Application of Computational Methods for the Design of New Potential Therapeutic Agents

A

DISSERTATION

by

Antonio Di Dato

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by Professor Caterina Fattorusso

Department of Pharmacy

University of Naples Federico II

Naples, Italy

2017

Part of this work have been published in the following articles:

Publications:

1. Andreoli M., Persico M., Kumar V., Pepe A, Kumar A., Alegria A.E., Petrella L., Sevciunaite L., Camperchioli A., Mariani M., Scambia G., Orteca N., **_Di Dato A_**., Novellino E., Malhotra S.V.,* Ferlini C.,* Fattorusso C.* 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. *J. Med. Chem.*, **2014**, *57*, 7916-7932.

2. Persico M., Petrella L., Orteca N., **_Di Dato A_**., Mariani M., Andreoli M., De Donato M., Scambia G., Novellino E., Ferlini C.,* Fattorusso C.* GTP is an allosteric modulator of the interaction between the guanylate-binding protein 1 and the prosurvival kinase PIM1. *Eur. J. Med. Chem.,* **2015**, *91*, 132-144.

3. Sonawane D.P., Persico M., Corbett Y., Chianese G., **_Di Dato A_**., Fattorusso C.,* Taglialatela-Scafati O., Taramelli D., Trombini C., Dhavale D.D., Quintavalla A., Lombardo M.* New antimalarial 3-methoxy-1,2-dioxanes: optimisation of cellular pharmacokinetics and pharmacodynamics properties by incorporation of amino and N-heterocyclic moieties at C4. *RSC Advances*, **2015**, *5*, 72995-73010.

4. Persico M., **_Di Dato A_**_,_ Orteca N, Fattorusso C.,* Novellino E., Andreoli M., Ferlini C. From protein communication to drug discovery. *Curr. Top. Med. Chem.,* **2015**, *15*, 2019-2031.

5. Santoro A. M., Cunsolo A., D'Urso A., Sbardella D., Tundo G. R., Ciaccio C., Coletta M.,* Diana D., Fattorusso R.,* Persico M., **_Di Dato A_**., Fattorusso C.,* Milardi D.,* Purrello R.* Cationic porphyrins are tunable gatekeepers of the 20S proteasome. *Chem. Sci.,* **2016**, *7*, 1286-1297.

6. Persico M., **_Di Dato A_**., Orteca N., Cimino P., Novellino E., Fattorusso C.* Use of integrated computational approaches in the search for new therapeutic agents. *Mol. Inform.*, **2016**, *35*, 309-325.

7. Fattorusso C., Persico M., Rondinelli F., Orteca N., **_Di Dato A_**_._ Computer-Aided Drug Discovery from Marine Compounds: Identification of the Three-Dimensional Structural Features Responsible for Antimalarial Activity. *Prog. Mol. Subcell. Biol.*, **2017**, *55*, 105-158.

8. **_Di Dato A_**., Cunsolo A., Persico M., Santoro A. M., D'Urso A., Milardi D., Purrello R.,* Stefanelli M., Paolesse R.., Tundo G. R., Sbardella D., Fattorusso C*., Coletta M.* Electrostatic map of proteasome α-rings encodes the design of allosteric porphyrin-based inhibitors able to affect 20s conformation by cooperative binding. *Sci. Rep.*, **2017**, *7*, 17098 (doi:10.1038/s41598-017-17008-7).

ACKNOWLEDGEMENTS

After a demanding period of three years, today is the day: writing this note of thanks is the finishing touch on my dissertation. It has been a period of intense learning for me, not only in the scientific field, but also on a personal level. I would like to acknowledge the people who have supported and helped me so much through the past three years of my Ph.D. life.

Thus, I would like to begin with my supervisor Prof. Caterina Fattorusso, who is not only a superb mentor but also my respectable scientific "mother". I have been greatly impressed by her enthusiasm and productivity in research work as well as her devoted pursuit of scientific perfection. This challenging work could not be completed without her dedication, time and energy. The joy and keenness she puts into her own research was contagious and boosted my motivation. I am also thankful for the excellent example she has provided as a successful person and professor.

I am also thankful to Prof. Marco Persico for the very helpful discussions and for sharing challenges and enjoyable situations in both work and life. He has been a source of friendship as well as good advice and collaboration. I found a lot of qualities in him that one should try to imbibe to become a good scientist as well as a better person.

My extensive gratitude to all of the people who warmly contributed with their stories, anecdotes and backgrounds. In particular, I would like to thank all my Ph.D. colleagues: Maria, Elisa, Simona, Alessia, Carlo, Alessandro, Eugenio and Maria Luisa, with whom I have shared moments of deep anxiety but also of big excitement.

My deepest acknowledgment to Prof. Modesto Orozco and all the members of the Orozco Lab for their immense contribution to my personal and professional growth in my stage at the Institute for Research in Biomedicine (IRB Barcelona). The group has been a source of friendships as well as excellent advice.

I would like to thank all my family, especially my mom and my twin for their unconditional support and empathy. I would be nothing without your love and support. My family is not only the pillar of my life, but also the foundation of who I am.

Last but not least, I would like to thank "mi amorcito" for your encouragement and support when I needed it the most. You came into my life at the end of this adventure, but you fully embraced it. Your love, concern, backup, continuous encouragement, and enthusiasm with every new finding and goal achieved really help me overcome most of the difficulties throughout my work. With you my life has resumed color.

ABSTRACT

Computer-aided drug discovery (CADD) represents a very useful tool to search for potential drug candidates and plays a strategic role in the discovery of new potential therapeutic agents for both pharmaceutical companies and academic research groups. Nevertheless, the modelling of biological systems still represents a challenge for computational chemists, and, at present, a single computational method able to face such challenge is not available. Computational tools are therefore evolving in the direction of combining molecular-mechanic (MM), molecular dynamics (MD), and quantum-mechanical (QM) approaches in order to achieve an overall better simulation of the actual molecular behaviour. In addition, many sampling methods have been developed and applied for the characterisation and comparison of the collective motions of protein structures related to the dynamics of proteins, protein folding and ligand-protein docking simulations. This prompted us, as computational medicinal chemists, to develop various CADD approaches, depending on the specific case under study, integrating theoretical and experimental data. In particular, the research activity carried out during the three years of my PhD led to: i) the development of three-dimensional (3-D) pharmacophore models for the analysis of 3-D structure-activity relationships (SARs) of bioactive compounds, ii) the identification of new molecular targets, iii) the simulation of large-scale protein conformational changes, iv) the simulation of protein/protein and ligand/protein interactions, and v) the design of new bioactive compounds. Computational studies were always performed in the frame of multi-disciplinary projects guided by a unique research strategy, which involved several international and national research groups, and were carried out by integrating and validating our computational studies with the experimental data coming from the other researchers involved in the various projects.

The results obtained enabled to: i) identify a new class of anticancer agents against paclitaxel resistant cancer cells, ii) provide important information on the mechanism of action of cationic porphyrins, a novel class of proteasome conformational regulators with great potentiality as "lead" pharmacophores, and iii) optimise the cellular pharmacokinetic and pharmacodynamic properties of a new series of antimalarial agents.

In addition, I spent a training period abroad of eight-months at the Institute of Research in Biomedicine (IRB) in Barcelona, under the supervision of prof. Modesto Orozco, during which I have had the opportunity to extend my computational background by learning and, then, performing metadynamic and MD simulations, investigating the open/close conformational transition of 20S human proteasome by molecular dynamics simulations.

List of Contents

In accordance with the standard scientific protocol, I will use the personal pronoun "we" to indicate my scientific collaborators and myself.

Chapter 1

Background

1.1 Brief Overview of Drug Discovery

"A lucky accident dropped the medicine into our hands"; this is a sentence of how a publication of the year 1886 from Arnold Cahn and Paul Hepp in the Centralblatt für Klinische Medizin started off. The history of drug discovery is studded by serendipity events, in which extensive knowledge of biological systems were lacking and a lot of working hypotheses were inaccurate in many cases. Historically, drugs were discovered through an experimental screening of compound libraries of molecule, generally natural compounds, which bind the natural target and modulate their activity. Subsequently, the design of new drugs was based on lead optimisation through repetitive chemical modifications to enhance both their potency and pharmacological proprieties. Today the case of lucky accidents have been left on the back-burner, with the aim to identify compounds with pharmacological interest, that can be used in the treatment of diseases, by using a rational approach.[1] Since drugs mediate their effect in the human body by particular interactions of the drug molecule with its target(s), then, to design new drugs, the comprehension of the etiology of the disease and the structure of the molecular target(s) where the ligand (drug) will bind, represent crucial issues. Actually, the first stage of the drug discovery pipeline is, indeed, represented by the identification/selection of the molecular target(s) for the pathology of interest (Figure 1).
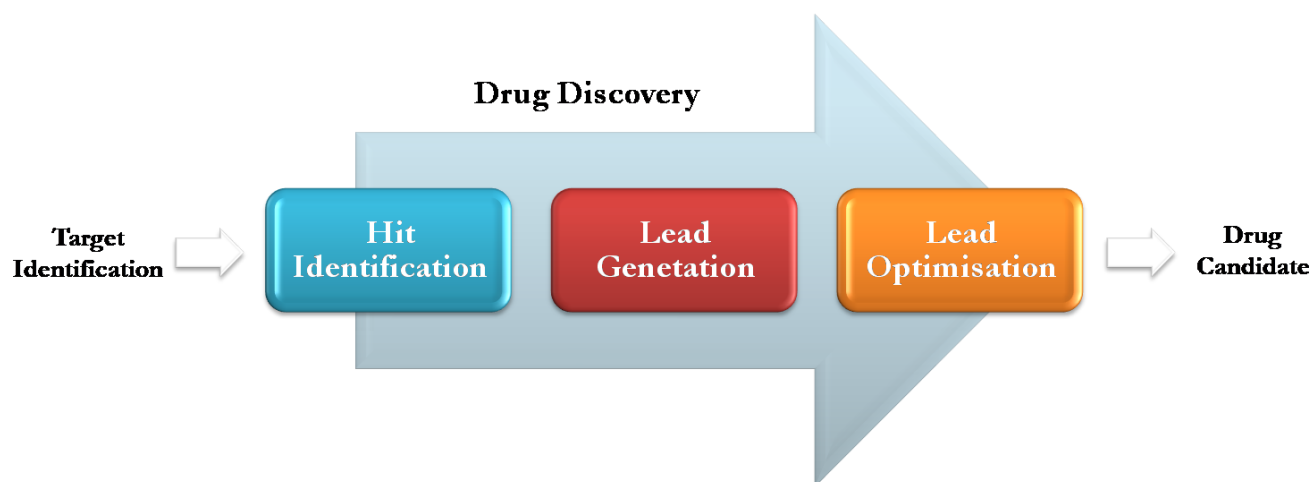
Figure 1. General description of Drug Discovery pipeline.

Drug targets are frequently badly defined in the literature, both for pharmaceutical product placed on the market and for potential therapeutic agents in discovery and development. In this regard, a revised complete map of molecular targets of approved drugs was published, in which were considered a total of 893 human and pathogen-derived biomolecules by means of which 1,578 Food and Drug Administration (FDA) approved drugs effect.[2] Among the 893 biomolecules targeted by the drugs, there are 667 human genome-derived proteins and 189 pathogen-genome derived proteins (Table 1 and Figure 2).

Table 1 | Molecular targets of FDA-approved drugs.

Drug target class	Targets			Drugs		
	Total targets	Small molecule drug targets	Biological drug targets	Total drugs	Small molecules	Biologics
Human protein	667	549	146	1,194	999	195
Pathogen protein	189	184	7	220	215	5
Other human biomolecules	28	9	22	98	63	35
Other pathogen biomolecules	9	7	4	79	71	8

Table adapted from Santos R. et al. (**2017**) (http://dx.doi.org/10.1038/nrd.2016.230).

Analysis of these drug targets highlighted the continued predominance of protein families towards the various human disease areas (Table 1 and Figure 2).
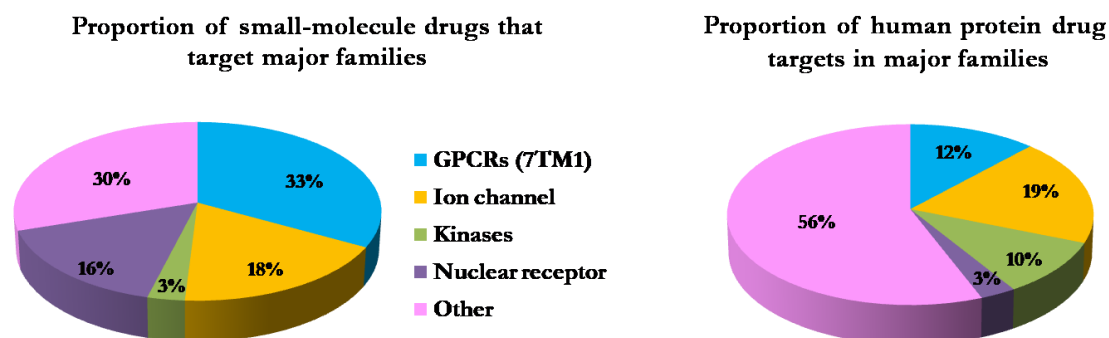
Figure 2. Right: distribution of human drug targets by gene family. Left: distribution by the fraction of drugs targeting those families. Figure adapted from Santos R. et al. (**2017**) (http://dx.doi.org/10.1038/nrd.2016.230).

Once the biomolecular target has been identified/selected, drug discovery activities typically proceed with 1) the *hit identification* phase, in which chemical compounds with a good likelihood of activity against the target are identified, 2) the *lead generation* phase, in which hit compounds are tailored in potency toward the

target, and finally 3) the *lead optimisation* phase, in which lead compounds are optimised, by generating druggable molecules (*i.e.*, capable to employ their pharmacological features in patients)[3] (Figure 1).

The first step aims at the identification of a molecular scaffold that will be recognised by the selected biological target (the protein) and design a model to reproduce the pathology under study. The goal of the second step, lead generation, is to select a lead compound, that is, a molecule that will manifest drug-like features in terms of target potency and selectivity, generally obtained through several rounds of hit modifications and structure-activity relationship (SAR) analysis. The third step is the optimisation of lead compound chemical-physical properties in such a way that it may move to the drug development phase, SARs are rationally explored in the lead optimisation step, too. Progressively, the drug goes beyond the preclinical stage of animal pharmacology and toxicology analysis, as well as formulation, studies of stability, and quality control measures, the whole process (drug discovery and preclinical trials) may take from three to six years. Finally, if the drug is promoted in all the steps previously described, the drug pass to clinical trials in humans, and, if approved, goes to the market after an additional period of about ten years.

1.2 Computer Aided-Drug Discovery (CADD)

To bring a drug to the market is a long run process that costs billions of dollars, it is a lengthy, costly, and difficult process with a high attrition rate of new therapeutic discovery.[4,5] Indeed, the probability of a failure in the drug discovery is high, about 90% of the new drugs entering clinical trials fail to get the FDA-approval and reach the consumer market. In this scenario, computer-aided drug discovery (CADD) gained close attention both in the pharmaceutical industry and at the academic level. CADD can operate as a virtual shortcut, contributing in the expedition of the discovery process and potentially decreasing the cost of research and development. Nowadays, the massive increase in computational power in combination with efficient software, is leading to more targeted approaches for the development of novel drugs[6] (Figure 3).
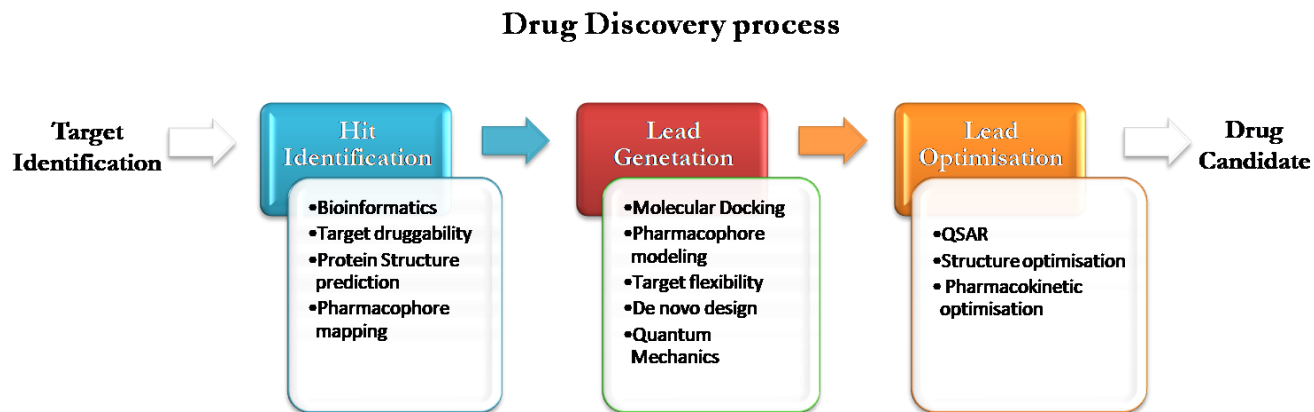
Drug Discovery process

Figure 3. Drug discovery pipeline composed of three phases: "*hit identification*", "*lead generation*" and "*lead optimisation*". Multiple computational approaches can be used to each phase of drug discovery and development pipeline.

Accordingly, in the last few decades, the impact of computational chemistry on drug discovery has been amplified and several theoretical methods that were once not available for efficient drug discovery are widely used for the hit identification and lead generation steps. Currently, CADD is broadly employed in the biomedical area, as well as, to accelerate and simplify hit identification, *hit-to-lead* selection, optimise the pharmacokinetic and toxicity profiles, and escape safety problems. Overall, computational methods are in favour of the identification of chemical scaffolds that may interfere in an efficient way with the target functions, generating a positive pharmacological response. Biochemical and pharmacological data on the new compounds can be used to compute predictions in order to lay the groundwork for more suitable models that can be built, driving to the design of refined compounds.[7]

1.2.1 Classification of CADD approaches

CADD methods can be broadly classified into two groups, such as, Target-Based Drug Discovery (TBDD) and Ligand-Based Drug Discovery (LBDD). As a general rule, the target-based approaches are employed when structure about the molecular target and its experimental data are available.[8] These approaches include structural and bioinformatics analysis, docking calculations, and molecular dynamic simulations.[9] The ligand-based approaches,[10] on the contrary, calculate the physical-chemical properties of bioactive molecules which are characterised by a common mechanism of action, and can be used also when the target structure is unknown. This kind of approach often drive to the generation a pharmacophore model specifying the structural characteristics held accountable for the observed biological activity. At any rate, both approaches have the purpose to rationalise available SARs and to design novel compounds.

1.2.1.1 Target-Based Drug Design (TBDD)

The TBDD is based on the capability to define and study 3-D biomolecular structures. The basic idea of this approach is that the likelihood of a molecule to interact with a given target, showing a desired biologic effect, is subordinate to its tendency to interact with a specific binding site on protein itself.[8] To be able to use TBDD tools is necessary to known the target structures. The information about the target are generally acquired experimentally by Nuclear Magnetic Resonance (NMR), Electron Microscopy or X-ray diffraction methods. X-ray crystallography is still considered today the most important tool of structural biology, supplying a lot of high-resolution protein structures, in the states of unbound/bound receptor and ligand (Brookhaven Protein Data Bank, PDB, www.rcsb.org).[11] However, while the analysis of a singular X-ray structure in many cases provides considerable information about the active site(s) and function of protein, it can also induce to consider the proteins as inflexible entities. Differently, protein flexibility is recognised as critical for protein-ligand or protein-protein interactions and protein conformational changes are essential in order to carry out its biological functions.[12] In spite of their limitations, X-ray structures can even so supply important information regarding intrinsic dynamic of proteins,[12] for instance, making a comparison of several structures resolved under distinct conditions. Furthermore, it is possible to achieve additional understanding of the protein-ligand interaction dynamics by combining the structural and bioinformatics analysis of the resolved X-ray structures with NMR studies.[13]

The huge gap between the number of available protein sequences and experimentally solved protein structures could be filled by the aid of computational methods. Theoretical structure prediction can be divided into two extreme camps: ab-initio method[9,10,14] and homology modelling.[15,16] The first approach has fold prediction from physical chemistry principles as one of its goals. The second method predicts the 3-D structure of a given protein sequence based primarily on its sequence similarity to one or more proteins of known structures.[17] Regarding the first approach, its main limitation is that we do not know precisely the physical procedure by which a protein chain obtains its native 3-D structure. Hydrophobic and electrostatic forces play critical roles, as well as in living systems protein-protein interactions (PPIs). But exactly how these proprieties create the final shape of a protein is still a riddle. The biologist Cyrus Levinthal, in the late 1960s, mentioned that one might assume that an unfolded amino acid chain to arrange a stable 3-D conformation by identifying the minimum energy situation. However, considering rotational atomic bonds and all the degrees of freedom for an amino acid chain, it should be necessary more time compared to the age of the universe for even a simple protein in order to assume its appropriate native conformation.[18] Conversely, also intricate proteins are structured within seconds or even less, showing undoubtedly that do not find an energy of global minimum (GM) by means of a random search of precise conformation.

Indeed, they always seem to find a fold-up with the same shape enough to be crystallised. In this respect, many scientists have esteemed "Levinthal's Paradox", proposing that proteins do not necessarily yearn for energy of GM but ideally safeguard sufficiently deep energy wells to survive possible biological perturbations, by giving up energy in an efficient manner to several functions. "Proteins are obviously not optimised for stability", probably because they require an unfolded structure in several situations.[19]

Nowadays, homology modelling plays a central role in determining protein structure; its importance has been steadily expanding because of the increasing reliability and accuracy of the method. In fact, a protein sequence with over 30% identity to a known structure can often be predicted with an accuracy equivalent to a low-resolution X-ray structure.[17] The ongoing efforts in solving protein structures by homology modelling are progressing toward a new step in the development of this technique, the aim is to predict the "native state of proteins," and not merely to "predict the structure" (with regard to the α-carbon scaffold) of a target sequence. Indeed, numerous studies have showed the effect of conformational diversity also on tamplate-based modelling approaches,[20] since a given template can have several conformers sampling an ample conformational space.[21]

In any case, with the purpose of simulating ligand-protein interactions, latest mechanistic models have to be put into context.[22-25] Proteins show an ensemble of conformational states/substates in solution and the interaction with other proteins and/or smaller ligands, influences these native states/substates by stabilising only one of them.[22] In fact, in accordance with Le Châtelier's principle, the protein-ligand complex formation moves equilibrium in favour of the regeneration of the interacting conformers. In this context, in recent time developed mechanistic models, such as "*conformational-selection*"[23] and "*induced-fit*"[24] (Figure 4), overcome the old but still popular "*lock-and-key*" model.
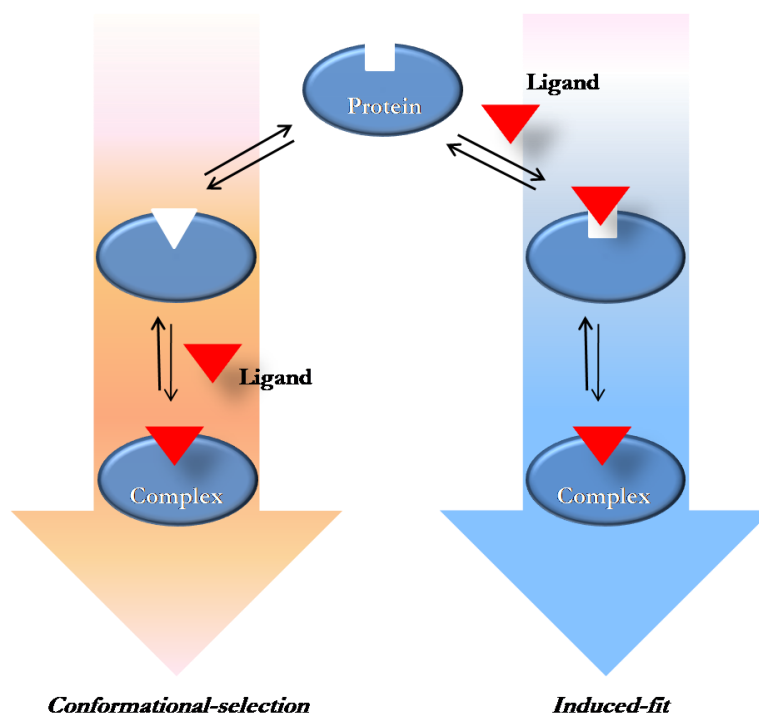
Figure 4. "*Conformational-selection*" and "*induced-fit*" models of binding. The two different models are characterised by the the binding site conformation when the ligand binds. During "*conformational-selection*" the ligand selects the configuration of the binding site whereas during "*induced-fit*" the bound form of the complex is shaped after binding of the ligand.

The "*conformational-selection*" model relies on the fact that, as stated above, the selected protein conformation pre-exists in solution, so that the ligand selectively recognises it, causing a shift of the equilibrium toward that conformation. On the other hand, once a ligand binds to the target protein in the bioactive conformation and interacts with key amino acids, the protein conformation may be modified or become locked, relying on the mechanism of action of ligand-protein complex.[26] According to the "*induced-fit*" model, a protein undergoes a conformational change upon ligand binding and this conformational rearrangement results in a protein-ligand complex with tighter binding. Taking into account both models, the overall scenario, includes a first molecular recognition step (*conformational-selection*) followed by a subsequent reorganisation of the formed complex to reach its closest energy minimum (*induced-fit*).[25]

For the above reasons, to understand the molecular mechanisms underlying conformational changes of proteins is of primary importance for the design of new drugs, and, consequentially, in ligand-protein interaction simulations is essential to take into account protein flexibility. Many computational methods have been developed and applied for the characterisation and comparison of the collective motions of protein structures related to the dynamics of proteins,[27-30] protein folding[31-34] and ligand-protein docking simulations.[35,36] Proteins are complex biomolecules and they usually have energy landscapes with many local energy minima separated by high-energy barriers, and this is a restricting factor for whole sampling as

simulations can come to a standstill in local minima. The collective conformational transitions through which proteins move, globally take place in microseconds to seconds timescales while molecular dynamic is circumscribed to discrete time-steps, no more than a few femtoseconds. Therefore, in spite of multi-thread parallelisation computing means, a classic Molecular Dynamic (MD) simulation is impracticable to sample such conformational transitions and a more suitable techniques are required. Consequently, a variety of innovative and complex techniques have been developed and applied in order to overcome the free energy sampling problem, such as, parallel tempering (using Monte Carlo method to define probabilistic exchanging systems states),[37] metadynamics,[38,39] simulated annealing (SA),[40] discrete molecular dynamics (dMD),[41,42] and a multitude of new sampling algorithms are constantly flourishing.[43,44] The more suited method shall be selected on the basis of physical and biological system properties under study. As an example, while classic MD produces a single trajectory at a specific temperature, allowing kinetic characteristics to be straight obtained, SA (as well as a lot of improved sampling methods) does not, nevertheless, it is suitable for very flexible systems and is marked by an enhanced sampling efficiency.[45,46] With a view to improving the quality of the dynamic simulations, hybrid approaches have also been developed[47-51] and introduced in many molecular docking software until today, enlarging the degree of protein flexibility esteemed throughout the conformational sampling. Of course, once again, no individual method attained optimum results, according to type of the system under consideration and the application of research.[52]

Finally yet importantly, to generate dependable model, it is essential to set up the MD simulation in the framework of a broader study of both the ligand and the protein target, just as the investigation of the molecular mechanism of action. In fact, the calculation protocol has to be tailored to the system, taking into account whether the ligand binding site is known or not, and, in the former case, if it arises in the allosteric or in an orthosteric site, what sort of biological response is brought about by the formation of each complex, etc. Equally important is to take into consideration the target as a functional biological entity, to evaluate if it show an activity as monomer or dimer or part of a multi-protein complex, as well as, to recognise both of the nature and the number of Protein Interaction Domains (PIDs) and Short Linear functional Motifs (SLiMs) (http://elm.eu.org/).[53,54] Indeed, a lot of ligand-protein interactions take place with conserved hot spot residues who belong to PIDs and/or SLiMs.[55-57] Therefore, such investigation is especially relevant when the ligand binding site is not known and/or the effect is explained through an allosteric mechanism.

1.2.1.2 Ligand-Based Drug Discovery (LBDD)

The LBDD approach entails the examinations of biologically active compounds known to be charachterised by the same mechanism of action. It is regarded as an indirect approach to the drug discovery since it does not required knowledge of the target structure at issue. There are two distinct approaches of LBDD:

- the selection of compounds based on chemical similarity to known active compounds,
- the build of a Structure-Activity Relationship (SAR) 3-D model that foretells biological activity by means of the chemical structure.

LBDD is based on the similar property principle which asserts that molecules with comparable structures have a clear likelihood to have comparable properties.[58]

Molecular characteristics can consist of properties such as molecular weight, geometry, shape, surface areas, ring content, rotatable bonds, inter-atomic and bond distances, atom types, planar systems, electronegativities, polarisabilities, symmetry, atom distribution, topological charge indices, functional group composition, aromaticity index, solvation properties, and many others.[59] These descriptors are engendered through the use of quite some methods, among which Molecular-Mechanics (MM) or Quantum-Mechanics (QM) tools[60,61] and are categorised in accordance with the "dimensionality" of the chemical description from which they are computed.[62] In this way, the most important descriptions for building predictive pharmacophores are easily identified. The same basic rules are efficient for molecular similarity search, where geometrical and physico-chemical likeness among active compounds are employed to recognise new compounds with pharmacological activity of many that can be virtually screened.[63]

Structural and electronic descriptors are employed to build a QSAR, where the collection of active and/or inactive ligands are evidenced by a set of mathematical descriptors capable to depict the physicochemical and structural features of such compounds. A model is therefore generated to identify the relationship among those descriptors and their experimental activity, maximising the predictive power. Lastly, the model is used to foresee activity for a library of test compounds that were coded with the same descriptors.

The International Union of Pure and Applied Chemistry (IUPAC) officially described a pharmacophore as "the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response".[64] In terms of activity, it is the spatial positioning of functional groups that a molecule (or drug) must include in order to obtain a sought-after biological response. Therefore, an efficient pharmacophore will consist of information about functional groups that interact with the target, type of non-covalent interactions and inter-atomic distances. Most common properties employed to describe pharmacophores are H-bond

acceptors/donors, basic/acidic groups, partial charge, aliphatic and aromatic hydrophobic moieties. A 3-D-pharmacophore would reproduce how key amino acids are arranged in the binding site of a target protein.[65] For instance, a key amino acid residue which is acting as H-bond acceptor should be in the proximity of a H-bond donor in the pharmacophore model, promoting the protein-ligand interaction. Pharmacophore features have been employed widely in drug discovery for virtual screening, *de novo* design, lead optimisation[66] and a target binding site pharmacophore model can be employed to sift a compound library to identify putative hits.

Chapter 2

Presentation of the Work

Computer-Aided Drug Discovery (CADD) still represents a challenge for computational chemists, and, at present, a single computational method able to face such challenge is not available. This prompted us, as computational medicinal chemists, to develop various CADD approaches, depending on the specific case under study, integrating the theoretical and experimental data in a unique research strategy.

In the present dissertation, I review the research activity carried out during the three years of my PhD, which included: i) the development of 3-D-pharmacophore models for the analysis of 3-D structure-activity relationships (SARs) of bioactive compounds, ii) the identification of new protein targets, iii) the simulation of large-scale protein conformational changes, iv) the simulation of protein/protein and ligand/protein interactions, and v) the design of new bioactive compounds.

The overall methodological approach used for the Target-Based studies is reported in the flowchart below (Figure 5).
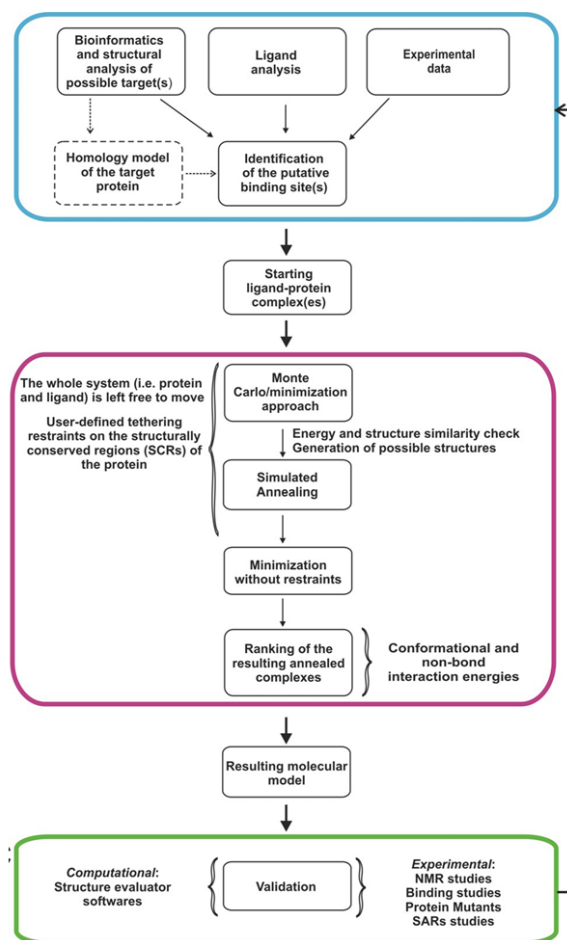
Figure 5. Schematic depiction of the methodological approach used in the molecular modelling studies. Dashed square indicates a discretionary step. Figure reproduced from Persico M. et al. (**2016**) with permission of WILEY-VCH Verlag GmbH & Co. KGaA (http://dx.doi.org/10.1002/minf.201501028).

The first step consists in the investigation of the electronic and conformational details of the ligands under study (which can be either known or putative or newly designed), and it is performed by a combination of simulated annealing/Molecular-mechanics (SA/MM) and Quantum-Mechanics (QM) calculations. Ligand analysis also aims at the identification of the common structural features possibily involved in target recognition. When the considered ligand is represented by a protein structure, it is subjected to the same kind of analysis performed on the protein target(s). The target protein structure(s) is subjected to a detailed structural and bioinformatics analysis, such as: i) inspection of all the available 3-D-structures, including investigation of the molecular interactions with co-crytallysed ligands and postion and conservation of water molecules, ii) sequence alignment (PROMALS3-D),[67] iii) mapping of experimentally identified functional amino acids, iv) secondary structure prediction calculations (http://www.predictprotein.org/), v) identification of structurally conserved regions (SCRs), and vi) identification of Protein Interaction Domains (PIDs) and Short Linear functional Motifs (SLiMs).[53,54]

The bioinformatics analysis becomes crucial when the experimental structure of the target is not complete or not available at all; on this occasion, it has to be verified if is still possible to generate a protein homology model by relying on the experimental structure of closely connected proteins.

The Homology Building technique includes the following main steps:

- identification of the adequate template structure(s),

- sequence alignment between target and template,

- copying coordinates for confidently aligned target and template portions,

- building missing atom coordinates of the target structure,

- evaluation and model refinement.

The information acquired from the analysis of the ligand and target structures (cyan box in Figure 5) will constitute the starting point to set up the subsequent dynamic docking calculations (magenta box in Figure 5). During the docking procedure, the hypothesised starting complex is subjected to disruptive effect in order to generate a user-defined number of possible complexes. The docking procedure is based on a combined Monte Carlo/SA approach[57,68,69] The Monte Carlo method is useful for simulating systems with many coupled degrees of freedom, and, based on statistical mechanics, it generates suitable distinct conformations of a system, which correspond, in this case, to as many ligand-protein complex structures. The obtained Monte Carlo complexes are then subjected to the SA temperature-reducing scheme to improve the ligand-fixing into the binding site.[70] Indeed, SA is an approach that reduces a minimisation of a function of large number of variables to the statistical mechanics of equilibration (annealing).

Throughout the docking calculations, the conformational space of both of the ligand and the protein binding domain area is explored. Importantly, the putative binding domain area is enlarged to the whole protein target, thus allowing to fully investigate ligand binding sites and/or modes, as well as, ligand-induced protein large-scale conformational changes. On the other hand, since these latter are frequently dominated by motions of conserved structural/functional protein domains, then, conformational constraints are applied to backbone hydrogen bonds (α-helices) and φ and ψ torsion angles (β-sheets) of those secondary structures classified as highly conserved by the previous structural and bioinformatics analysis. Water molecules can be constrained with a force constant too, in accordance with their function and conservation. The rest of the system is free to move without restrictions.

The obtained complexes are processed to a final structure optimisation without any conformational

restraint and, then, classified by their conformational and non-bonded interaction energy. The choice of the method to be used for the calculation of non-bonded interactions is selected in accordance to the results of test-docking calculations.

At the end (green box in Figure 5), the quality of the resulting docked structures is calculated by inspecting structural parameters using structure-validation software (*e.g.*, Procheck[71] and Molprobity[72]). The validity of the model is anyway constantly challenged and integrated by experimental data coming from rationally designed experimental studies (*i.e.,* structural, biochemical, pharmacological) (green box in Figure 5).

A somehow divergent, Ligand-Based, computational protocol has been applied for the design of redox-active chemotherapeutics, which couples the complete sampling of the conformational space of the ligands, by using a combination of empirical (MD/MM) and semi-empirical (*i.e.,* PM6 and PM7;[73,74] Eigenvector Following routine[75]) (summarised in Figure 6).
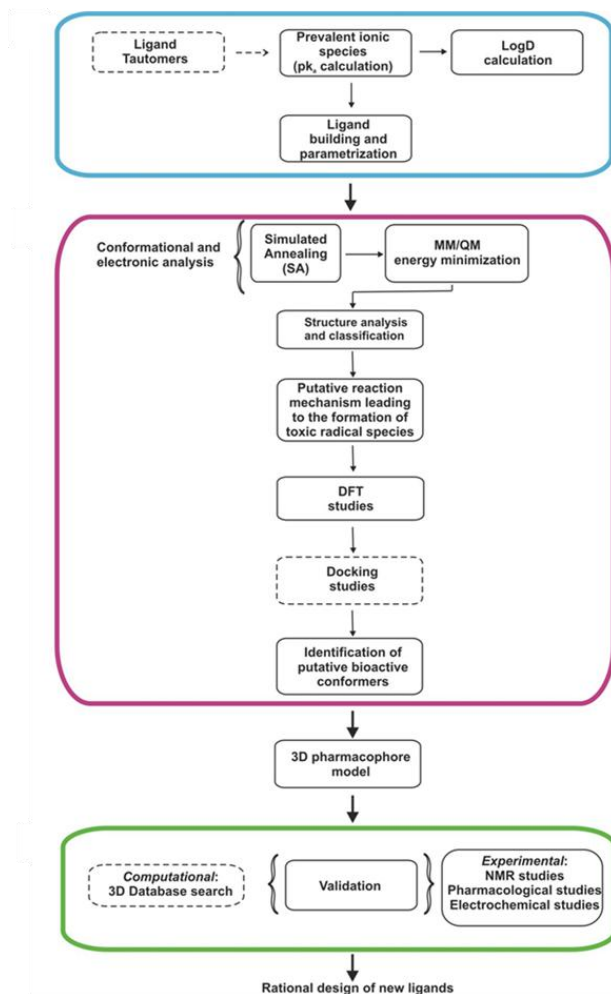

Figure 6. Schematic depiction of the CADD approach for redox-active chemotherapeutics. Dashed square indicates an discretionary step. Figure reproduced from Persico M. et al. (**2016**) with permission of WILEY-VCH Verlag GmbH & Co. KGaA (http://dx.doi.org/10.1002/minf.201501028).

According to the case under consideration, the ligands can be evaluated in the neutral, ionised, and/or radical state. As preliminary step, the predominant ionic state(s) at the evaluated pH(s) of all the doable tautomeric forms of the ligands under study and their log D value(s) are calculated (cyan box in Figure 6). For instance, in the case of the antimalarial endoperoxides described in Chapter 5, all the possible tautomeric forms of the ligands are considered, as well as, their prevalent ionic species at physiological, cytoplasmic, and *Plasmodium falciparum* (*Pf*) food vacuole (FV) pH values (*i.e.*, 7.4, 7.2, and 5.5). Then, the conformational space of the ligands is sampled through numerous cycles of SA calculations followed by MM/QM energy minimisation (magenta box in Figure 6). The number of cycles and the temperature range to be used are chosen by considering the number and the nature of torsional degrees of the ligands. The employ of a very high initial temperature permit to examine the conformational space overstepping any torsional barriers. In fact, at this point, the purpose is to sample the conformational space in a comprehensive way, by recognising all possible minima and obtain the starting structures for the following QM calculations (*i.e.*, semiempirical and DFT). In this view, to increase the variance of the MM starting conformers, more than one set of SA/MM calculations can be carried out, modifying the value of the dielectric constant in accordance with the simulated chemical environments. The MM conformers are then subjected to QM full geometry optimisation. In QM calculations, solvation effects are computed by the Polarizable Continuum Model (PCM)[76-78] in which the solvent is represented by infinite dielectric medium characterised by the relative dielectric constant of the bulk. QM conformers within 5 kcal/mol from the GM are examined and assembled into families taking into account their torsional angle and inter-atomic distance values. Considering the available SAR data, a possible redox reaction mechanism driving to the formation of toxic radical species can be hypothesised. Therefore, starting from the structures of the chosen QM conformers, the distinct species along the supposed reaction pathway of the hypothesised mechanism of action are produced and subjected to additional calculations, this may cosist of docking and DFT studies in complex with the redox partner (in this case free heme). Based upon the accomplished results supposed bioactive conformers are recognised[79] for each ligand, and a pharmacophore model can be generated (magenta box in Figure 6). The computational models require to be constantly integrated with experimental data in an iterative process.

The developed pharmacophore model can be then employed to carry out pharmacophore-based searches in 3-D-molecular databases (green box in Figure 6),[80,81] selecting the resulting hits on the basis of the compatibility of their presented biological activity (PubChem BioAssay database)[82,83] with the activity profile of the search compounds. The information acquired from this analysis can be employed either to additional confirm the 3-D-pharmacophore model and to recognise new molecular scaffolds useful for the rational design of novel ligands.

My research activity has been carried out in the frame of several multidisciplinary projects, each guided by a coordinated computational/experimental strategy. These studies involved several international and national research groups, and were carried out by integrating and validating our computational studies with the experimental data coming from the other researchers involved in the various projects. The projects in which I partecipated, together with the subject of the study and the heads of the other research groups involved, are herein listed:

"Allosteric Inhibitors of the GBP1-PIM1 Complex"
- Dr. Antonio E. Alegria, Department of Chemistry, University of Puerto Rico at Humacao, USA;
- Dr. Cristiano Ferlini, Biomedical Research Center of Danbury Hospital Research Istitute, USA;
- Dr. Sanjay V. Malhotra, Frederick National Laboratory for Cancer Research, USA;
- Dr. Ajay Kumar, School of Environmental Affairs, Universidad Metropolitana, San Juan, Puerto Rico, USA;
- Prof. G. Scambia, Dipartimento di Ginecologia, Università Cattolica del Sacro Cuore di Roma.

"Cationic Porphyrins as Tunable Gatekeepers of the 20S Proteasome"
- Prof. Milardi, Istituto di Biostrutture e Bioimmagini (IBB) - CNR UOS di Catania;
- Prof. A. M. Santoro, Istituto di Biostrutture e Bioimmagini (IBB) - CNR UOS di Catania;
- Prof. R. Purrello, Dipartimento di Scienze Chimiche, Università di Catania;
- Prof. M. Coletta, Dipartimento di Scienze Cliniche e Medicina Traslazionale, Università di Roma Tor Vergata;
- Prof. R. Fattorusso, Dipartimento di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Seconda Università degli Studi Napoli;
- Prof. R. Paolesse, Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma Tor Vergata-Via della Ricerca Scientifica.

"Design of Redox-active Chemotherapeutics"
- Dr. D. D. Dhavale, Garware Research Centre, Department of Chemistry, University of Pune;
- Prof. O. Taglialatela-Scafati, Dipartimento di Farmacia, Università di Napoli "Federico II";
- Prof. D. Taramelli, Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano;

- Prof. C. Trombini, Dipartimento di Chimica "Giacomo Ciamician", Alma Mater Studiorum, Università di Bologna.

In addition, I spent a training period of eight-months at the Institute of Research in Biomedicine (IRB) in Barcelona, under the supervision of prof. Modesto Orozco, during which I had the opportunity to extend my computational background by learning and, then, applying innovative molecular dynamics algorithms to simulate large-scale protein conformational transitions, such as, those occurring in the 20S human proteasome.

Chapter 3

3.1 Allosteric Inhibitors of the GBP1-PIM1 Complex

The management of solid malignancies is in many cases characterised by natural occurrence of drug resistance. Frequently, despite the initial response to therapy, cancer cells evolve a resistant phenotype that leads to a fatal advancement of the illness. Mechanisms that cause drug resistance are subordinated by several functional pathways. Biochemical studies showed that, in the case of paclitaxel resistant ovarian cancer cells, the overexpression of III β-tubulin plays a key role in the evolution of drug resistance to paclitaxel, inasmuch promotes the incorporation of the large-GTPase GBP1 (Guanylate-Binding Protein 1) into microtubules.[84-87] Once in the cytoskeleton, GBP1 binds to prosurvival kinases PIM1 (Proviral Integration of Moloney virus) and initiates a signalling pathway that induces resistance to paclitaxel.[88,89] GBP1 is a large GTP-binding protein, belonging to the Dynamin and Mx family of GTPases,[90] and it is structurally divided into two domains: a compact globular αβ-domain (amino acids 1-278; LG domain) and an elongated purely α-helical domain (amino acids 312-592; helical domain). These two domains are linked by a short intermediate region (amino acids 279-311), which consists of a two-stranded β-sheet and a α-helix.[91-95] Similar other members of the Dynamin family, throughout GTP hydrolysis GBP1 undergoes structural changes driving to self-assembly. Intra-protein communication is fundamentals to the protein allosteric mechanisms, allostery has now been studied on the inside of a single protein domain[96] and its explanation pushed out to involve any shift in protein structure and function at one site consequent from modification at another.

Biochemical studies demonstrated that i) the intermediate region plays a significant role in GTPase activity through allosteric interactions, thus acting as an internal GAP (GTPase Activating Protein),[97] and ii) the GTPase activity is coupled to structural changes in the C-terminal sub-domain α12/α13.[98-100] The interruptions of critical ionic contacts between the LG domain and helical domain has been showed to cause an increased GTPase activity, GMP production, and GBP1 oligomer formation.[98] Moreover, the determination of PIDs[53] and SLiMs,[54] including maintained hot spot residues which have a functional and structural ability of adaptation, can account for the ability of GBP1 to accommodate not only other proteins with complementary binding domains and/or motifs (*i.e.*, protein-protein interactions; PPIs), but also small compounds based on different scaffolds (*i.e.*, protein-ligand interactions).[101] Consequently, many studies showed that ligands are recognised by proteins employing well-defined binding sites, complementary in shape and physicochemical features to the accommodated small ligands.[102] Accordingly,

if two binding sites of distinct proteins share a common motif, it is probable that ligands or ligand portions that bind at one binding site will also be identified in the respective part of the other binding site.[103-105] Following this rationale, starting from the natural lead podophyllotoxin (**1**; Figure 7) and considering semi-synthetic analogues, such as, Etoposide (**2**; Figure 7), our investigations led to the identification of a new podophyllotoxin derivative (NSC756093; **3** in Figure 7) as the first, potent inhibitor of the GBP1-PIM1 interaction, indicating the interface between the catalytic and the helical domain of GBP1 as the binding site, whose occupancy is able to prevent the formation of the GBP1-PIM1 complex through an allosteric mechanism.[106]
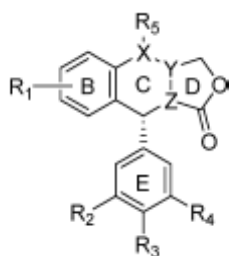

Compound	X	Y-Z	R$_1$	R$_2$	R$_3$	R$_4$	R$_5$
1	CH	CH-CH	6,7-OCH$_2$O	OCH$_3$	OCH$_3$	OCH$_3$	-OH
2	CH	CH-CH	6,7-OCH$_2$O	OCH$_3$	OH	OCH$_3$	
3	N	C=C	6-OCH$_3$	H	H	H	-CH$_2$CH$_2$OH

Figure 7. Structures of Podophyllotoxin (**1**), Etoposide (**2**), and NSC756093 (**3**). Figure adapted from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

In order to identify specific inhibitors of the III β-tubulin/GBP1/PIM1 functional gateway in paclitaxel resistant cells, a series of forty-four 4-aza-podophyllotoxins (4-APTs) derivatives was synthesised by the research group of Dr. M. Sanjay and was screened in the NCI-60 cell panel (Table 2).

Table 2 | Structures of 4-Aza-podophyllotoxins (4-APTs).
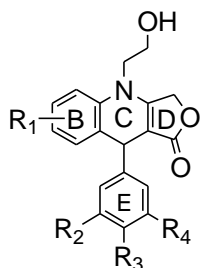
Compound	R_1	R_2	R_3	R_4
NSC750210	6,7-OCH$_2$O	OCH$_3$	OCH$_3$	OCH$_3$
NSC750716	6,7-OCH$_2$O	OCH$_3$	OCH$_3$	H
NSC750717	6,7-OCH$_2$O	OCH$_3$	H	H
NSC750211	6,7-OCH$_2$O	H	H	H
NSC750718	6,7-OCH$_2$O	OCH$_3$	OH	OCH$_3$
NSC750499	6,7-OCH$_2$O	H	OCH$_3$	H
NSC756083	6,7-OCH$_2$O	Cl	H	H
NSC750212	6,7-(CH$_2$)$_3$	OCH$_3$	OCH$_3$	OCH$_3$
NSC751500	6,7-(CH$_2$)$_3$	OCH$_3$	OCH$_3$	H
NSC750719	6,7-(CH$_2$)$_3$	OCH$_3$	H	H
NSC750213	6,7-(CH$_2$)$_3$	H	H	H
NSC751501	6,7-(CH$_2$)$_3$	OCH$_3$	OH	OCH$_3$
NSC751502	6,7-(CH$_2$)$_3$	H	OCH$_3$	H
NSC756084	6,7-(CH$_2$)$_3$	Br	H	H
NSC756085	6,7-(CH$_2$)$_3$	Cl	H	H
NSC756086	6,7-(CH$_2$)$_3$	Cl	Cl	H
NSC750720	6,7-O(CH$_2$)$_2$O	OCH$_3$	OCH$_3$	OCH$_3$
NSC750721	6,7-O(CH$_2$)$_2$O	OCH$_3$	OCH$_3$	H
NSC750722	6,7-O(CH$_2$)$_2$O	OCH$_3$	H	H
NSC750723	6,7-O(CH$_2$)$_2$O	H	H	H
NSC751503	6,7-O(CH$_2$)$_2$O	OCH$_3$	OH	OCH$_3$
NSC751504	6,7-O(CH$_2$)$_2$O	H	OCH$_3$	H
NSC756087	6,7-O(CH$_2$)$_2$O	Br	H	H
NSC756088	6,7-O(CH$_2$)$_2$O	Cl	H	H
NSC759089	6,7-O(CH$_2$)$_2$O	Cl	Cl	H
NSC756090	6-OCH$_3$	OCH$_3$	OCH$_3$	OCH$_3$
NSC756091	6-OCH$_3$	OCH$_3$	OCH$_3$	H
NSC756092	6-OCH$_3$	OCH$_3$	H	H
NSC756093	6-OCH$_3$	H	H	H
NSC756094	6-OCH$_3$	H	OCH$_3$	H
NSC756095	6-OCH$_3$	Br	H	H
NSC756097	6,7-OCH$_3$	OCH$_3$	OCH$_3$	OCH$_3$
NSC756098	6,7-OCH$_3$	OCH$_3$	OCH$_3$	H

NSC756099	6,7-OCH$_3$	OCH$_3$	H	H
NSC759100	6,7-OCH$_3$	H	H	H
NSC756102	6,7-OCH$_3$	H	OCH$_3$	H
NSC756103	6,7-OCH$_3$	Br	H	H
NSC756104	6,7-OCH$_3$	Cl	H	H
NSC756105	6,7-OCH$_3$	Cl	Cl	H
NSC756106	6-CH$_2$CH$_3$	OCH$_3$	OCH$_3$	OCH$_3$
NSC756108	6-CH$_2$CH$_3$	H	H	H
NSC756110	6-CH$_2$CH$_3$	Br	H	H
NSC756111	6-CH$_2$CH$_3$	Cl	H	H
NSC756112	6-CH$_2$CH$_3$	Cl	Cl	H

Table adapted from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Thirty-one compounds exhibited significant growth inhibition and were subjected to the COMPARE Analysis method.[107]

Results evidenced that some of these ligands are active on paclitaxel-resistant cells with a spectrum of cytotoxicity similar to that of other anticancer drugs, all interfering with the dynamicity of the cytoskeleton. Moreover, it was possible to identify five clusters with different structural homologies (Figure 8).
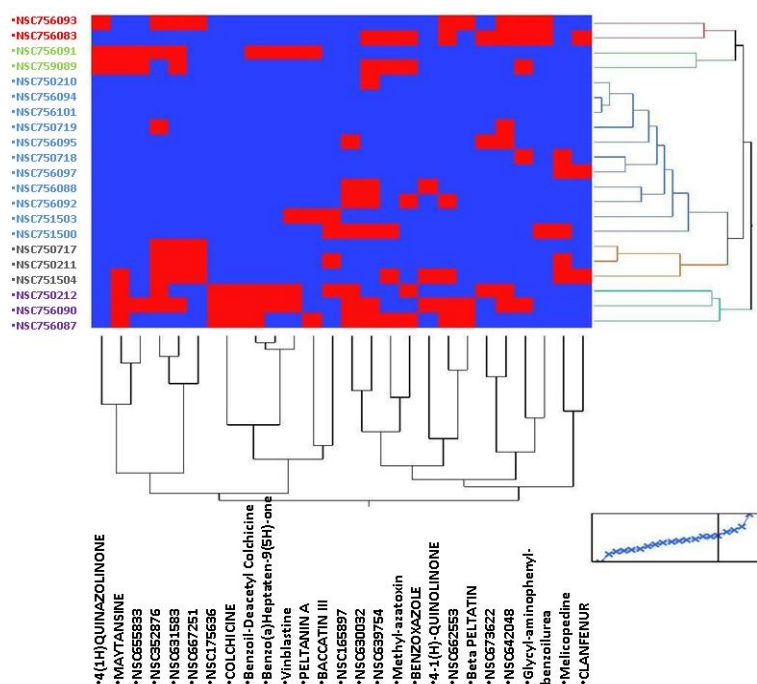

Figure 8. Schematic representation showing the most active twenty-one 4-APTs obtained by the COMPARE analysis. On the vertical axis, with different colours are represented the five different 4-APTs clusters. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Starting from these results, and in order to evaluate the ability of these compounds to interfere with the GBP1-PIM1 complex, the twenty-one 4-APTs found to be more active have been tested in Surface Plasmon Resonance (SPR) assays, carried out by the research group of Dr. C. Ferlini (Danbury Hosptial, USA). Since tubulin isoforms (α and β) are not usable for *in vitro* assays, as the their active conformation requires post-transcriptional modifications not obtainable with current recombinant technology, these studies were performed considering only GBP1 and PIM1, which can be expressed *in vitro*. The SPR results showed that i) the signal generated by the GBP1:PIM1 interaction is dose-dependent in a range of 8.25-140 nM with a KD of 38 ± 14 nM and ii) only the NSC756093 (**3**; Figure 7) compound is capable of inhibiting the GBP1:PIM1 interaction at 65%, with a correlation between concentration and inhibition rates (Figure 9).
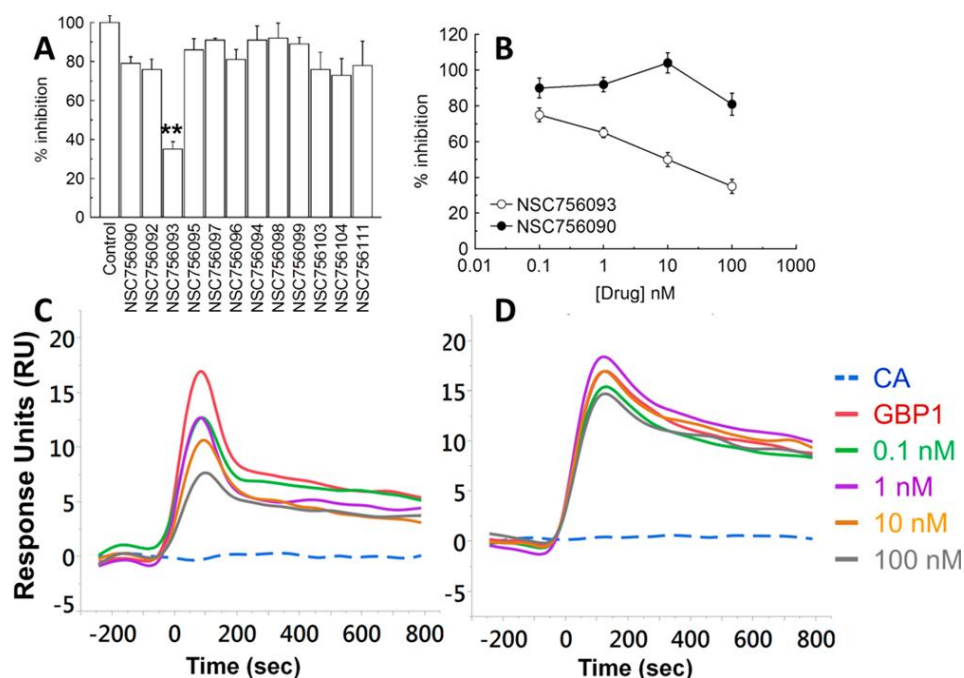

Figure 9. **A** Graphic representing the eleven compounds that were capable of producing an inhibition of the binding around 10-20%, while only **3** was able to inhibit 65% of the GBP1:PIM1 interaction. **B** The inhibition of the GBP1:PIM1 interaction by **3** was dose-dependent and statistically significant as compared with the control without drug (p < 0.001, Anova). Representative biosensograms are shown in **C** (NSC756093) and **D** (NSC756090). In all the experiments, CA was used as negative control and the maximum signal was calculated with GBP1 without inhibitors. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

To further confirm the ability of **3** to inhibit the GBP1:PIM1 interaction, co-immunoprecipitation (co-IP) studies in SKOV3 ovarian carcinoma resistant to paclitaxel, were also performed. The cells were treated for three hours with 100 nM both of the active compound (**3**) and of an inactive compound (NSC756090). The results have shown that treatment with **3** inhibits GBP1:PIM1 interaction even *in vitro*, while treatment

with Dimethylsulfoxide (DMSO) or inactive compound (NSC756090) does not result in any modulation of the GBP1:PIM1 interaction (Figure 10).
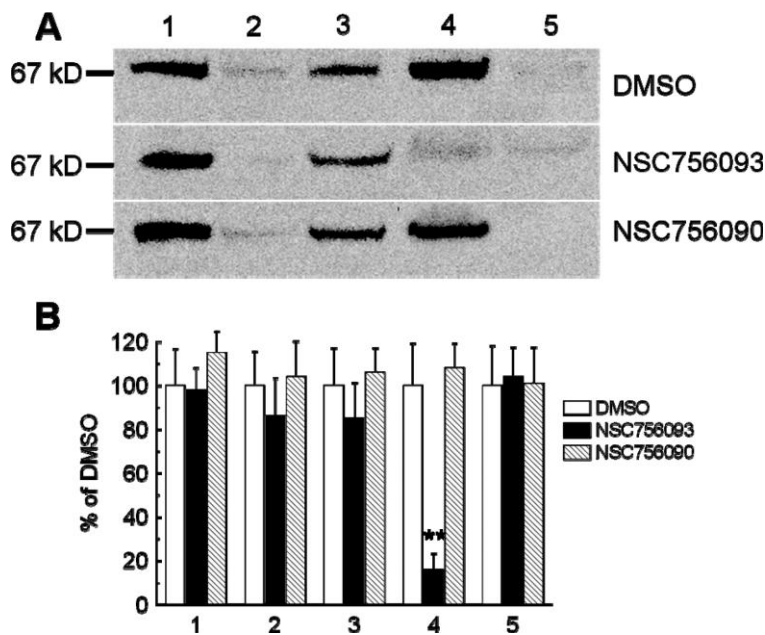

Figure 10. A Representative co-IP of PIM1 and GBP1 in SKOV3 cell line treated with **3** (3 h at 100 nM). The signal was revealed using anti-GBP1 antibody detected as a specific band of 67 kD. *Lane 1*: input of SKOV3 lysate. *Lane 2*: flow through co-IP. *Lane 3*: flow through control. *Lane 4*: co-IP. PIM1:GBP1 (antibody anti-PIM1). *Lane 5*: co-IP. Negative control (antibody anti-IGg). The presence of the signal in lane 4 means no interference in the GBP1:PIM1 binding in DMSO and the inactive NSC756090; the absence of detectable signal in the presence of **3** means that the compound is able of inhibiting the GBP1:PIM1 interaction. **B** Bar chart showing the densitometric analysis of the experiment shown in A, performed in two independent experiments. A significant (double asterisks = $p < 0.001$, Anova) suppression of the co-IP was noticed in both experiments only with the compound **3**. Figure reproduced from from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

In the next paragraph, I desscribe some computational approaches performed in order to investigate the molecular recognition mechanism and identify a putative binding site of new class of potential drugs.

