UNIVERSITAT JAUME·I

Ph.D. Thesis

Inhibition studies on the *human* 20S proteasome: molecular insights from a computational approach

Author

Natalia Serrano Aparicio

Supervisors

Vicent Moliner Ibáñez

Katarzyna Patrycja Świderek

Castelló de la Plana (Spain), January 2022

Doctoral Program in Theoretical Chemistry and Computational Modelling

Universitat Jaume I Doctoral School

Inhibition studies on the *human* 20S proteasome: molecular insights from a computational approach

Report submitted by Natalia Serrano Aparicio in order to be eligible for a doctoral degree awarded by the Universitat Jaume I

Natalia Serrano Aparicio	Prof. Vicent Moliner Ibáñez	Dr. Katarzyna Patrycja Świderek
Firmado por SERRANO APARICIO NATALIA - 73087468T el día 14/01/2022 con un certificado emitido por AC FNMT Usuarios	VICENTE MOLINER IBAÑEZ - NIF:25388050Y Firmado digitalmente por VICENTE MOLINER IBAÑEZ - NIF:25388050Y Fecha: 2022.01.14 15:07:41 +01'00'	KATARZYNA PATRYCJA| SWIDEREK Digitally signed by KATARZYNA PATRYCJA| SWIDEREK Date: 2022.01.14 15:00:12 +01'00'
(Ph.D. candidate)	(Supervisor)	(Supervisor)

Castelló de la Plana (Spain), January 2022

Funding:

- Spanish Ministerio de Economía y Competitividad for a doctoral FPI grant, BES-2016- 078029.
- Spanish Ministerio de Economía y Competitividad and FEDER funds through the project CTQ2015-66223-C2.
- Universitat Jaume I through the project UJI·B2017- 31 and UJI-A2019-04.
- Spanish Ministerio de Ciencia, Innovación y Universidades through the grant PGC2018-094852-B-C21 and PID2019- 107098RJ-I00.
- Generalitat Valenciana through the grant AICO/2019/195 and SEJI/2020/007.
- The computer resources at Pirineus and the technical support provided by the Consorci de Serveis Universitaris de Catalunya through the project RES-QSB-2020-3-0012.

List of publications

(1) Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, 164, 399–407. DOI. 10.1016/j.ejmech.2018.12.062. Impact Factor: 5.537

(2) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, 11, 3575–3589. DOI: 10.1021/acscatal.0c05313. Impact Factor: 13.084 (2020)

(3) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. On the Origin of the Different Reversible Characters of Salinosporamide A and Homosalinosporamide A in the Covalent Inhibition of the Human 20S Proteasome. *ACS Catal.* **2021**, 11806–11819. DOI: 10.1021/acscatal.1c02614. Impact Factor: 13.084 (2020)

(4) Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog. In preparation for submission.

This thesis has been accepted by the co-authors of the publications listed above that have waved the right to present them as a part of another PhD thesis

A mi madre, mi padre y mi hermano
Gracias por todo vuestro apoyo

Abstract

The *human* 20S proteasome activity and malfunction have been related to numerous diseases justifying that this biomolecular system has been validated as a protein target for inhibition in the treatment of cancer. However, and despite the good results in clinical treatment observed for already approved the first and second generation of 20S proteasome inhibitors, the detailed molecular mechanism of inhibition is incomplete. Therefore, computational studies can be of great value for clarifying the mode of action of such inhibitors, leading to a better understanding of the system and the improvement of the current molecules towards new compounds in the drug design process. The present doctoral thesis is devoted to the understanding of the mode of action of the main types of covalent inhibitors currently under investigation for the 20S proteasome. For this purpose, molecular dynamics (MD) simulations with hybrid QM/MM potentials have been applied to the study of two main classes of inhibitors within the β5 active site of the *human* 20S proteasome. The first type of pharmacophore investigated was the α,β-epoxyketones class, where dihydroeponemicin was selected as the compound under study. The results obtained showed the differences between different possible products and demonstrated that the inhibition mechanism works in a three steps process, where the initial activation step goes through the opening of the inhibitor epoxide ring by the attack of the N^{Thr1} atom, rendering a seven-membered ring product. The second major class was γ-lactam-β-lactone inhibitors, where salinosporamide A and different analogs of this class have been studied. First, the molecular mechanism of action of salinosporamide A was clarified, showing the crucial role of the Lys33-Aps17 dyad to activate the $O^{\gamma Thr1}$ nucleophile and initiating the reactivity. Then the origin of the reversibility of the analog homo-salinosporamide A was explained through the investigation of the inhibition mechanism and the possibility of ester linkage hydrolysis, showing that the nature of the reversibility has its origin in the regress of the reaction process. Finally, the S1 pocket phenyl analog, which shows worse activity in *in vitro* studies than salinosporamide A was compared by computing the binding energies of both compounds obtained by applying different computational techniques, and the reactivity was characterized following the same mechanism. Therefore, a set of computational tools have been applied to better understand the mode of action of these inhibitors, ranging from classical MD simulations and

molecular docking to high quality hybrid QM/MM simulations. This methodology allowed finally to characterize the free energy landscape for the inhibition mechanism of these compounds and to provide the structures necessary to analyze and understand deeply the inhibition process in the β5 active site of the 20S proteasome, providing valuable knowledge to optimize the studied compounds into more efficient inhibitors.

Acknowledgements

I would like to thank my supervisors Prof. Vicent Moliner and Dr. Katarzyna Świderek for all their help and advice during the course of my PhD degree. In addition, I am very grateful for giving me the opportunity of working in the Pistol Ribozyme project, and I would like to extend my gratitude to Prof. Joan Bertran, working with him was an amazing experience, and Prof. Iñaki Tuñón, for his valuable comments.

Additionally, I would like to express my gratitude to Dr. Alessio Lodola, who was an amazing supervisor during my research stay in Parma. To Dr. Laura Scalvini and Gian Marco Elisi for all their help and advice in the lab, and the rest of the group of Prof. Marco Mor, that gave me a warm welcome as part of the group during this period.

I would like to thank too to my tutor Dr. Sergio Martí for all his help and being always available for questions and fixing anything that would be not working at that moment. And to Dr. Raquel Castillo for all her help with the complex bureaucracy system that runs UJI.

During the course of this thesis there has been many people present. First, I would like to thank to Daria, Miquel and Eduardo, for being always there during this journey, it is amazing having you as friends. I am confident that we are going to keep making the most amazing memories together.

I would like to thank too to all the people that has been in Biocomp Group during this time, and I have not mentioned yet; Kemel, Isabel, Maite, Javi, Marco, Michał, Sergio B., Ezequiel, Adrián, Santiago. And to all the members of Prof. Juan Andrés, who through collaboration has hosted wonderful co-workers from Brasil sharing our lab, Thiago, Marisa, Renan, Amanda, Rafa, Camila, Ivo, Manoel, Leonelio. Thank you for all the good moments we had during this time.

Then, I would like to mention all the friends that I already knew before starting this journey and are always present despite the distance. The very good friends I met in Zaragoza, Alberto (what would be of me without those Skype meetings), Julen, Raquel, Goicoa. Specially I would like to thank to Diego, who has been there during most of this journey and despite that we have taken separate paths I will always be grateful. And I cannot

forget my childhood friends in Alcañiz; Mari, Nadia, Patri, Luz, Xio, Elisa, Gloria, Eva, thank you for all your wisdom.

And finally, I would like to thank to my family. Without them I would have never arrived here. Thank you to my mom, Fina, and my dad, José Ramón, who are always ready to listen and understand me, and to my brother, Jesús, who is the best brother I could ever wish.

Contents

Chapter 1. Introduction

The chemistry in living organisms is highly complex and works generally in mild conditions. Enzymes are the main players that rule the complex machinery that is the cellular metabolism, a coordinated task where multiple reactions catalyzed by enzymes work concertedly or stepwise to produce or degrade cellular macromolecules.[1] Enzymes are mainly proteins, although there are some RNA-based biomolecules called ribozymes that show the ability to catalyze specific biochemical reactions. Certain reactions in an aqueous solution can take from hours to years if they even occur, whereas in the presence of enzymes the same process accelerates to the fs/ms scale. The Transition State Theory (TST)[2–4] explains that rates in chemical reactions can be related to the equilibrium between an initial, reactants, and a final state, a transition state (TS) that is the highest potential energy point on the path. The reaction rate is related to the product of TS concentration formed from the initial state and the turnover frequency towards the final state. The sequence of each enzyme is prearranged to adopt the precise structure for binding the natural substrate and stabilizing the chemical reaction TS, reducing thus the energetic cost of the process.[5] Therefore the enzyme acts as a facilitator of the reaction progression by reducing the barrier of the process.[6] This has applications in many areas where enzymes can act as biocatalysts in detergents, leather, or food processing and in medicine. Enzymatic catalysis, therefore, is a key area for both academia and industry. Nowadays, the most important drug targets in medicinal chemistry are enzymes and receptors.

Before enzymes were known, a "life-force" was observed propelling the chemical reactions in living organisms. The first enzymes to be discovered were hydrolases, and it was thought that enzymes could only catalyze this type of reaction.[5] Afterwards, different roles that can be played by enzymes were discovered, until the current classification by the Enzyme Commission (EC) was done, that groups enzymes in seven main functional groups: oxidoreductases (EC.1.x.x.x), transferases (EC.2.x.x.x), hydrolases (EC.3.x.x.x), lyases (EC.4.x.x.x), isomerases (EC.5.x.x.x), ligases (EC.6.x.x.x), and translocases (EC.7.x.x.x).

The chemical reactions within cells are usually linked in complex cascades. Some molecules are commonly in steady-state concentrations and different enzymes and regulators are synthesized and degraded as needed. In this process, first, the substrate enters the binding pocket of the protein, and

then it is ready for the catalytic reaction to occur. There are molecules, different than the enzyme and the substrate, that can intervene in this process. In essence, there are regulator molecules that can either enhance or inhibit their target enzyme activity. Because each enzyme undergoes several stages during the catalytic cycle, there are different targets available for inhibition.[7]

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E \bullet S \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E + P \qquad \text{Eq. 1.1}$$

In enzymatic catalysis, two main steps are usually considered as shown in Eq. 1.1, where the k_1 and k_{-1} correspond to the kinetic constants for the binding and unbinding towards the Michaelis complex (E•S), respectively, the equilibrium process where the substrate is recognized entering the binding pocket, known as the binding step. The k_2 and k_{-2} correspond to the formation and regress towards the final product, respectively, the reactivity step that involves chemical transformation passing through the intermediates and products formed along the process. Therefore, any molecule that can interfere in the catalytic cycle avoiding its completion will act as an enzyme inhibitor.

1.1. Protein degradation

Substrate binding is explained by Koshland's theory of induced fit.[8,9] This theory retains part of the previous Key-Lock model idea, where it is proposed that the substrate has the precise shape to fit the enzyme binding pocket. In the new theory, the substrate-binding causes the readaptation of the enzyme geometry. In this way, the catalytic groups get properly aligned during substrate binding. This theory improves the former since it allows to explain why some enzymes can have a wide range of substrates and how allosteric regulation works. In the case of proteases, the binding pockets (S* shown in Figure 1.1) adapt to the specific substrate while the substrate accommodates.

Proteases are the subgroup of hydrolases responsible for catalyzing protein breakdown. The active site of a protease contains a nucleophilic residue or molecule responsible for the reactivity, leading to two main types of

enzymatic catalysis for this type of enzyme. Proteases are classified according to their nucleophiles. Aspartic, glutamic, and metalloproteases use an activated water molecule in the active site as a nucleophile to attack the scissile peptide bond rendering two segments of the protein. Whereas serine, cysteine, and threonine proteases use these respective residues as their active site nucleophiles. In these cases, the mechanism goes through an acyl-enzyme intermediate formed upon attack of the nucleophile to the carbonyl carbon, leading to the N-termini part of the product leaving due to the breaking of the peptide bond. Finally, hydrolysis by a water molecule breaks the acyl-enzyme intermediate, allowing the release of the C-termini part of the product and the regeneration of the enzyme active site.[10]

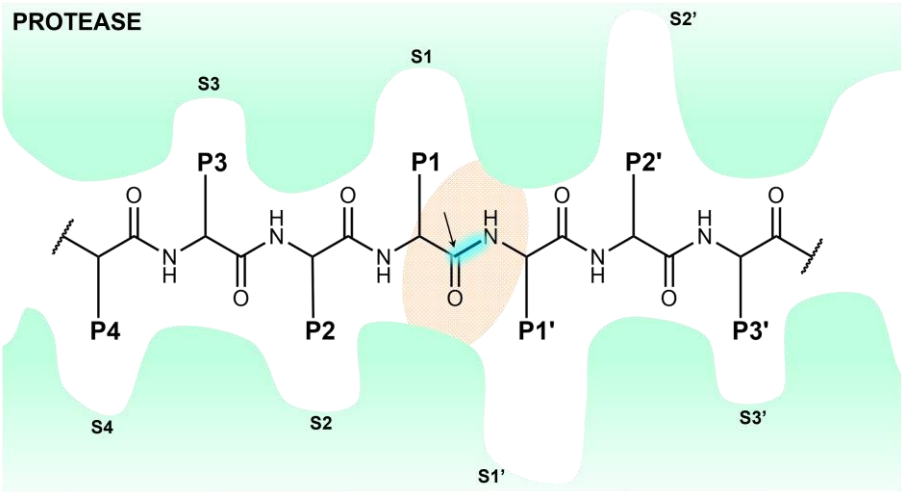


Figure 1.1 General representation of the polypeptide substrate in the proteases binding pocket. The scissile bond (highlighted in blue) is positioned in the active site (orange area) properly oriented towards the catalytic residues. The side chains of the substrate amino acids are interacting within the respective pockets of the enzyme.

Some proteins require post-translational modification for activating their function. This removal of some residues from the C- or N- terminus of a polypeptide chain is done by proteases, and is common for activating enzymes that participate in blood coagulation, digestion, and programmed cell death.[11] Besides, certain polypeptides are the precursor of active peptide hormones, such as EGF (Epidermal Growth Factor) and insulin, and are generated by proteolysis. Additionally, proteases carry the turnover of misfolded or denatured proteins, normal proteins whose concentration

must be decreased, and extracellular proteins are taken up by the cell. With such delicate tasks, maintaining control of protease activity is paramount and there is a broad and diverse range of mechanisms to do it.

Eukaryotic cells contain intracellular proteins with a wide life span range and have several intracellular proteolytic pathways for maintaining cellular homeostasis. One major pathway is for the degradation of extracellular proteins and aged or defective organelles. This is realized in lysosomes, membrane-limited organelles whose acidic interior is filled with hydrolytic enzymes. Then there are a few cytosolic mechanisms, where the ubiquitin-mediated protein degradation by the proteasome is the most important.

Table 1.1. Some examples of proteases implicated in disease progression. Table adapted from ref. 11.

Class	Protease	Disease indication
Metalloproteases	Angiotensin Converting Enzyme (ACE)	Hypertension, myocardial infarction
	Tumor necrosis factor alpha activating enzyme (TACE)	Cancer growth and development
Cysteine proteases	SARS-CoV-2 M^{pro} SARS-CoV-2 PL^{pro}	COVID-19
	Cathepsin B, Cathepsin L	Cancer invasion, growth, and angiogenesis, COPD and emphysema
Serine proteases	Neutrophil elastase	COPD, cystic fibrosis, emphysema
	Plasma Kallikrein	Hereditary angioedema, chronic inflammation, asthma
Aspartic proteases	HIV proteases	HIV infection
Threonine proteases	Proteasome	Cancer growth and progression, chronic kidney diseases, type I diabetes mellitus, sepsis

Proteases are controlled at the transcriptional level expressed only as required,[12] produced as inactive precursors,[13] modified with other post-translational changes such as phosphorylation,[14] regulated by cofactors,[15]

isolated from substrates in vesicles[16] or granules,[17] and by natural protease inhibitors.[18] Despite the existence of these preventing mechanisms, the malfunction of proteolytic pathways can be related to many diseases (see examples in Table 1.1), validating human proteases as therapeutic targets.[19] Likewise, numerous proteases of pathogenic organisms from the five major classes of disease-causing agents, namely bacteria, viruses, fungi, eukaryotes, and prions, are currently targeted for therapeutic intervention and viewed as potential drug targets.[20] Therefore search for the protease inhibitors is an extensive area of study. Many of the designed molecules against the activity of these proteins are already used effectively clinically.[20,21]

1.2. The process of protein inhibition

Enzyme inhibitors can be generally classified considering different aspects of inhibition; i.e. because of their site of action, or their mechanism of action.[22] According to their **site of action**: competitive inhibition is defined in the situations when the inhibitor binds in the same active site as the substrate, therefore inhibition affects substrate binding but not the reactivity. Non-competitive inhibition occurs when regulation happens in allosteric binding sites. When the inhibitor binds in the allosteric pocket, the enzyme structure adapts in such a manner that the reactivity with the substrate is prevented. Ideally, substrate binding is not affected, but active site distortions are almost inevitable and the binding strength of the substrate changes. When both, binding and reactivity are affected it is known as mixed inhibition. Finally, when the inhibitor recognizes and binds the enzyme-substrate complex the process is known as uncompetitive inhibition and affects both catalytic stages.

According to the classification by **mechanism of action,** two main types of inhibitors can be described, non-covalent and covalent. Non-covalent inhibitors undergo only the binding step to the enzyme, the step shown in Eq. 1.2. This type of inhibitor binds into the enzyme binding pocket assuring a stable network of weak interactions and results always in reversible inhibition.[23] In contrast, covalent inhibitors undergo both steps shown in Eq. 1.3, binding and reactivity,[24] forming a covalent adduct that usually gives them greater inhibitory strength. Moreover, depending on the weak interactions that the inhibitors establish with the binding site and the electrophilicity of the inhibitor reactive warhead the reversible and irreversible enzyme inactivation can be defined among the inhibitors of this class.[25]

Non-covalent

$$E + I \underset{k_{of}}{\overset{k_{on}}{\rightleftarrows}} E \bullet I \qquad \text{Eq. 1.2}$$

Covalent

$$E + I \underset{k_{of}}{\overset{k_{on}}{\rightleftarrows}} E \bullet I \xrightarrow{k_{inac}} E - I \qquad \text{Eq. 1.3}$$

The use of protein inhibitors is wide, starting from research studies where inhibitors have allowed understanding protein mode of action and are used to modulate protein activity in experiments, finishing on key medical applications. Enzymes now comprise one-third or more of the discrete drug targets found within large pharmaceutical company portfolios,[7] and roughly half of all marketed drugs are enzyme inhibitors.[26,27] These enzyme inhibitors drugs can have different targets. For instance, they can work against microorganisms, like penicillin and cephalosporin C, where the target is located in a crucial enzyme for the bacteria, thus preventing their growth or killing them, against viruses, like acyclovir (herpes), zidovudine, and saquinavir (human immunodeficiency virus, HIV) preventing their replications or against the human enzymes. However, targeting a single enzyme and managing off-target effects can be very difficult.

