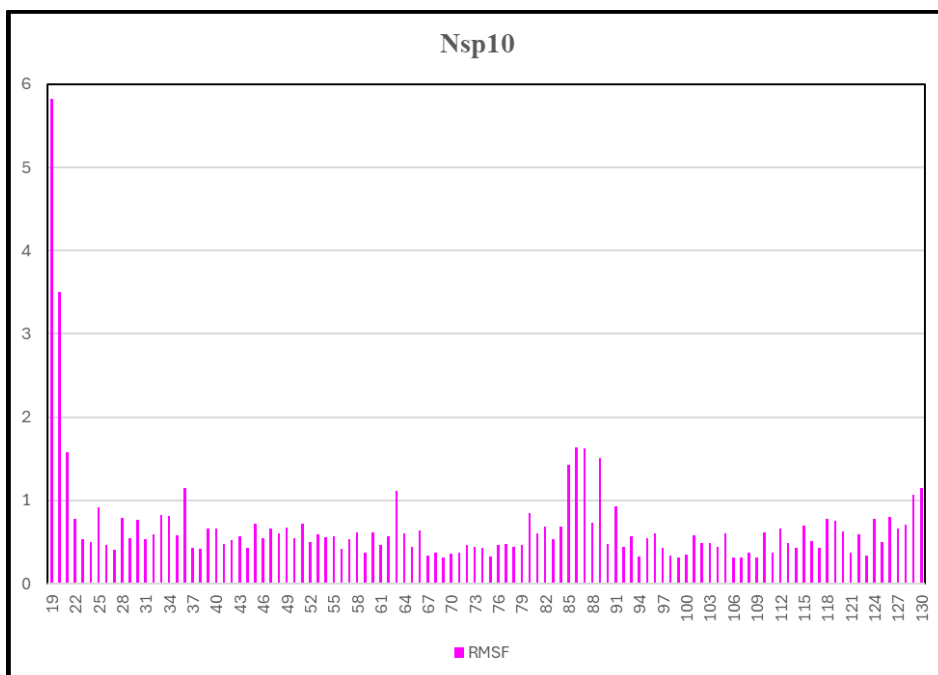


Figure 8. RMSF values (Å) of each residue calculated for Nsp10.

Interestingly, the $\alpha 1/\alpha 1\text{--}\alpha 2$ loop and the $\alpha 4\text{--}\beta 3$ loop contain the hotspot residues Phe19, Val21, and Phe89 (Figure 5) and are associated with local secondary structure variability. Specifically, residues Phe19–Ala20 adopt an α -helical conformation in approximately 10% of the analyzed structures and a coil conformation in the remaining 90%. Similarly, residues Pro86–Gly88 adopt an α -helical conformation in 21% of the structures and a coil conformation in 79% (Table 4).

Table 4. SARS-CoV-2 Nsp10 regions exhibiting changes in secondary structure

Residues	Secondary Structure (%)		
	α -helix	β -sheet	Coil
Phe19-Ala20	10	0	90
Ile55-Thr56	0	95	5
Gly70-Cys73	94	0	6
Pro86-Gly88	21	0	79
Thr101-Cys103	87	0	13
Thr115-Val116	0	2	98
Trp123-Lys124	0	2	98
Cys130	2	0	98

7.2.4. SARS-CoV-2 Nsp14 Hot Spot Residue Prediction.

Shown below are the predicted computational hotspots of SARS-CoV-2 Nsp14 (Figure 9), the predicted hotspots across the different functional states (Figure 10A), and the calculated average residue fluctuations (RMSF) in the analyzed structures (Figure 10B and Appendix E : Table 4).

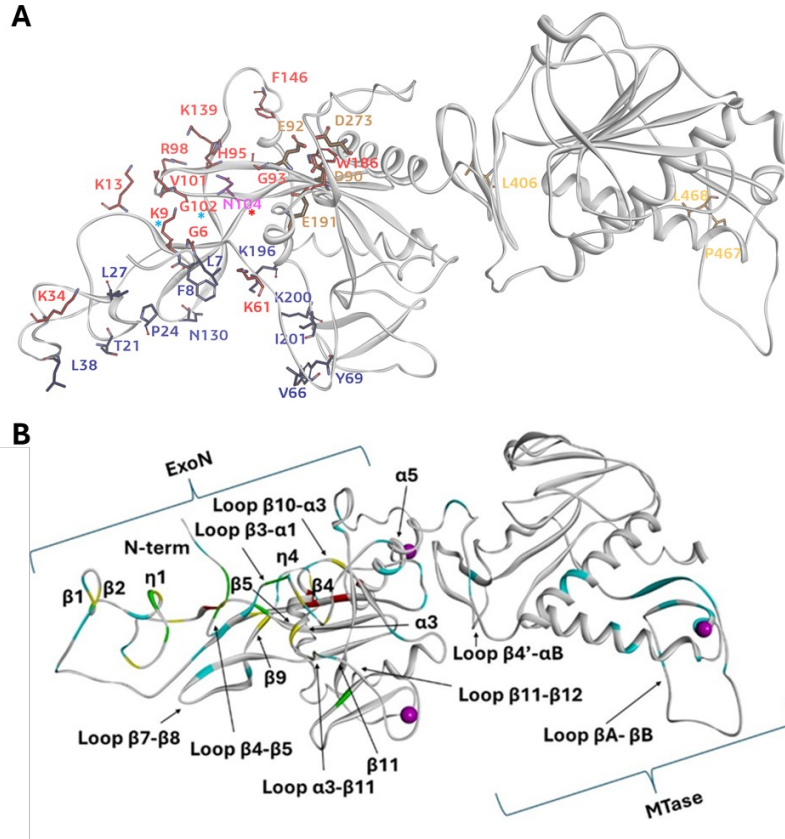


Figure 9. Nsp14 from the Cryo-EM structure of SARS-CoV-2 nsp10-nsp14 (WT)-RNA complex (only Nsp14 is showed, PDB ID: 7N0B). Nsp14 is displayed in solid ribbon and colored in white. Upper panel: Residues that are hot spot are showed in stick and colored accordingly with the binding partner: light blue for nsp10, red for RNA (high affinity binding site), golden for Mg^{2+} , yellow for the macrocomplex (Nsp13) and purple for the macrocomplex (nsp9). The colored asterisks represent hot spot residues that are also at the interface with: blue for Nsp10 and red for RNA(High affinity binding site). Lower panel: protein backbone is colored according to $\Delta\Delta G_{mut}$ values (cyan=very low; green=low; yellow=medium; red=high),

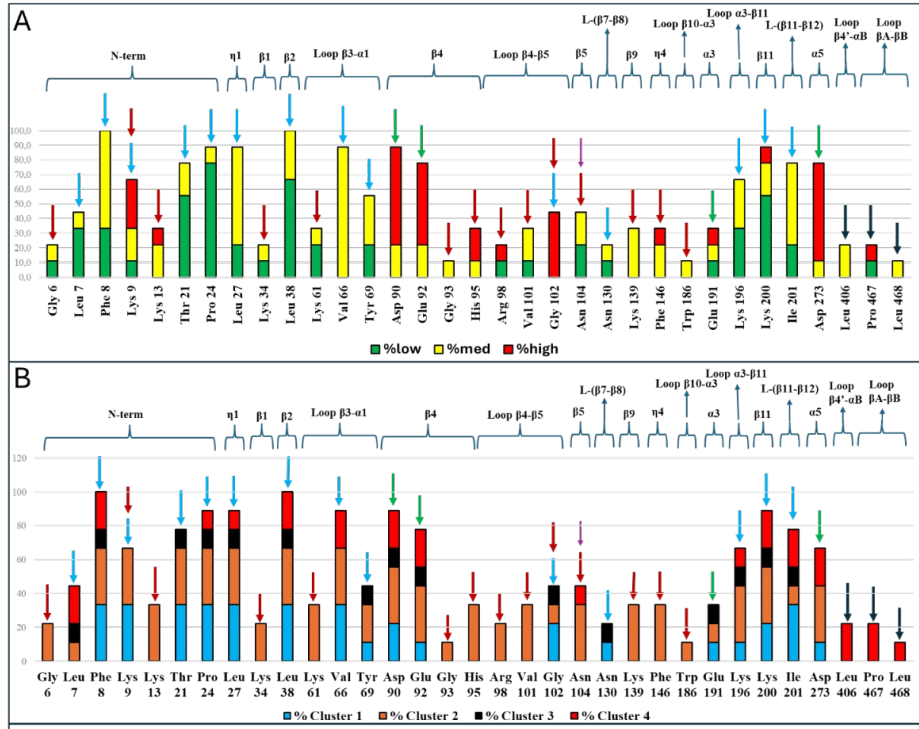


Figure 10. A: percentage of SARS CoV-2 nsp14 predicted hotspot residues in nsp10-nsp14 complexes. Bars are color-coded according to the mutation energy ($\Delta\Delta G_{mut}$): red ($\Delta\Delta G_{mut} \geq 2$ kcal/mol, high), yellow ($1 \leq \Delta\Delta G_{mut} < 2$ kcal/mol, medium), and green ($0.5 \leq \Delta\Delta G_{mut} < 1$ kcal/mol, low). B: percentage of SARS CoV-2 nsp14 predicted hotspot residues in nsp10-nsp14 complexes according to the functional/conformational states. The clusters are defined in Table 1. Arrows indicate the functional ligand involved in the interaction and are colored as follows: blue = nsp10; red = RNA; green = Mg^{2+} ; navy blue = nsp12 or nsp13 (macrocomplex); magenta = Nsp9 (macrocomplex).

The presence of viral proofreading exonuclease nsp14-nsp10 complex mainly challenges the development of nsp12 (RNA-dependent RNA polymerase)-targeted nucleoside inhibitors and causes a decreased efficacy of nucleoside analogue drugs [36,37]. Thus, exploiting nsp14-nsp10 interaction is a promising strategy to develop more effective pan coronavirus inhibitors.

In this study the Exon domain of nsp14 contains considerably large set of hotspot residues than the MTase domain. This is not surprising because nsp14-ExoN is the only domain that interacts with nsp10 and that this interaction maintains its structural and functional integrity. Apparently, the disordered nature of nsp14 has been revealed from the crystal structure of SARS-CoV-2 nsp14 solved without nsp10 (PDB:7R2V) [38]. Herein, in the long loop region between Asn3 to Pro24 of nsp14 Exon, more than half of the residues are identified as hotspot in the analyzed structures. Among these residues, Leu7, Phe8, Lys9, Thr21 and Pro24 persisted in higher than 40% of the structures. Alanine substitution to the basic amino acids Lys9, Lys61, and Lys139 (all are also predicted hotspots in our CAS analysis), were shown to affect the ExoN activity with pronounced defect observed in K9A and K61A mutations [39]. It is noteworthy that Thr21 and Pro24 make interactions with multiple residues located in the helix-3 and helix-4. Interestingly, enough, those nsp10 residues in helix-3 (Ala71, Ser72), helix-4 (Arg78), loop (Gly94) that interact with nsp14 hotspots (Thr21 and Pro24) also interact with nsp16 residues and are conserved in MERS, SARS and SARS-CoV-2. More importantly, some of the first zinc coordinating residues C74 and C77 of nsp10 are placed in the same helix-3. Thus, the above-mentioned nsp10 and nsp14 Exon regions could be targets for small molecules or peptides. Additionally, residues Phe8,

Lys9, Pro24, and Val66 of nsp14 are identical residues in SARS CoV-2, SARS CoV, MERS, and RaTG13 [12].

As shown in Figure 10A, the effects of mutations on Nsp14 binding affinity ($\Delta\Delta G_{\text{mut}}$) are associated not only with its interaction with Nsp10 but also with interactions involving RNA, Mg^{2+} , Nsp12, Nsp13, and Nsp9. CAS concentrates hotspots in the ExoN region rather than in MTase, consistent with nsp10 being the obligate structural co-factor for ExoN folding and activity.

The disordered nature of apo nsp14 (e.g., SARS-CoV-2 apo; 7R2V) rationalizes why the N-terminal loop Asn3–Pro24 harbors multiple hotspots (Leu7, Phe8, Lys9, Thr21, Pro24): these residues either stabilize nsp10 binding or engage RNA/ Mg^{2+} , and alanine variants at K9, K61, K139 are known to compromise ExoN [39].

The data also recover conserved hotspots Leu27 ($\eta 1$) and Leu38 ($\beta 2$), which contact nsp10 Lys43/Leu45 and sit within a region (Lys34–Leu38) showing notable flexibility, matching the interface plasticity implied by RMSF and secondary structure variability.

The identified hotspot residues, mapped onto the nsp14 surface for each functional cluster and color-coded according to their $\Delta\Delta G_{\text{mut}}$ values, are shown in Figure 11.

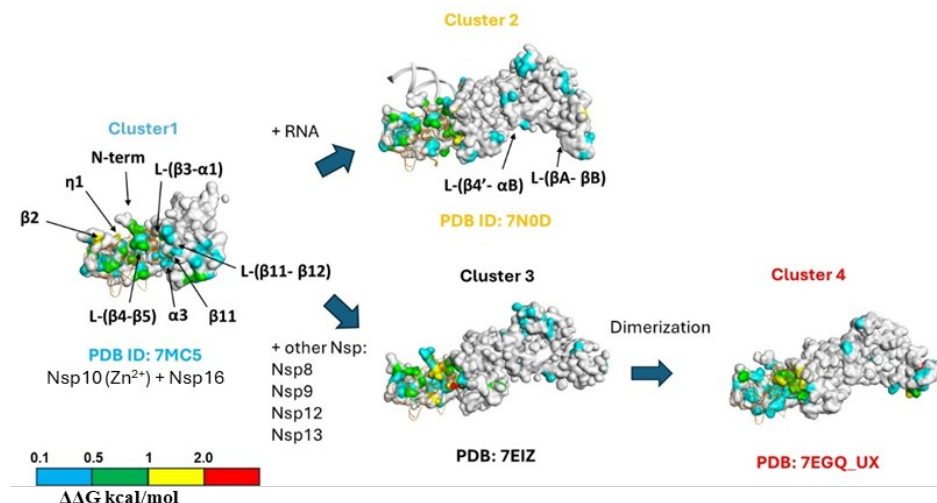


Figure 11. The best resolution structure for each functional cluster is displayed. The solvent-accessible surface of nsp14 is displayed in solid and colored on the bases of the following $\Delta\Delta G_{mut}$ range: cyan for residue with $\Delta\Delta G_{mut}$ between 0.1 kcal/mol and 0.5 kcal/mol; green for residues with $\Delta\Delta G_{mut}$ between 0.5 kcal/mol and 1.0 kcal/mol; yellow for residues with $\Delta\Delta G_{mut}$ between 1.0 kcal/mol and 2.0 kcal/mol; red for residues with $\Delta\Delta G_{mut}$ above 2.0 kcal/mol. Nsp10 is displayed as line ribbons and colored in brown.

The ExoN N-terminal loop, comprising residues Asn3 to Pro24 (Appendix E: Table 13), contains seven predicted hotspot residues across different structures. This loop also represents nearly half of the N-terminal region of nsp14 (residues 1–60), which has been experimentally shown to be disordered in the absence of nsp10 [40]. Among the predicted hotspots, Gly6 and Lys13 are involved exclusively in RNA binding (cluster 2), whereas Lys9 participates in interactions with both nsp10 (hydrogen bond with Ala1) and RNA (ionic interaction) (clusters 1 and 2) (Figure 10B).

All other residues are engaged solely in interactions with nsp10. Phe8 is present in all nsp10–nsp14 complexes, showing 66.7% medium and 33.3% low mutation energy, with $\Delta\Delta G_{\text{mut}}$ values ranging from 0.55 to 1.74 kcal/mol. It is well aligned across all structures and interacts with Thr5 (hydrogen bond), Leu14 (π –alkyl interaction), and His80 (CH–O or π – π interaction) of nsp10 (Appendix E: Table 14). Leu7, Lys9, Thr21, and Pro24 are also predicted in more than 40% of nsp10–nsp14 structures and represent energetically significant residues. RMSF analysis revealed that the N-terminal region of Nsp14 undergoes major conformational changes (Figure 12).

The first helix of the ExoN domain ($\eta 1$; Thr25–Lys32) contains the predicted hotspot residue Leu27 ($\Delta\Delta G_{\text{mut}}$: 0.74–1.36 kcal/mol), which appears in 89% of the analyzed structures. The first β -sheet of the ExoN domain ($\beta 1$; Table 7) includes Lys34 ($\Delta\Delta G_{\text{mut}}$: 0.79–1.03 kcal/mol), a predicted hotspot involved in RNA interaction (cluster 2) (Figure 10B). Leu38, which is in the short beta-sheet between Leu38 to Cys39, is another conserved hotspot in all nsp10–nsp14 analyzed structures. Leu38 of nsp14 makes interactions with nsp10 hotspot residues Lys43 and Leu45; the analysis of the RMSF values from Lys34 to Leu38 evidences a major conformational change of this region (Figure 12). The long loop of the nsp14 Exon domain that stretches from Ser56 to Ile74 contains multiple hot spot residues (differing in their $\Delta\Delta G_{\text{mut}}$ values and rate of occurrence) like Lys61, Val66, and Tyr69, with the former involved in the interaction with RNA while the other two are located at the binding interface with nsp10. A recent study shows how Lys61 represents one of the critical

residues to position the substrate RNA for exonucleolytic degradation and its Alanine mutant variant could severely affect the Exon activity [39]. Val66 is also identified as conserved hot spot residue among SARS CoV-2, SARS CoV, and MERS CoV [41-42]. The short β -sheet spanning Leu38–Cys39 exhibits major conformational changes across all structures and contains Leu38 as a hotspot residue, predicted in all cases with 33.3% medium and 66.7% low mutation energy ($\Delta\Delta G_{\text{mut}}$: 0.70–1.20 kcal/mol). The long $\beta 3$ – $\alpha 1$ loop (Ser56–Ile74) harbours multiple energetically significant hotspots. Among these, Lys61, Val66, and Tyr69 appear in more than 30% of the complexes analyzed. Notably, Val66 is present in 89% of the structures (except for PDB 7EIZ; cluster 3), consistently showing medium mutation energy ($\Delta\Delta G_{\text{mut}}$: 1.43–1.87 kcal/mol). This residue is well aligned across hotspot structures and maintains a conserved π -alkyl interaction with Phe16 of nsp10 (Appendix E: Table 15). Interestingly, Val66 is also identified as a common hotspot in both MERS-CoV and SARS-CoV [32].

The $\beta 4$ – $\beta 5$ loop contains five predicted hotspot residues—Gly93, His95, Arg98, Val101, and Gly102—all appearing in less than 45% of the analyzed structures and primarily involved in RNA binding (cluster 3) (Figure 10 B). Gly102 also participates in interactions with the N-terminal region of nsp10 (clusters 1–3) (Figure 10B). In some structures the loop between Thr97 to Thr103 of nsp14 contains predicted hotspot residues, such as, Arg98 and Val101, involved in RNA binding, and Gly102 interacting with both nsp10 and RNA. The nearby residue Asn104 is also a predicted hotspot in more than 40% of the analyzed structures. Another long helix (His188-Lys196) in the Exon region contains the predicted hot spot residue Lys196 at the interface with nsp10. Lys200 and Ile201 are predicted hot spot residues in more than 60% of the structures and are

located in the small β -sheet Val199-Lys200 and in the loop Ile201-Thr215, which is in the interstices of the zinc and magnesium binding sites [12].

Notably, the catalytic residues Asp90 and Glu92, which coordinate Mg^{2+} , are also predicted hotspots in this region. Within this hotspot cluster, Glu92, His95, and Arg98 exhibit significant side-chain reorientation. His95 is identified as a hotspot in only three structures (22.2%), all belonging to the nsp10–nsp14–RNA complexes (cluster 2) (Figure 10B); in all other structures, it contributes to local stabilization. Previous studies have demonstrated that His95 plays a role in the ExoN–RNA interaction and facilitates the 3′-to-5′ exoribonuclease activity of SARS-CoV-2 nsp10–nsp14 [43], which may explain its hotspot behaviour in these complexes.

Similarly, the β 5 strand (Asn104–Phe111) contains Asn104 as the only hotspot residue, displaying 22.2% medium and 22.2% low mutation energy ($\Delta\Delta G_{mut}$: 0.60–1.63 kcal/mol) (Figure 10A). This residue is involved in interactions with RNA (all structures of cluster 2) or with nsp9 (cluster 4) (Figure 10B). Additionally, Lys139 (β 9), Phe146 (η 4), and Trp186 (loop β 10– α 3) were predicted as hotspot residues due to their involvement in RNA interactions (cluster 2). Conversely, Glu191 (loop β 10– α 3), which interacts with Mg^{2+} , was identified as a hotspot residue in all structures where it was not mutated (7DIY, cluster 1; 7N0B, cluster 2; and 7EIZ, cluster 3), except for 7EGQ (cluster 4; macro complex dimeric form for proofreading activity).

Asn130 (loop β 7– β 8) was identified as a hotspot in only two structures, exhibiting 11% medium and 11% low mutation energy ($\Delta\Delta G_{mut}$ 0.52–1.00 kcal/mol) (Figure 10A). This residue forms a hydrogen bond either with Lys93 (Nsp10; PDB ID: 7EIZ) or with Asp82, His83, and Gly88

(Nsp10; PDB ID: 7MC5) (Appendix E: Table 16). Lys196, located in the long helix between His188 and Lys196 ($\alpha 3$), is the only hotspot residue identified in six structures, exhibiting 33.3% medium and 33.3% low mutation energy ($\Delta\Delta G_{mut}$ 0.55–1.47 kcal/mol) (Figure 10A). This residue participates in a π -alkyl interaction with Phe19 of Nsp10 (Appendix E: Table 17). Then, the short β -sheet Val199 to Lys200 ($\beta 11$), has Lys200 predicted as a hotspot residue in 89% of structures ($\Delta\Delta G_{mut}$: 0.54-2.21 kcal/mol) with the only exception of the PDB 7MC5 ($\Delta\Delta G_{mut}$: 0.32 kcal/mol). This residue is always involved in an alkyl-alkyl interaction with Val21 (nsp10) and in the PDB 7EIZ, where the effect of its mutation is high ($\Delta\Delta G_{mut}$: 2.21 kcal/mol), it is involved in the interaction with Asp82 (ionic interaction) as well as Ala18 and Ala20 (hydrogen bond) (Appendix E: Table 18). Similarly, Ile201 located in the loop $\beta 11$ - $\beta 12$ (Ile201-Thr215), was predicted as the only energetically significant residue in the 78% of the structures (56% medium and 22% low; $\Delta\Delta G_{mut}$: 0.61-1.53 kcal/mol) (Figure 10A). This residue is involved in the interaction with nsp10 establishing hydrogen bond with Phe19 and Val21 as well as alkyl-alkyl interaction with Ala20 (Appendix E: Table 19). Ile201 was identified as common hotspot in MERS CoV and SARS CoV [41,42].

Finally, except the residues of the active sites and those involved in the interaction with RNA, the hinge region and MTase domain of nsp14 structure of SARS-CoV-2 do not contain significant predicted hotspot residues. Indeed, only Leu406 (loop $\beta 4'$ - αB) as well as the residues Pro467 and Leu468 on the loop βA - βB was predicted as hotspot residues since all are involved in the interaction with nsp13 or nsp12 only in the macro complex active for proofreading activity (dimeric form; PDB ID: 7EGQ). In particular, Leu406 establishes a hydrogen bond with Gln886 (nsp12);

Pro467 establishes an alkyl-alkyl interaction with Leu92 (Nsp13), a π -alkyl interaction with Phe90 (nsp13), and a hydrogen bond with Ser904 (nsp13); Leu468 interacts with Val905 (nsp13) through an alkyl-alkyl interaction.

Among the hotspots identified in the nsp14 Exon domain, Thr5, Thr21, His26, Phe60, and Asn130, are nsp10 interacting residues both in the crystal structure of SARS CoV (PDB:5C8S) [17] and cryo-EM structure of SARS CoV-2 nsp10-nsp14 complex (PDB: 7EGQ,7EIZ) [12]. This study also demonstrates that Nsp9 and nsp12 NiRAN provides a platform to recruit nsp10-nsp14 into the RTC through nsp9 contacts with Thr103 and His95 of nsp14, and nsp12 NiRAN contacts with Lys13 and Arg98 of nsp14 [12]. Here, His95 is predicted as a hotspot in some of the analyzed structures.

7.2.5 Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp14 Hot Spots

In order to analyze putative allosteric effects, a preliminary assessment of residue flexibility was carried out through RMSF analysis, considering both the analyzed complexes and the corresponding apo-form structures (Figure 12, Appendix E: Table 4).