3.2 Identification of NSC756093 Putative Binding Site

Starting from this biological results and in order to identify a potential binding site for **3** on GBP1 and/or PIM1, a detailed structural and bioinformatics analysis was performed. Since compounds **1**, **2**, and **3** share a common structural skeleton, consisting of the polycyclic B-D system and the phenyl ring E (Figure 7), then, not only on the X-ray structures of the two protein targets (*i.e.*, GBP1 and PIM1) were considered in the anlysis, but also those of i) **2** in complex with TopoisomeraseIIβ (TopoIIβ) and DNA (PDB ID: 3QX3) and ii) **1** in complex with Tubulin (PDB ID: 1SA1).

The comparison of the **2**/TopoIIβ/DNA complex with the **1**/Tubulin complex by superimposing the common B-D polycyclic system of the ligands revealed a striking overall similarity (Figure 11) with both **1** and **2** mediating molecular interactions at the interface between two different protein domains, such as:

- monomers of α and β Tubulin for **1**,
- Winged Helix Domain (WHD) and Topoisomerase-Primase domain (TOPRIM) for **2** (in such a case DNA is also finds itself implicated).
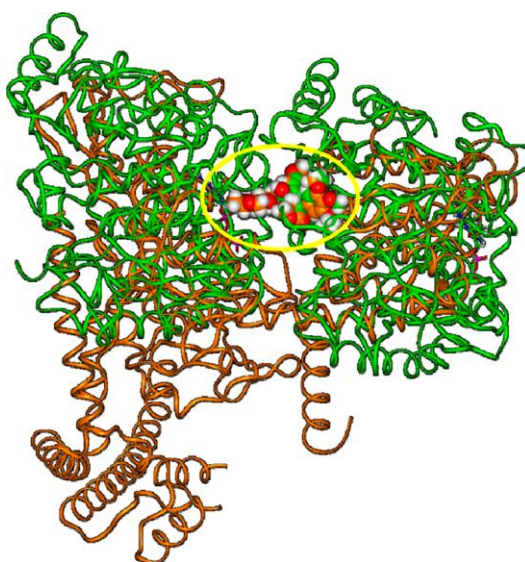

Figure 11. Superimposition of the X-ray structure of the **2**-TopoIIβ complex (orange; PDB ID: 3QX3) on the X-ray structure of **1**-Tubulin complex (green; PDB ID: 1SA1) by fitting the A-D rings of the ligands. The yellow circle highlights the position of the fitted ligands. Proteins are displayed as ribbons, ligands are displayed as CPK. DNA is omitted, for clarity of presentation. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

On the other hand, the results obtained by the Eukaryotic Linear Motifs Resource (ELM) server[53] revealed that the binding sites of the two ligands are characterised by several common consensus sequences contained in functional protein-protein interaction motifs (Figures 12 and 13).
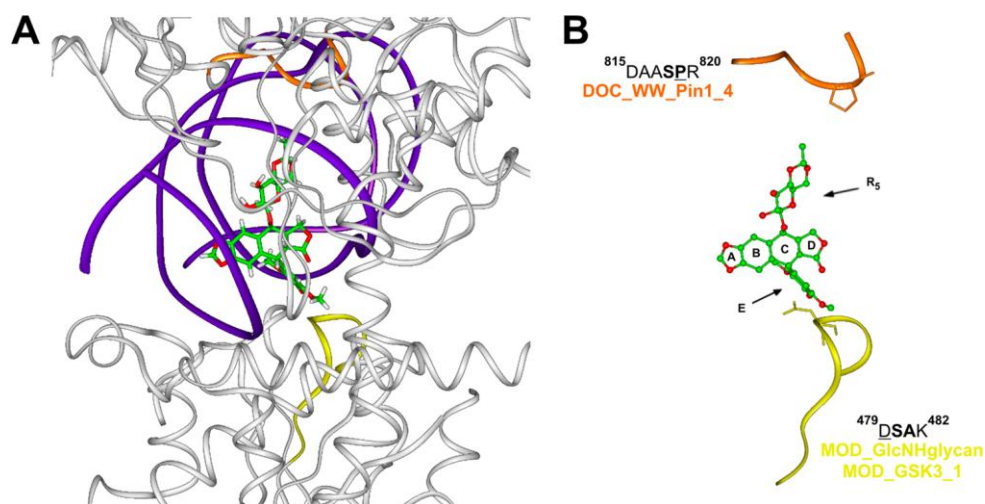
Figure 12. A Overview of the **2** binding site in TopoIIβ-DNA complex (PDB ID: 3QX3). TopoIIβ (white) and DNA (violet) are displayed as ribbons. **2** is displayed in stick and coloured by atom type (C = green, O = red, H = white). **B** Detailed view of protein motifs involved in etoposide binding site. Key interacting residues are displayed in stick and underlined in the sequence; key consensus sequence residues are in bold. **2** is displayed as ball and stick and coloured by atom type. Protein motifs involved in the binding site are displayed as ribbons and coloured: DOC_WW_Pin1_4 motif (orange); MOD_GlcNHglycan and MOD_GSK3_1 motifs (yellow). Hydrogens are omitted for sake of clarity. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).
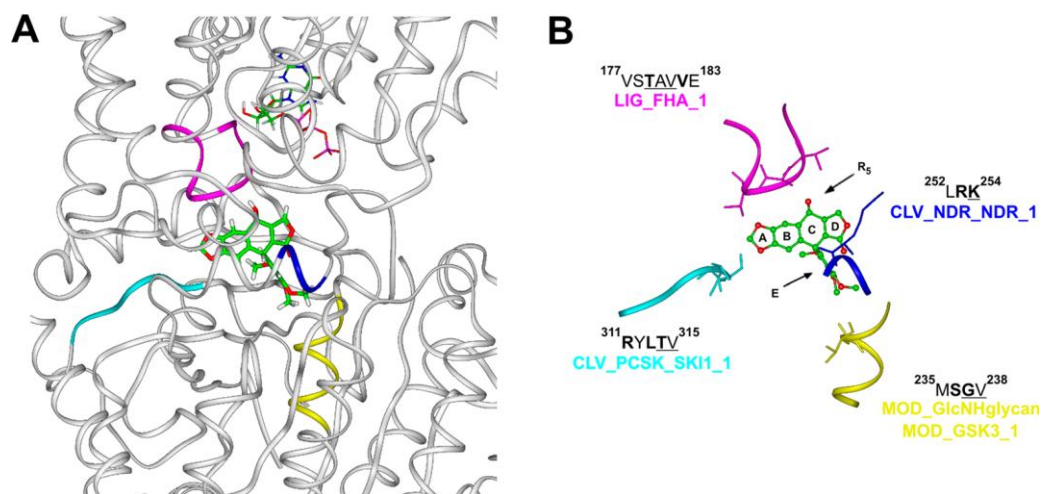

Figure 13. A Overview of **1** binding site on α/β Tubulin heterodimer (PDB ID 1SA1). Tubulin is displayed as ribbons and coloured in white. **1** and α-Tubulin GTP nucleotide are displayed in stick and coloured by atom type (C = green, O = red, N = blue, P = magenta, H = white). **B** Detailed view of protein motifs involved in **1** binding site. Key interacting residues are displayed in stick and underlined in the sequence; key consensus sequence residues are in bold. **1** is displayed as ball and stick and coloured by atom type. Protein motifs involved in this binding site are displayed as ribbons and coloured: LIG_FHA_1 motif (magenta); CLV_NDR_NDR_1 motif (blue); CLV_PCSK_SKI1_1 motif (cyan); MOD_GlcNHglycan and MOD_GSK3_1 motifs (yellow). Hydrogens are omitted for sake of clarity. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

It is noteworthy that similar ligand substructures (*e.g.*, phenyl ring E) were found to bind to similar consensus sequences, while different ligand substructures (*e.g.*, substituent R_5) bind to different consensus

sequences (Figures 12B and 13B).

The same structural and bioinformatics analysis carried out on TopoIIβ and Tubulin was also performed on the X-ray structures of PIM1 and GBP1. In both proteins, numerous consensus sequences of functional protein motifs have been identified, but only in GBP1 a 3-D dimensional combination of protein motifs similar to that found in active sites **1** and **2** was found (Figure 14).
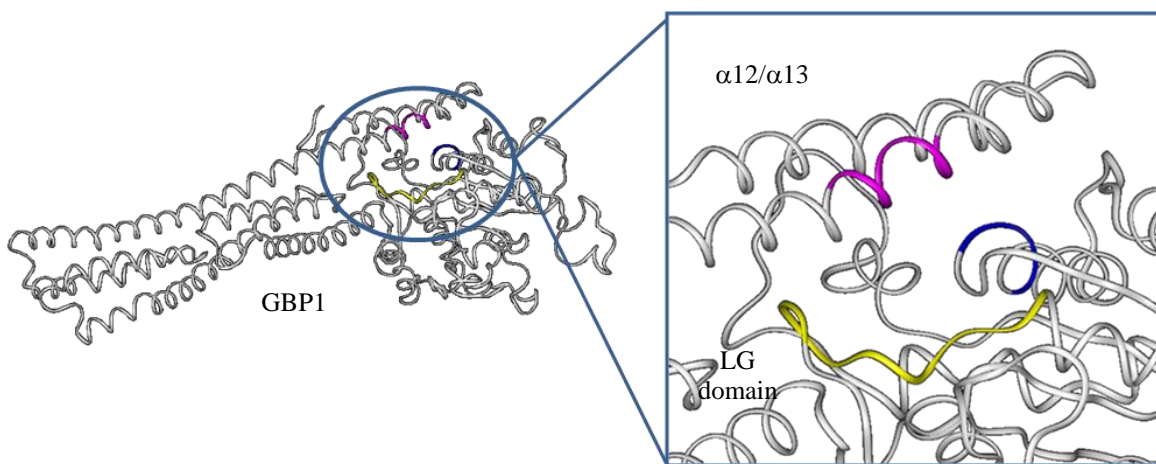

Figure 14. 3-D combination of protein patterns found in the GBP1 X-ray structure (PDB ID: 1DG3).

It is important to point out that these motifs are located at the interface of two GBP1 structural domains, such as, the LG domain and the helical domain. According to these results, this site was hypothesised as the possible interaction site for our aza-podophyllotoxin derivative **3**.

In order to validate the **3**:GBP1 interaction hypothesis, dynamic docking studies have been performed on the putative complex using a Monte Carlo/SA approach. The whole system was left free to move throughout the course of docking calculations, only the backbone of Structurally Conserved Regions (SCRs) was restrained by a force constant to avoid unrealistic results (for details see Chapter 2 and the Experimental Section). Calculations were performed preserving the water molecules present in the reference crystallographic structure of GBP1 (PDB ID: 1DG3). In particular, in order to evaluate the role played by the water molecules in the catalytic site, two sets of docking calcualtions were performed: one where the water molecules in the GBP1 nucleotide binding site were left fully free to move during the simulation, and the other where they are constrained by applying a tether force. Finally, in order to test the thermodynamic stability of the obtained complexes, they were subjected to a simulated annealing (SA) procedure. The resulting complexes were, then, classified according to their conformational energy values and the lowest energy complex was chosen as the most representative one.

The quality of the molecular model generated by docking calculations was evaluated using the Procheck server (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/),[71] and the results obtained showed a

structure quality similar to that of the reference GBP1 crystallographic structure (PDB ID: 1DG3) both for the acid backbone and side chains (Ramachandran Plot and Main-chain parameters; Appendixes I and II respectively).

The results obtained from dynamic docking calculations showed that:

- **3** remains linked to the interface between the LG domain and helical domain even if, after the simulated annealing procedure, the ligand rotates by ~ 180 ° compared to the position adopted in the Monte Carlo docked complex (Figure 15),

- **3** binding site contains three consensus sequences belonging to functional protein motifs also identified on the binding site of **1** and **2** (Figure 16).
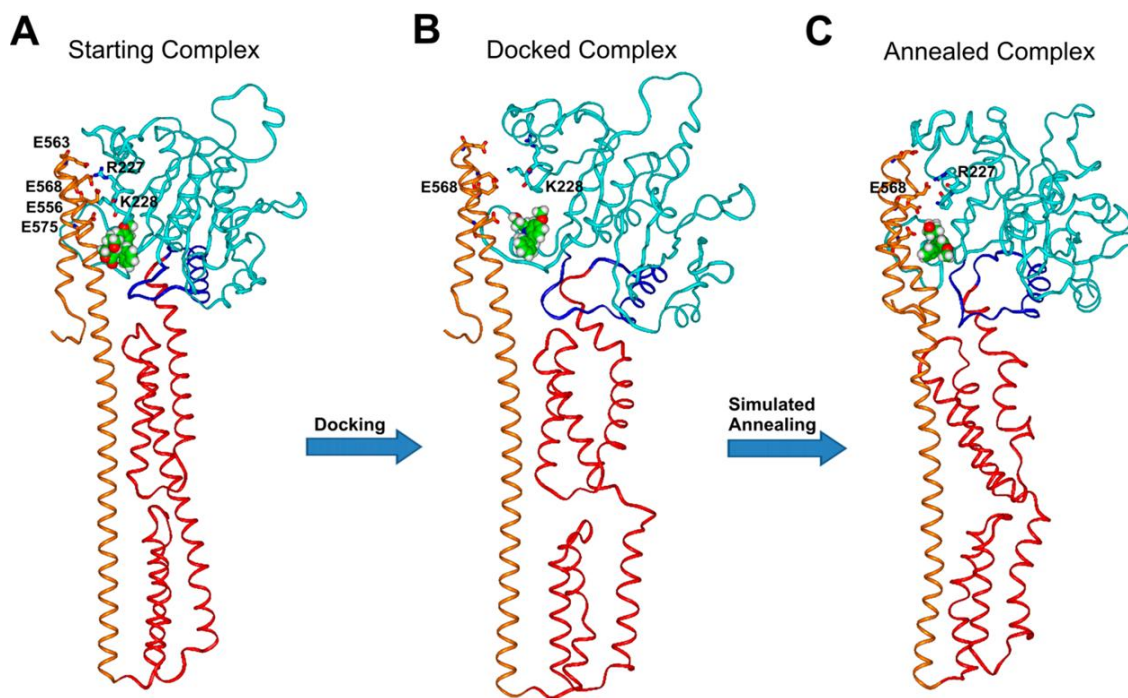
Figure 15. Overall view of **3**-GBP1 starting **A**, docked **B**, and annealed **C** complexes. GBP1 structure is displayed as ribbons, where the LG domain is in cyan, the connecting region in blue, the helical domain in red and α12/α13 in orange. **3** is displayed as CPK and coloured by atom type (C = green, O = red, N = blue, H = white). Key interacting residues between LG domain and α12/α13 are displayed in stick, coloured by atom type and labelled. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

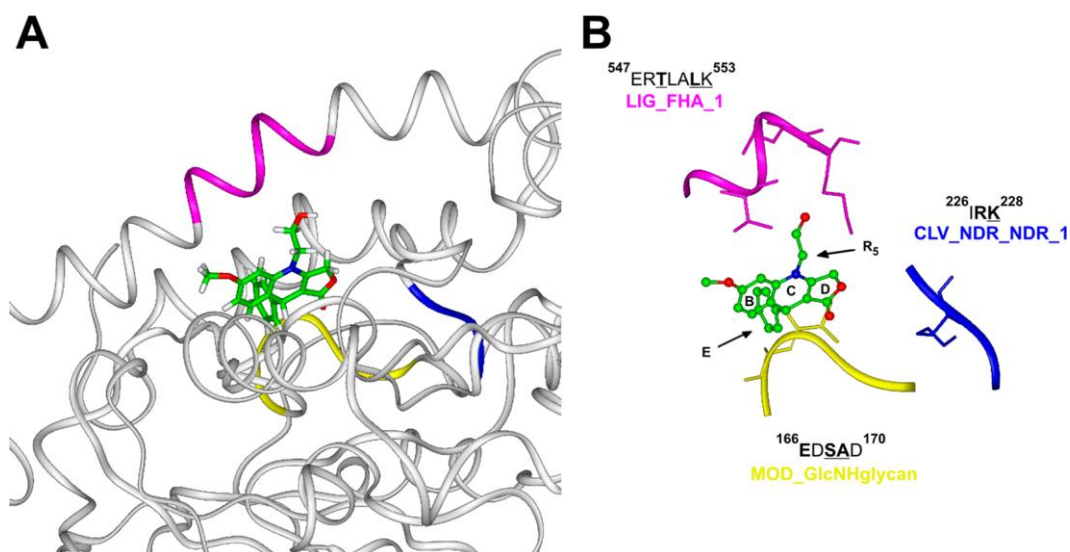**Figure 16. A** Overview of **3** binding site in GBP1 (white) in the final annealed complex. GBP1 is displayed as ribbons. **3** is displayed in stick and coloured by atom type (C = green, O = red, N = blue, H = white). **B** Detailed view of protein motifs involved in **3** binding site. Key interacting residues are displayed in stick and underlined in the sequence; key consensus sequence residues are in bold. **3** is displayed as ball and stick and coloured by atom type. Protein motifs involved in this binding site are displayed as ribbons and coloured: LIG_FHA_1 motif (magenta); CLV_NDR_NDR_1 motif (blue); MOD_GlcNHglycan motif (yellow). Hydrogens are omitted for sake of clarity. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Comparison of the results of docking studies with the results obtained from the bioinformatics and structural analysis performed on the **2**-TopoIIβ and **1**-Tubulin complexes showed other interesting results regarding the interactions occurring between the consensus sequences identified on the three binding sites (GBP1, Tubulin, and TopoIIβ) and the respective ligand substrates considered (Table 3).

Table 3 | Consensus sequences of the functional motifs found in the binding site of **1**, **2**, and **3**.

Ligand Sub-structures	Consensus[a] Sequences	Protein Target
Ring D	^{252}L**RK**^{254}	Tubulin
	^{226}I**RK**^{228}	GBP1
Ring E	^{235}M**SG**V^{238}	Tubulin
	^{479}D**SA**K^{482}	TopoIIβ
	^{166}**E**DS**A**D^{170}	GBP1
R_5	^{177}VS**T**AV**V**E^{183}	Tubulin
	^{547}**ERT**LA**LK**^{553}	GBP1

[a]motif amino acids involved in molecular interaction are highlighted in bold.

In particular, the phenyl ring E of the three ligands interacts with a protein motif having a consensus sequence maintained in the three binding sites and coloured in yellow in Figures 12, 13, and 16. The polycyclic system of **2** establishes π-π interactions and H-bonds with surrounding nucleotides (Figure 6) similarly to what happens to the **1** polycyclic system in Tubulin and the **3** polycyclic system in GBP1 (nucleotides are replaced by aromatic amino acids). On the other hand, the D ring, preserved just in **1** and **3,** in both binding sites, is oriented toward the same consensus sequence, blue-coloured in Figures 12 and 16 (Table 3). In addition, even smaller and fairly similar groups such as R_5 of **1** and **3** (Table 3) are oriented to a consensus sequence belonging to a linear pattern present in both binding sites and coloured in magenta in Figure 12 and 16. Thus, similar ligand substructures correspond to protein motifs containing common consensus sequences. On the contrary, specific ligand substructures such as the substituent R_5 present in **2** (Table 3), represented by a bulky glycoside residue, or as A ring of **1**, substituted with a 6-OCH_3 group in **3**, are oriented toward protein motifs that do not have similar consensus sequences to those found in the other complexes.

All of these results indicate that the molecular skeleton, composed of the B-D rings and the pendant phenyl ring E, is able to interfere at the interface of two protein domains. This interface is characterised by the presence of various consensus sequences contained in functional protein motifs and the different selectivity for a specific ligand differ in according to the substituents introduced on the common skeleton. Conversely, even small changes in the ligand structure can drive the binding to a different combination of protein motifs, that is, to different protein targets. Finally, the same ligand could recognise a similar pattern of protein motifs on different protein targets. In this view, it is to be noted that **2** is able to inhibit the TopoIIβ also by binding to the ATPase domain.[108]

The obtained **3**-GBP1 annealed complex, shown that **3** is almost completely "embedded" in the GBP1 binding pocket with only the D ring and its upper part, including the R_5 substituent, partially exposed to the solvent (Figure 17).
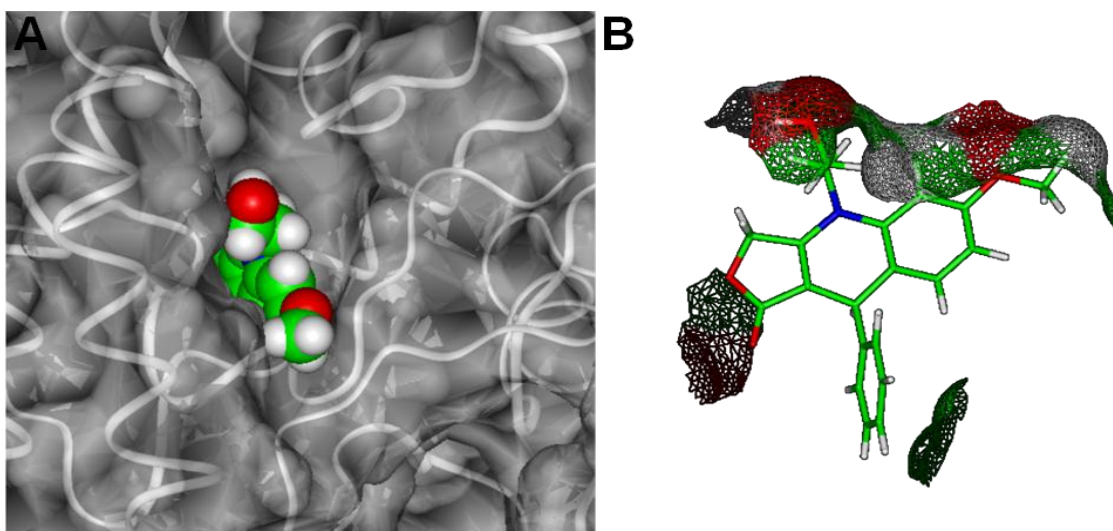
Figure 17. A. **3**-GBP1 annealed complex. The ligand is displayed as CPK and coloured by atom type (C = green, O = red, N = blue, H = white), the solvent accessible surface of the protein is displayed as solid (transparency = 50%) and coloured in white, protein backbone is displayed as white ribbon. **B.** solvent accessible surface of **3** in complex with GBP1. Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Accordingly, **3** is the only compound capable of inhibiting the GBP1:PIM1 interaction, as well as, the less bulky compound of the series. In fact, all the other derivatives have the bulk substituents on B and E rings compared to **3** (Table 2). In order to rationalise the inactivity of all the 4-APTs tested and, at the same time, to validate the **3**:GBP1 interaction model, the NSC756092, NSC756094, NSC756095, NSC759100, and NSC756108 compounds were subjected to a thorough conformational analysis that included dynamic and molecular-mechanic calculations followed by mechanical calculations (for details see the Experimental Section). As shown in Table 2, all these compounds show a minimum extra-volume at the several substituents, differently by **3**. Conformers within 5 kcal/mol from the global energy minimum (GM) were overlapped, through the common structural B-D skeleton, on **3** in complex with GBP1. All possible orientations of their substituents were considered and all possible steric impediments with the protein were analysed. It turned out that the replacement of the E ring in the meta position, with either a methoxy group (NSC756092) or a bromine atom (NSC756095), is not tolerated due to a steric clash with the T549 and L579 side chains, and with the A169 and V172 backbone. The side chain of L579 and the A169 backbone are also involved in steric clashes with the methoxy substituent present in the "*para*" position of E ring in the compound NSC756094. On the other hand, replacement on B ring with a further methoxy group in position 7, such as in NSC759100 compound, causes steric impediments with the S173, T279, K278 and K280 backbones. Finally, replacing the methoxy group on C6 with an ethyl chain, such as in the NSC756108 compound, shown steric clashes with the L278 and E545 side chains (Figure 18).
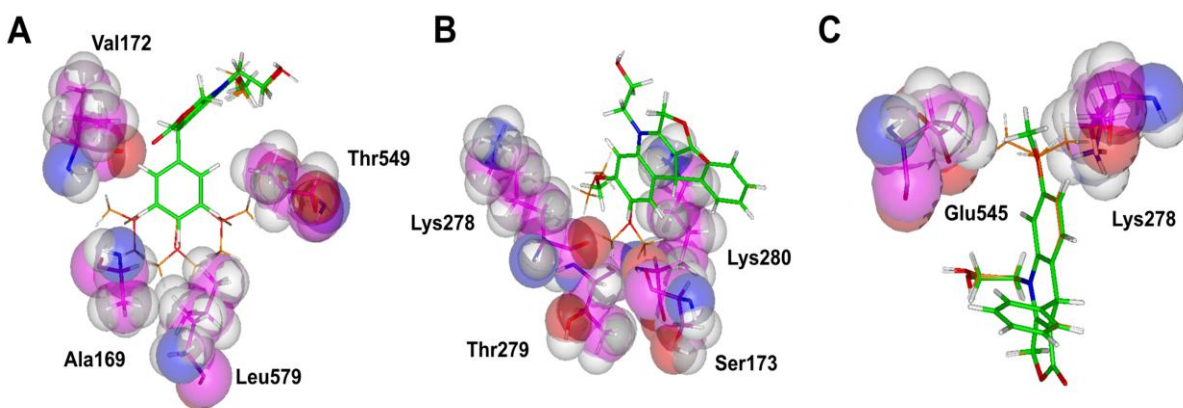
Figure 18. Superimposition of **A** NSC756095 and NSC756094, **B** NSC756100 and **C** NSC756108 PM7 conformers (carbons in orange) on the bioactive conformation of **3** (carbons in green). Bioactive conformation of **3** is displayed in sticks while conformers are displayed as lines. GBP1 amino acids (carbons in magenta) involved in steric clashes are labelled and displayed as CPK (transparency = 50 %). Heteroatoms are coloured by atom type (O = red, N = blue, H = white). Figure reproduced from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Although at first glance, the ethylene chain and the methoxy group may occupy a similar molecular volume, however, they are oriented differently to the plane of the polycyclic system. The methoxyl group, due to the conjugation of oxygen with the aromatic ring, is placed on the same plane as the polycyclic system (sp2 geometry). Conversely, the ethylene chain protrudes outside the ring plane (sp3 geometry). In accordance with the observed steric clashes, the energy minimisation of all possible GBP1 complexes with NSC756092, NSC756094, NSC756095, NSC759100, and NSC756108 conformers did not provide any acceptable solution since the ligand bioactive conformation showed a ΔE from the GM of more than 10 kcal/mol.

According to the results obtained from docking calculations, **3** links to the interface between LG domain and helical domain of GBP1, where the key interactions that regulate the GBP1 function preside. In fact, the conformational equilibrium that exists between LG and the helical domain governs the catalytic activity of GBP1 and its oligomerisation ability.[91,97] In other GTPases (*e.g.,* Ras, Rho) this regulation is carried out by other protein partners (*e.g.,* GAP and GEF).[109] Analysing all conformational changes resulting from GBP1 dynamic docking calculations, a significant movement of the α4' (LG domain) respect to α12/α13 (helical domain) is observed. Thus compound **3**, by influencing the interactions between the LG and the helical domain of GBP1, could induce GBP1 to adopt conformations incompatible with PIM1 binding.

To validate the results obtained from molecular modelling studies, a series of GBP1 mutants (single and double deletions in the human sequence) were designed and expressed and their effect of the formation of the GBP1:PIM1 complex was evaluated. In addition, in order to evaluate the impact of these mutations on GBP1 oligomerisation, kinetic analysis have been performed both in the presence of GDP and the non-hydrolysable analogue of GTP, GppNHp (1 µM in PBST) (Tabella 4).

Table 4 | Kinetic analysis of a panel of GBP1 mutants for the GBP1:PIM1 interaction assessed with SPR technology.

Protein/Mutants	KD (nM) in PBST (1µM GppNHp*)	KD (nM) in PBST (1µM GDP)
GBP1	67±20	30±10
R227E/K228E	>10,000	>10,000
R48A	29±2.1	21±0.1
K51A	50±10	26±11
D184A	80±5	77±3
Q72A	62±2	161±17
D103N	21±1	15±0.1
R227A/R228A	137±14	61±4
1-481 (C-terminal deletion)	52±4	68±8

*Guanosine 5'-[β,γ-imido]triphosphate trisodium salt hydrate, non-hydrolysable GTP analogue.
Table adapted from Andreoli M. et al. (**2014**) with permission of ACS Publications (http://dx.doi.org/10.1021/jm5009902).

Indeed, it is clear from the literature that GBP1 wild-type (wt) is a monomer both in its free form and when binds to GDP; differently, it is a dimer when binds either to GTP and non-hydrolysable GppNHp, while it is a tetramer during GTP and GDP hydrolysis, as well as, when in complex with the transition state mimetics GDP-AlF$_3$ and GMP-AlF$_4$.[91,110] On the other hand, the other mutants reduce the GBP1 ability to oligomerise and therefore affect its catalytic activity being located close to the catalytic site (R48A, K51A, D184A) or switches regions (Q72A and D103A).[92,94,98,110]

The results obtained using the designed mutants evidenced that only the R227E/K228E double mutant is able to determine a complete decoupling of the GBP1-PIM1 complex, both in the presence of GDP and GppNHp, concluding that residues R227 and K228 play a crucial role in:

- the catalytic activity,
- the ability of omo-oligomerisation of GBP1,[98]
- the ability to interact with PIM1.

The R227E/K228E and R227A/K228A double mutants is located at the putative **3** binding site on GBP1, at the interface between the LG and the helical domain (Figure 19).
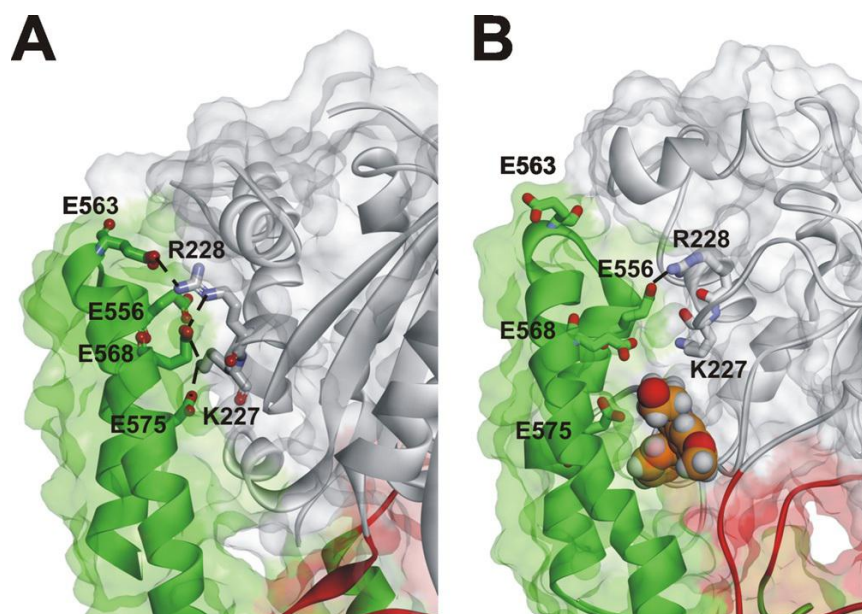
Figure 19. Overview of GBP1 (**A**; PDB ID: 1DG3) and **3**-GBP1 annealed complex **B**. GBP1 structure is displayed as solid ribbons, with the LG domain in grey, the connecting region in red, and α12/α13 of helical domain in green. **3** is displayed as CPK and coloured by atom type (C = orange, O = red, N = blue, H = white). The solvent accessible surface of both GBP1 and **3**-GBP1 annealed complex is displayed as solid (transparency=90%). Key interacting residues between LG domain and α12/α13 are displayed in stick, coloured by atom type and labelled. Figure reproduced from Persico M. et al. (**2016**) with permission of Wiley-VCH Verlag GmbH & Co. KGaA (https://doi.org/10.1002/minf.201501028).

The positive charges present on the side chains of R227 and K228 (α4' helix; LG domain) are coupled to four negative charges present on the side chains of residues of glutamate E556, E557, E563 and E568 (α12/α13 helices; helical domain). Consequently, in R227E/K228E mutant, the substitution of R227 and K228 with two negatively charged glutamate residues induce a moving away of the LG domain from the helical domain. Accordingly, the R227A/K228A double mutant, is unable to determine the same decoupling effect on protein-protein interaction. The results obtained from mutagenesis studies are in agreement with the results obtained from docking calculations since also **3** unsettle such interactions between the α4' helix and α12/α13 helices (Figure 19). Indeed, the presence of **3** induces the break of the interactions between R227 and K228 and the glutamate residues present on α12, as occurs in the mutant R227E/K228E, thus, inhibiting the formation of GBP1-PIM1 complex.

The obtained results support the hypothesis that the pocket where the 4-APTs bind is an allosteric site capable of controlling both the GBP1 active conformation and its ability to interact with protein partners, such as PIM1.

The investigation has proceeded with the modelling of the GBP1-PIM1 complex[111] by employing a similar docking protocol, adjusted for protein-protein interaction studies[112] (see Experimental Section).

Calculations were performed considering the GBP1 and PIM1 structures as "ligand" and "target", alternatively.

Results showed that the calculated PIM1 binding site partially overlap with **3** (or GTP) binding site, and, more importantly, it interacts with GBP1 regions indicated to be critical for the allosteric regulation of GTP hydrolysis, including the inter-region (Figure 20).
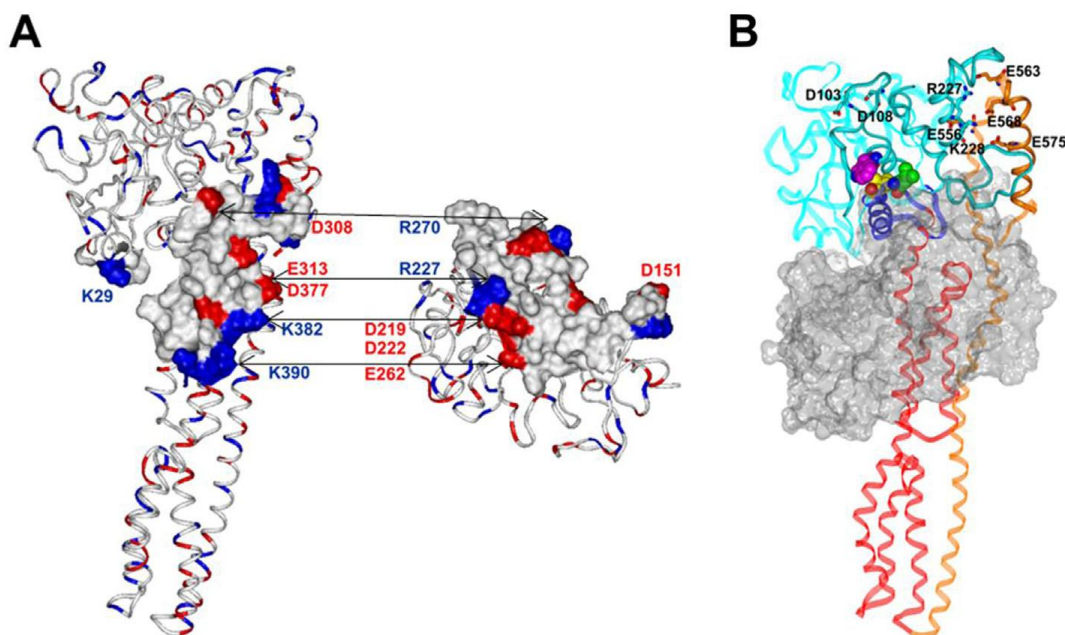

Figure 20. A. Ionic interactions occurring between GBP1 (left) and PIM1 (right) in the calculated structure of the GBP1-PIM1 complex. Involved residues are labelled and interactions are evidenced as arrows with the exception of K29-D151 interaction, for clarity of presentation. Residues with a net positive or negative charge are coloured in blue or red, respectively. Proteins are displayed as ribbons, the solvent accessible surface of the contacting region of the two proteins is shown as solid. **B.** Overall view of GBP1-PIM1 complex. GBP1 structure is displayed as ribbons. The LG (Large GTPase) domain is in cyan, the connecting region in blue, the helical domain in red and α12/α13 in orange. W114 (magenta), I304 (yellow) and L309 (green) are displayed as CPK and coloured by atom type (O = red and N = blue). D103, D108 and key interacting residues between LG domain and α12/α13 are displayed in stick, coloured by atom type and labelled. The solvent accessible surface of PIM1 is displayed as solid (transparency = 75%) and coloured in grey. Figure reproduced from Persico M. et al. (**2015**) with permission of Elsevier (http://dx.doi.org/10.1016/j.ejmech.2014.07.093).

In the calculated GBP1-PIM1 complex, the interaction of R227 and K228 with the cluster of negatively charged residues on the helical domain are conserved. Comparing the two obtained docked complexes (*i.e.*, **3**-GBP1 and PIM1-GBP1) and the available GBP1 X-ray structures, it can be observed that the simultaneous binding of **3** and PIM1 to GBP1 is not possible.

Moreover, our results indicate that the conformational changes provoked by GTP hydrolysis are not compatible with the formation of the GBP1-PIM1 complex too. To further investigate this issue, we decided to model the Full Length (FL) structure of human GBP1 in complex with a mimetic of the transition state of GTP hydrolysis (GDP•AlF$_3$).[111] Indeed, only the X-ray structure of the LG domain in

complex with GDP•AlF₃ is available and the FL model was built by combining the coordinates of two GBP1 X-ray structures and then adding several missing loops (Figure 21, A and B). Two sets of Monte Carlo/SA calculations were performed and two final docked complexes were obtained (Figure 21, C and D). The second set of calculations (**complex 2**; Figure 21D) was performed tethering the R48 residue in the catalytic site to the position found in the X-ray structure of the LG domain in complex with the GTP transition state analogue (Figure 21B).
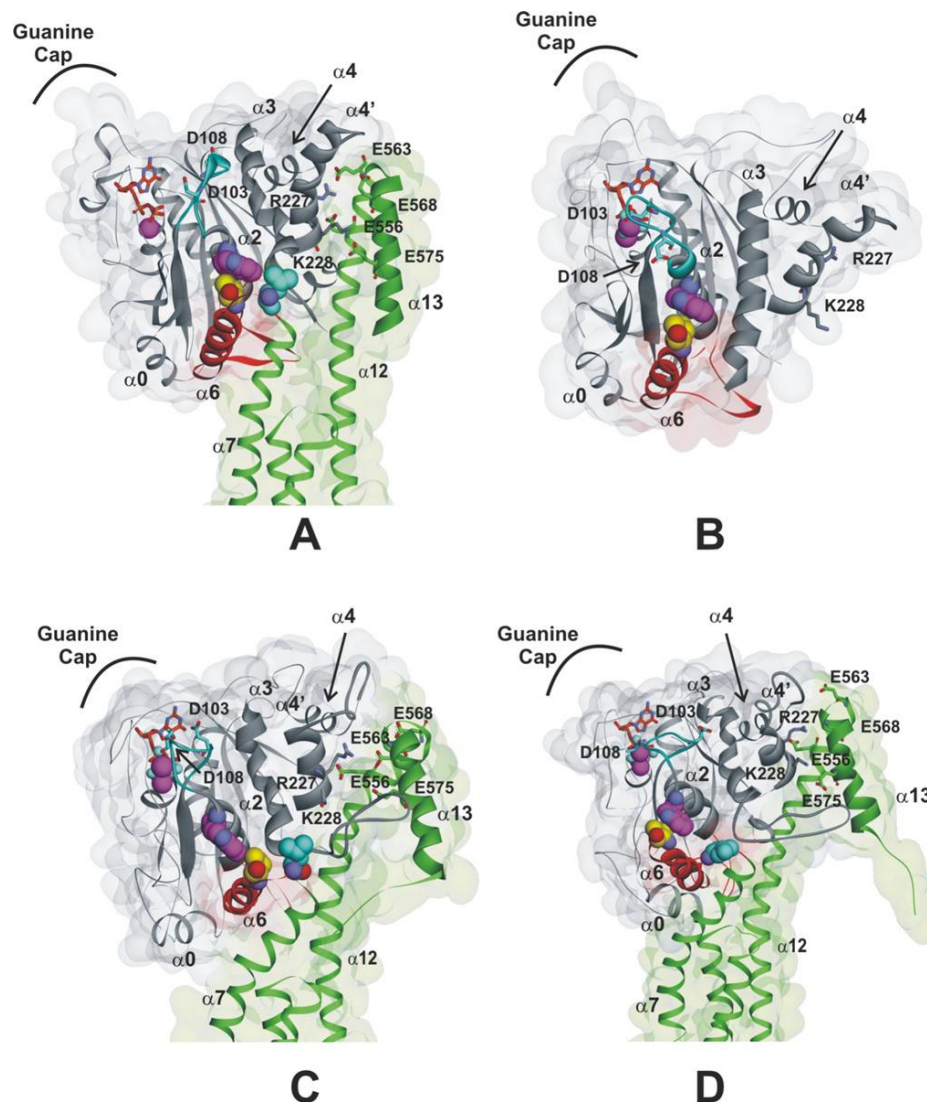

Figure 21. Overview of LG domain of: **A** GppNHp-bound GBP1[FL] (PDB ID: 1F5N); **B** GDP•AlF₃ bound GBP1[LG] (PDB ID: 2B92); **C** GDP•AlF₃ bound GBP1[FL] **complex 1** and **D** GDP•AlF₃ bound GBP1[FL] **complex 2**. GBP1 structure is displayed as ribbons in lines, where the LG domain is in grey (except the SwitchII loop in cyan), the connecting inter-region in red, the helical domain in green. The Guanine Cap, the α0, α2, α3, α4, α4', α6, α7, α12 and α13 helices are labelled. The SwitchII loop, the α2, α3, α4', α6 and α13 helices are evidenced as solid ribbons. W114 (magenta), I304 (yellow) and L309 (cyan) are displayed as CPK and coloured by atom type (O = red and N = blue). D103, D108 and key interacting residues between LG domain and α12/α13 are displayed in stick, coloured by atom type and labelled. The solvent accessible surface of GBP1 is displayed as solid (transparency = 90 %). GppNHp and GDP•AlF₃ (orange) are displayed in stick and coloured by atom type. Figure reproduced from Persico M. et al. (**2016**) with permission of Wiley-VCH Verlag GmbH & Co. KGaA (https://doi.org/10.1002/minf.201501028).

Subsequently, two sets of Monte Carlo/SA calculations were performed and two final docked complexes were obtained (Figure 21, C and D). The second set of calculations (**complex 2**; Figure 21D) was performed tethering a R48 residue in catalytic site, since the latter crucial for the catalysis and otherwise exposed to the solvent, to the position found in the X-ray structure of the LG domain in complex with the GTP transition state analogue (Figure 21B).

GBP1 conformational changes observed in the resulting complexes (Figure 21, C and D) are in agreement with the reported experimental data on the coupling of GTP hydrolysis (LG domain) with structural changes taking place at the C-terminal helical domain, and transmitted by the inter-region to the catalytic site.[98-100] In particular, similarly to the binding of **3**, GTP hydrolysis was hypothesised to inhibit PIM1:GBP1 interaction through an allosteric mechanism involving the ionic contacts between the LG domain (R227 and K228 on the α4' helix) and the helical domain (α12 and α13) of GBP1. Such conformational changes are not compatible with the modelled GBP1-PIM1 complex, indicating that GTP is likely to compete with PIM1 for GBP1 binding.

This hypothesis was finally challenged by experimental studies on the effects of the addition of GTP on the GBP1-PIM1 interaction.[111] It resulted that GBP1 efficiency in binding PIM1 is decreased by the GTP (and not GDP) rate. This makes sense considering that the cytoskeletal gateway of paclitaxel resistance represents a pro-survival adaptation mechanism which is activated in hypoxia, when GTP level decrease and GDP level increase.[88] Since GBP1 binds to GTP and pro-survival kinases using alternative conformational states, or, in other words, GTP hydrolysis and PIM1 binding stabilise alternative GBP1 conformations, then, GBP1 functions can be regulated depending on the rate of the interacting partners. In this view, it has been demonstrated[98,99] that GTP hydrolysis is also coupled to GBP1 oligomerisation and localisation, indeed, GBP1 in the free form, as well as, in the presence of GMP and GDP, is a monomer, while it is a dimer with the non hydrolysable GTP analogue GppNHp, and forms larger oligomers, most likely tetramers, with GTP and the GTP transition state analogue GDP•AlF$_3$. Accordingly, decreased GTP levels can switch the function(s) of GBP1 favoring its binding to pro-survival kinase, such as, PIM1. Taken together, the results of this investigation shed some light on GBP1 allosteric mechanisms regulating its:

- oligomerisation (with related cellular localisation),
- GTPase activity,
- binding to pro-survival kinases.

Knowledge of these mechanisms led to the identification of a new anticancer agent against paclitaxel-resistant cancer cells able to inhibit GBP1:PIM1 interaction and paves the way for the development of new allosteric inhibitors of such PPi, capable to block the conformation of GBP1 in the "GTP status".

Chapter 4

4.1 Cationic Porphyrins as Tunable Gatekeepers of the 20S Proteasome

Proteasomes are barrel-shaped proteolytic assemblies involved in the regulation of the physiological degradation of intracellular macromolecules. The anomalous proteins are habitually degraded by the ubiquitin-proteasome system (UPS), which it is necessary for the non-lysosomal degradation and clearance of short-lived, misfolded, mutant and damaged proteins in eukaryotic cells.[113] This is accomplished through a series of enzyme-mediated reactions that first identify and covalently link abnormal proteins with multiple ubiquitin molecules as a signal for degradation.

The central role of proteasome enzymes in oncogenes processing explains why its inhibition has become a new promising strategy in cancer therapeutics. The human proteasome consists of an empty cylindrical 20S proteolytic core particle (CP), made up by non-catalytic and catalytic protein subunits, which can be capped by one or two regulatory particles (RPs) (*i.e.*, 19S, PA28, and PA200), consisting in additional protein complexes. The RPs have the function of recognising polyubiquitin tags attached to protein substrates and initiate the degradation process, which are multisubunit proteases found in the cytosol, perinuclear regions and nucleus of eukaryotic cells.

The 20S CP is made up by four packed rings, each including seven protein subunits, classified into α and β types: two α-subunit rings (antechambers) border two central β-subunit rings (catalytic chambers)[114-116] (Figure 22).
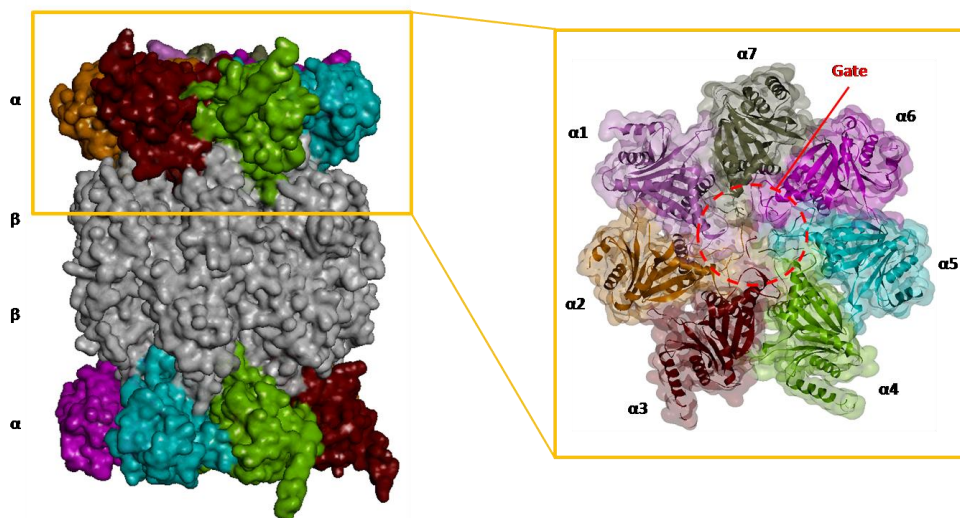
Figure 22. Left: 3-D-structure of 20S human proteasome (PDB ID: 4R3O); α subunit rings are labelled and coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7); β subunit rings are labelled and coloured in grey. Right: detailed top view of the α subunit ring; α subunits are coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7); the entrance gate of the CP channel is circled in red.