1.2.1. Covalent inhibitors

The development of covalent inhibitors has shown a significant increase in the last 50 years, as can be seen in Figure 1.2. Currently, at least 50 covalent inhibitors have been approved by the U.S. Food and Drug Administration (FDA) and numerous drug candidates are progressing through clinical trials.[28] But initially, the development of covalent inhibitors was considered with skepticism among the pharmaceutical industry.[29,30] The main issue for the use of covalent inhibitors as potential drugs was that the covalent modification of the protein may cause immunotoxicity problems that produce a hapten recognized as foreign by the immune system.[11] Furthermore, they demonstrated other disadvantages like for instance the potential off-target effects through a nonselective reaction that can cause adverse drug responses; the possibility of acquiring resistance by receptor mutation; difficulty for inhibition level modulation when short-duration inhibition is desired, or the lack of inhibition duration if the enzyme's half-life in a cell is short; and the difficulty on balancing reactivity and selectivity that makes the design of these inhibitors even more complex.[22]

Despite these limitations, covalent inhibitors have some key advantages as the possibility of targeting protein-protein interactions, the ATP binding site, or shallow binding pockets.[23,30–34] Targeting a unique nucleophilic residue allows to achieve selectivity for highly homologous proteins with increased ligand efficiency;[30] high potency with low dose caused by the

complete blockade of the target;[35] sustained duration of action due to the nonequilibrium condition leading to less frequent dosing, since the pharmacodynamic half-life of the drug is much longer than its pharmacokinetic half-life;[35] not all types of covalent modifications of proteins result in toxicity and otherwise undesirable pharmacokinetic properties can often be tolerated as pharmacodynamics properties of these inhibitors outlast measurable inhibitor concentration in the plasma;[36] covalent modification allows to help identify the proteins targeted, and many natural products covalently bind their natural target.

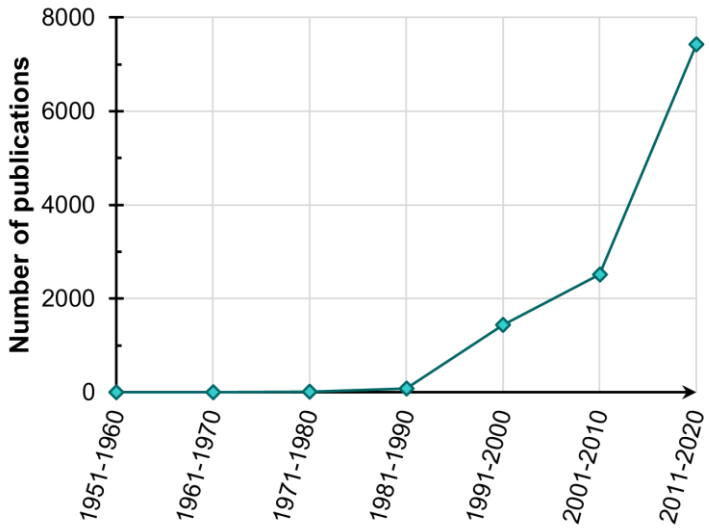


Figure 1.2. Number of publications per decade on covalent drugs. Results of the search (October 2021) on the ISI Web of Science™ using as keywords "covalent drug" in the Topic criteria.

Moreover, to ensure safety a list of requirements has emerged for designing the ideal covalent inhibitor, as shown by already successful covalent drugs like aspirin, penicillin, omeprazole, etc. that are on the market.[36] The two main requirements originate in the mechanism of action. Initially, optimal noncovalent binding interactions are necessary to improve the selectivity for the target binding pocket. In addition, mild electrophilic functionality, only reactive within the target binding site, allows the minimal formation of reactive metabolites.[23,25] These two requirements allow high selectivity with minimum doses when the targeted protein is turned over slowly, which should make safe covalent inhibitors adding some therapeutic advantages

compared to noncovalent drugs.[23,25] As shown in Figure 1.3, there is a wide range of compounds that covalently modify their targets. Approximately one-third of all approved drugs against enzyme activity are covalent inhibitors.[30,37]

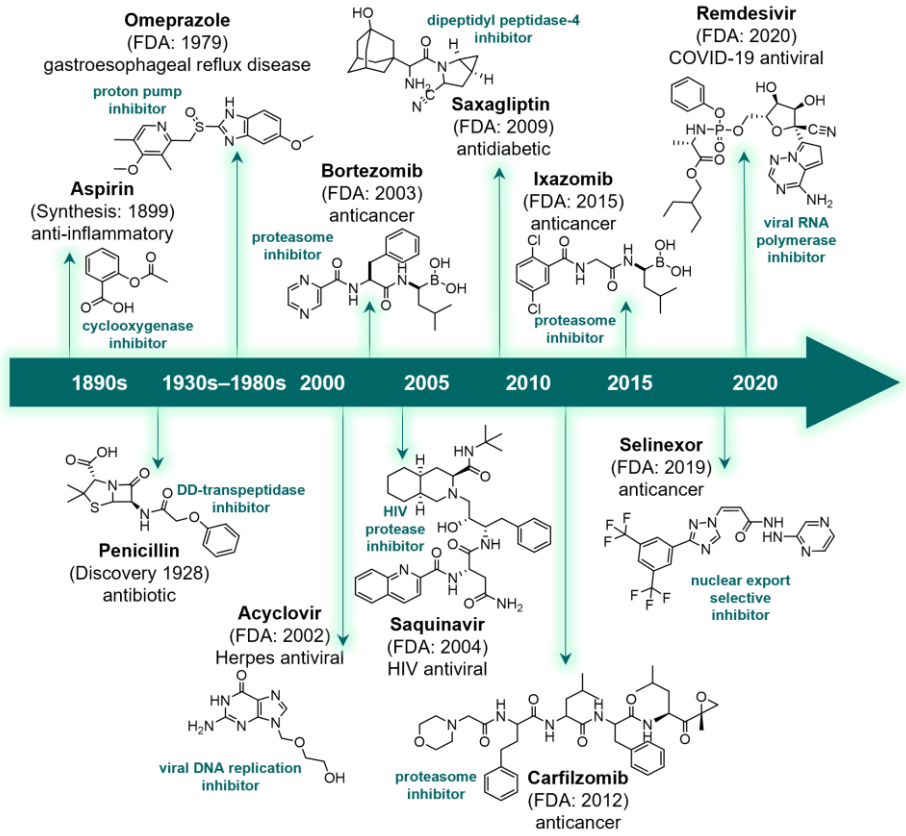
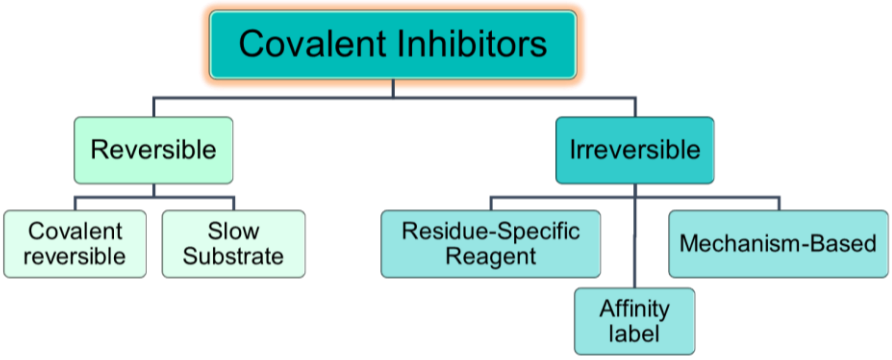


Figure 1.3. Some examples of marketed covalent inhibitors and their protein target. Detailed list can be found in ref. 38. Figure adapted from ref. 36.

Covalent inhibitors may be classified into two broad categories according to whether their adducts with the protein targets are functionally reversible on the biological timescale, upon dialysis, competition with the excess substrate, or extended incubation times.[22,25] **Reversible covalent inhibitors** allow the recovery of the protein faster than the turnover rate of the protein.[23] On the other hand, **irreversible covalent inhibitors** form adducts with a kinetic half-life longer than the protein biosynthesis or even do not dissociate from the protein during its lifetime.[23] Within these two groups, further classification can be considered, as shown in Figure 1.4.

Among reversible covalent inhibitors, two classes can be considered: **covalent reversible inhibitors** and **slow substrates**. The first class is the most common, where the reversibility comes from the regress of the inhibition mechanism to inhibitor unbinding. The second class does not use a reversible reaction mechanism, but the inhibitor acts as a substrate with a long-lived covalent intermediate which eventually is further catalyzed by the enzyme, rendering a product or inactivated inhibitor and the recovery of the enzyme.[22]



Class of inhibitor		Scheme
Reversible	Covalent Reversible	$E + I \underset{}{\overset{K_I}{\rightleftharpoons}} E \bullet I \overset{k_{inact}}{\rightleftharpoons} E - I$
	Slow Substrate	$E + I \overset{K_I}{\rightleftharpoons} E \bullet I \overset{k_{inact}}{\longrightarrow} E - I \overset{slow}{\longrightarrow} E + P$
Irreversible	Residue-Specific Reagent	$E + I \overset{k_{inact}}{\longrightarrow} E - I$
	Affinity label	$E + I \overset{K_I}{\rightleftharpoons} E \bullet I \overset{k_{inact}}{\longrightarrow} E - I$
	Mechanism-Based	$E + I \overset{K_I}{\rightleftharpoons} E \bullet I \overset{k_{cat*}}{\longrightarrow} E \bullet I^* \overset{k_{inact}}{\longrightarrow} E - I$

Figure 1.4. Classification of the most common mechanism for covalent inhibition. Scheme adapted from ref. 22.

Irreversible inhibitors can be classified into three major categories: **residue-specific reagents, affinity labels,** and **mechanism-based enzyme**

inactivators. Residue-specific reagents are reactive compounds that rely on chemoselectivity for the enzyme nucleophile instead of noncovalent affinity to the binding site and are used mostly *in vitro* as biochemical tools. Affinity labels increase their site selectivity by coupling a reactive group that provides noncovalent binding affinity,[39] and can be classical, quiescent, or photoaffinity labels depending on if they are activated by a weak electrophile, favoring "off-pathway" mechanism or light, respectively. Finally, covalent **mechanism-based enzyme inactivators** bind to the active site of enzymes and are processed by the normal catalytic mechanism to produce a reactive species that results in covalent bond formation.[22]

Furthermore, since covalent inhibitors contain different types of electrophilic groups designed to react with the protein nucleophile, the "**warhead**" classification is widely used as well. The warhead, or pharmacophore, and its electrophilicity are key for the reactivity step and will determine the mechanism of the reaction. There are molecules with a variety of warheads including epoxides, esters, ketones, aziridines, α, β-unsaturated carbonyls, nitriles, carbonitriles, aldehydes, α-ketoamides, boronic acids, etc.[25,29,40]

1.2.2. Design of covalent inhibitors

Frequently, the strategy for developing a protease inhibitor involves the use of competitive inhibitors that mimic the natural substrate and can block or limit normal substrate turnover.[19] As a matter of fact, most protease inhibitors clinically used have been developed employing this strategy.[11] Another option for proteases is to use a natural inhibitor since there are small molecules in nature that target certain proteases and can be optimized directly as drug leads.[7,18,41–46] Therefore, to design a covalent inhibitor there are usually three steps.[36] First, the target structure is analyzed, giving crucial information for optimizing the selectivity according to the binding pocket and the present nucleophile. Catalytic serine, threonine, and cysteine residues have all been targeted by covalent inhibitors.[47,48] Second, established inhibitors with high potency are used as leads for initial tests. Finally, the selected lead compound is optimized by adding a more specific "warhead" for the present nucleophile in the enzyme or improving the sidechains interacting within the binding site.

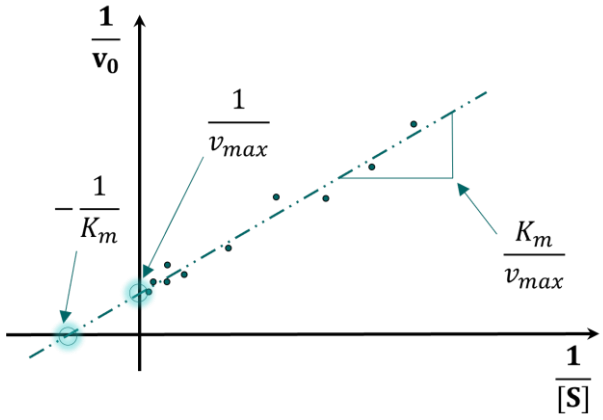


Figure 1.5. An example of a Lineweaver-Burk plot. [S]: substrate concentration. v_0: initial velocity of an enzyme reaction. v_{max}: maximum velocity of the reaction. K_M: Michaelis-Menten constant or enzyme affinity.

During the design, there are key molecular descriptors essential for understanding the type of inhibition and evaluating the quality of the designed drug. During *in vitro* assays, enzyme kinetics are extremely useful in determining the properties of the inhibitor. The Lineweaver-Burk diagram[49] (or double-reciprocal plot) is a commonly used graphical tool to calculate the kinetic parameters (see Figure 1.5). It represents the inverse of the reaction initial velocity ($1/v_0$) as a function of the inverse of the substrate concentration ($1/[S]$). These plots allow determining the type of inhibition by obtention of the Michaelis-Menten constant, K_M, and the maximum velocity, v_{max}, from the inhibited enzyme result compared to the no inhibition process.[50]

For covalent inhibition, these molecular descriptors obtained from kinetic assays are the equilibrium constant that measures inhibitor binding affinity, K_I; and the inactivation rate constant, k_{inact}. The overall potency of the covalent inhibitor on the enzyme is expressed as the ratio of k_{inact}/K_I.[30,33,51] Furthermore, there is growing interest in binding kinetics with an emphasis on residence time and dissociation rates.[38,52,53] As K_I corresponds to k_{on}/k_{off} ratio when inhibition is reversible, dissociation plays a key role when it is not correlated with K_I. In such cases, the half-live of the inhibitor can give further information. Association rates of drugs with nanomolar K_I values and dissociation half-lives less than 10 min are most likely diffusion-

controlled, while the association rates of drugs with longer half-lives must involve a rate-limiting conformational movement.[54]

Since measuring inhibition kinetics is very demanding, a widely used magnitude to compare the activity of tested compounds is the IC_{50} value, determined by varying inhibitor concentration at a single set preincubation time. The value of IC_{50} corresponds to the required inhibitor concentration that decreases in half (50%) the rate of product formation under the specific assay conditions [55,56] (see Figure 1.6) From a practical viewpoint, ranking compounds within a structural series can be useful, but IC_{50} values can be misleading due to their inherent dependence on experimental conditions and other processes different from just the chemical step.[25,30,56,57] Moreover, this magnitude will not indicate a change in the inhibition mechanism in a series of lead compounds and, since this type of inhibition is time-dependent, the data will not correlate with the pharmacodynamics as gives no information about "residence time" on the target.[52]

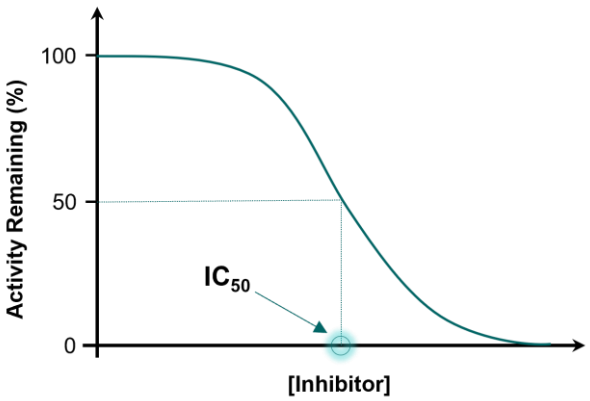


Figure 1.6. Determination of IC_{50}. Percent of enzymatic activity remaining as a function of inhibitor concentration.

Therefore, the k_{inact}/K_I ratio is a more appropriate indicator of covalent inhibitors' potency.[30,58] Furthermore, an additional descriptor can be evaluated to address how effectively the binding of an inhibitor to the target provides the effectively pharmacological response, the biochemical efficiency.[38,59] This value is defined as the ratio of the K_I obtained in a binding assay to the IC_{50} in a physiologically relevant functional assay when the unit indicates an efficient binding couple to the physiological response.[54] Isotopic substitution can also be used to further understand the mechanism

since it allows to obtain differences in the equilibrium and rate constants due to the isotopic substitution of some atoms involved in the process and can give further insight either in binding[60,61] and reactivity processes.[54,62]

In vivo inhibition studies, where the molecule is tested against whole cells or tissues is a natural step forward in testing the most promising candidates. The common descriptor obtained from these studies is the EC_{50}, which represents the concentration of inhibitor required to reduce a certain monitored cellular effect by 50%.[63] This gives a clue of the inhibitor potency in the cellular environment.