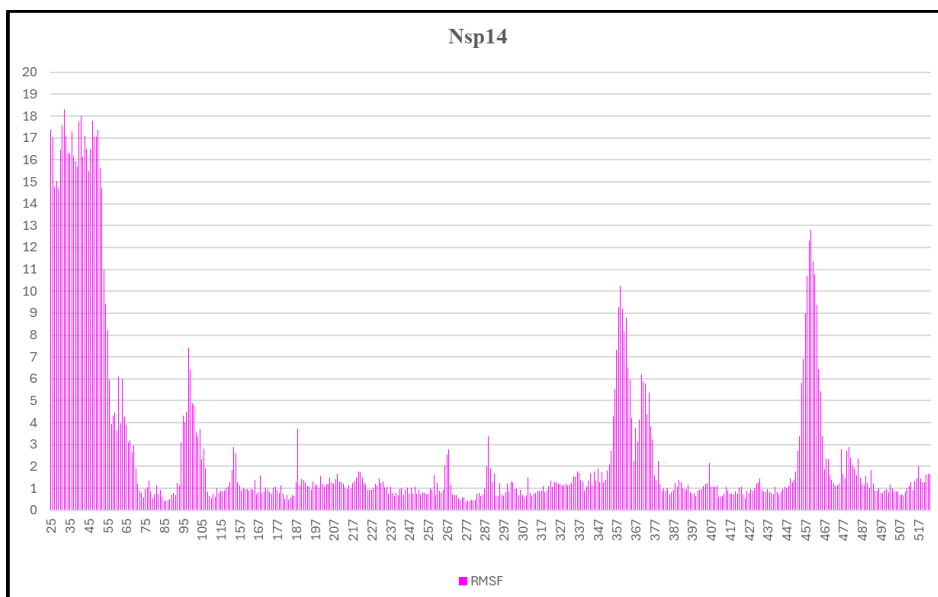


Figure 12. RMSF values (Å) of each residue calculated for Nsp14.

This analysis identified four highly flexible regions in Nsp14: (i) $\eta 1$ to the $\beta 3$ – $\alpha 1$ loop (aa 26–81), hereafter referred to as region I; (ii) $\beta 4$ to $\beta 5$ (aa 86–111), region II; (iii) the $\alpha 1'$ – $\beta 1'$ loop to the αA – $\beta 3'$ loop (aa 321–380), region III; and (iv) βA to $\alpha 2'$ (aa 446–502), region IV (Table 9). These regions exhibited a high degree of flexibility (Figure 8B; Table 8). In addition, seven other regions displayed moderate flexibility: (v) $\beta 6$ to the $\eta 5$ – $\alpha 2$ loop (aa 116–174), region V; (vi) $\beta 10$ – $\alpha 3$ loop to $\beta 14$ (aa 186–237), region VI; (vii) $\alpha 4$ – $\alpha 5$ loop (aa 258–269), region VII; (viii) $\alpha 5$ – $\alpha 1'$ loop (aa

287–301), region VIII; (ix) $\beta 3'$ – $\beta 4'$ loop (aa 386–395), region IX; (x) $\beta 4'$ – αB loop (aa 401–432), region X; and (xi) $\beta 5'$ to the C-terminus (aa 505–523), region XI (Appendix E: Table 3 and 4; Figure 12).

Interestingly, all identified hotspot residues—except for Asp90, Trp186, and Asp273—are located within these flexible regions. Specifically, they are distributed across regions I, II, IV, V, VI, and X (Figures 8B and 9B). Furthermore, the regions exhibiting the highest flexibility (I–IV), together with region VI, correspond to those showing the greatest secondary structure variability (Appendix E: Table 5).

7.2.6. SARS-CoV-2 Nsp16 Hot Spot Residues Prediction.

It has been suggested that the number of computationally predicted hotspots increases proportionally with the size of the interaction interface [43]. This appears to hold true for nsp16, which possesses a relatively large

surface area [32] and contains numerous predicted hotspot residues, although the mutation energy and frequency of occurrence of individual residues vary considerably among different structures (Figures 13–16).

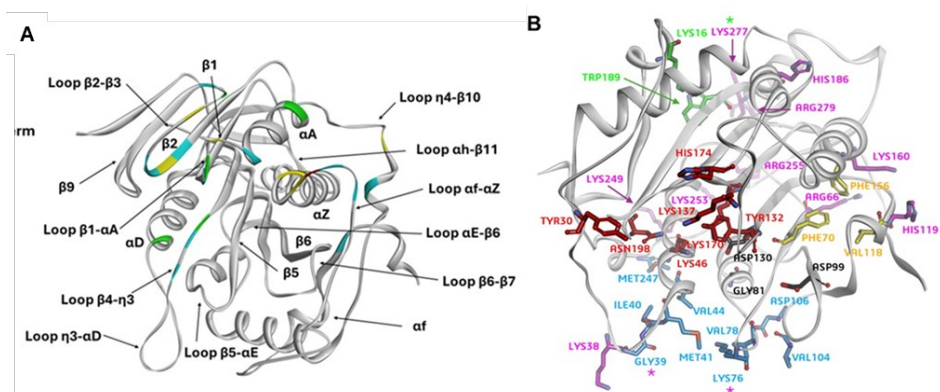


Figure 13. Nsp16 X-ray structure of SARS-CoV-2 nsp16 methyltransferase in complex with nsp10 and the SAM analogue TO383 (only Nsp16 is shown, PDB ID: 8A23). Nsp16 is displayed in solid ribbon and colored in white. A) Protein backbone is colored according to $\Delta\Delta G_{mut}$ values (cyan=very low; green=low; yellow=medium; red=high) (B) Residues that are hot spot are shown in stick and colored accordingly with the binding partner: light blue for nsp10, green for RNA (low affinity binding site), purple for SO_4^{2-} (mimicking the phosphate groups of RNA), black for SAM/SAH/SAM-like inhibitors, yellow for Nsp13 (RTC), and red for RNA (high affinity binding site). The purple asterisks represent two residues (G39 and K76) that are also at the interface with SO_4^{2-} .

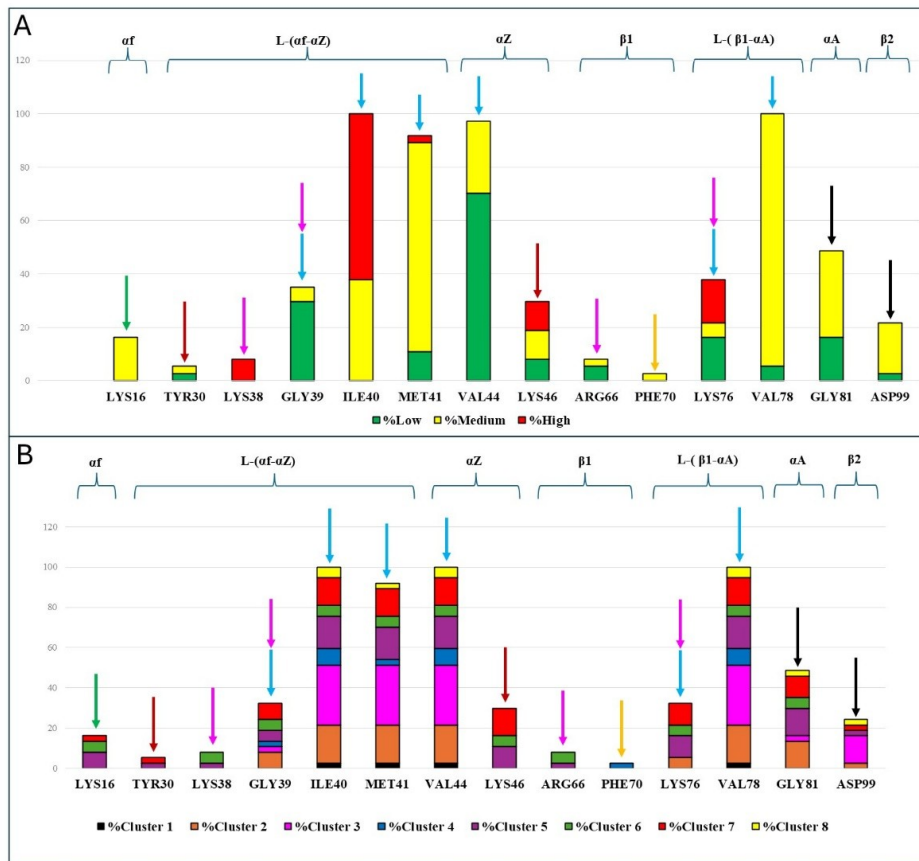


Figure 14. A: percentage of SARS CoV-2 nsp16 predicted hotspot residues in nsp10-nsp16 complexes. Bars are color-coded according to the mutation energy ($\Delta\Delta G_{mut}$): red ($\Delta\Delta G_{mut} \geq 2$ kcal/mol, high), yellow ($1 \leq \Delta\Delta G_{mut} < 2$ kcal/mol, medium), and green ($0.5 \leq \Delta\Delta G_{mut} < 1$ kcal/mol, low). B: percentage of SARS CoV-2 nsp16 predicted hotspot residues in nsp10-nsp16 complexes according to the functional/conformational states. The clusters are defined in Table 2. Arrows indicate the functional ligand involved in the interaction and are colored as follows: blue = nsp10; magenta = SO_4^{2-} ; yellow arrows = allosteric inhibitors; green = low affinity binding site (LBS) RNA; red = high affinity binding site (HBS) RNA; black = SAM/SAH/SAM-like inhibitors.

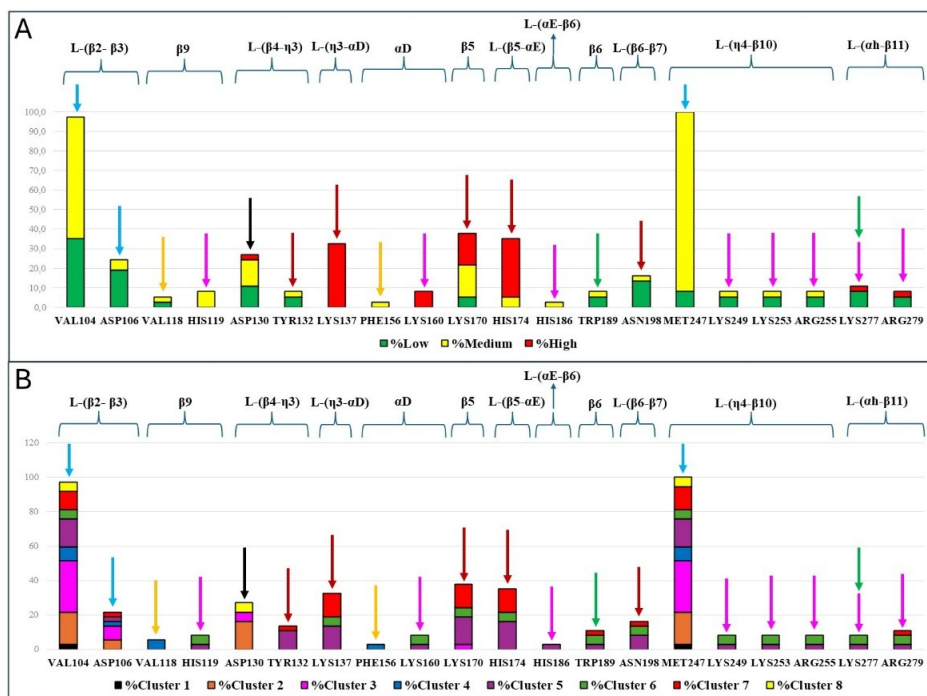


Figure 15. A: percentage of SARS CoV-2 nsp16 predicted hotspot residues in nsp10-nsp16 complexes. Bars are color-coded according to the mutation energy ($\Delta\Delta G_{\text{mut}}$): red ($\Delta\Delta G_{\text{mut}} \geq 2$ kcal/mol, high), yellow ($1 \leq \Delta\Delta G_{\text{mut}} < 2$ kcal/mol, medium), and green ($0.5 \leq \Delta\Delta G_{\text{mut}} < 1$ kcal/mol, low). B: percentage of SARS CoV-2 nsp16 predicted hotspot residues in nsp10-nsp16 complexes according to the functional/conformational states. The clusters are defined in Table 2. Arrows indicate the functional ligand involved in the interaction and are colored as follows: blue = nsp10; magenta = SO_4^{2-} ; yellow arrows = allosteric inhibitors; green = low affinity binding site (LBS) RNA; red = high affinity binding site (HBS) RNA; black = SAM/SAH/SAM-like inhibitors.

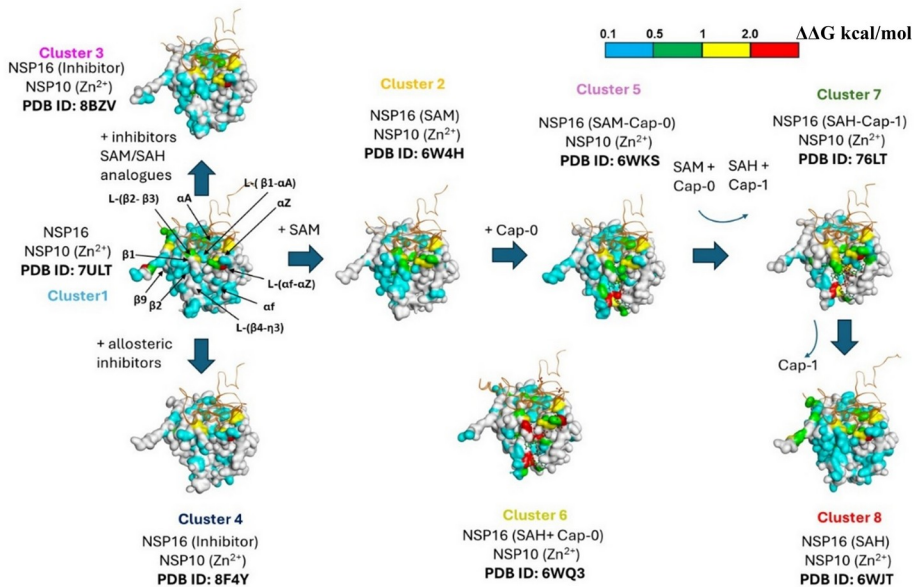


Figure 16. The best resolution structure for each conformational/functional cluster is displayed. The solvent-accessible surface of Nsp16 is displayed in solid and colored on the bases of the following $\Delta\Delta G_{mut}$ range: cyan for residue with $\Delta\Delta G_{mut}$ between 0.1 kcal/mol and 0.5 kcal/mol; green for residues with $\Delta\Delta G_{mut}$ between 0.5 kcal/mol and 1.0 kcal/mol; yellow for residues with $\Delta\Delta G_{mut}$ between 1.0 kcal/mol and 2.0 kcal/mol; red for residues with $\Delta\Delta G_{mut}$ above 2.0 kcal/mol. Nsp10 is displayed as line ribbons and colored in brown.

The other complexes that were extensively studied are the nsp10-nsp16 complexes. So far, there is no unliganded coronaviral nsp16 protein. At the PPI interface, the residues Pro37, Ile40, Val44, Ala45, Thr48, Val84, Leu244, and Met247 of nsp16 accommodates Val42 and Leu45 of nsp10, which deeply immersed into the hydrophobic pocket of nsp16 structure [44]. In our predictions, residues Ile40, Val78, and Met247 of nsp16 are conserved hotspots in all structures with comparatively high $\Delta\Delta G_{bind}$ suggesting the significance of these residues. Met41, Val44, and Val104 of nsp16 are also predicted hotspots in more than 90% of the analyzed

structures (Figures 13A and 14A). Ile40 and Met247 of nsp16 interact with the key hotspot residue Leu45 of nsp10. The long loop between Asp130 to Glu147, projecting outward in the nsp16 structure, contains three hot spot residues, such as, Asp130 (interacting with SAH or SAM) together with Tyr132 and Lys137 interacting with RNA (high affinity binding site) (Figures 13A and 14A). nsp16 Helix-3 (Met41 to Asn53) comprises significant hotspot residues: Met41 and Val44 at the interface with nsp10, and Lys46 in the RNA high affinity binding site, all with comparably higher mutation energy and higher percentage of occurrence.

The first α -helix αf comprises Lys16, located in the low affinity RNA binding site (LBS) [32] which is predicted as hotspot residue only in the structures characterized by the presence of a nucleotide within this site (16%; $\Delta\Delta G_{mut}$: 1.13-1.78 kcal/mol). The long αf - αZ loop (residues 17–41) flanks the αf and αZ helices of Nsp16. Its flexible segment, corresponding to residues Asp26–Lys38, surrounds the nucleotide-binding site (see below for RMSF details, Tyr30). Tyr30 is predicted as a hotspot residue in less than 10% of the analyzed nsp16 complexes—specifically, in two structures belonging to clusters 5 and 7 (Figures 13B and 14B) where it participates in RNA interactions ($\Delta\Delta G_{mut}$: 0.7–1.43 kcal/mol). The residue Ile40 is conserved hotspot residue in all structures with 60% high and 40% medium mutation energy ($\Delta\Delta G_{mut}$: 1.05-2.27 kcal/mol). Met41 is predicted in 90% of the analyzed structures with 78%medium and 11%low mutation energy ($\Delta\Delta G_{mut}$: 0.63-2.05 kcal/mol) (Figures 13A and 14A). In this substructure all the other hotspots residues including Gly39 and Lys38 appear in less than 30% of nsp16 complexes. The αZ helix extends from Met42 to Leu54. This long helix comprises energetically significant hotspot residues, including Val44 that is conserved in 98% of structures with 30%medium and 68%low mutation energy ($\Delta\Delta G_{mut}$: 0.69-1.15

kcal/mol), Lys46 that is predicted in 28% of the complexes belonging only to the functional/conformational clusters 5-7 where this residue is involved in the interaction with the Cap0 substrate ($\Delta\Delta G_{mut}$: 0.58-2.98 kcal/mol) (Figures 13 and 14).

The $\beta 1$ strand contains two predicted hotspot residues. Arg66 was identified as a hotspot in only three of the analyzed structures, with $\Delta\Delta G_{mut}$ values ranging from 0.74 to 1.02 kcal/mol and is consistently involved in interactions with sulfate ions (Figures 13A). Conversely, Phe70 was predicted as a hotspot residue only in structure PDB 8F4S ($\Delta\Delta G_{mut}$ = 1.28 kcal/mol), where it interacts with an allosteric inhibitor.

The adjacent loop region (Gly71–Ala79; loop β – αA) includes two additional significant hotspot residues, Lys76 and Val78. Lys76 was predicted as a hotspot in approximately 40% of the analyzed structures, forming either ionic interactions with the substrate/product or hydrogen bonds with Asn40 of nsp10 ($\Delta\Delta G_{mut}$ = 0.58–3.88 kcal/mol). Val78, conserved across all analyzed structures, exhibits 95% medium and 5% low mutation energy ($\Delta\Delta G_{mut}$ = 0.58–1.75 kcal/mol) (Figures 13A). It is positioned at the αA junction, oriented toward the hotspot residue Val42 of nsp10, with which it establishes alkyl–alkyl interactions (Appendix E: Table 20). The αA helix of nsp16 contains one predicted hotspot residues such as Gly81 which appears in 48% of the analyzed structures ($\Delta\Delta G_{mut}$: 0.60-1.75 kcal/mol) (Figures 13A). This residue establishes a hydrogen bond with the cofactor SAM/SAH (Appendix E: Table 21). The short β -sheet $\beta 2$ (Leu94–Asp99), contains one hotspot residue Asp99 predicted in 24% of structures. Also this residue establishes a hydrogen bond with the cofactor SAM/SAH or inhibitors (Appendix E: Table 22).

Similarly, the loop $\beta 2$ - $\beta 3$ has residues predicted as a hotspot including Val104 and Asp106. Val104 is predicted as hotspots in 36 (97%) of structures with 62% medium and 35% low mutation energy ($\Delta\Delta G_{\text{mut}}$: 0.51-1.66 kcal/mol) (Figures 14A). This residue is involved in the interaction with Nsp10 establishing alkyl-alkyl interaction with Ala71 and Cys77 (Appendix E: Table 15). On the other hand, Asp106 appears in less than 25% of structures, although this residue preserves the same interactions in all the structures (hydrogen bond with Ala71 e Gly94 of nsp10) (Appendix E: Table 24). A substructure of nsp16 between Ser109 to Ser129, comprised of loops, β -sheets and a helix do not contain energetically significant predicted hotspots. The long region characterized by loop $\beta 4$ - $\eta 3$, a 3_{10} -helix ($\eta 3$) and loop $\eta 3$ - αD has three hotspot residues including residue Asp130, Tyr132, and Lys137. Though, except Lys137, others residues appear in less than 25% of structures and seems less significant hotspots. Lys137 is predicted as a hotspot in 32% of structures, all having the substrate (Cap-0) or the product (Cap-1) with $\Delta\Delta G_{\text{mut}}$ values between 5.20 to 6.06 kcal/mol. Indeed, this residue is involved in ionic interactions with Cap-0 or Cap-1. In general, this region has major conformational change, projects outward in the Nsp16 structural arrangement and located opposite to Nsp10 interface region (see RMSF value of Lys137; Table 10).

From the protein structure of the nsp10-nsp16 hetero-dimer solved in the presence of SAH (the byproduct), Asn198 directly coordinates Mg^{2+} , and its substitution by alanine completely abolished the 2'-O MTase activity [45]. Asn198 is a predicted hot spot residue in 16% of analyzed structures. A molecular dynamic simulation study on the nsp10-nsp16 hetero-dimer (PDB ID: 6W4H), showed a relatively large fluctuation the loop Ser74–Pro80, Val104–Cys115, and Asp130–Cys148 near the interface of nsp16 and nsp10 [46]. Accordingly, the predicted hot spot residues Lys76 (at the

nsp10 interface) and Lys137 (high affinity RNA binding site) are characterized by high RMSF values.

Binding free energy analysis [46], crystallographic and biochemical studies in SARS-CoV-2 [47], SARS-CoV and MERS-CoV shows an increase in stability and binding affinity of SAM to nsp16 in the presence of nsp10 [48]. In particular, the interaction of nsp10 Lys93 with nsp16 Ser105 might contribute to SAM binding. Lys93 of nsp10 is predicted as a hot spot residue in both nsp10-nsp14 and nsp10-nsp16 complexes and in all the analyzed functional states.

The experimentally determined structure of the Nsp10-Nsp16 dimer in complex with Sinefungin (PDB: 6YZ1), a panMTase inhibitor, evidences the interaction of nsp16 with the inhibitor through Asp99, Asn101, and Asp114 [49]. In our analysis Asp99 is predicted as a hot spot residue in 20% of the analyzed structures and in most conformational states. Of note, it has been reported that short peptides that disturb the interaction between nsp10 and nsp16 of SARS-CoV [50], mouse hepatitis virus (MHV) [48] could specifically inhibit the 2'-O MTase activity.

Besides the interface region, the conserved nature of the SAM binding site, makes it promising to the development of pan-antiviral inhibitors that could specifically target this druggable pocket.

Different metal ions play a role in the nsp16 2'-O-methyl transfer reaction in a dose dependent manner. Biochemical studies reported magnesium dependent activity of Nsp10-16 2'-O MTase [15,51,52]. In SARS-CoV and MERS-CoV, nsp16 can be activated by Mn^{2+} or Mg^{2+} and to a lesser extent by Ca^{2+} [15,20,53,54]. In SARS CoV-2 both Mg^{2+} and Mn^{2+} increased MTase activity but not Ca^{2+} , and Na^{+} [55]. Moreover, alanine mutation of any residue of the 2'O-MTase catalytic tetrad (Lys46-Asp130-

Lys170-Glu203) results in an inactive enzyme [32,55]. In MERS CoV, it has been further illustrated that alanine substitution of Lys46, Lys170, Glu203, Asn133, Lys38, Tyr47, or Tyr181 concomitantly reduced MTase activity and RNA binding [51]. From recent study, genetically engineered mutated SARS CoV-2 active site D130A mutation resulted in decreased enzymatic activity of SARS CoV-2 NSP16 [56]. Similar study in SARS-CoV-2 also showed that all the active site mutants except for D130A completely destroyed the MTase activity of nsp16 [57]. In our CAS mutagenesis study Lys38, Lys46, and Lys170 are predicted as hot spot residues in the RNA high affinity binding site while Asp130 directly interacts with SAM.

Similarly, it is of particular importance to note that Met131-Asn138 surrounds the nucleotide binding site. The rest of nsp16 substructure do not have energetically important predicted hotspots except residue Met247. Met247 is the starting residue in the long loop $\eta 4\text{-}\beta 10$, and 100% conserved in all complexes with 92%medium and 8%low mutation energies ($\Delta\Delta G_{\text{mut}}$: 0.6-1.46 kcal/mol) (Figures 14A). This residue establishes alkyl-alkyl interaction with Leu45 of nsp10 (predicted as hotspot residue) (Appendix E: Table 25).

7.2.7. Root-Mean-Square Fluctuation (RMSF) of SARS-CoV-2 Nsp16 Hot Spots.

In order to analyze putative allosteric effects also on nsp16, a preliminary assessment of residue flexibility was carried out through RMSF analysis, considering in this case only the analyzed complexes as the corresponding

apo-form structures have not been determined (Figure 17; Appendix E: Table 6).