The outer rings comprise seven different α-subunits, none of which has catalytic activity, but which serve as an anchor for the multisubunit ATPase-containing PA700 (19S) regulator (700 kDa) that binds to form a complex referred to as the 26S proteasome (Figure 23).
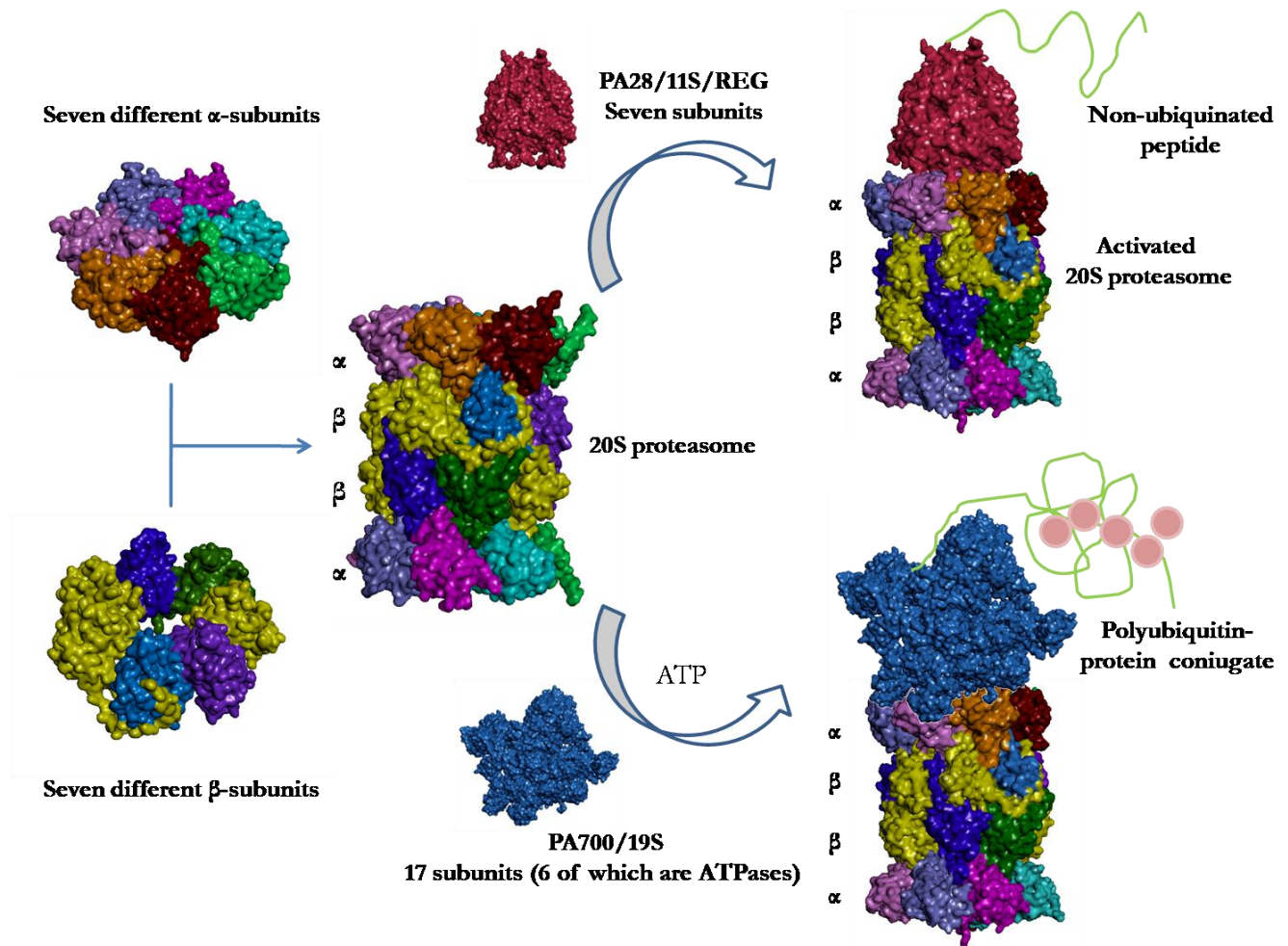
Figure 23. Schematic representation of composition and organisation of the 20S proteasome and their intracellular activators. Figure adapted from McNaught K. S. et al. (**2001**), Nature (http://dx.doi.org/10.1038/35086067).

Linking of the 19S regulatory complex serves two functions: it opens the channel through the 20S proteasome, which is normally gated by the amino termini of the α-subunits, and it unfolds ubiquitinated proteins to allow entry to the catalytic core; both processes require ATP.[117] The PA 28 (11S) regulatory (REG) complex can also bind to the 20S proteasome and open the channel through the complex, but this process is ATP-independent, and mediates the degradation of non-ubiquitinated short peptides(Figure 23).[117]

In the catalytic chambers, there are two duplicates of three different proteolytically active β subunits (*i.e.*, β1, β2, and β5) exhibiting caspase-like (PGPH-L), trypsin-like (T-L), and chymotrypsin-like (ChT-L)

activity.[118-120] (Figure 24).

Figure 24. 3-D-structure of 20S human proteasome (PDB ID: 4R67) in complex with Bortezomid; α subunit rings are labelled and coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7); the β subunit rings are 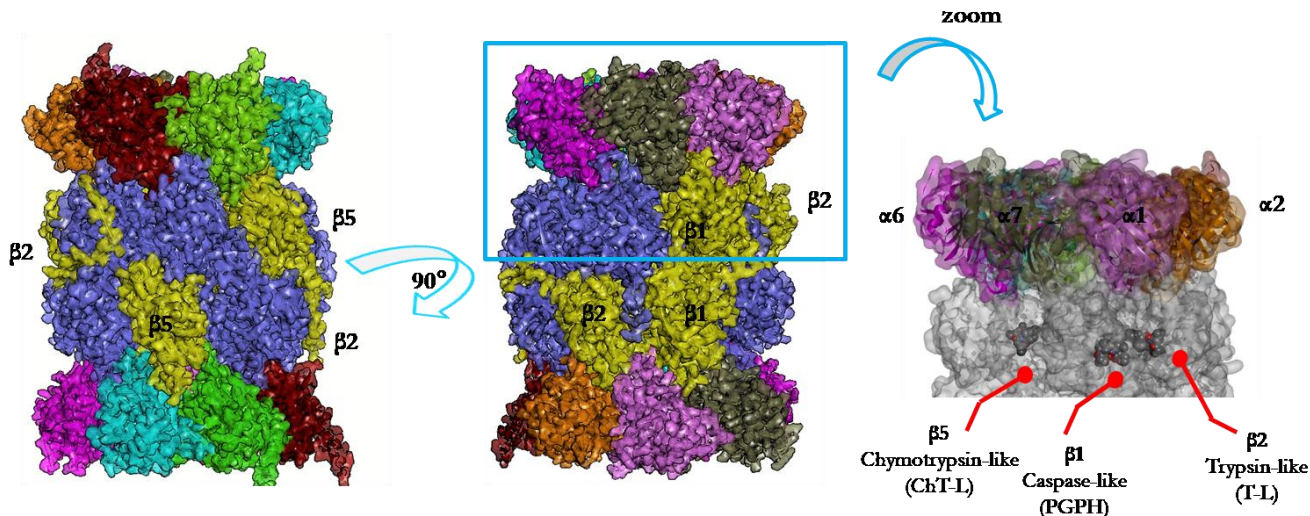coloured in grey, with the exception of the catalytic β subunits labelled and coloured in yellow. The ligands are displayed in CPK and coloured by atom type (C = grey, O = red and N = blue).

The access of proteasome substrates to the catalytic site(s) is finely regulated through the gating mechanism. The opening of the gate implies conformational changes displacing the α-subunit tails and it is induced by the RPs, although it can occur also in absence of them.[121] Indeed, it has been recently demonstrated that an equilibrium between the open and closed CP conformation exists *in vitro* in absence of any ligand. The modulation of this equilibrium by exogenous ligands represents an attractive therapeutic approach for the cure of a number of diseases, first of all, cancer. Indeed, considering the relevant increase of the intracellular CP upon oxidative stress, its inhibition may bring about the accumulation of misfolded proteins in cancer cells, and, ultimately, lead to cell-cycle arrest and cell death.[122-124] Bortezomib, a dipeptide boronate acting as competitive covalent inhibitor of all proteasome catalytic sites, was approved by the FDA for the treatment of several hematological tumors and, in particular, multiple myeloma. Unfortunately, the use of Bortezomib as an anticancer agent showed severe limitations due to its side effects and resistance phenomena.[125,126] Therefore, the search of new molecules, that inhibit proteasome activity by binding to sites far away from the catalytic centers, may be a promising anticancer strategy. In this context, in 2012 the first scientific work evidencing the anti-proteasome ability of cationic porphyrins was published.[127] These studies showed that micromolar amounts of some of the tested porphyrins (Figure 25) reversibly inhibit the catalytic activities of 20S. In particular, in this paper was reported a selection of porphyrins comprising the tetra-cationic porphyrin (H$_2$T4 - *meso*-tetrakis(4-N-methylpyridyl) porphyrin), the tetra-anionic porphyrin (H$_2$TPPS - *meso*-tetrakis(4-sulphonatophenyl)-

porphyrin) and tris- and bi-cationic substituents (Figure 25).
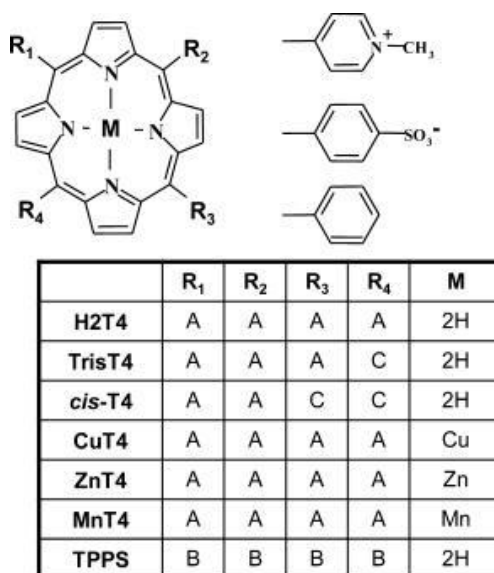
Figure 25. Chemical structure of all the porphyrin derivatives used in previous study. Figure reproduced from Santoro A. M. et al. (**2012**), *J. Am. Chem. Soc.* (http://dx.doi.org/10.1021/ja300781u).

From this preliminary screening some structure activity relatioships (SARs) came to light: i) the 20S inhibitory activity is not due to the metal-coordination properties of the porphyrin system, ii) the 20S inhibitory activity increase with the number of positively charged nitrogens present in the structure (Figure 26), with no inhibitory activity observed for the neutral or negatively charged analogues.
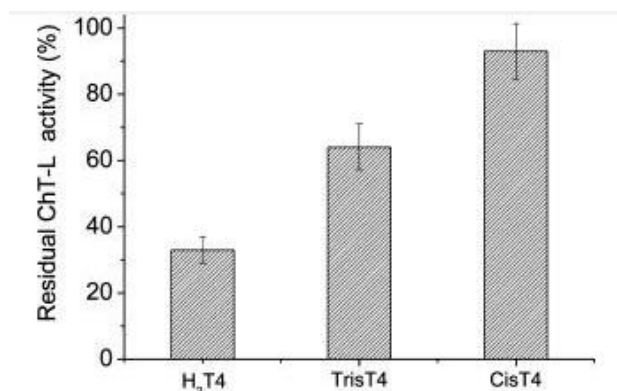
Figure 26. Residual chymotryptic-like (ChT-L) activity of 20S proteasome in cell lysates with di-, tri- and tetra-cationic porphyrins at 1 µM. Data are related to control sample normalised to 100%. Figure reproduced from Santoro A. M. et al. (**2012**), *J. Am. Chem. Soc.* (http://dx.doi.org/10.1021/ja300781u).

Accordingly, the most potent compound resulted the tetra-cationic derivative H₂T4, showing similar potency against the three proteasome catalytic activities (PGPH-L, T-L, and ChT-L) (Figure 27). Generally, the proteasome ChT-L activity is mostly probed in order to measure the activity of proteasome; indeed, previous reports indicate that the chymotryptic substrate is the most proteasome-specific of those examined[128] (Figure 27).
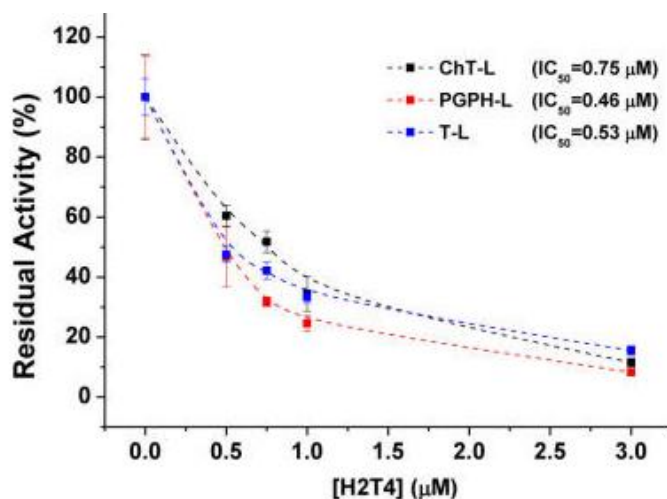
Figure 27. Concentration-response plot of H₂T4 for Chymotryptic-like (ChT-L), Caspase-like (PGPH-L), and Tryptic-like (T-L) residual activities on human 20S from cell lysates. Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

Inspired by these results, we performed our studies on the title compound - H₂T4 - and two of its isomers, *ortho*-H₂T4 and *meta*-H₂T4[129] (Figure 28).

Cmp	X	W	J	Y	Z
H₂T4	N-CH₃	N-CH₃	N-CH₃	CH	CH
meta-H₂T4	CH	CH	CH	N-CH₃	CH
ortho-H₂T4	CH	CH	CH	CH	N-CH₃

Figure 28. Structures of the cationic porphyrins derivatives: H₂T4, *meta*-H₂T4, and *ortho*-H₂T4.

In order to investigate the interaction of H₂T4 with 20S human proteasome at the molecular level, computational studies were integrated and validated with experimental data coming from the other

research groups involved in the project, such as:

- Prof. Milardi, Istituto di Biostrutture e Bioimmagini (IBB) - CNR UOS di Catania;

- Prof. A. M. Santoro, Istituto di Biostrutture e Bioimmagini (IBB) - CNR UOS di Catania;

- Prof. R. Purrello, Dipartimento di Scienze Chimiche, Università di Catania;

- Prof. M. Coletta, Dipartimento di Scienze Cliniche e Medicina Traslazionale, Università di Roma Tor Vergata;

- Prof. R. Fattorusso, Dipartimento di Scienze e Tecnologie Ambientali, Biologiche e Farmaceutiche, Seconda Università degli Studi Napoli;

- Prof. R. Paolesse, Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma Tor Vergata- Via della Ricerca Scientifica.

4.2 The Electrostatic Map of Proteasome α-Rings Encode the Rules for Designing Allosteric Porphyrin-based Inhibitors.

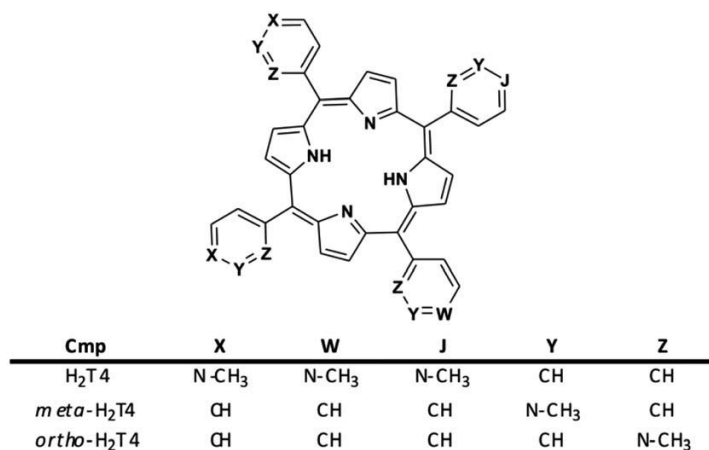*In vitro* inhibitory data (IC_{50}s) of the the new derivatives (Figure 28) confirmed and expanded previous SARs indicating that, not only the number, but also the relative positioning of the protonated pyridine nitrogen atoms, play a key role on the inhibition of the catalytic activities of 20S (Figure 29).
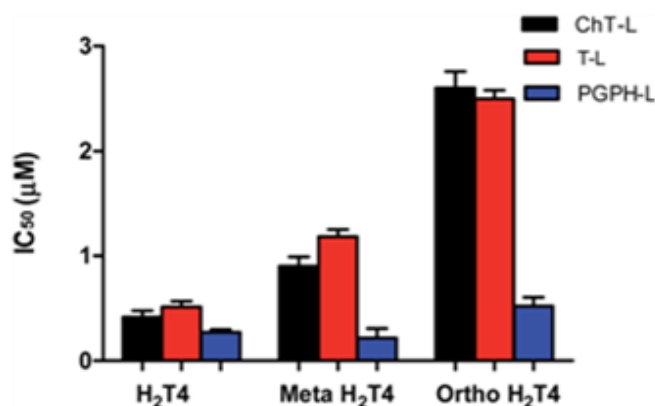
Figure 29. Comparison of the IC_{50} values of *para*-H$_2$T4 - *meso*-tetra(4-*N*-methyil-pyridyl) porphyrin -, *meta*-H$_2$T4 - *meso*-tetra(3-*N*-methyilpyridyl) porphyrin - and *ortho*-H$_2$T4 - *meso*-tetra(2-*N*-methyil-pyridyl) porphyrin - determined for the ChT-L (black), T-L (red), and PGPH-L (blue) peptidase activities of the CP. Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

Thus, at first, a 3-D-pharmacophore model of the lead compound H$_2$T4 was generated by assuming the four protonated nitrogen atoms as crucial interaction points, taking into account the H$_2$T4 X-ray structures present in the Cambridge Structural Database (CSD; http://www.ccdc.cam.ac.uk) (Figure 30A).
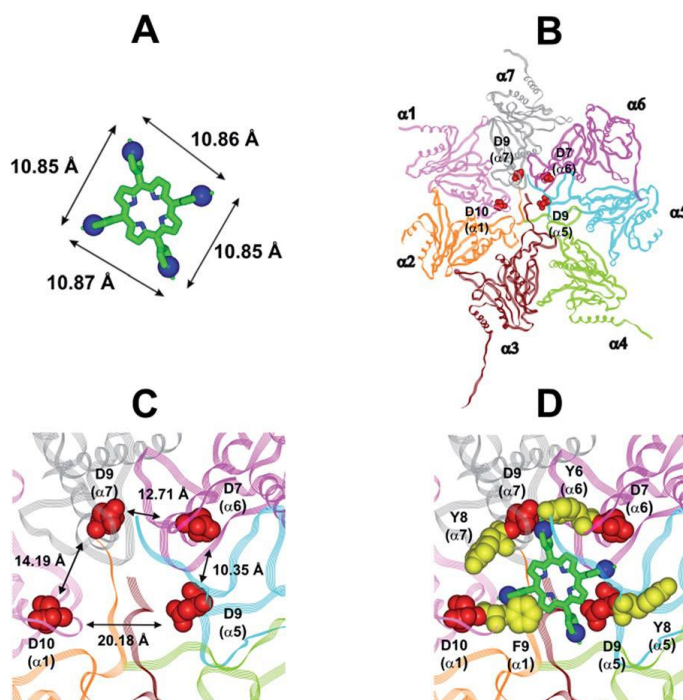
Figure 30. A. H$_2$T4 pharmacophore and related inter-atomic distances. The experimentally determined structure of H$_2$T4 (CSD code: OBOZAI) is displayed in stick with the pyridine nitrogen atoms evidenced in CPK. **B**. Top view of the 20S human proteasome (PDB ID: 4R3O); only the α ring is shown for clarity of presentation. **C.** Zoom of the top view of the 20S human proteasome (PDB ID: 4R3O); the four Asp residues are displayed as CPK and coloured in red and the suitable inter-residue distances for a possible interaction with the H$_2$T4 pharmacophore are reported. **D**. Positioning of H$_2$T4 in the putative binding site; the potential interactions between H$_2$T4 and the 20S proteasome are shown: the amino acid residues involved in ionic and cation-π interactions are displayed as CPK and coloured in red and yellow, respectively. H$_2$T4 is coloured by atom type (C: green; N: blue). The a subunits are coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7). Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

Then, I assessed the capability of the 20S human proteasome to arrange the planar, rigid and positively charged pharmacophore of H$_2$T4 by mapping the spatial positioning of negatively charged amino acids on the protein surface, as well as in the known functional and inhibitor binding sites (pharmacophore fitting procedure). The structural and bioinformatics analysis was carried out taking into account all the experimentally determined conformational states (closed, semi-closed, and open) and the sequence homologies among the different species were computed using the PROMALS3-D server (http://www.prodata.swmed.edu/promals3-D/promals3-D.ph).[67] Firstly, the available experimentally determined structures of 20S were downloaded from the Brookhaven Protein Data Bank (http://www.rcsb.org) and subjected to a structural and bioinformatics analysis. Then, an homology model of the 20S human

proteasome α-subunits was built. It must be mentioned here that, after the completion of molecular modelling studies, the X-ray structure of the human 20S proteasome was solved and deposited in the Brookhaven Protein Data Bank (PDB ID: 4R3O). Consequently, the structural parameters were compared using a structure evaluator software (*i.e.,* MolProbity)[72] which calculated the degree of similarity between the model and the X-ray structure. It resulted a 100% match of secondary structures (*i.e.*, α-helices, turns, and β-strands) and a RMSd value on all α-carbons of 0.61 Å. Acquired outcome throw out all the 20S catalytic sites and addressed us to suppose the substrate gate as the most plausible ligand binding site, for the competitive inhibition of 20S catalytic activities (Figure 30B), due to the presence of four Asp residues on the N-terminal tails of α1, α5, α6, and α7 which showed a reasonable spatial positioning for a conceivable interaction with the positively charged H_2T4 nitrogens (Figure 30C). Moreover, the four nearby aromatic residues could help the binding by instituting polarised π-π interaction with the pyridine rings (Figure 30D). In this way, H_2T4 was situated at its supposed binding site on the human proteasome model and the acquired complex was utilised as beginning structure for the completely dynamic docking protocol. Docking calculations were performed using the CVFF force field[130] and the Cell_Multipole method for non-bonded interactions.[131] In spite our docking calculations requires a beginning complex, it has to be pointed out that to completely investigate all conceivable binding sites and modes, the binding domain was defined as the whole α-ring of the human 20S proteasome and distance constraints were applied only to structurally conserved regions (see below and the Experimental Section for details). This enabled a complete exploration of:

- ligand binding sites and/or modes,
- ligand-induced protein large-scale conformational changes.

Initially, a Monte Carlo/minimisation method[69] for the stochastic conformational sampling was used and a maximum of 20 adequate complexes, by selecting the minimum energy state from the previous cycle, were generated. In order to guarantee a broad variability of the structures to be progressively minimised, an energy tolerance value of 10^6 kcal mol^{-1} from the previous structure was utilised. Following, an energy test and a structure similarity evaluation were adopted to select the 20 acceptable structures, with an energy range of 50 kcal mol^{-1} and rms tolerance = 0.3 kcal Å^{-1} respectively. Subsequently, the obtained complexes were subjected to simulated annealing (SA) calculations, during which the temperature of the system was linearly decreased from an initial (500 K) to a final (300 K) temperature. A temperature equivalent to 500 K was applied with a view to overcoming torsional barriers, in this manner allowing a complete rearrangement of both the protein and the ligand (see Experimental Section for details). All calculations were performed by applying constraints only to H-bonds (α-helices) and ψ and φ torsion angles (β-sheets) of some structurally conserved regions, by employing several force constant values in conformity with the

results of secondary structure prediction calculations (http://www.predictprotein.org/). Specifically, a force constant of 100 (high), 10 (medium), and 1 (low) kcal mol^{-1} was used to H-bonds (α-helices) and ψ and φ torsion angles (β-sheets) by following secondary structure prediction score. The remaining atoms of the protein was left without constraints and free on its movement. The quality validation of the obtained complexes was then estimated by employing an all-atom structure valuation software (*i.e.*, MolProbity).[72] Docking studies revealed interesting outcomes about H_2T4 capacity to inhibit the 20S proteasome behaving like a "plug" of the CP entrance gate. In the first place, all the complexes obtained by this dynamic docking procedure, either by Monte Carlo and SA calculations, always presented H_2T4 bound to the CP access channel, interacting with the located N-terminal residues of the α-subunits (Figure 31).
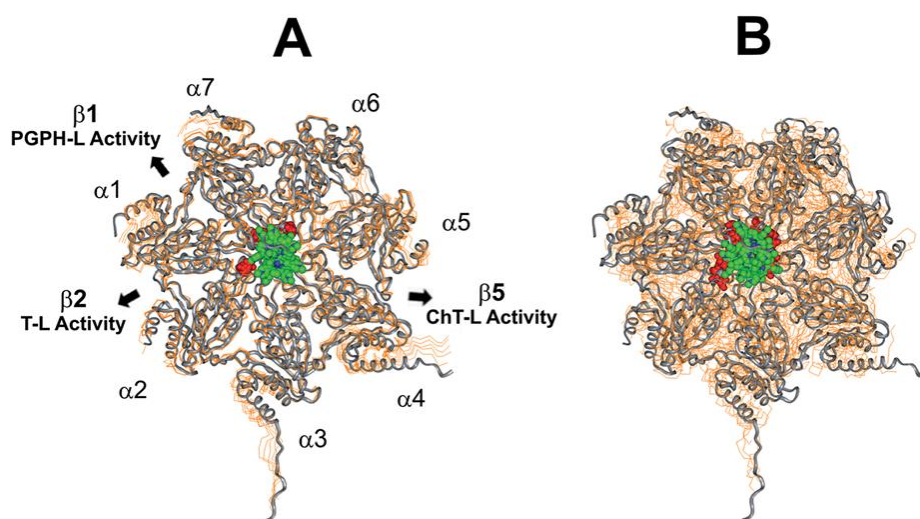
Figure 31. Top view of the dynamic docking results obtained for H_2T4: **A** Monte Carlo and **B** SA. The backbone of the starting complex is displayed as solid ribbons and coloured in grey, the one of the calculated complexes is displayed as line ribbons and coloured in orange. The cluster of negatively charged residues at the entrance gate of the CP channel is displayed as CPK and coloured in grey (starting complexes) and red (calculated complexes). The porphyrin ligands are coloured by atom type (C: green; N: blue; H: white) and displayed as CPK. In (A) the α-subunits and the position of the catalytic b subunits are labelled. Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

It is noteworthy that the H_2T4 binding mode of the selected final SA complex (the one characterised by the most favourable non-bonded interaction energy) significantly differed from the (starting) Monte Carlo complex (Table 5; Figure 32).

Table 5 | Non-bonded interaction energies (kcal/mol) of the 20S-H_2T4 complexes obtained by Monte Carlo and SA calculations.

Cplx	Non-bonded interaction energies (kcal/mol)	
	Monte Carlo simulation	Simulated Annealing simulation
1	-58.009	-63.827
2[a]	-36.088	-92.754
3	-48.788	-51.330
4	-54.795	-71.749
5	-43.954	-51.868

[a]Selected complex

Table adapted from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

In the Monte Carlo complex H_2T4 binds to the gate assuming a position perpendicular to the α-ring plane and interacts with only two of the four Aspartate residues (Figure 32A). After the SA procedure, H_2T4 resulted fully stuck into the gate, parallel to the α-subunits plane, interacting with the whole cluster of functional residues of the predicted binding site (α1 (F9, D10), α3 (Y5), α5 (Y8, D9), α6 (Y6, D7), and α7 (Y8, D9)) (Figure 32B).
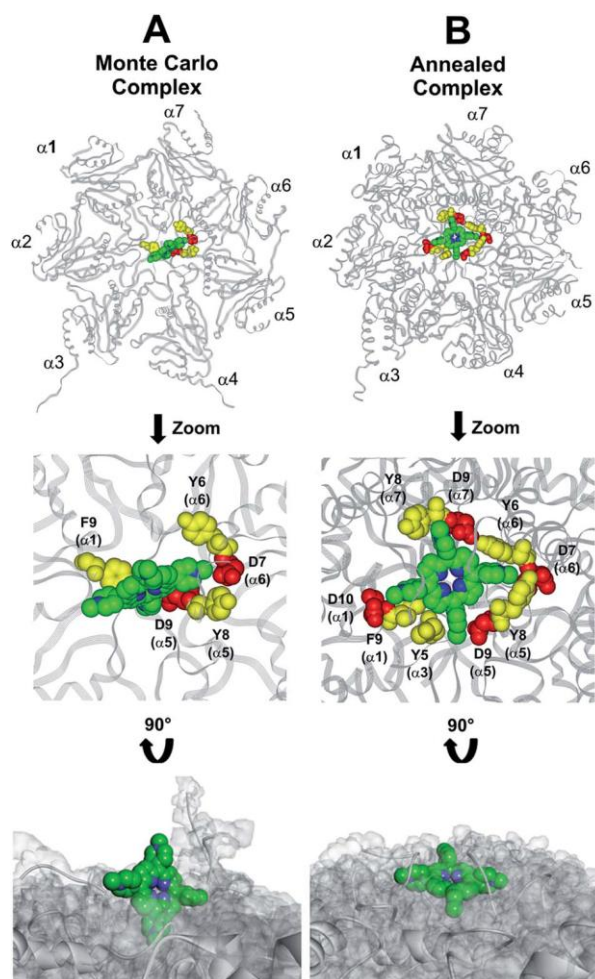
Figure 32. Results obtained by the dynamic docking procedure. H_2T4-20S Monte Carlo **A** and SA **B** complexes. The α-subunits are displayed as line ribbons and coloured in grey. H_2T4 is displayed as CPK and coloured by atom type (C: green; N: blue). The amino acid residues involved in ionic and cation-π interactions are displayed as CPK and coloured in red and yellow, respectively. Protein van der Waals volume is displayed as transparent surface (bottom representation). Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

Importantly, on the basis of these results, an experimental kinetic investigation on the interaction between H_2T4 and 20S proteasome by stopped flow UV spectroscopy and NMR studies was carried out by professor M. Coletta and R. Fattorusso, respectively (Figure 33A and B).
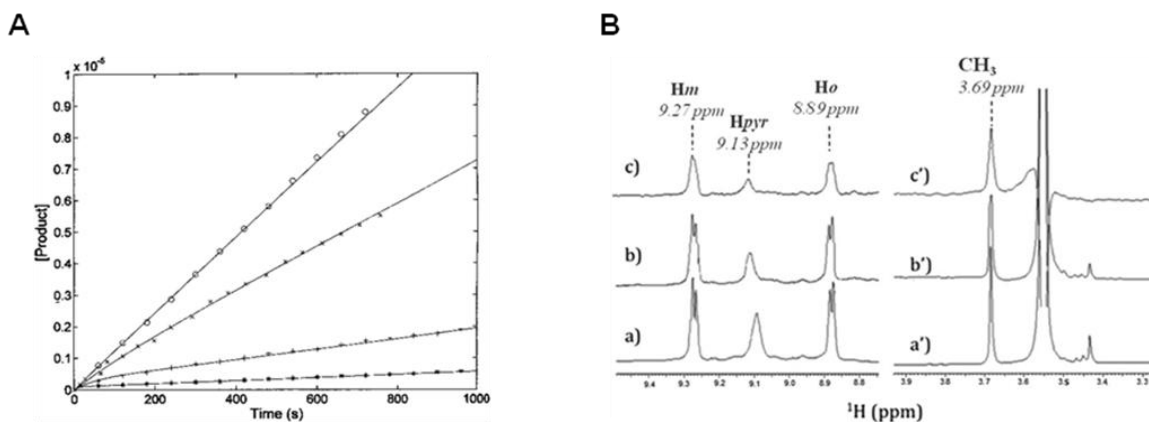
Figure 33. A. Kinetics of fluorescent product formation, obtained by mixing 2nM 20S proteasome, 100 mM of substrate Suc-LLVY-AMC and different concentrations of H_2T4 porphyrin, namely 0 (O), 0.3 mM (X) 3 mM (+) and 10 mM (*). **B**. Low-field (left) and high-field (right) ^1H NMR spectra of H_2T4 (172 mM) (a) in Tris-HCl (20 mM) pH 7.6, NaCl (150 mM) solution, and (b) in the presence of 20S (860 nM); (c) ^1H NMR STD spectrum of H_2T4 in the presence of 20S (860 nM). 1H chemical shift assignment has been reported for H_2T4 in presence of 20S proteasome. Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

The kinetic investigations highlighted the ability of H_2T4 to inhibit the CP by a competitive mechanism and by a multiphasic pattern. In particular, the resulting kinetic data confirmed that there is a first faster interaction of H_2T4 with the N-terminal tails of the α-subunits, likely corresponding to the first encountering of the porphyrin with the α-subunits plane (Figure 32A), followed by a tighter binding to the closed CP gate, which corresponds to the inhibition event and can be referred to the formation of the more stable porphyrin-protein complex (Figure 32B). On the other hand, the NMR studies showed that the major driving force in porphyrin/proteasome interaction are the methyl-pyridine moieties of H_2T4.

For the purpose of validating our protein-ligand complex model, *meta*-H_2T4 and *ortho*-H_2T4 derivatives were exposed to the same dynamic docking procedure used to H_2T4 obtaining interesting results too. Indeed, the docking simulation, contrarily to what observed for H_2T4 and in agreement with the varied spatial positioning of their protonated nitrogen atoms, revealed that *meta*-H_2T4 and *ortho*-H_2T4 do not remain bound to the identified N-terminal Aspartate residues at the 20S gate (Figure 34). In the complex presenting the most favourable non-bonded interaction energy (Figura 34 and Table 6), they bind at a side of the gate, interacting only with F9 (α1) and Y6 (α6) (*meta*-H_2T4) and Y8 (α5) (*ortho*-H_2T4).

Table 6 | Non-bonded interaction energies (kcal/mol) of the 20S-*meta*-H$_2$T4 and 20S-*ortho*-H$_2$T4 complexes obtained by SA calculations.

Cplx	Non-bonded interaction energies (kcal/mol)	
	20S-*meta*-H$_2$T4 complexes	20S-*ortho*-H$_2$T4 complexes
1[a]	-44.569	-108.642
2	-71.200	-89.385
3	-46.433	-84.523
4[b]	-88.350	-30.610
5	-62.243	-37.593
6	-83.360	-82.561
7	-36.456	-72.742
8	-60.625	-52.663
9	-49.381	-79.814
10	-58.207	-104.267
11	-47.554	-50.418
12	-38.331	-48.609
13	-58.209	-44.056
14	-35.057	-81.299
15	-44.122	-48.068
16	-0.774	-49.181
17	0.574	-80.797
18	-24.133	-56.376
19	0.314	-60.809
20	-75.163	-70.525

[a]Selected complex for *ortho*-H$_2$T4; [b]Selected complex for *meta*-H$_2$T4
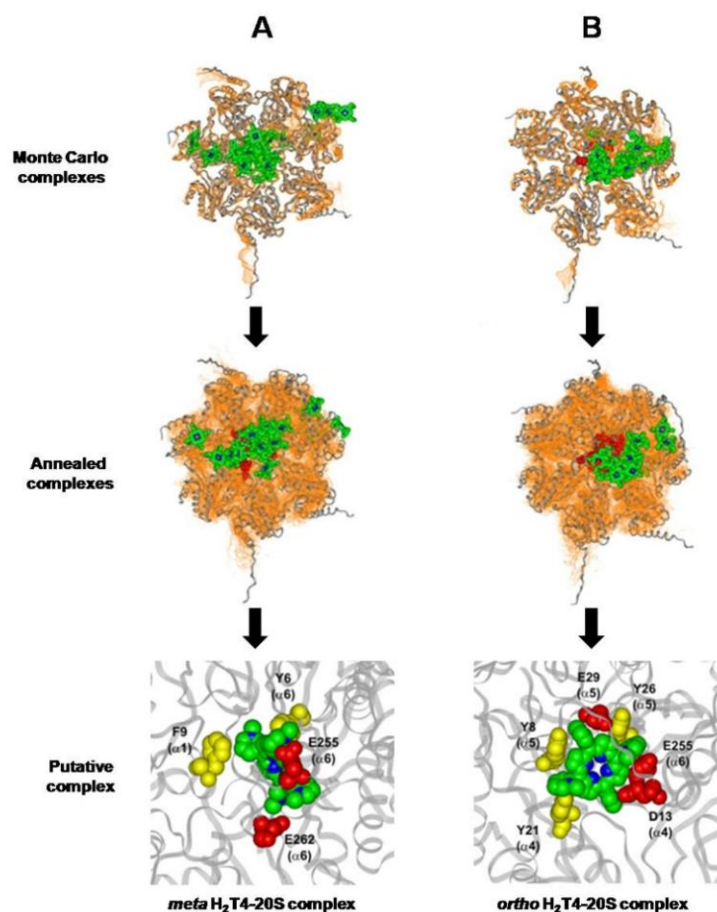
Figure 34. Comparison of the dynamic docking results obtained for **A** *meta*-H$_2$T4, and **B** *ortho*-H$_2$T4. The starting complex is displayed as solid ribbons and coloured in grey, the one of the calculated complexes is displayed as line ribbons and coloured in orange. The cluster of negatively charged residues at the entrance gate of the CP channel is displayed as CPK and coloured in grey (starting complexes) and red (calculated complexes). The porphyrin ligands are coloured by atom type (C: green; N: blue, and H: white) and displayed as CPK. In the zoomed view the amino acid residues involved in ionic and cation-π interactions are coloured in red and yellow, respectively, and displayed as CPK. Figure reproduced from Santoro A. M. et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5sc03312h).

Curiously, *ortho*-H$_2$T4 make interactions with two negatively charged residues in the groove between α5 and α6 subunits, which are known to be engaged in the binding with the positively charged C-terminal tails of regulatory proteins (RPs).[132-135] Therefore, the different binding modes of the three isomers appears to mirror their different inhibitory potencies. Indeed, as depicted in Figure 29, the *in vitro* CP inhibition assay evidenced the ability of *meta*-H$_2$T4 and *ortho*-H$_2$T4 derivatives to inhibit the three catalytic activities to different extents and with lower inhibitory potency with respect to H$_2$T4. According to our docking simulation, *meta*-H$_2$T4 and *ortho*-H$_2$T4 could modulate the catalytic activity through a more refined tight fit with proteasome allosteric regulation compared to H$_2$T4 (which blocks gate access), being the three proteasome catalytic activities subordinated to the opening of the gate by means of various allosteric mechanisms, involving the grooves between adjacent protein subunits.

On the premise of these outcomes, it is conceivable to propose that H$_2$T4 binding to the 20S α-subunits

could influence the catalytic activities both clogging the entry of all substrates and modulating the conformational equilibrium of the 20S human proteasome, bringing about a shift in the equilibrium of the "open-to-close" structural transition. These peculiar mechanisms of action coupled with their versatile chemistry, propose cationic porphyrins as a new class of CP conformational regulators with great potentialities as lead pharmacophores. Moreover, the key role played by electrostatics in driving porphyrin:proteasome interactions suggested that these highly versatile inhibitors may be fine-tuned by conjugating the parent frame with different charged moieties, endowing special properties to the molecule.

In this view, the negatively charged residues of the human 20S implicated in ionic interactions with the positively charged amino acids of PA200, PA28α, and 19S RPs have been mapped, discovering putative binding sites for cationic porphyrin-based allosteric inhibitors.

In particular, a bioinformatics and structural analysis was carried out to identify the negatively charged residues of the human 20S CP which are implicated in ionic interactions with positively amino acids of the RPs. In fact, the interaction with such negative residues from small-molecule ligands may have an effect on the allosteric mechanism correlating the opening/closing of the gate with the activation of proteolytic activity.[136,137] In this analysis, we took into account the three RPs of 20S human proteasome: PA28α (REGα/11S), PA200, and 19S.

PA28 is an ATP-independent activating complex of human 20S, which consists of seven structural subunits with three recognised isoforms (PA28α-γ).[138] For the purpose of mapping the interactions of the human 20S with PA28, the X-ray crystallography complex between *Saccharomyces cerevisiae* 20S with *Trypanosoma Brucei* PA26 (human PA28 homolog)[139] was employed as structure of reference. The residues of the complex that are engaged in electrostatic interactions between the CP (negative) and the RPs (positive) were detected. At the end, the information was translated to the relating human homologs employing the X-ray structures of PA28α and 20S human (closed and open states) (Figure 35A and Table 7).
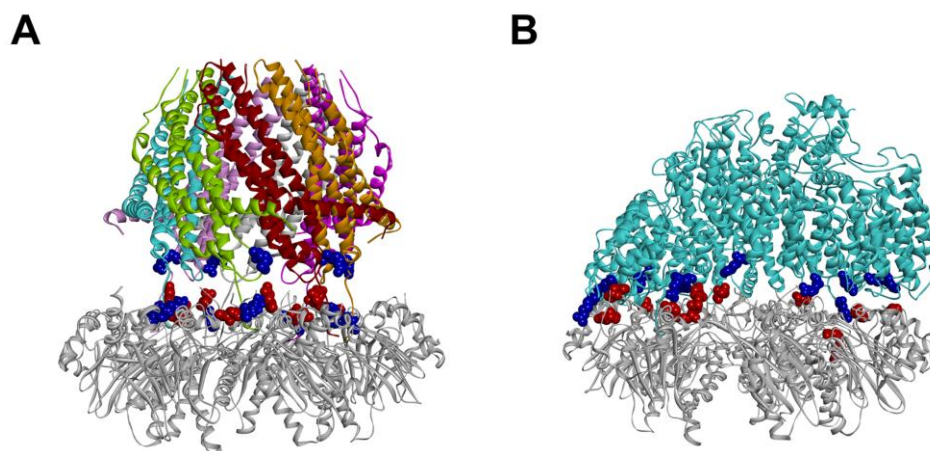
Figure 35. Homology models of 20S human (α-ring) in complex with **A** PA28 and **B** PA200. The residues involved in ionic interactions are evidenced (CPK render) and coloured in red (negative, 20S) and blue (positive, PA28). Proteins are displayed as solid ribbons; PA28 (A) is coloured by subunit type. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Table 7 | Ionic interactions between 20S human (negative) and PA28α(positive) obtained using as reference structure the X-ray complex of 20S yeast proteasome with *Trypanosoma Brucei* PA26 (PDB ID: 1ZTQ).

20S yeast proteasome (negative)	Trypanosoma Brucei PA26[a] (positive)	20S human proteasome (negative)	human PA28α (positive)[a]
E22 (α1)	K100 (Subunit 7)	E19 (α1)	R141 (Subunit 7)
-	-	E26 (α1)	K245 (Subunit 7)
-	-	E23 (α2)	K245 (Subunit 1)
E16 (α3)	K100 (Subunit 2)	E15 (α3)	R141 (Subunit 2)
-	-	E22 (α3)	K245 (Subunit 2)
D14 (α4)	K100 (Subunit 3)	D13 (α4)	R141 (Subunit 3)
-	-	E20 (α4)	K245 (Subunit 3)
E18 (α5)	K100 (Subunit 4)	E18 (α5)	R141 (Subunit 4)
-	-	E25 (α5)	K245 (Subunit 4)
-	-	E23 (α6)	K245 (Subunit 5)
D17 (α7)	K100 (Subunit 6)	D18 (α7)	R141 (Subunit 6)
-	-	E25 (α7)	K245 (Subunit 6)

[a]PA26 and PA28 subunits are numbered according to PDB structures 1ZTQ and 1AVO, respectively.
Table reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Contrary to the prior RP, despite PA200 is an ATP-independent activator protein too, the stimulation of proteasomal hydrolysis of PGPH-L substrates is nearly three times greater compared to the hydrolysis of ChT-L and T-L substrates.[140] The ionic interactions arising between PA200 and 20S human (open/closed states) were recognised making use of the X-ray complex of *Saccharomyces cerevisiae* 20S with the RP Blm10[139] (human PA200 homolog) as template structure. Acquired results are reported in Figure 35B and Table 8.

Table 8 | Ionic interactions between 20S human (negative) and PA200 (positive) obtained using the X-ray complex of 20S yeast proteasome with yeast Blm10 (PDB ID: 4V7O) as reference structure.

20S yeast proteasome (negative)	yeast Blm10 (positive)	20S human proteasome (negative)	human PA200 (positive)
-	-	E175 (α2)	K1025
-	-	E200 (α2)	K711 K878
-	-	E58 (α3)	K1069
-	-	D202 (α3)	R1074
-	-	E209 (α3)	K1069
E173 (α4)	R1798	E170 (α4)	R1502
E7 (α5)	K589	E7 (α5)	K411
D9 (α5)	K2112	D9 (α5)	R1811
E18 (α5)	K2111	E18 (α5)	R1810
E25 (α5)	K2112	E25 (α5)	R1811
E27 (α6)	K2111	E27 (α6)	R1810
D203 (α6)	R2048	E202 (α6)	R1742
-	-	D204 (α6)	R1740

Table reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Finally, 19S is an ATP-dependent RP complex that stimulates roughly three-fold all catalytic activities of 20S human.[141] It is composed by various ATPase and non-ATPase subunits that are put in order according to a cylindrical-shaped assembly. Using cryo-electron microscopy (Cryo-EM) techniques, Chen et al.[142] achieved significant information about proteasome 3-D shape by solving four different 20S-19S complexes (called S_{A-D}): from the ground state (S_A) to the opening of the 20S gate channel (S_D). The four structures were employed in our analysis in order to examine the ionic interactions taking place at 20S-19S interface (Figure 36 and Table 9).
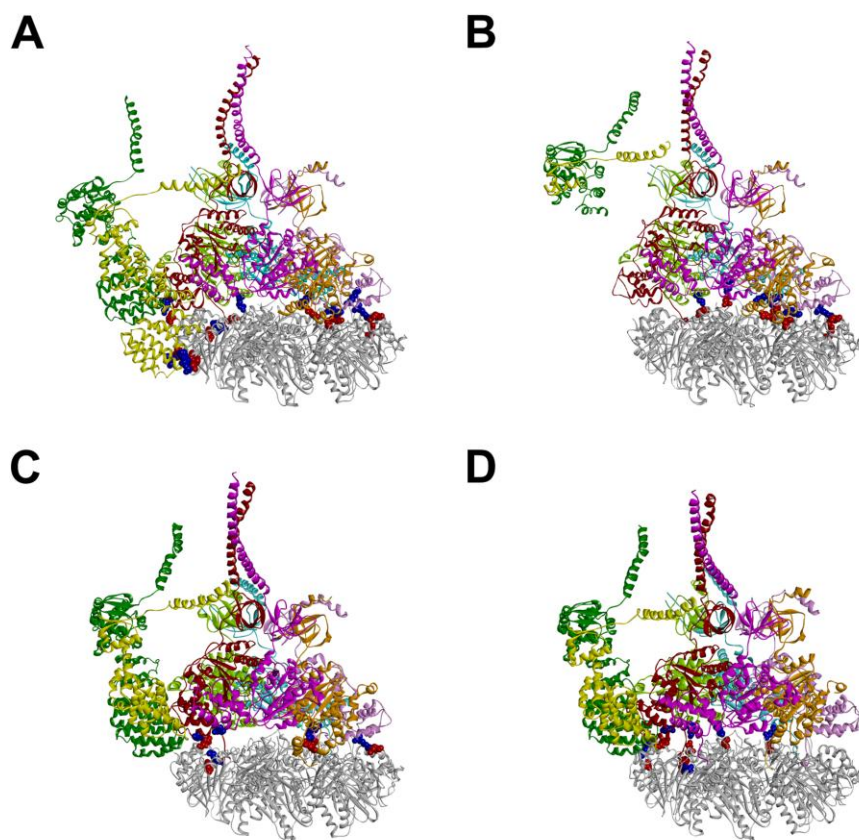
Figure 36. Ionic interactions occurring between human 20S and 19S according to Cryo-EM data Side view of: S$_A$, S$_B$, S$_C$, and S$_D$ conformational/functional states. The residues involved in ionic interactions between 20S (negative: red) and 19S (positive: blue) are displayed as CPK. Proteins are displayed as solid ribbons and coloured as follows: 20S α-ring (grey), 19S Rpt1 (pink), Rpt2 (orange), Rpt3 (brown), Rpt4 (light green), Rpt5 (cyan), Rpt6 (magenta), Rpn5 (green), and Rpn6 (yellow). Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Table 9 | Ionic interactions between 20S human (negative) and the RP protein complex 19S (positive) in the S$_{A-D}$ states (PDB IDs: 5T0G; 5T0H; 5T0I; 5T0J).

20S human proteasome (negative)	19S human (positive)
D58 (α1)	R372 (Rpt4)
D155 (α1)	K418 (Rpt3)
E174 (α1)	R366 (Rpt3)
	K408 (Rpt3)
	K404 (Rpt3)
D236 (α1)	K36 (Rpn5)
E244 (α1)	K36 (Rpn5)
E49 (α2)	K418 (Rpt3)
E183 (α2)	K121 (Rpn6)
E185 (α2)	R122 (Rpn6)
D228 (α2)	K84 (Rpn6)
E15 (α3)	R333 (Rpt2)
E24 (α4)	K422 (Rpt1)

E18 (α5)	K430 (Rpt5)
E29 (α5)	R351 (Rpt1)
E51 (α5)	R375 (Rpt1)
E207 (α5)	R375 (Rpt1)
E173 (α6)	R386 (Rpt5)
D9 (α7)	K387 (Rpt4)
E176 (α7)	K369 (Rpt4)
D203 (α7)	K193 (Rpt5)

Table reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

The achieved results enabled to map the spatial positioning of the negatively charged residues of 20S human engaged in ionic interactions with the all three RPs. Curiously, it emerged a regular disposition of such amino acids on the outer layer of the α-ring (Figure 37), revealing a structural code which clarifies the important role played by ionic interactions in the allosteric regulation of proteasome catalytic activity.[143]
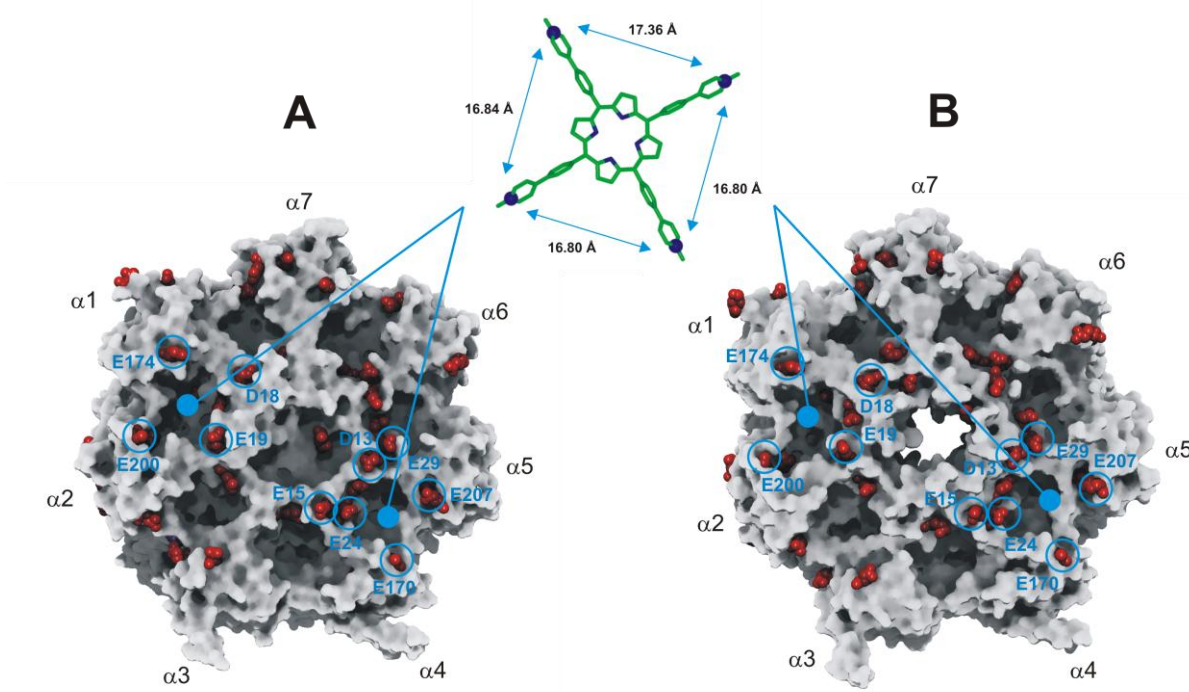
Figure 37. Putative binding sites of pTMPyPP4 on the α-ring surface of 20S human in the closed **A** and open **B** conformation. The putative interacting residues are labelled and circled in blue. The ligand is displayed in stick and coloured by atom type; hydrogens are omitted for sake of clarity. The inter-atomic distances among the pyridine nitrogen atoms (evidenced in CPK) are reported. The 20S human proteasome (top view) is displayed as Connolly surface and coloured in grey except for negatively charged residues involved in ionic interactions with PA28, PA200, and 19S RPs which are displayed as CPK and coloured in red. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

This structural analysis allowed us to identify the 20S negatively charged residues placed at the rim of the α1-α2 and α4-α5 grooves as supposed binding sites for cationic ligands having a planar conformation of

the positively charged groups, similar to **1** (Figure 38).
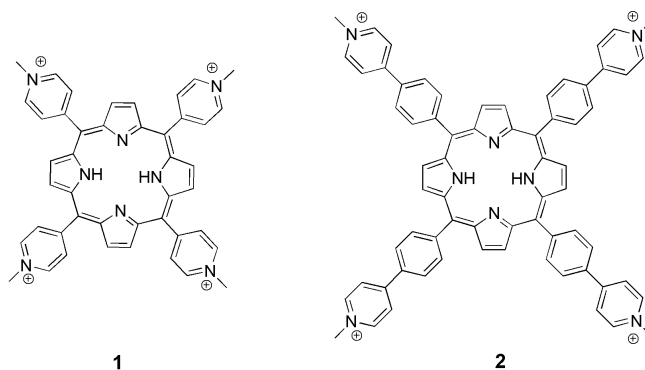
Figure 38. Chemical structures of cationic porphyrins H$_2$T4 (**1**) and pTMPyPP4 (**2**). Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Nevertheless, according to the inter-atomic distances, to allow the electrostatic interactions with the negatively charged amino acids of the protein, a greater distance (17 Å) between the protonated pyridine nitrogen atoms of the ligand is necessary with respect to that showed by **1** (11 Å). Consequently, a new polycationic porphyrin, having in the meso positions four phenylpyridine moieties (*meso*-tetrakis(4-N-methylphenyl pyridyl)-porphyrin pTMPyPP4 (**2**) has been designed and synthesised (see Figure 38). In fact, in **2** a phenyl ring is inserted between the pyridine substituent and the porphyrin planar core, thus leading to a distance among the adjacent charged groups of approximately 17 Å (Figure 37). Albeit additional putative binding sites have also been identified at the α1-α2 groove (Figure 37), however, the presence of a large number of negatively charged residues characterises the α4-α5 groove, thus proposing this as the more likely binding site for **2**. To corroborate this finding, **2** was subjected to dynamic docking calculations in complex with 20S human.

It should be underlined that the docking calculations have been carried out considering the whole 20S human proteasome structure available experimentally, including α-β-β-α rings for closed conformation and α-β rings for open conformation. Despite the extensive calculation time and the difficulty to model such a system due to the large number of residues, our approach allowed to explore the intricate mechanism underlying the dynamic of such protein complex.