When the identification of a new lead compound is necessary, the dominant technique is to use the methods described above for conducting the screening of large libraries of chemicals against a biological target.[64] This can be done *in vitro* (high-throughput screening), or *silico* (virtual screening). The latter consists of the computational screening of large libraries of chemicals for a particular target. With this technique, the goal is to find compounds of interest that rank well in the library to further test them experimentally in a more efficient manner. This type of calculation gives a qualitative result since the computational cost is reduced to the minimum to afford the screening of large libraries. However, to obtain more reliable computational results it is necessary to improve the description of the systems to be computed and therefore increase the cost of the calculations. In recent years computational studies that play a key role in understanding certain aspects of inhibition at a molecular level, including the characterization of the mechanisms of covalent adduct formation have been extensively exploited in drug design projects.[65–67]

1.2.3. Computational methods in drug discovery

Computational modeling is commonly used for performing structure-activity relationship studies, further applied to optimize a lead covalent inhibitor.[36] These techniques are especially useful in covalent inhibitor studies since the characterization of the inhibitor mechanism is key for understanding and optimizing the promising compound.[47,65–67] For the development of covalent-modifier drugs, computational studies usually start with a model preparation assisted by structural data deposited in the Protein Data Bank (https://www.rcsb.org/).[68] These structures provide

initial pivotal information about localization of possible binding pockets, binding pose adapted by ligands used in crystallographic studies, or residues playing the role of nucleophile in the active site. If there is no experimental structural data available, homology models can be constructed to obtain an initial atomistic representation of the protein target.[69]

Usually, the strategy is to gain an understanding of the covalent modifier molecular mechanism while comparing the predicted results with the available experimental data, so later the gathered knowledge is applied to design new inhibitors. However, and despite this design process can be very precise regarding the protein target, it has much more difficulties in the evaluation of the absorption, distribution, metabolism, and elimination properties (ADME) of the drug. *In silico* techniques have been widely used to estimate ADME properties of chemical compounds.[70] But the quality of the experimental data varies heavily, especially in the case of human systems.[71] Therefore, different types of approaches can be applied to deal with the biological complexity of ADME properties modeling. These range from a traditional quantitative structure-activity relationship (QSAR) approaches, using statistical methods or machine learning algorithms,[70] to more computationally intensive calculations based on quantum mechanics atomistic modelling. However, it is not possible to predict the full range of ADME properties with one single technique, and a combination of two or more models based on different principles can give higher confidence in the results.[71]

As commented in the previous section, the covalent inhibition process goes through two main stages, in the first one the inhibitor associates or binds to the target binding site, while in the second stage the covalent modification occurs in one or more reaction steps. The key magnitude associated with the first step is the binding Gibbs free energy (ΔG_{bind}), which can be related to the K_I, through Eq. 1.4, where C° is the standard state concentration.[72]

$$\Delta G_{bind}(T) = -RTlnK_I C° \qquad \text{Eq. 1.4}$$

For the reactivity step, it is crucial to characterize the full inhibition mechanism, for understanding the "warhead" mode of action within the target active site. This is essential for further improvement steps and for identifying the reaction rate-limiting step of the process, that is associated

with the experimental k_{inact}. The TST can be used to relate the activation Gibbs free energy computed for the rate-limiting step ($\Delta G^{\ddagger}$) to the k_{inact} through Eq. 1.5,[73,74] where k_{inact} corresponds to k_{TST} dependent on the temperature (T) and the reaction coordinate (ξ) employed in the simulations.

$$k_{TST}(T,\xi) = \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\ddagger}(T,\xi)}{k_B T}\right) \qquad \text{Eq. 1.5}$$

These two key magnitudes can be obtained using different approaches distinguished by the number of approximations in the model. Consequently, the precision of the obtained result and the computational cost will be different. ΔG_{bind} is usually obtained for large databases of compounds using docking algorithms,[75–77] molecular dynamics simulations, or enhanced sampling or free energy perturbation methods (FEP),[78–81] that can be used at different levels of theory. The value of $\Delta G^{\ddagger}$ can be obtained using quantum chemical cluster approaches that use models[82] of limited size mostly reduced to the binding site area including catalytic residues and the substrate or inhibitor and its very close surrounding. Nevertheless, and despite their accuracy in determining reaction mechanism or studies of stereoselectivity,[82–84] these types of models placed in the vacuum or a continuum solvent model do not provide a rigorous description of the overall protein environment and its influence on the reaction.[85] Alternatively, computationally more expensive quantum mechanics / molecular mechanics (QM/MM) methods, that account for the effects of the full protein, can be used. These more advanced methods have played a significant role in understanding and predicting the mechanisms, kinetics, and thermodynamics of covalent modification because they allow to include in the calculations the effect of the electrostatic environment of the full protein, which has been recently accepted to be the driving force of enzymatic catalysis.[86–90]

Another typical use of computational methods is the prediction of pKa of a target residue or the estimation of membrane permeability of a drug. Lastly, knowing the binding and reactivity free energies allows optimizing drug leads through "warhead" design or side-chain modification, obtaining afterwards an estimate of the inhibitory potency of the new compound.

1.3. The 26S proteasome

The ubiquitin-proteasome system (UPS) is responsible for hydrolyzing more than 90% of all cytosolic proteins, where the 26S proteasome, a large ~2.5 MDa molecular complex, is the central player of this non-lysosomal protein degradation pathway.[91] The 26S proteasome is formed by two subcomplexes, the catalytic core particle (CP) or 20S proteasome, and one or more 19S regulatory particles (RP) that serve as activators. These different parts of the complex can be distinguished using the apparent sedimentation coefficient, which corresponds to 20S, 19S, and 26S respectively for CP, RP, and the complex formed by the binding of RP to CP.[92] The UPS was discovered in 1969 when an ATP-dependent proteolysis process was identified,[93] and later, connected to a stable polypeptide, named ubiquitin for being present in all tissues and eukaryotic organisms.[94,95] Eventually, the intracellular proteolytic system was identified as a big multimolecular protein initially called "cylindrin",[96] finally named as the 20S proteasome[97–99] and confirmed to be the catalytic core of the ATP dependent proteolysis in the UPS.[100] The characterization of the ubiquitin system conjugation and its role in specific proteolysis labeling was awarded by the Nobel Prize in Chemistry in 2004 to Avram Hershko, Aaron Ciechanover, and Irwin Rose.[101]

The UPS degrades proteins in a multistep process, as shown in Figure 1.7 (B), where tagged proteins are linked by covalent bonds to the ubiquitin in presence of ATP.[102,103] This small protein binds to the 19S RP of the 26S proteasome, where the substrate protein is unfolded and inserted into the multicatalytic chamber, the 20S CP, to be rapidly degraded to smaller polypeptides.[104–106]

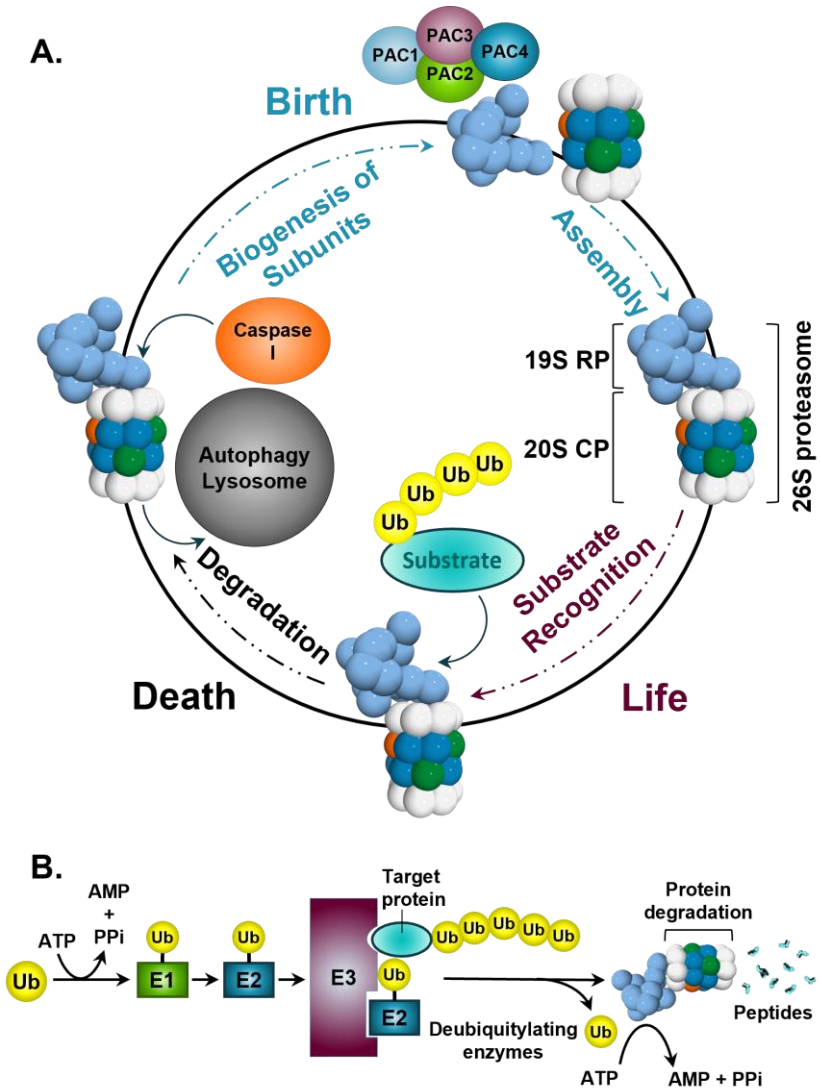


Figure 1.7. **A**. The life cycle of the proteasome, from cellular biogenesis to proteasome function and cell degradation of the macro complex. Figure adapted from ref. 107. **B**. Proteasomal-dependent degradation of cellular proteins within the UPS. Figure adapted from ref. 108. 26S proteasome structure obtained from PDB code: 5M32.[109]

1.3.1. Biological role: relationship between the proteasome and human diseases

The 20S proteasome's primary role is to degrade key regulatory proteins and aberrant proteins.[63] Therefore, it procures the maintenance of cellular homeostasis controlling protein turnover.[106] Since the UPS is the main protein degradation pathway in the cytosol, it has a wide spectrum of protein substrates, and its function intervenes in multiple crucial cellular processes via protein degradation.[110] It takes part in cell cycle progression,[111–113] cellular signalling,[114,115] genome integrity,[116,117] regulation of endocrine pathways,[118,119] apoptosis,[120,121] transcriptional regulation,[122,123] metabolism regulation,[1,124] immune responses,[125,126] and many human diseases pathogenesis.[127,128] Thus, all these functions make the proteasome emerge as a crucial target for therapeutic intervention in many diseases such as neurodegenerative diseases;[129,130] cancer;[131–133] autoimmune diseases; developmental disorders; cystic fibrosis; diabetes; cardiac diseases; atherosclerosis; or aging.[134] Additionally, some infectious diseases as Chaga's disease,[135] Malaria[136–138], and Tuberculosis[139] have been related to proteasomal function. Hence, depending on the relationship between proteasomal activity and the disease it will be required a different type of drug, i.e. a proteasome-activating or -modulating compound or a proteasome inhibiting compound.[140]

Aging is a progressive and irreversible, but not pathological, phenomenon that consists of a spectrum of changes at molecular, cellular, and organism level.[110] Therefore, an age-related decrease in proteasome activity can be considered the natural answer to a decrease in the rate of protein synthesis due to the same phenomenon.[140] A steady decrease in proteostasis can favor the aggregation of damaged and misfolded proteins, formation of inclusion bodies[141], and development of neurodegenerative diseases such as Alzheimer's[142] and Parkinson's disease,[143] amyotrophic lateral scleroiss[144] and Huntington's disease.[145] Therefore, enhancement of proteasome activity, by using proteasome activators or modulator compounds, has many therapeutic potentials but is still a relatively unexplored field.[146] On the other hand, accelerated protein degradation is a key feature of many malignancies, such as chronic kidney diseases, type I diabetes mellitus, sepsis, cancer cachexia, and starvation.[140] Therefore, inhibition of proteasome activity has a complex role in apoptosis, leading to pro- and

anti-apoptotic effects.[147] Proteasomes activate the NFκB pathway inducing anti-apoptotic members of the BCL2 family,[148] and degrade pro-apoptotic proteins,[149,150] and the negative cell cycle regulators.[151] Rapidly growing cells such as cancer cells show an induction of proliferation facilitated by these anti-apoptotic proteasomal effects.[140] Therefore, proteasome inhibitors target upregulated cells and disrupt these mechanisms inducing apoptosis in cancer. As will be explained below, there are currently three approved drugs by the FDA that act as proteasome inhibitors in multiple myeloma cells, and new drugs are in clinical trials for this and other types of cancer cells. Proteasome inhibitors have shown to have promising therapeutic effects in the treatment of cystic fibrosis too, where proteasome inhibition can disrupt the degradation of the misfolded cystic fibrosis transmembrane conductance regulator;[152] or autoimmune diseases, where the involvement of (immuno)proteasome in antigen presentation and the capacity of proteasome inhibitors for suppressing the NFκB pathway leads to the cynical exploration of proteasome inhibitors as treatment for autoimmune diseases such as rheumatoid arthritis, Sjörgren's syndrome and sclerodema.[153]

Finally, some parasite-related infectious diseases rely on rapid protein turnovers, such as the malaria parasite *Plasmodium falciparum* or the protozoan parasite *Trypanosoma cruzi* causing Chaga's disease. These parasites contain a functional eukaryotic proteasome that can be selectively targeted by proteasome inhibitors.[135–138] The bacterial proteasome of the *Mycobacterium tuberculosis* human pathogen can be selectively targeted as well, opening the path for the development of new drugs for tuberculosis.[139]

1.3.2. Structure of the 20S proteasome

The 20S proteasome appears in all three kingdoms of life, and despite a few fundamental differences in structure and function, the basic structure and subunit arrangement are conserved during evolution.[154–156] Except bacteria not including actinomycetes, which harbor complexes of two homohexameric rings with one active subunit each,[157] this protease complex consists of four stacked homoheptameric rings forming a cylinder of 15x11 nm and approximated molecular mass of 720 kDa, obtained from *Thermoplasma acidophilum*[158] and the eukaryote *Sacharomyces cerevisiae*.[159] There are two types of basic subunits, namely known as α and β. The two outer

rings are formed by seven α subunits and the two inner rings by seven β subunits. The tertiary structure is the same for all known α and β subunits, involving two antiparallel five-stranded β sheets flanked by α helices and probably evolved from a common ancestor.[158,160] While archaeal 20S proteasomes have only one type of α and β subunit,[158] the eukaryotic 20S proteasome has seven different types of α (1–7) subunits and seven different types of β (1–7) subunits (see Figure 1.8 A-B).[159] The α subunits have a key role in the complex assembly and form the entry gate to the interior of the proteasome.[161,162] The β subunits bear the catalytic activity, for this reason, they are synthesized as inactive precursor proteins, with N-terminal propeptide which is cleaved in the final biosynthesis step by intramolecular autolysis, as shown in Figure 1.7 (A).[163]

In eukaryotes, there are only three active β subunits that bear different specificity according to their S1 binding site pocket. The β5 active sites are named chymotrypsin-like and have a hydrophobic S1 pocket characterized by the presence of Met45; the β2 active sites are known as trypsin-like because of the Asp55 residue in their S1 specificity pocket and cut after basic residues; finally, the β1 active sites have caspase-like specificity with Arg45 residue in the S1 pocket and cleave preferentially after acidic residues.[164,165] Besides their differences in the specificity binding site pockets, the catalytic residues are highly conserved in certain key aspects: all active sites use an N-terminal threonine as a nucleophile;[106] the catalytic triad Thr1-Lys33-Asp17 has been proposed as key residues for the protease activity;[166] the hydrogen bonding surrounding the catalytic triad, as can be seen in Figure 1.8 (C), helps to maintain the active site structure; and the oxyanion hole residue, Gly47, that interacts with the protein substrate during the mechanism.

The eukaryotic 20S proteasome is present in all types of cells either in the cytosol or the nucleus. As explained above, the 20S proteasome can be named as 20S CP when running within the UPS, which is the primary route for proteasomal degradation. However, proteins can also be targeted by the 20S CP itself, in a ubiquitin-independent manner.[167] During ubiquitin-mediated proteolysis, within the UPS, the 20S CP assembles with the 19S RP forming the 26S proteasome. Moreover, at a higher species level, there is an increase in complexity and specificity in structure and function, shown by the presence of three classes of core particles in vertebrates: the

constitutive proteasome (cCP), the immunoproteasome (iCP), and the thymoproteasome (tCP).[168,169]

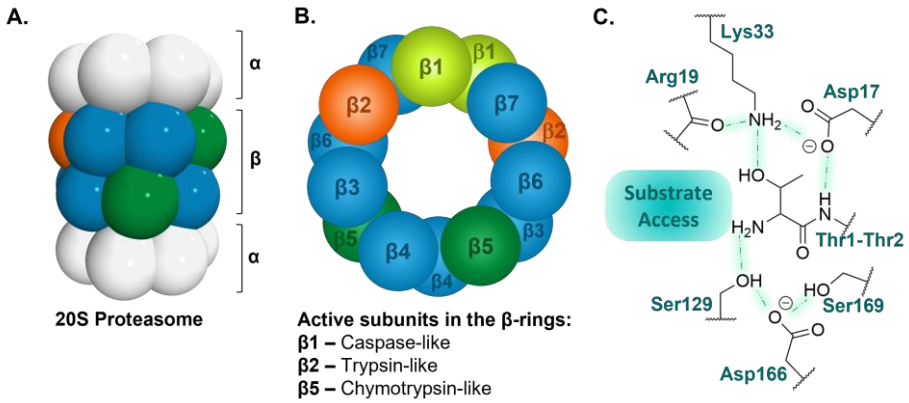


Figure 1.8 **A**. 20S proteasome representation showing α and β rings arrangement. 20S proteasome structure obtained from PDB code: 5M32.[109] **B**. β rings top representation showing the relative positions of the active subunits (β5, β2, β1). **C**. Active site H bonding network around Thr1. Figure adapted from ref. 170.

1.3.3. 20S proteasome inhibition

The fact that the 20S proteasome participates in so many cellular processes encouraged the development of its inhibitors.[171] Numerous inhibitions and site-directed mutagenesis studies have provided much of the current knowledge about the complex catalytic mechanism of this protease that was extensively used in drug design.[172] The first natural molecule with inhibitory effects against the 20S proteasome was lactacystin reported in 1991 and obtained from *Streptomyces* strain.[173] Afterwards, the first synthetic inhibitors for the 20S proteasome were developed,[171] these were designed to mimic the substrates specific for the β5 active site.[174] Additionally, it was found that β5 inhibition produces significant proteolysis decline, therefore studies solely concentrated on β5 inhibition.[175] However, the toxicity and innate or acquired resistance observed in patients of the first-generation proteasome inhibitors,[176] lead to the development of combined therapies to confer a broader inhibition profile by targeting more than one active site.[177] Therefore, nowadays studies take into consideration β1 and β2 active sites too.

Most known proteasome inhibitors are peptide-like molecules bearing an electrophilic "warhead" that reversibly or irreversibly inhibit the catalytic $O^{\gamma Thr1}$ by covalent inhibition. Therefore, proteasome inhibitors are commonly classified by their characteristic "warhead" in seven classes (see Figure 1.9): aldehydes, boronic acids, α,β-epoxyketones, α-ketoaldehydes (glyoxals), vinyl sulfones, vinyl amides (syrbactins), and β-lactones.[178] Aldehydes (MG-132) have great value as research tools, but no medical potential due to their off-target activity towards serine and cysteine proteases.[158,174,179,180] Boronic acids are among the most potent inhibitors despite their high reactivity and associated side effects, bortezomib (PS-341) was synthesized with MG-132 as a lead compound and was the first approved proteasome inhibitor by the FDA.[181–184] α,β-epoxyketones (carfilzomib)[185,186] and α-ketoaldehydes[187] inhibit, irreversibly and reversibly respectively, the 20S proteasome by forming two covalent bonds with both $O^{\gamma Thr1}$ and N^{Thr1}. Vinyl sulfones[188] and vinyl amides[189] are Michael acceptor inhibitors forming an ether bond with $O^{\gamma Thr1}$ and have off-target effects with cysteine proteases. Finally, β-lactones are small molecules that form an ester bond with $O^{\gamma Thr1}$ and cleave their β-lactone ring, they can have a reversible or irreversible mode of action depending on the side chains substitution,[190] which makes them of great interest for tunning proteasome inhibition. At this moment salinosporamide A is the most potent candidate within this class of inhibitors.