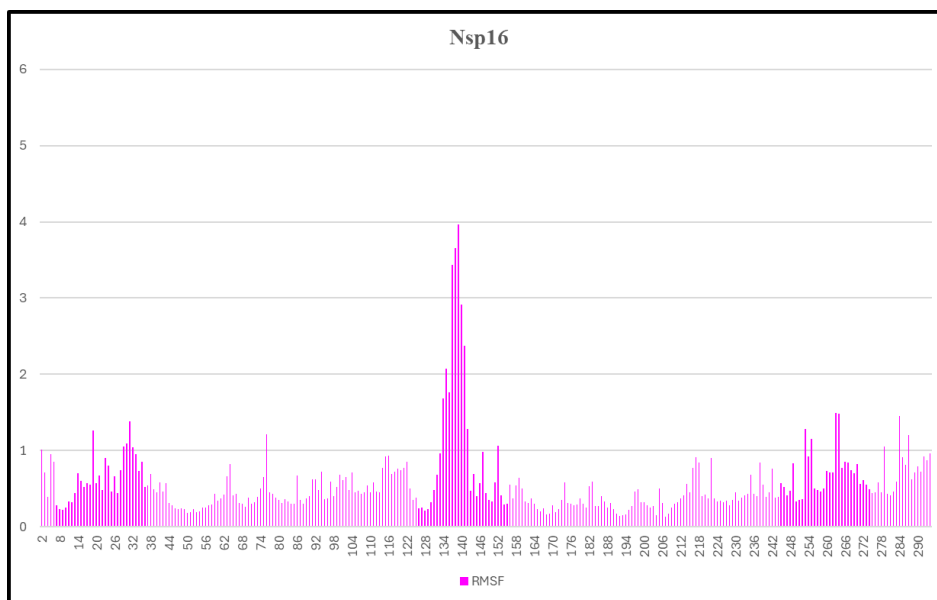


Figure 17. RMSF values (Å) of each residue calculated for Nsp16.

This analysis revealed one highly flexible region in Nsp16, namely, η 3 and the η 3– α D loop (aa134–148; Figure 8C; Table 10). In addition, three other regions showed a moderate degree of flexibility: the α F– α Z loop (aa17–41), the η 4– β 10 loop (aa246–257), and the β 10– η 5 loop/ η 5 (aa261–266) (Figure 8C; Table 10). Interestingly, all these regions, with the exception of the β 10– η 5 loop and η 5, contain hotspot residues (i.e., Tyr30, Lys137, Lys253, and Arg255; Figure 12). Moreover, the η 3– α D loop and the β 10– η 5 loop/ η 5 are associated with changes in secondary structure. Specifically, residues Pro134–Thr136 adopt an α -helix conformation in 93% of the analyzed structures and a coil conformation in the remaining 7%. Similarly, residues Glu264–Glu266 are found in an α -helix

conformation in 70% of the structures and in a coil conformation in 30% (Appendix E: Table 7).

7.2.8. Structural and Functional Comparison with Other Human Coronaviruses (HCoVs)

The structural analysis was conducted by integrating SARS-CoV-2 nsp10, nsp14, and nsp16 structures listed in Table 1 with those reported in Appendix E: Table 8.

In this context, SARS-CoV-2 structures in apo form were included, together with nsp10, nsp14, and nsp16 structures from other human coronaviruses (HCoVs) that shared at least 70% sequence identity with their SARS-CoV-2 counterparts. The non-SARS-CoV-2 structures were initially subjected to clustering based on their functional state, as detailed in Tables 5 and 6. Structural clustering was performed using the *Align Structures* protocol implemented in Discovery Studio 2025 (Dassault Systèmes BIOVIA, San Diego), enabling a systematic comparison of conformational features among homologous proteins. The results of the clustering analysis are summarized in Appendix E: Tables 26–28 and illustrated in Figures 18–20.

Table 5. Functional clusters of the nsp10-nsp14 complexes from other human coronaviruses (HCoVs)

Functional state	PDB ID	Cluster
------------------	--------	---------

Nsp10 bound to nsp14	5C8U (Sars-cov), 5NFY (Sars-cov)	1
Nsp10 bound to nsp14 and SAM	5C8T (Sars-cov)	6*
Nsp10 bound to nsp14, RNA and, SAH	5C8S (Sars-cov)	7*

*New cluster respect to functional clusterization of Sars-cov-2 structures.

Table 6. Functional clusters of the nsp10-nsp16 complexes from other human coronaviruses (HCoV)s

Functional state	PDB ID	Cluster
Nsp10 bound to nsp16	5YN5 (Mers-cov)	1
Nsp10 bound to nsp16 and the cofactor SAM	5YN6 (Mers-cov) 3R24 (Sars-cov)	2
Nsp10 bound to nsp16 and the inhibitors analogues of SAM/SAH	2XYR (Sars-cov) 5YNB (Mers-cov)	3
Nsp10 bound to nsp16, SAM and Cap-0	5YNI (Mers-cov) 5YNM (Mers-cov)	5
Nsp10 bound to nsp16, SAH and Cap-0	5YNO (Mers-cov) 5YNQ (Mers-cov)	6
Nsp10 bound to nsp16 and the cofactor SAH	2XYQ (Sars-cov) 2XYV (Sars-cov) 5YN8 (Mers-cov)	8
Nsp10 bound to nsp16 and Cap-0	5YNF (Mers-cov) 5YNJ (Mers-cov)	9*
Nsp10 bound to nsp16 and the inhibitors analogues of SAM/SAH and Cap-0	5YNP (Mers-cov)	10*

Regarding the nsp10-nsp14 complexes, two structures were assigned to the same functional cluster (Cluster 1), corresponding to the “nsp10 bound to nsp14”, already identified for SARS-CoV-2. In contrast, two structures were classified into new functional clusters: “Nsp10 bound to nsp14 and SAM” (PDB ID: 5C8T; cluster 6) and “Nsp10 bound to nsp14, RNA, and SAH” (PDB ID: 5C8S; cluster 7) (Table 6). Conversely, the majority of nsp10-nsp16 complexes (12 structures) were grouped into the same clusters as those identified for SARS-CoV-2. Only three complexes diverged, forming two new functional clusters: “Nsp10 bound to Nsp16

and Cap-0” (PDB IDs: 5YNF and 5YNJ; Cluster 9) and “Nsp10 bound to Nsp16 and inhibitors analogous to SAM/SAH and Cap-0” (5YNP; Cluster 10) (Table 6).

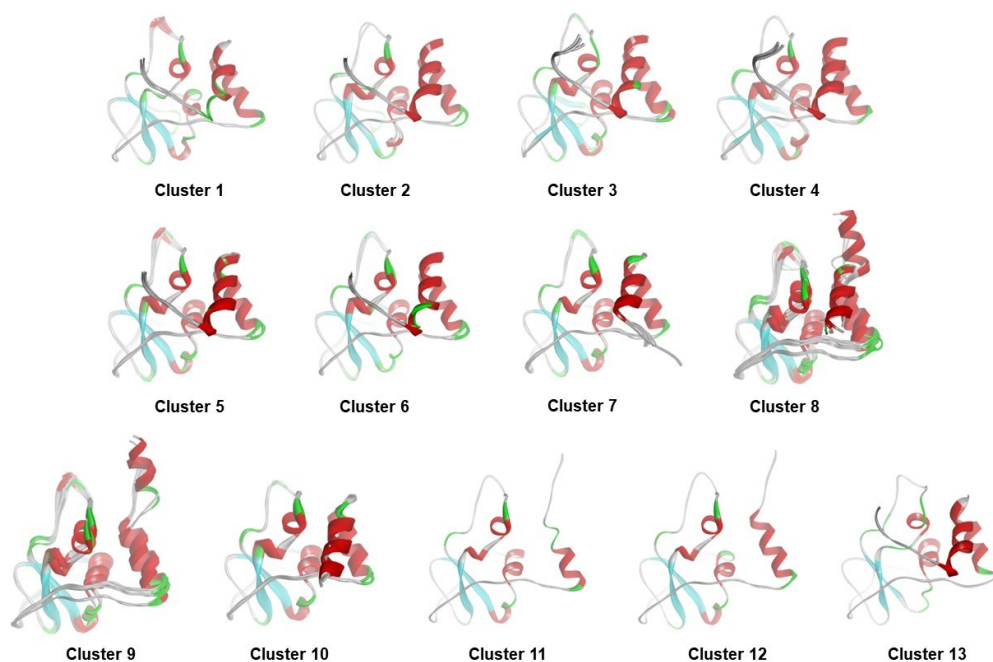


Figure 18. Structural clusters of nsp10. The clusters are defined in Appendix E: Table 26.

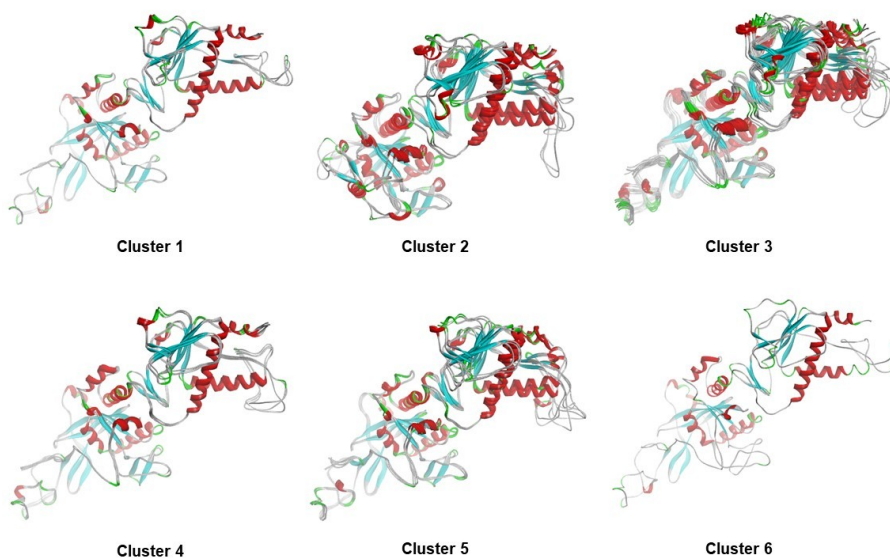


Figure 19. Structural clusters of nsp14. The clusters are defined in Appendix E: 27

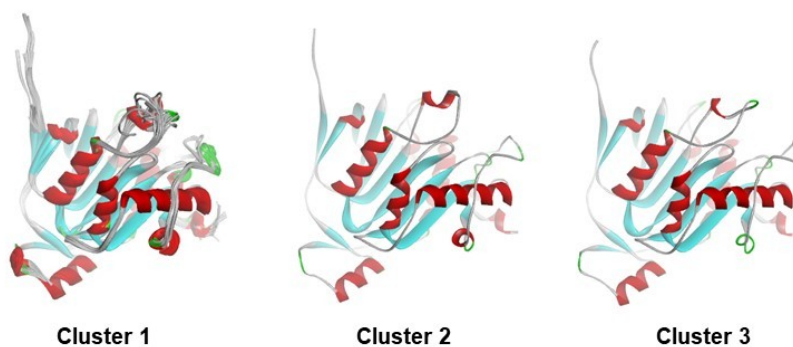


Figure 20. Structural clusters of nsp16. The clusters are defined in Appendix E: Table 28.

Regarding nsp10, the SARS-CoV-2 structures derived from complexes with Nsp14 were distributed across four structural clusters (Clusters 1, 2, 5, and 13; Table 6), in agreement with the four functional clusters previously defined. As for the SARS-CoV-2 nsp10 structures derived from complexes with nsp16, with the exception of structures 7LW3 (Cluster 2)

and 7LW4 (Cluster 3), which exhibit enzyme expansion upon 2'-O methylation (Figure 16) [42], all others were grouped into only two structural clusters (Clusters 8 and 9; Table 5 and 6), reflecting their high structural similarity.

Regarding nsp14, structural clustering revealed six distinct clusters, four of which (Clusters 1, 3, 5, and 6) were consistent with the classification based on functional states (Figure 17; Tables 5). Cluster 2 grouped the apo-form structures of nsp14, while Cluster 4 included the monomers of the dimeric SARS-CoV structures 5C8S, 5C8T, and 5C8U. The remaining monomers from these three structures, along with SARS-CoV structure 5NFY, were grouped, according to the same functional state, with the SARS-CoV-2 structures 7DIY, 7MC5, and 7MC6.

Regarding nsp16, the structural clustering evidenced that the conformations of nsp16 are very similar in the different functional states and among the different viral species. Indeed, all structures were grouped in the same cluster (cluster 1; Table 6) with the only exception of the structures 7LW3 (Cluster 2) and 7LW4 (cluster 3) where expansion of the enzyme upon 2'-O methylation was shown (Figure 20) [47]. In agreement with the RMSF analysis the structures are very similar between them.

Finally, an analysis of the alignment of Nsp10, Nsp14 and Nsp16 of the sequences of most closely related viruses of Sars-cov-2 was performed (Figures 21-23). Interestingly most of the identified hot spot residues falls in "identical" group of residues. We also have crossed this analysis with the hot spot results and their relative "strength" in term of percentage.

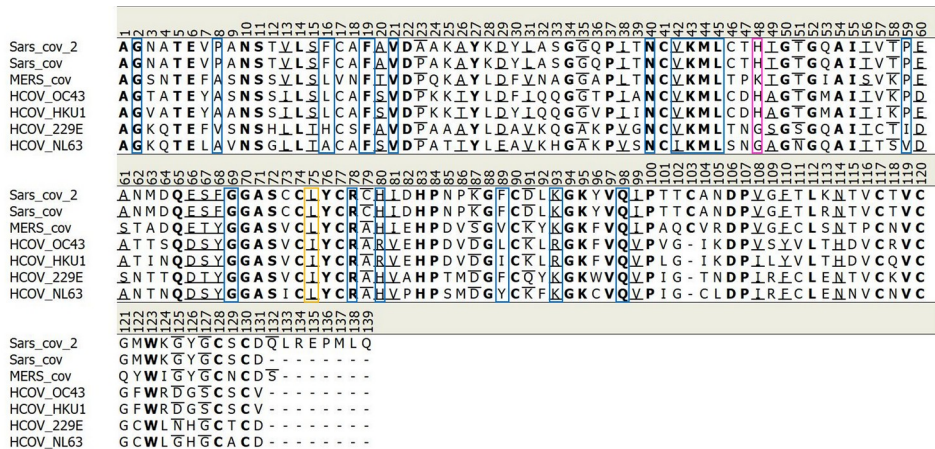


Figure 21. Sequence alignment for nsP10 protein of Sars-cov2, Sars-cov, Mers-cov, HCoVOC43, HCoVHKU1, HCoV229E and HCoVNL63. The residue's font indicates the degree of conservation: Bold = Conserved; Underline = Strong similarity; Overline = Weak similarity. Hot spots are highlighted by colored boxes that are colored according to the interaction partner (see Figure 4).

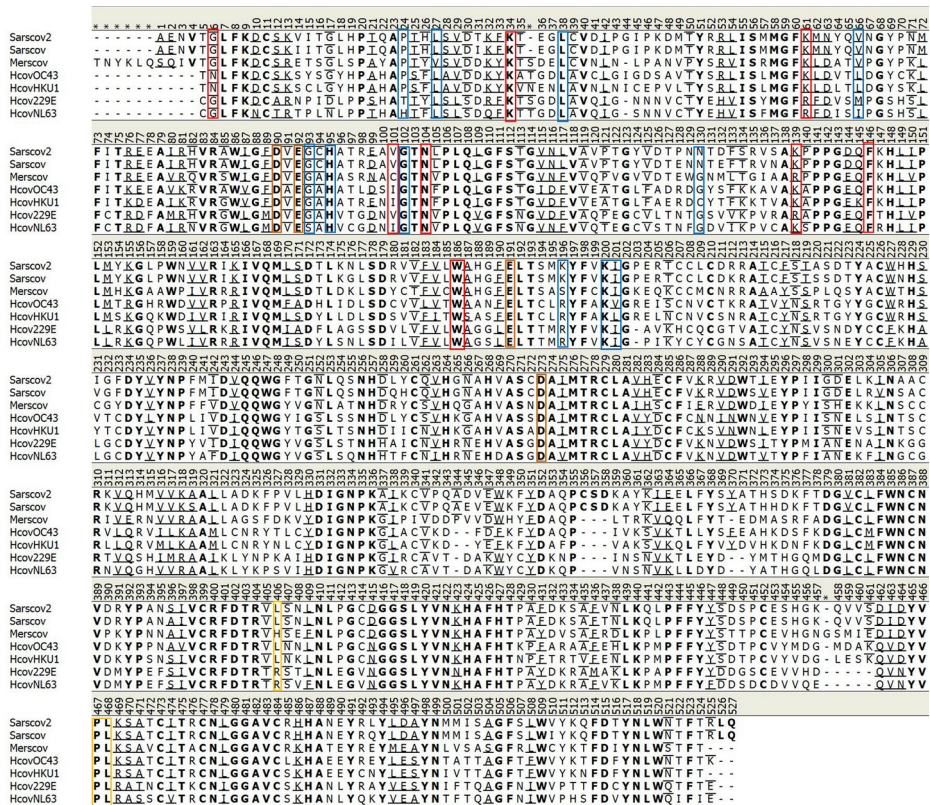


Figure 22. Sequence alignment for nsP14 protein of Sars-cov2, Sars-cov, Mers-cov, HCoVOC43, HCoVHKU1, HCoV229E and HCoVNL63. The residue's font indicates the degree of conservation: Bold = Conserved; Underline = Strong similarity; Overline = Weak similarity. Hot spots are highlighted by colored boxes that are colored according to the interaction partner (see Figure 10).

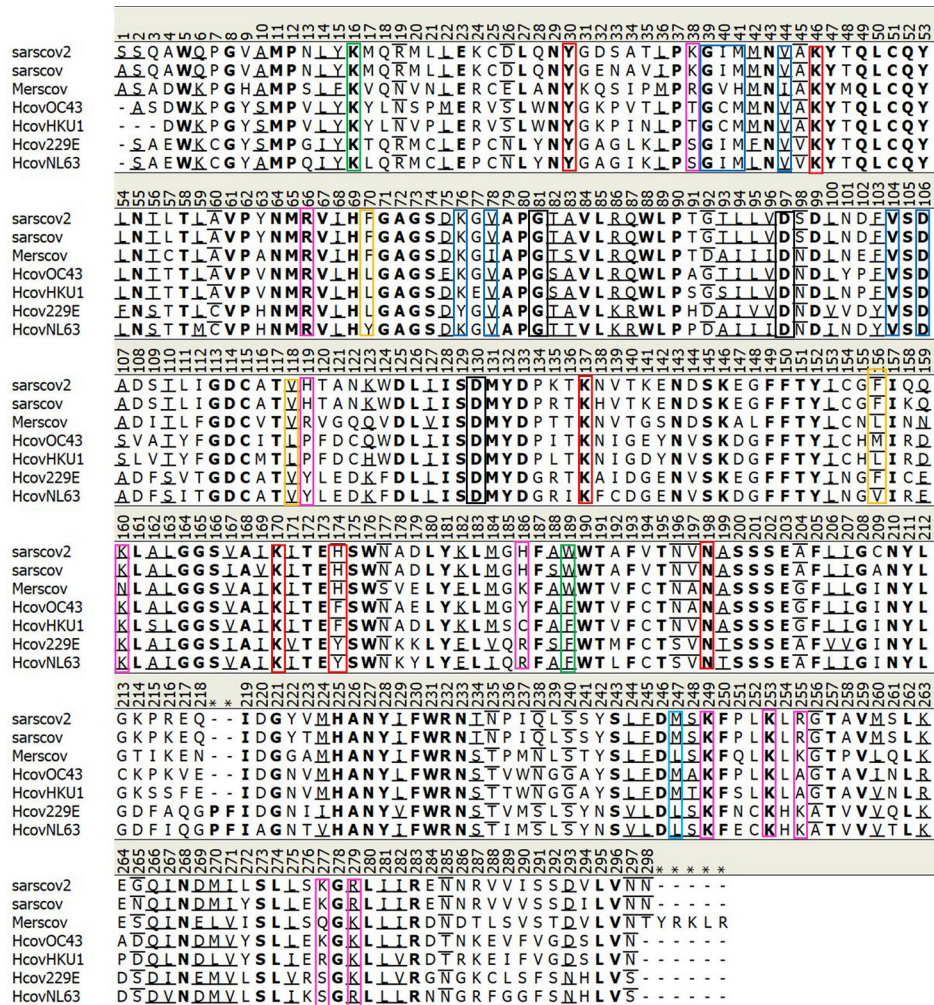


Figure 23. Sequence alignment for nsp16 protein of Sars-cov2, Sars-cov, Mers-cov, HCoVOC43, HCoVHKU1, HCoV229E and HCoVNL63. The residue's font indicates the degree of conservation: Bold = Conserved; Underline = Strong similarity; Overline = Weak similarity. Hot spots are highlighted by colored boxes that are colored according to the interaction partner (see Figure 13).

Regarding nsp10, the analysis showed that, considering the total length of the sequence alignment, 57,5% of the residues on a total of 146 are conserved, in particular 34,2% are identical, 15,8% have strong similarity and 7,5% have a weak similarity. The remaining 42,5% of the residues are not conserved. Considering the identified hot spots of nsp10, on a total of

19 residues identified 68,5% falls in the conserved group, with 47,4% identical and 21,1% with strong similarity. 31.6% of these hot spots are not conserved. Among these Hot Spot residues, Gly2, Phe19, Met44, Gly69, Arg78, His80 and Lys93 are not only in the “identical” group, but are also, in general, the residues with the higher percentage in the structures we considered with also (Gly2, Leu45, Gly69, Lys93) the highest mutation energy. Pro8, Phe16, His48, Pro59, Phe89 and Tyr96 fall in the group of not conserved. Pro8, Phe16 and Pro59, despite 100% of hot spot presence, did not have high energy values. His48 is present as hot spot only in the complex with Nsp16 at 10%.

About nsp14, the analysis showed that, considering the total length of the sequence alignment, 64,6% of the residues on a total of 535 are conserved. In particular 35,9 are identical, 19,1 have strong similarities and 9,6 have weak similarities. The remaining 31,6% of the residues are not conserved. Considering the identified hot spots of nsp14, on a total of 33 residues identified 78,8% falls in the conserved group, with 51,5% identical, 21,2 with strong similarities and 6,1 % with weak similarity. 21,2% of these hot spots are not conserved. Among the identical residues Lys9, Leu38, Asp90, Glu92, Lys200 and Asp273 are the residues with the highest percentage in the structure we considered with also at least half of the percentage as “high”. Also, here most of the not conserved hot spot residues have a low (20%-30%) percentage of presence in the structure or have medium or low energy only.

The analysis of Nsp16 showed that, considering the total length of the sequence alignment, 68,2% of the residues on a total of 305 are conserved. 40,3% are identical, 21,3 % have strong similarities and 6,6% have weak similarities. The remaining 31,8% of the residues are not conserved.

Considering the identified hot spots of nsp14, on a total of 33 residues identified 73,5 % falls in the conserved group, with 44,1 % identical, 21,5% with strong similarities and 5,9 % with weak similarity. 26,5 % of these hot spots are not conserved. Among the identical residues the most notable are Val104 with almost 100 percent presence as hot spot, and Lys137 that has 30% presence as hot spot with all high energy. For the group of not conserved hot spots, the generally have a low percentage of presence as hot spot with low or medium energy.

7.3. Conclusions.

The design of peptidomimetic therapeutics may be challenging due to the relatively large interface area involved in protein-protein interactions. The identification of hot regions formed by small clusters of hot spot residues opened new perspective in this challenge. Some of these hot spots may be involved in the alternative or simultaneous binding of different protein functional ligands and maybe present only in specific conformational/functional states. In this scenario, the results of the present CAS mutagenesis study, performed on the full-length sequence of all available experimentally determined structures of Sars-CoV-2 nsp10, nsp14, and nsp16, allowed to draw a comprehensive picture of the druggable binding sites of these key players of coronavirus transcription and replication complex. Our results are in line with experimental data reported in literature and provide additional information with respect to previously reported CAS analyses. Apart from the interacting residues hot spots are also predicted in the nearby regions and as expected, hot regions may include binding sites for alternative or simultaneous ligands, according to the protein functional state, providing essential tools for the design strategy. Finally, the presence, among the predicated hot spot

residues, of functional residues conserved across different coronaviruses allows ligands potentially endowed with the desired target specificity.

7.5. References.

1. Lise, S., et al. "Prediction of hot spot residues at protein–protein interfaces by combining machine learning and energy-based methods." *BMC Bioinformatics*, vol. 10, no. 1, 2009, p. 365.
2. DeLano, W. L. "Unraveling hot spots in binding interfaces: progress and challenges." *Current Opinion in Structural Biology*, vol. 12, no. 1, 2002, pp. 14-20.
3. Cunningham, B. C., and J. A. Wells. "Comparison of a structural and a functional epitope." *Journal of Molecular Biology*, vol. 234, no. 2, 1993, pp. 554-563.
4. Kozakov, Dima, et al. "Structural Conservation of Druggable Hot Spots in Protein–Protein Interfaces." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 33, 2011, pp. 13528–33.
5. Wells, J. A. "Additivity of mutational effects in proteins." *Biochemistry*, vol. 29, no. 36, 1990, pp. 8509-8517.
6. Guharoy, G., and P. Chakrabarti. "Empirical energy-based analysis of protein–protein interfaces identifies strongly interacting residues." *Journal of Computer-Aided Molecular Design*, vol. 23, no. 11, 2009, pp. 895-902.
7. Smilova, Mihaela D., et al. "Fragment Hotspot Mapping to Identify Selectivity-Determining Regions between Related Proteins." *Journal of Chemical Information and Modeling*, vol. 62, no. 2, 2022, pp. 284-294
8. Wu, F., et al. "A New Coronavirus Associated with Human Respiratory Disease in China." *Nature*, vol. 579, 2020, pp. 265-69.
9. Da Silva, S. J. R., et al. "Role of Nonstructural Proteins in the Pathogenesis of SARS-CoV" *Journal of Medical Virology*, vol. 92, 2020, pp. 1427-35.
10. 3.Arya, R., et al. "Structural Insights into SARS-CoV-2 Proteins." *Journal of Molecular Biology*, vol. 433, 2021, article 166725.
11. Malone, B., et al. "Structures and Functions of Coronavirus Replication–Transcription Complexes and Their Relevance for SARS-CoV-2 Drug Design." *Nature Reviews Molecular Cell Biology*, vol. 23, 2022, pp. 21-39.
12. Yan, Liming, et al. "Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading." *Cell* 184.13 (2021): 3474-3485. Gao, Y., et al. "Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus." *Science*, vol. 368, 2020, pp. 779-82.