To perform docking calculations on both, the closed and the open 20S structure, the molecular model of 20S human proteasome in the open conformation was built and optimised by MM (see the Experimental Section for details). On the other hand, the previously developed model of 20S human[129] was used to perform docking calculations with 20S human in the closed conformation. Afterward, to reproduce the

dynamic of the ligand-protein recognition event, the previously described docking protocol based on a Monte Carlo-Metropolis/simulated annealing (SA) approach was applied.[68,69,70]

Calculations were carried out starting from both the closed and the open 20S conformations, giving rise to two sets of results (Figures). The obtained complexes from Monte Carlo and SA were classified according to their ligand-protein interaction energies (Group Based non-bonded interactions method)[144] (Figures 39 and 40). Moreover, the quality of the chosen complexes was then evaluated by using structure evaluator software (*i.e.*, Molprobity).[72]

Many of the obtained SA complexes showed the ligand bound to the α4-α5 groove, with only six obtained complexes (using as starting structure the closed conformation of 20S human proteasome) exhibiting the position of ligand in the substrate channel at the level of the first β-ring, or at the interface between the two β-rings (Figures 39C).

Figure 39. A. Dynamic docking results obtained for **2** using as starting structure the closed conformation of 20S human proteasome. **B**. Top view of dynamic docking results without the two rings of β subunits and the second ring of α subunits. **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 coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7, α-subunits of the second ring and all β-subunits). The backbone of the calculated complexes is displayed as line ribbons and coloured in orange. The porphyrin ligands are coloured by atom type (C: green and N: blue) and displayed as CPK. In B the α subunits and the catalytic β-subunits are labelled. In C the catalytic β-subunits are labelled and coloured in pink (β1), violet (β2), and cyan (β5). The last six generated complexes (each docked complex is generated from the last accepted structure, see the Experimental Section for details) presented the ligand positioned in the substrate channel at the level of the first β-ring, or at the interface between the first and the second β-ring (evidenced in **C**); however, the binding to the α4- α5 groove (evidenced in B) is favoured according to the calculated non-bonded interactions energies. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).
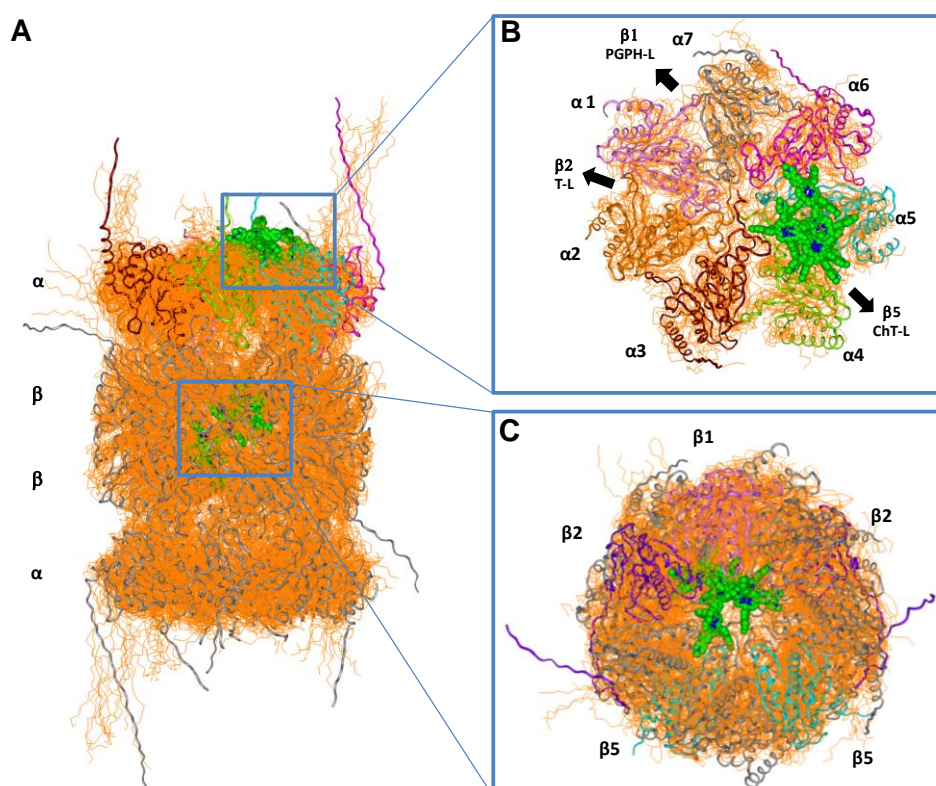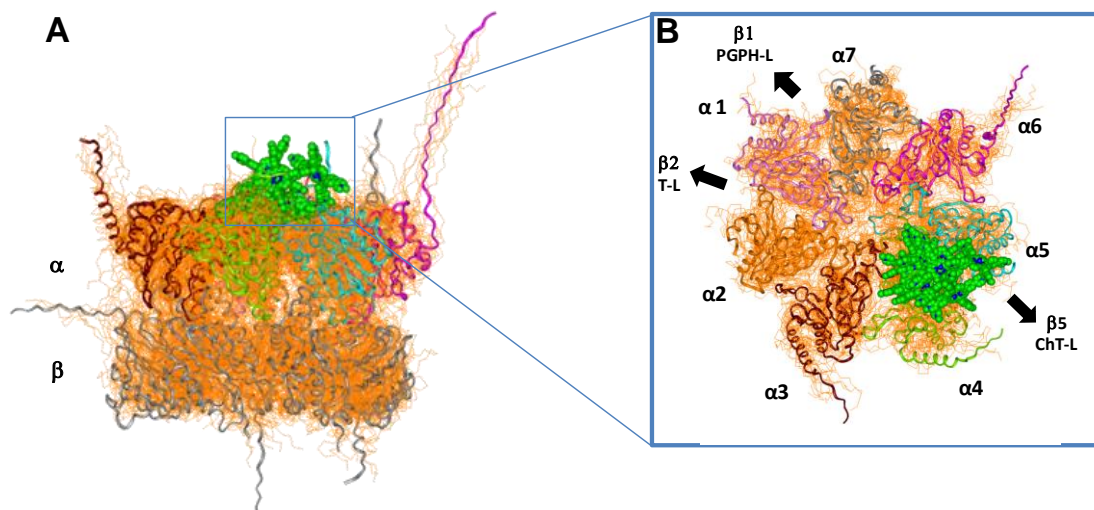
Figure 40. A. Dynamic docking results obtained for **2** using as starting structure the open conformation of 20S human proteasome. **B**. Top view of dynamic docking results without the ring of β subunits. The backbone of the starting complex is displayed as solid ribbons and coloured in pink (α1), orange (α2), brown (α3), light green (α4), cyan (α5), magenta (α6), and grey (α7 and β subunits). The backbone of the calculated complexes is displayed as line ribbons and coloured in orange. The porphyrin ligands are coloured by atom type (C: green and N: blue) and displayed as CPK. In B the α subunits and the catalytic β subunits are labelled. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Anyway, in accordance with the calculated non-bonded interaction energies (Tables 10 and 11), the complex with the best compromise between the non-bonded interaction energies obtained by Monte Carlo and SA calculations has been selected. According to the calculated non-bonded interaction energies, not only the two selected complexes but also the other favourable resulting SA complexes showed, in any case, the ligand placed within the α4-α5 groove.

Table 10 | Non-bonded interaction energies (kcal/mol) of the 20S-**2** complexes obtained by Monte Carlo and SA calculations using as starting structure the closed conformation of 20S.

Cplx	Non-bonded interaction energies (kcal/mol)	
	Monte Carlo simulation	Simulated Annealing simulation
1	-42.363	-66.518
2	-42.948	-81.754
3[a]	-53.260	-76.404
4	-50.670	-50.246
5	-25.029	-60.942
6	-6.877	-57.008
7	-6.649	-65.689
8	-4.466	-21.349
9	-6.976	-63.463
10	-17.344	-59.432

11	-3.415	-50.117

Table 11 | Non-bonded interaction energies (kcal/mol) of the 20S-**2** complexes obtained by Monte Carlo and SA calculations using as starting structure the open conformation of 20S.

Cplx	Non-bonded interaction energies (kcal/mol)	
	Monte Carlo simulation	Simulated Annealing simulation
1	-14.601	-56.056
2	-6.522	-55.024
3	-5.914	-0.147
4	-0.689	-62.805
5	0.672	-16.006
6	0.075	-42.347
7	-0.904	-70.354
8	-0.234	-60.497
9a	-3.970	-66.822
10	-1.321	-61.119
11	-1.646	-13.116

Considering the interactions established by **2** in the best docked complexes, when the 20S closed conformation as starting structure was used, the polycationic porphyrin **2** shows interactions with the α4-α5 groove by establishing (Figure 41A):

- ionic interactions with six negatively charged residues *(i.e.,* E20(α4), E24(α4), E170(α4), E25(α5), E29(α5), and E207(α5); all involved in the binding to RPs),

- cation-π interactions with three aromatic residues (*i.e.,* F17(α4), Y21(α4), and Y26(α5)) and one positively charged residue (*i.e.,* R53(α5)),

- a π -π interactions with a histidine residue (*i.e.,* H15(α4)) (Figure A).
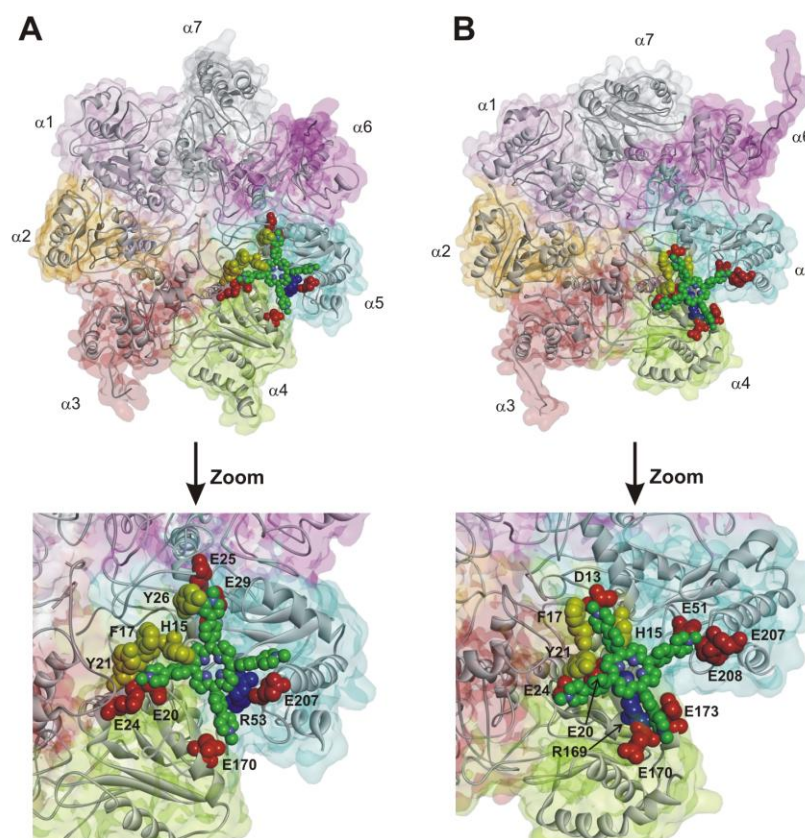
Figure 41. The 20S-**2** docked complexes obtained starting from the closed **A** or open **B** conformation of 20S. The α-subunits are displayed as ribbons and coloured in grey. **2** is displayed as CPK and coloured by atom type (C: green; N: blue). The amino acid residues involved in ionic, cation-π and π-π interactions are displayed as CPK and coloured in red (negatively charged residues), yellow (aromatic residues), and blue (positively charged residue). The protein van der Waals volume is displayed as transparent surface and coloured by subunit type. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7). Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Moreover, the binding of the ligand at the α4-α5 groove induced conformational changes which are not circumscribed to the α-ring, but also affect the interface with the β5-subunit and, in turn, the shape of the S3 catalytic site pocket (Figure 42).
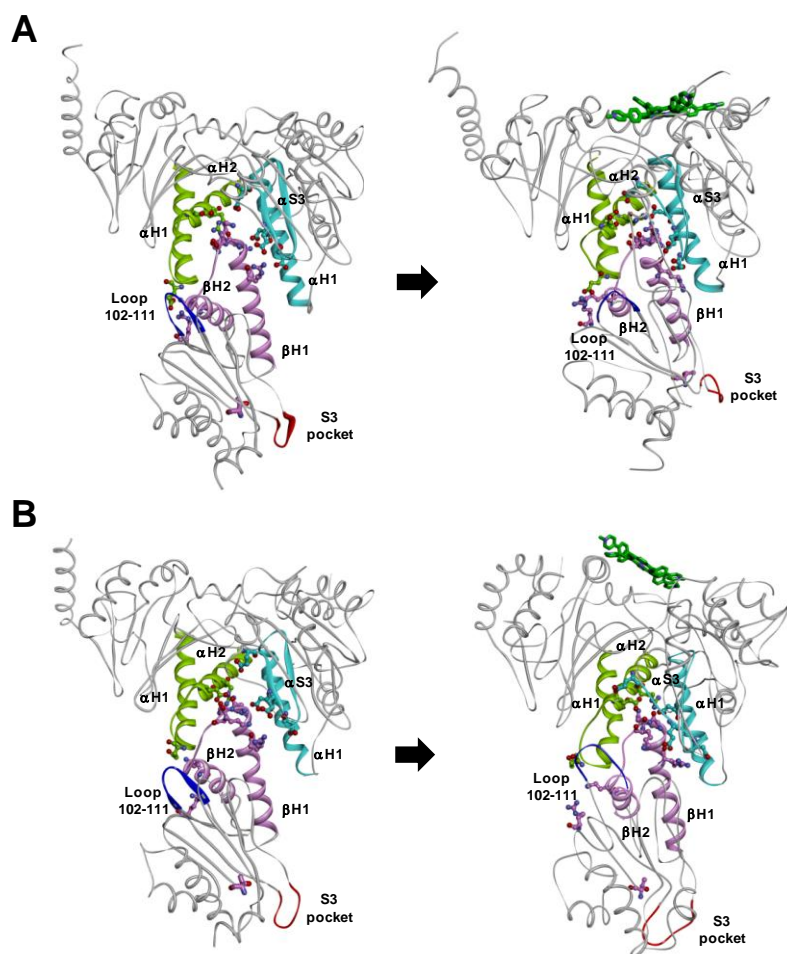
Figure 42. α4, α5, and β5 subunits before (left) and after (right) docking calculations in complex with **2**. **A**. Starting from the 20S closed state. **B**. Starting from the 20S open state. The protein is displayed as grey ribbons. αH1 and αH2 (α4-subunit; light green), αS3, αS4 and αH1 (α5-subunit; cyan), βH1 and βH2 (β5-subunit; pink), the loop region aa102-111 (β5-subunit; blue), and the S3 pocket (β5-subunit; red), are evidenced as solid ribbons. Key residues are displayed in ball and sticks. **2** is displayed as stick and coloured by atom type (C = green and N = blue). In (**A**) the conformational changes at α4-α5 groove caused, in the region at the interface among the subunits α4, α5 and β5, the break of the ionic interaction between E65 (αS3 strand of α5-subunit) and K71 (loop between the helices βH1 and βH2 of β5-subunit) which moved to interact with E69 (αS3 strand of α5-subunit). In turn, this change moved the helix αH2 (aa104-121) of the subunit α4 toward both the β5-subunit, determining the formation of a H-bond between R109 (α4) and N70 (β5), and the α5-subunit determining the formation of an ionic interaction between R109 (α4) and D90 (α5). Moreover, this movement modified the positioning of the βH2 helix with respect the loop region (aa102-111) within the β5-subunit. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Noteworthy, the regions involved in the calculated conformational changes were previously indicated to be implicated in the allosteric regulation of the 20S catalytic activity by NMR studies.[136,142] In particular, **2** provoked conformational changes involving the closing of the loop which describes the S3 pocket (*i.e.*, aa20-30) in proximity to the catalytic threonine (Figure 42). This structural movement is opposite compared to that watched in the shift from the closed to the open conformation of 20S, accounting for

the ability of **2** to allosterically inhibit proteasome catalytic activity. The binding of **2** at the α4-α5 groove influenced the conformation of the other α-grooves too, by changing the spatial positioning of the negatively charged residues that make interaction with the RPs (Figure 43).
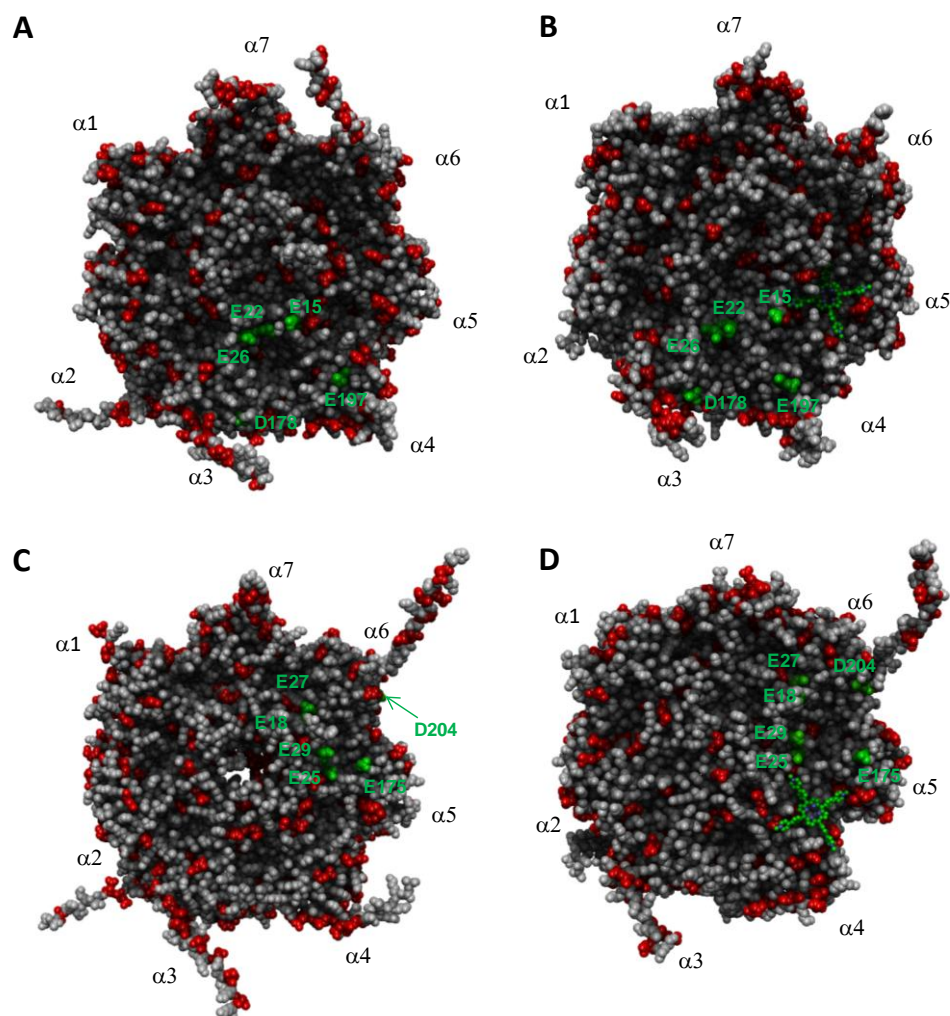
Figure 43. Top view of: **A** 20S human closed state, **B** 20S human in complex with **2** (Cplx_3), **C**) 20S human open state, and D) 20S human in complex with **2** (Cplx_9). In **B** and **D 2** is displayed in ball and stick and coloured by atom type (C = green; N = blue). The 20S human proteasome is displayed as CPK and coloured in grey. All negatively charged residues are coloured in red with the exception of those present in the groove α3-α4 (**A** and **B**) or in the groove α5-α6 (**C** and **D**) which are evidenced in green and labelled. Hydrogens are omitted for sake of clarity. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

Thanks to this conformational change, the negatively charged amino acids at the α3-α4 groove, result in a spatial arrangement capable to interact with the four positively charged pyridine of **2**, enabling the supposed binding of a second ligand molecule (Figure 43B).

In the second set of docking results, achieved using 20S open conformation as starting protein structure, **2** interacts with the α4-α5 groove establishing:

- ionic interactions with six residues also found to interact with RPs (*i.e.*, D13(α4), E20(α4), E24(α4), E170(α4), E51(α5), and E207(α5)),

- cation-π interactions with two aromatic residues (*i.e.*, F17(α4) and Y21(α4)) and one positively charged residue (*i.e.*, R169 (α4)),

- a π -π interaction with H15(α4) (Figure 41B).

In this case too, immediately after the binding of **2** to the α4-α5 groove, conformational changes were induced to the whole α-ring surface, revealing the formation, also in this case, of a further putative ligand binding site, which, this time, is located at the negatively charged residues identified to interact with RPs at the α5-α6 groove (Figure 43-D *vs*. 43C). Furthermore, the α4 N-terminal tail moved into the substrate gate, making interaction with the backbone carbonyl groups of residues G8 (α1), R3(α6), and N4 (α6) thanks to charge assisted H-bonds. This prompts the positioning of α1, α4, and α6 N-terminal tails at the substrate gate, inducing its steric blockage. The docking of **2** with human 20S in the open conformation caused lesser conformational changes at the interface with β5-subunit (Figure 42B *vs*. 42A) compared to the first set of docking results (20S in the closed conformation). In fact, the positioning of the βH2 helix with respect the loop region (aa102-111), as well as the positioning of the loop region (aa20-30) with respect to the catalytic threonine, resulted similar, to the end simulation, to those present in the starting (open) structure (Figure 42B).

Overall, the docking results indicate that **2** may bind at the α4-α5 groove of 20S, not only in the closed but also in the open conformation. Moreover, the binding of **2** means conformational changes promoting the accommodation of a second ligand molecule. It have to bear in mind that bioinformatics and structural analysis showed another possible binding site at the α1-α2 groove, which involved the negative residues implicated to the anchoring of RPs, thus, three supposed binding site for **2** were identified on each α-ring surface of the 20S human CP.

Importantly, on the basis of these results the inhibition mechanism of 20S human proteasome by **2** was investigated. Results evidenced that: i) **2** is able to inhibit the three catalytic activities at the same extent, with IC_{50} values very similar to **1** (around 2 μM) (Figure 44) , ii) **2** displays a purely non-competitive inhibition mechanism (Figure 45A) contrarily to what observed for **1** which acts as a competitive inhibitor. These experimental evidences perfectly matched our computational hypothesis.
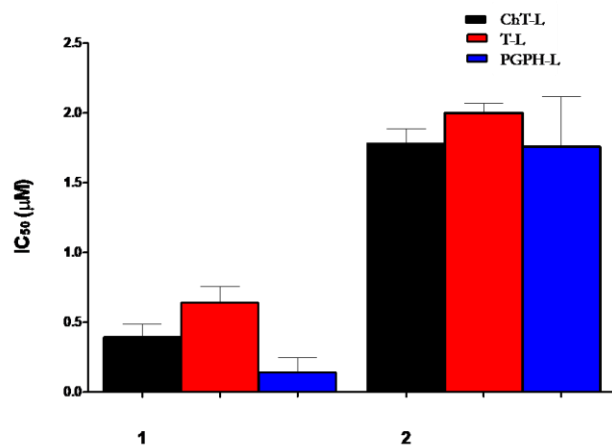
Figure 44. Comparison of the IC_{50} values of **1** and **2** determined for the ChT-L (black), T-L (red) and PGPH-L (blue) peptidase activities of the CP. Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

An almost pure non-competitive inhibitory mechanism was also observed when tested against yeast 20S proteasome (see Figure 45B), confirming that this porphyrin acts through a similar molecular mechanism against both human and yeast 20S proteasome. On the contrary, in the case of the yeast mutant α-3ΔN, where the first nine amino acids of the α-3 N-terminal tail have been removed, seemingly blocking the 20S proteasome in the "open" conformation,[145,146] the porphyrin behave as a competitive inhibitor (see Figure 45C).
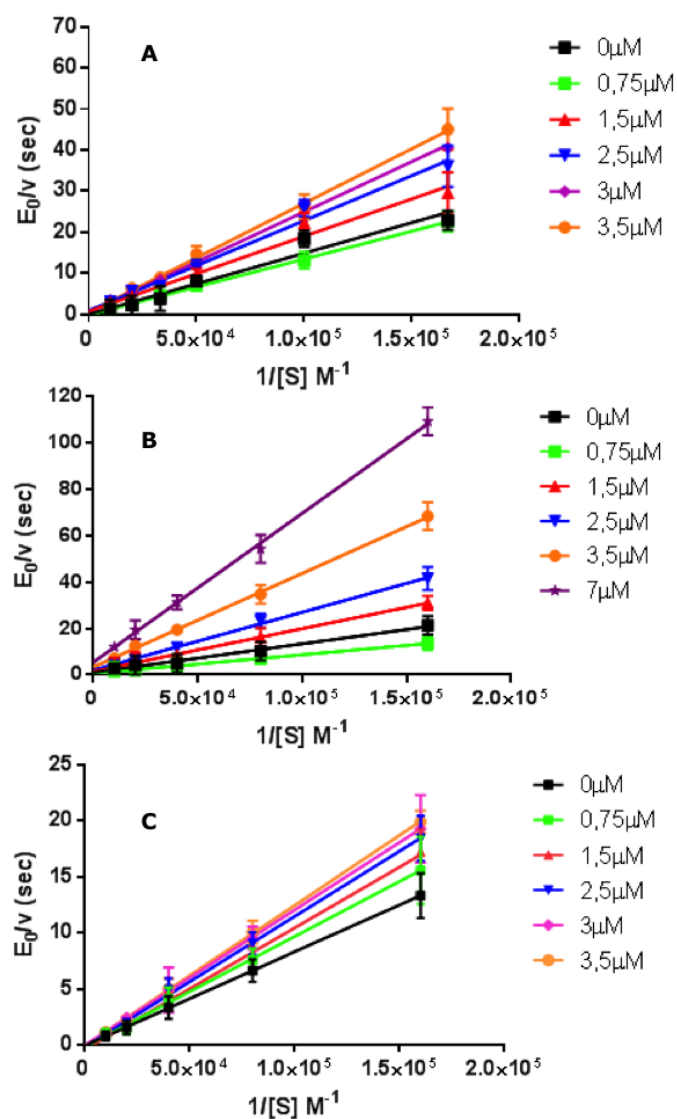
Figure 45. Double reciprocal Lineweaver-Burk plot of substrate Suc-LLVY-AMC enzymatic processing in the absence of inhibitor (black) and in the presence of increasing amounts of **2** (as indicated), for various concentrations of the substrate, by 20S human proteasome (*panel* A); by 20S yeast proteasome (*panel* B); by α-3ΔN 20S yeast mutant (*panel* C). Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

.

In addition, the analysis of the kinetic data, obtained from stopped-flow kinetic experiments conducted by prof. M. Coletta research group, revealed the cooperative binding of at least three inhibitor molecules to two different clusters of binding sites on human 20S, envisaging the possibility that this porphyrin is a conformational modulator of the 20S proteasome, likely affecting the open/closed equilibrium, in full accordance with *in silico* results (Figure 46). This experimental evidence confirmed the results obtained by computational studies, which indicated the presence of three potential binding sites on both proteasome faces, one of which formed after the binding of the first ligand molecule to the first (higher affinity) binding site located at the α4-α5 groove.
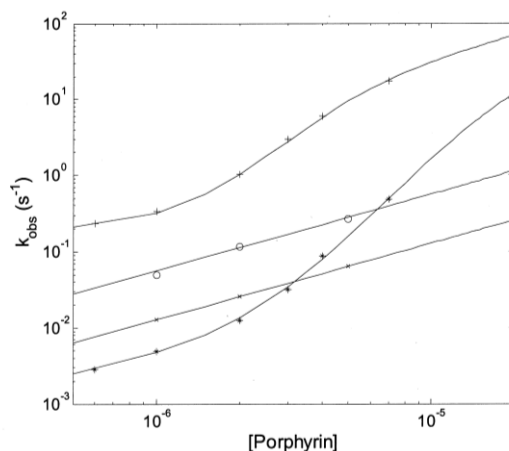
Figure 46. Porphyrin concentration dependence for the binding rate constants to 20S human proteasome at 37°C of **1** (o, fast phase, x, slow phase, *see ref.* 89 and **2** (+, fast phase, *, slow phase). Figure reproduced from Di Dato A. et al. (**2017**) with permission from the Nature (http://dx.doi.org/10.1038/s41598-017-17008-7).

In fact, from a broader outlook of parameters reported in Table 12, it emerges that binding of porphyrin(s) to the first cluster provoked a shift of the conformational equilibrium from the closed (which is prevalent in the absence of substrate) in favour of the open structure (as displayed by the higher affinity for the open state, see Table 12); differently, binding to the second cluster produces the opposite shift (as depicted by the higher affinity for the closed state, see Table 12), in a manner that completely porphyrin-bound proteasome regain to a "closed" conformation, which is not necessarily the same structure of initial one, differentiated by a limited enzymatic activity.

Table 12 | Kinetic parameters for the reaction of **2** and **1** with the human 20S proteasome.

Kinetic parameters			
2		**1**[a]	
^1k_c (M^{-1}s^{-1})	1.3×10^2	$^1k_{on}$ (M^{-1}s^{-1})	5.6×10^4
^1k_o (M^{-1}s^{-1})	2.8×10^6	$^2k_{on}$ (M^{-1}s^{-1})	1.3×10^4
^1K_c (M^{-1})	7×10^3	^3k (s^{-1})	3.2×10^{-3}
^1K_o (M^{-1})	1.8×10^7		
$^1k_{off}$ (s^{-1})	0.22		
^1L	4×10^{-6}		
^2k_c (M^{-1}s^{-1})	6.8×10^5		
^2k_o (M^{-1}s^{-1})	6.75		
^2K_c (M^{-1})	3.1×10^6		
^2K_o (M^{-1})	2.5×10^3		
$^2k_{off}$ (s^{-1})	2.7×10^{-3}		
^2L	7.3×10^4		

[a]from ref. 133

Importantly, the computational hypothesis resulted in agreement with both literature data and our biochemical data (including the hypothesis to host three binding sites for each α-ring structure). Moreover, our results confirm the hypothesis that gating dynamics are involved in driving conformational movements of the catalytic β-subunits, offering an insight for the design of increasingly refined porphyrin-based allosteric proteasome inhibitors.

4.3 Investigating the Open/Close Conformational Transition of 20S Human Proteasome by Molecular Dynamics Simulations

The allosteric connection between α-ring conformational changes (opening/closing of the gate) and β-ring catalytic activities (including substrate access, enzymatic reaction, and product release), is finely tuned by inter-related molecular interactions which can be triggered "from the outside" (the α-ring) or "from the inside" (catalytic β-subunits).[147] An augmenting number of alternative 20S conformational states, depending on external factors (*i.e.*, the experimental technique, changes in solvent environment, presence of ligands, etc.), have been now characterised. These evidences prove the existence of large-scale conformational transitions regulating 20S functions; however, no information on how such transitions occur at the molecular level is actually available. Affecting the 20S human proteasome conformational equilibrium is the key molecular mechanism of our allosteric porphyrin inhibitors, as well as of other modulators, including 20S RPs, thus, the availability of a molecular model of such a transition could have a dramatic impact on future researches. Regrettably, the gap between the transition and the currently accessible simulation times prevents the employ of direct-unbiased MD approaches to investigate large-scale protein conformational changes. Therefore, a simplification of the system and/or of the performed calculations is necessary, so that less computer time is needed for each simulation step, leading to an artificial increase of the available computational power and a greater sampling of the studied system. On these bases, discrete molecular dynamics (dMD),[41,42] which employs discrete step function potentials in place of the continuous potential used in traditional MD, has emerged as a simplification of traditional MD. During dDM, the simulation engine solves the ballistic equations of motion for only those particles participating in a collision, instead of solving Newton's equations of motion for every particle in the system. Because fewer calculations are performed per time step, the dMD technique allows longer time and length scales to become accessible in the simulation of large biomolecules. Moreover, the use of coarse-grained (CG) protein models[148] extends the computational advantage of dMD, as the simulation of a CG system requires fewer resources and goes faster than that for the same system in all-atom representation. As a result, an increase of orders of magnitude in the simulated time and length scales can be achieved. On

these bases, in order to simulate the opening/closing of 20S human proteasome, I spent eight-months at the Molecular Modelling and Bioinformatics laboratory (Institute for Research in Biomedicine; IRB Barcelona) coordinated by prof. Modesto Orozco, where I applied an innovative dMD approach. In particular, dMD was used in the framework of the GOdMD software (http://mmb.irbbarcelona.org/GOdMD), a new algorithm to trace conformational transitions in macromolecules that was developed by the research group of prof. Orozco.[149]

GOdMD is based on the use of discrete Molecular Dynamics (dMD)[41,42] and biasing techniques based on a combination of essential dynamics and Maxwell-Demon sampling techniques.[150] The whole procedure applied by GOdMD is depicted in the flowchart in Figure 47.
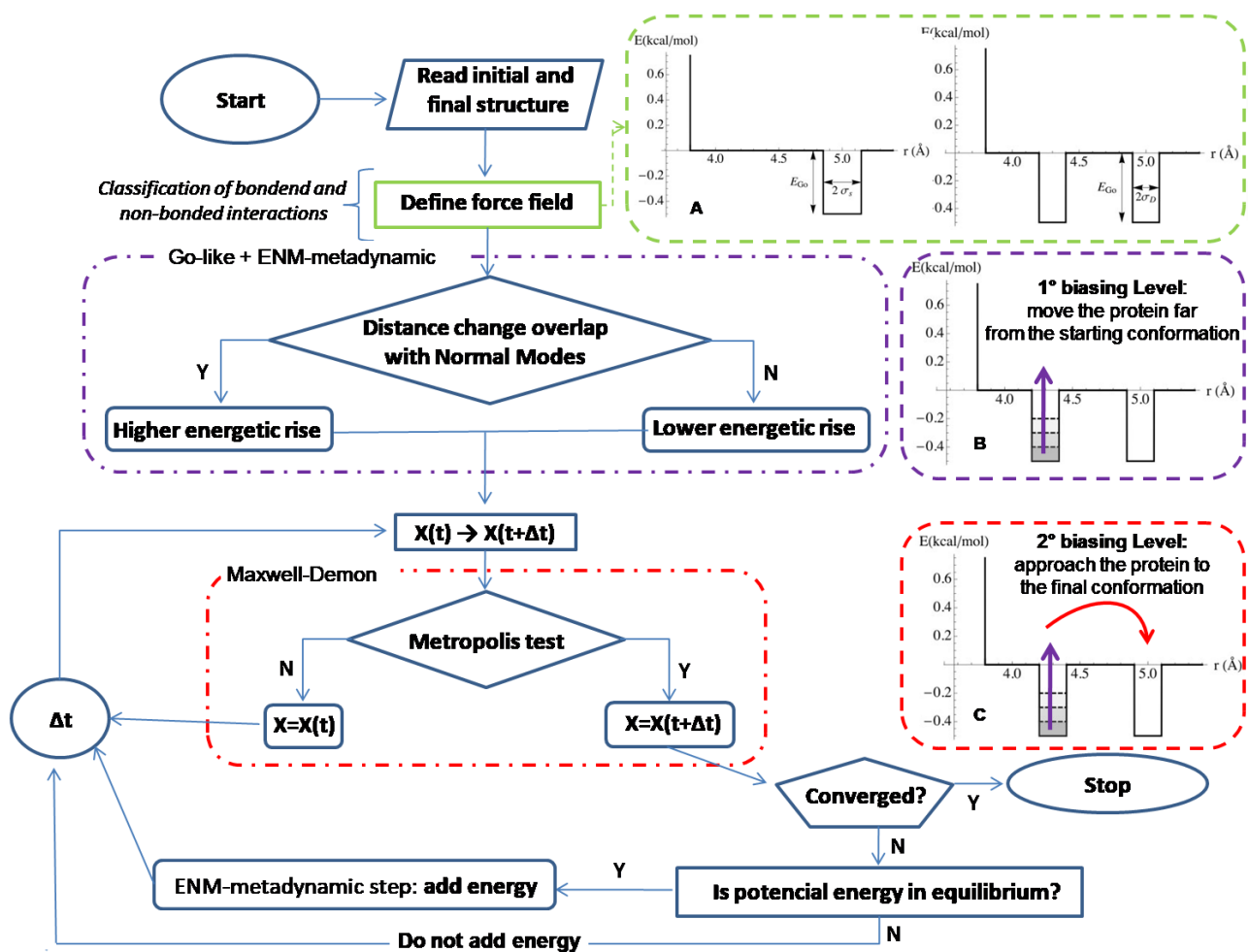

Figure 47. Flowchart of the basic GOdMD algorithm.

In particular, the dMD approach uses two paths to simplification: the first through variation of the interaction potential and the second by a search algorithm. The interactions between atoms are computed by the simplest possible potential function, a *square-well* potential. This facilitated method represents a CG

model of the potential function itself, and quite curtails the number of calculations required. Unlike MM and MD simulations, which are driven by physical forces, dMD simulations are led by collision events. Indeed, the basic dMD formalism supposes that particles move to constant-velocity regime until they collide with another atom. Consequently, the interaction potential between the two atoms differs and, accordingly, the velocities of the two interacting atoms vary immediately. These changes of velocity must be in accordance with the energy and momentum conservation laws. Then, the potential energy is rough out with progressive discontinuous functions of the particle-particle distance instead of continuous functions employed in standard MD. The searching of these collision events gives rise to the second major simplification present in dMD. Instead of searching every possible collisions, the tracking algorithm is circumscribed to a local search; only those atoms inside the same local area are taken into account as having possible collisions. Once collisions are recognised, only the atoms taking part in collision are required to be updated after every event, thereby fewer calculations required per time-step and significantly improving ability of sampling.

GOdMD adopt a CG representation of the protein, in accordance with an *one-bead* model, where only Cαs of the protein are explicitly represented as single spheres and with the entire protein represented by *beads-on-a-string* polymer. The potentials make use of a single square well with infinite walls to define the bonded interactions between consecutive beads. The non-bonded interactions, instead, are characterised by a multiple well Go-like model,[151,152] dependent on the experimental inter-particle distance and classified in two categories based in presenting or not a different inter-particle distance between the initial (A) and final (B) conformation. Indeed, when both structures shows a distinct inter-particle distance, a double well potential is defined (Figure 47A), centred in the related experimental values (rA and rB). When, on the contrary, the distances are similar in both conformations, a single and ample well (50% larger) is defined, centred at ½ (rA + rB) (Figure 47A). This scheme allows completing transition paths since target minimum behaves as attractor and makes possible to use softer biasing schemes. To avoid the over-restraining of the system with physically irrelevant interactions, the potential energy was defined only for those particles within a cutoff (rA, rB > radii of gyration/2). Moreover, the height of the wells for all non-bonded interactions is 0.5 kcal/mol, permitting the particles to dash off from the wells if necessary.

In order to guarantee an enrichment of sampling along the transition path, two biasing schemes are used by GOdMD algorithm. In particular, the first level of biasing is delineated to move the protein far away from the starting conformation (Figure 47B), by combining metadynamics[38] with Elastic Network Model (ENM),[153,154] in such a way that:

- the wells associated to inter-particle distances showing large changes along the first five essential deformation modes are filled faster;

- the wells associated to inter-particle distances that are not coupled to the essential deformation modes of the protein are filled slowly.

The ENM analysis is performed only on the initial structure irrespective of the target structure, avoiding the production of too linear pathways. Moreover, the ENM-metadynamics guarantees to explore alternative ways to run away from the original structure minima, expanding the possibility to sample favoured pathways as described by the essential deformation modes.

The second level of biasing is designed, instead, to address the trajectory toward the final structure, without inserting energy restraints (Figure 47C). For this purpose, a Maxwell Demon biasing algorithm was used, which increases the sampling in the transition direction and along the essential deformation modes. Following the Maxwell-Demon approach, the bias toward the target structure is defined by the calculation of a progress variable correlated to RMSd value. Accordingly, after at a certain simulation step (t), a progress variable Γ (dependent on the total number of residues, target structure, the sampled conformation and a weight function dependent on the inter-particle distance and the size of protein) is calculated and compared with that obtained in a previous accepted movement (t + Δt). The simulation frame (t + Δt → t) is then accepted or rejected based on a Metropolis test, where acceptance probability is a function of the RMSd between the structure sampled at time X(t) and the target structure, including also a variable that is adjusted to guarantee an acceptation rate of 70% and the time frame (Δt) is typically 100 time steps. These large acceptance rates, combined with a weight function having a maximum at ~15Å, enables the system not to explore direct pathways that could lead to stressed structures and provides the best local geometries.

If the simulation frame is rejected, the structure sampled at time X(t) continues the simulation and it is redirected to the Metropolis test again. If the simulation frame is accepted, the convergence value is verified. Considering that the protein structures are in constant moving by thermal noise, it would be unrealistic to obtain a zero RMSd value when a protein structure is intrinsically oscillating at room temperature. On this basis, a convergence principle based on the slope of RMSd to target structure respect to time has been adopted. This criterion averts to obtain stressed structures that would be only an untrue minimum as a result of the strong bias necessary to reach close to zero RMSd values. On the other hand, if this criterion is not met, the algorithm reassess the potential energy and add or not energy based on Go-like scheme described above, subjecting the structure sampled at simulation steps and, then, at Metropolis test.

In order to trace the transition between known equilibrium conformations of 20S human proteasome, our previously developed[129,155] molecular models of human 20S proteasome, in the closed and in the open state, were taken as starting and target structures. To perform the GOdMD calculation both structures should

contain the same number of residues and ordering should match the appropriate sequence alignment, thus, only two packed rings (*i.e.*, one α-ring and one β-ring) were considered for each structure of 20S human proteasome (it has to be herein recalled that, being 20S structured as a dimer, the other two missing rings are identical). All trajectories were generated with random seed number, applying a temperature of 300 K and a simulation time of 5000,000 ps.

Overall, 80 trajectories were obtained, considering the open → closed and closed → open transitions for both 20S wild type and the α-3ΔN mutant (lacking the first N-terminal residues of the α3-subunit). The most representative structures have been selected by evaluating the RMSd value. The conformational changes spanned by these simulations was analysed using Root-Mean-Square Fluctuations (RMSF).

Firstly, the closed → open transition of 20S human proteasome was simulated and twenty trajectories were generated. At the end of the simulation the trajectory with the lowest overall RMSd value was selected as the best simulated transition (FigureV). Interestingly, according to the selected trajectory, all the generated complexes presented a linear decrease of the RMSd value to the target structure (Figure 48).

Trajectory	Cas RMSd value (Å)
1	1.50
2	1.48
3	1.41
4	1.44
5	1.50
6	1.48
7	1.50
8	1.46
9	1.47
10	1.42
11	1.47
12	1.34
13	1.24
14	1.49
15	1.50
16	1.50
17	1.50
18	1.49
19	1.49
20	1.49

Figura 48. RMSd values of twenty generated trajectories with random seed numbers and Plot of RMSd for selected trajectory (highlighted in red).

The dMD simulation shows a concerted opening movement involving α and β subunits, allowing α3 N-terminal tail, which deletion leads to an opening of the channel, to get out of the gate channel. On the

other hand, since α2, α3, and α4 N-terminal tails are involved in molecular interactions with each other, then, the conformational change of α3 N-terminal tail affects also the position of α2 N-terminal tail. This latter shows a marked readiness to switch spatial positioning in absence of α3 N-terminal tail and, in accordance with the intimate connection among the N-terminal portions at the access of gate, the α2 N-terminal tail residues (aa1-8) show a partial (not complete) movement in adopting the positioning towards open conformation as well as α4 N-terminal tail residues (aa1-7) (Figure 49A). Moreover, the RMSd values between the last frame of the considered trajectory and the target structure (open conformation of the 20S human proteasome) showed that also some residues at the β2-β3 interface (T-like catalytic activity) are not able to reach their target position (aa136-146 β3) (Figure 49B).

Figure 49. Overall view of the 20S human proteasome structure (α and β rings). The Cαs are explicitly represented as single spheres and coloured by RMSF. Red values indicate areas of less fluctuation (RMSF < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å), and blue values indicate areas of more fluctuation (> 7 Å).

This aspect highlights the existing connection between α and β rings, accounting for the experimental evidences about 20S allostery.[155,136,147]

In addition, Root-Mean-Square Fluctuations (RMSF) were calculated in order to obtain a measure of the average atomic mobility of Cαs atoms during the simulation and identify structure local changes (Figure 50).
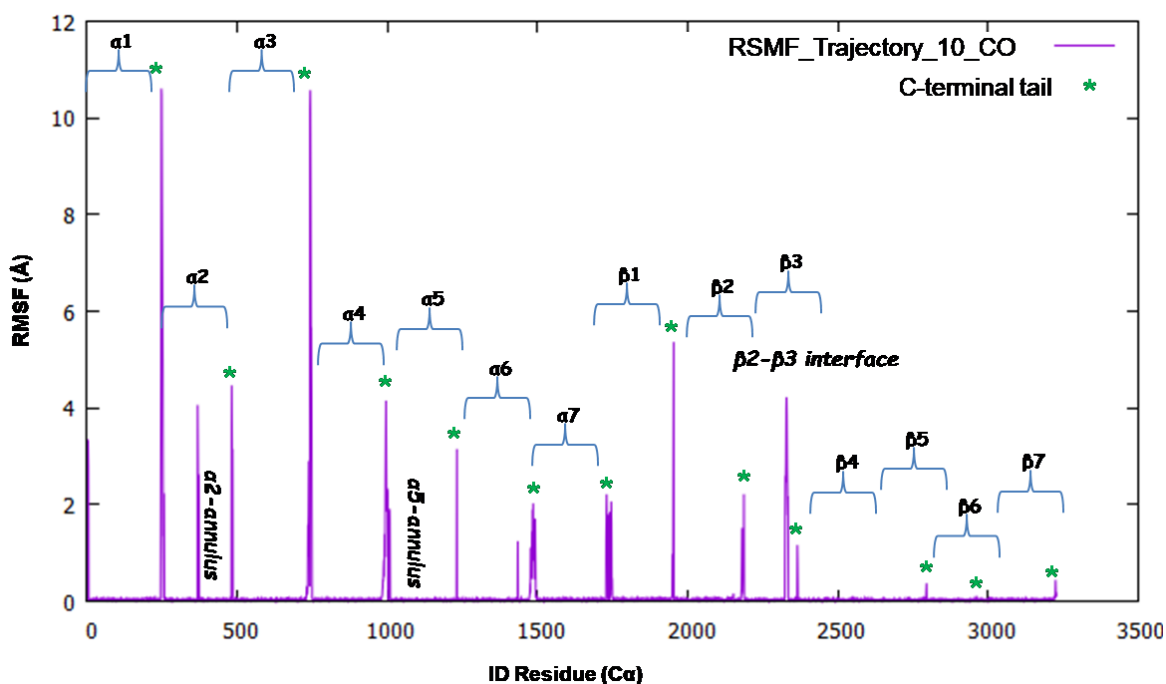
Figure 50. Plot of RMSF for selected trajectorie *vs.* 20S human proteasome target structure (open conformation).

Excluding some C-terminal tails (reported with green asterisks), which correspond to disordered and flexible protein regions, the graph evidenced as the most pronounced local changes those occurring at α5-subunit and β2-β3 interface (S3 catalytic site pocket). In particular, regarding the affected α5 structural region, it matches to loop forming part of the so called *a-annulus*, that is, a narrow channel that is located slightly below the surface of the α ring. Substrates enter the proteasome through a gated pore in the centre of α-ring[156] and their passage is restricted by the *a-annulus*, which is in direct connection with the β-ring and ensures that substrates are substantially unfolded before they can enter the catalytic site. Accordingly, our results showed conformational changes transmitted from the α-ring (α5) to the interface with the β-ring, and extending to the S3 active site pocket of the T-like catalytic activity (β2-β3 subunits). Importantly, NMR studies evidenced that the same 20S structural regions are involved in the regulation of proteasome activation.[136,142]

Then, the opposite conformational transition, *i.e.*, open → closed state, was investigated applying the same simulation procedure. Obtained results have showed a value of 1.49 Å for the trajectories with the lowest RMSd and the trajectory 20 was selected as the most representative.(Figure 51)

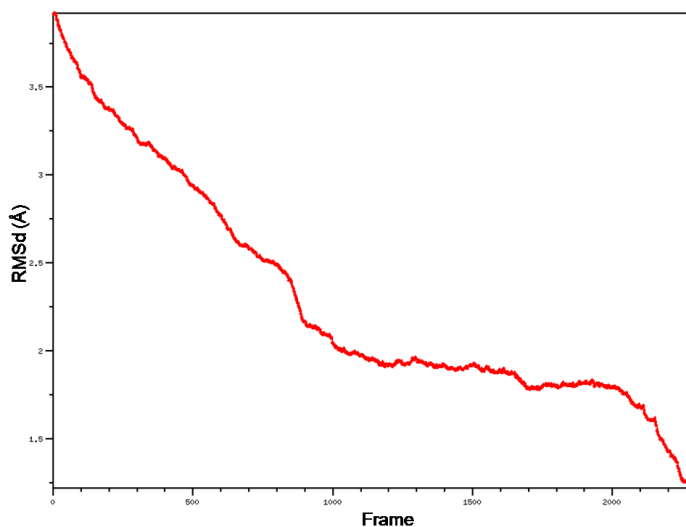| Trajectory	Cαs RMSd value (Å)
1	1.79
2	1.49
3	1.83
4	1.50
5	1.50
6	1.49
7	1.85
8	1.50
9	1.50
10	1.50
11	1.49
12	1.88
13	1.84
14	1.50
15	1.85
16	1.50
17	1.50
18	1.50
19	1.80
20	1.49

Figura 51. RMSd values of twenty generated trajectories with random seed numbers and Plot of RMSd for selected trajectory (highlighted in red).

The selected transition confirms the coordinated conformational changes of the α and β rings, as well as, a trend toward the fitting to the target structure of α2 and α4 N-terminal tails (Figures 52 and 53). Differently, the α3 N-terminal tail appears more constrained, underlining its structural role as clog of the proteasome central pore.
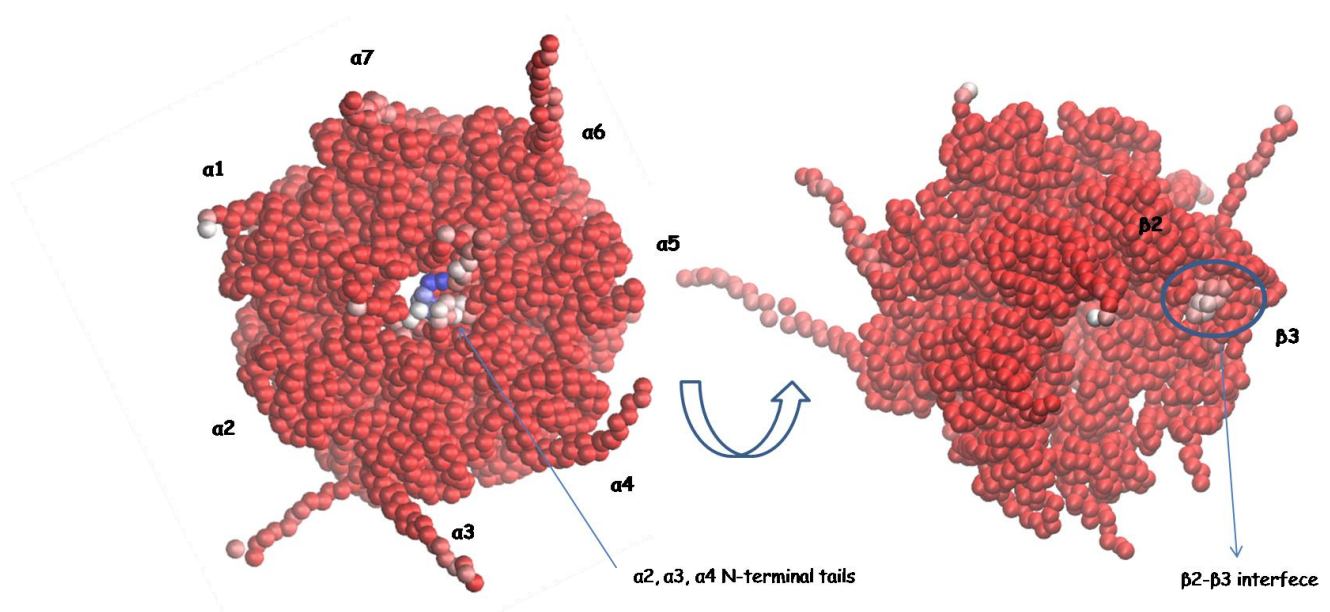
Figure 52. Top vision of the 20S human proteasome structure (α and β rings); The Cαs are explicitly represented as single spheres and coloured by RMSF. Red values indicate areas of less fluctuation (RMSF < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å),, and blue values indicate areas of more fluctuation (> 7 Å).

Figure 53. Plot of RMSF for selected trajectorie *vs.* 20S human proteasome target structure (closed conformation).

Our results indicate that 20S catalytic activities are auto-inhibited by the outer α-ring subunits, with a key role played by the N-terminal tail of the α3-subunit. In the latent state, it prevents substrate entry by

imposing topological closure. Importantly, biochemical and structural analysis showed that deletion of the N-terminal tail of the α3-subunit (α-3ΔN mutant) leaves the channel permanently open to the proteolytically active interior chambers.[145] On the other hand, the correct repositioning of the α3-subunit N-terminal tail in to the gate is not achieved in the open → closed transition simulation, suggesting that an additional intermediate conformational state is probably involved, which may explain as the mechanism of the gate opening is finely regulated by a sequence of perfectly coordinated events.

Thus, to further investigate the structural role of the N-terminal tail of the α3-subunit on 20S closed/open equilibrium, the structure of the α-3ΔN mutant was built by the deletion of nine residue (^{1}MSRRYDSRT^9) from the N-terminal tail in the two 20S human proteasome structures (closed and open states). The resulting models were subjected to the same dMD procedure previously applied to the wild type structure. The best trajectory obtained for the α-3ΔN mutant in the simulation of the closed → open transition show an RMSd value of 1.37 Å (Figure 54).