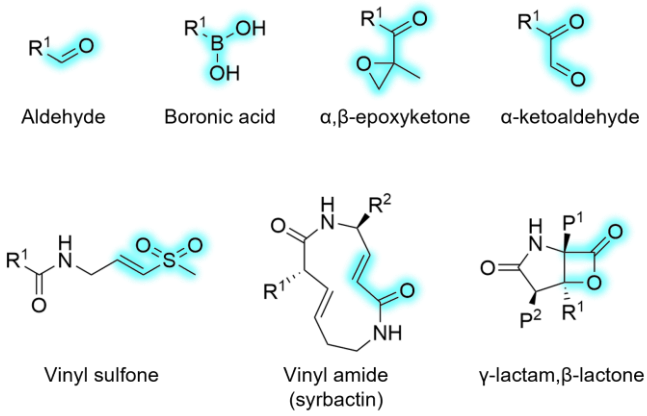


Figure 1.9 Schematic representation of 20S proteasome inhibitor warheads. The functional group is highlighted in blue. R^1 and R^2 correspond to the variable part of the compounds. P^1 and P^2 correspond to the side chains in γ-lactam-β-lactones that target the corresponding S pockets.

Up to now, there are three proteasome inhibitors approved by the FDA (see Figure 1.10). Bortezomib (Velcade; Takeda Pharmaceuticals) is a boronic acid that was the first approved drug in 2003 for the treatment of relapsed and refractory multiple myeloma (RRMM),[191,192] because of its higher toxicity for cancerous cells than in healthy cells. The success of this first approved inhibitor came with natural and acquired resistance and dose-limiting toxicity. This led to the development of the second generation of proteasome inhibitors with improved pharmacokinetics.[193] Carfilzomib (Kyprolis®; Onyx Pharmaceuticals / Takeda) is an α,β-epoxyketone that was approved in 2012 for treatment of RRMM. Both, bortezomib and carfilzomib, are drugs of intravenous administration, which requires administration in a clinic. Ixazomib (Ninlaro®; Takeda Pharmaceuticals) is a boronic acid that was the next to be approved in 2015 for RRMM treatment, currently the only orally bioavailable approved proteasome inhibitor.[194]

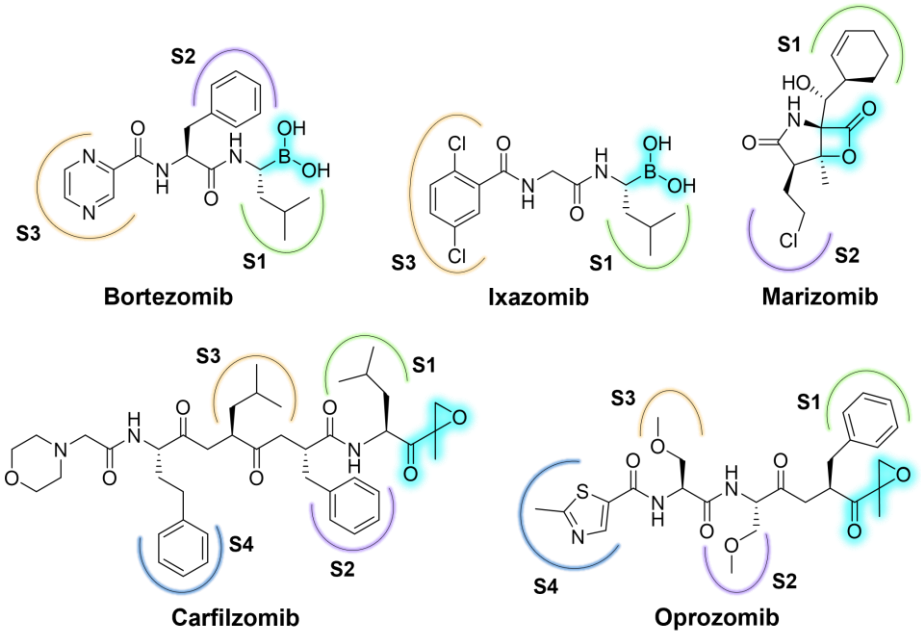


Figure 1.10 Representation of the 20S proteasome inhibitors approved by FDA (Bortezomib, Carfilzomib, Ixazomib) or currently in clinical trials (Marizomib, Oprozomib).

Despite the success of these first inhibitors, the efforts for improving proteasome inhibitors continue. In fact, two new compounds are under clinical trials: Marizomib (NPI-0052, salinosporamide A) and oprozomib

(ONX-0912). Marizomib is a natural proteasome inhibitor extracted from the marine actinomycete *Salinospora tropica*. It is a highly selective γ-lactam-β-lactone that irreversibly inhibits the three different active sites of the proteasome, with different potency, being more potent in the β5 and β2 active sites.[172] Marizomib is in clinical trials for multiple myeloma (MMM) (Phase 2), newly diagnosed glioblastoma (Phase 3), lymphomas (Phase 1), and advanced cancer (Phase 1). On the other side, oprozomib is the next generation of α,β-epoxyketone inhibitors, which has been developed to have an orally bioavailable inhibitor of this class.[195] Oprozomib is in clinical trials for MMM (Phase1b/2), solid tumors (Phase1), RRMM (Phase 1), and advanced non-central nervous system (CNS) malignancies (Phase 1).

1.4. Computational studies of the 20S proteasome

Modeling the 20S proteasome is a difficult task because of the size and complexity of this giant molecular system. Taking as an example a human 20S proteasome structure (PDB: 5LF1),[196] it contains 47872 protein heavy atoms (before adding hydrogen atoms, counterions, and solvation water molecules). As explained before, the 20S proteasome consists of four heptameric rings stacked in a cylinder-like shape. This leads to a total of 28 protein chains with an average of 230 residues. Therefore, building a model for such a system is challenging and requires considering the type of calculations that are achievable taking into account the size of the system. Although the available results of experimental studies involving the 20S proteasome are extensive, the number of computational studies done on the very same system is limited. Moreover, just a few attempts were done to study the system using advanced computational techniques.[196–202] These studies can be classified by the size of the model employed and the computational approach used by the researchers to solve particular problems. Previous studies have been based on two main types of potentials: Mata and co-workers[196,197] reported their results based on QM-cluster calculations, and hybrid QM/MM potentials were used with different approaches. Mihalovits et al.[198] reported potentials of mean force (PMFs) results corrected by QM-cluster calculations at a higher level of theory; Wei et al.[199–201] use the FEP methodology for their reported results; and Saha et al.[202] use the empirical valence bond (EVB) methodology for obtaining their results. The differences in methodology also determine the size of the system employed. While QM-cluster models in the reported studies contain only a few residues of the system, reported hybrid QM/MM models vary in size, consisting of one or two subunits of the 20S proteasome where only the active site chains are selected, or including the full 20S proteasome.

For instance, small QM-cluster models used by Mata and co-workers to study the inhibition mechanism of dihydroeponemycin [153] explain the existence of a seven-member ring product of inhibition observed in a crystallographic structure published in the same study. The results were obtained accounting only the participation of the catalytic residues (Thr1-Lys33-Asp17) placed in a continuum water solvent surrounding. In this

study, the free energy profile was obtained by computing the stationary points along the inhibition path at the B3LYP-D3(BJ)/def2-SVP level of theory.[203–206] Afterwards, corrections computed from frequency analysis under the rigid-rotor harmonic approximation at the same level of theory were applied. Another study involving QM-cluster model by Mata and co-workers[197] was designed to model the acid-base equilibria of the β2 active site with a substrate model, the inhibitors with the epoxy- and boronic-acid warheads. Additionally, in this study, replica-exchange constant pH molecular mechanics simulations with a larger model consisting of one subunit in the β2 active site and the inhibitor were reported. The level of theory employed in cluster model calculations was the B3LYP-D3(BJ)/def2-SVP level of theory, as in their previous study. The obtained results show that coupled acid-base equilibria need to be considered when modeling inhibition, although the coupling between Lys33, and Thr1 is not affected by the presence of the studied inhibitors. Nevertheless, it is important to take into account when using QM-cluster models that this approach can give reliable energetics when the size of the model is large enough.[207] However, in this case the size of the model was too small. Studies are showing consistent results employing QM-cluster models and QM/MM methodologies,[84,207–209] but this conclusion can be very dependent on the studied enzyme and a generalization should be avoided.[207]

A study based on the use of a small model containing just the β5i subunit of the protein was recently reported by Mihalovits et al.[198] where the chemical reaction of oxathiazolones covalent inhibitors selective for the immunoproteasome was reported. The authors employed a series of umbrella sampling (US) MD simulations on selected windows of previous steered MD (SMD) simulations on the assumed reaction steps. The calculations were based on the use of hybrid QM/MM potentials, with the QM subset of atoms (the ligand, Thr1, and the CαH-NH fragment of Thr2) described with the DFTB3[210] semiempirical method, and the FF14SB force-field was used to treat the rest of the protein. Finally, the obtained reactant, intermediates, and product along the path were modeled as a cluster of the QM atoms to correct the DFTB3 energies. An optimization at B3LYP/6-31G++(d,p) in continuum water solvent of each QM-cluster was done, followed by gas-phase single-point calculations at DFTB and ωB97XD/aug-cc-PVTZ levels of theory. The final free energy profile, in terms of the PMF, allowed identifying the rate limiting step. Additionally,

the equilibrium constant (K_I) and the k_{inact}/K_I ratio was reported for a set of compounds based on classical thermodynamic integration. This approach ignores the contribution of the neighbouring subunit of the active site (β6), which participates in non-covalent interactions of the substrate,[211] and assumes the possible errors related to QM-cluster models explained above.

Wei et al.[199–201] used a model consisting of two subunits of the protein, the β5 active site, and the contiguous β6, which is part of the chymotrypsin-like binding pocket. Here, the free energy profile of the mechanism was computed performing Monte Carlo (MC) molecular dynamic simulations and FEP[212] calculations along a previously computed minimum energy path (MEP)[213] at B3LYP/AMBER level of theory. According to their results, the number of steps of the obtained mechanism highly depended on the initially computed MEP. The results for the inhibition mechanism of the epoxyketone epoxomicin[199] and the vinyl amide syringolin A[200] were reported, as well as the peptide metabolism reaction pathway.[201] As later pointed out by Saha et al.,[202] the free energy calculations of the work of Wei et al. were uncoupled from the reaction coordinate, which might lead to problems of evaluation of crucial electrostatic contributions.

In fact, Saha et al. employed a larger model consisting of the full 20S proteasome to study the peptide metabolism reaction pathway considering as substrate the small artificial fluorogenic peptide, Suc-LLVY-AMC.[202] In this case the EVB[214–216] was employed to generate solution potential energy surfaces (PESs). These were calibrated by using a previous *ab initio* study,[217] and *ab initio* B3LYP/6-31G(d)//HF/6-31G(d) results were combined with Langevin dipoles[218] calculations and a careful comparison to available experiments, using a small model. Finally, the FEP combined with umbrella sampling (US) sampling technique[219] was employed to obtain the final EVB free energy surfaces (FESs), in water, in the small model, and within the full protein. The activation barriers and reaction free energies of the entire reaction profile were evaluated for the catalytic steps, i.e., acylation and deacylation, finding two possible pathways where the nucleophilic attack was always the rate-limiting step. Since the experimental measurement accounts only for the acylation step and the observed barrier corresponds to a k_{cat} of 18.6 kcal·mol^{-1},[220] the reported result for the acylation step, 14.5±2 kcal·mol^{-1}, underestimated the experimental value. Therefore,

although the EVB method involves full configurational averaging and the change of the charges of the reacting (QM) atoms during the sampling process, the initial PESs were calibrated using acid-base catalysis in serine proteases.[217] Additionally, this type of study relay on experimental information for the generation of solution potential surfaces, which depends on the analysis done.[217] Finally, it is noticeable that the reported results by Saha et al. show step wise mechanisms when a proton transfer occurs before or after a heavy-heavy atom bond formation or breaking, but these step wise processes are not confirmed since the transition states localized along the mechanism were not provided.

Bibliography

(1) Deberardinis, R. J.; Thompson, C. B. Cellular Metabolism and Disease: What Do Metabolic Outliers Teach Us? *Cell* **2012**, *148*, 1132–1144.

(2) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.

(3) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Trans. Faraday Soc.* **1935**, *31*, 875–894.

(4) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(5) Heckmann, C. M.; Paradisi, F. Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *ChemCatChem* **2020**, *12*, 6082–6102.

(6) Buller, A. R.; Townsend, C. A. Intrinsic Evolutionary Constraints on Protease Structure, Enzyme Acylation, and the Identity of the Catalytic Triad. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E653 LP-E661.

(7) Auld, D. S. Carboxypeptidase A. In *Handbook of Proteolytic Enzymes*; Rawlings, N. D., Salvesen, G. B. T.-H. of P. E. (Third E., Eds.; Academic Press, 2013; Vol. 1, pp 1289–1301.

(8) KOSHLAND, D. E. Enzyme Flexibility and Enzyme Action. *J. Cell. Comp. Physiol.* **1959**, *54*, 245–258.

(9) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **1958**, *44*, 98–104.

(10) López-Otín, C.; Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **2008**, *283*, 30433–30437.

(11) Scott, C. J.; Taggart, C. C. Biologic Protease Inhibitors as Novel Therapeutic Agents. *Biochimie* **2010**, *92*, 1681–1688.

(12) Costantino, C. M.; Ploegh, H. L.; Hafler, D. A. Cathepsin S Regulates Class II MHC Processing in Human CD4 + HLA-DR + T Cells . *J. Immunol.* **2009**, *183*, 945–952.

(13) Khan, A. R.; James, M. N. G. Molecular Mechanisms for the Conversion of Zymogens to Active Proteolytic Enzymes. *Protein Sci.* **1998**, *7*, 815–836.

(14) Allan, L. A.; Clarke, P. R. Apoptosis and Autophagy: Regulation of Caspase-9 by Phosphorylation. *FEBS J.* **2009**, *276*, 6063–6073.

(15) Di Cera, E. Thrombin. *Mol. Aspects Med.* **2008**, *29*, 203–254.

(16) Ra, H. J.; Parks, W. C. Control of Matrix Metalloproteinase Catalytic Activity. *Matrix Biol.* **2007**, *26*, 587–596.

(17) Korkmaz, B.; Moreau, T.; Gauthier, F. Neutrophil Elastase, Proteinase 3 and Cathepsin G: Physicochemical Properties, Activity and Physiopathological Functions. *Biochimie* **2008**, *90*, 227–242.

(18) Rawlings, N. D.; Tolle, D. P.; Barrett, A. J. Evolutionary Families of Peptidase Inhibitors. *Biochem. J.* **2004**, *378*, 705–716.

(19) Turk, B. Targeting Proteases: Successes, Failures and Future Prospects. *Nat. Rev. Drug Discov.* **2006**, *5*, 785–799.

(20) Agbowuro, A. A.; Huston, W. M.; Gamble, A. B.; Tyndall, J. D. A. Proteases and Protease Inhibitors in Infectious Diseases. *Med. Res. Rev.* **2018**, *38*, 1295–1331.

(21) Abbenante, G.; Fairlie, D. Protease Inhibitors in the Clinic. *Med. Chem.* **2005**, *1*, 71–104.

(22) Tuley, A.; Fast, W. The Taxonomy of Covalent Inhibitors. *Biochemistry* **2018**, *57*, 3326–3337.

(23) Noe, M. C.; Gilbert, A. M. Targeted Covalent Enzyme Inhibitors. In *Annual Reports in Medicinal Chemistry*; Desai, M. C. B. T.-A. R. in M. C., Ed.; Academic Press, 2012; Vol. 47, pp 413–439.

(24) Malla, T. R.; Tumber, A.; John, T.; Brewitz, L.; Strain-Damerell, C.; Owen, C. D.; Lukacik, P.; Chan, H. T. H.; Maheswaran, P.; Salah, E.; Duarte, F.; Yang, H.; Rao, Z.; Walsh, M. A.; Schofield, C. J. Mass Spectrometry Reveals Potential of β-Lactams as SARS-CoV-2 Mproinhibitors. *Chem. Commun.* **2021**, *57*, 1430–1433.

(25) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie - Int. Ed.* **2016**, *55*, 13408–13421.

(26) Imming, P.; Sinning, C.; Meyer, A. Drugs, Their Targets and the Nature and Number of Drug Targets. *Nat. Rev. Drug Discov.* **2006**, *5*, 821–834.

(27) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730.

(28) Sutanto, F.; Konstantinidou, M.; Dömling, A. Covalent Inhibitors: A Rational Approach to Drug Discovery. *RSC Med. Chem.* **2020**, *11*, 876–884.

(29) Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discov. Today* **2015**, *20*, 1061–1073.

(30) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of

Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.

(31) Potashman, M. H.; Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J. Med. Chem.* **2009**, *52*, 1231–1246.

(32) Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C.; Johnson, T. O.; Joslyn, C.; Kath, J. C.; Niessen, S.; Roberts, L. R.; Schnute, M. E.; Wang, C.; Hulce, J. J.; Wei, B.; Whiteley, L. O.; Hayward, M. M.; Cravatt, B. F. A Road Map to Evaluate the Proteome-Wide Selectivity of Covalent Kinase Inhibitors. *Nat. Chem. Biol.* **2014**, *10*, 760–767.

(33) Kalgutkar, A. S.; Dalvie, D. K. Drug Discovery for a New Generation of Covalent Drugs. *Expert Opin. Drug Discov.* **2012**, *7*, 561–581.

(34) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem.* **2010**, *2*, 949–964.

(35) Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009**, *52*, 225–233.

(36) Ghosh, A. K.; Samanta, I.; Mondal, A.; Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **2019**, *14*, 889–906.

(37) Robertson, J. G. Mechanistic Basis of Enzyme-Targeted Drugs. *Biochemistry* **2005**, *44*, 5561–5571.

(38) Swinney, D. C. Biochemical Mechanisms of Drug Action: What Does It Take for Success? *Nat. Rev. Drug Discov.* **2004**, *3*, 801–808.

(39) Plapp, B. V. [25] Application of Affinity Labeling for Studying Structure and Function of Enzymes. In *Methods in Enzymology*; Purich, D. L. B. T.-M. in E., Ed.; Academic Press, 1982; Vol. 87, pp 469–499.

(40) Gehringer, M.; Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2019**, *62*, 5673–5724.

(41) Turk, B.; Turk, D.; Salvesen, G. Regulating Cysteine Protease Activity: Essential Role of Protease Inhibitors As Guardians and Regulators. *Current Pharmaceutical Design*. 2005, pp 1623–1637.

(42) Brew, K.; Dinakarpandian, D.; Nagase, H. Tissue Inhibitors of Metalloproteinases: Evolution, Structure and Function. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **2000**, *1477*, 267–283.

(43) Dubrez-Daloz, L.; Dupoux, A.; Cartier, J. IAPs: More than Just Inhibitors

of Apoptosis Proteins. *Cell Cycle* **2008**, *7*, 1036–1046.

(44) Lee, C.; Bongcam-Rudloff, E.; Sollner, C.; Jahnen-Dechent, W.; Claesson-Welsh, L. Type 3 Cystatins; Fetuins, Kininogen and Histidine-Rich Glycoprotein. *Front. Biosci.* **2009**, *14*, 2911–2922.