13. Yan, L., et al. "A Mechanism for SARS-CoV-2 RNA Capping and Its Inhibition by Nucleotide Analog Inhibitors." *Cell*, vol. 185, 2022, pp. 4347-60.
14. Kikkert, M. "Innate Immune Evasion by Human Respiratory RNA Viruses." *Journal of Innate Immunity*, vol. 12, 2020, pp. 4-21.
15. Kasuga, Y., et al. "Innate Immune Sensing of Coronavirus and Viral Evasion Strategies." *Experimental & Molecular Medicine*, vol. 53, 2021, pp. 723-36.
16. Liang, J., et al. "How the Replication and Transcription Complex Functions in Jumping Transcription of SARS-CoV-2." *Frontiers in Genetics*, vol. 13, 2022, article 904513.
17. Yan, L., et al. "Coupling of N7-Methyltransferase and 3'-5' Exoribonuclease with SARS-CoV-2 Polymerase Reveals Mechanisms for Capping and Proofreading." *Cell*, vol. 184, 2021, pp. 3474-85.
18. Park, G. J., et al. "The Mechanism of RNA Capping by SARS-CoV-2." *Nature*, vol. 609, no. 7928, 2022, pp. 793-98.
19. Yan, W., et al. "Structural Biology of SARS-CoV-2: Open the Door for Novel Therapies." *Signal Transduction and Targeted Therapy*, vol. 7, 2022, article 26.
20. Bouvet, M., et al. "Coronavirus Nsp10, a Critical Co-factor for Activation of Multiple Replicative Enzymes." *Journal of Biological Chemistry*, vol. 289, 2014, pp. 25783-96.
21. Decroly, E., et al. "Coronavirus Nonstructural Protein 16 Is a Cap-0 Binding Enzyme Possessing (Nucleoside-2' O)-Methyltransferase Activity." *Journal of Virology*, vol. 82, 2008, pp. 8071-85.
22. Ma, Y., et al. "Structural Basis and Functional Analysis of the SARS Coronavirus nsp14-nsp10 Complex." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, 2015, pp. 9436-41.
23. Imprachim, N., et al. "Crystal Structures and Fragment Screening of SARS-CoV-2 NSP14 Reveal Details of Exoribonuclease Activation and mRNA Capping and Provide Starting Points for Antiviral Drug Development." *Nucleic Acids Research*, vol. 51, 2023, pp. 475-92.
24. Saramago, M., et al. "New Targets for Drug Design: Importance of nsp14/nsp10 Complex Formation for the 3'-5' Exoribonucleolytic Activity on SARS-CoV-2." *FEBS Journal*, vol. 288, 2021, pp. 5130-46.
25. Inniss, N. L., et al. "Discovery of a Druggable, Cryptic Pocket in SARS-CoV-2 nsp16 Using Allosteric Inhibitors." *ACS Infectious Diseases*, vol. 9, 2023, pp. 1918-31.
26. Wang, H., et al. "Emerging Variants of SARS-CoV-2 NSP10 Highlight Strong Functional Conservation of Its Binding to Two Non-structural Proteins, NSP14 and NSP16." *Elife*, vol. 12, 2023, article RP87884.

27. Takada, K., et al. "Genomic Diversity of SARS-CoV-2 Can Be Accelerated by Mutations in the nsp14 Gene." *iScience*, vol. 26, no. 3, 2023, article 106210
28. J. Liang, J. Shi, S. Chen, G. Duan, F. Yang, Z. Cheng, X. Li, J. Ruan, D. Mi, S. Gao, "How the replication and transcription complex functions in jumping transcription of SARS-CoV-2", *Front. Genet.* **2022**, *13*, 904513.
29. Y. Chen, H. Cai, J. A. Pan, N. Xiang, P. Tien, T. Ahola, D. Guo, "Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase", *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3484.
30. N. S. Ogando, P. El Kazzi, J. C. Zevenhoven-Dobbe, B. W. Bontes, A. Decombe, C. C. Posthuma, V. Thiel, B. Canard, F. Ferron, E. Decroly, "Structure–function analysis of the nsp14 N7–guanine methyltransferase reveals an essential role in Betacoronavirus replication", *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2108709118.
31. Rosas-Lemus, M., et al. "High-Resolution Structures of the SARS-CoV-2 2'-O-Methyltransferase Reveal Strategies for Structure-Based Inhibitor Design." *Science Signaling*, vol. 13, 2020, article eabe1202.
32. Hsu, J. C.-C., et al. "Translational Shutdown and Evasion of the Innate Immune Response by SARS-CoV-2 NSP14 Protein." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, 2021, article e2101161118.
33. Lugari, A., et al. "Molecular Mapping of the RNA Cap 2'-O-Methyltransferase Activation Interface between Severe Acute Respiratory Syndrome Coronavirus nsp10 and nsp16." *Journal of Biological Chemistry*, vol. 285, 2010, pp. 33230-41.
34. Ke, M., et al. "Short Peptides Derived from the Interaction Domain of SARS Coronavirus Nonstructural Protein nsp10 Can Suppress the 2'-O-Methyltransferase Activity of nsp10/nsp16 Complex." *Virus Research*, vol. 167, 2012, pp. 322-29.
35. Agostini, M. L., et al. "Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease." *mBio*, vol. 9, no. 2, 2018, article e00221-18.
36. Jockusch, S., et al. "Sofosbuvir Terminated RNA Is More Resistant to SARS-CoV-2 Proofreader than RNA Terminated by Remdesivir." *Scientific Reports*, vol. 10, 2020, article 16577.
37. Czarna, A., et al. "Refolding of Lid Subdomain of SARS-CoV-2 nsp14 upon nsp10 Interaction Releases Exonuclease Activity." *Structure*, vol. 30, 2022, pp. 1050-60.
38. Moeller, N. H., et al. "Structure and Dynamics of SARS-CoV-2 Proofreading Exoribonuclease ExoN." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 119, 2022, article e2106379119.

39. Lin, S., et al. "Crystal Structure of SARS-CoV-2 nsp10 Bound to nsp14-ExoN Domain Reveals an Exoribonuclease with Both Structural and Functional Integrity." *Nucleic Acids Research*, vol. 49, 2021, pp. 5382-92.
40. Sarma, H., and G. N. Sastry. "A Computational Study on the Interaction of NSP10 and NSP14: Unraveling the RNA Synthesis Proofreading Mechanism in SARS-CoV-2, SARS-CoV, and MERS-CoV." *ACS Omega*, vol. 7, 2022, pp. 29995-30006.
41. Arabi-Jeshvaghani, F., et al. "Analysis of Critical Protein-Protein Interactions of SARS-CoV-2 Capping and Proofreading Molecular Machineries towards Designing Dual Target Inhibitory Peptides." *Scientific Reports*, vol. 13, 2023, article 350.
42. Liu, C., et al. "Structural Basis of Mismatch Recognition by a SARS-CoV-2 Proofreading Enzyme." *Science*, vol. 373, 2021, pp. 1142-46.
43. Keskin, O., et al. "Hot Regions in Protein-Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues." *Journal of Molecular Biology*, vol. 345, 2005, pp. 1281-94.
44. Krafcikova, P., et al. "Structural Analysis of the SARS-CoV-2 Methyltransferase Complex Involved in RNA Cap Creation Bound to Sinefungin." *Nature Communications*, vol. 11, 2020, article 3717.
45. Viswanathan, T., et al. "A Metal Ion Orients SARS-CoV-2 mRNA to Ensure Accurate 2'-O Methylation of Its First Nucleotide." *Nature Communications*, vol. 12, 2021, article 3287.
46. Sk, M. F., et al. "Computational Investigation of Structural Dynamics of SARS-CoV-2 Methyltransferase-Stimulatory Factor Heterodimer nsp16/nsp10 Bound to the Cofactor SAM." *Frontiers in Molecular Biosciences*, vol. 7, 2020, article 590165.
47. Viswanathan, T., et al. "Structural Basis of RNA Cap Modification by SARS-CoV-2." *Nature Communications*, vol. 11, 2020, article 3718.
48. Wang, Y., et al. "Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide in Vitro and in Vivo to Reduce Replication and Pathogenesis." *Journal of Virology*, vol. 89, 2015, pp. 8416-27.
49. Martin, B., et al. "The Methyltransferase Domain of the Sudan Ebolavirus L Protein Specifically Targets Internal Adenosines of RNA Substrates, in Addition to the Cap Structure." *Nucleic Acids Research*, vol. 46, 2018, pp. 7902-12.
50. Ke, M., et al. "Short Peptides Derived from the Interaction Domain of SARS Coronavirus Nonstructural Protein nsp10 Can Suppress the 2'-O-Methyltransferase Activity of nsp10/nsp16 Complex." *Virus Research*, vol. 167, 2012, pp. 322-29.
51. Martin, B., et al. "The Methyltransferase Domain of the Sudan Ebolavirus L Protein Specifically Targets Internal Adenosines of RNA Substrates, in Addition to the Cap Structure." *Nucleic Acids Research*, vol. 46, 2018, pp. 7902-12.

52. Aouadi, W., et al. "Binding of the Methyl Donor S-Adenosyl-L-Methionine to Middle East Respiratory Syndrome Coronavirus 2'-O-Methyltransferase nsp16 Promotes Recruitment of the Allosteric Activator nsp10." *Journal of Virology*, vol. 91, no. 5, 2017, article e02217-16.
53. Decroly, E., et al. "Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2'-O-Methyltransferase nsp10/nsp16 Complex." *PLoS Pathogens*, vol. 7, 2011, article e1002059.
54. Bouvet, M., et al. "RNA 3'-End Mismatch Excision by the Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10/nsp14 Exoribonuclease Complex." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, 2012, pp. 9372-77.
55. Minasov, G., et al. "Mn(2+) Coordinates Cap-0-RNA to Align Substrates for Efficient 2'-O-Methyl Transfer by SARS-CoV-2 nsp16." *Science Signaling*, vol. 14, 2021, article eabh2071.
56. Schindewolf, C., et al. "SARS-CoV-2 Uses Nonstructural Protein 16 to Evade Restriction by IFIT1 and IFIT3." *Journal of Virology*, vol. 97, 2023, article e01532-22.
57. He, M., et al. "Reconstitution of RNA Cap Methylation Reveals Different Features of SARS-CoV-2 and SARS-CoV Methyltransferases." *Journal of Medical Virology*, vol. 96, 2024, article e29411.

Chapter 8. Rational PROTAC Design and Modelling

8.1. Introduction.

In this chapter, I will describe the new skills I acquired and the results I achieved during my six-month stay at the Department of Chemistry,

Biology and Biotechnology of the University of Perugia, from 16th April to 31st September 2025. During this period, I participated in a research project under the supervision of Prof. Gabriele Cruciani also scientific director of Molecular Discovery Ltd. (<https://www.moldiscovery.com/>) and Molecular Horizon S.r.l. (<https://www.molhorizon.it/>).

The project focused on the rational design of new PROTAC linkers connecting an E3 ligase binder with a protein of interest, the bromodomains belonging to the BET family. The computational approach was based on the use of Molecular Interaction Fields (MIFs), and the software tools applied included RAINDROP, developed by the research group of Prof. Cruciani.

I was directly involved in the development and application of the RAINDROP procedure, as well as in the analysis and interpretation of the results. I also participated in the regular meetings of Prof. Cruciani's research group, both at the University of Perugia and at Molecular Discovery Ltd., where I had the opportunity to interact with other researchers working on related topics, including PROTAC design, cheminformatics, and drug metabolism prediction.

Additionally, I attended internal seminars such as "Mass spectrometry: principles and applications to Omics" held by Prof. Maria Fedorova (Technische Universität Dresden) from 13th to 17th May, in which I learned the application of mass spectrometry in omics fields. Moreover, I participated to the school organized by Prof. Cruciani's Group and Molecular Discovery Ltd. "Spring Italian Training for AI in Drug Design (SPRINT-AIDD)". Molecular Discovery Ltd. Bettona, Perugia, Italy. 25th-28th May 2025 in which I learned new predictive methods for physiochemical properties chemical compounds by using software

produced by Molecular Discovery Ltd. such as FLAP, Volsurf and Metasite.

8.1.1. Proteolysis Targeting Chimera (PROTACs).

As a traditional treatment method, chemotherapy plays an irreplaceable role in the cancer treatment process. However, traditional anticancer drugs face significant disadvantages, including poor selectivity and the tendency to easily develop drug resistance [1-3]. Consequently, targeted therapy for cancer has attracted considerable attention [4-6]. The discovery of new targets and Small Molecule Inhibitors (SMIs) has emerged as a powerful treatment strategy [7]. In particular, the development of small molecule kinase inhibitors has achieved great success in cancer treatment [8]. Nevertheless, this targeted therapy strategy, much like chemotherapy, faces the persistent problem of drug resistance, which remains the main limitation urgently requiring resolution in cancer therapy [2,9].

In recent years, a novel strategy targeting disease-related proteins for degradation has gained attention. This strategy involves Proteolysis Targeting Chimeras (PROTACs), which are also known as bivalent chemical protein degraders. PROTACs are heterobifunctional molecules designed to degrade specific endogenous proteins through the E3 ubiquitin ligase pathway [10].

Structurally, a PROTAC molecule connects a protein of interest (POI)-binding ligand and an E3 ubiquitin ligase (E3) ligand using an appropriate linker [11-14]. PROTACs execute their mechanism of action by utilizing the ubiquitin-proteasome system (UPS) to ubiquitinate and degrade the target protein (Figure 1).

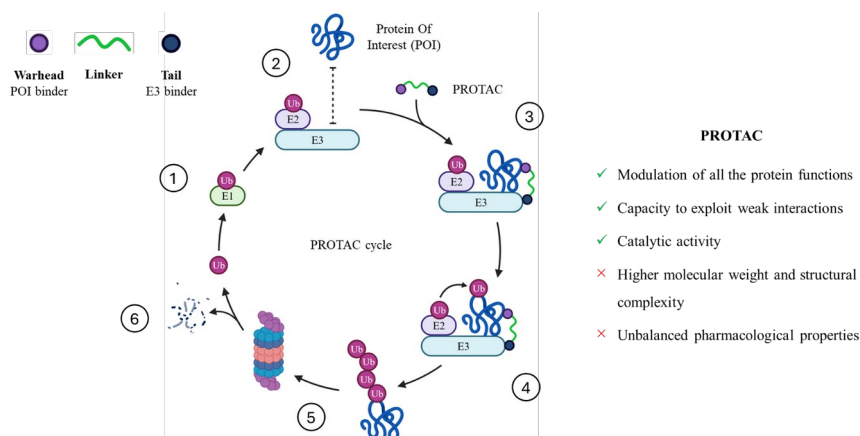


Figure 1. Schematic representation of PROTAC mechanism of action and main advantages and disadvantages of the PROTAC approach. Figure taken from: Thomas, Bedwyr ab Ion, et al. "Central nervous system targeted protein degraders." *Biomolecules* 13.8 (2023): 1164.

The UPS is a crucial mechanism for specifically degrading proteins, and it is vital for maintaining the homeostasis of cell proteins and regulating numerous cellular processes, such as cell cycle control and apoptosis [15]. The UPS relies on ATP and consists of two steps: the polyubiquitination of the target protein and the proteolysis of the polyubiquitin by the 26S proteolytic enzyme complex [16]. Once PROTAC molecules simultaneously combine the target protein with the E3 ligase, they form a ternary complex. This complex induces the E3 ligase to ubiquitinate the target protein, initiating the degradation process. The ubiquitinated target protein is then recognized and degraded by the 26S proteasome [17].

The potential benefits of PROTAC technology may compensate for the shortcomings of traditional drug therapy [18,19]. One of the biggest advantages of PROTAC technology is its potential to target proteins previously considered "undruggable". Unlike SMIs, which inhibit the biological activity of protein targets, PROTACs function by degrading the target proteins. This mechanism means that the ligand of the target protein

in a PROTAC does not necessarily have to bind to the active site of the protein, thereby overcoming a key limitation of most SMIs [20-22].

Crucially, PROTAC technology demonstrates a great superiority in overcoming resistance caused by target mutation or overexpression when compared with SMIs. This technology has been applied to degrade various targets, including Androgen Receptor (AR), Estrogen Receptor (ER), BTK, BET, and BCR-ABL, specifically to overcome drug resistance [19].

Research into PROTAC technology has rapidly demonstrated its viability in degrading target proteins. This strategy has achieved remarkable progress since its concept was first developed in 2001. Currently, several PROTACs have advanced into clinical trials. The first oral PROTACs, **ARV-110** (Figure 1) and its derivative **ARV-471**, have shown encouraging results in clinical trials for prostate and breast cancer treatment, respectively [23].

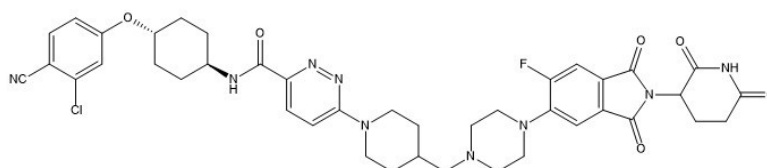


Figure 1. Chemical structure of the oral PROTAC ARV-110. The chirality of the stereocenters is crucial for its biological activity.

Despite these achievements, challenges remain in the design and application of PROTACs. Although there are more than 600 E3 ubiquitin ligases, fewer than 10 have been successfully utilized to degrade target proteins using current PROTAC technology (e.g., CRBN, VHL, MDM2, and IAPs).

The effectiveness of a PROTAC is highly dependent not only on the ligands for the POI and E3 ligases but also on the length and chemical properties of the linker connecting them. Moreover, safety is another

challenge, as PROTACs may have off-target effects, though high selectivity may be achieved by utilizing tissue-specific E3 protein expression.

8.1.2. Molecular Interaction Fields.

One of the most established and versatile concepts in drug discovery is the Molecular Interaction Field (MIF). MIFs were specifically conceived to describe molecular interactions of pharmaceutical relevance. A MIF describes the spatial variation of the interaction energy between a molecular target and a chosen probe. This computational procedure, first described by P. J. Goodford [24], involves calculating the interaction of a probe group (such as water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl) with a protein of known structure at sample positions within a grid throughout and around the macromolecule (Figure 2).

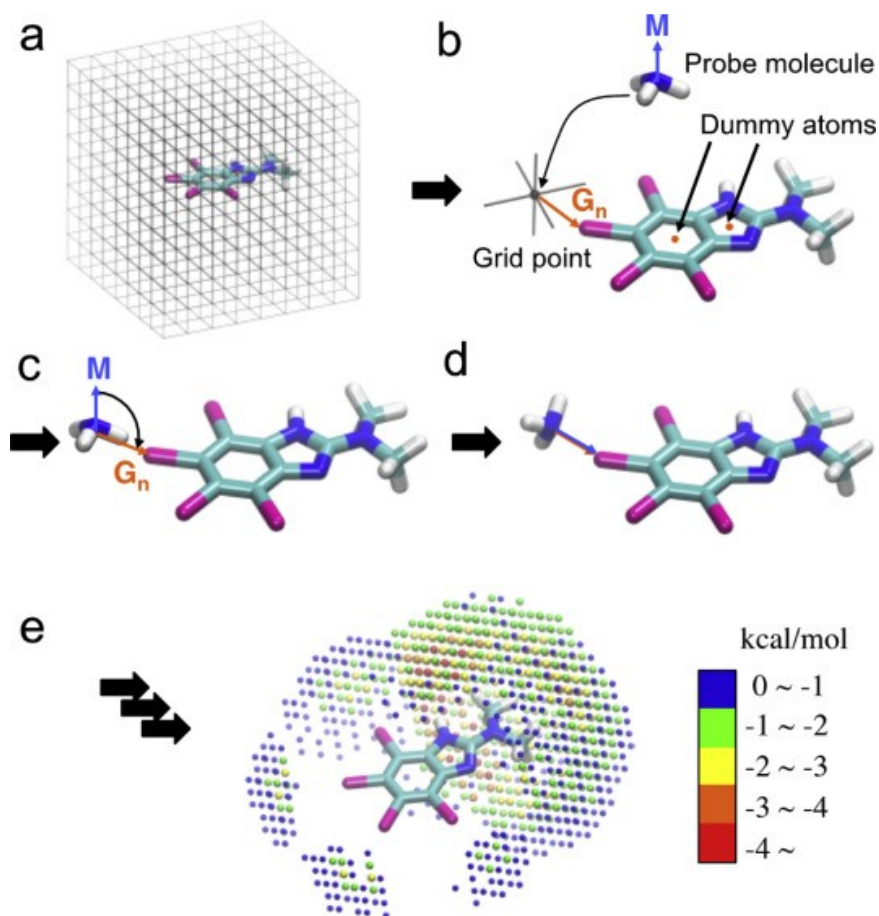


Figure 2. An example of MIF calculation (a-d) and result (e).

The methodology is implemented in software like the GRID program. GRID yields an array of energy values, which are displayed as three-dimensional (3D) energy contour surfaces together with the protein structure. The process involves the following steps: i) a regular array of "GRID points" is established throughout and around the protein. The distance between grid points can be varied and defines the resolution of the calculated interactions. ii) the potential energy of a selected probe, which may be water, an atom, or a group (e.g., methyl group, hydroxyl group, amine nitrogen, carboxy oxygen.) is calculated at each GRID point. In the

GRID program, the nonbonded interaction energy is computed as the sum of three different components.

$$E_{xyz} = \sum E_{lj} + \sum E_{el} + \sum E_{hb}$$

The first term (Lennard-Jones) describes repulsion and dispersion forces. Repulsion dominates when the distance is small (large positive), while attraction (from the term) results in modest negative energies. The second term (Electrostatic) is calculated pairwise between the probe group and every protein atom, irrespective of their separation, as it does not diminish rapidly with distance. The magnitude of this term is critically sensitive to the spatial dielectric behavior of the environment. The third term (Hydrogen bond) is a direction-dependent function. For maximal interaction, the probe is rotated at each grid point to optimize the hydrogen bond energy. If the Lennard-Jones term indicates repulsion but the hydrogen bond term is favorable, the latter is set to zero. Contours at negative energy levels delineate regions of attraction between the probe and the protein. These favorable sites are often found at known ligand binding clefts, but they also enable other regions of attraction to be identified, thereby facilitating the interpretation of protein-ligand energetics and providing significant value for drug design. This is of importance in the case of PROTACs, where the identification of new possible site of binding is crucial in the design of the warhead that will bind the protein of interest.

8.1.3. Rational Design of PROTACs using the RAINDROP procedure

The field of rational drug design has evolved significantly with the emergence of PROTACs as a promising therapeutic strategy. Their successful design requires accurate prediction of the structure and stability of the ternary complex formed by the PROTAC molecule, the E3 ubiquitin ligase, and the protein of interest (POI). The *RAtional DRug design Of Protac* (RAINDROP) is a software designed to achieve this goal.

RAINDROP integrates structural, energetic, and conformational analyses to predict PROTAC structures capable of promoting efficient ternary complex formation. The process begins with the preparation of structural inputs for both the E3 ligase–ligand and the target protein–ligand complexes. These models are used to define the spatial orientation and physicochemical context in which the linker will operate.

The linker plays a pivotal role in determining the ternary complex geometry. In RAINDROP, linkers are defined through their chemical formula expressed in SMILES notation and are subjected to an extensive Conformational Sampling phase. The procedure systematically explores the rotatable bonds along the linker to generate thousands of conformational hypotheses consistent with the steric and energetic constraints of both protein–ligand complexes. Each atom of the linker is associated with two distinct conformational spaces, one related to the E3 ligase and one to the target protein, allowing a detailed mapping of compatibility between the two binding sites.

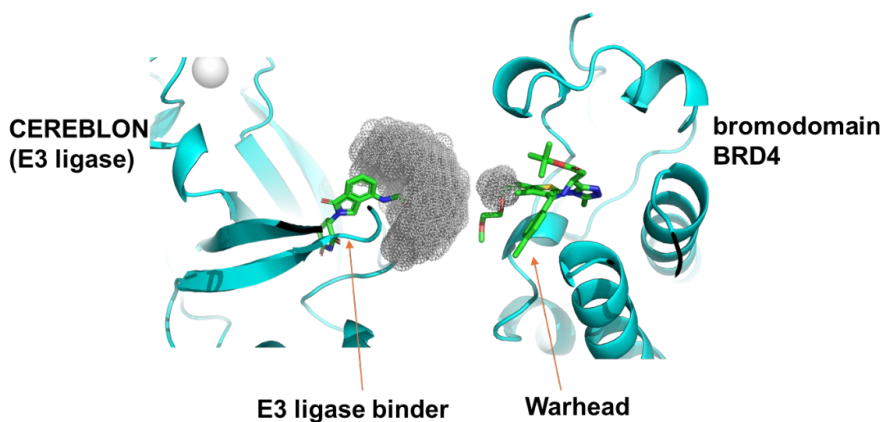


Figure 3. Conformational sampling step. The proteins are shown as blue ribbons. The warhead and E3 ligase binder are shown as green stick. The results of conformational search for the linker in both the proteins are displayed as grid surfaces (grey).