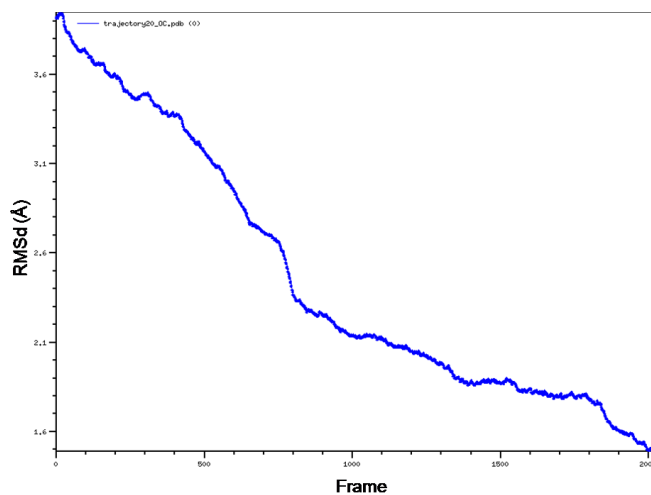| Trajectory	C$_{\alpha s}$ RMSd value (Å)
1	1.50
2	1.50
3	1.49
4	1.50
5	1.48
6	1.50
7	1.37
8	1.41
9	1.45
10	1.50
11	1.46
12	1.44
13	1.43
14	1.50
15	1.48
16	1.50
17	1.45
18	1.42
19	1.49
20	1.50
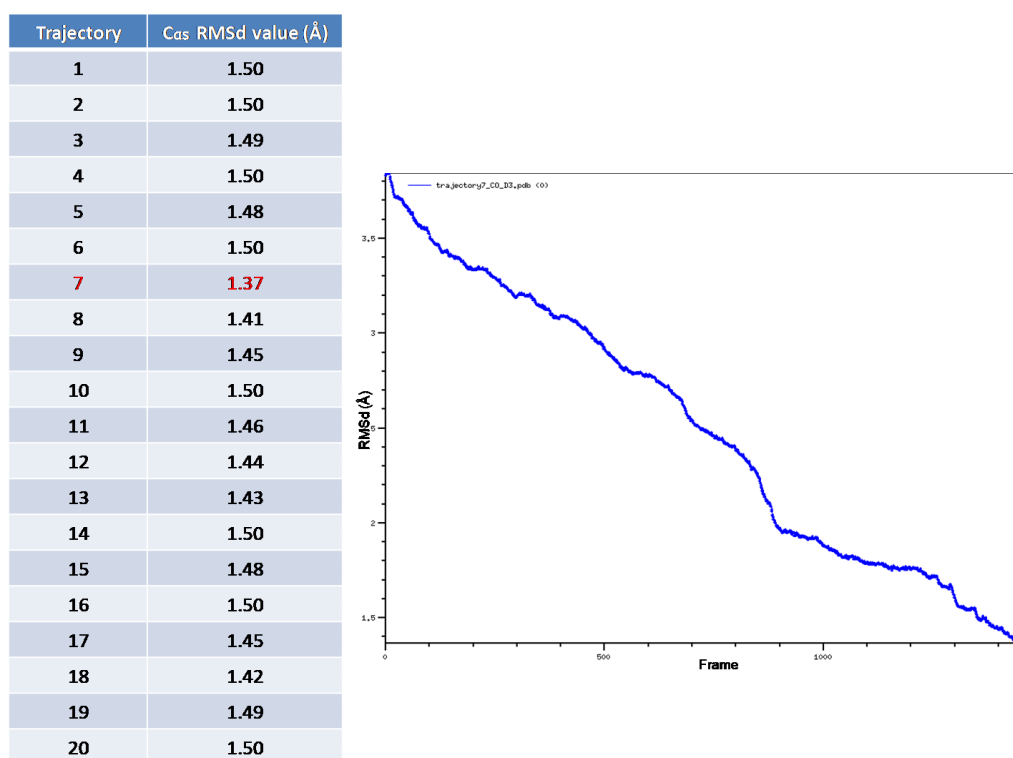
Figura 54. RMSd values of twenty generated trajectories with random seed numbers and Plot of RMSd for selected trajectory (highlighted in red).

According to this simulation, the α4 N-terminal tail of the mutant gets out of the proteasome gate in place of the α2 N-terminal tail, which instead remain deeply into the central pore. This behaviour, once again,

emphasises the close relationship between α2 and α3 N-terminal tails. Indeed, the lack of α3 N-terminal tail does not allow α2 N-terminal tail to detach from central gate (Figures 55 and 56).
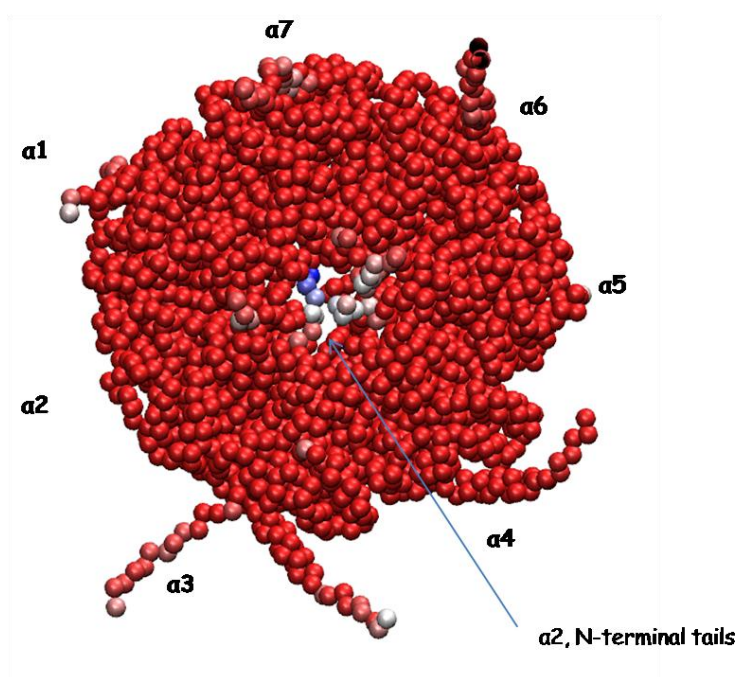
Figure 55. Top vision of the 20S human proteasome structure (α and β rings); The Cαs are explicitly represented as single spheres and coloured by RMSF. Red values indicate areas of less fluctuation (RMSF < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å), and blue values indicate areas of more fluctuation (> 7 Å).
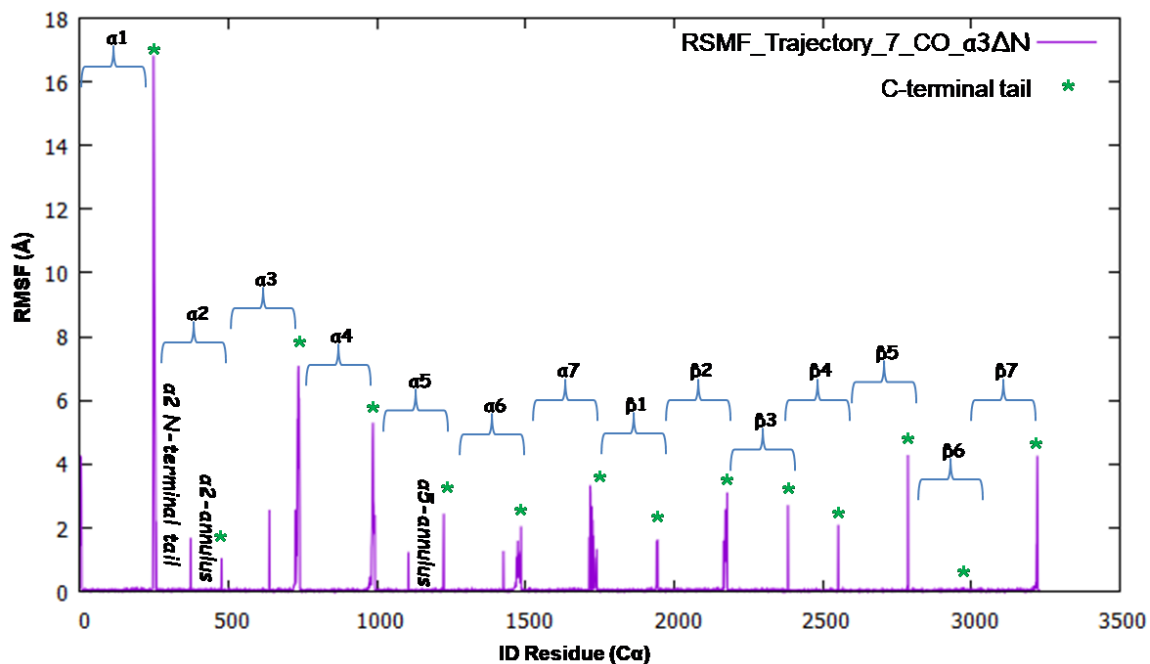
Figure 56. Plot of RMSF for selected trajectorie *vs.* 20S human proteasome target structure of α-3ΔN mutant (open conformation).

On the other hand, the best trajectory obtained during the simulation of the open → closed transition exhibits a RMSd value of 1.34 Å (Figure 57) and the same tendency for the α2 and α4 N-terminal tails to correctly adopt the positioning of the target (closed) conformation, as observed for the wild type structure (Figures 58 and 59)

Trajectory	Cαs RMSD value (Å)
1	1.50
2	1.50
3	1.48
4	1.46
5	1.42
6	1.50
7	1.49
8	1.83
9	1.49
10	1.34
11	1.49
12	1.46
13	1.42
14	1.42
15	1.50
16	1.35
17	1.38
18	1.49
19	1.41
20	1.43
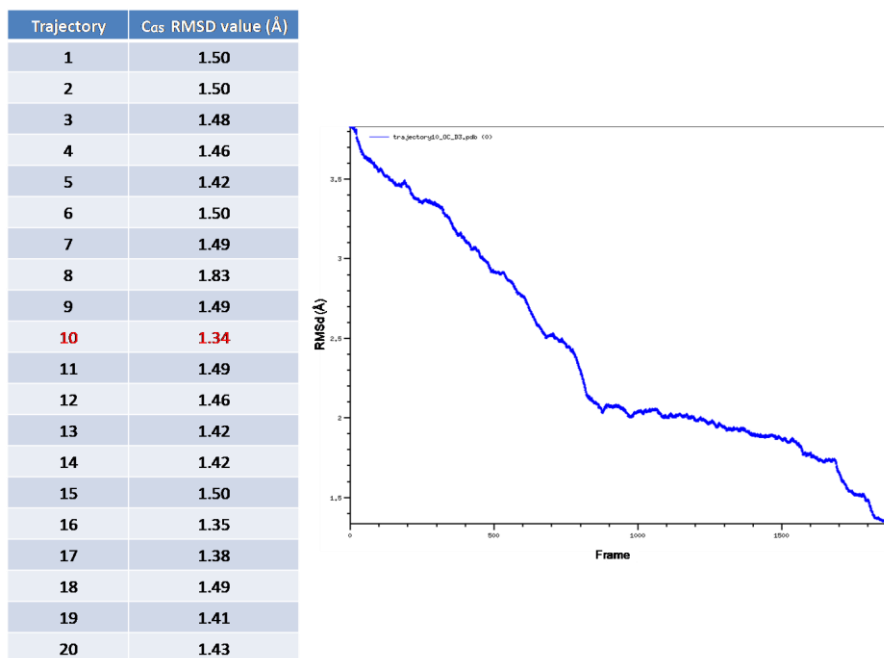
Figura 57. RMSd values of twenty generated trajectories with random seed numbers and Plot of RMSd for selected trajectory (highlighted in red).
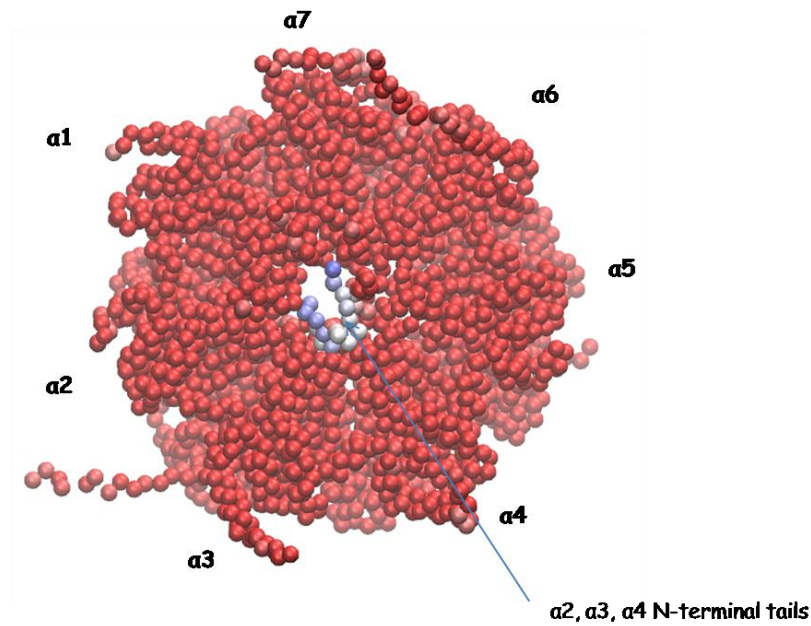

Figure 58. Top vision of the 20S human proteasome structure (α and β rings); The Cαs are explicitly represented as single spheres and coloured by RMSF. Red values indicate areas of less fluctuation (RMSF < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å), and blue values indicate areas of more fluctuation (> 7 Å).
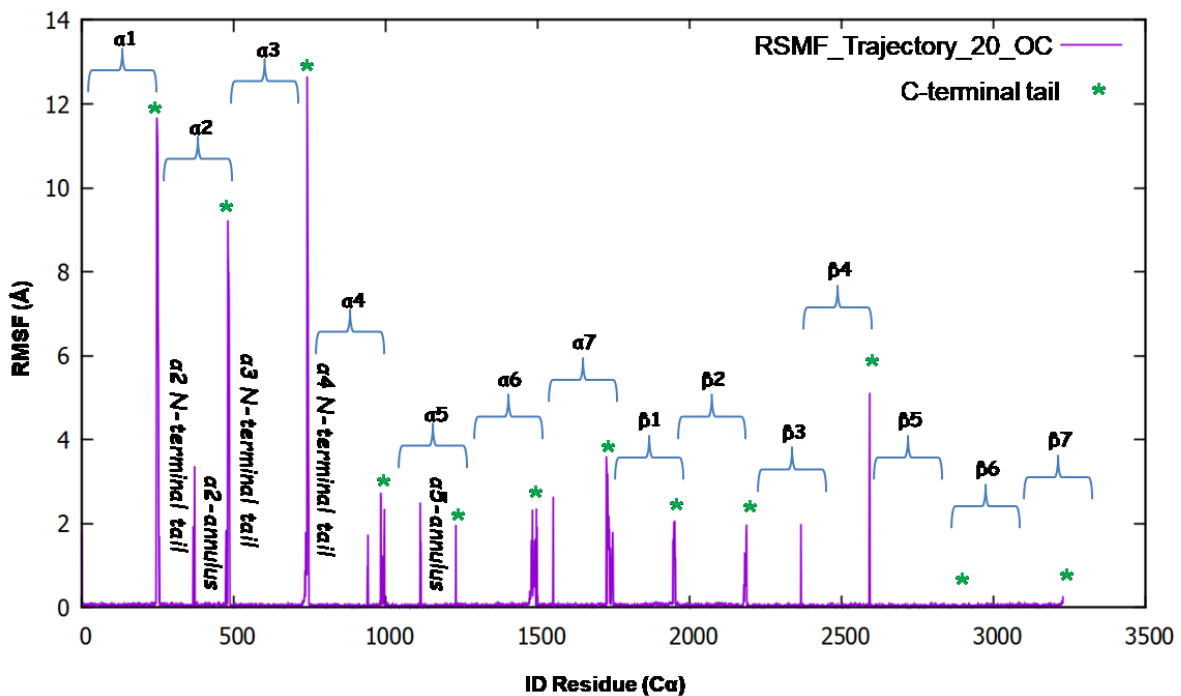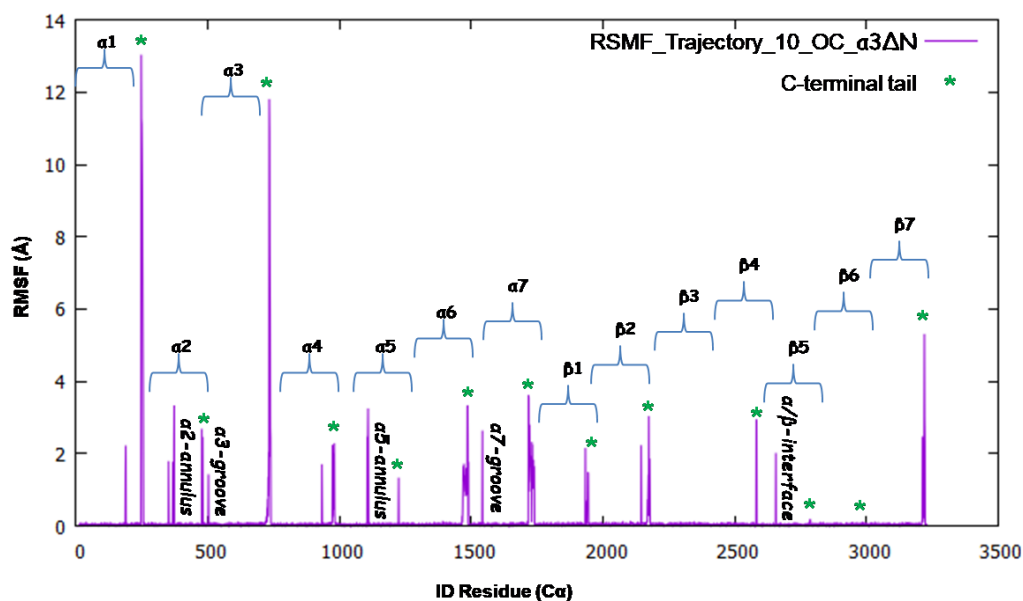
Figure 59. Plot of RMSF for selected trajectorie *vs.* 20S human proteasome target structure of α-3ΔN mutant (closed conformation).

These studies provided the first molecular model accounting for the key role played by the α3 N-terminal tail in the regulation of the 20S conformational equilibrium.

4.3.1 Full atom simulation

MD simulations on full atom structures of 20S human proteasome, either in the closed and in the open state, have been also performed in order to explore essential structure fluctuations in both conformational states. All simulations were performed with the GROMACS 5.0 software package.[157] The flowchart below describes the general procedure for setting up an MD run. (Figure 60).
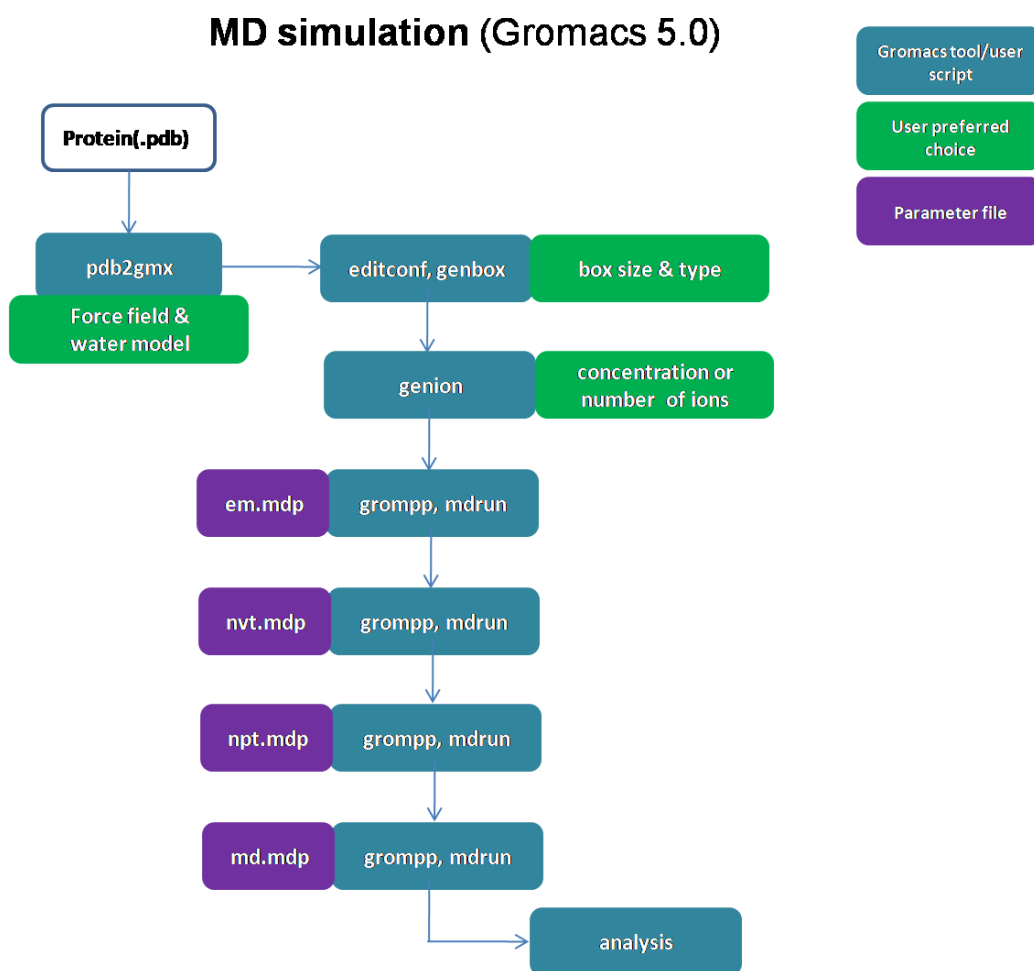
Figure 60. Flowchart of Gromacs procedure.

The Amber ff99SB-ILDN force field[158] was used, in which side chain torsion parameters of four residues (isoleucine, leucine, aspartate, and asparagine) have been improved. The LINCS[159] algorithm was used to constrain bond lengths, allowing a time step of 2 fs.

Following the flowchart in Figure T, the initial structure coordinate file is converted into a GROMACS readable format (pdb2gmx), generating two files that contain the molecular structure and the topology file (*i.e.*, molecular parameters, force field and charges). Then, a *dodecahedron box* is generated around the protein with a buffer of 1.4 nm between the outside of the protein and the edge of the box (editconf). Moreover, a water box, based upon the dimensions/box type, is created and the number of water molecules needed to solvate the box of the given dimensions (genbox) have been added with the TIP3P (transferable intermolecular potential 3P) model.[160] Subsequently, it was necessary to neutralize the system prior to simulations and, for this reason, counter ions have been added according to the total charge of the system (genion). At this point, the protein has been subjected to energy minimisation using the Steepest Descent algorithm (etmol = 1 and emstep = 0.01). In order to rearrange the water molecules towards protein,

before the starting of MD simulation, the backbone restraints on whole protein (except for the modelled N- and C-terminal portions) were applied using the 'define = -DPOSRES' line command. The MD simulation procedure starts with equilibration, carried out in two steps. First, an NVT (constant Number of atoms, Volume, and Temperature) simulation is performed in order to bring the system to the final temperature of 300K. The parameter file for the NVT simulation calls for an MD run of 1000000 steps with a 0.02 ps time-step (a total of 2000 ps = 2 ns). Temperature coupling has been performed using the V-rescale method and simulated annealing procedure. In order to be sure velocities (and therefore temperature) are evenly distributed across water and protein, protein and water atoms are coupled to separate baths (separate coupling).

After the equilibration to a target temperature, the system is next relaxed into a constant pressure ensemble. Indeed, a NPT (constant Number of atoms, Pressure, and Temperature) simulation is carried out to allow the system to find the correct density. In particular, five NPT equilibration steps each of 0.5 ns have been performed (a total of 2.5 ns), decreasing gradually the constant forces of position restraints as follows: step1/9000 kJ mol^{-1} nm^{-2}, step2/1000 kJ mol^{-1} nm^{-2}, step3/100 kJ mol^{-1} nm^{-2}, step4/10 kJ mol^{-1} nm^{-2}, and step5 without restraints. The V-rescale thermostat has been preserved and the Berendsen barostat[161] has been used to couple pressure isotropically (the same in all directions) to a value of 1.0 bar (from step1 to step3). Then, a Parrinello-Rahman barostat[162] allowed a variable cell shape during the simulation (from step4 to step5). Finally, being this run a continuation of the previous run, "gen_vel" has been set to "no" and "continuation" has been set to "yes". Once the system was properly equilibrated, the parameter file for MD production looks almost exactly like that of the last NPT equilibration step (step5) for a 10,000,000 MD steps with a 2 fs time-step, totaling 20,000 ps (or 20 ns) of sampling. Subsequently, such MD run was extended for further 25,000,000 MD steps (50,000 ps = 50 ns) for a total time of 70 ns.

It should be underlined that a very large number of amino acid residue composes the 20S human proteasome structures, by consequence, the MD calculations requested an extensive computational time despite a multiple parallelization scheme (*i.e*, 96 processors used).

The Principal Component analysis (PCA) has been used in order to extract the most important elements in the obtained data, using a covariance matrix constructed from atomic coordinates that describe the accessible number of degrees of freedom of the protein, such as the cartesian coordinates that define atomic displacements in each conformation comprising a trajectory.[163] Thus, the "essential" motions were extracted from the set of sampled conformations and the obtained eigenvectors of a covariance matrix has been analysed. RMSF values were calculated in order to quantify the essential dynamics of the 20S human proteasome. The obtained results showed interesting details. Both the α-ring and β-ring appear to be more

susceptible to structural fluctuations when the 20S human proteasome is in the closed conformation (Figure 61).
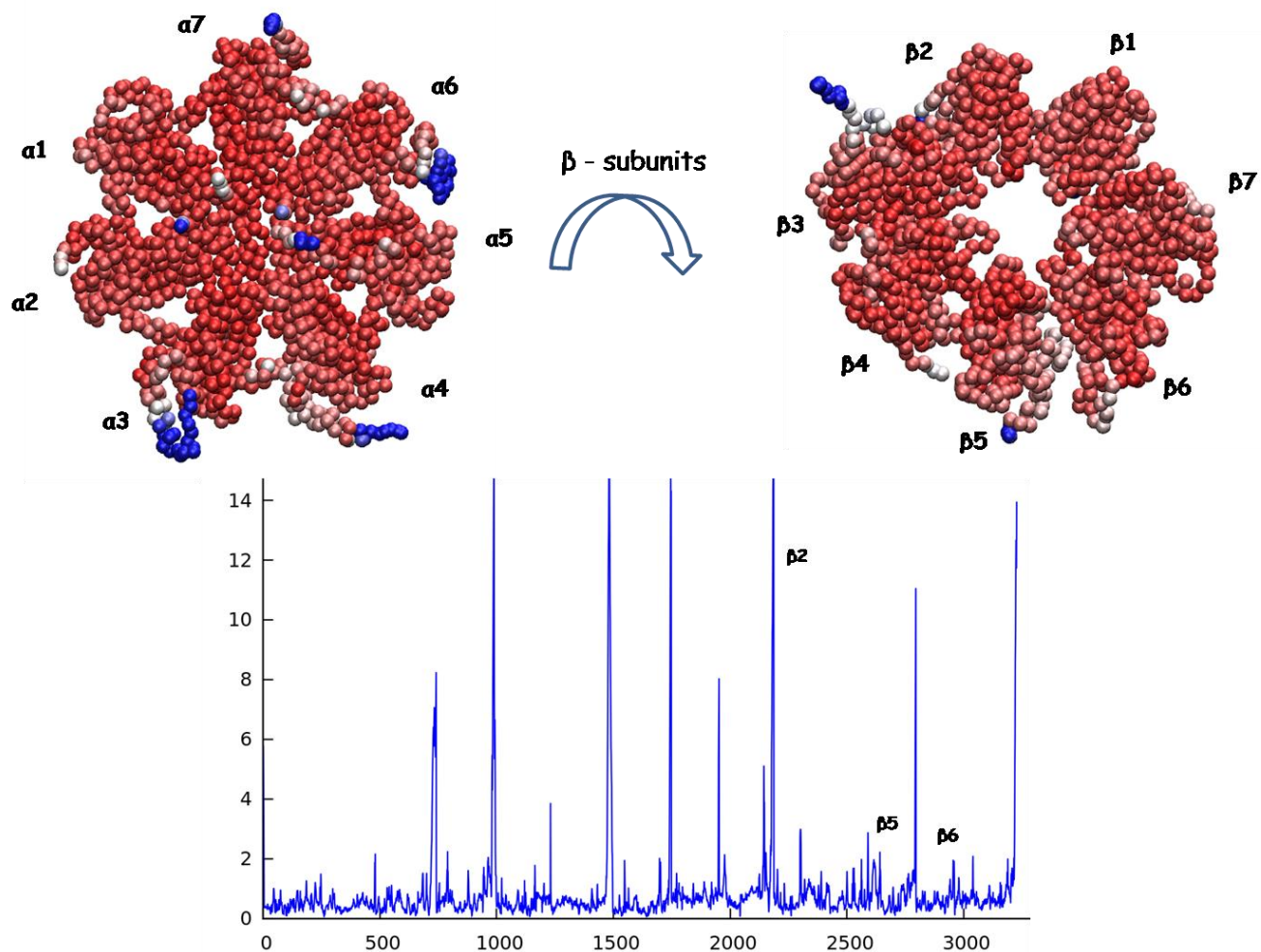
Figure 61. Top vision of the 20S human proteasome structure (α and β rings); the Cαs are explicitly represented as single spheres and coloured by PCA RMSF. Red values indicate areas of less fluctuation (RMSF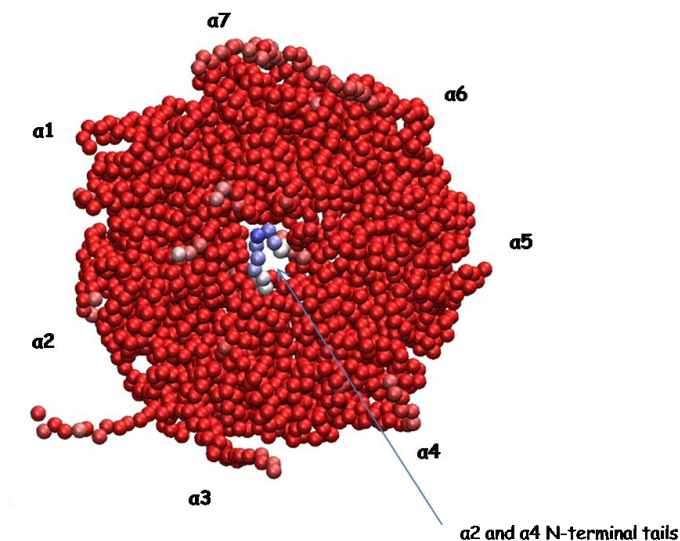 < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å), and blue values indicate areas of more fluctuation (> 7 Å). Plot of PCA RMSF for 20S human proteasome (α and β rings) for 70 ns of simulation.

In particular, the α-ring showed most conformational changes at the grooves (and C-terminal tails) level while the N-terminal tails kept their initial positions, confirming their role as gate blocker. On the other hand, the β-ring showed conformational changes at catalytic subunits level and more precisely at the β5-β6 subunit interface. Indeed, the β6-loop looks as if it opens up in order to allow the entrance and/or release of the substrate. Moreover, the β2 C-terminal tail was the only to show a significant RMSF value, confirming highly dynamic and eager to act as structural communication between the α and β rings. Indeed, during the simulation, the long β2 C-terminal tail makes contacts with the α2 and α3 subunits,

notably, its positioning has never been experimentally determined and these results suggest that it could represent the structural portion designated to the allosteric communication between α and β rings.

The same study has been performed also considering the 20S human proteasome in the open conformation (Figure 62).
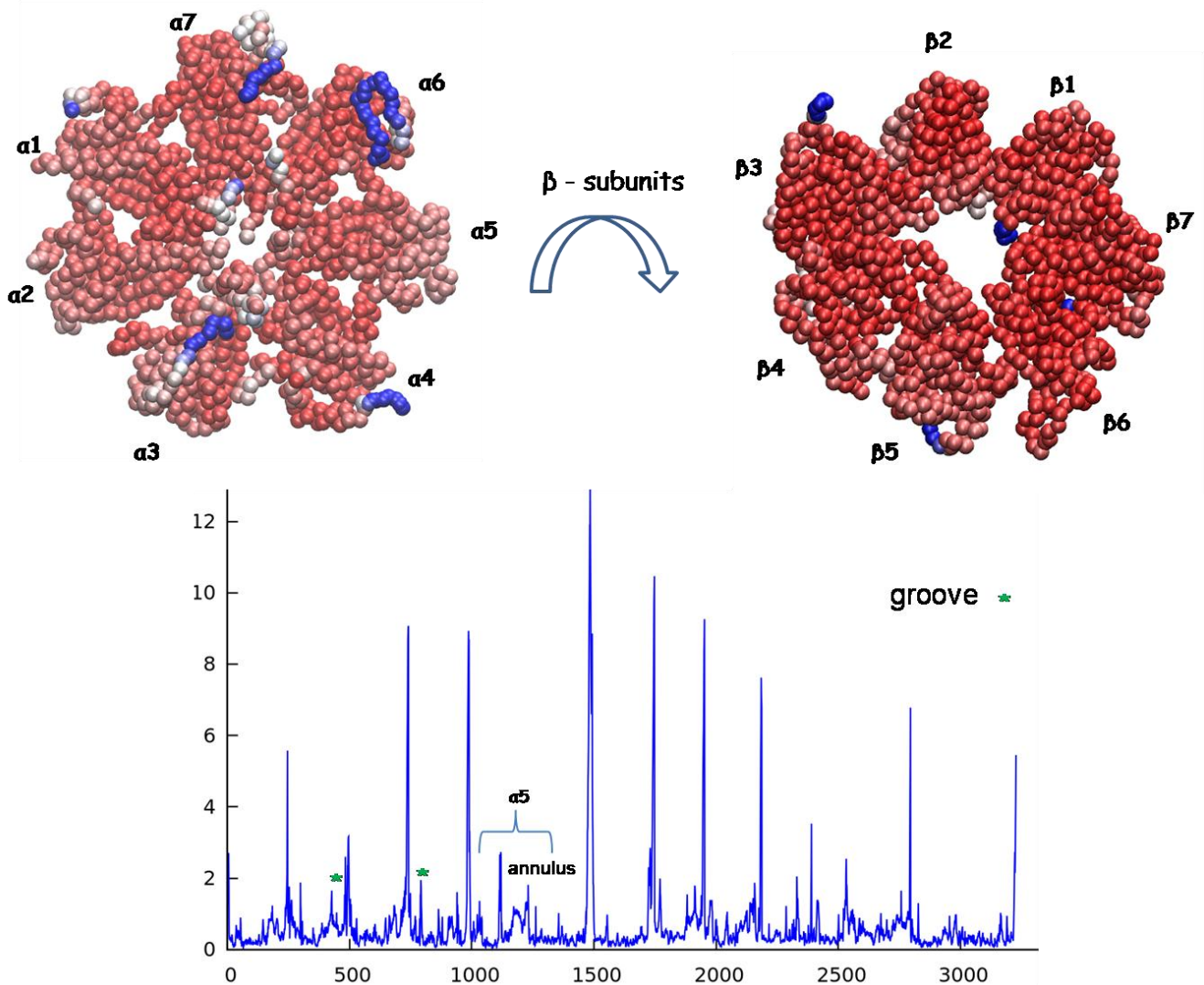
Figure 62. Top vision of the 20S human proteasome structure (α and β rings); The Cαs are explicitly represented as single spheres and coloured by PCA RMSF. Red values indicate areas of less fluctuation (RMSF < 1 Å), white indicate areas of medium fluctuation (1 Å < RMSF < 7 Å), and blue values indicate areas of more fluctuation (> 7 Å). Plot of PCA RMSF for 20S human proteasome (α and β rings) for 70 ns of simulation.

Differently from the previous results, the α-ring showed large conformational fluctuations in several protein regions; in fact, the α1-α2 and α3-α4 grooves showed a higher RMSF value with respect to other α-subunit interfaces. Moreover, the α1 and α5-α7 N-terminal tails tend to adopt the closed conformation while those of α2, α3, and α4 are consistently in the open conformation, in accordance with their intimate

structural relationship, as described above. In addition, two peculiar aspects that could be of interest for the mechanism of gate closing/opening, also emerged from these results. In particular: i) the N-terminal tails showed a co-operative movement with the α-*annulus* and groove regions, ii) the α3 and α5 N-terminal tails are involved in molecular interactions with the α3 and α7 C-terminal tails, respectively. This could account for the auto-regulation of 20S, which it is capable to adopt a closed/open conformation also in the absence of RPs. Unfortunately, the lack of accurate information about the biological and structural role of the 20S C-terminal tails allows only to formulate hypotheses, so far.

Regarding the β-ring, the results show that conformational changes occur at β2 C-terminal tail and β5-β6 interface, congruent with those observed in the closed conformation, although highlighting a more marked movement of the β2-subunit.

Finally, by comparing the RMSF values of the closed and open conformation, significant differences in local fluctuations can be observed (Figure 63).
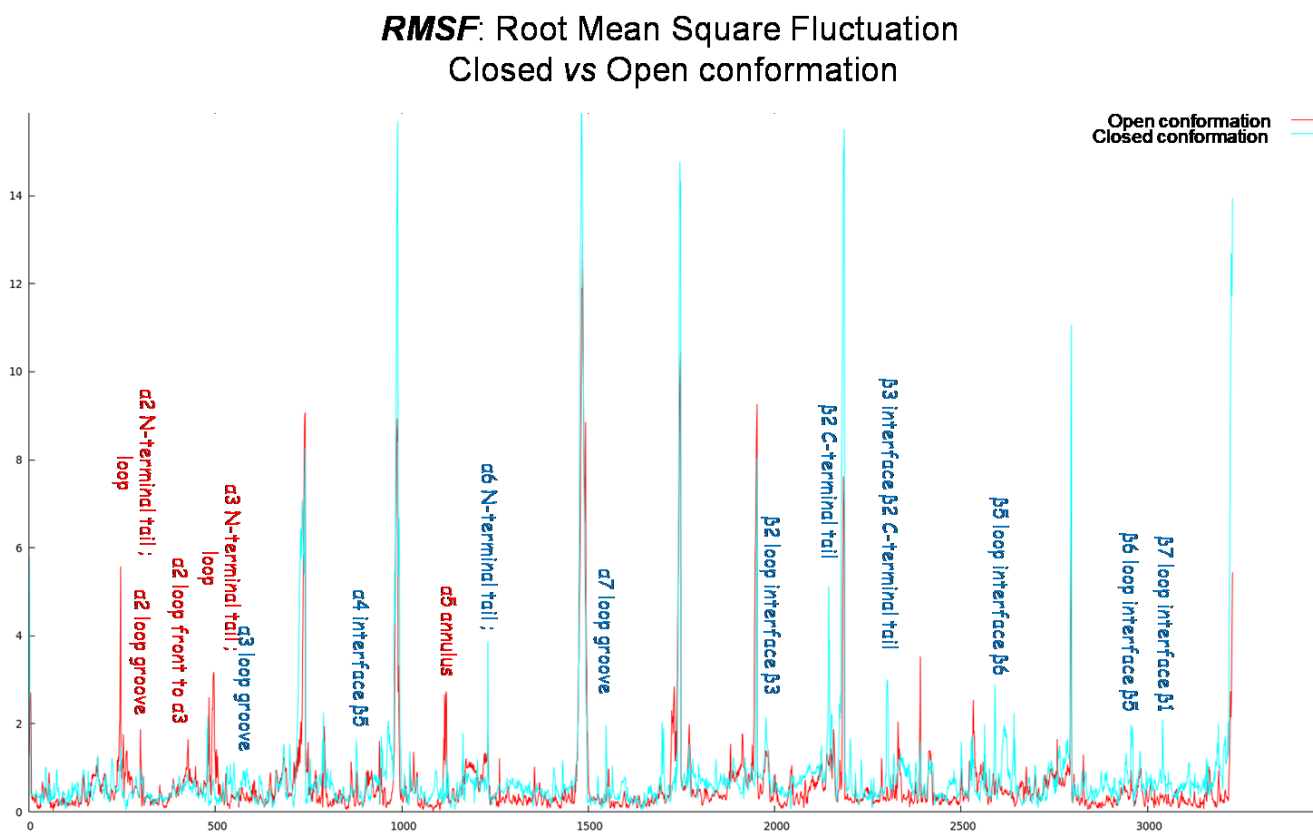
Figure 63. PCA RMSF plot of the backbone atoms of 20S human proteasome *vs.* amino acids number. The closed conformation is shown in cyan; the open conformation is shown in red.

In fact, the closed conformation of 20S human proteasome showed greater fluctuations during the MD simulation at the β-ring and other specific regions, such as:

- α3-subunit: structural region belonging to α3-α4 groove;

- α4-subunit: α-helix in connection with β5-subunit;

- α7-subunit: structural region belonging to α6-α7 groove.

Differently, the open conformation of 20S human proteasome showed major conformation changes at α-ring level, such as:

- α2-subunit: structural loop belonging to annulus region and structural region belonging to α1-α2 groove.

- α3-subunit: structural loop belonging to annulus region and structural loop belonging to α3-α4 groove;

- α5-subunit: structural loop belonging to annulus region.

In conclusion, conventional molecular dynamics simulations and discrete molecular dynamics simulations were performed to investigate the structural characteristics for the open and closed 20S CP states and the dynamic transition mechanism of the two states, respectively. The results of this study show that the predominant structural differences between the open and closed 20S states were localised at gate, grooves and α-β interface, underlining the allosteric network occurs from α-ring to the β-ring via the N-terminal tails and annulus region. In fact, the structural regions involved in the conformational changes were previously proved[136] to be involved in the allosteric regulation of the 20S catalytic activity. Taken together, this observations represent a further confirmation of obtained dynamic docking results and provides useful insights into understanding the dynamics of the transition mechanism of the closed 20S CP state to the open 20S state. In all these contexts, the availability of chemical tools that allow the control and modulation of the functional dynamics of proteins (and their networks) can have fundamental impact on physiology studies and the development of novel mechanism-based therapeutics.

Chapter 5

5.1 Design of Redox-active Chemotherapeutics

Natural compounds including redox-active chemical function(s) are frequently cytotoxic and only some of them show a profile of selectivity appropriate to become structural leads for the development of new chemotherapeutic agents. A perfect redox-active cytotoxic agent should behave as a "sleeping bomb" which is activated only in a precise cellular environment (*e.g.*, cancer, microbial, or infected cells).[164,165] However, the molecular mechanisms of such selective cytotoxicity are even now not well-know. This is caused by the intricacy of the chemical reactions taking place in biological systems[165] and by the difficulty to precisely predict the molecular reactivity of the ligands in such environments.[166-168] In addition, in many cases, these compounds do not present a well-defined protein target.[164] Therefore, the chemical reaction driving to their bio-activation represents the key molecular interaction to be examined. When a series of active analogues is available but the protein target is unknown or not specific, only ligand-based approaches can be used to address hit identification/optimisation. One of the most common application is the development of a pharmacophore model based on the SARs of available ligands.[169] Considering all putative mechanism of action, the in-depth analysis of the conformational behaviour of the considered ligands is a key point for the identification of the features responsible for the identified biological activity. Indeed, the molecular recognition is driven by the conformational equilibria of the interacting species present in solution. Obviously, this is true also when the mechanism of action pertains to a redox reaction. The majority of the conformations that a ligand can hypothetically acquire are energetically not allowed and are not observed experimentally. The difficult task for a computational chemist is to comprehend the structural and energetic dissimilarity among these conformations, recognising the prevalent and the bioactive ones (which may show a higher energy with respect to their global minima).[170,171] In the case of redox-active ligands, the bioactive conformations are those capable to interact with the redox counterpart (*e.g.*, radical species, metal ions, etc.) accountable for the bio-activation reaction permitting simultaneously the appropriate spatial orientation of all reaction partners. In this context, in order to enhance the biological activity by inserting rational structural modifications capable to regulate the redox behaviour, the identification of the bioactive conformation must be combined with the capability to accurately calculate the electronic parameters. Accordingly, the computational approach (summarised in Figure 5 Chapter 2) was applied to design new antimalarial agents and to the investigate their redox mechanism of action. Malaria is a disease of worldwide involvements and eradicating malaria is one of the biggest programs of the World Health Organization (WHO).[172] An alarming rise in the number of lethal cases has been taken

note in recent years and it is especially due to the propagation of multi-drug resistant strains of *Plasmodium falciparum* (*Pf*), making less effective the available drugs.[173] Malaria parasite is a micro-organism highly sensitive to oxidative stress forced to live in a pro-oxidant environment including oxygen and iron, which elicit a large quantity of reactive oxygen species (ROS). This imposes a serious threat to the parasite,[165,174] which developed adaptation and survival strategies to live under these pro-oxidant conditions. Therefore, identification of redox-active cytotoxic molecules which can be activated in the parasite environment represents a rational approach to design new antimalarial drugs.

The lead compound in the family of antimalarial endoperoxides is Artemisinin (Figure 64), a natural compound isolated from the plant *Artemisia annua*.
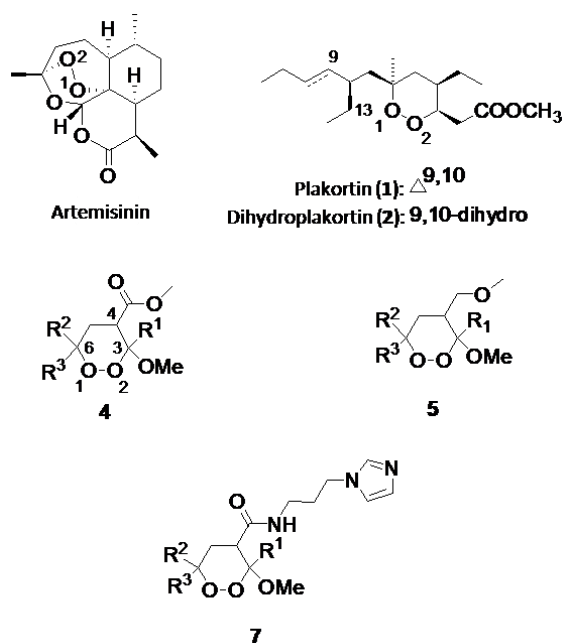
Figure 64. Structure of Artemisinin , Plakortin (**1**), Dihydroplakortin (**2**) and the scaffold of the new synthetic 3-methoxy-1,2-dioxanes (**4**-**7**). R^1, R^2, and R^3 are hydrogens or alkyl chains. Figure adapted from Persico M. et al. (**2016**) with permission of WILEY-VCH Verlag GmbH & Co. KGaA (http://dx.doi.org/10.1002/minf.201501028).

Artemisinin and its semi-synthetic derivatives are the principal components of Artemisinin-based Combination Therapies (ACTs), the first-line treatment for *Pf* malaria in disease-endemic countries. The neurotoxicity and the known cases of clinical resistance related to the use of artemisinins[175,176] strongly incite the discovery, as soon as possible, of feasible alternatives to the existing therapeutic options. Therefore, there is an impelling need for novel and affordable antimalarial drugs with high effectiveness against resistant strains and action broad stage mode.

In this regard, the research group of my tutor C. Fattorusso investigated the molecular mechanism of action of plakortin (**1**, Figure 64) and dihydroplakortin (**2**, Figure 64), simple endoperoxide-containing

polyketides isolated from the Caribbean sponge *Plakortis simplex*,[177] possess an interesting *in vitro* antimalarial activity on CQ-Resistant strains (CQ-R) of *Pf*.[178,179] The synthesis of several plakortin semi-synthetic derivatives, in order to gain insight into the structural requirements of these simple 1,2-dioxanes for exhibiting antimalarial activity, showed the crucial role both of the endoperoxide function and of 1,2-dioxane ring conformation, together with the absence of stereoselectivity in the antimalarial mechanism of action.[180] These data strongly revealed that **1**, similarly to artemisinin, does not interact with a specific protein target, but rather impedes heme detoxification which occurs in the food vacuole (FV), compromising the plasmodium antioxidant defences. On these bases, in a recently reported investigation[181-183] based on an exhaustive computational and experimental study, it was proposed that molecules belonging to the plakortin family, immediately after interaction with Fe(II), undergo a dissociative electron transfer (DET) of the endoperoxide bond, generating an oxygen-centered radical. Then, a fast "through space" 1,4- or 1,5- intramolecular radical shift to an alkyl side chain carbon atom occurs through a radical cascade that leads to the putative toxic species for the *Plasmodium* environment (Figure 65).
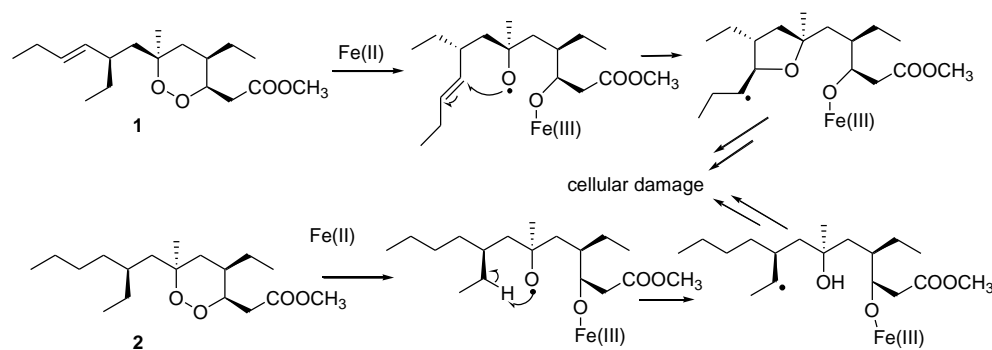
Figure 65. Proposed antimalarial mechanism of action of plakortin (**1**) and di-hydroplakortin (**2**). Figure reproduced from reference Chianese et al. (**2014**), *Bioorg. Med. Chem.* (http://dx.doi.org/10.1016/j.bmc.2014.07.034).

This putative antimalarial mechanism of action was confmed by SARs studies carried out on further plakortin analogues and the obtained information allowed to develop a pharmacophoric model of plakortins (Figure 66).[180,184-186]
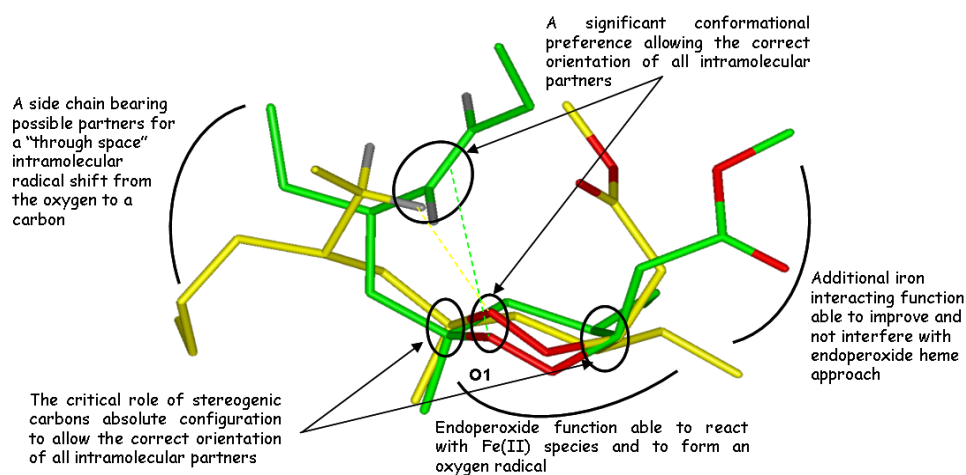
Figure 66. 3-D pharmacophore model of plakortin analogues. Figure reproduced from Taglialatela-Scafati O. (**2010**), *Org. Biomol. Chem.* (http://dx.doi.org/10.1039/b918600j).

Starting from these results, a new series of antimalarial endoperoxides was developed, characterised by a 3-methoxy-1,2-dioxane scaffold (Figure 67), a low cost synthesis and an *in vitro* antimalarial activity on *Pf* CQ-resistant strains similar to that of the lead compound plakortin.[187-189]

Computational and SAR studies performed on these new synthetic 1,2-dioxane derivatives showed that, as in the case of plakortins, the antimalarial activity is related to their capability to react with Fe(II) and to develop a carbon centered radical on the alkyl chain located on C3 and/or C6 by means of an intra-molecular hydrogen shift. Furthermore, the 3-D-dimensional-SAR study demonstrated the active role of the C4 substituent to determine the antimalarial activity, allowing to achieve IC_{50} on CQ-R strain in the low micromolar range and, at the same time, very low toxicity against human cells.

5.2 Optimisation of Cellular Pharmacokinetic and Pharmacodynamic Properties of New Antimalarial 3-methoxy-1,2-dioxanes

Starting from the results described in the previous paragraph and since the amine chains are known to play a decisive role both on cellular pharmacokinetic and pharmacodynamic properties of different antimalarials,[190-192] a new series of compounds with general structure **9**[193,194] (Figure 67) were designed and synthesised, inserting a substituted amine chain at C4 of 3-methoxy-1,2-dioxane scaffold.[193]
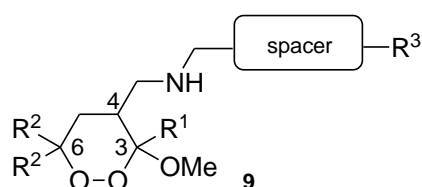
Figure 67. General structure of 4-methylamino-1,2-dioxanes **9** (R^3 = amino group containing molecular scaffold or N-heterocyclic moiety). Figure reproduced from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

In particular, a methylamino group, which, by means of a flexible spacer, binds a second amino-including scaffold or a N-heterocyclic moiety was introduced. Among the polyamine chains inserted at C4, also the aromatic substructure of chloroquine or primaquine was considered, leading to the production of new endoperoxide-quinoline hybrids. The antimalarial activity and the structures of the new compounds, modified at C4 side chain, are reported in Tables 13 and 14.

Table 13 | Structure, yield, c Log D values, antimalarial activity against CQ-S (D10) and CQ-R (W2) *Pf* strains, and cytotoxicity against HMEC-1 of 3,4-cis-3-methoxy-4-aminomethyl-1,2-dioxanes **9a-j**.
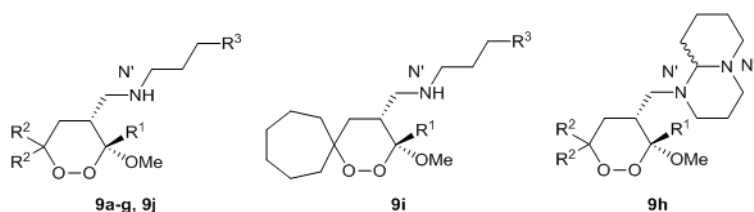
3,4-*cis*-9 (Y %)[b]	R^1	R^2	R^3	D10 IC$_{50}$ (μM)[c]	W2 IC$_{50}$ (μM)[c]	RI[d]	HMEC-1 IC$_{50}$ (μM)[c]	D10 SI[e]	W2 SI[e]	c Log D pH 7.4	c Log D pH 7.2	c Log D pH 5.5
9a (91)	*n*-Bu	Me	—N⟅N (imidazole)	0.42 ± 0.09	0.21 ± 0.04	0.5	> 59	> 140	> 280	0.5	0.3	-1.3
9b (83)	Me	*n*-Bu	—N⟅N	0.37 ± 0.05	0.4 ± 0.1	1.1	-	-	-	1.9	1.7	0.1
9c (83)	*n*-Bu	*n*-Bu	—N⟅N	0.24 ± 0.06	0.2 ± 0.1	0.8	5.14 ± 0.08	21.4	25.7	3.2	3.0	1.4
9d (85)	*n*-Bu	Me	NMe$_2$	0.6 ± 0.1	0.39 ± 0.02	0.7	-	-	-	-0.3	-0.6	-1.9
9e (88)	Me	*n*-Bu	NMe$_2$	0.54 ± 0.09	0.33 ± 0.05	0.6	3.624 ± 0.001	6.7	11.0	1	0.8	-0.3
9f (85)	*n*-Bu	*n*-Bu	NMe$_2$	0.8 ± 0.1	0.6 ± 0.3	0.8	-	-	-	2.4	2.1	0.8
9g (86)	*n*-Bu	Me	—N⟅O (morpholine)	1.4 ± 0.4	0.55 ± 0.06	0.4	-	-	-	0.2	0.0	-0.6
9h (85)	Me	*n*-Bu	-	0.6 ± 0.1	0.4 ± 0.2	0.7	-	-	-	5.4	5.3	4.0
9i (86)	Me	-	—N⟅N	2.1 ± 0.8	0.7 ± 0.1	0.3	-	-	-	0.4	0.2	-1.3
9j (65)	Me	Me	—N⟅N	>10	>10	-	-	-	-	-0.9	-1.1	-2.6
CQ[f]	-	-	-	0.05 ± 0.02	0.7 ± 0.2	14	> 38	> 760	> 54	-	-	-
DHA[g]	-	-	-	0.005 ± 0.001	0.0024 ± 0.0004	0.5	3.1 ± 0.3	620	1292	-	-	-

[a]ACD/Percepta software (version 14.0.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada) [b]Isolated yields. [c]Data are the mean ± SD of three different experiments in duplicate. [d]RI = Resistance Index: IC$_{50}$ ratio for CQ-R/CQ-S strains of *Pf*. [e]SI = Selectivity Index: IC$_{50}$ ratio for HMEC-1/CQ-R or CQ-S strains of *Pf*. [f]CQ: chloroquine. [f]DHQ: di-hydroartemisinin. Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Table 14 | Structure, yield, c Log D values, antimalarial activity against CQ-S (D10) and CQ-R (W2) *Pf* strains, and cytotoxicity against HMEC-1 of 3,4-cis-3-methoxy-4-aminomethyl-1,2-dioxanes **9k**-**s**.