(45) Moreau, T.; Baranger, K.; Dadé, S.; Dallet-Choisy, S.; Guyot, N.; Zani, M. L. Multifaceted Roles of Human Elafin and Secretory Leukocyte Proteinase Inhibitor (SLPI), Two Serine Protease Inhibitors of the Chelonianin Family. *Biochimie* **2008**, *90*, 284–295.

(46) Ascenzi, P.; Bocedi, A.; Bolognesi, M.; Spallarossa, A.; Coletta, M.; Cristofaro, R.; Menegatti, E. The Bovine Basic Pancreatic Trypsin Inhibitor (Kunitz Inhibitor): A Milestone Protein. *Current Protein & Peptide Science*. 2005, pp 231–251.

(47) Awoonor-Williams, E.; Walsh, A. G.; Rowley, C. N. Modeling Covalent-Modifier Drugs. *Biochim. Biophys. Acta - Proteins Proteomics* **2017**, *1865*, 1664–1675.

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

(49) Lineweaver, H.; Burk, D. The Determination of Enzyme Dissociation Constants. *J. Am. Chem. Soc.* **1934**, *56*, 658–666.

(50) Nelson, D. L.; Cox, M. M.; Lehninger, A. L. *Lehninger: Principles of Biochemistry*, 4th ed.; W. H. Freeman and Co: New York, 2013.

(51) Schwartz, P. A.; Kuzmic, P.; Solowiej, J.; Bergqvist, S.; Bolanos, B.; Almaden, C.; Nagata, A.; Ryan, K.; Feng, J.; Dalvie, D.; Kath, J. C.; Xu, M.; Wani, R.; Murray, B. W. Covalent EGFR Inhibitor Analysis Reveals Importance of Reversible Interactions to Potency and Mechanisms of Drug Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 173–178.

(52) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discov.* **2006**, *5*, 730–739.

(53) Tummino, P. J.; Copeland, R. A. Residence Time of Receptor - Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, *47*, 5481–5492.

(54) Swinney, D. C. Molecular Mechanism of Action (MMoA) in Drug Discovery. In *Annual Reports in Medicinal Chemistry*; Macor, J. E. B. T.-A. R. in M. C., Ed.; Academic Press, 2011; Vol. 46, pp 301–317.

(55) Yung-Chi, C.; Prusoff, W. H. Relationship between the Inhibition

Constant (KI) and the Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.

(56) Holdgate, G. A.; Meek, T. D.; Grimley, R. L. Mechanistic Enzymology in Drug Discovery: A Fresh Perspective. *Nat. Rev. Drug Discov.* **2018**, *17*, 115–132.

(57) Breinbauer, R. *Evaluation of Enzyme Inhibitors in Drug Discovery. By Robert A. Copeland.*, 2nd ed.; Wiley: Hoboken N.J., 2005; Vol. 44.

(58) Mah, R.; Thomas, J. R.; Shafer, C. M. Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorganic Med. Chem. Lett.* **2014**, *24*, 33–39.

(59) Swinney, D. Biochemical Mechanisms of New Molecular Entities (NMEs) Approved by United States FDA During 2001-2004: Mechanisms Leading to Optimal Efficacy and Safety. *Current Topics in Medicinal Chemistry*. 2006, pp 461–478.

(60) Świderek, K.; Paneth, P. Binding Isotope Effects. *Chem. Rev.* **2013**, *113*, 7851–7879.

(61) Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and Hydrophilic Binding Sites of HIV-1 RT. *J. Phys. Chem. B* **2015**, *119*, 917–927.

(62) Mercedes-Camacho, A. Y.; Mullins, A. B.; Mason, M. D.; Xu, G. G.; Mahoney, B. J.; Wang, X.; Peng, J. W.; Etzkorn, F. A. Kinetic Isotope Effects Support the Twisted Amide Mechanism of Pin1 Peptidyl-Prolyl Isomerase. *Biochemistry* **2013**, *52*, 7707–7713.

(63) Patrick, G. L. *An Introduction to Medicinal Chemistry*, 6th ed.; Oxford University Press: New York NY, 1995.

(64) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862–865.

(65) Singh, J.; Petter, R. C.; Kluge, A. F. Targeted Covalent Drugs of the Kinase Family. *Curr. Opin. Chem. Biol.* **2010**, *14*, 475–480.

(66) Leproult, E.; Barluenga, S.; Moras, D.; Wurtz, J. M.; Winssinger, N. Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors. *J. Med. Chem.* **2011**, *54*, 1347–1355.

(67) Cohen, M. S.; Zhang, C.; Shokat, K. M.; Taunton, J. Biochemistry: Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science (80-.).* **2005**, *308*, 1318–1321.

(68) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(69) Lohning, A. E.; Levonis, S. M.; Williams-Noonan, B.; Schweiker, S. S. A Practical Guide to Molecular Docking and Homology Modelling for Medicinal Chemists. *Current Topics in Medicinal Chemistry*. 2017, pp 2023–2040.

(70) Shen, J.; Cheng, F.; Xu, Y.; Li, W.; Tang, Y. Estimation of ADME Properties with Substructure Pattern Recognition. *J. Chem. Inf. Model.* **2010**, *50*, 1034–1041.

(71) Butina, D.; Segall, M. D.; Frankcombe, K. Predicting ADME Properties in Silico: Methods and Models. *Drug Discov. Today* **2002**, *7*, S83–S88.

(72) Woo, H. J.; Roux, B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825–6830.

(73) Olsson, M. H. M.; Mavri, J.; Warshel, A. Transition State Theory Can Be Used in Studies of Enzyme Catalysis: Lessons from Simulations of Tunnelling and Dynamical Effects in Lipoxygenase and Other Systems. *Philos. Trans. R. Soc. B Biol. Sci.* **2006**, *361*, 1417–1432.

(74) Roston, D.; Cui, Q. QM/MM Analysis of Transition States and Transition State Analogues in Metalloenzymes. In *Methods in Enzymology*; Voth, G. A. B. T.-M. in E., Ed.; Academic Press, 2016; Vol. 577, pp 213–250.

(75) Yuriev, E.; Agostino, M.; Ramsland, P. A. Challenges and Advances in Computational Docking: 2009 in Review. *J. Mol. Recognit.* **2011**, *24*, 149–164.

(76) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H. Structure-Based Virtual Screening for Drug Discovery: A Problem-Centric Review. *AAPS J.* **2012**, *14*, 133–141.

(77) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.

(78) Wang, J.; Deng, Y.; Roux, B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

(79) Michel, J.; Essex, J. W. Prediction of Protein-Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. *J. Comput. Aided. Mol. Des.* **2010**, *24*, 639–658.

(80) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.;

Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.

(81) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(82) Siegbahn, P. E. M.; Himo, F. The Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 323–336.

(83) Siegbahn, P. E. M.; Himo, F. Recent Developments of the Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *Journal of Biological Inorganic Chemistry*. 2009, pp 643–651.

(84) Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **2017**, *139*, 6780–6786.

(85) Paasche, A.; Schirmeister, T.; Engels, B. Benchmark Study for the Cysteine-Histidine Proton Transfer Reaction in a Protein Environment: Gas Phase, COSMO, QM/MM Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1765–1777.

(86) Warshel, A.; Papazyan, A. Electrostatic Effects in Macromolecules: Fundamental Concepts and Practical Modeling. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211–217.

(87) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210–3235.

(88) Ruiz-Pernía, J. J.; Martí, S.; Moliner, V.; Tuñón, I. A Novel Strategy to Study Electrostatic Effects in Chemical Reactions: Differences between the Role of Solvent and the Active Site of Chalcone Isomerase in a Michael Addition. *J. Chem. Theory Comput.* **2012**, *8*, 1532–1535.

(89) Galmés, M.; García-Junceda, E.; Świderek, K.; Moliner, V. Exploring the Origin of Amidase Substrate Promiscuity in CALB by a Computational Approach. *ACS Catal.* **2020**, *10*, 1938–1946.

(90) De Raffele, D.; Martí, S.; Moliner, V. Understanding the Directed Evolution of de Novo Retro-Aldolases from QM/MM Studies. *ACS Catal.* **2020**, *10*, 7871–7883.

(91) Hyams, J. S. *Gastroenterology and Nutrition*, 2nd ed.; Elsevier Saunders: Philadelphia PA, 2001; Vol. 13.

(92) Tanaka, K. The Proteasome: Overview of Structure and Functions. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.* **2009**, *85*, 12–36.

(93) Etlinger, J. D.; Goldberg, A. L. A Soluble ATP Dependent Proteolytic System Responsible for the Degradation of Abnormal Proteins in Reticulocytes. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 54–58.

(94) Ciehanover, A.; Hod, Y.; Hershko, A. A Heat-Stable Polypeptide Component of an ATP-Dependent Proteolytic System from Reticulocytes. *Biochem. Biophys. Res. Commun.* **1978**, *81*, 1100–1105.

(95) Goldstein, G.; Scheid, M.; Hammerling, U.; Schlesinger, D. H.; Niall, H. D.; Boyse, E. A. Isolation of a Polypeptide That Has Lymphocyte Differentiating Properties and Is Probably Represented Universally in Living Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 11–15.

(96) Harris, J. R. The Isolation and Purification of a Macromolecular Protein Component from the Human Erythrocyte Ghost. *BBA - Protein Struct.* **1969**, *188*, 31–42.

(97) Hase, J.; Kobashi, K.; Nakai, N.; Mitsui, K.; Iwata, K.; Takadera, T. The Quaternary Structure of Carp Muscle Alkaline Protease. *BBA - Enzymol.* **1980**, *611*, 205–213.

(98) Arrigo, A. P.; Tanaka, K.; Goldberg, A. L.; Welch, W. J. Identity of the 19S "prosome" Particle with the Large Multifunctional Protease Complex of Mammalian Cells (the Prot Easome). *Nature* **1988**, *331*, 192–194.

(99) Wilk, S.; Orlowski, M. Cation-Sensitive Neutral Endopeptidase: Isolation and Specificity of the Bovine Pituitary Enzyme. *J. Neurochem.* **1980**, *35*, 1172–1182.

(100) Eytan, E.; Ganoth, D.; Armon, T.; Hershko, A. ATP-Dependent Incorporation of 20S Protease into the 26S Complex That Degrades Proteins Conjugated to Ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 7751–7755.

(101) The Nobel Prize in Chemistry 2005 NobelPrize.org. Nobel Prize Outreach AB 2021 https://www.nobelprize.org/prizes/chemistry/2005/summary/ (accessed 2021 -09 -04).

(102) Hershko, A.; Ciechanover, A.; Heller, H.; Haas, A. L.; Rose, I. A. Proposed Role of ATP in Protein Breakdown: Conjugation of Protein with Multiple Chains of the Polypeptide of ATP-Dependent Proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 1783–1786.

(103) Ciechanover, A.; Elias, S.; Heller, H. Characterization of the Heatstable Polypeptide of the ATP-Dependent Proteolytic System from Reticulocytes. *J. Biol. Chem.* **1980**, *255*, 7525–7528.

(104) DeMartino, G. N.; Proske, R. J.; Moomaw, C. R.; Strong, A. A.; Song, X.; Hisamatsu, H.; Tanaka, K.; Slaughter, C. A. Identification, Purification, and Characterization of a PA700-Dependent Activator of the Proteasome. *J. Biol. Chem.* **1996**, *271*, 3112–3118.

(105) Adams, G. M.; Crotchett, B.; Slaughter, C. A.; DeMartino, G. N.; Gogol, E. P. Formation of Proteasome-PA700 Complexes Directly Correlates with Activation of Peptidase Activity. *Biochemistry* **1998**, *37*, 12927–12932.

(106) Thibaudeau, T. A.; Smith, D. M. A Practical Review of Proteasome Pharmacology. *Pharmacol. Rev.* **2019**, *71*, 170–197.

(107) Livneh, I.; Cohen-Kaplan, V.; Cohen-Rosenzweig, C.; Avni, N.; Ciechanover, A. The Life Cycle of the 26S Proteasome: From Birth, through Regulation and Function, and onto Its Death. *Cell Res.* **2016**, *26*, 869–885.

(108) Ciechanover, A. Intracellular Protein Degradation: From a Vague Idea through the Lysosome and the Ubiquitin-Proteasome System, and onto Human Diseases and Drug Targeting (Nobel Lecture). *Angew. Chemie Int. Ed.* **2005**, *44*, 5944–5967.

(109) Haselbach, D.; Schrader, J.; Lambrecht, F.; Henneberg, F.; Chari, A.; Stark, H. Long-Range Allosteric Regulation of the Human 26S Proteasome by 20S Proteasome-Targeting Cancer Drugs. *Nat. Commun.* **2017**, *8*, 1–8.

(110) Mishra, R.; Upadhyay, A.; Prajapati, V. K.; Mishra, A. Proteasome-Mediated Proteostasis: Novel Medicinal and Pharmacological Strategies for Diseases. *Med. Res. Rev.* **2018**, *38*, 1916–1973.

(111) Hershko, A.; Ciechanover, A. The Ubiquitin System. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.

(112) Peters, J. M. The Anaphase-Promoting Complex: Proteolysis in Mitosis and Beyond. *Mol. Cell* **2002**, *9*, 931–943.

(113) Bassermann, F.; Eichner, R.; Pagano, M. The Ubiquitin Proteasome System - Implications for Cell Cycle Control and the Targeted Treatment of Cancer. *Biochim. Biophys. Acta - Mol. Cell Res.* **2014**, *1843*, 150–162.

(114) Bonvini, P.; Rosa, H. D.; Vignes, N.; Rosolen, A. Ubiquitination and Proteasomal Degradation of Nucleophosmin-Anaplastic Lymphoma Kinase Induced by 17-Allylamino-Demethoxygeldanamycin: Role of the Co-Chaperone Carboxyl Heat Shock Protein 70-Interacting Protein. *Cancer Res.* **2004**, *64*, 3256–3264.

(115) Didcock, L.; Young, D. F.; Goodbourn, S.; Randall, R. E. The V Protein of Simian Virus 5 Inhibits Interferon Signalling by Targeting STAT1 for Proteasome-Mediated Degradation. *J. Virol.* **1999**, *73*, 9928–9933.

(116) Desai, S. D.; Li, T. K.; Rodriguez-Bauman, A.; Liu, L. F.; Rubin, E. H. Ubiquitin/26s Proteasome-Mediated Degradation of Topoisomerase I as a Resistance Mechanism to Camptothecin in Tumor Cells. *Cancer Res.* **2001**, *61*, 5926–5932.

(117) Ratner, J. N.; Balasubramanian, B.; Corden, J.; Warren, S. L.; Bregman, D. B. Ultraviolet Radiation-Induced Ubiquitination and Proteasomal Degradation of the Large Subunit of RNA Polymerase II: Implications for Transcription- Coupled DNA Repair. *J. Biol. Chem.* **1998**, *273*, 5184–5189.

(118) Rui, L.; Fisher, T. L.; Thomas, J.; White, M. F. Regulation of Insulin/Insulin-like Growth Factor-1 Signaling by Proteasome-Mediated Degradation of Insulin Receptor Substrate-2. *J. Biol. Chem.* **2001**, *276*, 40362–40367.

(119) Sun, X. J.; Goldberg, J. L.; Qiao, L. Y.; Mitchell, J. J. Insulin-Induced Insulin Receptor Substrate-1 Degradation Is Mediated by the Proteasome Degradation Pathway. *Diabetes* **1999**, *48*, 1359–1364.

(120) McDonald, E. R.; El-Deiry, W. S. Suppression of Caspase-8- and -10-Associated RING Proteins Results in Sensitization to Death Ligands and Inhibition of Tumor Cell Growth. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6170–6175.

(121) Suzuki, Y.; Nakabayashi, Y.; Takahashi, R. Ubiquitin-Protein Ligase Activity of X-Linked Inhibitor of Apoptosis Protein Promotes Proteasomal Degradation of Caspase-3 and Enhances Its Anti-Apoptotic Effect in Fas-Induced Cell Death. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8662–8667.

(122) Miller, D. M.; Thomas, S. D.; Islam, A.; Muench, D.; Sedoris, K. C-Myc and Cancer Metabolism. *Clin. Cancer Res.* **2012**, *18*, 5546–5553.

(123) Gregory, M. A.; Hann, S. R. C-Myc Proteolysis by the Ubiquitin-Proteasome Pathway: Stabilization of c-Myc in Burkitt's Lymphoma Cells. *Mol. Cell. Biol.* **2000**, *20*, 2423–2435.

(124) Kelly, J. M.; Summers, M.; Park, H. S.; Milligan, L. P.; McBride, B. W. *Cellular Energy Metabolism and Regulation*; Academic Press: New York, 1991; Vol. 74.

(125) Bhoj, V. G.; Chen, Z. J. Ubiquitylation in Innate and Adaptive Immunity. *Nature* **2009**, *458*, 430–437.

(126) Zhang, M.; Wang, L.; Zhao, X.; Zhao, K.; Meng, H.; Zhao, W.; Gao, C. TRAF-Interacting Protein (TRIP) Negatively Regulates IFN-β Production and Antiviral Response by Promoting Proteasomal Degradation of TANK-Binding Kinase 1. *J. Exp. Med.* **2012**, *209*, 1703–1711.

(127) Glickman, M. H.; Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, *82*, 373–428.

(128) Sakamoto, K. M. Ubiquitin-Dependent Proteolysis: Its Role in Human Diseases and the Design of Therapeutic Strategies. *Mol. Genet. Metab.* **2002**, *77*, 44–56.

(129) Layfield, R.; Lowe, J.; Bedford, L. The Ubiquitin–Proteasome System and Neurodegenerative Disorders. *Essays Biochem.* **2005**, *41*, 157–171.

(130) McKinnon, C.; Tabrizi, S. J. The Ubiquitin-Proteasome System in Neurodegeneration. *Antioxidants Redox Signal.* **2014**, *21*, 2302–2321.

(131) Dou, Q. P.; Smith, D. M.; Daniel, K. G.; Kazi, A. Interruption of Tumor Cell Cycle Progression through Proteasome Inhibition: Implications for Cancer Therapy. *Prog. Cell Cycle Res.* **2003**, *5*, 441–446.

(132) Papandreou, C. N. The Proteasome as a Target for Cancer Treatment: Focus on Bortezomib. *Am. J. Cancer* **2005**, *4*, 359–372.

(133) Dick, L. R.; Fleming, P. E. Building on Bortezomib: Second-Generation Proteasome Inhibitors as Anti-Cancer Therapy. *Drug Discov. Today* **2010**, *15*, 243–249.

(134) Schmidt, M.; Finley, D. Regulation of Proteasome Activity in Health and Disease. *Biochim. Biophys. Acta - Mol. Cell Res.* **2014**, *1843*, 13–25.