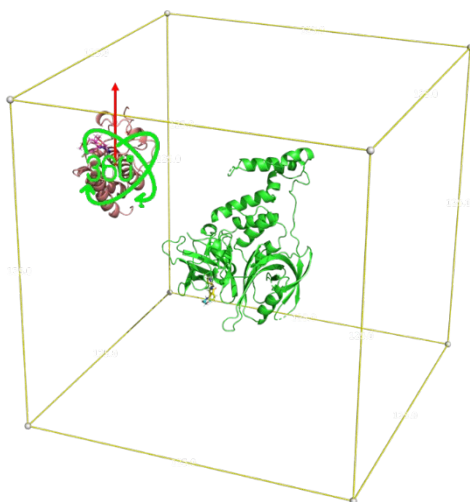


Figure 4. Example of the exploration of the space by the free protein (upper left) respect to the fixed protein (center). The proteins are enclosed in the GRID box; the GRID points are not displayed for clarity of presentation.

In the simulation phase, proteins, ligands, and linker conformers are positioned within a unified reference frame. One molecular complex is

fixed, while the other undergoes rotational and translational movements to explore all possible orientations (Figure 4).

Molecular complementarity is evaluated using *GRID-based Molecular Interaction Fields (MIFs)*, where the calculated interaction energies represent hydrophobic, donor, or acceptor interactions, depending on the type of probe, between the protein and the chemical probes (Figure 5).

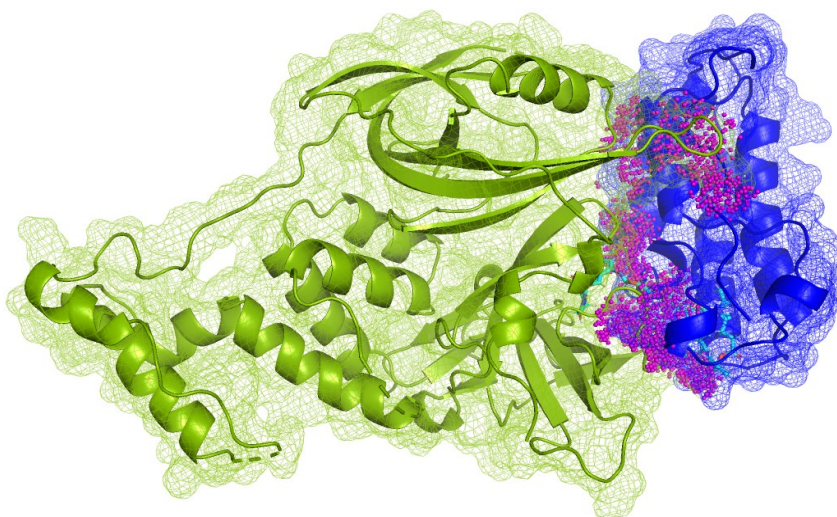


Figure 5. Example of molecular shape complementarity evaluated using MIFs. The red dots indicate site of hydrophobic interaction.

The scoring function quantifies the overall stability and compatibility of the ternary complex. Positive contributions arise from contact surface complementarity, hydrophobic and hydrogen-bonding energies, and the spatial intersection of linker conformational spaces. Negative

contributions penalize steric clashes, and mismatched polar or hydrophobic environments.

The aim of this study was to develop a computational procedure [ref. Ignatov] based on the use of Molecular Interaction Fields (MIFs) and RAINDROP as computational tools, capable of predicting trends in molecular properties such as dissociation constant and cooperativity by analyzing the model generated for the ternary complex.

8.2. Results and Discussion.

The workflow proposed by *Ignatov et al.* [25] represents a benchmark methodology for the structural prediction of PROTAC-mediated ternary complexes. In the original model, the authors introduced two major innovations: i) the generation of half-linker clouds, i.e., ensembles of linker conformations split into two halves associated with the E3 ligase ligand and the target protein ligand, respectively; ii) a modified Fast Fourier Transform (FFT)-based docking algorithm designed to simultaneously explore the conformational space of both proteins and the linker, excluding geometrically incompatible poses a priori.

In the present work, the conceptual structure of that workflow was retained but implemented through a distinct computational approach using the RAINDROP software. The new procedure was tested on known PROTAC degraders and their targets as training set.

8.2.1 Structure selection and docking procedure.

The unbound (apo) structures of the bromodomains belonging to the BET family, namely BRD2, BRD3 and BRD4, were selected as proteins of interest to be linked to an E3 ligase, and the well-characterised PROTAC degraders MZ1 and AT1 were chosen as reference ligand structures (Figure 6).

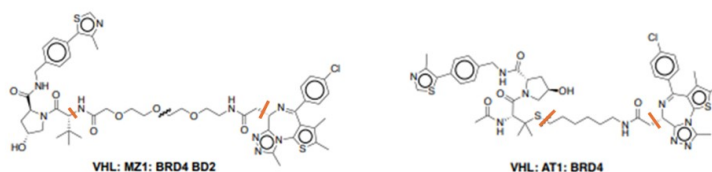


Figure 6. Chemical structures of the two PROTACs used in this study. The orange lines represent the separation point between the ligands and the linker. In both structures the leftmost ligand binds E3 ligase VHL, while the rightmost ligand binds the Protein of interest, Bromodomain 4.

This choice was motivated by two factors: i) The BET bromodomains present a set of closely homologous binding partners (BD1/BD2 domains for each of BRD2/3/4), which makes them ideal for studying subtle variations in cooperativity and dissociation behaviour; ii) The degraders MZ1 and AT1 have been extensively studied: for example, the crystal structure of BRD4 BD2 in complex with VHL and MZ1 was reported, showing significant cooperative ternary complex formation [26].

The chosen structures and related PDB ID are reported in the experimental section. The structures of the BRD proteins and its MZ1 or AT1 warhead ligand along with VHL E3 ligase protein and its ligand were subjected to

homology docking by the web server LigTBM (<http://ligtbm.cluspro.org/>) [27].

The resulting 20 poses for each docking run were ranked according to a weighted sum of homology score (a machine learning term that uses descriptors of the ligand and the protein) and an energy score (derived from statistical potentials such as the *Potential of Mean Force (PMF)*). The best pose for each docking run was retained as starting point for the generation of ternary complex by using the RAINDROP software.

8.2.3 Model generation and Correlation Analysis.

The previously generated structures were used as input for the generation of ternary complex models with the software and procedure of RAINDROP (see experimental section for details). For each set of models of ternary complex obtained, an analysis was performed in order to evaluate the possibility of predicting the trend of dissociation constant and cooperativity of the ternary complex as observed with experimental values (Table 1). A selection of the structures was made based on energetic evaluation. The obtained set of models were selected according to two consecutive filters: best 50 quantile of PPI energy and best 50 quantile of PROTAC energy. This was made in order to eliminate all the improbable structures (not favourable PPI interactions) but also to prevent structures with good protein complementarity but poor PROTAC conformation.

The molecular parameter used by *Ignatov et al.* was obtained by applying the formula $e^{-0.1(\text{SASA})}$, in which SASA refers to the solvent accessible surface of the hydrophobic atoms of the linker. In this way, it was possible to obtain a score for each set of models representing the sum of

contribution of each model of the set (weighted sum in Table 1). This value was then correlated to the experimental data reported in Table 1.

Table 1. Dissociation constant and cooperativity compared to the weighted sum of acceptable poses for the Von Hippel-Lindau (VHL) - PROTAC - bromodomain ternary complexes.

PROTAC	Target	ITC ¹		SPR ²		Weighted sum of poses
		K _d	α	K _d	α	
MZ1	BRD2 BD1	24	2,9	23	1,3	10,5
MZ1	BRD2 BD2	28	2,3	0,9	32	10,9
MZ1	BRD3 BD1	19	3,5	12	2,4	16,3
MZ1	BRD3 BD2	7	10,7	8	3,6	23,6
MZ1	BRD4 BD1	28	2,3	30	9	6,2
MZ1	BRD4 BD2	3,7	17,6	1	22	31,1
MZ1	BRD2 BD1 KEA	12	7,9			18,7
MZ1	BRD4 BD2 QVK	14,9	4,2			2,5
AT1	BRD2 BD1	280	1,4	402	0,3	0,89
AT1	BRD2 BD2	78	4,1	27	4,1	0,54
AT1	BRD3 BD1	207	1,5	133	0,8	1,43
AT1	BRD3 BD2	79	4,3	163	0,7	0,89
AT1	BRD4 BD1	390	1	578	0,2	0,37
AT1	BRD4 BD2	46	7	24	4,7	2,54
AT1	BRD2 BD1 KEA	52	7			1,24
AT1	BRD4 BD2 QVK	160	2			0,1

^aDetermined by isothermal calorimetry ([26]). ^bDetermined by surface plasmon resonance ([28]). Taken from ref. [25]

This procedure, following an analogous approach, was implemented using MIFs to identify the grid points satisfying specific hydrophobic or hydrophilic probe requirements. In this way, also the SASA of the hydrophilic portions of the molecule were taken into account. Each valid MIF point was multiplied by 0.00496, corresponding to the conversion factor between grid points and surface area, in order to reproduce the

SASA-based weighting applied by *Ignatov et al.* To isolate the solvent-exposed regions, the surface of the PROTAC was subtracted by the portion overlapping with the protein, yielding the final hydrophobic or hydrophilic exposed surface, depending on the probe used. The total area was then obtained by summing all grid points. For each selected set of ternary complexes, we calculated the MIF surface by using the hydrophobic probe DRY and CRY (this latter considers also aromatic moieties) together with the hydrophilic OH and H₂O probe (1.4 Å radius). At first, the calculations were performed on the best 1000 structures, then on the best 500, and then on the best 250, in order to check if the reduction of the number of selected complexes could reduce the noise in the data.

In any case, results showed that there was no clear correlation between the obtained sum of surfaces calculated by MIFs for each probe used, and the dissociation constant or cooperativity of the complexes (Figures 7 and 8).

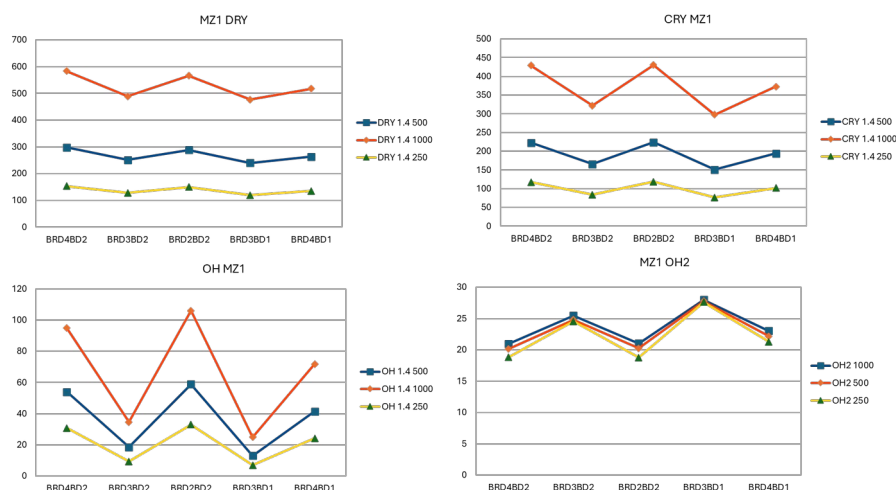


Figure 7. Graph for each probe used for MZ1 bearing complexes. The x-axis reports the selected complexes in order of stability (dissociation constant) from left to right. The y-axis reports the value of the sum of MIF derived surfaces.

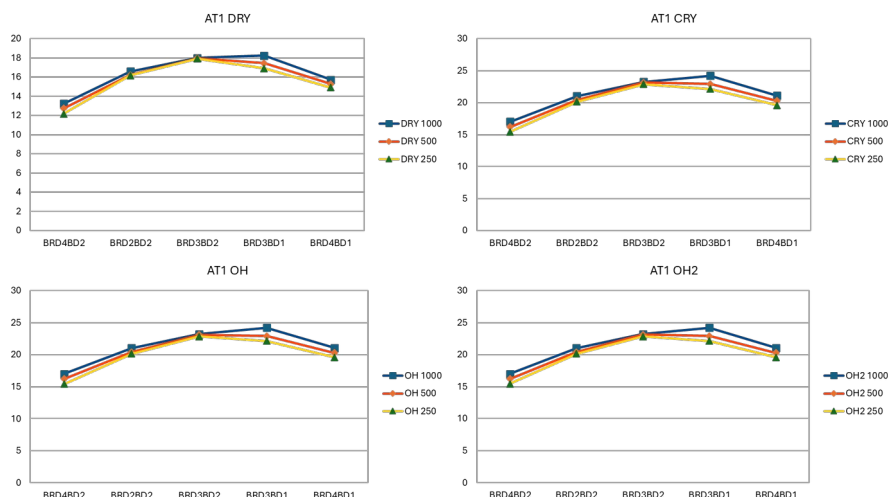


Figure 8. Graph for each probe used for AT1 bearing complexes. The x-axis reports the selected complexes in order of stability (dissociation constant) from left to right. The y-axis reports the value of the sum of MIF derived surfaces.

We also evaluated the mean of the ratio of the DRY and OH2 probe calculated surfaces for each complex of each ternary complex considered (Figure 9). In this case as well there was no clear correlation.

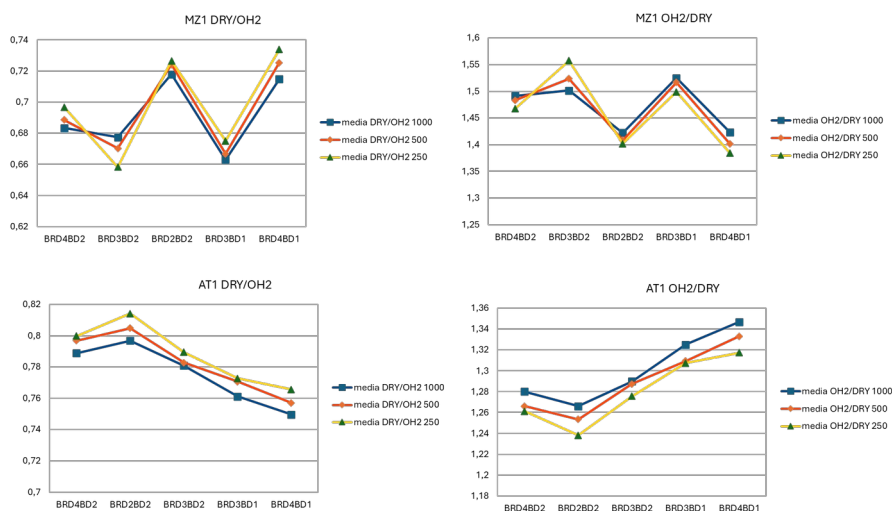


Figure 9. Graphs of the mean value of the ratio between surfaces calculated by the DRY and OH2 probe. The x-axis reports the selected complexes in order of stability from left to right. The y-axis reports the value of the mean ratio.

On the other hand, this analysis evidenced a similar trend for all probes, thus this issue was further investigated. In particular, since each complex should show “favorable” hydrophobic surface (i.e. the smaller the resulting calculated area is the better, since there is less hydrophobic portions exposed) or, on the contrary, a greater exposition of the hydrophilic surface, the correlation of these two parameters was analyzed (cutoff best 500 structures) (Figure 10).

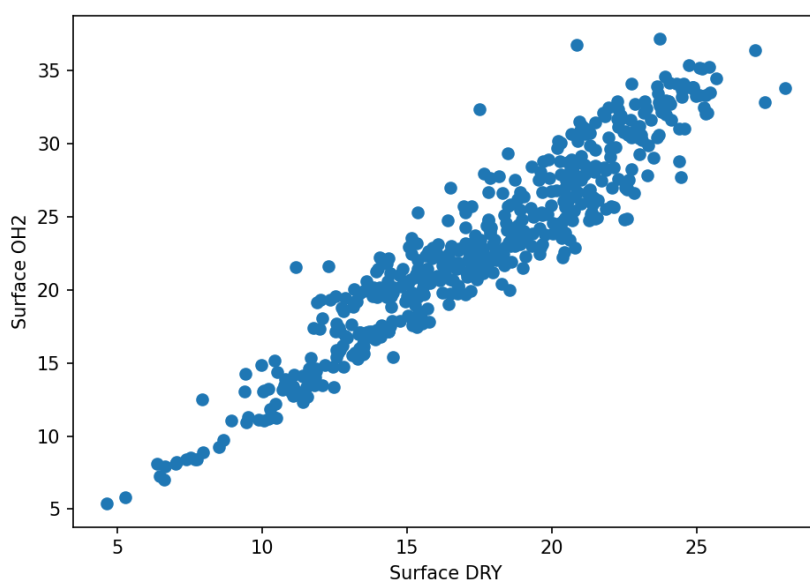


Figure 10. Correlation of calculated surfaces using the DRY (X axis) or OH2 (Y axis) probes obtained for the best 500 models of BDR3BD1-AT1-VHL complex. The results for the other models were very similar.

This analysis confirmed that the two parameters are collinear in the selected models of the complex, meaning that they are correlated to such

an extent that it becomes challenging to isolate and interpret their individual effects on the outcome.

A more restrictive selection of the structures was applied, using only the top 100 and top 20 structures, but similar results were obtained (Figure 11).

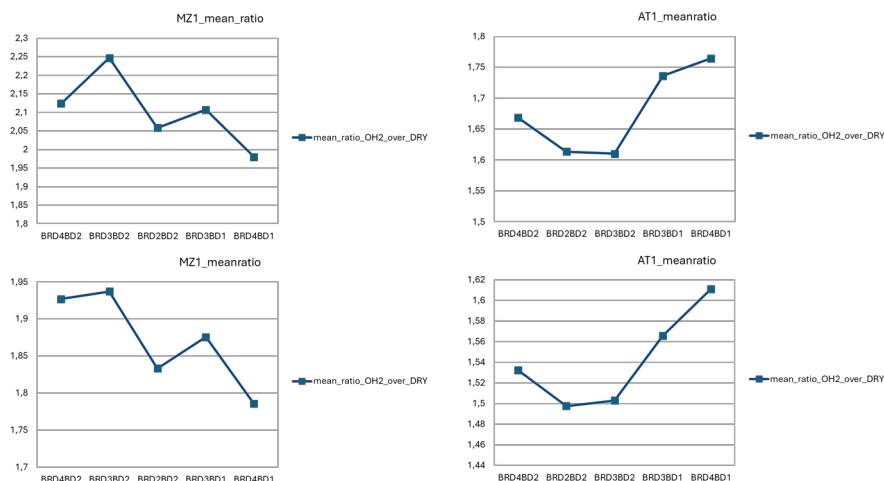


Figure 11. Graphs of the mean value of the ratio between surfaces calculated by the DRY and OH2 probe. The x-axis reports the selected complexes in order of stability from left to right. The y-axis reports the value of the mean ratio. Upper part: top 20 structures. Lower part: top 100 structures.

In summary, our results show that, unlike the trend reported by *Ignatov et al.*, no clear correlation was observed between experimental constant of dissociation and cooperativity and the MIF calculated surfaces. This lack of correlation is primarily attributed to the collinearity among the MIF surfaces obtained using different probes, which complicates the identification and interpretation of their individual contributions to the observed outcome.

To address this limitation, future refinements could include the improvement of the criteria used to select the structures produced by

docking studies and a tighter coupling between energetic and scoring functions, allowing the MIF distribution to directly influence the energetic ranking of the models. In addition, introducing dynamic refinement steps, such as short molecular dynamics simulations or local conformational sampling around the best-scoring poses could select the most favorable poses in term of surface exposition.

Finally, the integration of statistical weighting schemes, for example, clustering models according to MIF density overlap or interface residue recurrence, could help reveal hidden correlations between energy, surface complementarity, and potential cooperativity.

8.3. Conclusions.

The primary objective of this study was to reproduce and evaluate a benchmark computational methodology using RAINDROP software and Molecular Interaction Fields (MIFs) to predict key properties of PROTAC ternary complexes, specifically the dissociation constant and cooperativity. The focus was placed on the well-characterized degraders MZ1 and AT1 in complex with the VHL E3 ligase and BET bromodomains (BRD2, BRD3, and BRD4). The results clearly demonstrated a lack of correlation between the obtained sum of MIF surfaces (using probes such as DRY, CRY, OH, and OH2) and the experimental constant of dissociation and cooperativity of the complexes. This failure was mainly attributed to the collinearity among the MIF surfaces generated by different probes within the ensemble of ternary complexes. This distribution prevented the identification of a dominant, energetically favourable interaction pattern based on surface exposure, even after rigorous energetic filtering. To improve the predictive accuracy

of *in silico* PROTAC modelling, future refinements should focus on addressing this issue, such as introducing enhanced complex production (short molecular dynamics simulations or local conformational sampling around the best-scoring poses) and targeted complex selection, or clustering the models based on MIF density overlap.

8.5. References.

1. Mangal, S., Gao, W., Li, T., and Zhou, Q. (2017). Pulmonary Delivery of Nanoparticle Chemotherapy for the Treatment of Lung Cancers: Challenges and Opportunities. *Acta Pharmacol. Sin.* 38, 782–797. doi:10.1038/aps.2017.34
2. Dong, J., Qin, Z., Zhang, W.-D., Cheng, G., Yehuda, A. G., Ashby, C. R., Jr., et al. (2020). Medicinal Chemistry Strategies to Discover P-Glycoprotein Inhibitors: An Update. *Drug Resist. Updates.* 49, 100681. doi:10.1016/j.drup.2020.100681
3. Yuan, L., Xu, Z. Y., Ruan, S. M., Mo, S., Qin, J. J., and Cheng, X. D. (2020). Long Non-coding RNAs towards Precision Medicine in Gastric Cancer: Early Diagnosis, Treatment, and Drug Resistance. *Mol. Cancer.* 19, 96. doi:10.1186/s12943-020-01219-0
4. Zhou, Y., Chen, R., Luo, X., Zhang, W. D., and Qin, J. J. (2020). The E2 Ubiquitin-Conjugating Enzyme UbcH5c: an Emerging Target in Cancer and Immune Disorders. *Drug Discov. Today* 25 (11), 1988–1997. doi:10.1016/j.drudis.2020.09.015
5. Qi, S. M., Cheng, G., Cheng, X. D., Xu, Z., Xu, B., Zhang, W. D., et al. (2020). Targeting USP7-Mediated Deubiquitination of MDM2/MDMX-P53 Pathway for Cancer Therapy: Are We There yet?. *Front Cel Dev Biol.* 8, 233. doi:10.3389/fcell.2020.00233
6. Yu, D. H., Xu, Z. Y., Mo, S., Yuan, L., Cheng, X. D., and Qin, J. J. (2020). Targeting MDMX for Cancer Therapy: Rationale, Strategies, and Challenges. *Front. Oncol.* 10, 1389. doi:10.3389/fonc.2020.01389
7. Dong, J., Zhang, Q., Wang, Z., Huang, G., and Li, S. (2018). Recent Advances in the Development of Indazole-Based Anticancer Agents. *ChemMedChem.* 13, 1490–1507. doi:10.1002/cmdc.201800253
8. Wu, P., Nielsen, T. E., and Clausen, M. H. (2015). FDA-approved Small-Molecule Kinase Inhibitors. *Trends Pharmacol. Sci.* 36, 422–439. doi:10.1016/j.tips.2015.04.005
9. Xu, J. L., Yuan, L., Tang, Y. C., Xu, Z. Y., Xu, H. D., Cheng, X. D., et al. (2020). The Role of Autophagy in Gastric Cancer Chemoresistance: Friend or Foe?. *Front. Cel Dev Biol.* 8, 621428. doi:10.3389/fcell.2020.621428

10. Potjewyd, Frances, et al. "Degradation of polycomb repressive complex 2 with an EED-targeted bivalent chemical degrader." *Cell chemical biology* 27.1 (2020): 47-56.
11. Buckley, D. L., Raina, K., Darricarrere, N., Hines, J., Gustafson, J. L., Smith, I. E., et al. (2015). HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem. Biol.* 10, 1831–1837. doi:10.1021/acscchembio.5b00442
12. Zhang, X., Thummuri, D., He, Y., Liu, X., Zhang, P., Zhou, D., et al. (2019). Utilizing PROTAC Technology to Address the On-Target Platelet Toxicity Associated with Inhibition of BCL-XL. *Chem. Commun.* 55, 14765–14768. doi:10.1039/c9cc07217a
13. Kregel, S., Wang, C., Han, X., Xiao, L., Fernandez-Salas, E., Bawa, P., et al. (2020). Androgen Receptor Degraders Overcome Common Resistance Mechanisms Developed during Prostate Cancer Treatment. *Neoplasia.* 22, 111–119. doi:10.1016/j.neo.2019.12.003
14. Vollmer, S., Cunoosamy, D., Lv, H., Feng, H., Li, X., Nan, Z., et al. (2020). Design, Synthesis, and Biological Evaluation of MEK PROTACs. *J. Med. Chem.* 63, 157–162. doi:10.1021/acs.jmedchem.9b00810
15. Nam, T., Han, J. H., Devkota, S., and Lee, H. W. (2017). Emerging Paradigm of Crosstalk between Autophagy and the Ubiquitin-Proteasome System. *Mol. Cell.* 40, 897–905. doi:10.14348/molcells.2017.0226
16. Nandi, D., Tahiliani, P., Kumar, A., and Chandu, D. (2006). The Ubiquitin-Proteasome System. *J. Biosci.* 31, 137–155. doi:10.1007/bf02705243
17. Zou, Y., Ma, D., and Wang, Y. (2019). The PROTAC Technology in Drug Development. *Cell Biochem Funct.* 37, 21–30. doi:10.1002/cbf.3369
18. Toure, M., and Crews, C. M. (2016). Small-Molecule PROTACS: New Approaches to Protein Degradation. *Angew. Chem. Int. Ed.* 55, 1966–1973. doi:10.1002/anie.201507978
19. Sun, X., and Rao, Y. (2020). PROTACs as Potential Therapeutic Agents for Cancer Drug Resistance. *Biochemistry.* 59, 240–249. doi:10.1021/acs.biochem.9b00848
20. Neklesa, T. K., Winkler, J. D., and Crews, C. M. (2017). Targeted Protein Degradation by PROTACs. *Pharmacol. Ther.* 174, 138–144. doi:10.1016/j.pharmthera.2017.02.027
21. Guo, J., Liu, J., and Wei, W. (2019). Degrading Proteins in Animals: "PROTAC"tion Goes In Vivo. *Cell Res.* 29, 179–180. doi:10.1038/s41422-019-0144-9
22. Schapira, M., Calabrese, M. F., Bullock, A. N., and Crews, C. M. (2019). Targeted Protein Degradation: Expanding the Toolbox. *Nat. Rev. Drug Discov.* 18, 949–963. doi:10.1038/s41573-019-0047-y

23. Qi, Si-Min, et al. "PROTAC: an effective targeted protein degradation strategy for cancer therapy." *Frontiers in pharmacology* 12 (2021): 692574.
24. Goodford, Peter J. "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules." *Journal of medicinal chemistry* 28.7 (1985): 849-857.
25. Ignatov, Mikhail, et al. "High accuracy prediction of PROTAC complex structures." *Journal of the American Chemical Society* 145.13 (2023): 7123-7135.
26. Gadd, Morgan S., et al. "Structural basis of PROTAC cooperative recognition for selective protein degradation." *Nature chemical biology* 13.5 (2017): 514-521.
27. Kotelnikov, Sergei, et al. "Accurate ligand–protein docking in CASP15 using the ClusPro LigTBM server." *Proteins: Structure, Function, and Bioinformatics* 91.12 (2023): 1822-1828.
28. Roy et al. SPR-measured dissociation kinetics of PROTAC ternary complexes influence target degradation rate. *ACS Chemical Biology* 14, 361–368 (2019)

Experimental Section

Molecular modeling calculations were performed on the CPU/GPU hybrid High-Performance Computing Cluster (10 Twin servers, for a total of 560 Intel® Xeon® Gold processors (128 GB RAM), 64 AMD® EPYC® processors and 2 GPU NVIDIA® Tesla® V100) and on the High-Performance Computing Cluster (6 Twin servers for a total of 12 nodes each equipped with Intel® Xeon® QuadCore E5520 CPU, 36 GB RAM). The molecular modeling graphics were carried out on a personal computer equipped with an Intel (R) Core (TM) i7-8700 processor and SGI Octane 2 workstations.