3,4-*cis*-9 (Y %)[b]	R[1]	R[2]	n	D10 IC$_{50}$ (μM)[c]	W2 IC$_{50}$ (μM)[c]	RI[d]	HMEC-1 IC$_{50}$ (μM)[c]	D10 SI[e]	W2 SI[e]	c Log D[a] pH 7.4	pH 7.2	pH 5.5
9k (65)	*n*-Bu	Me	1	0.050 ± 0.002	0.11 ± 0.04	2.2	5.5 ± 0.7	110	50	1.9	1.6	0.4
9l (62)	Me	*n*-Bu	1	0.072 ± 0.004	0.11 ± 0.03	1.5	3.31 ± 0.08	46	30	3.1	2.9	1.7
9m (65)	*n*-Bu	*n*-Bu	1	0.09 ± 0.01	0.13 ± 0.02	1.4	5.49 ± 0.05	61	42	4.6	4.4	3.1
9n (61)	*n*-Bu	Me	2	0.05 ± 0.01	0.13 ± 0.02	2.6	2.5 ± 0.2	50	19	2.1	1.9	0.7
9o (60)	Me	*n*-Bu	2	0.074 ± 0.003	0.13 ± 0.01	1.8	2.24 ± 0.01	30	17	3.3	3.1	2.0
9p (64)	*n*-Bu	*n*-Bu	2	0.09 ± 0.01	0.11 ± 0.03	1.2	-	-	-	4.8	4.5	3.4
9q (63)	Me	Me	2	0.048 ± 0.006	0.34 ± 0.09	7.1	25 ± 2	521	74	0.8	0.5	-0.6
9r (60)	*n*-Bu	Me	-	0.5 ± 0.1	0.31 ± 0.07	0.6	-	-	-	2.8	2.7	2.5
9s (62)	Me	*n*-Bu	-	0.94 ± 0.04	0.5 ± 0.1	0.5	-	-	-	4.1	4.0	3.7
CQ[f]	-	-	-	0.05 ± 0.02	0.7 ± 0.2	14	> 38	> 760	> 54	-	-	-
DHA[g]	-	-	-	0.005 ± 0.001	0.0024 ± 0.0004	0.5	3.1 ± 0.3	620	1292	-	-	-

[a]ACD/Percepta software (version 14.0.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada) [b]Isolated yields. [c]Data are the mean ± SD of three different experiments in duplicate. [d]RI = Resistance Index: IC$_{50}$ ratio for CQ-R/CQ-S strains of *Pf*. [e]SI = Selectivity Index: IC$_{50}$ ratio for HMEC-1/CQ-R or CQ-S strains of *Pf*. [f]CQ: chloroquine. [f]DHQ: di-hydroartemisinin
Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

The results presented in Tables 13 and 14 evidenced that imidazol-1-yl derivatives **9a**-**c** exhibited IC$_{50}$ values in the high nanomolar range (200-400 nM) against both parasite strains, an activity greater than the whole library of endoperoxides previously reported. The need of butyl chain(s) either at position C3 or C6 of the 1,2-dioxane ring is confirmed by the absence of activity showed by **9j**. Moreover, the substitution of the butyl chains at C6 with a spiro-cycloheptane ring (**9i**) reduced the antimalarial activity (Table 13). In fact, the corresponding spiro-derivatives **4i** and **7i**, owning a methyl ester and a methyl amide substituent at C4, respectively, were both completely inactive (Table 15).

Table 15 | Structure, $c \log D$ values, and antimalarial activity of **4i**, **7i** and **9i** against CQ-S (D10) and CQ-R (W2) *Pf* strains.
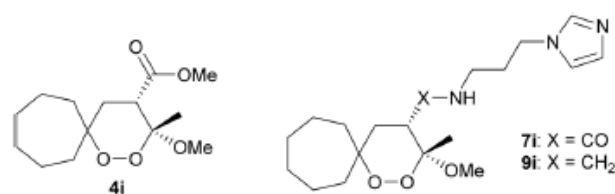
| | D10 IC$_{50}$ | W2 IC$_{50}$ | $c \log D^b$ | | |
	(μM)^a	(μM)^a	pH 7.4	pH 7.2	pH 5.5
4i	> 10	> 10	3	3	3
7i	> 10	> 10	2.1	2.0	0.6
9i	2.1 ± 0.8	0.7 ± 0.1	0.4	0.2	-1.3

A small but important loss of antimalarial activity is linked to the substitution of the imidazole ring with simple aliphatic tertiary amines (**9d-h**), especially when the tertiary amine is inserted into a morpholine ring (**9g**). Generally, all the compounds showed in Table 13 were most active against W2 (CQ-R) compared to D10 (CQ-S) *Pf* strains (RI ≤ 1; Table 13) and exhibited variable cytotoxicity against human cells HMEC-1. In particular, compound **9a**, as CQ, was very safe when tested for cytotoxicity against the human cell line. The hybridisation of the 1,2-dioxane scaffold with 4-amino-7-chloroquinoline (**9k-q**) generated a series of molecules with significantly higher activity against both *Pf* strains (Table 14). It should be noted that, as in the case of the derivatives reported in Table 13, the compound that did not exhibit any significant toxicity against HMEC-1 cells (**9q**, Table 14) was the one described by the lowest $c \log D$ value at physiological pH. Concerning the antimalarial activity, although it is complicated to discern the role of the two pharmacophoric moieties (*i.e.,* 1,2-dioxane and 4-amino-quinoline); the low IC$_{50}$ values of this series of compounds against the D10 strain (CQ-S) appear to be caused by quinoline moiety. This hypothesis is firstly confirmed by compound **9q**. In fact, having only methyl substituents at C3 and C6, the 1,2-dioxane moiety of **9q** cannot take part to the antimalarial activity; however, this compound was active as CQ on D10 (CQ-S) strain (Table 14). Moreover, **9q**, due to its lower activity against W2 (CQ-R) strain, showed an extended resistance index (RI) than to **9k-p** (Table 14). Compounds **9k-p** were indeed significantly more active than CQ against the W2 (CQ-R) strain with a levelling effect of IC$_{50}$ at the 100 nM, which could denote the participation of the endoperoxide pharmacophore to the studied antimalarial activity.

Therefore, a synergic and/or a resistance reverting effect of the two pharmacophores appears to take place in the case of the W2 (CQ-R) *Pf* strain. The 8-Aminoquinoline including hybrids **9r,s** are intriguing, too, for the known potential activity against the *Pf* hepatic stage and as transmission blocking agents. The compounds **9r,s** exhibited increased IC_{50} with respect to **9k-p** (Table 14) against both CQ-R and CQ-S *Pf* strains.

In order to investigate the SARs of this new series of antimalarial endoperoxides and to identify their 3-D structural features accountable for the antimalarial activity, all the compounds reported in Tables 13 and 14 (**9a-j** and **9k-s**, respectively) were subjected to computational studies. These studies started with the estimation of the ionic forms at pH 7.4, 7.2, and 5.5 (blood, cytoplasm, and *Pf* FV) (Table 16). Subsequently, in order to investigate the role on antimalarial activity of a probable accumulation in the FV, the distribution coefficient (Log *D*) values at blood, cytoplasm, and FV pH, were calculated using the ACD/Percepta software (version 14.0.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada). Results are reported in Tables 17 and 18. Log *D* is an expression of the lipophilicity of the compounds which reveals the equilibrium concentration of ionic forms at a given pH. It is usually accepted[195,196] that molecules with reasonable lipophilicity (Log *D* 0-3) show a good balance between permeability and solubility and are ideal for cell membrane permeation. When the Log *D* is included among a 3-5 value, the compounds still show a good membrane permeability, but absorption is lower owing to their lower solubility.

Table 16 | Calculated ionic forms of 3-methoxy-4-aminomethyl-1,2-dioxanes **9a-s**.

Compound	Ionic form (%)[a]		
	pH7.4	pH 7.2	pH 5.5
9a	DP(15) P(85)	DP(22) P(78)	DP(93) P(7)
9b	DP(15) P(85)	DP(22) P(78)	DP(93) P(7)
9c	DP(15) P(85)	DP(22) P(78)	DP(93) P(7)
9d	DP(80) P(20)	DP(86) P(14)	DP(100)
9e	DP(80) P(20)	DP(86) P(14)	DP(100)
9f	DP(80)	DP(86)	DP(100)

	P(20)	P(14)	
9g	DP(2) P(98)	DP(3) P(97)	DP(56) P(44)
9h	P(68) N(32)	P(77) N(23)	P(99) N(1)
9i	DP(15) P(85)	DP(22) P(78)	DP(93) P(7)
9j	DP(15) P(85)	DP(22) P(78)	DP(93) P(7)
9k	DP(14) P(86)	DP(21) P(79)	DP(100)
9l	DP(14) P(86)	DP(21) P(79)	DP(100)
9m	DP(14) P(86)	DP(21) P(79)	DP(100)
9n	DP(11) P(89)	DP(16) P(84)	DP(100)
9o	DP(11) P(89)	DP(16) P(84)	DP(100)
9p	DP(11) P(89)	DP(16) P(84)	DP(100)
9q	DP(11) P(89)	DP(16) P(84)	DP(100)
9r	P(100)	P(100)	P(100)
9s	P(100)	P(100)	P(100)

[a]Percentage of ionic form in brackets. P protonated form; DP di-protonated form (ACD/Percepta software, version 14.0.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada).

Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Table 17 | c Log D values of 3-methoxy-4-aminomethyl-1,2-dioxanes **9a**-**j**.

Compound	c Log D		
	pH7.4	pH 7.2	pH 5.5
9a	0.5	0.3	-1.3
9b	1.9	1.7	0.1

Compound			
9c	3.2	3.0	1.4
9d	-0.3	-0.6	-1.9
9e	1	0.8	-0.3
9f	2.4	2.1	0.8
9g	0.2	0.0	-0.6
9h	5.4	5.3	4.0
9i	0.4	0.2	-1.3
9j	-0.9	-1.1	-2.6

Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Table 18 | c Log D values of 3,4-cis-3-methoxy-4-aminomethyl-1,2-dioxanes **9k-s**.

Compound	c Log D		
	pH7.4	pH 7.2	pH 5.5
9k	1.9	1.6	0.4
9l	3.1	2.9	1.7
9m	4.6	4.4	3.1
9n	2.1	1.9	0.7
9o	3.3	3.1	2.0
9p	4.8	4.5	3.4
9q	0.8	0.5	-0.6
9r	2.8	2.7	2.5
9s	4.1	4.0	3.7

Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

On the other hand, hydrophilic compounds (Log D < 0) show good solubility but inadequate membrane permeability.[195,196] Therefore, in conformity with the data reported in Tables 17 and 18, an indisputable degree of accumulation in the FV of the parasite is expected for all compounds, there is no precise correlation between the calculated Log D (c Log D) values and the antimalarial activity. A similar observation has been reported for 1,2,4-trioxaquines, molecular hybrids including both a quinoline and 1,2,4-trioxane pharmacophore moieties. In the new series of antimalarial endoperoxides, it was true for the

endoperoxide derivatives **9a-j** but not for the quinoline-hybrids **9k-s** (Table 13). The c Log D values did not give reasons for the activity trend observed with the insertion of the amino chain at C4, such as, 3-(1H-imidazol-1-yl)-N-methylpropan-1-amine (**9a-c**) > N1,N1,N3-trimethylpropane-1,3-diamine (**9d-f**) ≈ octahydro-1-methyl-1H-pyrido[1,2-a] pyrimidine (**9h**) > N-methyl-3-morpholinopropan-1-amine (**9g**). Furthermore, compound **9c**, conversely from **9a** and **9b**, should be capable to diffuse back to the parasite cytoplasm across the FV membrane. This characteristic did not interfere with the antimalarial activity of **9a-c**, which is almost identical (Table 13). To investigate the role of the C4 amino-imidazole chain in the antimalarial activity of compounds **9a-c** (the amido-imidazole analogue of **9a** was inactive),[189] compound **9j** was synthesised. Its complete absence of activity proved that the presence of at least one butyl chain at C3 is essential for antimalarial activity (**9j** *vs.* **9a**, Table 13). However, due to the unfavourable c Log D value at physiological pH (*i.e.*, -0.9, Table 17), the observed inactivity of **9j** could be attributed to cellular pharmacokinetics. To get an amino-imidazole analogue with increased c Log D without affecting the pharmacodynamic features, was used the 3-methoxy-1,2-dioxaspiro[5.6]dodecane moiety. Indeed, in the ester series (Table 15), the spiro derivative resulted inactive despite its favourable c Log D, proving that the spiro-cycloheptane substituent at C6 do not possess the pharmacodynamic requirements for antimalarial activity. Thus, the C4 substituent of the inactive analogues **4i** and **7i** was replaced with the amino-imidazole chain, synthesising the analogue **9i** (Tables 13 and 17). The amino-imidazole derivative **9i** showed that, when there is a sufficient lipophilicity, just the introduction of the amino-imidazole chain at C4 is sufficient to restore some antimalarial activity (Table 13). In the end, in agreement with the data reported in Tables 17 and 18, the minimum c Log $D_{7.4}$ value for activity against human HMEC-1 cell line seems to be +1. Indeed, the compounds **9a** and **9q**, with a c Log $D_{7.4}$ value < 1, did not have any significant activity. In order to continue to rationalise the SARs of these new antimalarial endoperoxides their conformational and electronic properties were investigated. Prevalent ionic forms at cytoplasm and *Pf* FV pH were subjected to an in depth conformational analysis, which included a molecular dynamics (simulated annealing, SA) procedure followed by molecular-mechanics (MM) energy minimisation, and the subsequent quantum mechanics (QM) full geometry optimisation (MOPAC, PM7) of the resulting MM minima (see Experimental Section for details). The resulting conformers were ranked by their potential energy values and grouped into families on the basis of their 1,2-dioxane ring conformation (Table 19).

Table 19 | Occurrence rate (%) of 1,2-dioxane ring conformations considering PM7 conformers within 5 kcal mol^{-1} from the GM.

Compound	Ionic form[a]	Chair A	Chair B	Skew boat A	Skew boat B
9a	P	68	0	0	32
	DP	17	0	0	83
9b	P	17	0	0	83
	DP	0	0	0	100
9c	P	0	0	0	100
	DP	0	0	0	100
9d	P	65	0	0	35
	DP	0	0	0	100
9e	P	71	0	0	29
	DP	0	0	0	100
9f	P	78	0	0	22
	DP	0	0	0	100
9g	P	49	0	7	45
	DP	0	0	0	100
9h[b]	P	9	0	9	82
9h[c]	P	23	6	0	71
9i	P	63	5	0	32
	DP	0	0	0	100
9j	P	41	0	0	59
	DP	0	0	0	100
9k	P	36	15	12	36
	DP	50	0	0	50
9l	P	50	0	20	30
	DP	0	0	0	100
9m	P	29	13	29	29
	DP	0	0	0	100
9n	P	50	6	19	25
	DP	33	17	0	50
9o	P	25	0	12.5	62.5
	DP	0	50	0	50
9p	P	25	0	0	75

	DP	0	0	0	100
9q	P	45	7	13	35
	DP	20	0	0	80
9r[b]	P	64	12	3	21
9r[c]	P	55	17	0	28
9s[b]	P	50	0	0	50
9s[c]	P	56	0	0	44

[a]DP = di-protonated form, P = protonated form. [b]Diastereomer with (R)-configured stereocenter in the side-chain. [c]Diastereomer with (S)-configured stereocenter in the side-chain.
Table adapted from Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

These computational studies evidenced that the series of derivatives **9** presents particular conformational parameters than to the plakortin and the previous synthetic analogues[184-189] Indeed, the presence of the amino chain at C4 stabilises the Skew Boat B conformation of the 1,2-dioxane ring, which becomes by far the most populated conformer at the FV pH (Table 19). This boat-like conformation of the 1,2-dioxane ring facilitates accessibility of endoperoxide oxygens, reproducing the artemisinin endoperoxide ring conformation (Figure 68).
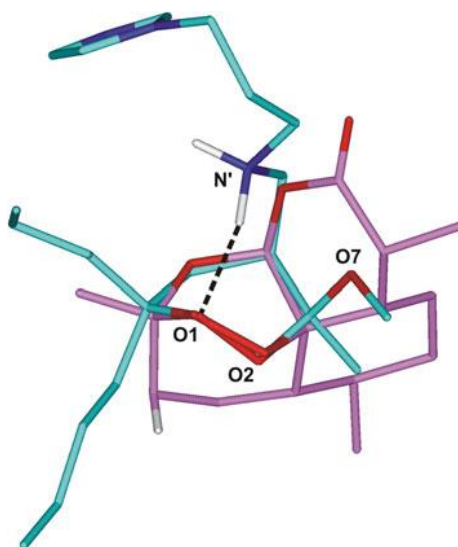
Figure 68. Superimposition of PM7 GM of **9b** (cyan) in protonated form on artemisinin X-ray structure (pink; CSDS code: QNGHSU). The molecules are coloured by atom type (O = red, N = blue, and H = white). Hydrogens are omitted for sake of clarity, with the exception of those involved in the H-bonds (highlighted by a black dashed line). Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

The structural analysis of all conformers within 5 kcal/mol from the global minimun (GM) pointed out that the di-amino derivatives **9d**-**f** and the amino-imidazole derivatives **9a**-**c**, **9j** (Table 13), present GM conformers displaying comparable conformational characteristics, not only in the protonated (N') but also

in di-protonated (N' and N") forms. These conformers resulted to be characterised by the Skew Boat B conformation of the 1,2-dioxane ring which favours the formation of a network of intra-molecular H-bonds, engaging both O1 and the secondary amine N' (Figure 69).
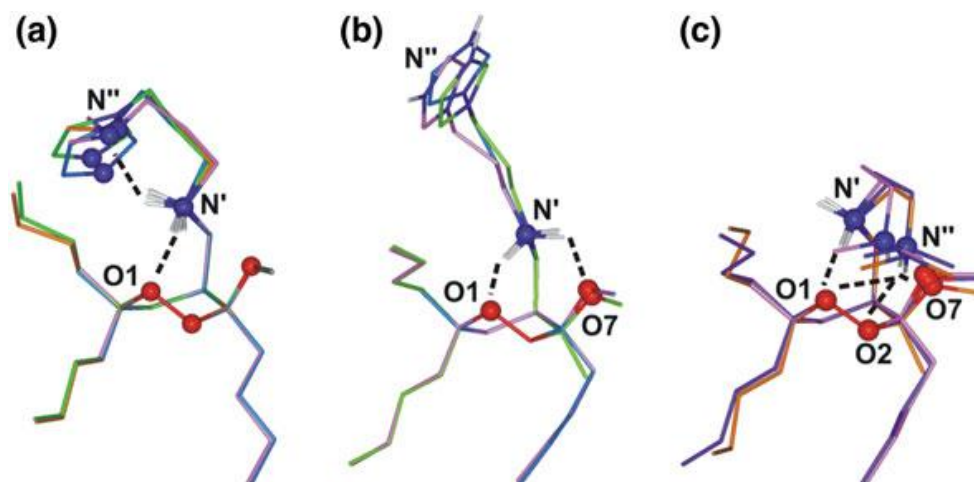
Figure 69. Superimposition of GM conformers of: (**a**) **9a** (light blue), **9b** (light green), **9c** (yellow), **9d** (pink), **9e** (orange), and **9f** (violet); all protonated on N'; (**b**) **9a** (light blue), **9b** (light green), and **9c** (pink) in di-protonated form; (**c**) **9d** (pink), **9e** (orange), and **9f** (violet); protonated on N' and N". The molecules are coloured by atom type (O = red, N = blue, and H = white). The heteroatoms involved in H-bonds (highlighted by a black dashed line) are displayed in ball. Hydrogens are omitted for sake of clarity, with the exception of those involved in H-bonds. Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Additionally, these low-energy conformers presented, with the exception of **9j**, appropriate intra-molecular distances for a H-shift from a butyl chain to the putative oxygen radical, in agreement with the developed pharmacophoric model (Figures 69 and 70).
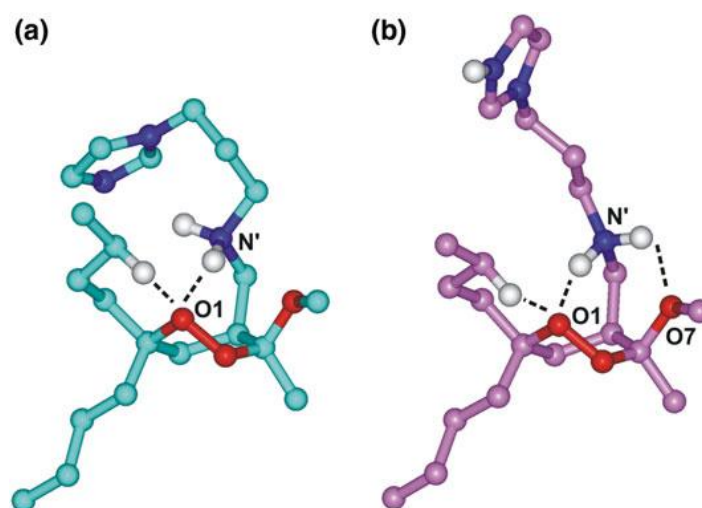
Figure 70. PM7 GM of **9c** presenting intra-molecular distances suitable for the 1,5-H shift in protonated (cyan; **a**) and di-protonated (pink; **b**) form. Atoms are coloured by atom type (O = red, N = blue, and H = white). Hydrogens are omitted for sake of clarity, with the exception of those involved as possible partners in a "through-space" intra-molecular radical shift and in

the H-bonds (highlighted by a black dashed line). Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

In order to clarify the effect of this conformational behaviour on antimalarial activity, a dynamic docking study of **9c**-heme complex was performed. The protonation state both of **9c** and heme has been evaluated taking into consideration the FV pH, where heme digestion predominantly occurs.[197] Both heme parameters and atomic partial charges were attributed employing the QM method (PM7). Throughout the dynamic docking calculation, which combines Monte Carlo and SA methods, all rotatable bonds of **9c** and heme are let completely free to move. The docking results showed only one possible complex structure (Figure 71). The resulting docked complex was then subjected to a full geometry optimisation by the semi-empirical QM method PM7 (for details see Experimental Section).
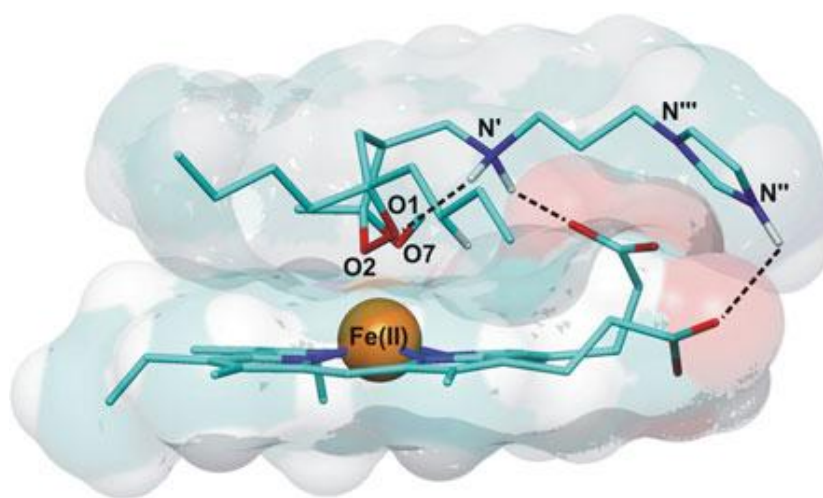
Figure 71. Docking results of **9c** (protonated on N' and N'') in complex with heme. The molecules are coloured by atom type (C = cyan; O = red; N = blue; Fe = orange and H = white). Iron atom vdW volume is shown (scaled by 50% for clarity of presentation). Hydrogens are omitted for sake of clarity, with the exception of those involved in H-bonds (highlighted by a black dashed line). Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

The obtained conformation of **9c** exhibited the key characteristics (*i.e.*, Skew Boat B and O1-N' interaction) of the GM conformer identified by the conformational search (Figure 69b). In addition, the both protonated nitrogens N' and N'' set up charge-assisted H-bonds with the ionised and neutral propionic groups of heme, analogously to what mentioned for the protonated amine chain of CQ.[198] As stated above, at *Pf* FV pH, when the two basic centers are protonated, compounds **9a-f** exhibited a higher propensity for the Skew Boat B conformation and the H-bond between N' and O1 (Table 17 and Figure 69). For the compounds **9a-c**, this is caused by a shortage of an additional electron donating group (*i.e.*, not protonated imidazole) contending with the endoperoxide oxygen O1 the favourable interactions (*i.e.*, cation-π) with the protonated secondary amine N' (Figures 70a *vs*. 70b). On the contrary, the compounds

9d-**f** resulted to be characterised by an additional stabilisation of the Skew Boat B conformation, due to the interaction of the distal and protonated tertiary amine group (N'') with O1, O2, and O7 (Figures 69a *vs.* 69c). Consequently, the only thing different between the amino-imidazole and the di-amino derivatives (**9a-c** and **9d-f**, respectively) is that, in the di-protonated form, the GM conformers assume a lying conformation of the substituent at C4, with the protonated imidazole ring orienting far away from the remaining molecule portion. Therefore, the **9a-c** GM conformers reflect the heme interacting conformation obtaining from docking simulations (Figure 71). On the other hand, in the **9d-f** GM conformers the protonated tertiary amine N'' is bonded through a H-bond to O1, O2, and O7 (Figures 69a *vs.* 69b). As a consequence, when the di-amino derivatives (**9d-f**) approach heme, the distal protonated amine N'' should first of all disrupt the intra-molecular H-bonds to form ionic interactions with the propionate groups. This could explain the total greater antimalarial activity of **9a-c** with respect to **9d-f** (Table 13). In agreement with the apparent pK_a values (Table 16), differently from compounds **9a-f**, the less active morpholine analogue **9g** results even mono-protonated at the acidic FV pH (44%). In this form, the **9g** GM conformer does not adopt either the Skew Boat B conformation or the O1-N' interaction, and the first conformer having such characteristics is at ≈ 1 kcal/mol from the GM. On the contrary, in the di-protonated form, **9g** showed the same conformational features of **9d-f** (Figure 69c), with the protonated tertiary amine (N'') establishing H-bonds with O7. For this reason, even in the case of di-protonated **9g**, the morpholine nitrogen could be involved in an extra-interaction with heme propionate after dsirupting its intra-molecular H-bond. A specific case is constituted by compound **9h** (W2 IC_{50} = 400 nM, Table 13), analysed as a mixture of two different diastereomers due to the stereogenic carbon of the bi-cyclic ring. In this analogue, the two basic nitrogens (single pK_{a1}: 7.56 ± 0.20 and single pK_{a2}: 7.24 ± 0.20) are forced into a bi-cyclic structure and, in agreement with the calculated apparent pK_a value, only the mono-protonated form is available at cytoplasm and FV pHs (Table 16). In order to investigate the role of the bicyclic ring on the antimalarial activity, **9h** diastereomers were subjected to the same docking procedure applied to **9c** (see for details Experimental Section). These studies evidenced, also in this case, only one possible complex structure for each **9h** diastereomer. According to the conformational analysis results, even though not exhibiting the H-bond between N' and O1, the two diastereomers still showed the Skew Boat B of the 1,2-dioxane ring as bioactive conformation with the two conformationally restricted amine nitrogens appropriately well-positioned to interact with the negatively charged heme propionate group (Figure 72).
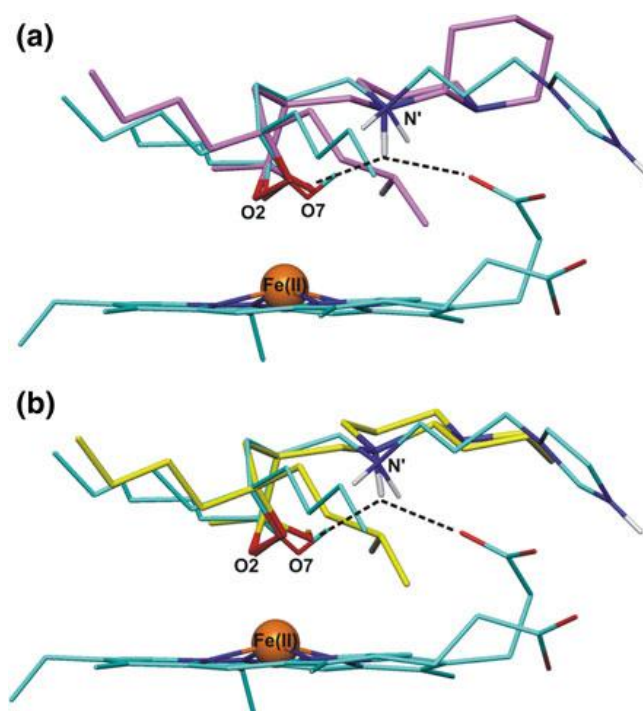
Figure 72. Superimposition of the putative bioactive conformer of **9h** (**a**) (pink): (R)-configured stereocenter in the side-chain; (**b**) (yellow): (S)-configured stereocenter in the side-chain) on the docking results of **9c** in complex with heme (cyan). The molecules are coloured by atom type (O = red; N = blue; Fe = orange; H = white). Iron atom vdW volume is shown (scaled by 50% for clarity of presentation). Hydrogens are omitted for sake of clarity, with the exception of those involved in H-bonds (highlighted by a black dashed line). Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Lastly, also the amino-imidazole spiro derivative (**9i**; Table 13) present the di-protonated form at FV pH (Table 16). The GM conformer of this form shows the Skew Boat B and the O1-N' interaction, but, in agreement with experimental peroxide oxygen-heme iron distances collected from the Cambridge Structural Database,[199] the presence of the stiff spiro-cycloheptane at C6 compromises the endoperoxide approach to heme (Figure 73).
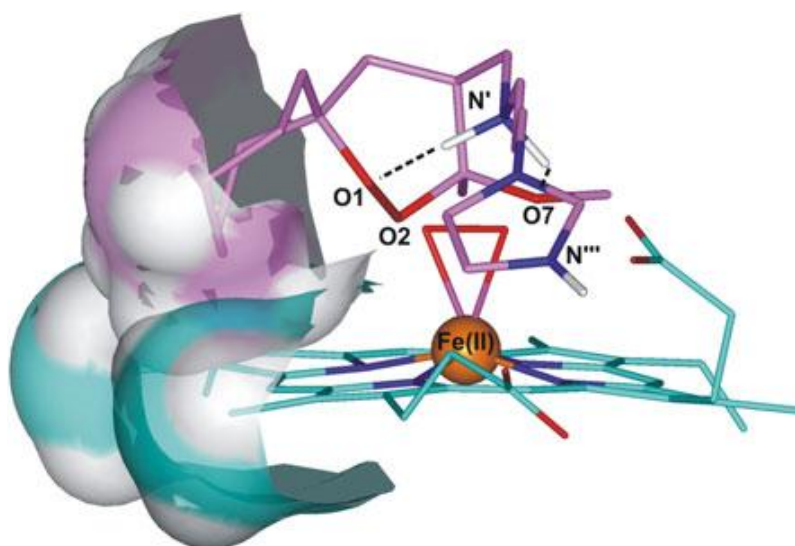
Figure 73. Di-protonated form of **9i** in complex wih heme superimposed (coordination O2/O7) on the X-ray structure of peroxo-bridged heme complex (CSD code: UKACIS). The molecules are coloured by atom type (C = pink and cyan for PM7 conformer and heme, respectively; O = red; N = blue; Fe = orange; H = white). The solvent accessible surface of atoms responsible for steric hindrance between **9i** and heme is shown (transparency = 50%). Iron atom vdW volume is shown (scaled by 50% for clarity of presentation). Hydrogens are omitted for sake of clarity, with the exception of those involved in the intra-molecular H-bonds (highlighted by a black dashed line). Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

The conformational behaviour of the endoperoxide-aminoquinoline hybrids (**9k-s**; Table 14) was examined by employing the same computational procedure used for **9a-j**. In these compounds, the bulky quinoline moiety produced a less homogenous conformational behaviour of the low energy conformers of **9k-s**. In fact, because of hydrophobic interaction between the quinoline system and the butyl chain(s), the conformational properties are affected by the substitution at C6 (Table 19). However, the **9k-s** GM conformers are still characterised by the Skew Boat B, in the mono-protonated and di-protonated forms. The endoperoxide-7-chloroquinoline hybrids (**9k-q**; Table 14) are completely present in the di-protonated form at FV pH, while they are mainly (≈ 80%) mono-protonated at cytoplasm pH (Table 16). Taking into account the mono-protonated form, N' presents a H-bonded with O1 and O7 in **9k**, **9n**, and **9q** (Figure 74a), and only with O1 in **9l**, **9m**, **9o**, and **9p** (Figure 74b).
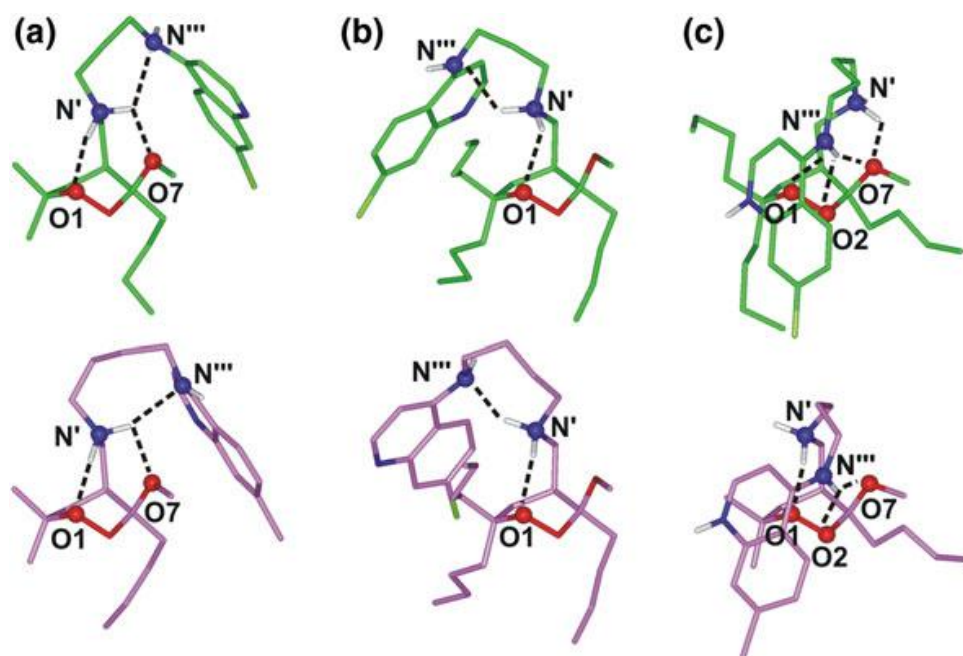
Figure 74. GM conformers of: (**a**) **9k** (green) and **9n** (pink) (protonated on N′); (**b**) **9m** (green) and **9p** (pink) (protonated on N′); (**c**) **9m** (green) and **9q** (pink) (protonated on N′ and N″). The molecules are coloured by atom type (O = red, N = blue, and H = white). The heteroatoms involved in H-bonds (highlighted by a black dashed line) are displayed in ball. Hydrogens are omitted for sake of clarity, with the exception of those involved in H-bonds. Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

Differently, in the di-protonated form N′ presents a H-bonded with O1 (**9k-m**) or O7 (**9n-q**), while N‴ continues to interact with O1, O2, and O7 (Figure 74c). Therefore, the precise pattern of intra-molecular H-bonds is induced by the butyl chains at C6 in the protonated forms (Figures 74a and 74b), and by the length of the alkyl chain that connects the two protonated amino groups in the di-protonated forms (Figure 74c). It should be noted that, the **9k-q** GM conformers exhibit the quinoline nitrogen (N″) accessible to heme iron interaction, while the endoperoxide function is sterically obstructed by the quinoline ring (Figure 74). The primaquine-type hybrids (**9r** and **9s**) only appear in the mono-protonated form at FV and cytoplasm pH (Table 16); moreover, as showed for the mono-protonated GM conformers (**9k-q**), due to the presence of the butyl chains at C6 N′ is H-bonded with only O1, or O1 and O7 (Figure 75).
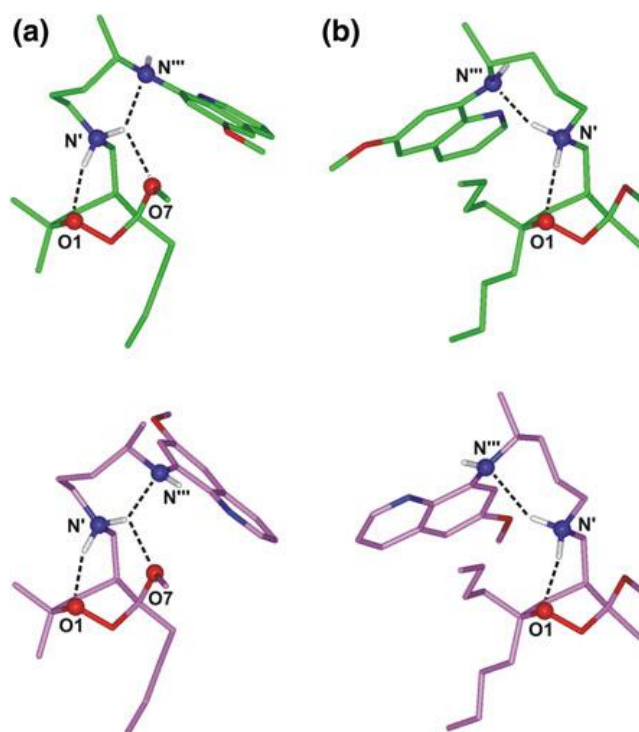
Figure 75. GM conformers of: (**a**) **9r** diastereomer with (S)-configured stereocenter in the side-chain (green) and diastereomer with (R)-configured stereocenter in the side-chain (pink); (**b**) **9s** diastereomer with (R)-configured stereocenter in the side-chain (green) and diastereomer with (S)- configured stereocenter in the side-chain (pink). The molecules are coloured by atom type (O = red, N = blue, and H = white). The heteroatoms involved in H-bonds (highlighted by a black dashed line) are displayed in ball. Hydrogens are omitted for sake of clarity, with the exception of those involved in H-bonds. Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

In opposite way to **9k-q**, in **9r-s** hybrids the quinoline ring does not hamper the endoperoxide group (Figures 74a and 74b *vs*. Figures 75a and 75b).

Taken all together, these results evidenced the key role played by the amine chain at C4 on the antimalarial activity. In this view, the SARs of our new compounds was examined. Some amines are well-known to be efficient as antimalarials[200,201] or to be present as additional pharmacophoric features in different antimalarial compounds.[190-192] Accordingly, the new series of endoperoxides **9a-s** showed an improved antimalarial activity with respect to the previous analogues, thanks to the amine substituent at C4. This enhancement in the activity can be assigned to cellular pharmacokinetics and/or pharmacodynamics. Polyamines play a key role as radical scavengers in biological environments[202] and their capability to shuttle a free-radical from oxygen to carbon atoms has been extensively studied.[203-207] Thus, the polyamine substituent could promote the activity of molecular scaffolds, the activity of which is rigorously connected to their redox potential by playing a critical role in the formation/propagation of toxic radical species. Additionally, for compounds presumed to interact with heme in free form, a basic chain could be implicated in favourable interactions with heme propionate groups. In this respect, the conformational analysis carried out on molecules of the series **9** unveiled a strong penchant of the protonated secondary

amine to form H-bond with O1, stabilising a 1,2-dioxane ring conformation which facilitates peroxide oxygen accessibility. This was proved by dynamic docking calculation on **9c**-heme complex, pointing out favourable interactions between the polyamine chain and the heme propionate groups. Once this pre-reactive complex was formed, a potential scenario is that, after the Fe(II) catalysed production of the distonic radical anion, the secondary amine could, at the same time, protonate the oxygen anion and shuttle the oxygen radical to an adjacent carbon through a hydrogen radical shift.[203-207] Additionally, the presence of such multiple basic functionalities may favour the accumulation of compounds **9a**-**s** in the acidic FV of the parasites. More than be reached thanks to a polyamine transporter,[208,209] most probably this could be achieved to an insufficiency of the back diffusion across the membrane of a single or double protonated derivative. Calculated pK_a and Log D values at cytoplasm and FV pH suggested that compounds **9a**-**s** can accumulate in the parasite FV but no linear correlation was observed between the antimalarial activity and the c Log D values, pointing out that additional parameters also influence the antimalarial activity. It is important to underline that most of the low energy conformers (*i.e.*, within 5 kcal/mol from the GM) of compounds **9a**-**s** show an intra-molecular H-bond between the two amine functions. The formation of intra-molecular H-bonds taking part in ionised groups, can "mask" the molecule charge and influence its passage across cellular membranes. The subgroup of 1,2-dioxanes **9k**-**s** hybridises two biologically active entities, the 1,2-dioxane scaffold and a quinoline-based pharmacophore. In the recent years examples of artemisinin-quinolines hybrids have been reported.[210-212] A recent study shows the most intriguing results achieved with hybrid molecules including fully synthetic 1,2,4-trioxanes or 1,2,4-trioxolane and an aminoquinoline derivative.[213,214] These hybrid molecules show a double mode of action, a concept concerned to as "covalent biotherapy", thus they favour heme alkylation with the peroxy entity, and heme stacking with the aminoquinoline moiety resulting in a strengthened inhibition of haemozoin formation.[215,216] Killing the parasite by two different mechanisms has also the advantage of reducing the development of resistance.[217] For the hybrids **9k**-**p**, the increased activity of CQ compared to the peroxide moiety is probable to play an important role in the antimalarial activity against *Pf* CQ-S strains. Contrarily, **9k**-**p** are more active compared to chloroquine against *Pf* CQ-R strains and resulting SARs suggest that the peroxide pharmacophore is responsible for this increased antimalarial activity. It cannot be ruled out the possibility that the improved activity against CQ-R strains is ascribable to the modification of the basic side chain of CQ, which would not to be recognised by the mutated CQ-transporter. With respect to CQ, primaquine is well-known to be poorly active *in vitro* against *Pf* infected erythrocytes, not only on CQ-S but also on CQ-R strains (W2 IC$_{50}$ = 3.3 μM).[212] Accordingly, the antimalarial activity showed by the hybrids **9r** and **9s** has to be ascribed to the mechanism of action of the peroxide group, while the primaquine moiety acts, in this specific case, as an amine-containing C4 substituent. In summary, optimisation of the cellular

pharmacokinetics of this series of antimalarial peroxides by means of the introduction of a substituted amino side-chain at C4, appears not to be the only factor accountable for their improved antimalarial activity (Figure 76).
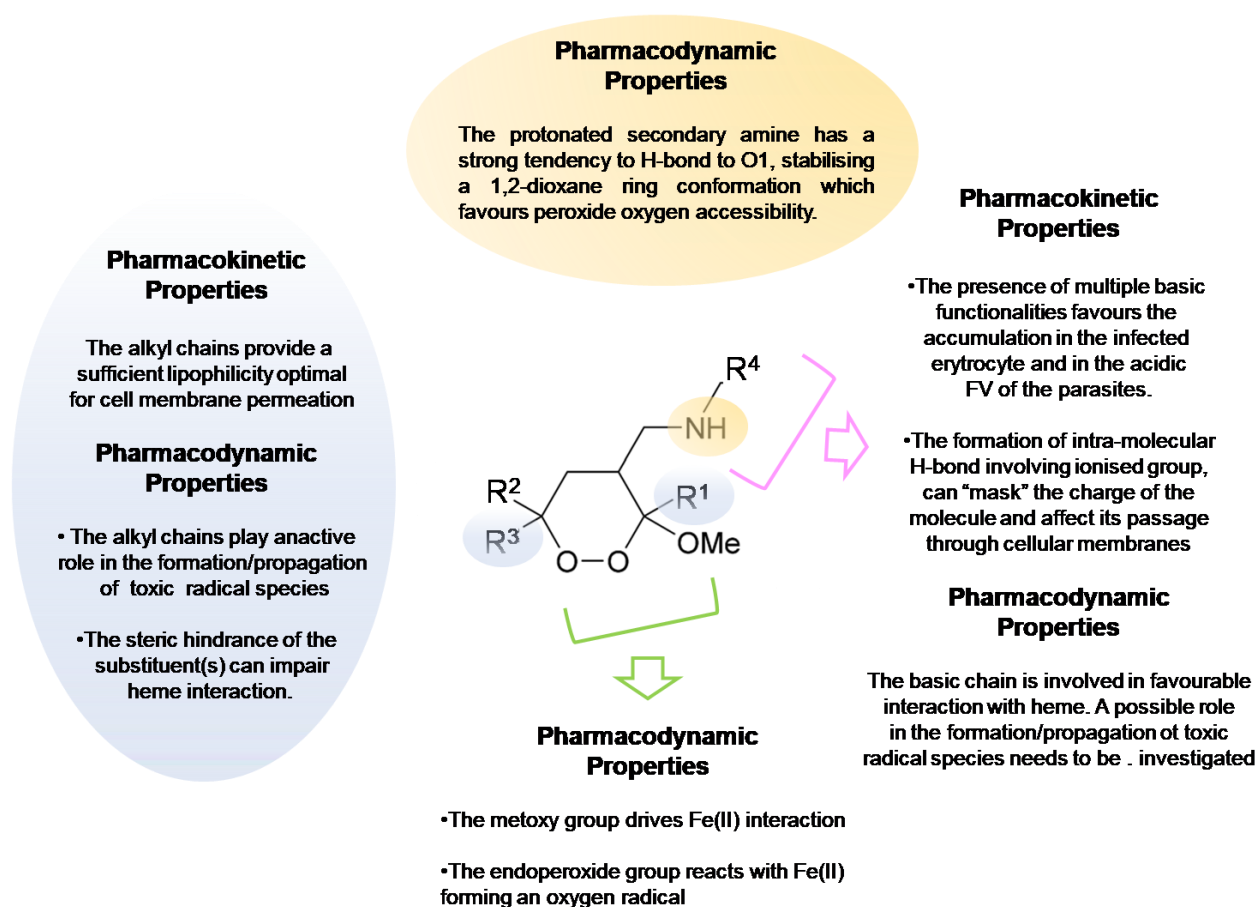
Figure 76. Drawings of the SARs of the 4-methylamino-1,2-dioxanes **9a-s**. Figure adapted from reference Sonawane et al. (**2015**) with permission from the Royal Society of Chemistry (http://dx.doi.org/10.1039/c5ra10785g).

As previously discussed,[187-189] the substituent at C4 could also affect the pharmacodynamic properties, having an effect on the ability of the peroxide bridge to undergo reductive cleavage and to generate and/or to propagate the putative toxic radical chain. Molecular modelling studies put forward that the increased antimalarial activity is correlated to the stabilisation of a conformation which promotes the assumed reductive cleavage of the endoperoxide bridge with a probable participation of the inserted amine functions in toxic radical formation and/or propagation. This is well-supported by the fact that substituting the amide group with an amino group increased the antimalarial activity and determined distinct SARs. Indeed, the **9a** amido-imidazole analogue (which presents one butyl chain at C3; W2 (CQ-R) IC$_{50}$ = 0.21 μM) resulted to be inactive as antimalarial, while the **9b** amido-imidazole analogue (which presents two butyl chains at C6, W2 (CQ-R) IC$_{50}$ = 0.4 μM) was one of the most active derivative (W2

(CQ-R) IC_{50} = 0.5 µM).[188] Therefore, substituting the amide with an amino group increased the antimalarial activity in an independent manner from the position of the butyl chain(s). However, the lack activity of **9j** proved the necessity of at least one butyl chain at C3 (**9j** *vs.* **9c**), the activity of **9i** compared to its analogues **4i** and **7i** pointed out that the secondary amine function at C4 is enough to restore some antimalarial activity and that the inactivity of **9j** is probably due to an unfavourable Log *D*. In this respect, it has to be underlined that, in agreement with *c* Log $D_{7.4}$ values, the minimum threshold of lipophilicity to be active against infected erythrocytes and normal human cells, is ~ -1 and ~ +1, respectively. Accordingly, compound **9a**, which has a *c* Log $D_{7.4}$ = 0.5, presents a IC_{50} of 210 nM against *Pf* W2 strain (CQ-R) and is inactive on human HMEC-1 cells, showing an optimal SI.

In conclusion, the introduction of a substituted amine chain on easily accessible 3-methoxy-1,2-dioxane scaffold led to a significant improvement of the antimalarial activity as result of a combination of pharmacokinetic and pharmacodynamics effects that, rationalised in the light of the developed pharmacophoric model, can steer the development of further optimised 1,2-dioxane antimalarials.

Chapter 6

EXPERIMENTAL SECTION: MOLECULAR MODELLING STUDIES

Molecular modelling calculations were performed on SGI Origin 200 8XR12000 and E4 Server Twin 2 x Dual Xeon 5520, equipped with two nodes. Each node: 2 x Intel Xeon QuadCore E5520, 2,26Ghz, 36 GB RAM. The molecular modelling graphics were carried out on a personal computer equipped with Intel(R) Core (TM) i7-4790 processor and SGI Octane 2 workstations.

6.1 Sampling procedure of the conformational space of the ligands

6.1.1 Theoretical Background

The detailed analysis of the ligand conformational behaviour is a central task for the identification of the structural features responsible for the observed biological activity and, then, for the molecular design of new optimised analogues.[170] At this aim, several conformational sampling methods were developed. These sampling methods fall into two broad categories: systematic search and stochastic methods.[171] The choice of the most suitable method depends on the flexibility of the ligands under study (*i.e.*, number of torsional angles). Indeed, although the systematic search can be very effective for molecules with limited conformational flexibility, the exponential growth of the search space with the number of rotatable bonds limits its utility as a general conformational sampling technique. Then, for flexible molecules, it is necessary to use stochastic methods, such as simulated annealing (SA),[40,218] which rely on probabilistic descriptions to generate conformations of a molecule and to aid in locating the global minimum. Accordingly, in order to exhaustively sample the conformational space of the ligands considered in my research projects, a computational protocol which combines SA method with MM energy minimisation calculations was applied.[218]

SA is based on a connection between statistical mechanics and the process of crystallisation. If a physical system is heated until it melts, and then cooled slowly, the entire arrangement can be made to produce the most stable arrangement, and not get trapped in a local minimum. The idea here is to randomly rotate bonds and bend rings within the molecule until a preferential (minimum energy) geometry is attained. Initially, during SA simulation, the molecule is considered to be in a high-temperature system; this means that it has enough energy to move from low to high energy conformations. This is important because often

the global minimum conformation of a molecule may be very different from the initial conformation. As more conformations are explored, the temperature of the system decreases, making the molecule less inclined to move out of low energy conformations, and thus looking more closely at other minima. Accordingly, in SA, the temperature is altered in time increments from an initial temperature to a final temperature by adjusting the kinetic energy of the structure (by rescaling the velocities of the atoms) for identifying the minimum conformers.

6.1.2 Conformational sampling procedure

The conformational space of all compounds is sampled through a number of cycles of simulated annealing calculations ($\varepsilon = 80*r$ and $\varepsilon = 1$). The number of cycles was selected on the bases of the number of torsional angles of the compounds. The following protocol was applied: the system was heated to 1000 K over 2000 fs (time step of 3.0 fs); a temperature of 1000 K was applied to the system for 2000 fs (time step of 3.0 fs) to surmount torsional barriers.; successively, temperature was linearly reduced to 300K in 1000 fs (time step of 1.0 fs). Resulting conformations were then subjected to Molecular-mechanics (MM) energy minimisation within the Insight 2005 Discover Module until the maximum RMSd was less than 0.001 kcal/Å, using Conjugate Gradient[219] as minimisation algorithm. The resulting MM conformers were then subjected to a full geometry optimisation by semiempirical calculations, using the quantum mechanical method PM7[72] in the MOPAC2012 package[73] and EF (eigenvector following routine)[74] as geometry optimisation algorithm. GNORM value was set to 0.01. To reach a full geometry optimisation, the criterion for terminating all optimisations was increased by a factor of 100, using the keyword PRECISE.

6.2 Monte Carlo/Simulated Annealing based docking protocol

6.2.1 Theoretical Background

Understanding the molecular mechanisms underlying conformational changes of proteins is a critical issue for the design of new drugs and for the investigation of their mechanism of action. Accordingly, in protein-protein and ligand-protein interaction simulations is essential to consider protein flexibility. At this aim, as already reported in the paragraph 1.2.1 "Classification of CADD approaches", several computational methods have been developed for the investigation of ligand-protein or protein-protein interactions.[35,36] The choice of the most suitable method depends on the biological and physical

characteristics of the protein under study and in order to improve the overall quality of the simulations, hybrid approaches have also been developed.[47-51] On the bases of the above considerations, as docking procedure, a Monte Carlo minimisation/SA[57,68,69] based docking protocol was applied (Affinity, SA_Docking; Insight 2005, Accelrys, San Diego, CA). Monte Carlo minimisation method is a combination of random moves and energy minimization.[69] This method moves the ligand by random combination of translation, rotation, and torsional changes. The random move of ligand samples both the orientational and conformational spaces of the ligand with respect to the target protein. The Monte Carlo moves are useful in getting over energy barriers, while energy minimisation is for locating energy minima. Then, following the principles of statistical mechanics, Monte Carlo simulation produces acceptable different conformations of ligand-protein or protein-protein complexes, which will be then subjected to the SA calculation. The combination of such sampling method with SA allows to evaluate ligand-induced large scale conformational changes enhancing the fixing of the ligand into the binding site.[70]

6.2.2 Dynamic docking procedure

The dynamic docking procedure formally requires a reasonable starting structure. Accordingly, the starting complex was subjected to a preliminary energy minimisation to generate roughly docked starting structure (Steepest Descent algorithm, maximum RMSd = 1 kcal/Å; ε=1; Discover_3 Module of Insight2005). During these calculations, in order to fully explore all possible binding sites/modes (*i.e.*, ligand flexibility) and to investigate the putative large-scale ligand induced motions in the protein (*i.e.*, protein flexibility), the binding domain area was defined as the whole of protein target.