(135) Zmuda, F.; Sastry, L.; Shepherd, S. M.; Jones, D.; Scott, A.; Craggs, P. D.; Cortes, A.; Gray, D. W.; Torrie, L. S.; De Rycker, M. Identification of Novel Trypanosoma Cruzi Proteasome Inhibitors Using a Luminescence-Based High-Throughput Screening Assay. *Antimicrob. Agents Chemother.* **2019**, *63*, e00309-19.

(136) Cromm, P. M.; Crews, C. M. The Proteasome in Modern Drug Discovery: Second Life of a Highly Valuable Drug Target. *ACS Cent. Sci.* **2017**, *3*, 830–838.

(137) Aminake, M. N.; Arndt, H. D.; Pradel, G. The Proteasome of Malaria Parasites: A Multi-Stage Drug Target for Chemotherapeutic Intervention? *Int. J. Parasitol. Drugs Drug Resist.* **2012**, *2*, 1–10.

(138) Le Chapelain, C.; Groll, M. Rational Design of Proteasome Inhibitors as Antimalarial Drugs. *Angew. Chemie - Int. Ed.* **2016**, *55*, 6370–6372.

(139) Gandotra, S.; Schnappinger, D.; Monteleone, M.; Hillen, W.; Ehrt, S. In Vivo Gene Silencing Identifies the Mycobacterium Tuberculosis Proteasome as Essential for the Bacteria to Persist in Mice. *Nat. Med.* **2007**, *13*, 1515–1520.

(140) Dahlmann, B. Role of Proteasomes in Disease. *BMC Biochem.* **2007**, *8*, S3.

(141) Rideout, H. J.; Larsen, K. E.; Sulzer, D.; Stefanis, L. Proteasomal Inhibition Leads to Formation of Ubiquitin/α-Synuclein-Immunoreactive Inclusions in PC12 Cells. *J. Neurochem.* **2001**, *78*, 899–908.

(142) Keller, J. N.; Hanni, K. B.; Markesbery, W. R. Impaired Proteasome Function in Alzheimer's Disease. *J. Neurochem.* **2000**, *75*, 436–439.

(143) McNaught, K. S. .; Jenner, P. Proteasomal Function Is Impaired in Substantia Nigra in Parkinson's Disease. *Neurosci. Lett.* **2001**, *297*, 191–194.

(144) Kabashi, E.; Agar, J. N.; Taylor, D. M.; Minotti, S.; Durham, H. D. Focal Dysfunction of the Proteasome: A Pathogenic Factor in a Mouse Model of Amyotrophic Lateral Sclerosis. *J. Neurochem.* **2004**, *89*, 1325–1335.

(145) Zhou, H.; Cao, F.; Wang, Z.; Yu, Z. X.; Nguyen, H. P.; Evans, J.; Li, S. H.; Li, X. J. Huntingtin Forms Toxic NH2-Terminal Fragment Complexes That Are Promoted by the Age-Dependent Decrease in Proteasome Activity. *J. Cell Biol.* **2003**, *163*, 109–118.

(146) Njomen, E.; Tepe, J. J. Proteasome Activation as a New Therapeutic Approach To Target Proteotoxic Disorders. *J. Med. Chem.* **2019**, *62*, 6469–6481.

(147) Naujokat, C.; Hoffmann, S. Role and Function of the 26S Proteasome in Proliferation and Apoptosis. *Lab. Investig.* **2002**, *82*, 965–980.

(148) Catz, S. D.; Johnson, J. L. Transcriptional Regulation of Bcl-2 by Nuclear Factor KB and Its Significance in Prostate Cancer. *Oncogene* **2001**, *20*, 7342–7351.

(149) Li, B.; Dou, Q. P. Bax Degradation by the Ubiquitin Proteasome-Dependent Pathway: Involvement in Tumor Survival and Progression. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3850–3855.

(150) Breitschopf, K.; Zeiher, A. M.; Dimmeler, S. Ubiquitin-Mediated Degradation of the Proapoptotic Active Form of Bid. A Functional Consequence on Apoptosis Induction. *J. Biol. Chem.* **2000**, *275*, 21648–21652.

(151) Reed, S. I. Ratchets and Clocks: The Cell Cycle, Ubiquitylation and Protein Turnover. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 855–864.

(152) Norez, C.; Bilan, F.; Kitzis, A.; Mettey, Y.; Becq, F. Proteasome-Dependent Pharmacological Rescue of Cystic Fibrosis Transmembrane Conductance Regulator Revealed by Mutation of Glycine 622. *J. Pharmacol. Exp. Ther.* **2008**, *325*, 89–99.

(153) Verbrugge, E. E.; Scheper, R. J.; Lems, W. F.; de Gruijl, T. D.; Jansen, G. Proteasome Inhibitors as Experimental Therapeutics of Autoimmune Diseases. *Arthritis Res. Ther.* **2015**, *17*, 17.

(154) Grziwa, A.; Baumeister, W.; Dahlmann, B.; Kopp, F. Localization of Subunits in Proteasomes from Thermoplasma Acidophilum by Immunoelectron Microscopy. *FEBS Lett.* **1991**, *290*, 186–190.

(155) Hilt, W.; Heinemeyer, W.; Wolf, D. H. Studies on the Yeast Proteasome Uncover Its Basic Structural Features and Multiple in Vivo Functions. *Enzym. Protein* **1993**, *47*, 189–201.

(156) Tamura, T.; Nagy, I.; Lupas, A.; Lottspeich, F.; Cejka, Z.; Schoofs, G.; Tanaka, K.; De Mot, R.; Baumeister, W. The First Characterization of a Eubacterial Proteasome: The 20S Complex of Rhodococcus. *Curr. Biol.* **1995**, *5*, 766–774.

(157) Bochtler, M.; Ditzel, L.; Groll, M.; Huber, R. Crystal Structure of Heat Shock Locus V (HslV) from Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6070–6074.

(158) Löwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal Structure of the 20S Proteasome from the Archaeon T. Acidophilum at 3.4 Å Resolution. *Science (80-.).* **1995**, *268*, 533–539.

(159) Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. Structure of 20S Proteasome from Yeast at 2.4 Å Resolution. *Nature* **1997**, *386*, 463–471.

(160) Zwickl, P.; Grziwa, A.; Piihler, G.; Lottspeich, F.; Baumeister, W.; Dahlmann, B. Primary Structure of the Thermoplasma Proteasome and Its Implications for the Structure, Function, and Evolution of the Multicatalytic Proteinase. *Biochemistry* **1992**, *31*, 964–972.

(161) 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.

(162) 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.

(163) Groll, M.; Heinemeyer, W.; Jäger, S.; Ullrich, T.; Bochtler, M.; Wolf, D. H.; Huber, R. The Catalytic Sites of 20S Proteasomes and Their Role in Subunit Maturation: A Mutational and Crystallographic Study. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10976–10983.

(164) Dick, T. P.; Nussbaum, A. K.; Deeg, M.; Heinemeyer, W.; Groll, M.; Schirle, M.; Keilholz, W.; Stevanović, S.; Wolf, D. H.; Huber, R.;

Rammensee, H. G.; Schild, H. Contribution of Proteasomal β-Subunits to the Cleavage of Peptide Substrates Analyzed with Yeast Mutants. *J. Biol. Chem.* **1998**, *273*, 25637–25646.

(165) Kisselev, A. F.; Garcia-Calvo, M.; Overkleeft, H. S.; Peterson, E.; Pennington, M. W.; Ploegh, H. L.; Thornberry, N. A.; Goldberg, A. L. The Caspase-like Sites of Proteasomes, Their Substrate Specificity, New Inhibitors and Substrates, and Allosteric Interactions with the Trypsin-like Sites. *J. Biol. Chem.* **2003**, *278*, 35869–35877.

(166) Huber, E. M.; Heinemeyer, W.; Li, X.; Arendt, C. S.; Hochstrasser, M.; Groll, M. A Unified Mechanism for Proteolysis and Autocatalytic Activation in the 20S Proteasome. *Nat. Commun.* **2016**, *7*, 10900.

(167) Ben-Nissan, G.; Sharon, M. Regulating the 20S Proteasome Ubiquitin-Independent Degradation Pathway. *Biomolecules*. 2014, pp 862–884.

(168) Murata, S.; Sasaki, K.; Kishimoto, T.; Niwa, S. I.; Hayashi, H.; Takahama, Y.; Tanaka, K. Regulation of CD8+ T Cell Development by Thymus-Specific Proteasomes. *Science (80-.).* **2007**, *316*, 1349–1353.

(169) Groettrup, M.; Kirk, C. J.; Basler, M. Proteasomes in Immune Cells: More than Peptide Producers? *Nat. Rev. Immunol.* **2010**, *10*, 73–78.

(170) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, *11*, 3575–3589.

(171) Rock, K. L.; Gramm, C.; Rothstein, L.; Clark, K.; Stein, R.; Dick, L.; Hwang, D.; Goldberg, A. L. Inhibitors of the Proteasome Block the Degradation of Most Cell Proteins and the Generation of Peptides Presented on MHC Class I Molecules. *Cell* **1994**, *78*, 761–771.

(172) Groll, M.; Huber, R. Inhibitors of the Eukaryotic 20S Proteasome Core Particle: A Structural Approach. *Biochim. Biophys. Acta - Mol. Cell Res.* **2004**, *1695*, 33–44.

(173) Ōmura, S.; Fujimoto, T.; Otoguro, K.; Matsuzaki, K.; Moriguchi, R.; Tanaka, H.; Sasaki, Y. Lactacystin, a Novel Microbial Metabolite, Induces Neurito-Genesis of Neuroblastoma Cells. *J. Antibiot. (Tokyo).* **1991**, *44*, 113–116.

(174) Goldberg, A. L. Functions of the Proteasome: From Protein Degradation and Immune Surveillance to Cancer Therapy. *Biochem. Soc. Trans.* **2007**, *35*, 12–17.

(175) 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.

(176) Lü, S.; Wang, J. The Resistance Mechanisms of Proteasome Inhibitor Bortezomib. *Biomark. Res.* **2013**, *1*, 13.

(177) C. Potts, B.; X. Albitar, M.; C. Anderson, K.; Baritaki, S.; Berkers, C.; Bonavida, B.; Chandra, J.; Chauhan, D.; C. Cusack, J.; Fenical, W.; M. Ghobrial, I.; Groll, M.; R. Jensen, P.; S. Lam, K.; K. Lloyd, G.; McBride, W.; J. McConkey, D.; P. Miller, C.; T.C. Neuteboom, S.; et al. Marizomib, a Proteasome Inhibitor for All Seasons: Preclinical Profile and a Framework for Clinical Trials. *Current Cancer Drug Targets*. 2011, pp 254–284.

(178) Huber, E. M. Introduction. In *Structural and Functional Characterization of the Immunoproteasome*; Huber, E. M., Ed.; Springer Theses TS; Springer International Publishing: Cham, 2013; pp 1–18.

(179) Vinitsky, A.; Michaud, C.; Orlowski, M.; Powers, J. C. Inhibition of the Chymotrypsin-like Activity of the Pituitary Multicatalytic Proteinase Complex. *Biochemistry* **1992**, *31*, 9421–9428.

(180) Kisselev, A. F.; Goldberg, A. L. Proteasome Inhibitors: From Research Tools to Drug Candidates. *Chem. Biol.* **2001**, *8*, 739–758.

(181) Adams, J.; Palombella, V. J.; Sausville, E. A.; Johnson, J.; Destree, A.; Lazarus, D. D.; Maas, J.; Pien, C. S.; Prakash, S.; Elliott, P. J. Proteasome Inhibitors: A Novel Class of Potent and Effective Antitumor Agents. *Cancer Res.* **1999**, *59*, 2615–2622.

(182) An, B.; Goldfarb, R. H.; Siman, R.; Ping Dou, Q. Novel Dipeptidyl Proteasome Inhibitors Overcome Bcl-2 Protective Function and Selectively Accumulate the Cyclin-Dependent Kinase Inhibitor P27 and Induce Apoptosis in Transformed, but Not Normal, Human Fibroblasts. *Cell Death Differ.* **1998**, *5*, 1062–1075.

(183) Masdehors, P.; Omura, S.; Merle-Beral, H.; Mentz, F.; Cosset, J. M.; Dumont, J.; Magdelénat, H.; Delic, J. Increased Sensitivity of CLL-Derived Lymphocytes to Apoptotic Death Activation by the Proteasome-Specific Inhibitor Lactacystin. *Br. J. Haematol.* **1999**, *105*, 752–757.

(184) Adams, J. Proteasome Inhibition in Cancer: Development of PS-341. *Semin. Oncol.* **2001**, *28*, 613–619.

(185) Groll, M.; Kim, K. B.; Kairies, N.; Huber, R.; Crews, C. M. Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of α',β'-Epoxyketone Proteasome Inhibitors [12]. *J. Am. Chem. Soc.* **2000**, *122*, 1237–1238.

(186) Kyung, B. K.; Myung, J.; Sin, N.; Crews, C. M. Proteasome Inhibition by

the Natural Products Epoxomicin and Dihydroeponemycin: Insights into Specificity and Potency. *Bioorganic Med. Chem. Lett.* **1999**, *9*, 3335–3340.

(187) Gallastegui, N.; Stein, M.; Schmidt, B.; Kloetzel, P. M.; Huber, R.; Groll, M. Elucidation of the α-Keto-Aldehyde Binding Mechanism: A Lead Structure Motif for Proteasome Inhibition. *Angew. Chemie - Int. Ed.* **2011**, *50*, 542–544.

(188) Bogyo, M.; McMaster, J. S.; Gaczynska, M.; Tortorella, D.; Goldberg, A. L.; Ploegh, H. Covalent Modification of the Active Site Threonine of Proteasomal β Subunits and the Escherichia Coli Homolog HslV by a New Class of Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6629–6634.

(189) Groll, M.; Schellenberg, B.; Bachmann, A. S.; Archer, C. R.; Huber, R.; Powell, T. K.; Lindow, S.; Kaiser, M.; Dudler, R. A Plant Pathogen Virulence Factor Inhibits the Eukaryotic Proteasome by a Novel Mechanism. *Nature* **2008**, *452*, 755–758.

(190) Groll, M.; Huber, R.; Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of β-Lactone Ring Opening and a Mechanism for Irreversible Binding. *J. Am. Chem. Soc.* **2006**, *128*, 5136–5141.

(191) Dou, Q. P.; Goldfarb, R. H. Bortezomib Millennium Pharmaceuticals. *IDrugs*. August 2002, pp 828–834.

(192) Kane, R. C.; Bross, P. F.; Farrell, A. T.; Pazdur, R. Velcade ® : U.S. FDA Approval for the Treatment of Multiple Myeloma Progressing on Prior Therapy . *Oncologist* **2003**, *8*, 508–513.

(193) Goldberg, A. L. Alfred L. Goldberg: Probing the Proteasome. *Trends Cell Biol.* **2016**, *26*, 792–794.

(194) Moreau, P.; Masszi, T.; Grzasko, N.; Bahlis, N. J.; Hansson, M.; Pour, L.; Sandhu, I.; Ganly, P.; Baker, B. W.; Jackson, S. R.; Stoppa, A.-M.; Simpson, D. R.; Gimsing, P.; Palumbo, A.; Garderet, L.; Cavo, M.; Kumar, S.; Touzeau, C.; Buadi, F. K.; et al. Oral Ixazomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. *N. Engl. J. Med.* **2016**, *374*, 1621–1634.

(195) Roccaro, A. M.; Sacco, A.; Aujay, M.; Ngo, H. T.; Azab, A. K.; Azab, F.; Quang, P.; Maiso, P.; Runnels, J.; Anderson, K. C.; Demo, S.; Ghobrial, I. M. Selective Inhibition of Chymotrypsin-like Activity of the Immunoproteasome and Constitutive Proteasome InWaldenström Macroglobulinemia. *Blood* **2010**, *115*, 4051–4060.

(196) Schrader, J.; Henneberg, F.; Mata, R. A.; Tittmann, K.; Schneider, T. R.; Stark, H.; Bourenkov, G.; Chari, A. The Inhibition Mechanism of Human

20S Proteasomes Enables Next-Generation Inhibitor Design. *Science (80-.).* **2016**, *353*, 594–598.

(197) Uranga, J.; Hasecke, L.; Proppe, J.; Fingerhut, J.; Mata, R. A. Theoretical Studies of the Acid-Base Equilibria in a Model Active Site of the Human 20S Proteasome. *J. Chem. Inf. Model.* **2021**, *61*, 1942–1953.

(198) Mihalovits, L. M.; Ferenczy, G. G.; Keserű, G. M. Mechanistic and Thermodynamic Characterization of Oxathiazolones as Potent and Selective Covalent Immunoproteasome Inhibitors. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4486–4496.

(199) Wei, D.; Lei, B.; Tang, M.; Zhan, C. G. Fundamental Reaction Pathway and Free Energy Profile for Inhibition of Proteasome by Epoxomicin. *J. Am. Chem. Soc.* **2012**, *134*, 10436–10450.

(200) Wei, D.; Tang, M.; Zhan, C. G. Fundamental Reaction Pathway and Free Energy Profile of Proteasome Inhibition by Syringolin A (SylA). *Org. Biomol. Chem.* **2015**, *13*, 6857–6865.

(201) Wei, D.; Fang, L.; Tang, M.; Zhan, C. G. Fundamental Reaction Pathway for Peptide Metabolism by Proteasome: Insights from First-Principles Quantum Mechanical/Molecular Mechanical Free Energy Calculations. *J. Phys. Chem. B* **2013**, *117*, 13418–13434.

(202) Saha, A.; Oanca, G.; Mondal, D.; Warshel, A. Exploring the Proteolysis Mechanism of the Proteasomes. *J. Phys. Chem. B* **2020**, *124*, 5626–5635.

(203) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(204) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(205) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(206) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(207) Prejanò, M.; Marino, T.; Russo, N. QM Cluster or QM/MM in Computational Enzymology: The Test Case of LigW-Decarboxylase . *Frontiers in Chemistry* . 2018, p 249.

(208) Himo, F. Quantum Chemical Modeling of Enzyme Active Sites and Reaction Mechanisms. *Theor. Chem. Acc.* **2006**, *116*, 232–240.

(209) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601–3658.

(210) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method: Third-Order Expansion of the Density Functional Theory Total Energy and Introduction of a Modified Effective Coulomb Interaction. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.

(211) Harshbarger, W.; Miller, C.; Diedrich, C.; Sacchettini, J. Crystal Structure of the Human 20S Proteasome in Complex with Carfilzomib. *Structure* **2015**, *23*, 418–424.

(212) Orlando, A.; Jorgensen, W. L. Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions. *Acc. Chem. Res.* **2010**, *43*, 142–151.

(213) Zhang, Y.; Liu, H.; Yang, W. Free Energy Calculation on Enzyme Reactions with an Efficient Iterative Procedure to Determine Minimum Energy Paths on a Combined Ab Initio QM/MM Potential Energy Surface. *J. Chem. Phys.* **2000**, *112*, 3483–3492.