1. Pka calculations

The apparent pKa and clogD values (pH 7.4) of compounds were calculated using the ACD/pKa GALAS algorithm of ACD/Percepta software (ACD/Percepta, Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2017, <http://www.acdlabs.com> (accessed 5 September 2022)). In particular it was calculated for compounds 1a–1d, 2a–2d, 3a–3d, and 4a–4d of GAPDH inhibitors, for compounds **9a-i** and **10a-b** of the TSP derivatives and for DM1 and DM2 compounds of hCav3.1 channels inhibitors. Then, the percentage of neutral/ionized forms was computed at pH 7.2 (cytoplasm pH value) using the Handerson–Hasselbalch equation and the compounds were built accordingly.

2. Molecule preparation

2.1 Molecule Building

The molecular models of the new compounds were built using Small Molecule tool of Discovery Studio 2017 or 2023. (Dassault Systèmes BIOVIA, San Diego, CA, USA, 2017). This tool was used to build i) compounds **1a–1d**, **2b–2d**, **3b–3d**, and **4b–4d** of *pf*GAPDH inhibitors; ii) **9a-i** and **10a-b** of TSP derivatives; iii) H2TTPS; iv) DM1 and DM2; v) the THIA compounds.

2.2 Forcefield

The atomic potentials and charges were assigned using CFF forcefield [1] (*pf*GAPDH and THIA compounds), CVFF forcefield [2] with MNDO semi-empirical 1 SCF calculation for partial charge assignment [3] (H2TTPS) and CHARMM forcefield [4] (TSPs derivatives, DM1 and DM2, THIA targets, nsp10, nsp14 and nsp16 proteins of Sars-cov-2)

3. Conformational analysis

3.1 Conformational Sampling at MM level

The conformational space of each small-molecule ligand was explored at the molecular mechanics level prior to quantum-mechanical refinement and docking studies. Initial 3D structures were generated as described in the molecule building section and subjected to force-field-based conformational searches using Discovery Studio (Dassault Systèmes BIOVIA) or Insight 2005 (Accelrys).

Two main strategies of MM sampling were adopted:

i) Systematic torsional search.

In the case of the PfGAPDH inhibitor series (Chapter 2), an initial set of low-energy conformers was generated by a systematic variation of selected rotatable bonds. All $\text{sp}^3\text{-sp}^3$ and $\text{sp}^3\text{-sp}^2$ single bonds not part of rings were rotated in 60° increments. The structures were first minimized by MM using the CFF force field ($\epsilon = 80\cdot\text{r}$) until the maximum RMS derivative was below $0.001 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$ (Conjugate Gradient algorithm[5]). Duplicates were removed using an RMSD cutoff of $0.01 \text{ \AA}$, and structures within an energy window of $10^6 \text{ kcal}\cdot\text{mol}^{-1}$ above the global minimum were retained as a diverse starting set for further refinement.

ii) Random Boltzmann Jump search.

Conformational sampling for the TSP pyrrole derivatives (9a–i, 10a–b and reference 1, Chapter 3), the Cav3.1 ligands DM1 and DM2 (Chapter 5) and the THIA derivatives (Chapter 6) was carried out by the Boltzmann Jump random search algorithm. Up to 300–400 random conformations of each compound were generated. Trial structures were accepted or rejected based on the Metropolis criterion at 300 K, biasing the sampling towards the low-energy regions of the potential energy surface. Subsequently, series-specific energy thresholds were employed in the selection: $20 \text{ kcal}\cdot\text{mol}^{-1}$ above the global minimum for the TSP derivatives, and $10^6 \text{ kcal}\cdot\text{mol}^{-1}$ for DM1/DM2 and THIA compounds. Accepted conformations were minimized by MM (Conjugate Gradient[5], $\epsilon = 80\cdot\text{r}$) down to an RMS gradient below 0.001 or $0.01 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$, appropriate for each series.

A distinct protocol was used for porphyrin H_2TPPS (Chapter 4), whose conformational flexibility is dominated by collective deformations of the macrocycle and the sulfonate groups. The conformational space was explored in that case by simulated annealing (SA) using the CVFF force

field and MNDO-derived partial charges. From the starting tetra-anionic form, 200 SA cycles were carried out ($\epsilon = 80 \cdot r$). During each cycle, the system was first heated to 1000 K during 2000 fs (time step 1.0 fs), then equilibrated at 1000 K during 2000 fs to surmount torsional barriers, and finally cooled linearly to 300 K in 1000 fs (decrease of $0.5 \text{ K} \cdot \text{fs}^{-1}$). All conformations generated by SA were subsequently minimized by MM (CVFF, $\epsilon = 80 \cdot r$, Conjugate Gradient) down to a maximum RMS force below $0.001 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$ and ranked by their relative energy.

3.2 MM-level conformer classification

The MM-minimized conformers were then classified according to their torsional features in order to obtain a description of the accessible conformational space.

For the TSP pyrrole derivatives (Chapter 3), the conformers were gathered into families and sub-families defined by the values of three key torsion angles (τ_1 , τ_2 and τ_3), which determine the relative orientation of the pharmacophoric aromatic rings (X, Y, Z) and the linker. The qualitative labels used are synperiplanar (C), antiperiplanar (T), synclinal (G^+/G^-) and anticlinal (A^+/A^-), with the following angular ranges:

0° to $\pm 30^\circ$ (C), 30° to 90° (G^+), -30° to -90° (G^-), 90° to 150° (A^+), -90° to -150° (A^-), and $\pm 150^\circ$ to 180° (T).

The combinations of the qualitative states of τ_1 - τ_3 defined conformational families, here denoted as TTT, TTC, TCC, TCT, TCG⁻, TCA⁻ and TCG⁺. These were further subdivided into subfamilies indexed by Roman numerals according to an additional torsion τ_5 .

For DM1 and DM2 (Chapter 5), classification proceeded likewise but focusing on τ_1 and τ_2 (and τ_3 for DM2), describing the relative orientation of the phenyl ring and carboxylate moiety. Owing to symmetry, τ_1 and τ_2 were mapped into the interval -90° to $+90^\circ$ by subtracting 180° from angles in the range 90° to 180° and adding 180° to angles in the range -180° to -90° , thus avoiding redundant representations of equivalent rotamers. Conformers were clustered according to the torsion sets thus reduced.

For the THIA derivatives (Chapter 6), MM conformers were sorted according to increasing energy and directly classified based on the values of all relevant torsion angles along the flexible scaffold, in order to identify distinct binding relevant orientations prior to docking and DFT refinement.

In the case of H₂TPPS, classification (Chapter 4) was based on the conformational energy and the interatomic distances between the sulfonate groups, because these distances directly modulate the spatial arrangement of the four negative charges and are critical for the definition of the sulfonate-based pharmacophore.

3.3 DFT refinement of MM conformers.

Selected MM conformers were refined at the DFT level by using Gaussian 16[6] to get more accurate geometries and relative energies for further pharmacophore and docking analyses.

The geometries for the TSP series (Chapter 3), DM1/DM2 (Chapter 5) and THIA derivatives (Chapter 6) were fully optimized using the B3LYP functional in conjunction with the 6-311++G(d,p) basis set[7,8]. The conductor-like polarizable continuum model (C-PCM)[9] was used throughout along with water as the solvent to include the solvent effects. The vibrational frequency calculations (keyword = freq) were used to

characterize all the stationary points in order to assure that no imaginary frequencies exist (energy minima) and for obtaining thermochemical corrections when required. Natural bond orbital analysis (NBO)[10] was employed to analyze the electronic distribution.

For each series, DFT-optimized conformers were ranked by their potential energy and the GM was identified; ΔE_{GM} relative to GM was then computed for each conformer. Further analysis was restricted to low-energy structures: $\Delta E_{\text{GM}} \leq 5 \text{ kcal}\cdot\text{mol}^{-1}$ for the TSP and DM1/DM2 series; ΔE_{GM} thresholds specific for the THIA set are given in the corresponding chapter.

The TSP set was re-assigned to the families and sub-families previously defined (i.e., TTT, TCC, TCT, etc.) within the same torsional criteria at both the MM and the DFT level in order to compare the results consistently at the classical and quantum-mechanical levels.

3.4 Pharmacophore-based geometric analysis

DFT-refined conformers were used to derive pharmacophore geometries and to compare ligand-based distance patterns with those observed in experimentally determined protein-protein interfaces or enzyme active sites. For the TSP derivatives presented in Chapter 3, centroids of the aromatic rings X, Y and Z were used as pharmacophore points. Distances between such centroids were computed for each low-energy DFT conformer ($\Delta E_{\text{GM}} \leq 2 \text{ kcal}\cdot\text{mol}^{-1}$) and compared with the distances between the centroids of hot-spot residues in coiled-coil and LxxLL motifs extracted from crystal structures. Coiled-coil hot spots were defined at positions i , $i+3$, $i+4$ and $i+7$ and centroids were computed using ring carbon atoms for aromatic residues and all heavy atoms of the side chains for aliphatic residues. TSP conformers were considered good

mimics of a given motif if two pharmacophore distances differed by less than 1.0 Å and a third one by less than 1.5 Å from the corresponding experimental values; conformers with $\text{RMSD} < 2.5$ Å upon superposition of the pharmacophore centroids onto the hot-spot centroids were classified as mimetic. For H₂TPPS (Chapter 4), the distances between the sulfonate groups in the global minimum conformer defined a square-planar, tetra-anionic pharmacophore that was fitted onto spatial arrangements of solvent-exposed positively charged residues on the human 20S proteasome. This mapping guided the generation of initial binding poses for subsequent docking and dynamic docking simulations. For the THIA derivatives presented in Chapter 6, DFT geometries were used to ensure consistent internal conformations during both docking and steered molecular dynamics simulations. More specifically, a set of key distances and orientations involving the reactive sulfur atom and the N2/N4 substituents with respect to the catalytic cysteine and subsite-defining residues—for example, S1, S2, S3 subsites—were monitored to assess whether the binding mode was preserved. In all, the conformational analysis workflow combined MM-level exhaustive or stochastic sampling with systematic torsional classification, DFT refinement, and pharmacophore-based distance analysis. The multilayered approach provided a consistent set of ligand conformations that formed the basis of subsequent input for docking, molecular dynamics, and structure-based bioinformatics throughout the thesis.

4. Protein structure selection, preparation and modelling

4.1. Selection of experimental structures

All structure-based calculations reported in this thesis were grounded, whenever possible, on experimentally determined protein structures. Crystal and cryo-EM models were retrieved from the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/>), using the macromolecule name and functional annotations as search criteria. For each target, the structures were chosen to satisfy three main requirements: (i) adequate resolution and refinement quality, (ii) maximal completeness of the amino acid sequence, and (iii) a functional state consistent with the biological question addressed in the corresponding chapter.

For PfGAPDH, the glycolytic enzyme investigated in Chapter 2, previously developed atomic models and the crystallographic complex with a glyceraldehyde 3-phosphate analogue (PDB ID: 1ML3) were used to define the architecture of the catalytic site and to guide the placement of the designed inhibitors. The analysis of protein–protein interaction motifs in Chapter 3 relied on coiled-coil and LxxLL-containing complexes, namely the GCN4 leucine zipper (PDB ID: 2ZTA), the GreA coiled coil (PDB ID: 1GRJ) and the LxxLL motif of nuclear receptor coactivator 5 bound to the estrogen receptor β (PDB ID: 2J7X).

In the case of the proteasome (Chapter 4), a panel of structures was assembled to cover the human 20S core particle in different conformational states and in complex with distinct regulatory particles. Semi-open human 20S structures (e.g. PDB ID: 5T0I) and 20S–19S, 20S–PA28 and 20S–PA200 assemblies were included to capture the range of gating and activation modes. The *Thermoplasma acidophilum* 20S proteasome in complex with the small-molecule activator ZYA (PDB ID:

8F7K) was additionally selected to transfer information on binding sites and interaction patterns to the human complex.

For voltage-gated calcium channels (Chapter 5), the T-type Cav3.1 channel was represented both in apo form (PDB ID: 6KZO) and in complex with the selective blocker Z944 (PDB ID: 6KZP), while L-type Cav1.1 structures in complex with verapamil and diltiazem (PDB IDs: 6JPA, 6JPB) were used for comparison of binding modes across channel families. The panel of cysteine proteases analysed in Chapter 6 comprised SARS-CoV-2 3CLpro and PLpro in complex with covalent or non-covalent inhibitors (e.g. PDB IDs: 7JKV, 7N89, 6WX4, 7JIW), together with papain (PDB ID: 6TCX), cathepsin L (PDB ID: 3OF8) and bromelain (PDB ID: 6YCG). Additional structures such as 7SF1 (3CLpro), 1PE6 (papain) and 7W34 (cathepsin L) were selected as templates for covalent docking.

For the non-structural proteins nsp10, nsp14 and nsp16 (Chapter 7), all PDB entries containing these proteins, either alone or in complexes, and deposited up to December 2023 were retrieved using the “Advanced Search Query Builder” interface. The initial list was filtered by imposing a sequence identity threshold of 70% to ensure that only structurally and functionally comparable models were retained. The final sets for each protein, and the corresponding functional states, are reported in the methodological tables of that chapter.

Finally, PROTAC modelling in Chapter 8 required representative bromodomain-containing BET proteins (BRD2, BRD3, BRD4) and the VHL E3 ligase. For each of these, bromodomains and VHL complexes with well-resolved ligands (e.g. AT1, MZ1) were selected as templates to

guide the construction of binary complexes subsequently used as input for ternary PROTAC assembly.

4.2. General preparation and protonation protocol

All selected PDB structures were subjected to a uniform preparation protocol before being used in docking, molecular dynamics or energetic analyses. This protocol was implemented using the “Prepare Protein” tools in Discovery Studio (versions 2017, 2023 and 2025), complemented by manual inspection and, where appropriate, by the use of Insight 2005 or related modules.

In a first step, non-functional small molecules introduced by crystallisation conditions, such as buffer components, salts or cryo-protectants, were removed. Substrates, cofactors and inhibitors directly related to the biological activity under study were retained when they provided essential information on the binding site geometry or on key interaction patterns. An exception was made in nsp16–nsp10 complexes, where sulfate ions mimicking RNA phosphate groups were maintained, as they defined low-affinity binding sites relevant to the functional interpretation.

Missing side chains were rebuilt using rotamer libraries under steric and hydrogen-bonding constraints, ensuring that the new conformations were compatible with the local environment. Internal backbone gaps, when present and functionally relevant, were subsequently treated by homology or loop modelling (Section 4.3), whereas disordered N- and C-terminal tails were often left truncated when they did not significantly affect the local structure of the binding sites.

Hydrogen atoms were added to all structures, and residue protonation states were assigned assuming a pH of 7.2. Protein ionisation and residue

pK_a values were evaluated to refine the assignment when necessary. Standard protonation patterns were used for acidic and basic residues (Asp/Glu deprotonated; Lys/Arg protonated), whereas the protonation of histidines was chosen based on the local hydrogen-bonding network and, when relevant, on metal coordination or catalytic roles. N- and C-termini were generally treated as charged (NH₃⁺ and COO⁻), unless the crystallographic construct corresponded to a truncated peptide where neutral capping was more appropriate, as in some Cav3.1 chain termini. Metal ions such as Ca²⁺ in Cav3.1 were retained in their crystallographic positions; in subsequent minimizations, harmonic restraints were applied to maintain their coordination geometry and avoid unrealistic displacements.

Protein atoms were parametrised using the CHARMM force field for all structural and energetic calculations, including the human 20S proteasome and its regulatory complexes, Cav channels, coronavirus proteases and non-structural proteins. In workflows where alternative force fields were used for ligands (e.g. CVFF or cff for specific small molecules), the protein environment remained described by CHARMM.

4.3. Homology modelling and rebuilding of missing regions

When the available experimental structures lacked entire regions that were essential for the planned simulations, homology modelling or focused rebuilding was employed to obtain complete and physically realistic models.

A representative example is provided by the semi-open human 20S proteasome used in Chapter 4. The experimental structure 5T0I lacks several N- and C-terminal segments, the loop spanning residues 128–133 in the $\alpha 5$ subunit, and multiple side chains. To reconstruct a complete

model, the sequence of 5T0I was first aligned with the full-length human α and β subunits downloaded from UniProt. Secondary structure prediction with PredictProtein was then used to define structurally conserved regions (SCRs) in each subunit, which served as a rigid core during the modelling process. The coordinates of these SCRs were assigned using the SCRs-AssignCoords procedure in the Homology module of Insight 2005, with 5T0I as template.

The missing $\alpha 5$ loop was generated with the Generate Loops procedure. In this step, a peptide backbone was built between the conserved residues flanking the missing segment, and ϕ/ψ angles were sampled to satisfy distance constraints between the loop ends and the anchoring SCR residues. Candidate loops were iteratively refined until convergence of the geometric constraints, filtered for steric clashes using a permissive overlap factor, and ranked by conformational energy. The loop with the lowest energy and acceptable steric profile was selected. Missing N- and C-terminal residues were added using the EndRepair command, and side chains were reconstructed with the Replace command. The resulting full-length model was then energy-minimised within the Discover_3 module, allowing all disordered N-/C-terminal segments and SCR side chains to move, while keeping the SCR backbone fixed to avoid large distortions of the overall fold. At each step, geometric quality was assessed using ProStat/Struct_Check and MolProbity[11], by monitoring backbone dihedrals, side-chain rotamers and hydrogen-bond patterns.

A second case of homology-based rebuilding was the construction of a human p14ARF model in Chapter 3, starting from the NMR structure of the mouse homologue p19ARF (PDB ID: 1HN3). The human p14ARF sequence was aligned with the mouse sequence (PROMALS 3D), and a

backbone model was generated in Discovery Studio by transferring the conserved core. Hot-spot residues corresponding to newly identified protein-recognition motifs on the α H2 helix of p14ARF were then introduced, and side chains were optimised. This model was subsequently used for pharmacophore superposition of TSP conformers and for the structural interpretation of potential protein–protein interaction interfaces.

In other systems, full homology modelling was not required, but local rebuilding was necessary. For Cav3.1 (PDB ID: 6KZP), for instance, missing side chains in the pore region and selectivity filter (e.g. I107, C109, D110, S111, Q112) were reconstructed, and termini of truncated chains were capped neutrally to avoid artefacts in electrostatics and docking. In 3CLpro and related proteases, alternative side-chain conformations present in the PDB files were removed, missing atoms were added, and short gaps were regularised prior to docking and steered molecular dynamics.

4.4. Structurally conserved regions and restraints in simulations

Across several systems, structurally conserved regions were explicitly identified and treated as a relatively rigid core during docking and dynamic simulations. This strategy was adopted to preserve the integrity of well-defined secondary structure elements while allowing local flexibility in loops, side chains and binding sites.