A Monte Carlo/minimisation approach for the random generation of a maximum of 20 acceptable complexes was used. During the first step, the ligand was moved by a random combination of translation, rotation, and torsional changes (Flexible_Ligand option, considering all rotatable bonds) to sample both the conformational space of the ligand and its orientation with respect to the protein target (MxRChange = 3 Å; MxAngChange = 180°). Moreover, van der Waals (vdW) and Coulombic terms were scaled to a factor of 0.1 to avoid very severe divergences in the Coulombic and vdW energies. If the energy of a complex structure resulting from random moves of the ligand was higher by the energy tolerance parameter than the energy of the last accepted structure, it was not accepted for minimisation. To ensure a wide variance of the input structures to be successively minimised, an energy tolerance value of 10^6 kcal mol^{-1}, from the previous structure has been used. After the energy minimisation step (conjugate gradient, 2500 iterations), the energy test, with an energy range of 50 kcal/mol, and a structure similarity check (rms tolerance = 0.3 kcal/Å) was applied to select the 20 acceptable structures. Each subsequent structure was generated from the last accepted structure. All the accepted complexes resulting from the Monte

Carlo/minimisation approach were subjected to a molecular dynamic simulated annealing protocol, including 5 ps of a dynamic run divided in 50 stages (100 fs each) during which the temperature of the system was linearly decreased from 500 to 300 K (Verlet velocity integrator; time step of 1.0 fs). In simulated annealing, the temperature is altered in time increments from an initial temperature to a final temperature. The temperature is changed by adjusting the kinetic energy of the structure (by rescaling the velocities of the atoms). Molecular dynamics calculations were performed using a constant temperature and constant volume (NVT) statistical ensemble, and the direct velocity scaling as temperature control method (temp window = 10 K). In the first stage, initial velocities were randomly generated from the Boltzmann distribution, according to the desired temperature, while during the subsequent stages initial velocities were generated from dynamics restart data. The temperature of 500 K was applied with the aim of surmounting torsional barriers, thus allowing an unconstrained rearrangement of the "ligand" and the "protein" binding site (initial vdW and Coulombic scale factors = 0.1). Successively temperature was linearly reduced to 300 K in 5 ps, and, concurrently, the vdW and Coulombic scale factors have been similarly increased from their initial values (0.1) to their final values (1.0). A final round of 10^5 minimisation steps (Conjugate Gradient algorithm)[219] followed the last dynamics steps, and the minimised structures were saved in a trajectory file. After this procedure, the resulting structures were analysed.

6.3 Allosteric Inhibitors of the GBP1-PIM1 Complex: Experimental Procedure

Conformational Analysis of the New Aza-podophyllotoxin Derivatives. The apparent pK_a values of aza-podophyllotoxin derivatives were calculated by using the ACD/pK_a DB, version 12.00, software (Advanced Chemistry Development Inc., Toronto, Canada).[220] All compounds were considered neutral in all calculations performed as a consequence of the estimation of percentage of neutral/ionised forms computed at pH 7.4 (physiological value) and pH 7.2 (cytoplasmic value) using the Handerson-Hasselbalch equation. The compounds were built using the Insight 2005 Builder Module (Accelrys Software Inc., San Diego, CA). Atomic potentials and charges were assigned using the CVFF force field.[130] The conformational space of the compounds was sampled using 200 cycles of SA and the dielectric constant of the vacuum ($\varepsilon = 1$) (for details see the procedure described in the paragraph 6.1 "*Sampling procedure of the conformational space of the ligands*"). Resulting conformers were grouped into families on the basis of their torsional angles and ranked by their potential energy values (*i.e.*, ΔE from the GM).

Bioinformatics Analysis. The experimentally determined structures of i) **1**-Tubulin complex (PDB ID: 1SA1), ii) **2**-TopoIIβ complex (PDB ID: 3QX3), (iii) GBP1 (PDB IDs: 2B8W, 2B92, 2BC9, 2D4H, 1DG3, 1F5N), and (iv) PIM1 (PDB IDs: 4MBL, 4K0Y, 4K18, 4K1B, 4JX7, 4JX3, 4GW8, 4ENX, 4ENY, 4DTK, 4AS0, 4ALW, 4ALV, 4ALU, 4A7C, 3VC4, 3VBW, 3VBX, 3VBY, 3VBV, 3VBT, 3VBQ, 3UIX, 3UMW, 3UMX, 3T9I, 3R00, 3R01, 3R02, 3R04, 3QF9, 3MA3, 3JXW, 3JY0, 3JYA, 3JPV, 3F2A, 3-DCV, 3CXW, 3CY2, 3CY3, 3C4E, 3BWF, 3BGP, 3BGQ, 3BGZ, 3A99, 2XIZ, 2XIY, 2XIX, 2XJ0, 2XJ2, 2XJ1, 2J2I, 2OI4, 2O3P, 2O63, 2O64, 2O65, 2OBJ, 2C3I, 2BZH, 2BZI, 2BZK, 2BZJ, 2BIK, 2BIL, 1YWV, 1YXS, 1YXT, 1YXU, 1YXV, 1YXX, 1YHS, 1YI3, 1YI4, 1XWS, 1XQZ, 1XR1) were downloaded from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/). Hydrogens were added to all the PDB structures assuming a pH of 7.2. These structures were analysed using Biopolymer and Homology Module of Insight 2005 (Accelrys, San Diego). Linear functional motifs present in i) Tubulin podophyllotoxin binding site (*i.e.*, within a 5 Å radius from any given ligand atom), ii) TopoIIβ etoposide binding site, iii) GBP1, and iv) PIM1 were identified by using the Eukaryotic Linear Motif server (http://elm.eu.org/),[51] a resource for predicting small functional sites in eukaryotic proteins. The identified motifs in GBP1 and PIM1 were logically intersected with the union of those present in the binding sites of Tubulin and Topoisomerase. This allowed us to identify the putative binding site of **3** in GBP1, characterised by three linear functional motifs present in Tubulin and/or TopoIIβ binding sites.

Modelling of GBP1 and PIM1. The molecular models of i) full-length human GBP1 in free form; ii) full-length human GBP1 in complex with GDP•AlF$_3$, and iii) full-length human PIM1 in free form structures were built as described below. All subsequent structural calculations were performed using the CVFF force field.[130]

Modelling of GBP1 in free form. The molecular model of full length human GBP1 in free form was built starting from the experimentally determined structure of full-length human free GBP1 (hGBP1^{FL}; PDB ID: 1DG3), which lacks four loops (aa63-73, 157-166, 190-193, and 244-256), the N-terminal region (aa1-5) and the C-terminal region (aa584-592). The sequence of 1DG3 was aligned with the sequence of hGBP1^{FL} downloaded from the UniProtKB/Swiss-Prot Data Bank (http://www.uniprot.org; entry P32455) by using the Multiple_Alignment algorithm (Homology Module, Accelrys, San Diego). Subsequently, the secondary structure prediction of the hGBP1^{FL} was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of the structurally conserved regions (SCRs) aa6-62, 74-156, 167-189, 194-243, and 257-583 of hGBP1^{FL} were accordingly assigned by the SCR-AssignCoords procedure (Homology Module) using 1DG3 as

template structure. The lacking loop segments aa63-73, 157-166, 190-193, and 244-256 were inserted by using the Generate Loops procedure. With the Generate Loops procedure, a peptide backbone chain is built between two conserved peptide segments using randomly generated values for all the loops' φ's and ψ's. The chain was defined starting from the N-terminal end of the loop being built; the Start and Stop Residues were defined as the SCR residues of the model protein at either end of the loop itself. The geometry about the base was described by the four distances between Cα and N-termini of the Start residue and the Cα and C-termini of the Stop Residues. In the process of closing the loop, the values for the generated φ's and ψ's are adjusted until the four distance criteria are met. Specifically, a function was defined for the distances in terms of the dihedral angles (Scale Torsions = 60). The differences between the desired distances and their current values were minimised using a linearised Lagrange multiplier method. After a series of 1000 iterations, the loop was closed, except in the case where the distances between the ends of the loop were not respected (Convergence = 0.05). The geometry at the base of the loop is then checked for proper chirality. Finally, the loops were screened on the basis of steric overlap violations. All loops that are found to have unacceptable contacts were rejected. Since successive calculations can correct some bad contacts, a fairly large overlap factor was used (Internal and External overlap = 0.6). A bump check of the 10 generated loops together with the evaluation of their conformational energy were used as selection criteria. The lowest conformational energy loop presenting no steric overlap with the rest of the protein, was selected. Finally, the coordinates of the N-terminal and C-terminal amino acids (aa1-5 and aa584-592, respectively) were assigned using the EndRepair command in Homology Module of Insight 2005. In particular, according to results obtained from the secondary structure prediction, the residues aa584-585 (with the standard geometry of an α-helix) and the residues 1-5 and 584-592 (with an extended chain conformation) were appended to the molecular model of hGBP1^{FL}. The molecular model of hGBP1^{FL} was completed inserting the water molecules of GBP1 experimentally determined structure (PDB ID: 1DG3) through the Modify/UnMerge and Modify/Merge commands (Biopolymer Module, Accelrys, San Diego). The entire GBP1 model was then subjected to a full energy minimisation within Insight 2005 Discover_3 Module (Steepest Descent algorithm, maximum RMSd = 10 kcal/Å; Conjugate Gradient algorithm, maximum RMSd = 1 kcal/Å; ε = 1). During the minimisation, only the whole disordered N- and C-terminal and loop backbone and side chains were left free to move, whereas the SCRs of GBP1 were fixed to avoid unrealistic results. Each step of refining procedure was followed by a structural check by using the Struct_Check command of the ProStat pull down in the Homology Module to verify the correctness of the geometry optimisation procedure before moving to the next step. Checks included φ, ψ, $\chi 1$, $\chi 2$, $\chi 3$ and ω dihedral angles, Cα virtual torsions, and Kabsch and Sander main chain H-bond energy evaluation. The final generated model was checked for quality using

Procheck structure evaluator software.[70] The obtained hGBP1^{FL} model was used for successive dynamic docking studies.

Modelling of GBP1 in complex with GDP•AlF₃. The molecular model of full-length human GBP1 in complex with GDP•AlF₃ was built using as template the experimentally determined structure of full-length human GBP1 in complex with the substrate analogue GppNHp (PDB ID: 1F5N), which lacks one loop (aa158-164), the N-terminal region (aa1-6) and the C-terminal region (aa584-592). The sequence of 1F5N was aligned with the sequence of hGBP1^{FL} downloaded from the UniProtKB/ Swiss-Prot Data Bank (http://www.uniprot.org; entry P32455) by using the Multiple_Alignment algorithm (Homology Module, Accelrys, San Diego). Then, the secondary structure prediction of the hGBP1^{FL} was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of the structurally conserved regions aa7-157 and aa165-583 of hGBP1^{FL} were accordingly assigned by the SCR-AssignCoords procedure of Homology using 1F5N as template structure. Thus, the obtained model of hGBP1^{FL}, in which the lacking loop segment aa158-164, was inserted by using the Generate Loops procedure (for the procedure see previous paragraph). Finally, the coordinates of the N-terminal and C-terminal (aa1-6 and aa584-592, respectively) were assigned using the EndRepair command in Homology Module of Insight 2005. In particular, the residues aa584-585 (with the standard geometry of an α-helix) and the residues aa1-6 and aa584-592 (with an extended chain conformation) were appended to the molecular model of hGBP1^{FL}. The obtained model of hGBP1^{FL} was completed inserting, through the Modify/UnMerge and Modify/Merge commands (Biopolymer Module, Accelrys, San Diego), i) GDP•AlF₃ and Mg^{2+} ion of hGBP1^{LG} experimentally determined structure (PDB ID: 2B92) and ii) the water molecules, combining those of hGBP1^{FL} in complex with the substrate analogue GppNHp (PDB ID: 1F5N) with those of hGBP1^{LG} in complex with GDP•AlF₃ (PDB ID: 2B92). The obtained model of GDP•AlF₃-bound hGBP1^{FL} was then subjected to a full energy minimisation within Insight 2005 Discover_3 Module (Steepest Descent algorithm, maximum RMSd = 10 kcal/Å; Conjugate Gradient algorithm, maximum RMSd = 1 kcal/Å; ε = 1). The partial charges to the nucleotide, were assigned by comparing partial charges assigned by CVFF force field with those estimated by MNDO[221] semi-empirical 1 SCF calculations for ionised nucleotide. In particular, CVFF force field partial charges were added to the algebraic difference between MNDO partial charges of the tri-anionic and di-anionic form for GDP. During the minimisation, quadratic distance or torsion restraints were applied on the SCRs to avoid unrealistic results. To identify the SCRs, the hGBP1^{FL} sequence (UniProtKB/Swiss-Prot Data Bank; entry code GBP1: P32455) was analysed using the Structure Prediction and Sequence Analysis Server Predict Protein (http://www.predictprotein.org/). In GBP1, sixteen α-helix and seven β-sheet secondary structures were predicted to be conserved (α0: aa25-33; α1: aa51-59; α2: aa112-123; α3: aa136-151; α3': aa169-172; α4:

aa198-207; a4': aa221-229; α5: aa260-272; α6: aa290-306; α7: aa312-343; α8: aa349-372; α9: aa375-426; α10: aa 431-450; α11: 456-479; α12: aa483-562; α13: aa566-585; β-1: aa11-15; β0: aa21-23; β1: aa38-44; β2: aa78-83; β3: aa92-97; β5: aa178-184; β6 aa235-238). Accordingly, for the α-helices, the distance between H-bond donors and acceptors was restrained within 2.5 Å by means of distance restraints. On the other hand, for the β-sheets, the φ and ψ torsional angles, according to the parallel or anti-parallel conformation, were restrained within -119° and +113°, or -139° and +135°, respectively (Restrain command; Discover_3 Module, Accelrys, San Diego). According to the reliability index values obtained from secondary structure prediction analysis, the following set of restraint force constants was used: i) force constants of 1 kcal/mol/Å^2-10 kcal/mol/Å^2 for reliability index values from 0 to 3, ii) force constants of 10 kcal/mol/Å^2-100 kcal/mol/Å^2 for reliability index values from 4 to 6, and iii) force constants of 100 kcal/mol/Å^2-1000 kcal/mol/Å^2 for reliability index values from 7 to 9. Moreover, during these calculations, the interaction distances among GDP•AlF$_3$, Mg^{2+} ion, the active site amino acids, and the active site water molecules were restrained to the values of reference crystal structure (*i.e.* PDB ID: 2B92) by means of quadratic distance restraints (100 kcal/mol/Å^2 -1000 kcal/mol/Å^2). Finally, GDP•AlF$_3$ and Mg^{2+} ion were tethered with a force constant of 100 kcal/mol/Å^2. GDP•AlF$_3$-bound hGBP1^{FL} model was subjected to the same structural check procedure previously described for the model of unbound GBP1, and then used for the subsequent dynamic docking studies.

Modelling of PIM1. The molecular model of full length human PIM1 was built using as template the experimentally determined structure of human PIM1 (hPIM1) in free form (PDB ID: 1XQZ), which lacks of N-terminal (aa1-122) and C-terminal (aa400-404) regions. The sequence of 1XQZ was aligned with the sequence of hPIM1 downloaded from the UniProtKB/Swiss-Prot Data Bank (http://www.uniprot.org; entry P11309) by using the Multiple_Alignment algorithm (Homology Module, Accelrys, San Diego). Secondly, the secondary structure prediction of the hPIM1 was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of i) the structurally conserved regions (aa123-399) of hPIM1 sequence were accordingly assigned by the SCR-AssignCoords procedure of Homology using 1XQZ as template structure and ii) the N-terminal and C-terminal amino acids (aa1-122 and aa400-404, respectively) were assigned using the EndRepair command of Homology Module. Finally, the obtained model of hPIM1 was completed inserting the water molecules of PIM1 experimentally determined structure (PDB ID: 1XQZ) through the Modify/UnMerge and Modify/Merge commands (Biopolymer Module, Accelrys, San Diego). PIM1 homology model was then subjected to the same full energy minimisation and structural check procedure previously described for the homology model of GBP1 in free form.

Docking studies. The docking studies on i) GBP1:PIM1 complex; ii) GBP1 in complex with GDP•AlF$_3$ and iii) GBP1 in Complex with **3** were performed as described below.

Docking studies on GBP1:PIM1 complex. In order to build a starting GBP1-PIM1 complex for dynamic docking studies, the experimentally determined structure of p110γ catalytic subunit in complex with H-Ras-1 (PDB ID: 1HE8) was downloaded from the Protein Data Bank (PDB; http://www. rcsb.org/pdb/). Hydrogens were added to the PDB structure assuming a pH of 7.2. This structure was analysed and compared with the obtained homology models of PIM1 and GBP1 (free form), by using Biopolymer and Homology Module of Insight 2005 (Accelrys, San Diego). In particular, a sequence alignment and a structural superimposition were performed between: i) the homology model of PIM1 and p110γ kinase domain (PDB ID: 1HE8) and ii) the homology model of GBP1 (free form) and H-Ras-1 (PDB ID: 1HE8). Water molecules overlapping the proteins in the obtained starting GBP1-PIM1 dimeric complex model were removed. The resulting structure was used for successive dynamic docking studies. The docking methodology is described in the paragraph 6.2 "Monte Carlo/Simulated Annealing based docking protocol". Although in the subsequent dynamic docking protocol all the systems were perturbed by means of Monte Carlo and simulated annealing procedures, nevertheless the dynamic docking procedure formally requires a reasonable starting structure. Accordingly, the starting model of GBP1-PIM1 complex was subjected to a preliminary energy minimisation to generate roughly docked starting structure (Steepest Descent algorithm, maximum RMSd = 10 kcal/Å; Conjugate Gradient algorithm, maximum RMSd = 1 kcal/Å; ε = 1). During the minimisation, the whole system was left free to move, whereas a quadratic distance or torsional restraints were applied on the SCRs to avoid unrealistic results (Restrain command; Discover_3 Module, Accelrys, San Diego). To identify the SCRs, we analysed the hGBP1^{FL} and hPIM1 sequences (UniProtKB/Swiss- Prot Data Bank; entry code GBP1: P32455; entry code PIM1: P11309) using the Structure Prediction and Sequence Analysis Server Predict Protein (http://www.predictprotein.org/). In GBP1, as reported in the paragraph "*Modelling of GBP1 in complex with GDP•AlF$_3$*," sixteen α-helix and seven β-sheet secondary structures were predicted to be conserved (α0: aa25-33; α1: aa51-59; α2: aa112-123; α3: aa136-151; α3': aa169-172; α4: aa198-207; α4': aa221-229; α5: aa260-272; α6: aa290-306; α7: aa312-343; α8: aa349-372; α9: aa375-426; α10: aa431-450; α11: aa456-479; α12: aa483-562; α13: aa566-585; β-1: aa11-15; β0: aa21-23; β1: aa38-44; β2: aa78-83; β3: aa92-97; β5: aa178-184; β6: aa235-238) In PIM1, seven α-helix and eight β-sheet secondary structures were predicted to be conserved (αC: aa178-187; αD: aa219-226; αE: aa232-251; αEF: aa300-304; αF: aa319-323; αH: aa352-362; αI: aa372-376; β1: aa129-134; β2: aa140-149; β3: aa153-161; β4: aa194-201; β5: aa206-213; β7: aa254-256; β8: aa264-267, β9: aa274-276). The parameters of the restraints applied on the SCRs during these

calculations were the same as those previously described in the paragraph "*Modelling of GBP1 in complex with GDP•AlF₃*". Moreover, in order to analyse the role of the water molecules within the binding sites of GBP1 and PIM1, during these calculations, also the water molecules within the active sites of GBP1 and PIM1 were tethered with a force constant of 100 kcal/mol/Å². In particular, all the water molecules within a 6 Å radius from any given nucleotide atom (*i.e.*, GppNHp for GBP1 (PDB ID: 1F5N); AppNHp for PIM1 (PDB ID: 1XR1)) were considered. Since the docking protocol included a Monte Carlo based conformational search of a "ligand" within a "protein", then the docking procedure was performed employing GBP1 as "ligand" and PIM1 as "protein". The binding domain area was defined as a subset including all residues of the "protein". All atoms included in the binding domain area were left free to move during the entire course of docking calculations, whereas, in order to avoid unrealistic results, a tethering restraint was applied on the SCRs of each protein. The set of restraints applied was the same as for the preliminary energy minimisation. The non-bonded interactions were calculated using Quartic_vdW_no_Coul method (vdW CUT_OFF: 5). The resulting docked structures were ranked by their conformational energy and were analysed by a structural check by using the Struct_Check command of the ProStat pulldown in the Homology Module to verify the correctness of their dihedral angles values. Structures characterised by unrealistic backbone geometry were discarded. Selected complexes were checked for quality using Procheck structure evaluator software.[70] In order to allow the whole relaxation of the proteins, the resulting docked complexes were then subjected to Molecular-mechanics (MM) energy minimisation within Insight 2005 Discover Module (Steepest Descent algorithm; ε = 1) firstly, until the maximum RMSd less than 0.5 kcal/Å using the set of structural restraints applied for docking calculations, secondly, until the maximum RMSd was less than 0.1 kcal/Å without structural restraints. Optimised complexes were checked for quality using Procheck structure evaluator software.[70] Finally, in order to test the thermodynamic stability of the resulting docked complexes, these latter were subjected also to a molecular dynamics SA protocol using the Cell_Multipole method for non-bonded interactions[69] and the dielectric constant of the water (ε = 80*r). A tethering restraint was applied on the SCRs of the complex and on the water molecules within the nucleotide binding sites. The set of structural restraints applied was the same as for previous docking calculations. In order to allow the whole relaxation of the complexes, the resulting annealed complexes were then subjected to Molecular-mechanics (MM) energy minimisation within Insight 2005 Discover Module using the same protocol reported for the resulting docked complexes. The optimised complexes were checked for quality using Procheck structure evaluator software.

Docking studies on GBP1 in complex with GDP•AlF₃. The starting GBP1-GDP•AlF$_3$ complex was subjected to dynamic docking studies (Affinity, SA_Docking; Insight 2005, Accelrys, San Diego). The applied docking and SA protocol as well as, the parameters were the same as those previously described in the "*Docking studies on GBP1-PIM1 complex*" paragraph. The docking protocol included a Monte Carlo-based conformational search of the "ligand" (*i.e.*, GDP•AlF$_3$) within the obtained homology model of GBP1. The binding domain area was defined as a subset including all residues of GBP1 protein (aa1-592). All atoms included in the binding domain area were left free to move during the entire course of docking calculations, whereas, in order to avoid unrealistic results, quadratic distance or torsional restraints were applied on the SCRs of the protein. The parameters of restraints applied on the SCRs were the same as for the preliminary energy minimisation (see paragraph "*Modelling of GBP1 in complex with GDP•AlF$_3$*"). Moreover, in order to preserve the key interactions among GDP•AlF$_3$, Mg^{2+} ion, the active site amino acids and the active site water molecules, during these calculations, GDP•AlF$_3$, Mg^{2+} ion and the atoms of amino acids or of water molecules involved in these interactions were tethered with a force constant of 100 kcal/mol/Å^2. The resulting docked and annealed complexes were checked for quality using Procheck structure evaluator software.

*Docking Studies on GBP1 in Complex with **3**.* The putative **3**-GBP1 complex was subjected to dynamic docking studies (Affinity, SA_Docking; Insight 2005, Accelrys, San Diego). The applied docking and simulated annealing protocol as well as, the parameters were the same as those previously described in the "*Docking studies on GBP1-PIM1 complex*" paragraph. The docking protocol included a Monte Carlo-based conformational search of the "ligand" (*i.e.*, **3**) within the obtained homology model of GBP1. The binding domain area was defined as a subset including all residues of GBP1 protein (aa1-592). All atoms included in the binding domain area were left free to move during the entire course of docking calculations, whereas, in order to avoid unrealistic results, quadratic distance or torsional restraints were applied on the SCRs of the protein. The parameters of restraints applied on the SCRs were the same as for the preliminary energy minimisation (see paragraph "*Modelling of GBP1 in complex with GDP•AlF$_3$*"). Moreover, in order to analyse the role of the water molecules within the binding site of GBP1, for each set of docking, two typologies of calculations were performed: one, where all water molecules were left free to move, another, where the water molecules within the active site of GBP1, were tethered with a force constant of 100 kcal/Å^2 as for the preliminary energy minimisation. The non-bonded interactions were calculated using three different methods: i) Quartic_vdW_no_Coul (vdW CUT_OFF: 5), ii) Cell_Multipole, and iii) Group Based (vdW and Coul CUT_OFF: 15). The resulting docked and annealed complexes were checked for quality using Procheck structure evaluator software.

6.4 Cationic Porphyrins as Tunable Gatekeepers of the 20S Proteasome: Experimental Procedure

Calculation of the chemical-physical properties of cationic porphyrins. The apparent pKa values were estimated by using the ACD/Percepta software.[220] The compounds were considered in their cationic form in all calculations performed, as a consequence of the estimation of percentage of neutral/ionised forms computed at pH 7.4 and pH 7.2 (physiological and cytoplasmic values, respectively) using the Handerson-Hasselbalch equation. Atomic potentials were assigned using the CVFF force field, while the partial charges were assigned using the partial charges estimated by MNDO semi-empirical 1 SCF calculations.[221] The conformational space of the compounds was sampled using 200 cycles of SA and the dielectric constant of the water ($\epsilon = 80*r$) (for details see the procedure described in the paragraph 6.1 "*Sampling procedure of the conformational space of the ligands*"). The resulting PM7 conformers were subsequently ranked by: i) conformational energy (ΔE from the GM < 5 kcal/mol), ii) inter-atomic distances between the charged nitrogen atoms, and iii) conformation of porphyrin ring.

Structural and bioinformatics analysis of 20S proteasome. The experimentally determined structures of 20S proteasome in free form and 20S human proteasome in complex with its regulatory particles (RPs), such as: i) 19S (PDB IDs: 5L4G (Cryo-EM; resolution 3.9 Å), 5GJQ (Cryo-EM; resolution 4.35 Å), 5GJR (Cryo-EM; resolution 3.5 Å), 5T0C (Cryo-EM; resolution 3.88 Å), 5T0G (Cryo-EM; resolution 4.4 Å), 5T0H (Cryo-EM; resolution 6.8 Å), 5T0I (Cryo-EM; resolution 8.0 Å), and 5T0J (Cryo-EM; resolution 8.0 Å)), ii) *Saccharomyces cerevisiae* 20S proteasome in complex with its RPs (*i.e.*, PA26, 19S and Blm10) and regulatory peptides (*i.e.*, PR-VI and Blm-Pep) (PDB IDs: 1Z7Q (X-ray diffraction; resolution 3.22 Å), 4V70 (X-ray diffraction; resolution 3.01 Å), 4CR2 (Cryo-EM; resolution 7.7 Å), 4CR3 (Cryo-EM; resolution 9.3 Å), 4CR4 (Cryo-EM; resolution 8.8 Å), 3JCO (Cryo-EM; resolution 4.8 Å), 3JCP (Cryo-EM; resolution 4.6 Å)), 4X6Z (X-ray diffraction; resolution 2.7 Å), and 4ZZG (X-ray diffraction; resolution 3.0 Å) were downloaded from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/). Hydrogens were added to all the PDB structures assuming a pH of 7.2. These structures were analysed using Biopolymer and Homology Modules of Insight 2005 (Accelrys, San Diego). In order to define the putative binding site of cationic porphyrins to the 20S proteasome, we analysed: i) the binding mode of inhibitors co-crystallised with 20S proteasome creating, for each inhibitor/enzyme complex, a subset around the ligand including all protein amino acids and water molecules having at least one atom within a 5 Å radius from any given ligand atom (Interface command of the Subset pulldown; Insight2005) and ii) the ionic interactions between 20S proteasome and its modulators creating, for each co-crystallised complex a subset around the 20S charged residues including all RP amino acids having at least one atom within a 5 Å radius

from any given 20S charged residue. In particular, in the analysis of yeast 20S proteasome in complex with the RPs PA26 and Blm10, in order to verify if such interactions are conserved between 20S human and its modulators, we considered both the X-ray structures of 20S yeast and homology models of 20S human (see "*Modelling of human 20S proteasome in the closed conformation*" and "*Modelling of human 20S proteasome in the open conformation*" paragraphs) superimposed on yeast 20S in complex with activators. In addition, a structural superimposition, where it was possible, between human (PA28, PDB ID:1AVO; X-ray diffraction; resolution 2.8 Å) and non human RPs (PA26, PDB ID: 1ZTQ). Then, the new identified charged residues were introduced in the yeast RPs (Replace Residue command; Biopolymer Module; Insight2005) and the putative ionic interactions were analysed using different charged residue side chain rotamers (Manual_Rotamer command; Biopolymer Module; Insight2005). Finally, all charged amino acids involved in these interactions were mapped on human 20S proteasome, in the closed and open conformation. Combining these results with those obtained from porphyrins pharmacophoric property analysis, the putative starting complexes were generated for the cationic porphyrins. In particular, we selected the gate as starting point for the docking studies of H_2T4, *meta*-H_2T4, and *ortho*-H_2T4 as well as the α4-α5 groove for the docking studies of **2.**

Modelling of 20S human proteasome. It must be mentioned here that, at the beginning of this research activity concerning the first class of cationic porphyrins (H_2T4, *meta*-H_2T4, *ortho*-H_2T4), the X-ray structure of the human 20S proteasome was not solved and only after our modelling studies it was determined and deposited in the Brookhaven Protein Data Bank (PDB ID:4R3O). All subsequent structural calculations were performed using the CVFF force field.[130]

Modelling of 20S human proteasome a1-a7 subunits. The molecular models of α1-α7 subunits of 20S human proteasome were built starting from the experimentally determined structure of 20S mouse proteasome (PDB ID: 3UNE). The sequences of 3UNE α1-α7 subunits were aligned with the sequences of 20S human proteasome α1-α7 subunits downloaded from the UniProtKB/Swiss-Prot Data Bank (http://www.uniprot.org; entry P60900 (α1); P25787 (α2); P25789 (α3); O14818 (α4); P28066 (α5); P25786 (α6) and P25788 (α7); by using the Multiple_Alignment algorithm, Homology Module, Accelrys, San Diego). Subsequently, the secondary structural prediction of 20S human proteasome α1- α7 subunits was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of the structurally conserved regions (α1: aa2-244; α2: aa2-233; α3: aa2-249; α4: aa2-240; α5: aa9-240; α6: aa4-241; α7: aa2-245) were accordingly assigned by the SCR-AssignCoords procedure (Homology Module, Insight 2005) using 3UNE as template structure. On

the other hand, the coordinates of the N- and C-terminal amino acids (α1: aa1 and aa245-246; α2: aa1 and aa234; α3: aa1 and aa250-261; α4: aa1 and aa241-248; α5: aa1-8 and aa241; α6: aa1-3 and aa242-263; α7: aa1 and 246-255) were assigned using the EndRepair command (Homology Module, Insight 2005). The obtained homology model was refined inserting the water molecules of experimentally determined structure 20S mouse proteasome α1-α7 subunits (PDB ID: 3UNE) through the UnMerge and Merge commands (Biopolymer Module, Accelrys, San Diego). The obtained homology model was then subjected to a full energy minimisation within Insight 2005 Discover_3 Module (Steepest Descent algorithm, maximum RMSd = 1 kcal/Å; ε=1). During the minimisation, only the whole disordered N- and C-terminals and the SCRs side chains were left free to move, whereas the SCRs backbone were fixed to avoid unrealistic results. Each step of refining procedure was followed by a structural check by using the Struct_Check command of the ProStat pulldown in the Homology Module to verify the correctness of the geometry optimisation procedure before moving to the next step. Checks included φ, ψ, χ1, χ2, χ3, and ω dihedral angles, Cα virtual torsions, and Kabsch and Sander main chain H-bond energy evaluation. The quality of the resulting complexes was then checked using Molprobity structure evaluator software[71] and compared to that of the reference PDB structure. The obtained homology model was used for successive dynamic docking studies (*i.e.,* H$_2$T4-20S complex, *meta*-H$_2$T4-20S complex and *ortho*-H$_2$T4-20S complex).

Modelling of human 20S proteasome in the closed conformation. The molecular model of 20S human proteasome in the closed conformation was built starting from the experimentally determined structure of 20S human proteasome (PDB ID: 4R3O), which lacks the N-terminal regions (α1: aa1; α2: aa1; α3: aa1; α4: aa1; α5: aa1-7; α6: aa1-3; α7: aa1), C-terminal regions (α1: aa246; α3: aa252-261; α4: aa245-248; α6: aa242-263; α7: aa247-255; β1: aa203-205; β2: aa221-234; β4: aa200-201; β5: aa202-204; β7: aa218-219), and the side chains of several residues (α1: aa2, 3, 45, 46, 55, 59, 62, 63, 146, 181,186-188, 196, 223, and 245; α2: aa4, 18, 36, 52-55, 60, 139-141, 143, 163, 176-178, 184, 193, 195, 196, 199, 200, 202, 226, 227, 230, and 233; α3: aa13, 37, 51, 52, 54, 177, 184, 194, 197, 201, 205, 210, 222, 229-231, 250, and 251; α4: aa27, 40, 42, 43, 46-50, 53, 54, 56, 61, 68, 84, 136-138, 141, 157, 163, 166, 169, 170, 174, 181-183, 185, 187-189, 192, 193, 195-199, 204, 207, 208, 211-213, 216, 218, 223-227, 229, and 231-244; α5: aa8, 86, 187, 192, 208, 209, and 231; α6: aa208, 217, 218, 237, 238, and 241; α7: aa109, 143, 203, 205, 207, 223, and 245; β1: aa201-202; β2: aa9, 180, 206, and 214; β3: aa47; β4: aa41, 95, 109, 185, 198, and 199; β5: aa106, 150, and 185; β6: aa45, 161, 166, and 200; β7: aa44, 156, 170, 195, and 216). The sequence of 4R3O was aligned with the sequences of 20S human proteasome downloaded from the UniProtKB/Swiss-Prot Data Bank (http://www.uniprot.org; entry P60900 (α1); P25787 (α2); P25789 (α3); O14818 (α4); P28066 (α5); P25786 (α6); P25788 (α7); P28072 (β1); Q99436 (β2); P49720 (β3); P49721 (β4); P28074 (β5); P20618 (β6), and P28070 (β7)) by using the

Multiple_Alignment algorithm (Homology Module, Accelrys, San Diego). Subsequently, the secondary structural prediction of 20S human proteasome was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of the structurally conserved regions (α1: aa2-245; α2: aa2-234; α3: aa2-251; α4: aa2-244; α5: aa8-241; α6: aa4-241; α7: aa2-246; β1: aa1-202; β2: aa1-220; β3: aa1-204; β4: aa1-199; β5: aa1-201; β6: aa1-213; β3: aa1-217) were accordingly assigned by the Structurally Conserved Regions (SCRs)-AssignCoords procedure (Homology Module, Insight 2005) using 4R3O as template structure. On the other hand, the coordinates of the lacking N-terminal and C-terminal amino acids were assigned using the EndRepair command (Homology Module, Insight 2005). The obtained homology model was completed inserting the missing residue side chains by using the Replace command (Biopolymer Module, Accelrys, San Diego), and the water molecules of experimentally determined structure 20S human proteasome (PDB ID: 4R69) through the UnMerge and Merge commands (Biopolymer Module, Accelrys, San Diego). The obtained homology model was then subjected to a full energy minimisation within Insight 2005 Discover_3 Module (Steepest Descent algorithm, maximum RMSd = 0.5 kcal/Å; ε = 80*r). During the minimisation, only the whole disordered N- and C- terminals and the SCRs side chains were left free to move, whereas the SCRs backbone were fixed to avoid unrealistic results. Each step of refining procedure was followed by a structural check by using the Struct_Check command of the ProStat pulldown in the Homology Module to verify the correctness of the geometry optimisation procedure before moving to the next step. Checks included φ, ψ, χ1, χ2, χ3, and ω dihedral angles, Cα virtual torsions, and Kabsch and Sander main chain H-bond energy evaluation. The quality of the resulting complexes was then checked using Molprobity structure evaluator software and compared to that of the reference PDB structure. The obtained homology model was used for successive dynamic docking studies (*i.e.,* **2**-20S complex).

Modelling of human 20S proteasome in the open conformation. According to the reference structure (PDB ID: 5T0J), it must be highlighted that the molecular model of 20S in the open conformation was built considering only two packed rings (*i.e.*, one α subunit ring and one β subunit ring). The molecular model of 20S human proteasome in the open conformation was built starting from the experimentally determined structure of 20S human proteasome (PDB ID: 5T0J), which lacks the N-terminal regions (α1: aa1-4; α2: aa1-2; α3: aa1; α4: aa1-2; α5: aa1-8; α6: aa1-3; α7: aa1-5), C-terminal regions (α1: aa245-246; α3: aa252-261; α4: aa241-248; α6: aa242-263; α7: aa246-255; β1: aa192-205; β2: aa221-234; β4: aa200-201; β5: aa202-204; β7: aa216-219), one loop region (α5: aa128-133), and the side chains of several residues (α1: aa45, 46, 55, 59, 63, 146, 181, 186-188, 196, and 223; α2: aa18, 36, 52-55, 60, 139-141, 143, 163, 176-178, 184, 193, 195, 196, 199, 200, 202, 226, 227, 230, and 233; α3: aa13, 37, 51, 52, 54, 177, 184, 194, 197, 201, 205, 210, 222,

229-231, and 250-251; α4: aa27, 40, 42, 46-50, 53, 136, 138, 141, 157, 163, 166, 169, 170, 174, 181-183, 185, 187, 189, 192-193, 195-199, 204, 207, 208, 212-213, 216, 218, 223, 225-227, 229, 232-237, and 239-240; α5: aa8, 86, 187, 192, 208-209, and 231; α6: aa208, 217-218, 237-238, and 241; α7: aa109, 143, 203, 205, 207, and 223; β2: aa9, 180, 206, and 214; β3: aa47; β4: aa41, 95, 109, 185, 198, and 199; β5: aa106, 150, and 185; β6: aa45, 161, 166, and 200; β7: aa44, 156, and 195).

The sequence of 5T0J was aligned with the sequences of 20S human proteasome downloaded from the UniProtKB/Swiss-Prot Data Bank (http://www.uniprot.org; entry P60900 (α1); P25787 (α2); P25789 (α3); O14818 (α4); P28066 (α5); P25786 (α6); P25788 (α7); P28072 (β1); Q99436 (β2); P49720 (β3); P49721 (β4); P28074 (β5); P20618 (β6), and P28070 (β7)) by using the Multiple_Alignment algorithm (Homology Module, Accelrys, San Diego). Subsequently, the secondary structural prediction of 20S human proteasome was performed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). The coordinates of the structurally conserved regions (α1: aa5-244; α2: aa3-234; α3: aa2-251; α4: aa3-240; α5: aa9-127 and aa134-241; α6: aa4-241; α7: aa6-245; β1: aa1-191; β2: aa1-220; β3: aa1-204; β4: aa1-199; β5: aa1-201; β6: aa1-213; β3: aa1-215) were accordingly assigned by the SCRs-AssignCoords procedure (Homology Module, Insight 2005) using 5T0J as template structure. The lacking loop segment in the α5 subunit (aa128-133) was inserted by using the Generate Loops procedure. With the Generate Loops procedure, a peptide backbone chain is built between two conserved peptide segments using randomly generated values for all the loops' φ's and ψ's. The chain was defined starting from the N-terminal end of the loop being built; the Start and Stop Residues were defined as the SCR residues of the model protein at either end of the loop itself. The geometry about the base was described by the four distances between Cα and N-termini of the Start residue and the Cα and C-termini of the Stop Residues. In the process of closing the loop, the values for the generated φ's and ψ's are adjusted until the four distance criteria are met. Specifically, a function was defined for the distances in terms of the dihedral angles (Scale Torsions: 60). The differences between the desired distances and their current values were minimised using a linearised Lagrange multiplier method. After a series of 1000 iterations, the loop was closed, except in the case where the distances between the ends of the loop were not respected (Convergence = 0.05). The geometry at the base of the loop is then checked for proper chirality. Finally, the loops were screened on the basis of steric overlap violations. All loops that are found to have unacceptable contacts were rejected. Since successive calculations can correct some bad contacts, a fairly large overlap factor was used (Internal and External overlap = 0.6). A bump check of the 10 generated loops together with the evaluation of their conformational energy were used as selection criteria. The lowest conformational energy loop presenting no steric overlap with the rest of the protein, was selected. Finally, the coordinates of the lacking N-terminal and C-terminal amino acids were assigned using the

EndRepair command (Homology Module, Insight 2005). The obtained homology model was completed inserting the missing residue side chains by using the Replace command (Biopolymer Module, Accelrys, San Diego). The obtained homology model was subjected to the same full energy minimisation and structural check procedure previously described in "*Modelling of human 20S proteasome in the closed conformation*", and, then, used for the subsequent dynamic docking studies. (*i.e.,* **2**-20S complex). During the minimisation, the whole disordered N- and C- terminals, the inserted loop region and the SCRs side chains were left free to move, whereas the SCRs backbone were fixed to avoid unrealistic results.

Docking studies on human 20S proteasome in complex with H₂T4, meta-H₂T4, ortho-H₂T4 and 2.
The putative starting complexes (H$_2$T4-20S human proteasome α1-α7 subunits; *meta*-H$_2$T4-20S human proteasome α1-α7 subunits, *ortho*-H$_2$T4-20S human proteasome α1-α7 subunits, and **2**-20S human proteasome in closed and open conformation) was subjected to dynamic docking studies. The docking methodology is described in the paragraph 6.2 "Monte Carlo/Simulated Annealing based docking protocol". During the Monte Carlo approach the dielectric constant of the vacuum ($\varepsilon = 1$) was used while, during the simulated annealing protocol, the dielectric constant of the water ($\varepsilon = 80*r$). The non-bonded interactions were calculated using the Cell Multipole method.[69] The binding domain area was defined as a subset including all residues of 20S human proteasome. All atoms included in the binding domain area were left free to move during the entire course of docking calculations, whereas, in order to avoid unrealistic results, a tethering restraint was applied on the SCRs of protein. To identify SCRs, the 20S human proteasome sequences were analysed using the Structure Prediction and Sequence Analysis server PredictProtein (http://www.predictprotein.org/). In α1 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa23-33; α2, aa85-104; α3, aa111-128; α4, aa172-183; α5, aa191-207; α6, aa232-243; β1, aa13-16; β2, aa38-43; β3, aa48-53; β4, aa68-72; β5, aa76-80; β6, aa135-143; β7, aa150-154; β8, aa160-168; β9, aa215-222; β10, aa226-229). In α2 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa20-30; α2, aa81-100; α3, aa107-124; α4, aa167-178; α5, aa184-198; α6, aa223-231; β1, aa9-13; β2, aa34-39; β3, aa44-49; β4, aa66-68; β5, aa72-76; β6, aa131-139; β7, aa145-149; β8, aa155-163; β9, aa208-214; β10, aa219-220). In α3 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa18-29; α2, aa80-100; α3, aa107-124; α4, aa168-178; α5, aa186-200; α6, aa230-248; β1, aa10-12; β2, aa33-38; β3, aa43-48; β4, aa66-68; β5, aa72-76; β6, aa131-139; β7, aa146-150; β8, aa157-164; β9, aa211-217; β10, aa224-227). In α4 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa17-27; α2, aa78-97; α3, aa104-121; α4, aa165-176; α5, aa183-198; α6, aa222-243; β1, aa6-10; β2, aa31-36; β3, aa41-46; β4, aa62-65; β5, aa69-73; β6, aa128-136; β7, aa143-147; β8, aa154-161; β9, aa206-212; β10, aa217-219). In α5

subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa22-32; α2, aa83-102; α3, aa109-120; α4, aa174-185; α5, aa191-206; α6, aa231-240; β1, aa11-15; β2, aa36-41; β3, aa46-51; β4, aa66-70; β5, aa74-78; β6, aa138-146; β7, aa152-156; β8, aa162-170; β9, aa215-221; β10, aa226-228). In α6 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa19-30; α2, aa79-98; α3, aa105-122; α4, aa165-176; α5, aa184-199; α6, aa226-236; β1, aa10-13; β2, aa35-39; β3, aa45-49; β4, aa62-66; β5, aa70-74; β6, aa129-137; β7, aa143-147; β8, aa154-161; β9, aa210-216; β10, aa221-223). In α7 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa22-32; α2, aa83-102; α3, aa109-124; α4, aa170-180; α5, aa187-202; α6, aa229-245; β1, aa13-15; β2, aa36-41; β3, aa46-51; β4, aa67-70; β5, aa74-78; β6, aa133-141; β7, aa148-152; β8, aa160-166; β9, aa212-219; β10, aa224-227). In β1 subunits, 5 α-helix and 11 β-sheet secondary structures were predicted to be highly conserved (α1, aa52-69; α2, aa76-90; α3, aa132-143; α4, aa149-166; α5, aa191-200; β1, aa2-8; β2, aa12-16; β3, aa25-27; β4, aa34-37; β5, aa41-45; β6, aa95-103; β7, aa110-114; β8, aa120-122; β9, aa124-128; β10, aa174-180; β11, aa185-189). In β2 subunits, 5 α-helix and 12 β-sheet secondary structures were predicted to be highly conserved (α1, aa52-67; α2, aa76-90; α3, aa131-142; α4, aa148-165; α5, aa194-201; β1, aa2-8; β2, aa12-17; β3, aa25-27; β4, aa33-37; β5, aa41-45; β6, aa95-103; β7, aa109-113; β8, aa123-127; β9, aa173-179; β10, aa184-186; β11, aa210-213; β12, aa215-225). In β3 subunits, 4 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa56-76; α2, aa83-97; α3, aa142-152; α4, aa159-175; β1, aa9-14; β2, aa18-24; β3, aa32-34; β4, aa41-44; β5, aa48-52; β6, aa103-111; β7, aa119-123; β8, aa134-138; β9, aa184-190; β10, aa194-199). In β4 subunits, 5 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa32-35; α2, aa50-70; α3, aa77-93; α4, aa136-147; α5, aa153-170; β1, aa3-8; β2, aa12-17; β3, aa26-27; β4, aa36-38; β5, aa42-46; β6, aa99-107; β7, aa114-118; β8, aa129-132; β9, aa178-184; β10, aa189-190). In β5 subunits, 6 α-helix and 11 β-sheet secondary structures were predicted to be highly conserved (α1, aa31-34; α2, aa49-69; α3, aa76-91; α4, aa132-143; α5, aa149-166; α6, aa189-200; β1, aa2-8; β2, aa12-16; β3, aa25-26; β4, aa35-37; β5, aa41-45; β6, aa96-104; β7, aa110-114; β8, aa120-122; β9, aa124-128; β10, aa174-180; β11, aa185-186). In β6 subunits, 4 α-helix and 11 β-sheet secondary structures were predicted to be highly conserved (α1, aa58-77; α2, aa85-99; α3, aa142-152; α4, aa168-184; β1, aa5-6; β2, aa11-16; β3, aa20-26; β4, aa34-36; β5, aa42-46; β6, aa50-54; β7, aa105-113; β8, aa120-124; β9, aa134-138; β10, aa193-199; β11, aa203-210). In β7 subunits, 6 α-helix and 10 β-sheet secondary structures were predicted to be highly conserved (α1, aa39-41; α2, aa57-77; α3, aa85-100; α4, aa143-154; α5, aa162-179; α6, aa207-216; β1, aa10-16; β2, aa20-25; β3, aa33-35; β4, aa42-45; β5, aa49-53; β6, aa107-115; β7, aa121-125; β8, aa136-139; β9, aa187-193; β10, aa198-199). Accordingly, for the alpha-helices, the distance between backbone H-bond donors and acceptors was restrained within 2.5 Å. On the other hand, for the β-sheets, the φ and ψ torsional angles, according to the parallel or anti-parallel

conformation, were restrained within -119° and +113°, or -139° and +135°, respectively (Restrain command; Discover_3 Module, Accelrys, San Diego). According to the reliability index values obtained from secondary structure prediction analysis, the following set of restraint force constants was used: i) force constants of 1 kcal/mol/Å^2-10 kcal/mol/Å^2 for reliability index values from 0 to 3, ii) force constants of 10 kcal/mol/Å^2-100 kcal/mol/Å^2 for reliability index values from 4 to 6, and iii) force constants of 100 kcal/mol/Å^2-1000 kcal/mol/Å^2 for reliability index values from 7 to 9. In order to allow the whole relaxation of the protein, the resulting annealed complexes were then subjected to MM energy minimisation without restraints (Steepest Descent algorithm; $\varepsilon = 80*r$) until the maximum RMSd was less than 0.5 kcal/Å (Discover Module; Insight 2005). The resulting docked and annealed structures were ranked by their conformational energy and analysed considering the non-bonded interaction energies between the ligand and the enzyme (vdW and electrostatic energy contribution; Group Based method; CUT_OFF = 10 and 100; $\varepsilon = 2*r$; Discover_3 Module of Insight2005). The complex with the best compromise between the non-bonded interaction energies obtained by Monte Carlo and SA calculations was selected as the structure representing the most probable binding mode.

6.5 Design of Redox-active Chemotherapeutics: Experimental Procedure

Conformational analysis of the new antimalarial endoperoxides. Apparent pK_a and log D values of new endoperoxides were calculated by using ACD/Percepta software.[220] Accordingly, percentage of neutral/ionised forms were computed at pH 7.4 (blood), at pH 7.2 (cytoplasm), and at pH 5.5 (*Pf* FV) using the Henderson-Hasselbalch.

The new endoperoxides were built, taking into account the prevalent ionic forms at the considered different pH values, using the Insight 2005 Builder Module (Accelrys Software Inc., San Diego). Since **9h**, **9r** and **9s** were tested as mixture of two diastereomers, both diastereomers of these compounds were built. Atomic potentials and charges were assigned using the CFF91 force field.[222] The conformational space of the compounds was sampled using 200 cycles of SA and the dielectric constant of the vacuum ($\varepsilon = 1$) (for details see the procedure described in the paragraph 6.1 "*Sampling procedure of the conformational space of the ligands*"). The resulting conformers were ranked by their potential energy values (*i.e.*, ΔE from the GM) and all conformers within 5 kcal mol^{-1} from the GM were classified on the basis of: i) 1,2-dioxane ring conformation; ii) intra-molecular H-bonds, and iii) distance between endoperoxide oxygens (O1 or O2) and possible partners for a "through space" intra-molecular radical shift ($\leq$ 3 Å). Then, the occurrence rates were calculated.

Docking procedure. In order to further investigate the role of the conformational behaviour on antimalarial activity, docking studies were carried out on **9c** di-protonated form and on **9r** diastereomer protonated form in complex with heme, using the docking methodology described in the paragraph 6.2 "Monte Carlo/Simulated Annealing based docking protocol" and the dielectric constant of the vacuum (ε = 1). Atomic potentials of heme were assigned using the Heme29.frc;[223] a force field including heme parameters and atomic partial charges were assigned using the QM method PM7. Heme apparent pK_a values were calculated by using ACD/Percepta software. Accordingly, one propionic chain was considered protonated and the heme net total charge was set at +1. Although during the subsequent dynamic docking protocol all the systems were perturbed by means of Monte Carlo and SA procedures, a reasonable starting structure is anyway required. Thus, the PM7 lowest energy conformer of **9c** and **9r** diastereomers meeting the hypothesised pharmacophoric requirements for antimalarial activity was selected as the ligand starting conformation. The ligand was then placed above the heme, taking as template the crystal structure of a peroxo-bridged heme-copper dinuclear complex (CSD code UKACIS) and considering the O2/O7 iron coordination complex. During the entire docking calculations, all atoms in the complex were left free to move with the exception of heme pyrrolic carbons, which were kept fixed and, in order to avoid unrealistic results, the distance between the iron and the interacting oxygens (*i.e.*, O2 and O7) was restrained within 2.0 Å (100-1000 kcal mol^{-1} Å^{-2}). The Cell_Multipole method[69] has been used to calculate the non-bonded interactions. The resulting annealed structures were ranked by their conformational energy. The complex with the best conformational energy was selected as structure representing the most probable binding mode. The selected complex was then subjected to a full geometry optimisation by semi-empirical calculations, using the QM method PM7[72] in the Mopac2012 package[73] and EF (Eigenvector Following routine)[74] as geometry optimisation algorithm. GNORM value was set to 0.01. To reach a full geometry optimisation the criteria for terminating all optimisations was increased by a factor of 100, using the keyword PRECISE.

List of References

1. Klebe G. Drug Design, Springer-Verlag Berlin, Heidelberg **2013**.

2. Santos R., Ursu O., Gaulton A., Bento A. P., Donadi R. S., Bologa C. G., Karlsson A., Al-Lazikani B., Hersey A., Oprea T. I., Overington J. P. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.*, **2017**, *16*, 19-34.

3. Palermo G., De Vivo M. Computational Chemistry for Drug Discovery; in *Encyclopedia of Nanotechnology*, **2015**, pp. 1-15.

4. Tamini N. A., Ellis P. Drug development: from concept to marketing! *Nephron Clin. Pract.*, **2009**, *113*, 125-131.

5. Di Masi J. A., Hansen R. W., Grabwoski H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.*, **2003**, *22*, 151-185.

6. Wilson and Gisvold's textbook of organic medicinal and pharmaceutical chemistry. - 12th ed. / edited by John M. Beale Jr., John H. Block, **2011**, 17-25.

7. Jorgensen W. L. The many roles of computation in drug discovery. *Science*, **2004**, *303*, 1813-1818.

8. Callaway E. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, **2015**, *525*, 172-174.

9. Leach A. R., Jhoti H., in *Structure-based Drug Discovery*, Springer, Berlin **2007**.

10. Lee C. H., Huang H. C., Juan H. F. Reviewing ligand-based rational drug design: the search for an ATP synthase inhibitor. *Int. J. Mol. Sci.*, **2011**, *12*, 5304-5318.

11. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E. *Nucleic Acids Res.*, **2000**, *28*, 235-242.

12. Marsh J. A., Teichmann S. A., Forman-Kay J. D. Probing the diverse landscape of protein flexibility and binding. *Curr. Opin. Struct. Biol.*, **2012**, *22*, 643-650.

13. Mittermaier A., Kay L. E. New tools provide new insights in NMR studies of protein dynamics. *Science*, **2006**, *312*, 224-228.

14. Villoutreix B. O., Lagorce D., Labb C. M., Sperandio O., Miteva M. A. One hundred thousand mouse clicks down the road: selected online resources supporting drug discovery collected over a decade. *Drug Discov. Today*, **2013**, *18*, 1081-1089.