(214) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.

(215) Kamerlin, S. C. L.; Warshel, A. The Empirical Valence Bond Model: Theory and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 30–45.

(216) Kamerlin, S. C. L.; Warshel, A. The EVB as a Quantitative Tool for Formulating Simulations and Analyzing Biological and Chemical Reactions. *Faraday Discuss.* **2010**, *145*, 71–106.

(217) Štrajbl, M.; Florián, J.; Warshel, A. Ab Initio Evaluation of the Potential Surface for General Base- Catalyzed Methanolysis of Formamide: A Reference Solution Reaction for Studies of Serine Proteases. *J. Am. Chem. Soc.* **2000**, *122*, 5354–5366.

(218) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(219) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(220) Kisselev, A. F.; Songyang, Z.; Goldberg, A. L. Why Does Threonine, and Not Serine, Function as the Active Site Nucleophile in Proteasomes? *. *J.*

Biol. Chem. **2000**, *275*, 14831–14837.

Chapter 2. Objectives

The 20S proteasome is a highly interesting enzyme because of its potential medical applications, in particular for cancer treatments. However, although it has been extensively studied over the years by different experimental and computational techniques, there are certain aspects of its operation that still require clarification. The present thesis is focused on these enigmatic areas and provides insights into the mechanism of inhibition of the 20S proteasome at the molecular level obtained based on the use of multiscale QM/MM methods. Computational studies on the inhibition of the 20S proteasome can have a significant impact on the field by clarifying the specific interactions created between the inhibitors and binding pocket of the enzyme and explaining chemical transformations occurring in its active sites. From the computational chemistry point of view, the 20S proteasome represents a challenging system because of its large size. From the biological point of view, the understanding and development of inhibitory compounds can have a direct impact on the development process of new drugs. The optimization of new lead compounds can benefit from the rationalization of the inhibitory mechanism, therefore the knowledge obtained in this work may be crucial in the design of new drugs.

Therefore, the main specific objectives of the present thesis are the following:

- Understanding the inhibition process of the β5 active site of the 20S proteasome at a molecular level.
- Clarifying the role of different types of warheads on the inhibition process.
- Understanding the mechanism of action of double reactive center epoxyketone inhibitors.
- Characterizing the mode of action of γ-lactam-β-lactone inhibitors such as salinosporamide A.
- Understanding the origin of reversibility and irreversibility of γ-lactam-β-lactone inhibitors, using as an example salinosporamide A and homo-salinosporamide A.
- Computing the binding free energies with different techniques for understanding the binding process of γ-lactam-β-lactone inhibitors.

Chapter 3. Computational framework

3.1. Structure and energy

The calculation of the energy of a molecular system mainly depends on the relative positions of the atoms that comprise it, and therefore its structure. Additionally, the stability of a structure is related to the possibility of finding a relative or an absolute minimum in the potential energy surface (PES). Therefore, through punctual energy and surface gradient calculations, it is possible to optimize the structure towards a potential energy minimum and to obtain the minimum energy value. This optimization of the structure towards a minimum and its consecutive obtention of the minimum energy can be done at different levels of theory depending on the employed Hamiltonian. The selection of the Hamiltonian usually dependents on the specific problem. Thus, when the goal is to study large systems such as proteins, the choice of the Hamiltonian is a key factor to consider. In this chapter, the employed Hamiltonians for the calculation of potential energies along the PES are going to be introduced and the advantages and disadvantages of their different energy definitions will be highlighted.

Although there is a wide range of techniques available for atomistic modeling, two broad categories can be distinguished: methods based on classical interatomic potentials and quantum mechanical methods based on electronic structure techniques. These two categories are mixed in the hybrid or multiscale QM/MM approach.

3.1.1. MM potentials

MM potentials are constructed by using classical mechanics (Newton's second law). The basic expression of these potentials is usually known as "force field", which are defined to describe molecular systems treating the atoms as spheres and the bonds as springs. Centered based on the Born-Oppenheimer approximation,[1] states that the electronic motions can be averaged since they occur much faster, and thus the energy surface is constructed through the atoms' positions and motions.[2] Initially, force fields were built on molecular spectroscopy data of small molecules and, as available data increased, it was noticed that some constants were transferable for different molecules containing the same atom-atom connections.[3] Therefore, in each force field there are basic definitions through a set of simple classical functions to represent the energy, and

parameters for each type of atom and all possible atom-atom relationships within the system is predefined, adjusted to optimize agreement with quantum calculations on small molecules and experimental data.[2] Thus, it is possible to obtain the energy of the system through the force-field equations and the established parameters.

The MM Hamiltonian or potential energy function (V_{MM}) commonly used by protein force fields can be expressed as a sum of two sets of terms, the bonding energy term ($V_{bonding}$) and the non-bonding term ($V_{non-bonding}$). This expression is sometimes called the "Class I" potential energy function.

$$V_{MM} = V_{bonding} + V_{non-bonding} \qquad \text{Eq. 3.6}$$

The bonding terms are typically bond stretching, angle bending, torsions, and out-of-plane deformations or improper torsions.

$$V_{bonding} = V_{stretch} + V_{bend} + V_{tors+improp} \qquad \text{Eq. 3.7}$$

The non-bonding terms correspond to Coulomb interaction terms between rigid point charges and the Lennard–Jones-type van der Waals terms.[4]

$$V_{non-bonding} = V_{elec} + V_{LJ} \qquad \text{Eq. 3.8}$$

In this thesis, the employed force field is the AMBER force field, an all-atom force field developed in the Kollman group[5] and further upgraded through parametrization strategy improvements and collaboration.[6,7] The general equation in the AMBER force field for obtaining the energy of the system contains the main terms explained above.

$$E_{bonds} = \sum_{bonds} k_d(d - d_0)^2 \qquad \text{Eq. 3.9}$$

$$E_{angles} = \sum_{angles} k_\theta(\theta - \theta_0)^2 \qquad \text{Eq. 3.10}$$

Within the bonds and angles terms all 1-2 interactions and 1-3 interactions, respectively, are comprised. The equations describe the bonds and angles with a harmonic potential by using Hook's law, and the parameters used

are the bond length and angle, d and θ, equilibrium values for bond length and angle, d_0 and θ_0, and the force constants defined for each bond and angle type, k_d, and k_θ, respectively.

$$E_{torsions+improper} = \sum_{dihedrals} k_\phi[1 + \cos(n\phi + \delta)] \quad \text{Eq. 3.11}$$

In the torsions expression, all 1-4 interactions are included. Additionally, the interactions where not all the atoms within the four atoms defining the torsion angle are connected by covalent bonds or improper torsion are computed within this term. These improper torsions serve to enforce planarity around sp2 central atoms. In Eq. 3.11, ϕ corresponds to the torsion, while *n* and δ correspond to the torsional multiplicity and phase, respectively, and finally, k_ϕ is the force constant for the torsions.

$$E_{electrostatic} = \sum_{\substack{non-bonded \\ pairs\ AB}} \frac{1}{4\pi\varepsilon_0}\frac{q_A q_B}{r_{AB}} \quad \text{Eq. 3.12}$$

$$E_{LJ} = \sum_{\substack{non-bonded \\ pairs\ AB}} 4\varepsilon_{AB}\left[\left(\frac{r_{eq,AB}}{r_{AB}}\right)^{12} - 2\left(\frac{r_{eq,AB}}{r_{AB}}\right)^{6}\right] \quad \text{Eq. 3.13}$$

The non-bonding terms comprise the electrostatic term, where partial charges, q_i, are introduced for each atom interacting through Coulomb's law; and the van der Waals term is computed using a Lennard-Jones potential. In expressions Eq. 3.12 and Eq. 3.13, r_{AB} corresponds to the distance between both interacting atoms, $r_{eq,AB}$ is the equilibrium distance between the same pair of atoms, ε_{AB} is the depth of the potential well or "dispersion energy" and ε_0 is the vacuum permittivity.[4] Finally, a scaling factor is applied to non-bonding interactions ranging from 0.4–1 for atoms related in 1-4 interactions and is set to zero for non-bonding interactions between atoms connected by bonds or angles.

The employed force field parameters in this thesis are the AMBER ff03[8] force field for treating the protein and the counterions added to neutralize the charge of the system, and the TIP3P[9] force field for the explicit solvent water molecules. Additionally, the General AMBER Force Field (GAFF)[10]

has been used to assign parameters for the ligands, employing the Antechamber software.[11]

These simplifications provided by using an effective potential built from the interactions between the atoms of the system allow to compute the molecular geometry of large systems at low cost and high efficiency.[4,12] Although the main disadvantage is the impossibility of accounting for electronic rearrangement since electrons are not considered within the force field equations, therefore the situations where atomic bonds are formed or broken cannot be studied with this type of potential.[13] At this point, it is important to point out the existence of recent more expensive polarizable force fields that allow the charge distribution to vary with the demands of the dielectric environment,[14] looking forward to accounting for the appropriate variations in charge distribution with the dielectric environment. One example is the general purpose polarizable force field AMOEBA developed by Ponder and co-workers.[15–19] These "next-generation" of force fields have focused on introducing an explicit electronic polarization model,[20] although the computational cost of these types of calculations is still too expensive for the QM/MM methods applied in this work.

3.1.2. QM potentials

QM potentials are built on the fundamental postulate of quantum mechanics that states that for any chemical system the application of an operator upon the wave function, Ψ, returns the observable properties of the system. The Schrödinger equation (see Eq. 3.14) returns the energy of the system when the Hamiltonian operator is applied to the wavefunction. Since quantum chemists are usually faced with many-electron problems, the electronic structure techniques or QM methods look for a solution of the Schrödinger equation at some level of approximation.

$$\widehat{H} \cdot \Psi = E \cdot \Psi \qquad \text{Eq. 3.14}$$

The typical form of the Hamiltonian operator useful in our studies to describe a chemical system of M nuclei and N electrons consists of the kinetic energy of the electrons and the nuclei, the attraction of the electrons

to the nuclei, and the interelectronic and internuclear repulsions, as expressed in the following equation

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{\mathrm{M}_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \qquad \text{Eq. 3.15}$$

where i and j states for the N electrons in the system, A and B for the M nuclei, and ∇_q^2 is the Laplacian operator, defined as the sum of differential operators in cartesian coordinates.

The first and most widely applied approximation is the Born-Oppenheimer approximation,[1] which considers that the electronic motion and the nuclear motion in molecules can be uncoupled as if the electrons in a molecule were moving in the field of fixed nuclei. Therefore, the nuclear terms from Eq. 3.15 can be removed and the expression simplified to the electronic Hamiltonian containing the kinetic energy of the electrons, the attraction of the electrons to the nuclei, and the interelectronic repulsions.

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \qquad \text{Eq. 3.16}$$

Therefore, the wavefunctions used will be describing only the electronic terms and in consequence, the total energy will be the sum of the obtained electronic energy and the nuclear repulsion term (final term in Eq. 3.15); and the Hamiltonian operators used will be referring to the electronic Hamiltonian as shown in Eq. 3.16.

QM methods are mainly based on two different principles, the study of the system through the wavefunction, represented by the Hartree-Fock (HF) approximation and are central to quantum chemistry, leading to the many wavefunction-based QM methods; and the study of the system through the electronic density, centered in the density functional theory (DFT), which was developed to avoid the calculation of the wavefunction.[21]

The origin of HF methods is in the Hartree product, which consists of a many-electron wavefunction built as a simple product of one-electron wavefunctions. Since the Hartree product does not satisfy the antisymmetry

principle, Slater determinants were introduced to improve the approximation by introducing exchange-correlation between parallel spins. Nevertheless, the motion of electrons with opposite spin is not correlated, therefore more accurate approximations are needed. The Hartree-Fock approximation is the first step towards solving the many-electron problem. This approximation is solved through the self-consistent field (SCF) method, where an initial guess is made for the wavefunction and the corresponding eigenvalues are obtained, the guess is improved iteratively until convergence. Additionally, to improve the quality of the obtained wavefunction, a finite basis set is introduced to expand the guess electron orbitals, although the more orbitals used in the basis set the slower is the calculation.[22] Furthermore, to deal with the electron correlation several wavefunction-based *ab initio* methods are used, named post-HF methods since they take their basis from the HF approximation. The most popular are the Møller and Plesset (MP) methods based on second-order perturbation theory, the ones based on configuration interaction (CI), quadratic (QCI), or coupled-cluster approaches (CC). These are capable to achieve high accuracy but are computationally more expensive approaches that are usually devoted to small QM systems.[23] On the other hand, to have an HF-based method with lower computational cost, semiempirical methods introduce some experimentally derived parametrizations to approximate the Hamiltonian and allow to study of larger QM systems.

Based on the electron density emerges the DFT, which is widely used since it scales less severely than HF methods and larger QM systems can be described than with post-HF methods, keeping the *ab initio* level accuracy. The starting point of DFT is the first Hohenberg-Kohn theorem, which proves that it is possible to determine the Hamiltonian operator from the electron density, and therefore all the properties of the system. By developing a functional that describes the energy of the system through the electron density (Eq. 3.17), the theorem shows that exists a unique functional $F_{HK}[\rho_0]$ that is only dependent on the electron density and therefore valid for any system.

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] = F_{HK}[\rho_0] + E_{Ne}[\rho_0] \qquad \text{Eq. 3.17}$$

where $E_{Ne}[\rho_0]$ is the attraction of the electrons to the nuclei. The electron-electron interaction contained in the $F_{HK}[\rho_0]$ functional can be divided in

the exact classical Coulomb part, $J[\rho]$, and the *non-classical* contribution to the electron-electron interaction, $E_{ncl}[\rho]$, of self-interaction correction, exchange, and correlation. This base theory would provide the exact solution of the system if the "universal" functional would be known, but further approximations to this theory are necessary to obtain an applicable functional to solve the system. These approximations are described by Kohn–Sham equations, which bring the applicability of DFT to real chemical problems.

$$F_{KS}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \qquad \text{Eq. 3.18}$$

where the KS approach computes the major fraction of the kinetic energy of the system using Slater type determinants, $T_S[\rho]$, valid only for non-interacting electrons. This energy expression does not give the real wavefunction if it is used in HF theory, but it corresponds to the exact ground-state electron density, providing a solution of the system in terms of DFT. Even though the kinetic energy computed corresponds to a non-interacting system, and exchange-correlation energy was introduced.

$$E_{XC}[\rho] = T_C[\rho] + E_{ncl}[\rho] \qquad \text{Eq. 3.19}$$

where $T_C[\rho]$ is added to the previous non-interacting system kinetic energy, $T_S[\rho]$, to obtain the true kinetic energy. This exchange-correlation term is the point where the different approximations are introduced for the different existing DFT functionals, leading to accurate results that have wide applications in computational chemistry and have been established as a standard tool in probing reaction mechanisms at the molecular level.[21]

In this thesis, two different levels of theory have been used in order to describe the QM system to obtain faster low-level energy and structure results, which are later on corrected or re-optimized at a higher level of theory. The Austin Model 1 (AM1)[24] semiempirical potential has been used as a low level of theory and the hybrid Minnesota DFT functional M06-2X[25] from the M06 suite density functionals has been used as a high level of theory description. The standard 6-31+G(d,p) basis set has been applied.

3.1.2.1. Semiempirical methods

Semiempirical methods are built based on molecular orbitals theory or HF method, where the eigenfunctions for a molecular system can be called molecular orbitals, and the wavefunction of the electrons depends only on the fixed nuclear coordinates.[26] Additionally, since there are systems which size is not accessible by *ab initio* methods, there are acceptable approximations that can be applied to allow their study. Therefore, modifications of the HF theory to simplify its formalism, usually by adopting a parametric form for some aspect of the calculation where the parameters are chosen to reproduce experimental data, are applied in the semiempirical approach.[26]

The major approximations taken are the *a priori* estimation of the two-electron integrals; approximation to zero the Coulomb integrals for pairs of electrons which basis functions are very far from each other and solving analytically the derivatives of the energy with respect to nuclear motion. All the introduced approximations provided a method able of optimizing TS geometries in a fast and robust manner for larger systems.

The AM1[24] semiempirical Hamiltonian has been used in this thesis as the low-level QM methodology. This method was built by modification of the neglect of diatomic differential overlap (NDDO) model and the modified neglect of diatomic overlap (MNDO) parameters. Since MNDO had poor performance in the prediction of hydrogen bonding geometries and energies, the nuclear repulsion term was modified as follows:

$$V_N(A,B) = V_{AB}^{MNDO} + \frac{Z_A Z_B}{r_{AB}} \sum_{i=1}^{4} [a_{A,i} e^{-b_{A,i}(r_{AB}-c_{A,i})^2} + a_{B,i} e^{-b_{B,i}(r_{AB}-c_{B,i})^2}] \quad \text{Eq. 3.20}$$

where up to four parameters (a, b, c) are introduced for each atom modifying the potential of mean force between two atoms. These modifications result in reasonably robust and fast QM calculations, representing a real improvement over MNDO, with no increase in the computer time needed.[24] AM1 reproduces better than its predecessor

hydrogen bonds and gives better estimates of reaction activation energies.[24] Therefore, since the computational cost remains low for QM calculations, it is suited for the PES exploration and QM/MM MD simulations with large biomolecular systems performed in this work.

3.1.2.2. DFT methods

As stated above, the main point of DFT functionals is to optimize the exchange-correlation term presented in Eq. 3.19. The local density approximation (LDA) is the model system in which approximate exchange-correlation functionals are based. The basic idea for this model is a hypothetical uniform electron gas, where the electrons move on a positive background charge distribution. When this model is extended to the unrestricted case the local spin-density approximation (LSDA) is obtained, where the degree of spin polarization is controlled by a spin-polarization parameter. These models represent the only system where the exchange and correlation energy functionals form is known exactly or with very high accuracy.[23] The major problem is that it works very well for simple metals, but it is not realistic for atoms or molecules. The next improvement for these functionals was to consider the gradient of the charge density in addition to the density, obtaining a way to account for the non-homogeneity of the true electron density. The functionals that include the gradients of the charge density are known as generalized gradient approximations (GGA). Finally, to arrive at a more accurate exchange-correlation energy, hybrid functionals use the exact KS exchange energy and approximate functionals only for the missing in the HF part, i.e., the electron correlation. Over the years, exchange-correlation functionals have been optimized against broad and diverse databases.

The M06-2X functional is a hybrid meta-GGA DFT functional developed by Truhlar and co-workers.[25] The exchange-correlation term introduced in this functional can be expressed as follows

$$E_{XC}^{hib} = \frac{X}{100} E_X^{HF} + \left(1 - \frac{X}{100}\right) E_X^{DFT} + E_C^{DFT} \qquad \text{Eq. 3.21}$$

where a percentage, X, of the exchange is the nonlocal HF exchange energy, E_X^{HF}, and the rest is the local DFT exchange energy, E_X^{DFT}, and the local

correlation energy comes from the DFT term, E_C^{DFT}. In the local DFT exchange energy, the exchange functional is obtained as a linear combination of the exchange energy density of the PBE[27] exchange model and the LSDA spin density approximation[28] for exchange. In the correlation energy term, the opposite-spin and parallel-spin correlations are treated differently.