In the human 20S proteasome (Chapter 4), the SCRs of each α and β subunit were defined on the basis of secondary structure prediction (PredictProtein) and sequence conservation. Segments forming α -helices and β -sheets with high reliability scores were designated as SCRs and were used throughout homology modelling, docking and dynamic docking. During these simulations, the distances between backbone hydrogen-bond

donors and acceptors within α -helices were restrained around typical values (~ 2.5 Å), and ϕ/ψ dihedral angles in β -sheets were confined within ranges consistent with parallel or antiparallel arrangements. Quadratic restraints were applied with force constants scaled according to the reliability index from PredictProtein, so that highly conserved regions were more strongly restrained than less conserved ones. In particular, the starting structure for homology modelling of semi-open conformation (PDB ID: 5T0I) lacks the N-terminal regions ($\alpha 1$: aa1-4; $\alpha 2$: aa1; $\alpha 3$: aa1; $\alpha 4$: aa1; $\alpha 5$: aa1-7; $\alpha 6$: aa1-3; $\alpha 7$: aa1-5), C-terminal regions ($\alpha 1$: aa245-246; $\alpha 3$: aa252-261; $\alpha 4$: aa241-248; $\alpha 6$: aa242-263; $\alpha 7$: aa246-255; $\beta 1$: aa192-205; $\beta 2$: aa221-234; $\beta 4$: aa200-201; $\beta 5$: aa202-204; $\beta 7$: aa216-219), one loop region ($\alpha 5$: aa128-133), and the side chains of several residues ($\alpha 1$: aa45, 46, 55, 59, 63, 146, 181, 186-188, 196, 223; $\alpha 2$: aa18, 36, 52-55, 60, 139-141, 143, 163, 176-178, 184, 193, 195, 196, 199, 200, 202, 226, 227, 230, and 233; $\alpha 3$: aa13, 37, 51, 52, 54, 177, 184, 194, 201, 205, 210, 222, 229-231, 250, and 251; $\alpha 4$: aa27, 40, 42, 46-50, 53, 136, 138, 141, 157, 163, 166, 169, 170, 177, 181-183, 185, 187, 189, 192, 193, 195-199, 204, 207, 208, 212-213, 216, 218, 223, 225-227, 229, 232-237, 239 and 240; $\alpha 5$: 8, 86, 187, 192, 208, 209, and 231; $\alpha 6$: aa208, 217, 218, 237, 238, and 241; $\alpha 7$: aa110, 144, 204, 206, 208, and 224; $\beta 2$: aa9, 180, 206, and 214; $\beta 3$: aa47; $\beta 4$: aa41, 95, 109, 185, 198, and 199; $\beta 5$: aa106, 150, and 185; $\beta 6$: aa45, 161, 166, and 200; $\beta 7$: aa44, 156, and 195). The h20S sequences were analyzed using the Structure Prediction and Sequence Analysis server PredictProtein (<http://www.predictprotein.org/>). In $\alpha 1$ subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved ($\alpha 1$, aa23-33; $\alpha 2$, aa85-104; $\alpha 3$, aa111-128; $\alpha 4$, aa172-183; $\alpha 5$, aa191-207; $\alpha 6$, aa232-243; $\beta 1$, aa13-16; $\beta 2$, aa38-43; $\beta 3$, aa48-53; $\beta 4$, aa68-72; $\beta 5$, aa76-80; $\beta 6$, aa135-143; $\beta 7$, aa150-154;

β 8, aa160-168; β 9, aa215-222; β 10, aa226-229). In α 2 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa20-30; α 2, aa81-100; α 3, aa107-124; α 4, aa167-178; α 5, aa184-198; α 6, aa223-231; β 1, aa9-13; β 2, aa34-39; β 3, aa44-49; β 4, aa66-68; β 5, aa72-76; β 6, aa131-139; β 7, aa145-149; β 8, aa155-163; β 9, aa208-214; β 10, aa219-220). In α 3 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa18-29; α 2, aa80-100; α 3, aa107-124; α 4, aa168-178; α 5, aa186-200; α 6, aa230-248; β 1, aa10-12; β 2, aa33-38; β 3, aa43-48; β 4, aa66-68; β 5, aa72-76; β 6, aa131-139; β 7, aa146-150; β 8, aa157-164; β 9, aa211-217; β 10, aa224-227). In α 4 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa17-27; α 2, aa78-97; α 3, aa104-121; α 4, aa165-176; α 5, aa183-198; α 6, aa222-243; β 1, aa6-10; β 2, aa31-36; β 3, aa41-46; β 4, aa62-65; β 5, aa69-73; β 6, aa128-136; β 7, aa143-147; β 8, aa154-161; β 9, aa206-212; β 10, aa217-219). In α 5 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa22-32; α 2, aa83-102; α 3, aa109-120; α 4, aa174-185; α 5, aa191-206; α 6, aa231-240; β 1, aa11-15; β 2, aa36-41; β 3, aa46-51; β 4, aa66-70; β 5, aa74-78; β 6, aa138-146; β 7, aa152-156; β 8, aa162-170; β 9, aa215-221; β 10, aa226-228). In α 6 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa19-30; α 2, aa79-98; α 3, aa105-122; α 4, aa165-176; α 5, aa184-199; α 6, aa226-236; β 1, aa10-13; β 2, aa35-39; β 3, aa45-49; β 4, aa62-66; β 5, aa70-74; β 6, aa129-137; β 7, aa143-147; β 8, aa154-161; β 9, aa210-216; β 10, aa221-223). In α 7 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa22-32; α 2, aa83-102; α 3, aa109-124; α 4, aa170-180; α 5, aa187-202; α 6, aa229-245; β 1, aa13-15; β 2, aa36-41; β 3, aa46-51; β 4,

aa67–70; β 5, aa74–78; β 6, aa133–141; β 7, aa148–152; β 8, aa160–166; β 9, aa212–219; β 10, aa224–227). In β 1 subunits, 5 α -helix and 11 β -sheet secondary structures were predicted to be highly conserved (α 1, aa52–69; α 2, aa76–90; α 3, aa132–143; α 4, aa149–166; α 5, aa191–200; β 1, aa2–8; β 2, aa12–16; β 3, aa25–27; β 4, aa34–37; β 5, aa41–45; β 6, aa95–103; β 7, aa110–114; β 8, aa120–122; β 9, aa124–128; β 10, aa174–180; β 11, aa185–189). In β 2 subunits, 5 α -helix and 12 β -sheet secondary structures were predicted to be highly conserved (α 1, aa52–67; α 2, aa76–90; α 3, aa131–142; α 4, aa148–165; α 5, aa194–201; β 1, aa2–8; β 2, aa12–17; β 3, aa25–27; β 4, aa33–37; β 5, aa41–45; β 6, aa95–103; β 7, aa109–113; β 8, aa123–127; β 9, aa173–179; β 10, aa184–186; β 11, aa210–213; β 12, aa215–225). In β 3 subunits, 4 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa56–76; α 2, aa83–97; α 3, aa142–152; α 4, aa159–175; β 1, aa9–14; β 2, aa18–24; β 3, aa32–34; β 4, aa41–44; β 5, aa48–52; β 6, aa103–111; β 7, aa119–123; β 8, aa134–138; β 9, aa184–190; β 10, aa194–199). In β 4 subunits, 5 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa32–35; α 2, aa50–70; α 3, aa77–93; α 4, aa136–147; α 5, aa153–170; β 1, aa3–8; β 2, aa12–17; β 3, aa26–27; β 4, aa36–38; β 5, aa42–46; β 6, aa99–107; β 7, aa114–118; β 8, aa129–132; β 9, aa178–184; β 10, aa189–190). In β 5 subunits, 6 α -helix and 11 β -sheet secondary structures were predicted to be highly conserved (α 1, aa31–34; α 2, aa49–69; α 3, aa76–91; α 4, aa132–143; α 5, aa149–166; α 6, aa189–200; β 1, aa2–8; β 2, aa12–16; β 3, aa25–26; β 4, aa35–37; β 5, aa41–45; β 6, aa96–104; β 7, aa110–114; β 8, aa120–122; β 9, aa124–128; β 10, aa174–180; β 11, aa185–186). In β 6 subunits, 4 α -helix and 11 β -sheet secondary structures were predicted to be highly conserved (α 1, aa58–77; α 2, aa85–99; α 3, aa142–152; α 4, aa168–184; β 1, aa5–6; β 2, aa11–16; β 3, aa20–26; β 4, aa34–36; β 5,

aa42–46; β 6, aa50–54; β 7, aa105–113; β 8, aa120–124; β 9, aa134–138; β 10, aa193–199; β 11, aa203–210). In β 7 subunits, 6 α -helix and 10 β -sheet secondary structures were predicted to be highly conserved (α 1, aa39–41; α 2, aa57–77; α 3, aa85–100; α 4, aa143–154; α 5, aa162–179; α 6, aa207–216; β 1, aa10–16; β 2, aa20–25; β 3, aa33–35; β 4, aa42–45; β 5, aa49–53; β 6, aa107–115; β 7, aa121–125; β 8, aa136–139; β 9, aa187–193; β 10, aa198–199).

A similar approach was employed for PfGAPDH in Chapter 2, where SCRs were identified and tethered during flexible docking with SA_Docking, allowing side-chain rearrangements and modest backbone adjustments near the active site while maintaining the overall domain organization. For nsp10, nsp14 and nsp16 (Chapter 7), the concept of a rigid core and flexible periphery was used more at the level of analysis than as an explicit restraint scheme: clustering, RMSF profiles and computational alanine scanning were interpreted in light of the distinction between conserved structural scaffolds and variable loops or terminal regions. In *Pf*GAPDH, 8 α -helix and 16 β -sheet secondary structures were predicted to be highly conserved (α 1, aa13–23; α 2, aa40–48; α 3, aa105–113; α 4, aa155–167; α 5, aa198–207; α 6, aa214–227; α 7, aa258–270; α 8, aa323–335 β 1, aa4–8; β 2, aa29–34; β 3, aa58–62; β 4, aa66–70; β 5, aa73–79; β 6, aa93–98; β 7, aa118–123; β 8, aa131–136; β 9, aa146–149; β 10, aa171–182; β 11, aa234–238; β 12, aa244–252; β 13, aa275–278; β 14, aa296–299; β 15, aa303–306; β 16, aa310–317).

In summary, the combination of careful selection of experimental structures, a uniform preparation and protonation protocol, targeted homology modelling of missing regions and explicit treatment of

structurally conserved cores provided a coherent structural framework for all docking, molecular dynamics and bioinformatics studies presented in this thesis.

5. Docking and dynamic docking protocols

In this thesis, several complementary docking strategies were employed to characterise ligand binding modes to enzymes, ion channels and large macromolecular assemblies. Rather than relying on a single protocol, the choice of method was adapted to the size and flexibility of the system, to the availability of experimental binding poses and to the specific questions addressed in each chapter. Broadly, three families of approaches were used: (i) flexible docking with Monte Carlo/energy minimization ; (ii) grid-based docking in pre-defined binding regions (LigandFit and CDOCKER), and (iii) docking followed by simulated annealing (“dynamic docking”) to account for large-scale rearrangements. In addition, a dedicated covalent docking workflow was implemented for reactive THIA analogues.

5.1. General considerations

All docking calculations were carried out on protein structures prepared as described in Section 5 and on ligand conformers generated and refined according to the conformational analysis protocol (Section 4). For each system, the docking setup involved three key steps:

i) Definition of starting structures and binding regions. When experimental complexes were available (e.g. PfGAPDH with a glyceraldehyde 3-phosphate analogue, Cav3.1 with Z944, 3CL^{pro} with GRL-2420), they were used to define the approximate binding site location, the orientation of key pharmacophoric elements and, in some cases, to calibrate scoring and grid parameters. For H₂TPPS-h20S (Chapter 4), initial binding poses were derived by superimposing the tetra-anionic sulfonate pharmacophore onto solvent-exposed positively charged residues identified at regulatory-particle interaction sites; ii) Sampling of ligand position and orientation. Depending on the protocol, ligand translational, rotational and torsional degrees of freedom were sampled using Monte Carlo moves (SA_Docking), random placement and rigid-body rotation in a grid (CDOCKER)[12], or shape-based placement combined with local minimisation (LigandFit)[13]. In all cases, the aim was to generate a diverse ensemble of low-energy poses compatible with steric constraints and basic electrostatic complementarity; iii) Energy evaluation, filtering and clustering. Generated poses were subjected to local energy minimisation and ranked according to appropriate energy functions, typically the sum of van der Waals and electrostatic contributions, with system-specific dielectric constants and cut-off distances. Redundant poses were removed by RMSD-based clustering of ligand heavy atoms, and only representative poses from each cluster were retained for further analysis.

Where relevant, the protein was treated as fully rigid, partially flexible (with flexible side chains and restrained backbone) or globally flexible with explicitly defined structurally conserved regions (SCRs) to prevent unphysical distortions. The criteria used to select final docking solutions are detailed below for each protocol.

5.2. Flexible docking (*PfGAPDH* and *h20S* proteasome)

For *PfGAPDH* inhibitors (Chapter 2) and for *H₂TPPS* binding to the human 20S proteasome (Chapter 4), flexible docking calculations were performed using the SA_Docking procedure in the Affinity module of the Insight 2005 suite. This protocol treats both ligand and protein side chains as flexible and is particularly suited for systems where induced fit and side-chain rearrangements are expected to play a significant role.

For each ligand, a subset of low-energy conformers obtained from the MM conformational analysis (Section 4) was first placed in an approximate binding orientation: in *PfGAPDH*, within the catalytic pocket defined by the co-crystallised glyceraldehyde 3-phosphate analogue; in *h20S*, by fitting the *H₂TPPS* sulfonate pharmacophore onto clusters of positively charged residues located in grooves between α -subunits (e.g. $\alpha 1$ – $\alpha 2$, $\alpha 3$ – $\alpha 4$). These rough complexes were then subjected to SA_Docking.

The docking procedure consisted of iterative Monte Carlo moves in which the ligand was randomly translated, rotated and rotated around its rotatable bonds. For each trial configuration, the non-bonded interaction energy was evaluated with scaled van der Waals and Coulombic terms (typically multiplied by a factor of 0.1 in the early stages) to avoid excessive energy penalties due to transient clashes. Configurations whose energy exceeded that of the last accepted structure by more than a predefined tolerance (up to 10^6 kcal·mol⁻¹ in the most permissive searches) were rejected without minimisation. Accepted configurations were subjected to local energy minimisation using the Conjugate Gradient algorithm[5], with dielectric constants tailored to the environment ($\epsilon = 1$ or $2 \cdot r$ within the protein interior).

To control the exploration of conformational space, an upper limit was imposed on the number of acceptable structures per run (typically 20), and a structural similarity filter was applied using an RMSD tolerance of 0.3 Å on ligand heavy atoms. Only poses with sufficiently low energy and distinct geometries were retained. In PfGAPDH and h20S calculations, the protein backbone in structurally conserved regions was restrained using quadratic positional restraints derived from PredictProtein reliability indices (Section 5), whereas side chains and non-conserved segments were left free to move. This ensured a realistic degree of plasticity at the binding site while preserving the global fold.

Docked complexes were finally ranked by their interaction energy (van der Waals + electrostatic) computed with either the group-based or atom-based method and analysed for geometrical consistency with known mechanistic features (e.g. appropriate orientation of catalytic residues, distances and angles compatible with hydrogen bonds or nucleophilic attacks). Only poses satisfying both energetic and geometric criteria were taken forward.

5.3. Grid-based docking: LigandFit and CDOCKER

Grid-based docking methods were used when the binding site could be clearly defined and when a larger number of starting conformers needed to be screened efficiently. Two implementations were adopted: LigandFit for 3CLpro (Chapter 6) and CDOCKER for Cav3.1 ligands DM1 and DM2 (Chapter 5).

In the Cav3.1 case, the experimentally determined structure of the channel in complex with Z944 (PDB ID: 6KZP) was first refined as described in Section 5. DM1 and DM2 DFT conformers (Section 4) were then placed in the channel pore by matching key pharmacophore elements

shared with Z944. Docking was performed using CDOCKER, a CHARMM-based docking tool that combines random rigid-body placement of the ligand within a user-defined binding region with local minimization. Two binding spheres were explored for each conformer: a sphere of 20 Å around the mass centre of the ligand (binding site BS1) and a sphere of 21.7 Å around the centre of the channel (binding site BS2), thereby probing both local and more extended binding environments.

For each docking run, up to 500 random orientations were generated by translating the ligand centre within the binding sphere and applying random rotations. A softened van der Waals potential was used to allow partial penetration into grid cells during this exploratory phase; orientations with an energy higher than 1000 kcal·mol⁻¹ were discarded. Pose generation proceeded until either 500 low-energy orientations were found or a maximum number of rejected orientations was reached. The remaining poses were minimised and clustered using a pose cluster radius of 0.5 Å. Representative poses from each cluster were then subjected to further MM minimisation with $\epsilon = 4 \cdot r$, allowing side-chain relaxation within the binding site while keeping the rest of the protein relatively rigid. Ligand–protein interaction energies were calculated for residues within 5 Å of the ligand, and poses were ranked accordingly.

For THIA derivatives binding to 3CLpro (Chapter 6), a shape-based grid docking approach was used via the LigandFit procedure (cff force field). Docking parameters (RMS thresholds for ligand–site matching, grid dimensions, dielectric constant, softened Lennard-Jones 9–6 potential) were calibrated by re-docking the covalent inhibitor GRL-2420 into 3CL^{pro} (PDB ID: 7JKV), deliberately perturbing its position and verifying that the protocol was able to recover the crystallographic binding

mode. DFT conformers of THIA-2, THIA-4, THIA-7, THIA-8 and THIA-10 were placed in the catalytic site so that the R substituent occupied the S1 subsite, the R' substituent the S2 subsite and the N2 nitrogen the oxyanion hole. For each conformer, 20 poses were generated, minimised (steepest descent followed by BFGS, $\epsilon = 80 \cdot r$), and filtered by RMSD (threshold 1.5 Å) and by distance criteria relative to the catalytic cysteine (e.g. S–S distance < 4 Å). The most favourable binding modes (BM-I–BM-IV) were defined based on the positioning of R and R' in the 3CLpro subsites, and only those consistent with a plausible pre-covalent orientation were retained.

5.4. Dynamic docking and simulated annealing refinement

In some systems, particularly in the 20S proteasome (Chapter 4), the conformational changes associated with ligand binding are not limited to local side-chain movements but may involve more extensive rearrangements, such as gate opening or inter-subunit breathing motions. To account for such effects, docking was complemented by short molecular dynamics simulations at variable temperature, i.e. simulated annealing (SA), starting from the docked complexes; this strategy is referred to here as “dynamic docking”.

For H₂TPPS–h20S complexes, representative docked poses obtained by SA_Docking were first minimised and then subjected to short SA runs in which the temperature was gradually decreased from 500 to 300 K over 5 ps. The simulations were carried out in the NVT ensemble using a Verlet velocity integrator with a time step of 1 fs. Initial velocities were drawn from a Boltzmann distribution at the starting temperature, and in subsequent stages they were taken from restart data. During the early stages at 500 K, van der Waals and Coulombic contributions were scaled

by a factor of 0.1 to facilitate crossing of torsional barriers and exploration of the conformational landscape. As the temperature was reduced, these scaling factors were increased back to 1.0, allowing the system to relax into local minima under full non-bonded interactions. Structurally conserved regions (Section 5.4) were restrained throughout, whereas binding-site residues and the ligand remained fully flexible.

At the end of the SA protocol, a final round of energy minimization ($\epsilon = 80 \cdot r$; RMS gradient $\leq 0.1 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$) was performed, and the resulting complexes were saved for analysis. Non-bonded interaction energies between H₂TPPS and h20S residues were computed, and complexes were ranked by ligand–protein interaction energy and filtered by ligand conformational energy ($\Delta E_{\text{GM}} \leq 10 \text{ kcal} \cdot \text{mol}^{-1}$). The selected complexes were subsequently examined to identify binding modes that were both energetically favourable and structurally consistent with known regulatory particle interaction sites.

5.5. Covalent docking of THIA derivatives

For THIA-3, THIA-4, THIA-6 and THIA-8, which were designed as electrophilic inhibitors of cysteine proteases (3CLpro, papain, cathepsin L), a covalent docking workflow was adopted (Chapter 6). In this case, the goal was not only to place the ligand in a pre-reactive pose but also to model the formation of a covalent bond between the THIA sulfur atom and the catalytic cysteine.

Covalent docking was performed using AutoDock4 with the flexible side-chain method. Receptors were prepared from experimental structures (e.g. 3CLpro: PDB ID 7SF1; papain: 1PE6; cathepsin L: 7W34) by removing co-crystallised ligands, ions and water molecules and by adding hydrogens. THIA ligands were parametrised using SwissParam[14] and

LigParGen[15], and converted into mol2 format. Docking simulations employed the Lamarckian genetic algorithm with default parameters, and ligand poses were ranked according to the AutoDock semi-empirical scoring function.

The resulting complexes were examined to identify plausible covalent geometries, and manual iterative adjustments were then performed in Coot [16] to refine the covalent bond and local geometry. Final complexes were inspected visually and used primarily for qualitative analysis and figure preparation, rather than for quantitative energetic comparison, which was reserved for non-covalent pre-reactive complexes and for SMD trajectories.

6. Molecular dynamics and steered molecular dynamics simulations

Molecular dynamics (MD) simulations were used in this thesis in a targeted way, primarily to refine docked complexes and to probe the early stages of ligand approach to catalytic sites under explicitly solvated conditions. In particular, steered molecular dynamics (SMD) was employed to investigate the binding/approach pathways of THIA derivatives towards the catalytic cysteine of SARS-CoV-2 3CL^{pro} (Chapter 6), building on the pre-reactive binding modes identified by docking.

Whereas docking and dynamic docking (Section 6) provided a static or quasi-static picture of favourable binding poses, MD and SMD were used to (i) relax protein–ligand complexes in an explicit solvent environment, (ii) assess the stability of the proposed binding mode, and (iii) explore possible exit/entry pathways compatible with the geometry of the active site and surrounding subsites.

6.1. System setup and equilibration

All MD and SMD simulations were carried out on protein–ligand complexes prepared as described in Sections 4 and 5, using force-field parameters derived from CHARMM for proteins and from the conformational/DFT protocols for the THIA ligands.

Starting from the selected docked complexes of THIA derivatives with 3CLpro (binding mode BM-I), each system was embedded in an explicit solvent box. A cubic periodic simulation box was constructed such that all protein atoms were at least 12 Å away from the box edges in every direction. The systems were solvated with TIP3P water molecules, and counterions (Cl^-) were added to neutralise the total charge, yielding a final model that approximates a dilute aqueous solution at physiological ionic strength.

All systems were parametrised using the CHARMM force field. This choice ensured consistency with the docking and bioinformatics analyses performed elsewhere in the thesis and provided a well-tested description of both protein and water interactions. THIA derivatives were treated with parameters compatible with CHARMM, as derived from the previous conformational and docking workflows.

Prior to production dynamics, each system underwent an initial energy minimization to remove unfavourable contacts introduced during solvation or ion placement. Minimization was performed using the Conjugate Gradient algorithm until the average RMS gradient fell below $0.1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$. During this phase, no external forces were applied apart from those implicit in the force field.

Following minimization, a short equilibration stage was carried out under an isothermal–isobaric (NPT) ensemble, with protein and ligand

atoms initially held fixed and only solvent molecules allowed to move. Temperature was maintained at 298 K using the Nosé–Hoover thermostat, and pressure was controlled at 1 atm via the Langevin piston method. This step allowed the solvent to relax around the solute and the box dimensions to adjust to a physically reasonable density before releasing constraints on the solute in the subsequent SMD simulations.

Throughout minimisation and equilibration, bonds involving hydrogen atoms were constrained using the SHAKE algorithm, allowing the use of a 1 fs time step. Long-range electrostatics were treated with the particle mesh Ewald (PME) method, and non-bonded interactions were truncated at a 12 Å cut-off, with appropriate switching functions to avoid discontinuities near the cut-off.

The equilibrated configurations obtained at the end of this procedure were used as starting points for the SMD runs.

6.2. Steered molecular dynamics protocol

Steered molecular dynamics was used to explore the initial approach and potential exit pathways of THIA derivatives relative to the 3CLpro active site, with a particular focus on the geometry of the catalytic dyad and subsites S1, S2 and S3. The aim was not to simulate the full reaction coordinate of covalent bond formation, but rather to probe whether the docked pre-reactive binding mode (BM-I) could be maintained under mechanical perturbation and to identify directions of motion that are compatible with the architecture of the pocket.

For each THIA–3CLpro complex, a 500 ps SMD simulation was carried out with a time step of 1 fs, starting from the equilibrated structure. The pulling protocol was defined by choosing a pair of atoms between which a harmonic steering potential was applied: the sulfur atom of the

catalytic cysteine (Cys145) and either (i) the para carbon of the R substituent (for THIA-4, THIA-7, THIA-8 and THIA-10), or (ii) the methylene carbon adjacent to the reactive centre (for THIA-2). A spring constant of $350 \text{ pN} \cdot \text{\AA}^{-1}$ (corresponding to approximately $5 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$) was used, and the pulling speed was set to $0.04 \text{ \AA} \cdot \text{ps}^{-1}$.

Based on the geometry of the active site and the subsites identified from docking, two main pulling directions were defined: (i) through the S3 subsite, which represents the widest channel leading out of the catalytic pocket, and (ii) along trajectories that traverse the “upper” or “lower” loops flanking the binding site, as determined from structural analysis (Chapter 6). These directions were chosen to test whether the ligand could plausibly enter or exit the pocket without requiring unrealistic deformations of the protein.

During SMD, several restraints were maintained to preserve key mechanistic features. As in the docking stage, a tethering restraint was applied to the hydrogen bond between Cys145 and His41, constraining the donor–acceptor distance within $3.0 \text{ \AA}$ with a force constant of $100 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$, thereby preserving the catalytic dyad geometry. In addition, dihedral restraints (force constant $100 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$) were applied to the torsion angles defining the THIA conformation obtained from DFT calculations, ensuring that the ligand maintained a chemically reasonable internal geometry even under mechanical perturbation.

Temperature was kept at 298 K (Nosé–Hoover thermostat), and pressure at 1 atm (Langevin piston), using the same NPT settings as in the equilibration phase. PME and SHAKE were retained for electrostatics and bond constraints, respectively. Trajectory frames were saved every 5 ps for subsequent structural and energetic analysis.

6.3. Trajectory analysis and criteria for binding mode preservation

The SMD trajectories were analysed to assess both the stability of the BM-I binding mode and the response of the protein to the applied steering forces.

For each simulation, non-bonded ligand–protein interaction energies (van der Waals + electrostatic contributions) were computed for all frames, using an atom-based cut-off scheme with a non-bonded list radius of 15 Å, higher and lower cut-off distances of 14 and 11 Å, respectively, and an effective dielectric $\epsilon = 2 \cdot r$. Water molecules that mediated ligand–protein interactions were included in the energy evaluation when they formed persistent bridges between THIA and active-site residues. Geometric interaction criteria (hydrogen bonds, hydrophobic contacts, π – π and cation– π interactions) were analysed using the Receptor–Ligand Interaction tools in Discovery Studio.

In parallel, root mean square fluctuation (RMSF) values were calculated for each protein residue over the course of the SMD, after aligning all trajectory frames to the initial structure using C α atoms. Residues with RMSF values greater than 2 Å were considered to exhibit notable flexibility, and their location relative to the active site and substrate-binding subsites was examined to identify regions that contribute to accommodating ligand motion.

To evaluate whether a given frame preserved the BM-I binding mode, a set of geometric criteria was applied: i) the distance between the THIA S1 sulfur atom and the sulfur atom of Cys145 remained below 4.0 Å; ii) the N2 nitrogen was within 4.5 Å of the catalytic cysteine hydrogen; iii) the N2 and N4 substituents occupied the S2 and S3 subsites identified in the

docking phase; iv) the hydrogen bond with Gln189, characteristic of the proposed binding mode, was maintained.

Frames satisfying all these conditions were classified as preserving BM-I, whereas frames in which one or more criteria were violated were considered to represent deviations or alternative transient arrangements. The time evolution of these criteria along the SMD trajectory was used to judge the robustness of the proposed binding mode under mechanical perturbation along the chosen pathways.