15. Wilson G. L., Lill M. A. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med. Chem.*, **2011**, *3*, 735-750.

16. Kuhn B., Guba W., Hert J., Banner D., Bissantz C., Ceccarelli S., Haap W., Körner M., Kuglstatter A., Lerner C., Mattei P., Neidhart W., Pinard E., Rudolph M. G., Schulz-Gasch T., Woltering T., Stahl M. A

Real-World Perspective on Molecular Design. *J. Med. Chem.*, **2016**, *59*, 4087-4102.

17. Xiang Z. Advances in Homology Protein Structure Modeling. *Curr. Protein Pept. Sci.*, **2006**, 7, 217-227.

18. Levinthal, C. How to fold graciously; in *Mössbauer Spectroscopy in Biological Systems. Proceedings of a Meeting Held at Allerton House, Monticello, Illinois.* (eds. DeBrunner J.T.P & Munck E.) (University of Illinois Press, Chicago, IL, USA, **1969**.

19. Thiel K.A. Structure-aided drug design's next generation. *Nat. Biotechnol.*, **2004**, *22*, 513-519.

20. Burra P. V., Zhang Y., Godzik A., Stec B. Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. *Proc. Natl. Acad. Sci. U.S.A.*, **2009**, *106*, 10505-10510.

21. Monzon A. M., Zea D. J., Marino-Buslje C., Parisi G. Homology modeling in a dynamical world. *Protein Sci.*, **2017**; (doi:10.1002/pro.3274).

22. Deschamps N., Simþes-Pires C. A., Carrupt P. A., Nurisso A. How the flexibility of human histone deacetylases influences ligand binding: an overview. *Drug Discov. Today*, **2015**, *20*, 736-742.

23. Lange O. F., Lakomek N. A., Farès C., Schrçder G. F., Walter K. F., Becker S., Meiler J., Grubmïller H., Griesinger C., de Groot B. L. Accessing ns-micros side chain dynamics in ubiquitin with methyl RDCs. *Science*, **2008**, *320*, 1471-1475.

24. Koshland D. E. The Key–Lock Theory and the Induced Fit Theory. *Angew. Chem. Int. Ed. Engl.*, **1995**, *33*, 2375-2378.

25. Boehr D. D., Wright P. E. Biochemistry. How do proteins interact? *Science*, **2008**, *320*, 1429-1430.

26. Drews J. Drug discovery: a historical prospective. *Science*, **2000**, *287*, 1960-1964.

27. Klepeis J. L., Lindorff-Larsen K., Dror R. O., Shaw D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin. Struct. Biol.*, **2009**, *19*, 120-127.

28. Rohrdanz M. A., Zheng W., Clementi C. Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Ann. Rev. Phys. Chem.*, **2013**, *64*, 295-316.

29. Li D., Liu M. S., Ji B., Hwang K., Huang Y. Coarse-grained molecular dynamics of ligands binding into protein: The case of HIV-1 protease inhibitors. *J. Chem. Phys.*, **2009**, *130*, 215102.

30. Matsunaga Y., Fujisaki H., Terada T., Furuta T., Moritsugu K. Minimum free energy path of ligand-induced transition in adenylate kinase. *PLoS Comput. Biol.*, **2012**, *8*, e1002555.

31. Freddolino P. L., Harrison C. B., Liu Y., Schulten K. Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat. Phys.*, **2010**, *6*, 751-758.

32. Schafer N. P., Kim B. L., Zheng W., Wolynes P. G. Learning To Fold Proteins Using Energy Landscape Theory. *Isr. J. Chem.*, **2014**, *54*, 1311-1337.

33. Campbell A. J., Lamb M. L., Joseph-McCarthy D. Ensemble-based docking using biased molecular dynamics. *J. Chem. Inf. Model.*, **2014**, *54*, 2127-2138.

34. Hubner I. A., Deeds E. J., Shakhnovich E. I. Understanding ensemble protein folding at atomic detail. *Proc. Natl. Acad. Sci. U.S.A.*, **2006**, *103*, 17747-17752.

35. Zhuravlev P. I., Papoian G. A. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q. Rev. Biophys.*, **2010**, *43*, 295-332.

36. Zeller F., Zacharias M. Substrate Binding Specifically Modulates Domain Arrangements in Adenylate Kinase. *Biophys J.*, **2015**, *109*, 1978-1985.

37. Sugita Y., Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **1999**, *314*, 141-151.

38. Laio A., Parrinello M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, **2002**, *99*, 12562-12566.

39. Cavalli A., Spitaleri A., Saladino G., Gervasio F. L. Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Acc. Chem. Res.*, **2015**, *48*, 277-285.

40. Tsallis C., Stariolo D. A. Generalized simulated annealing. *Physica A*, **1996**, *233*, 395-406.

41. Zhou Y., Karplus M. Folding thermodynamics of a model three-helix-bundle protein. *Proc Natl Acad Sci U S A* 1997, 94, 14429-14432.

42. Proctor E. A., Ding F., Dokholyan N. V. Discrete molecular dynamics. *WIREs Comput. Mol. Sci.*, **2011**, *1*, 80-92.

43. Fujisaki H., Moritsugu K., Matsunaga Y., Morishita T., Maragliano L. Extended Phase-Space Methods for Enhanced Sampling in Molecular Simulations: A Review. *Front. Bioeng. Biotechnol.*, **2015**, *3*, 125.

44. Shirts M. R., Chodera J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, **2008**, *129*, 124105.

45. Bernardi R. C., Melo M. C., Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta.*, **2015**, *1850*, 872-877.

46. Pan A. C., Weinreich T. M., Piana S., Shaw D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.*, **2016**, *12*, 1360-1367.

47. Liu P., Kim B., Friesner R. A., Berne B. J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, *102*, 13749-13754.

48. Roe D. R., Bergonzo C., Cheatham T. E. Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J. Phys. Chem. B.*, **2014,** *118*, 3543-3552.

49. Huang Z., Wong C. F., Wheeler R. A. Flexible protein-flexible ligand docking with disrupted velocity

simulated annealing. *Proteins*, **2008**, *71*, 440-454.

50. Melo M. C., Bernardi R. C., Fernandes T. V., Pascutti P. G. GSAFold: a new application of GSA to protein structure prediction. *Proteins*, **2012**, *80*, 2305-2310.

51. Kannan S., Zacharias M. Simulated annealing coupled replica exchange molecular dynamics--an efficient conformational sampling method. *J. Struct. Biol.,* **2009**, *166*, 288-294.

52. Antunes D. A., Devaurs D., Kavraki L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.*, **2015**, *28*, 1-13.

53. Dinkel H., Van Roey K., Michael S., Davey N. E., Weatheritt R. J., Born D., Speck T., Krìger D., Grebnev G., Kuban M., Strumillo M., Uyar B., Budd A., Altenberg B., Seiler M., Chemes L.B., Glavina J., Sànchez I. E., Diella F., Gibson T. J. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.*, **2014**, *42*, D259-D266.

54. Van Roey K., Dinkel H., Gibson R. J., Weatheritt T. J., Davey N. E. The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.*, **2013**, *6*, rs7.

55. Neduva V., Russell R. B. Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **2005**, *579*, 3342-3345.

56. Cohen G. B., Ren R., Baltimore D. Modular binding domains in signal transduction proteins. *Cell*, **1995**, *80*, 237-248.

57. Persico M., Di Dato A., Orteca N., Fattorusso C., Novellino E., Andreoli M., Ferlini C. From Protein Communication to Drug Discovery. *Curr. Top. Med. Chem.*, **2015**, *15*, 2019-2031.

58. Johnson M. A., Maggiora G. M. Concepts and Applications of Molecular Similarity. Wiley, New York, **1990**.

59. Cramer R. D., Patterson D. E., Bunce J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, **1988**, *110*, 5959-5967.

60. Acharya C., Coop A., Polli J. E., Mackerell A. D. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput. Aided Drug. Des.*, **2011**; *7*, 10-22.

61. Marrero-Ponce Y., Santiago O. M., López Y. M., Barigye S. J., Torrens F. Derivatives in discrete mathematics: a novel graph-theoretical invariant for generating new 2/3-D molecular descriptors. I. Theory and QSPR application. *J. Comput. Aided. Mol. Des.*, **2012**, *26*, 1229-1246.

62. Ekins S., Mestres J., Testa B. *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.*, **2007**, *152*, 9-20.

63. Leach A. R. Molecular Modelling: Principles and Applications, 2nd edn. Prentice Hall, New York, **2001**.

64. Wermuth C. G. Pharmacophores: historical perspective and viewpoint from a medicinal chemist,

Pharmacophores and Pharmacophore Searches. Wiley-VCH Verlag GmbH & Co, Weinheim, Germany **2006**, pp. 1-13.

65. Wolber, G., Seidel T., Bendix F., Langer T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug. Discov. Today*, **2008**, *13*, 23-29.

66. Yang S. Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug. Discov. Today.*, **2010**, 15, 444-450.

67. Pei J., Kim B. H., Grishin N. V. PROMALS3-D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **2008**, *36*, 2295-2300

68. Persico M., Di Dato A., Orteca N., Cimino P., Novellino E., Fattorusso C. Use of integrated computational approaches in the search for new therapeutic agents. *Mol. Inform.*, **2016**, *35*, 309-325.

69. Senderowitz H., Guarnieri F., Still W. C. A Smart Monte Carlo Technique for Free Energy Simulations of Multi conformational Molecules. Direct Calculations of the Conformational Populations of Organic Molecules. *J. Amer. Chem. Soc.*, **1995**, *117*, 8211-8219.

70. Tuckerman M. E., Martyna G. J. Understanding Modern Molecular Dynamics: Techniques and Applications *J. Phys. Chem. B.*, **2000**, *104*, 159-178.

71. Laskowski R. A., MacArthur M. W., Moss D. S., Thornton J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **1993**, *26*, 283-291.

72. Davis I. W., Leaver-Fay A., Chen V. B., Block J. N., Kapral G. J., Wang X., Murray L.W., Arendall W. B., Snoeyink J., Richardson J.S., Richardson D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **2007**, *35*, W375-W383.

73. Stewart J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.*, **2013**, 19, 1-32.

74. Stewart J. P. MOPAC2012; Computational Chemistry: Colorado Springs, CO, **2012**; (http://OpenMOPAC.net).

75. Baker J. An algorithm for the location of transition state. *J. Comput. Chem.*, **1986**, *7*, 385-395.

76. Lipparini F., Mennucci B. Perspective: Polarizable continuum models for quantum-mechanical descriptions. *J. Chem. Phys.*, **2016**, *144*, 160901.

77. Cossi M., Scalmani G., Rega N., Barone V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution *J. Chem. Phys.*, **2002**,*117*, 43-54.

78. Scalmani G., Barone V., Kudin K. N., Pomelli C. S., Scuseria G. E., Frisch M. Achieving linear-scaling computational cost for the polarizable continuum model of salvation. *Theor. Chem. Acc.*, **2004**, *111*, 90-100.

79. Marshall G. R., Barry C. D., Bosshard H. E., Dammkoehler R. A., Dunn D. A. In The Conformational Parameter in Drug Design: the Active Analog Approach. In Computer-Assisted Drug Desig, vol. 112,

(Eds: E. C. Olson, R. E. Christoffersen), American Chemical Society, **1979**, pp. 205-225.

80. Persico M., Ramunno A., Maglio V., Franceschelli S., Esposito C., Carotenuto A., Brancaccio D., De Pasquale V., Pavone L. M., Varra M., Orteca N., Novellino E., Fattorusso C. New anticancer agents mimicking protein recognition motifs. *J. Med. Chem.*, **2013**, *56*, 6666-6680.

81. Pavadai E., El Mazouni F., Wittlin S., de Kock C., Phillips M. A., Chibale K. Identification of New Human Malaria Parasite Plasmodium Falciparum Dihydroorotate Dehydrogenase Inhibitors by Pharmacophore and Structure-Based Virtual Screening. *J. Chem. Inf. Model.*, **2016**, *56*, 548-562.

82. Bolton E., Wang Y., Thiessen P. A., Bryant S. H., in Annual Reports in Computational Chemistry, Vol. 4 (Eds: R. A Wheeler, D.C Spellmeyer), American Chemical Society, Washington, DC **2008**, pp. 217-241.

83. Wang Y., Xiao J., Suzek T. O., Zhang J., Wang J., Zhou Z., Han L., Karapetyan K., Dracheva S., Shoemaker B. A., Bolton E., Gindulyte A., Bryant S. H. PubChem's BioAssay Database. *Nucleic Acids Res.*, **2012**, *40*, D400-D412.

84. Raspaglio.G., Filippetti F., Prislei S., Penci R., De Maria I., Cicchillitti L., Mozzetti S., Scambia G., Ferlini C. Hypoxia induces class III beta-tubulin gene expression by HIF-1alpha binding to its 3' flanking region. *Gene,* **2008**, *409*,100-108.

85. Ferlini C., Raspaglio G., Cicchillitti L., Mozzetti S., Prislei S., Bartollino S., Scambia G. Looking at drug resistance mechanisms for microtubule interacting drugs: does TUBB3 work? *Curr. Cancer Drug Targets,* **2007**, *7*, 704-712.

86. Seve P., Dumontet C. Is class III beta-tubulin a predictive factor in patients receiving tubulin-binding agents? *Lancet Oncol.* **2008**, *9*, 168-175.

87. Mariani M., Shahabi S., Sieber S., Scambia G., Ferlini C. Class III β-tubulin (TUBB3): more than a biomarker in solid tumors? *Curr. Mol. Med.*, **2011**, *11*, 726-731.

88. De Donato M., Mariani M., Petrella L., Martinelli E., Zannoni G. F., Vellone V., Ferrandina G., Shahabi S., Scambia G., Ferlini C. Class III β-tubulin and the cytoskeletal gateway for drug resistance in ovarian cancer. *J. Cell. Physiol.*, **2012**, *227*, 1034-1041.

89. Katsetos C. D., Draber P. Tubulins as therapeutic targets in cancer: from bench to bedside. *Curr. Pharm. Des.*, **2012**, *18*, 2778-2792.

90. Praefcke G. J., McMahon H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell. Biol.,* **2004**, *5*, 133-147.

91. Prakash B., Praefcke G. J., Renault L., Wittinghofer A., Herrmann C. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*, **2000**, *403*, 567-571.

92. Ghosh A., Praefcke G. J. Renault L.; Wittinghofer A., Herrmann C. How guanylate-binding proteins achieve assembly stimulated processive cleavage of GTP to GMP. *Nature*, **2006**, *440*, 101-104.

93. Kunzelmann S., Praefcke G. J., Herrmann C. Transient kinetic investigation of GTP hydrolysis catalyzed by interferon-gamma induced hGBP1 (human guanylate binding protein 1). *J. Biol. Chem.,* **2006**, *281*, 28627-28635.

94. Praefcke G. J. Kloep S., Benscheid U., Lilie H., Prakash B., Herrmann C. Identification of residues in the human guanylate binding protein 1 critical for nucleotide binding and cooperative GTP hydrolysis. *J. Mol. Biol.,* **2004**, *344*, 257-269.

95. Schwemmle M., Staeheli, P. The interferon induced 67-kDa guanylate- binding protein (hGBP1) is a GTPase that converts GTP to GMP. *J. Biol. Chem.*, **1994**, *269*, 11299-11305.

96. Volkman B. F., Lipson D., Wemmer D. E., Kern D. Two-state allosteric behavior in a singledomain signaling protein. *Science,* **2001**, *291*, 2429-2433.

97. Prakash B., Abdullah N., Srinivasan B., Modiano N., Cresswell P., Sau A. K. Role of individual domains and identification of internal gap in human guanylate binding protein-1. *Mol. Biol.*, **2009**, *386*, 690-703.

98. Vöpel T., Syguda A., Britzen-Laurent N.,. Kunzelmann S, Lüdemann M. B., Dovengerds C., Stürzl M., Herrmann C. Mechanism of GTPase-activity-induced self-assembly of human guanylate binding protein 1. *J. Mol. Biol.,* **2010**, *400*, 63-70.

99. Syguda A., Bauer M., Benscheid U., Ostler N., Naschberger E., Ince S., Sturzl M., Herrmann C. Tetramerization of human guanylate-binding protein 1 is mediated by coiled-coil formation of the C-terminal α-helices. *FEBS J.,* **2012**, *279*, 2544-2554.

100. Vöpel T., Kunzelmann S., Herrmann C. Nucleotide dependent cysteine reactivity of hGBP1 uncovers a domain movement during GTP hydrolysis. *FEBS Lett.,* **2009**, *583*, 1923-1927.

101. Wells J. A, McClandon C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **2007**, *450*, 1001-1009.

102. Weskamp N., Huellermeier E., Klebe G. Merging chemical and biological space: Structural mapping of enzyme binding pocket space. *Proteins,* **2009**, *76*, 317-330.

103. Kortagere S., Krasowski M. D., Ekins S. The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.*, **2009**, *30*, 138-147.

104. Kellenberger E., Hofmann A., Quinn R. J. Similar interactions of natural products with biosynthetic enzymes and therapeutic targets could explain why nature produces such a large proportion of existing drugs. *Nat. Prod. Rep.,* **2011**, *28*, 1483-1492.

105. Brylinski M., Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U. S. A.,* **2008**, *105*, 129-134.

106. Andreoli M., Persico M., Kumar V., Pepe A, Kumar A., Alegria A.E., Petrella L., Sevciunaite L., Camperchioli A., Mariani M., Scambia G., Orteca N., Di Dato A., Novellino E., Malhotra S.V., Ferlini C.,

Fattorusso C. 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. *J. Med. Chem.*, **2014**, *57*, 7916-7932.

107. Deng J. Z. Newman D. J., Hecht S. M. Use of COMPARE analysis to discover functional analogues of bleomycin. *J. Nat. Prod.,* **2000**, *63*, 1269-1272.

108. Leroy D., Kajava A. V., Frei C., Gasser S. M. Analysis of etoposide binding to subdomains of human DNA topoisomerase II alpha in the absence of DNA. *Biochemistry*, **2001**, *40*, 1624-34.

109. Vigil D., Cherfils J., Rossman K. L., Der C. J. Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy? *Nat. Rev. Cancer*, **2010**, *10*, 842-57.

110. Abdullah N., Srinivasan B., Modiano N., Cresswell P., Sau A. K. Role of individual domains and identification of internal Gap in human Guanylate Binding Protein-1 (hGBP1). *J. Mol. Bio.* **2009**, *386*, 690 - 703.

111. Persico M., Petrella L., Orteca N., Di Dato A., Mariani M., Andreoli M., De Donato M., Scambia G., Novellino E., Ferlini C., Fattorusso C. GTP is an allosteric modulator of the interaction between the guanylate-binding protein 1 and the prosurvival kinase PIM1. *Eur. J. Med. Chem.,* **2015**, *91*, 132-144.

112. Ferlini C., Cicchillitti L., Raspaglio G., Bartollino S., Cimitan S., Bertucci C., Mozzetti S., Gallo D., Persico M., Fattorusso C., Campiani G., Scambia G. Paclitaxel directly binds to Bcl-2 and functionally mimics activity of Nur77. *Cancer. Res.*, **2009**, *69*, 6906-6914.

113. Sherman M. Y., Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron*, **2001**, *29*, 15-32.

114. Coux O., Tanaka K., Goldberg A. L. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.*, **1996**, *65*, 801-847.

115. Baumeister W., Walz J., Zühl F., Seemüller E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **1998**, *92*, 367-380.

116. Glickman M. H., Rubin D. M., Coux O., Wefes I., Pfeifer G., Cjeka Z., Baumeister W., Fried V. A., Finley D. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3.*Cell*, **1998**, *94*, 615-623.

117. McNaught K. S., Olanow C. W., Halliwell B., Isacson O., Jenner P. Failure of the ubiquitin-proteasome system in Parkinson's disease. *Nat Rev Neurosci.*, **2001**, *2*, 589-594.

118. Arendt C. S., Hochstrasser M. Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. U. S. A.*, **1997**, *94*, 7156-7161.

119. Heinemeyer W., Fischer M., Krimmer T., Stachon U., Wolf D. H. The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.*, **1997**, *272*, 25200-

25209.

120. Kisselev A. F., Callard A., Goldberg A. L. Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *J. Biol. Chem.*, **2006**, *281*, 8582-8590.

121. Lehmann A., Jechow K., Enenkel C. Blm10 binds to pre-activated proteasome core particles with open gate conformation. *EMBO Rep.*, **2008**, *9*, 1237-1243.

122. Obeng E. A., Carlson L. M., Gutman D. M., Harrington W. J. Jr., Lee K. P., Boise L. H. Proteasome inhibitors induce a terminal unfolded protein response in multiple myeloma cells. *Blood*, **2006**, *107*, 4907-4916.

123. Nawrocki S. T., Carew J. S., Dunner Jr. K., Boise L. H., Chiao P. J., Huang P., Abbruzzese J. L., McConkey D. J. Bortezomib inhibits PKR-like endoplasmic reticulum (ER) kinase and induces apoptosis via ER stress in human pancreatic cancer cells. *Cancer Res.*, **2005**, *65*, 11510-11519.

124. Bianchi G., Oliva L., Cascio P., Pengo N., Fontana F., Cerruti F., Orsi A., Pasqualetto E., Mezghrani A., Calbi V., Palladini G., Giuliani N., Anderson K. C., Sitia R., Cenci S. The proteasome load versus capacity balance determines apoptotic sensitivity of multiple myeloma cells to proteasome inhibition. *Blood*, **2009**, *13*, 3040-3049.

125. Kuhn D. J., Orlowski R. Z., Bjorklund C. C. Second generation proteasome inhibitors: carfilzomib and immunoproteasome-specific inhibitors (IPSIs). *Curr. Cancer Drug Targets*, **2011**, *11*, 285-295.

126. Kumar S., Rajkumar S. V. Many facets of bortezomib resistance/susceptibility. *Blood*, **2008**, *112*, 2177-2178.

127. Santoro A. M., Lo Giudice M. C., D'Urso A., Lauceri R., Purrello R., Milardi D. Cationic Porphyrins Are Reversible Proteasome Inhibitors. *J. Am. Chem. Soc.,* **2012**, *134*, 10451-10457.

128. Rodgers K. J., Dean R. T. Assessment of proteasome activity in cell lysates and tissue homogenates using peptide substrates. *Int. J. Biochem. Cell. Biol.,* **2003**, *35*, 716-727.

129. Santoro A. M., Cunsolo A., D'Urso A., Sbardella D., Tundo G. R., Ciaccio C., Coletta M., Diana D., Fattorusso R., Persico M., Di Dato A., Fattorusso C., Milardi D., Purrello R. Cationic porphyrins are tunable gatekeepers of the 20S proteasome. *Chem. Sci.,* **2016**, *7*, 1286-1297.

130. Dauber-Osguthorpe P., Roberts V.A., Osguthorpe D.J., Wolff J., Genest M., Hagler A.T. Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins,* **1998**, *4*, 31-47.)

131. Ding H. Q., Karasawa N., Goddard W. A. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London non-bond interactions. *J. Chem. Phys.*, **1992**, *97*, 4309-4315.

132. Whitby F. G., Masters E. I., Kramer L., Knowlton J. R., Yao Y., Wang C. C., Hill C. P. Structural basis

for the activation of 20S proteasomes by 11S regulators. *Nature*, **2000**, *408*, 115-120.

133. Jankowska E., Gaczynska M., Osmulski P. A., Sikorska E., Rostankowski R., Madabhushi S., Tokmina- Lukaszewska M., Kasprzykowski F. Potential allosteric modulators of the proteasome activity. *Biopolymers*, **2010**, *93*, 481-495.

134. Witkowska J., Karpowicz P., Gaczynska M., Osmulski P. A., Jankowska E. Dissecting a role of a charge and conformation of Tat2 peptide in allosteric regulation of 20S proteasome. *J. Pept. Sci.*, **2014**, *20*, 649-656.

135. Groll M., Brandstetter H., Bartunik H. D., Bourenkow G., Huber R. Investigations on the maturation and regulation of archaebacterial proteasomes. *J. Mol. Biol.*, **2003**, *327*, 75-83.

136. Ruschak A. M., Kay L. E. Proteasome allostery as a population shift between interchanging conformers. *Proc Natl Acad Sci USA.*, **2012**, *109*, E3454-3462.

137. Osmulski, P. A., Hochstrasser, M., Gaczynska M. A tetrahedral transition state at the active sites of the 20S proteasome is coupled to opening of the alpha-ring channel. *Structure*, **2009**, *17*, 1137-1147.

138. Li J., Gao X., Ortega J., Nazif T. Joss L., Bogyo M., Steven A., Rechsteiner M. Lysine 188 substitutions convert the pattern of proteasome activation by REGγ to that of REGs α and β. *EMBO J.*, **2001**, *20*, 3359-3369.

139. Yao Y., Huang L., Krutchinsky A., Wong M. L., Standing K. G., Burlingame A. L., Wang C. C. Structural and functional characterizations of the proteasome-activating protein PA26 from Trypanosoma brucei. *J. Biol. Chem.*, **1999**, *274*, 33921-33923.

140. Ustrell V., Hoffman L., Pratt G., Rechsteiner M. PA200, a nuclear proteasome activator involved in DNA repair. *EMBO J.*, **2002**, *21*, 3516-3525.

141. Hoffman L., Rechsteiner M. Activation of the multicatalytic protease. The 11S regulator and 20S ATPase complexes contain distinct 30-kilodalton subunits. *J. Biol. Chem.,* **1994**, *269*, 16890-16895.

142. Chen S., Wu J., Lu Y., Ma Y. B., Lee B. H., Yu Z., Ouyang Q., Finley D. J., Kirschner M. W., Mao, Y. Structural basis for dynamic regulation of the human 26S proteasome. *Proc Natl Acad Sci U S A.*, **2016**, *113*, 12991-12996.

143. Karpowicz P., Osmulski P. A., Witkowska J., Sikorska E., Giżyńska M., Belczyk-Ciesielska A., Gaczynska M. E., Jankowska E. Interplay between Structure and Charge as a Key to Allosteric Modulation of Human 20S Proteasome by the Basic Fragment of HIV-1 Tat Protein. *PLoS One.*, **2015**, *10*, e0143038.

144. Steinbach P. J., Brooks B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comp. Chem.,* **1994**, *15*, 667-683.

145. Groll M., Bajorek M., Köhler A., Moroder L., Rubin D. M., Huber R., Glickman M. H., Finley D. A gated channel into the proteasome core particle. *Nat Struct Biol.*, **2000**, *7*, 1062-1067.

146. Kohler A., Bajorek M., Groll M., Moroder L., Rubin D. M., Huber R., Glickman M., Finley D. The substrate translocation channel of the proteasome. *Biochimie*, **2001**, *83*, 325-332.

147. Gaczynska M., Osmulski P. A. Harnessing Proteasome Dynamics and Allostery in Drug Design. *Antioxid. Redox Signal.*, **2014**, *21*, 2286-2301.

148. Kmiecik S., Gront D., Kolinski M., Wieteska L., Dawid A. E., Kolinski A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.*, **2016**, *116*, 7898-7936.

149. Sfriso P., Hospital A., Emperador A., Orozco M. Exploration of conformational transition pathways from coarse-grained simulations. *Bioinformatics*, **2013**, *29*, 1980-1986.

150. Sfriso P., Emperador A., Orellana L., Hospital A., Gelpí J. L., Orozco M. Finding conformational transition pathways from discrete molecular dynamics simulations. *J. Chem. Theory Comput.*, **2012**, *8*, 4707-4718.

151. Taketomi H., Ueda Y., Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Pept. Protein Res.*, **1975**, *7*, 445-459.

152. Ueda Y., Taketomi H., Gō N. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. Biopolymers, **1978**, *17*, 1531-1548.

153. Emperador A., Carrillo O., Rueda M., Orozco M. Exploring the suitability of coarse-grained techniques for the representation of protein dynamics. *Biophys J.*, **2008**, 95, 2127-2138.

154. Orellana L., Rueda M., Ferrer-Costa C., Lopez-Blanco J. R., Chacón P., Orozco M. Approaching elastic network models to molecular dynamics flexibility. *J. Chem. Theory Comput.*, **2010**, *6*, 2910-2923.

155. Di Dato A., Cunsolo A., Persico M., Santoro A. M., D'Urso A., Milardi D., Purrello R., Stefanelli M., Paolesse R.., Tundo G. R., Sbardella D., Fattorusso C., Coletta M. Electrostatic map of proteasome α-rings encodes the design of allosteric porphyrin-based inhibitors able to affect 20s conformation by cooperative binding. *Sci. Rep.*, **2017**. In Press; Accepted Manuscript.

156. Wenzel T., Baumeister W. Conformational constraints in protein degradation by the 20S proteasome. *Nat Struct Biol.,* **1995**, *2*, 199-204.

157. Abraham M. J., van der Spoel D., Lindahl E., Hess B. and the GROMACS development team, GROMACS User Manual version 5.0, www.gromacs.org (**2014**).

158. Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010; 1950 -1958.

159. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. *LINCS:* a linear constraint solver for molecular simulations. *J. Comput. Chem.* (1997) **18**, 1463-1472.

160. Jorgensen W. L., Chandrasekhar J., Madura J. D. Comparison of simple potential functions for

simulating liquid water. *J. Chem. Phys.*, **1983**, *79*, 926.

161. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **1984**, *81*, 3684.

162. Parrinello M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **1981**, *52*, 7182.

163. Balsera M. A., Wriggers W., Oono Y., Schulten K. Principal component analysis and long-time protein dynamics. *J. Phys. Chem.,* **1996**, *100*, 2567-2572.

164. Wondrak G. T. Redox-directed cancer therapeutics: molecular mechanisms and opportunities. *Antioxid. Redox Signal.*, **2009**, *11*, 3013-3069.

165. Goyal M., Alam A., Bandyopadhyay U. Redox regulation in malaria: current concepts and pharmacotherapeutic implications. *Curr. Med. Chem.*, **2012**, *19*, 1475-1503.

166. Baik M. H., Friesner R. A. Computing Redox Potentials in Solution: Density Functional Theory as A Tool for Rational Design of Redox Agents. *J. Phys. Chem. A*, **2002**, *106*, 7407-7412.

167. Donkers R. L., Workentin M. S. First Determination of the Standard Potential for the Dissociative Reduction of the Antimalarial Agent Artemisinin. *J. Phys. Chem. B*, **1998**, *102*, 4061-4063.

168. Johann L., Lanfranchi D. A., Davioud-Charvet E., Elhabiri M. A physico-biochemical study on potential redox-cyclers as antimalarial and anti-schistosomal drugs. *Curr. Pharm. Des.*, **2012**, *18*, 3539-3566.

169. Czechtizky W., Hamley P. in Small Molecule Medicinal Chemistry: Strategies and Technologies, Wiley, Hoboken **2015**.

170. Dunitz J. D., Burgi H.B. in Structure Correlation, VCH, Weinheim, **1994**. Leach A. R. in Reviews in Computational Chemistry, Vol. 2, (Eds: K. B. Lipkowitz, D. B. Boyd), VCH, New York **1991**, pp. 1-5.

171. Agrafiotis D. K., Gibbs A. C., Zhu F., Izrailev S., Martin E. Conformational sampling of bioactive molecules: a comparative study. *J. Chem. Inf. Model.*, **2007**, *47*, 1067-1086.

172. WHO, World Malaria Report **2015**, ISBN: 978924156515 8.

173. Arav-Boger R., Shapiro T. A. Molecular mechanisms of resistance in antimalarial chemotherapy: the unmet challenge. *Annu. Rev. Pharmacol. Toxicol.*, **2005**, *45*, 565-585.

174. Nepveu F., Turrini F. Targeting the redox metabolism of Plasmodium falciparum. *Future Med. Chem.*, **2013**, *5*, 1993-2006.

175. Li Q., Hickman M. Toxicokinetic and toxicodynamic (TK/TD) evaluation to determine and predict the neurotoxicity of artemisinins. Toxicology, **2011**, *279*, 1-9.

176. Dondorp A. M., Yeung S., White L., Nguon C., Day N. P. J., Socheat D., von Seidlein L. Artemisinin resistance: current status and scenarios for containment. *Nat. Rev. Microbiol.*, **2010**, *8*, 272-280.

177. Cafieri F., Fattorusso E., Taglialatela-Scafati O., Ianaro A. Metabolites from the sponge plakortis
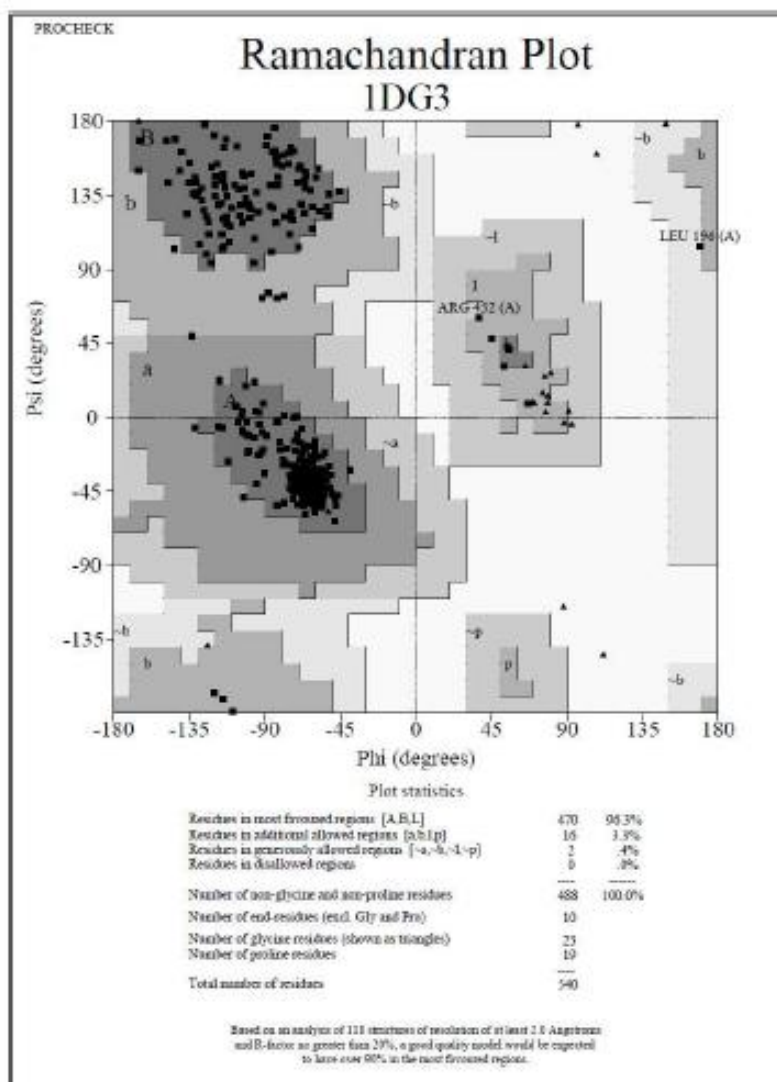
simplex. *Tetrahedron*, **1999**, *55*, 7045-7056.

178. Fattorusso E., Parapini S., Campagnuolo C., Basilico N., Taglialatela-Scafati O., Taramelli D. Activity against Plasmodium falciparum of cycloperoxide compounds obtained from the sponge Plakortis simplex. *J. Antimicrob. Chemother.*, **2002**, *50*, 883-888.

179. Campagnuolo C., Fattorusso E., Romano A., Taglialatela-Scafati O., Basilico N., Parapini S., Taramelli D. Antimalarial polyketide cycloperoxides from the marine sponge Plakortis simplex. *Eur. J. Org. Chem.*, **2005**, *23*, 5077-5083.

180. Fattorusso C., Campiani G., Catalanotti B., Persico M., Basilico N., Parapini S., Taramelli D., Campagnuolo C., Fattorusso E., Romano A., Taglialatela-Scafati O. Endoperoxide derivatives from marine organisms: 1,2-dioxanes of the plakortin family as novel antimalarial agents. *J. Med. Chem.*, **2006**, *49*, 7088-7094.

181. Jefford C. W. Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. *Curr. Med. Chem.*, **2001**, *15*, 1803-1826.

182. Jefford C. W. New developments in synthetic peroxidic drugs as artemisinin mimics. *Drug Discov. Today*, **2007**, *12*, 487-495.

183. Robert A., Benoit-Vical F., Dechy-Cabaret O., Meunier B. From classical antimalarial drugs to new compounds based on the mechanism of action of artemisinin. *Pure Appl. Chem.*, **2001**, *73*, 1173-1188.

184. Fattorusso C., Persico M., Calcinai B., Cerrano C., Parapini S., Taramelli D., Novellino E., Romano A., Scala F., Fattorusso E., Taglialatela-Scafati O. Manadoperoxides A-D from the Indonesian sponge Plakortis cfr. simplex. Further insights on the structure-activity relationships of simple 1,2-dioxane antimalarials. *J. Nat. Prod.*, **2010**, *73*, 1138-1145.

185. Fattorusso C., Persico M., Basilico N., Taramelli D., Fattorusso E, Scala F., Taglialatela-Scafati O. Antimalarials based on the dioxane scaffold of plakortin. A concise synthesis and SAR studies. *Bioorg. Med. Chem.*, **2011**, *19*, 312-320.

186. Chianese G., Persico M., Yang F., Lin H. W., Guo Y. W., Basilico N., Parapini S., Taramelli D., Taglialatela-Scafati O., Fattorusso C. Endoperoxide polyketides from a Chinese Plakortis simplex: further evidence of the impact of stereochemistry on antimalarial activity of simple 1,2-dioxanes. *Bioorg. Med. Chem.*, **2014**, *22*, 4572-4580.

187. Persico M., Quintavalla A., Rondinelli F., Trombini C., Lombardo M., Fattorusso C., Azzarito V., Taramelli D., Parapini S., Corbett Y., Chianese G., Fattorusso E., Taglialatela-Scafati O. A new class of antimalarial dioxanes obtained through a simple two-step synthetic approach: rational design and structure-activity relationship studies. *J. Med. Chem.*, **2011**, *54*, 8526-8540.

188. Persico M., Parapini S., Chianese G., Fattorusso C., Lombardo M., Petrizza L., Quintavalla A.,

Rondinelli F., Basilico N., Taramelli D., Trombini C., Fattorusso E., Taglialatela-Scafati O. Further optimization of plakortin pharmacophore: structurally simple 4-oxymethyl-1,2-dioxanes with promising antimalarial activity. *Eur. J. Med. Chem.*, **2013**, *70*, 875-886.

189. Lombardo M., Sonawane D. P., Quintavalla A., Trombini C., Dhavale D. D., Taramelli D., Corbett Y., Rondinelli F., Fattorusso C., Persico M., Taglialatela-Scafati O. Optimized Synthesis and Antimalarial Activity of 1,2-Dioxane-4-carboxamides. *Eur. J. Org. Chem.*, **2014**, *2014*, 1607-1614.

190. Coleman R. E., Nath A.K., Schneider I., Song G. H., Klein T. A., Milhous W. K. Prevention of sporongony of Plasmodium falciparum and P. Berghei in Anopheles stephensi mosquitoes by transmission-blocking antimalarials. *Am. J. Trop. Med. Hyg.*, **1994**, *50*, 646-653.

191. Charman S. A., Arbe-Barnes S., Bathurst I. C., Brun R., Campbell M., Charman W. N., Chiu F. C., Chollet J., Craft J. C., Creek D. J., Dong Y., Matile H., Maurer M., Morizzi J., Nguyen T., Papastogiannidis P., Scheurer C., Shackleford D. M., Sriraghavan K., Stingelin L., Tang Y., Urwyler H., Wang X., White K. L., Wittlin S., Zhou L., Vennerstrom J. L. Synthetic ozonide drug candidate OZ439 offers new hope for a single-dose cure of uncomplicated malaria. *Proc. Natl. Acad. Sci. U. S. A.*, **2011**, *108*, 4400-4405.

192. O'Neill P. M., Barton V. E., Ward S. A., Chadwick J. 4-Aminoquinolines: chloroquine, amodiaquine and next-generation analogues. In: Treatment and prevention of malaria, Springer, **2012**, pp 19-44.

193. Sonawane D.P., Persico M., Corbett Y., Chianese G., Di Dato A., Fattorusso C., Taglialatela-Scafati O., Taramelli D., Trombini C., Dhavale D.D., Quintavalla A., Lombardo M. New antimalarial 3-methoxy-1,2-dioxanes: optimisation of cellular pharmacokinetics and pharmacodynamics properties by incorporation of amino and N-heterocyclic moieties at C4. *RSC Advances*, **2015**, *5*, 72995-73010.

194. Fattorusso C., Persico M., Rondinelli F., Orteca N., Di Dato A. Computer-Aided Drug Discovery from Marine Compounds: Identification of the Three-Dimensional Structural Features Responsible for Antimalarial Activity. *Prog. Mol. Subcell. Biol.*, **2017**, *55*, 105-158.

195. Kerns E. H., Di L. Drug-Like properties: concepts, structure design and methods: from ADME to toxicity optimization. *Academic Press*, Amsterdam, **2008**, pp 1-526.

196. Han C., Davis C. B., Wang B. Evaluation of drug candidates for preclinical development: pharmacokinetics, metabolism, pharmaceutics, and toxicology. Wiley, New Jersey, **2010**, pp 1-289.

197. Egan T. J., Combrinck J. M., Egan J., Hearne G. R., Marques H. M., Ntenteni S., Sewell B. T., Smith P. J., Taylor D., Van Schalkwyk D. A., Walden J. C. Fate of haem iron in the malaria parasite Plasmodium falciparum. *Biochem. J.*, **2000**, *365*, 343-347.

198. Dodd E. L., Bohle D. S. Orienting the heterocyclic periphery: a structural model for chloroquine's antimalarial activity. *Chem. Commun.*, **2014**, *50*, 13765-13768.

199. Chishiro T., Shimazaki Y., Tani F., Tachi Y., Naruta Y., Karasawa S. Hayami S., Maeda Y. Isolation and crystal structure of a peroxo-bridged heme-copper complex. Angew *Chem. Int. Ed.,* **2003**, *42*, 2788-2791.

200. McCann P. P., Bacchi C. J., Hanson W. L., Cain G. D., Nathan H. C., Hutner S. H., Sjoerdsma A. Effect on parasitic protozoa of a-difluoromethylornithine an inhibitor of ornithine carboxylase. Adv. Polyamine Res., **1981**, *3*, 97-110.

201. Whaun J. M., Brown N. D. Ornithine decarboxylase inhibition and the malaria-infected red cell: a model for polyamine metabolism and growth. *J Pharmacol. Exp Ther.*, **1985**, *233*, 507-511.

202. Ha H. C., Sirisoma N. S., Kuppusamy P., Zweier J. L., Woster P. M., Casero R. A. The natural polyamine spermine functions directly as a free radical scavenger. Proc. *Natl. Acad. Sci. U. S. A.*, **1998**, *95*, 11140-11145.

203. Maeda Y., Ingold K. U. Knetic applications of electron paramagnetic resonance spectroscopy. The search for a dialkyl-aminyl rearrangement. Ring opening of N-cyclobutyl-N-n-propyl-aminyl. *J. Am. Chem. Soc.,* **1980**, *102*, 328-331

204. Griller D., Howard J. A., Marriott P. R., Scaiano J. C. Absolute rate constants for the reactions of tert-butoxyl, tert-butylperoxyl, and benzophenone triplet with amines: the importance of a stereoelectronic effect. *J. Am. Chem. Soc.*, **1981**, *103*, 619-623.

205. Tanko J. M., Friedline R., Suleman N. K., Castagnoli N. Tert-Butoxyl as a model for radicals in biological systems: caveat emptor. *J. Am. Chem. Soc.*, **2001**, *123*, 5808-5809.

206. Pischel U., Nau W. M. Switch-over in photochemical reaction mechanism from hydrogen abstraction to exciplex-induced quenching: interaction of triplet-excited versus singlet-excited acetone versus cumyloxyl radicals with amines. *J. Am. Chem. Soc.,* **2001**, *123*, 9727-9737.

207. Das K. C., Misra H. P. Hydroxyl radical scavenging and singlet oxygen quenching properties of polyamines. *Mol. Cell Biochem.*, **2004**, 262, 127-133.

208. Reguera R. M., Tekwani B. L., Balana-Fouce R. Polyamine transport in parasites: a potential target for new antiparasitic drug development. *Comp. Biochem. Physiol. Part. C. Toxicol. Pharmacol.*, **2005**, *140*, 151-164

209. Niemand J., Burger P., Verlinden B. K., Reader J., Joubert A. M., Kaiser A., Louw A. I., Kirk K., Phanstiel O., Birkholtz L. M. Anthracene-polyamine conjugates inhibit in vitro proliferation of intraerythrocytic Plasmodium falciparum parasites. *Antimicrob Agents Chemother.*, **2013**,*57*, 2874-2877.

210. Grellepois F., Grellier P., Bonnet-Delpon D., Begue J. P. Design, synthesis and antimalarial activity of trifluoromethylartemisinin-mefloquine dual molecules. *ChemBioChem.*, **2005**, *6*, 648-652.

211. Walsh J. J., Coughlan D., Heneghan N., Gaynor C., Bell A. A novel artemisinin-quinine hybrid with potent antimalarial activity. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 3599-3602.

212. Capela R., Cabal G. G., Rosenthal P. J., Gut J., Mota M. M., Moreira R., Lopes F., Prudencio M. Design
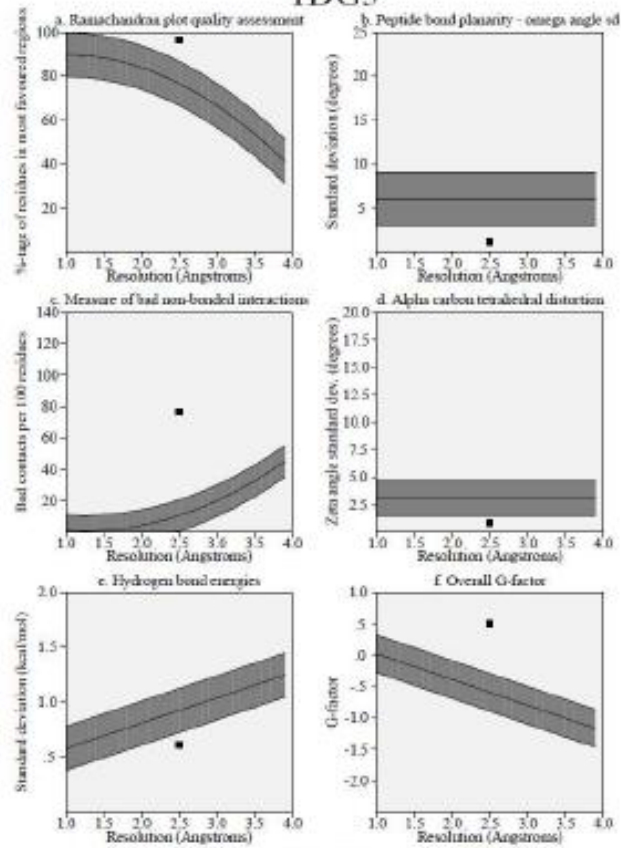
and evaluation of primaquine-artemisinin hybrids as a multistage antimalarial strategy. *Antimicrob. Agents Chemother.,* **2011**, *55*, 4698-4706.

213. Muregi F. W., Ishih A. Next-generation antimalarial drugs: hybrid molecules as a new strategy in drug design. *Drug. Dev. Res.*, **2010**, *71*, 20-32.

214. Flannery E. L., Chatterjee A. K., Winzeler E A. Antimalarial drug discovery-approaches and progress towards new medicines. *Nat. Rev. Microbiol.,* **2013**, *11*, 849-862.

215. Loup C., Lelievre J., Benoit-Vical F., Meunier B. Trioxaquines and heme-artemisinin adducts inhibit the in vitro formation of hemozoin better than chloroquine. *Antimicrob. Agents Chemother.,* **2007**, *51*, 3768-3770.

216. Cosledan F., Fraisse L., Pellet A., Guillou F., Mordmuller B., Kremsner P. G., Moreno A., Mazier D., Maffrand J. P., Meunier B. Selection of a trioxaquine as an antimalarial drug candidate. *Proc. Natl. Acad. Sci. U. S. A.*, **2008**, *105*, 17579-17584.

217. Araujo N. C., Barton V., Jones M., Stocks P. A., Ward S. A., Davies J., Bray P. G., Shone A. E., Cristiano M. L., O'Neill P. M. Semi-synthetic and synthetic 1,2,4-trioxaquines and 1,2,4-trioxolaquines: synthesis, preliminary SAR and comparison with acridine endoperoxide conjugates. *Bioorg. Med. Chem. Lett.,* **2009**, *19*, 2038-2043.

218. Kirkpatrick S., Gelatt C. D., Vecchi M. P. Optimization by simulated annealing. *Science*, **1983**, *220*, 671-680.

219. Fletcher, R. Unconstrained Optimization. In Practical Methods of Optimization; John Wiley & Sons: New York **1980**; Vol. 1.

220. ACD/Percepta, version 14.0.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2013, http://www.acdlabs.com.

221. Dewar M. J. S., Thiel W. Ground states of molecules. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.*, **1977**, *99*, 4899-4907.

222. Maple J. R., Hwang M. J., Stockfisch T. P., Dinur U, Waldman M., Ewig C. S., Hagler A. T. Derivation of Class-II Force Fields. 1. Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules.*J. Comput. Chem.*, **1994**, *15*, 162-182.

223. Holtje H. D., Fattorusso C. Construction of a model of the Candida albicans lanosterol 14-alpha-demethylase active site using the homology modelling technique. *Pharm. Acta Helv.,* **1998**, *72*, 271-277.

Appendix I: Ramachandran plots and Main Chain parameters of Human GBP1 X-ray structure (PDB ID: 1DG3).

PROCHECK

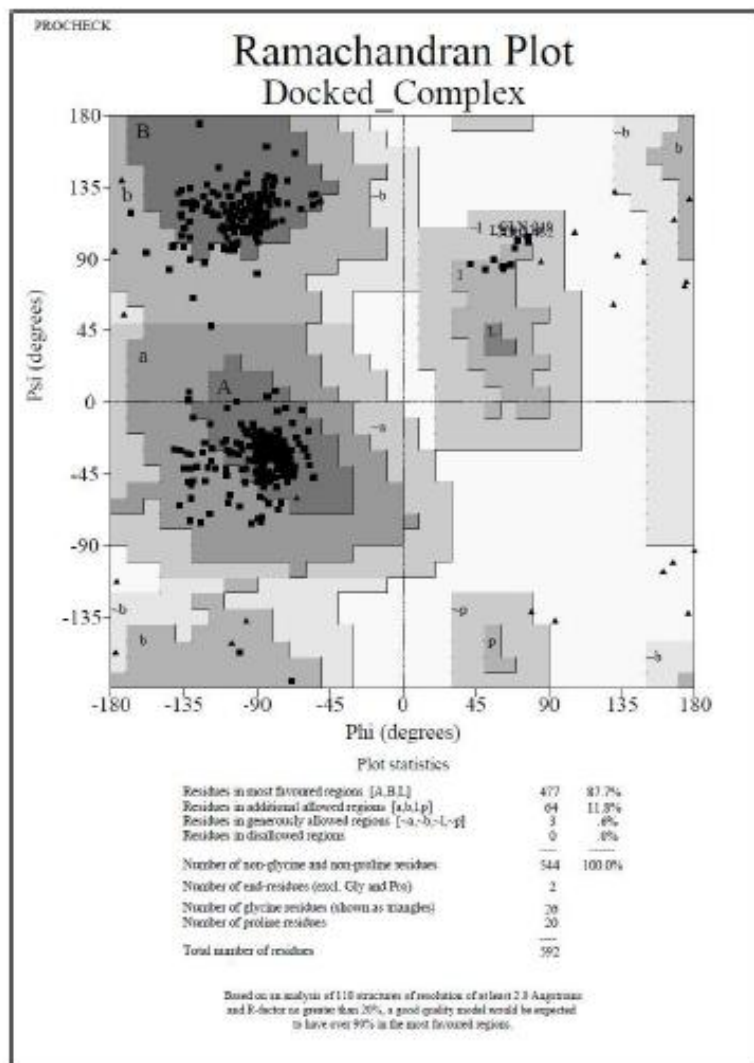# Main-chain parameters
1DG3

a. Ramachandran plot quality assessment

b. Peptide bond planarity - omega angle sd

c. Measure of bad non-bonded interactions

d. Alpha carbon tetrahedral distortion

e. Hydrogen bond energies

f. Overall G-factor

Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	488	96.3	76.6	10.0	2.0	BETTER
b. Omega angle st dev	534	1.2	6.8	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	415	76.5	10.3	10.0	6.6	WORSE
d. Zeta angle st dev	517	.8	3.3	1.6	-1.4	BETTER
e. H-bond energy st dev	402	.6	.9	.2	-1.6	BETTER
f. Overall G-factor	540	.5	-.6	.3	3.6	BETTER

Appendix II: Ramachandran plots and Main Chain parameters of Structure of NSC756093/GBP1 docked complex.

PROCHECK

Main-chain parameters
Docked_Complex
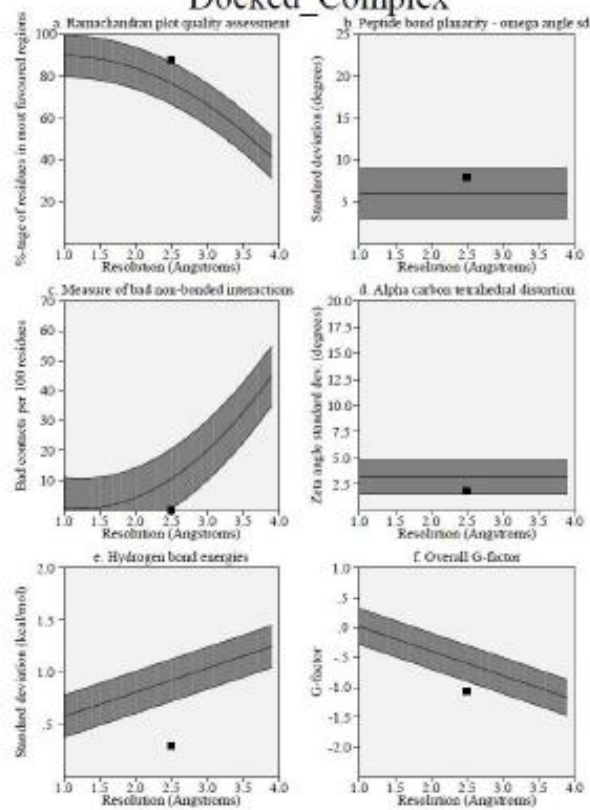
Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	544	87.7	76.8	10.0	1.1	BETTER
b. Omega angle st dev	590	7.9	6.0	3.0	.6	Inside
c. Bad contacts / 100 residues	0	.0	10.5	10.0	-1.1	BETTER
d. Zeta angle st dev	566	1.8	3.1	1.6	-.8	Inside
e. H-bond energy st dev	311	.1	.0	.2	-.1	BETTER
f. Overall G-factor	592	-1.1	-.6	.3	-1.6	WORSE