Finally, to ensure consistency between the energetic descriptions used in docking and in SMD, non-bonded interaction energies for the docked complexes (previously evaluated with the CFF force field) were recalculated with CHARMM and the same cut-off and dielectric settings as in the SMD analysis. This allowed a direct comparison between static docking results and dynamic SMD behaviour within a uniform force-field framework.

In summary, the MD and SMD simulations provided a dynamic complement to the docking studies, offering insight into the stability of the THIA binding mode, the flexibility of the 3CLpro active site, and the plausibility of specific entry and exit pathways within the constraints of the catalytic architecture.

7. Bioinformatics and structural analyses

Bioinformatics and structure-based analyses were used throughout this thesis to (i) extract geometric information from experimentally determined complexes, (ii) relate ligand-based pharmacophores to protein–protein interaction motifs, (iii) characterise binding sites across homologous proteins, and (iv) quantify the contribution of individual residues to protein–protein interfaces. Although the specific questions differed among

chapters, all analyses were grounded on a common workflow: structural alignment, identification of functionally relevant residues, and quantitative evaluation of distances, surfaces and interaction energies.

7.1. Analysis of protein–protein interaction motifs and ligand pharmacophores

In Chapter 3, bioinformatics and structural tools were combined to link the conformational space of tetra-substituted pyrrole (TSP) derivatives to known protein–protein interaction (PPI) motifs.

Experimentally determined structures of paradigmatic PPIs were first collected: the two-stranded parallel coiled coil of the GCN4 leucine zipper (PDB ID: 2ZTA), the antiparallel coiled coil of GreA (PDB ID: 1GRJ), and the LxxLL motif of nuclear receptor coactivator 5 in complex with the estrogen receptor β (PDB ID: 2J7X). These structures were analysed using the Simulation and Macromolecule tools of Discovery Studio.

For each motif, “hot spot” residues were identified based on their position within canonical patterns (coiled coil: i, i+3, i+4, i+7; LxxLL: i, i+3, i+4). The centroid of each hot-spot side chain was defined as follows: ring carbon atoms for aromatic residues and all heavy atoms of the side chain for aliphatic residues. Distances between centroids were computed to generate “distance maps” characteristic of each PPI motif.

These distance maps were then compared to the ligand-based pharmacophores derived from low-energy DFT conformers of TSP derivatives (Section 4). In TSPs, the centroids of the aromatic rings X, Y and Z were used as pharmacophoric points, and inter-centroid distances were calculated for all conformers within a chosen energy window ($\Delta E_{\text{GM}} \leq 2 \text{ kcal}\cdot\text{mol}^{-1}$). TSP conformers were considered compatible with a given PPI motif when two pharmacophoric distances differed by

less than 1.0 Å and a third one by less than 1.5 Å from the experimentally observed distances between hot-spot residues.

Conformers satisfying these criteria were grouped into pharmacophore families, and for each family an “average pharmacophore” was obtained by averaging the ring–ring distances. Representative conformers of each pharmacophore were superimposed on the PPI motifs by fitting the X–Y–Z centroids onto the hot-spot centroids, and the resulting root-mean-square deviation (RMSD) values were used to quantify mimetic quality. Conformers with RMSD < 2.5 Å were classified as good mimics of the corresponding coiled-coil or LxxLL motif.

A similar procedure was applied to the analysis of p14ARF (Chapter 3). After building a homology model of human p14ARF based on mouse p19ARF (PDB ID: 1HN3), potential protein-recognition motifs on the first two N-terminal helices (α H1, α H2) were identified using the Predict Sequence Properties protocol in Discovery Studio and consensus patterns of hydrophobic residues. The hot-spot residues of these motifs were mapped onto the model, centroids were calculated, and selected TSP conformers were superimposed by fitting their pharmacophore centroids onto the side-chain centroids of the putative hot spots. RMSD values were then used to assess how well individual conformers could mimic these newly identified motifs.

7.2. Surface mapping and interaction analysis on the 20S proteasome

In Chapter 4, bioinformatics and structural tools were used to characterise the surface of the human 20S proteasome (h20S) and to relate it to the binding patterns of regulatory particles and small-molecule activators.

First, a collection of experimentally determined complexes of h20S with regulatory particles (19S, PA28, PA200) and of *Thermoplasma acidophilum* 20S with ZYA was assembled. All structures were parametrised with CHARMM and processed with the Prepare Protein tool in Discovery Studio, adding missing side chains and assigning protonation states at pH 7.2.

Using the Receptor–Ligand Interactions tool, positively charged residues (Lys, Arg, protonated His) involved in ionic interactions with regulatory particles were identified from the contact patterns observed in h20S–19S, h20S–PA28 and h20S–PA200 complexes. In parallel, residues interacting with Rpt2 and Rpt3 HbYX motifs, PA28 activation loops and the ZYA activator in t20S were catalogued. Information from the t20S–ZYA complex (PDB ID: 8F7K) was transferred to h20S via secondary structure-based sequence alignment (Macromolecules tools, Discovery Studio). This allowed the mapping of ZYA-relevant positions onto the human proteasome.

The solvent-accessible surface area (SASA) of each identified positively charged residue was calculated using a probe radius of 1.4 Å, corresponding to a water molecule. Residues with a solvent accessibility higher than 25% were considered solvent-exposed and potential candidates for interaction with the tetra-anionic H₂TPPS. The spatial distribution of these residues was then examined across the closed, semi-open and open conformations of h20S, focusing in particular on grooves between α -subunits such as $\alpha 1$ – $\alpha 2$ and $\alpha 3$ – $\alpha 4$, which coincide with canonical regulatory particle docking sites.

These mapped surfaces provided the structural basis for generating initial H₂TPPS binding poses (Section 6) by fitting the square-planar

sulfonate pharmacophore to specific clusters of solvent-exposed, positively charged residues, thereby linking ligand design to the natural regulatory landscape of the proteasome.

7.3. Comparative analysis of Cav channels and cysteine proteases

In Chapters 5 and 6, structural and bioinformatics analyses were applied to families of related proteins to identify conserved binding site features and to rationalise ligand selectivity.

For Cav channels (Chapter 5), structures of Cav3.1 (apo and Z944-bound) and Cav1.1 (verapamil- and diltiazem-bound) were superimposed using $\text{C}\alpha$ -based 3D structural alignment (Align Structures protocol, Discovery Studio), with a distance cutoff and minimal aligned length chosen to ensure reliable backbone superposition. Following alignment, the pore region and the inner vestibule were compared in terms of side-chain composition and orientation, focusing on residues within 5 Å of the ligands. Non-bonded interaction energies between ligands and neighbouring residues were calculated for each complex (van der Waals + electrostatics), and residues contributing at least $-1 \text{ kcal}\cdot\text{mol}^{-1}$ were flagged as relevant contact points. The combination of structural superposition, residue-level energies and geometric interaction criteria enabled a comparative view of how different ligands exploit conserved and non-conserved elements of the channel architecture.

For cysteine proteases (Chapter 6), a multi-step alignment strategy was adopted. First, secondary structure-based sequence alignments of cathepsin L (PDB ID: 3OF8), papain (PDB ID: 6TCX) and bromelain (PDB ID: 6YCG) were generated using the Align123 algorithm. Structures were superimposed by fitting $\text{C}\alpha$ atoms of aligned residues, calculating pairwise RMSD values and iteratively refining the superposition by

retaining only C α pairs with RMSD < 1 Å. SARS-CoV-2 PL^{pro} (PDB ID: 6WX4) was then anchored to this superposition by fitting its catalytic residues (Cys–His–Asp) onto those of cathepsin L, and 3CL^{pro} (PDB ID: 7JKV) was added by superimposing the backbone atoms of the P1 and P2 residues of its peptide inhibitor onto the corresponding residues of the PL^{pro} inhibitor VIR251. Additional 3CL^{pro} and PL^{pro} structures (e.g. 7N89, 7JIW) were incorporated using the same scheme.

Once in a common frame of reference, the active sites and subsites (S1, S2, S3) of these proteases were analysed. DFT conformers of THIA derivatives were manually fitted into the active sites of PL^{pro}, cathepsin L, papain and bromelain by mapping the centroid of their N2 and N4 phenyl rings onto the centroids of the P1 and P2 side chains of substrate or inhibitor residues, and aligning the reactive sulfur atom with the carbon of the covalent bond in the reference inhibitor. The resulting complexes were inspected using geometric interaction criteria to identify hydrogen bonds, hydrophobic interactions and electrostatic contacts, and the SASA of the catalytic cysteine sulfur atom was calculated for each protease. This comparative analysis informed hypotheses on the accessibility and reactivity of the catalytic cysteine across different proteases and on the potential off-target profile of THIA derivatives.

7.4. Structural clustering, hotspot prediction and computational alanine scanning (nsp10, nsp14, nsp16)

Chapter 7 focused on SARS-CoV-2 non-structural proteins nsp10, nsp14 and nsp16, combining large-scale structural clustering, computational alanine scanning (CAS) and sequence alignment to identify hotspot residues at protein–protein interfaces and to relate them to functional states.

All selected structures containing nsp10, nsp14 or nsp16 (alone or in complexes) were prepared as described in Section 5. For the analysis of functional states, complexes such as nsp10–nsp16 and nsp10–nsp14 were first grouped according to their functional annotations (e.g. apo, substrate-bound, cofactor-bound). To compare structures across different viruses and complexes, each protein was extracted from the corresponding complex and subjected to structural clustering using the Align Structures protocol in Discovery Studio. The protocol employs the 3DMA program to perform structural alignment based on three-dimensional similarity. A $\text{C}\alpha$ distance cutoff of 0.3 Å and a minimum aligned length of 50 residues were used to define a common core, and a bin size of 20° for $\text{C}\alpha$ pseudo-torsion angles was adopted in the initial pairwise comparison. The “Fuzzy Bins” option was enabled to increase tolerance for local variations. Based on these criteria, structures were partitioned into structural clusters, each associated with a distinct conformational/functional state; proteins that could not be aligned with others at this level of stringency were excluded from the clustering.

Computational alanine scanning was performed on full-length nsp16–nsp10 and nsp14–nsp10 complexes using the Calculate Mutation Energy (Binding) protocol in Discovery Studio. For each residue, the mutation to alanine was modelled in both the bound and unbound states, and the change in binding free energy ($\Delta\Delta G_{\text{mut}}$) was computed as the difference between mutant and wild-type binding free energies. Binding free energies were estimated using a CHARMM-based energy function including van der Waals interactions, electrostatics (Generalised Born implicit solvent), a non-polar solvation term and an empirical side-chain entropy contribution. The pH-dependent mode was employed with pH 7.2 and ionic strength 0.1 M, accounting for changes in residue ionisation upon

mutation. Residues were classified as low ($0.5\text{--}1.0\text{ kcal}\cdot\text{mol}^{-1}$), medium ($1.0\text{--}2.0\text{ kcal}\cdot\text{mol}^{-1}$) or high ($>2.0\text{ kcal}\cdot\text{mol}^{-1}$) contributors to binding. Residues showing $\Delta\Delta G_{\text{mut}} \geq 1\text{ kcal}\cdot\text{mol}^{-1}$ in at least one structure were flagged as potential hotspots.

To interpret CAS results in a structural context, all nsp16–nsp10 complexes were superimposed using the SARS-CoV-2 heterodimer (PDB ID: 7ULT) as reference, and all nsp14–nsp10 complexes were superimposed on the crystal structure of SARS-CoV-2 nsp10 bound to the nsp14 ExoN domain (PDB ID: 7N0B). Initial secondary structure-based sequence alignment was performed using Align123 (DSC method; BLOSUM scoring; appropriate gap penalties), followed by C α -based superposition and iterative refinement restricted to C α pairs with RMSD $< 0.5\text{ \AA}$. On this common scaffold, the location and frequency of hotspots across structures and conformational clusters were analysed, together with the distribution of low-, medium- and high-impact $\Delta\Delta G_{\text{mut}}$ values.

In addition, RMSF values were computed to quantify the structural variability of nsp10, nsp14 and nsp16 across the ensemble of available structures. For each protein, non-conserved termini and poorly represented loops (e.g. the 123–154 loop in nsp14 ExoN) were excluded. The remaining core residues were superimposed as described above, and per-residue RMSF values were extracted. Combining RMSF profiles with CAS allowed the distinction between rigid, high-impact hotspots and more flexible, context-dependent contributors to binding.

7.5. Cross-species sequence alignment of nsp10, nsp14 and nsp16

Finally, to relate hotspot patterns and structural features to viral evolution, a cross-species alignment of nsp10, nsp14 and nsp16 was performed across selected human coronaviruses (Chapter 7). Sequences

were retrieved from UniProt for SARS-CoV-2, SARS-CoV, MERS-CoV, HCoV-HKU1, HCoV-OC43, HCoV-NL63 and HCoV-229E, selecting viruses that infect humans and are endemic; the first five belong to the Betacoronavirus genus, whereas HCoV-NL63 and HCoV-229E belong to Alphacoronavirus.

Multiple sequence alignment was carried out using the Align123 algorithm with parameters matching those used for structural alignment in the hotspot analysis. Residues were classified as “identical” when conserved in all sequences, as showing “strong similarity” when belonging to conservative substitution groups (STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW), and as showing “weak similarity” when belonging to broader substitution sets (e.g. CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY). This classification was then overlaid with structural hotspot maps and RMSF profiles to identify positions that are both structurally and evolutionarily constrained, and therefore of particular interest for understanding the robustness and potential druggability of nsp10–nsp14–nsp16 interaction networks.

Taken together, these bioinformatics and structural analyses provided a unifying framework to connect ligand conformations, protein–protein interaction motifs, proteasome regulatory surfaces, channel and protease binding sites, and coronavirus non-structural protein interfaces, supporting the structure-based interpretations and hypotheses developed in the individual chapters.

8. PROTAC ternary complex modelling (LigTBM and RAINDROP workflow)

PROTAC systems pose a distinct modelling challenge, as they require the simultaneous consideration of two protein partners (target and E3 ligase), two ligands and a flexible linker. In this thesis, the ternary assemblies formed by BET bromodomains (BRD2–4), the VHL E3 ligase and PROTAC molecules derived from AT1 and MZ1 were modelled using a two-stage procedure: (i) generation of binary protein–ligand complexes with LigTBM, and (ii) *ab initio* assembly and scoring of ternary complexes with RAINDROP. Post-processing, filtering and analysis of the resulting ensembles were carried out using custom Python scripts.

8.1. Binary complex generation with LigTBM

As a first step, high-quality binary complexes between each bromodomain and its cognate PROTAC warhead, and between VHL and its recruiting ligand, were generated. Experimentally determined structures from the BET family and VHL were selected (BRD2 BD2, BRD3 BD1/BD2, BRD4 BD1/BD2, VHL) and used as structural templates. For each system, the crystallographic coordinates of the protein and the bound ligand were separated into individual files (protein in PDB format, ligand in PDB/mol2 format).

The LigTBM (Ligand Template-Based Modelling) module was then employed to construct the binary complexes [17]. LigTBM uses the experimental protein–ligand structure as a template and performs template-guided docking of the PROTAC warhead into the known binding pocket. Briefly, the protocol combines: i) conformational sampling of the ligand guided by the template geometry and by a machine-learning model trained on structural descriptors and experimental complexes; ii)

optimization of ligand poses based on steric complementarity, hydrogen bonding and hydrophobic interactions within the binding pocket; iii) scoring and ranking of poses using an empirical energy function which includes terms for hydrogen bonds, hydrophobic contacts and a homology-inspired similarity contribution.

For each bromodomain and for VHL, LigTBM returned a set of plausible protein–warhead complexes. The top-scoring complexes, characterised by correct recapitulation of key template interactions and absence of substantial steric clashes, were selected as starting binary assemblies. These binary complexes, together with the explicit linker description in SMILES format and the atom indices of the attachment points on each warhead, were used as input for RAINDROP.

8.2. Ternary complex assembly with RAINDROP

The ternary complex assembly was performed with the RAINDROP protocol, which constructs PROTAC-mediated protein–protein complexes by systematically exploring the relative orientation of two pre-formed binary complexes under explicit linker constraints.

RAINDROP requires as input: (i) the E3–ligand complex (VHL + recruiting moiety), (ii) the target–ligand complex (BET bromodomain + warhead), both in PDB format, (iii) the linker as a SMILES string, and (iv) the indices of the atoms to which the linker is covalently attached on each ligand. In the first stage, one binary complex (typically VHL–ligand) is fixed at the centre of a three-dimensional reference grid. The second complex (bromodomain–ligand) is systematically rotated and translated across the grid; at each relative orientation, RAINDROP samples linker conformations and evaluates whether a geometrically feasible connection between the two warheads can be established without excessive strain.

For each sampled spatial configuration, the protocol evaluates: i) the overlap between the conformational spaces accessible to the linker, given the two attachment points and their separation; ii) the complementarity between the molecular interaction fields (MIFs) of the protein–protein surfaces, computed with the GRID force field; iii) potential steric clashes between the linker and protein surfaces.

Sampling is performed using a rotational step of 6° and a translational step of 1 Å, typically yielding on the order of 10,000–15,000 candidate ternary arrangements for each PROTAC system. Conformations that fail basic geometric or steric feasibility tests are discarded early, while those that are consistent with both linker geometry and protein–protein surface compatibility are retained for energetic evaluation.

For each retained pose, two energy-like quantities are computed: the protein–protein interaction energy (E_{PPI}), capturing the net stabilising contribution of the interface between the target protein and VHL in the presence of the PROTAC, and the internal conformational energy of the PROTAC within the ternary context (E_{PROTAC}), computed from MIF-based terms. Their sum defines a composite score:

$$POSI_e = E_{PPI} + E_{PROTAC}$$

which is used as a global indicator of pose quality. Ternary complexes with favourable E_{PPI} (stabilising protein–protein interactions) and acceptable E_{PROTAC} (realistic linker conformation) tend to exhibit low $POSI_e$ values.

8.3. Energetic filtering and clustering of ternary ensembles

The raw RAINDROP output consists of a large ensemble of ternary models with associated energies and descriptors. To extract a manageable

subset of high-quality poses, a two-stage filtering workflow was implemented in Python.

First, the RAINDROP summary file (containing E_PPI, E_PROTAC, POSI_e and auxiliary geometrical descriptors for each pose) was read into a pandas DataFrame. The apply_filters() function was then used to refine the dataset: i) Primary filter on protein–protein interaction energy. Poses were sorted according to E_PPI, and only those lying below a specified quantile threshold (e.g. the best x% of PPI energies) were retained. This step selects complexes with strong protein–protein interfaces, independently of linker conformational cost; ii) Secondary filter on PROTAC conformational energy. Within this PPI-filtered subset, a second quantile-based filter was applied to E_PROTAC to remove poses with unrealistic linker strain. The remaining configurations represent ternary assemblies with both favourable protein–protein interaction and reasonable PROTAC geometry. iii) Final ranking by POSI_e. The doubly filtered set was finally ordered by POSI_e, and a fixed number of top-ranked complexes (typically 1000, 500 or 250 per system) was selected for detailed analysis. This step provides a compact representation of the energetically most consistent ternary arrangements according to the RAINDROP scoring scheme.

When necessary, additional geometric criteria (e.g. minimum interface area, absence of large steric clashes) were applied to further refine the ensembles or to select representative poses for visual inspection.

8.4. Surface and interface analysis

To relate the ternary models to known principles of PROTAC-mediated degradation, and to quantify the characteristics of the protein–protein

interface, surface-based descriptors were computed using a custom `compute_surfaces` function integrated with the RAINDROP output.

The RAINDROP MIF-based surface data were parsed into pandas DataFrames, and for each set of surface atoms corresponding to a given protein or interface region, three classes of metrics were computed:

- i) Geometric surface metrics, e.g. total surface area or interface area, obtained by summing the contributions of individual surface atoms;
- ii) Energetic surface metrics, integrating local MIF-derived interaction energies over the surface, with missing or undefined values treated as zero;
- iii) Subset-specific metrics, such as energies or areas restricted to particular patches of the interface (e.g. around specific bromodomain helices or VHL loops).

By construction, if the input dataset for a given subset was empty, all metrics were set to zero, ensuring robust behaviour of the analysis code across systems and filters.

These surface descriptors were then combined with the energetic and geometrical information from RAINDROP (E_{PPI} , E_{PROTAC} , POSI_e , relative orientations) to characterise the ternary complexes along dimensions such as interface size, interaction strength, and PROTAC conformational cost. Representative poses were inspected visually to confirm that the predicted assemblies conformed to known structural features of VHL- and BET-bound PROTACs and to explore potential differences between AT1- and MZ1-like constructs.

References

1. Ewig, C.S., et al. "Derivation of Class II Force Fields. VIII. Derivation of a General Quantum Mechanical Force Field for Organic Compounds." *Journal of Computational Chemistry*, vol. 22, no. 15, 2001, pp. 1782-800, doi:10.1002/jcc.1131.
2. Dauber-Osguthorpe, Pnina, et al. "Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system." *Proteins: Structure, Function, and Bioinformatics* 4.1 (1988): 31-47.
3. Dewar, Michael JS, and Walter Thiel. "Ground states of molecules. 38. The MNDO method. Approximations and parameters." *Journal of the American Chemical Society* 99.15 (1977): 4899-4907.
4. Vanommeslaeghe, K., et al. "CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields." *Journal of Computational Chemistry*, vol. 31, no. 4, 2010, pp. 671-90.
5. Fletcher, R. Unconstrained optimization. In *Practical Methods of Optimization*; John Wiley & Sons Ltd.: New York, NY, USA, 1980; Volume 1, pp. 1–128.
6. Frisch, M.J., et al. Gaussian 16, Revision, C.01. Gaussian, Inc., 2016.
7. Becke, A.D. "Density-Functional Thermochemistry. III. The Role of Exact Exchange." *The Journal of Chemical Physics*, vol. 98, 1993, pp. 5648-52.
8. Lee, C., et al. "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density." *Physical Review B*, vol. 37, 1988, pp. 785-89.
9. Cossi, M., et al. "Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model." *Journal of Computational Chemistry*, vol. 24, 2003, pp. 669-81.
10. Reed, A.E., et al. "Natural Population Analysis." *The Journal of Chemical Physics*, vol. 83, 1985, pp. 735-46.
11. Davis, I.W., et al. "MolProbity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids." *Nucleic Acids Research*, vol. 35, Web Server issue, 2007.
12. Wu, G., et al. "A comparative study of methods for calculating protein–ligand binding affinities." *Journal of Computational Chemistry*, vol. 24, no. 13, 2003, pp. 1549-62.

13. Venkatachalam, C.M., et al. "LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites." *Journal of Molecular Graphics and Modelling*, vol. 21, no. 4, 2003, pp. 289-307, doi:10.1016/s1093-3263(02)00164-x.
14. Zoete, V., et al. "SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules." *Journal of Computational Chemistry*, vol. 32, no. 11, 2011, pp. 2359-68, doi:10.1002/jcc.21816.
15. Dodda, L.S., et al. "LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands." *Nucleic Acids Research*, vol. 45, Web Server issue, 2017, pp. W331-36, doi:10.1093/nar/gkx312.
16. Emsley, Paul, et al. "Features and development of Coot." *Biological crystallography* 66.4 (2010): 486-501.
17. Kotelnikov, Sergei, et al. "Accurate ligand–protein docking in CASP15 using the ClusPro LigTBM server." *Proteins: Structure, Function, and Bioinformatics* 91.12 (2023): 1822-1828.

Conclusions

This thesis explored various molecular systems under the hypothesis that protein interactions follow structural patterns that can be studied and modulated through computational approaches. Starting from the concept of a “molecular language,” the work aimed to design and identify new small-molecule modulators—either derived from natural products or obtained through rational design—capable of interfering with elements of this language, such as linear motifs, domains, or hot spots.

Within this framework, the study on nature-inspired 3-Br-Acivicin isomers and derivatives highlighted how stereochemistry influences molecular recognition and biological response, establishing its role as a key factor in the design of bioactive compounds. Further investigations focused on small-molecule scaffolds that mimic protein–protein interaction motifs, leading to the identification of pyrrole derivatives with anticancer activity and porphyrin compounds capable of activating the 20S proteasome. These ligands were shown to reproduce key recognition features at protein interfaces.

In parallel, thiadiazolidine and imidazopyridazine derivatives were studied as inhibitors of cysteine proteases and hCaV3.1 channels, respectively. Their activity was rationalized through DFT and docking studies, which revealed the preferential binding of specific conformational enantiomers, emphasizing the importance of structural complementarity and conformational behavior in modulating biological responses.

Computational alanine-scanning mutagenesis of coronavirus non-structural proteins led to the identification of conserved hot spot residues and novel druggable binding sites. Finally, the PROTAC design study demonstrated the application of Molecular Interaction Fields in optimizing

linker composition for the selective degradation of BET family bromodomains.

In conclusion, the studies presented in this thesis demonstrate that combining computational and experimental approaches provides valuable insights into molecular recognition and supports the rational design of new bioactive compounds. By integrating concepts such as molecular language, stereochemical control, and structure–function relationships, this work contributes to a deeper understanding of how small molecules interact with biological targets—offering a viable strategy to enhance drug discovery efforts.

Appendices

Appendices can be found in the following link:
<https://doi.org/10.5281/zenodo.17884351>.