This functional is recommended for main-group thermochemistry, kinetics, noncovalent interactions, and electronic excitation energies to valence and Rydberg states.[25] This makes it a good choice for the studies on protein reactivity present in this thesis since it considers a more realistic situation for the electron density than pure LSDA functionals, it includes the GGA approximation and it has been designed for the type of atoms and molecules that intervene in the reactivity steps studied in this work.

3.1.3. QM/MM potentials

Hybrid or multiscale QM/MM potentials were introduced by Karplus, Warshel, and Levitt in the seventies.[29,30] The development of their concept was later awarded by the Nobel Prize in Chemistry in 2013.[31] The QM/MM concept is based on combining the two types of potentials explained above to describe different parts of the same system, therefore giving a model that could have the advantages of both types of potentials. Within this approach, it is possible to model big systems such as macromolecules and to study their reactivity involving electron rearrangement. To do so, different layers are established in the model. The inner layer is devoted to describing the system at the highest possible level with QM potentials, and it includes only those atoms directly involved in the chemical reaction, i.e., electron rearrangement due to the formation or breaking of covalent bonds. The rest of the system forms the outer layer, and it is described with MM potentials, so a high number of atoms can be included in this part, allowing to study the dynamics of the system and how the flexibility of the system affects the QM layer. Therefore, a crucial issue in this type of potentials is how the boundaries between QM and MM layers are treated. This interface between both levels with different theoretical descriptions is a central aspect of this type of potentials. QM/MM modeling has been proved to be successful in the modeling of enzymes and solid surface studies, being able

to reach high levels of accuracy for obtaining results in enzymatic and other types of catalysis studies.[21]

There are different approaches to consider in the QM/MM schemes. First of all, the term boundary region is used to describe where the QM and MM areas encounter. Although there are QM/MM schemes that allow the boundary to change during the simulations as can be the adaptative partitioning, "hot spot" methods,[32] in this thesis the QM-MM partitioning is fixed and defined at the starting point. For fixed partitioning, there are two main types of schemes, subtractive and additive QM/MM schemes. The subtractive QM/MM scheme is based on summing the QM contribution of the inner system, the MM contribution of the full system, and subtracting the MM contribution of the QM system to avoid replication of its contribution to the energy. The ONIOM approach[33] is one of the most representative methods employing this scheme. Since it avoids the description of the QM-MM coupling the main advantage of this scheme is its simplicity, but it can neglect important contributions, as the Coulomb interactions between both layers, and requires a good parametrization method for the QM atoms. The standard additive QM/MM scheme is the one employed in this work, therefore it will be explained in more detail.

The basic Hamiltonian for the standard additive QM/MM scheme can be expressed as follows

$$\hat{H}_{QM/MM} = \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} + \hat{H}_{\mathrm{QM-MM}} \qquad \text{Eq. 3.22}$$

where the Hamiltonian corresponding to the QM and MM part, and the contribution of the boundary region or coupling term, $\hat{H}_{\mathrm{QM-MM}}$, are included. This final term is the key aspect of this type of scheme and contains all the information on how both layers influence each other.

The coupling term contains the bonded, electrostatic, and van der Waals interactions between QM and MM atoms, as shown in Eq. 3.23. The bonded and van der Waals are treated with the force field equations as described in section 3.1.1.

$$\hat{H}_{\mathrm{QM-MM}} = \hat{H}^{\mathrm{bond}}_{\mathrm{QM-MM}} + \hat{H}^{\mathrm{elec}}_{\mathrm{QM-MM}} + \hat{H}^{\mathrm{vdW}}_{\mathrm{QM-MM}} \qquad \text{Eq. 3.23}$$

The electrostatic coupling term involves the interaction between charge density from the QM level and the charge model used in the MM level. This term is usually the most relevant and contains more technical aspects. For this term different types of electrostatic embedding can be used, mechanical embedding, where the charges are treated at the MM level; electrostatic embedding, where the MM part point charges are included in the QM calculation and therefore polarize the QM part; and polarized embedding that depending on the level of polarization both parts influence each other. The most common type of embedding is electrostatic embedding[34] and is the one used in this thesis. Here, the MM part point charges are introduced as one-electron terms in the QM Hamiltonian, as shown in Eq. 3.24.

$$\hat{H}^{\text{elec}}_{\text{QM-MM}} = -\sum_{i=1}^{N}\sum_{J=1}^{L}\frac{q_J}{r_{iJ}} + \sum_{A=1}^{M}\sum_{J}^{L}\frac{q_J Q_A}{R_{AJ}} \qquad \text{Eq. 3.24}$$

where the MM point charges, q_J , are located at r_{iJ} distance from the QM electrons and the nuclear charges of the QM atoms, Q_A, are located at R_{AJ} distance from the MM point charges, in a system with L point charges, N electrons, and M QM nuclei. This type of embedding can be used independently of the type of QM Hamiltonian employed in the QM part, SE or DFT since it is described as an interaction of point charges with charge density.

As stated above, the bonding and van der Waals terms are calculated as given by the force field equations, as shown in Eq. 3.25 and Eq. 3.26.

$$\hat{H}^{\text{bonding}}_{\text{QM-MM}} = \sum_{bonds} k_d(d-d_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+\cos(n\phi+\delta)] \qquad \text{Eq. 3.25}$$

The bonding term involves the contribution of the bonds, angles, and torsions, dihedrals or improper torsions, interactions between the QM and MM parts of the model. This term is necessary when there are covalent bonds between the QM and MM layers. This type of truncation of the layers is usually avoided, if possible, but when it is necessary to break the QM

layer through a covalent bond it is strategically placed as far as possible from the reactivity involved atoms, by at least three bonds away, and usually involving non-polar aliphatic C-C bonds whose chemical character will not change along the reaction process. The description of these QM-MM bonded interactions can later be included in the MM bonding term for the energy.

$$\hat{H}^{\mathrm{vdW}}_{\mathrm{QM-MM}} = \sum_{\substack{non-bonded \\ pairs\ AB}} \varepsilon_{AB} \left[\left(\frac{r_{eq,AB}}{r_{AB}} \right)^{12} - 2 \left(\frac{r_{eq,AB}}{r_{AB}} \right)^{6} \right] \qquad \text{Eq. 3.26}$$

The van der Waals term involves the contribution of pair interactions consisting only of one atom from the QM part and one atom from the MM part. The interactions are treated with a Lennard-Jones potential as done for purely MM calculations.

In order to treat the covalent bonds truncated by the QM-MM boundary, the link atom approximation was used where an additional atom, the hydrogen atom, in this case, is introduced covalently bound where the QM part is cut to saturate the free valence.[30,35] As many link atoms as covalent bonds cut by the QM-MM boundaries are introduced. The link atoms appear only in the description of the QM part and are not involved in geometry-optimization or molecular-dynamics algorithms. Therefore, their position is not propagated according to forces acting on them. Although this strategy may cause overpolarization on the QM density since the introduced atom is spatially very close to the MM region. Link atoms are widely used and overpolarized QM density can be minimized by using small atom-centered basis sets.

Considering the expanse of the non-bonding interactions within the system, the calculation of the non-bonding contribution may get computationally heavy. To assure that the calculation is accurate without including extensive calculations for long-range non-bonding interactions the use of cutoffs for the non-bonding term is included. In this case, a *switch function* is used to gradually reduce the contribution of the non-bonding term. This function includes the definition of an internal cutoff of 14.5 Å and an external cutoff of 16.0 Å. Therefore, the non-bonding calculations between pairs of atoms are smoothly diminished so abrupt disruptions are avoided.

Finally, the expected value for the potential energy within this scheme can be expressed as shown in Eq. 3.27.

$$E_{QM/MM} = E_{QM}^{vac} + E_{MM} + E_{QM-MM}^{elect} + E_{\mathrm{QM-MM}}^{\mathrm{bond}} + E_{QM-MM}^{vdW} \qquad \text{Eq. 3.27}$$

From Eq. 3.27, by applying the Hamiltonian definition to the wavefunction the resulting energy can be expressed as shown in Eq. 3.28.

$$E = \frac{\langle \Psi | \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM-MM}}^{\mathrm{elec}} | \Psi \rangle}{\langle \Psi | \Psi \rangle} + E_{\mathrm{QM-MM}}^{\mathrm{bond}} + E_{\mathrm{QM-MM}}^{\mathrm{vdW}} + E_{\mathrm{MM}} \qquad \text{Eq. 3.28}$$

Taking into account that the MM part is described by a non-polarizable potential, the interaction contribution to the energy by each residue of the protein is given by Eq. 3.29.

$$E_{QM-MM}^{Int} = \sum \left\langle \Psi \left| \frac{q_{MM}}{r_{e,MM}} \right| \Psi \right\rangle + \sum \sum \frac{Z_{QM} q_{MM}}{r_{QM,MM}} + E_{QM-MM}^{vdW} \qquad \text{Eq. 3.29}$$

This interaction energy can be decomposed in a sum over residues, therefore the effect of the electrostatic potential of each residue in the protein over the QM part can be evaluated.

3.2. Statistical simulations: reactivity studies

This section is devoted to the computational methods and algorithms employed for the exploration of the reaction mechanisms of the different inhibitors of the 20S proteasome studied herein. The different techniques are exposed following the necessary steps for finally obtaining the free energy path at the high level of theory. Since all studied inhibitors are covalent inhibitors and their reactivity involves electronic rearrangement, all employed methods in this section use QM/MM potentials.

3.2.1. Exploration of the potential energy surface (PES)

The exploration of the potential energy landscape is essential in the employed methodology since it produces the starting points for the following molecular dynamics (MD) simulation step. The reaction mechanism is explored at the potential energy level for two main aspects, for generating the path on the potential energy surfaces (PESs) and for the localization of the transition states and optimization of the derived minima from them through intrinsic reaction coordinate (IRC) path exploration. These minima are essential for confirming the explored mechanism at the high level of theory, as will be explained below in section 3.2.1.2.

3.2.1.1. Generation of the PES

Based on the Born-Oppenheimer approximation, PES are representations of the potential energy of the particles (electron and nuclei) of the chemical system, where the energy landscape is presented depending on the positions of the atoms. The number of coordinates is related to the number of atoms, i.e., in a molecule with N atoms, the system would have 3N cartesian coordinates or 3N – 6 internal coordinates. To have a manageable number of coordinates for visualization, PES is expressed in terms of distinguished reaction coordinates, ξ_i, which are defined depending on the key varying distances of the explored chemical step. Hence, the importance of choosing the right distinguished reaction coordinates is key to properly explore the reaction mechanism. Once the visualization coordinates are selected, one or two reaction coordinates and the energy of the system in each point, the PES can be generated. There are two key properties to be considered to explore the PES: the gradient and the Hessian on singular points.

The gradient is the vector composed by the first derivative of the potential energy with respect to the nuclear coordinates, $\vec{r_i}$.

$$\overrightarrow{g_i} = \frac{\partial U}{\partial \vec{r_i}} \qquad \forall\, i = 1, \ldots, N \qquad \text{Eq. 3.30}$$

The Hessian is the symmetric matrix containing the second derivatives of the potential energy with respect to the nuclear positions of the atoms in the system. This can be expressed as shown in Eq. 3.31. To obtain the eigenvalues the second derivative matrix is diagonalized.

$$H_{ij} = \frac{\partial^2 U}{\partial \vec{r_i} \partial \vec{r_j}} \qquad \text{Eq. 3.31}$$

Finally, the PES projected in the selected distinguished reaction coordinates, ξ_i can be obtained by constraining the selected coordinate with a harmonic potential while the rest of the system is optimized, for each step of the mechanism. Some of the optimization algorithms employed to minimize the geometries of the system are the conjugate gradient algorithm[36] and the L-BFGS-B algorithm.[37] The resulting PES allows to identify the minimum energy path (MEP) and provides the key structures along the reaction mechanism used for further optimization such as local energy minima that correspond to reactant complex (R), possible intermediates (I), and product complex (P); and the saddle points of order one that define the transition states (TS). The AM1 semiempirical method was set for the QM subset of atoms for the PES calculations in this study.

3.2.1.2. Localization of TS structures

In order to fully characterize the key structures present along the MEP, a micro-macro iteration optimization scheme[38,39] was employed, which makes use of Baker's algorithm.[40] This is an efficient quasi-Newton-like algorithm designed for TS localization by invoking the Hessian mode following algorithm to refine the initial guess until convergency. Therefore, the Hessian matrix containing a sub-set of atom coordinates (usually, but not necessarily the QM atoms) is employed, and TS structures are characterized as the ones containing one imaginary frequency, that corresponds to the

vibrational motion of the chemical reaction step. The gradient of energy for the remaining movable atoms is usually maintained at less than 0.1 kcal·mol^{-1}.

Once the TS structure is localized, the intrinsic reaction coordinate[41] (IRC) path can be traced as the steepest-descent path down towards the corresponding local minima in mass-weighted Cartesian coordinates. This calculation runs over a two steps optimization algorithm that starts from the TS geometry and moves the system toward the two possible downhill directions of this saddle point. Then the next guess structure is optimized and iteratively shifted along the same direction until the final point is reached, where the local minimum is found.

The characterization of TS structures along the computed mechanisms was done at the SE level of theory in order to obtain a good starting structure at the computed PES. Then, this structure was optimized and characterized at the high level of theory, employing the M06-2X DFT functional, and IRC paths were traced to reach the two corresponding minima for each step. Therefore, confirmation of the explored reaction step was obtained. Additionally, since the explored mechanisms involve multiple steps, common intermediate structures from different TS were compared to ensure the connection between all explored steps.

3.2.2. Exploration of the free energy surface (FES)

The generation and exploration of FESs is the key step that allows comparing the computational results with experimental magnitudes. As commented in section 1.2.3, through TST it is possible to obtain the computational kinetic rate constants for the rate-limiting step in a reaction, and this can be directly compared with experimental kinetic measurements.

3.2.2.1. Umbrella sampling (US) method

In this thesis, all reaction free energy profiles have been computed in terms of the potential of mean force (PMF). The PMF is the free energy surface generated along a chosen coordinate that is constructed by averaging the force over all the possible configurations of the particles of the system acting on the fixed chosen coordinate. The calculation of the PMF requires a series of molecular dynamics simulations using an enhanced sampling

technique along the distinguished reaction coordinate variable, ξ. The umbrella sampling (US) approach[42–45] is employed to ensure that all regions along the reaction coordinate are sampled with similar probability. A biasing potential is used, and a set of US-MD simulations are done for equidistant values of ξ, covering all the range of this coordinate. Then the weighted mass histogram analysis method (WHAM)[46] is employed to obtain the normalized population along the coordinate.

The values of the variables sampled during the simulations are then gathered to construct a distribution function from which the PMF is obtained as a function of the distinguished reaction coordinate, $W(\xi)$. The PMF is related to the normalized probability of finding the system at a particular value of the chosen coordinate, as shown in Eq. 3.32.

$$W(\xi) = C - kTln \int \rho(r^N)\delta(\xi(r^N) - \xi)dr^{N-1} \qquad \text{Eq. 3.32}$$

Then, the activation free energy can be expressed as follows,

$$\Delta G^{\ddagger}(\xi) = W(\xi^{\ddagger}) - \left[W(\xi^R) + G_{\xi}(\xi^R)\right] \qquad \text{Eq. 3.33}$$

where the superscripts indicate the value of the reaction coordinate at the reactants (R), and at the TS (‡), and $G_{\xi}(\xi^R)$ is the free energy associated with setting the reaction coordinate to a specific value at the reactant state. Normally this last term makes a small contribution, and the activation free energy is directly estimated from the PMF change between the maximum of the profile and the reactant's minimum,

$$\Delta G^{\ddagger}(\xi) \approx W(\xi^{\ddagger}) - W(\xi^R) = \Delta W^{\ddagger}(\xi) \qquad \text{Eq. 3.34}$$

The selection of the reaction coordinate is usually trivial when the mechanism can be driven by a single internal coordinate or a simple combination (as the antisymmetric combination of two interatomic distances). However, this is not the case for the reaction steps explored in this thesis. Instead, it was necessary to obtain a much more computationally demanding 2D-PMF using two coordinates: ξ_1 and ξ_2. The 2D-PMF is related to the probability of finding the system at distinct values of these two coordinates,

$$W(\xi) = C' - kTln \int \rho(r^N)\delta(\xi_1(r^N) - \xi_1)\delta(\xi_2(r^N) - \xi_2)dr^{N-2} \quad \text{Eq. 3.35}$$

To estimate the activation free energy from this quantity, the one-dimensional PMF changes were recovered and a maximum probability reaction path on the 2D-PMF surface was traced integrating over the perpendicular coordinate.

The employed force constants in this thesis were set to 2500 $kJ \cdot mol^{-1} \cdot Å^{-2}$, when atom-atom bond distances were controlled, and to 10 $kJ \cdot mol^{-1} \cdot Å^{-2}$ when a dihedral angle between four atoms is used as a distinguished reaction coordinate. The window spacing for bond distances was set to 0.1 Å when involving heavy atom connections, i.e., not involving hydrogen atoms, and to 0.05 Å when hydrogen atoms were participating in the reaction coordinate. The window spacing for the dihedral angle in the reaction coordinate was set to 0.5º. The AM1 semiempirical Hamiltonian was used for treating the QM subsystem during the US-MD simulations. For obtaining the final PMF, an equilibration US-MD simulation of 5 ps was carried out, followed by 20 ps of production at the given temperature using the Langevin-Verlet algorithm[47] with a time step of 1 fs. Finally, all data collected were gathered and the FES was obtained through the WHAM approach.

3.2.2.2. Spline corrections

Up to this point, and due to the computational cost of running QM/MM MD simulations in every window of the PMFs, free energy calculations are restricted to the use of a low-level semiempirical Hamiltonian to describe the QM sub-set of atoms. To improve the level of theory of the results and to reduce the errors associated with the quantum level of theory employed during free energy simulations, high-level corrections are employed using DFT. The application of this type of corrections was initially proposed by Truhlar and co-workers for reactions in solution[48] and later implemented for biological systems.[49,50] This approach is based on a spline correction term interpolated to any value along the FES reaction coordinates. Through this correction term, a new energy function is defined as presented in Eq. 3.36.

$$E(\xi_1, \xi_2) = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\xi_1, \xi_2)] \quad \text{Eq. 3.36}$$

where S stands for the two-dimensional spline function, which is obtained through a correction term, ΔE_{LL}^{HL}, evaluated from the single-point energy difference between the high-level (HL) and the low-level (LL) calculation of the QM part. As stated above, the AM1 semiempirical Hamiltonian was used as the LL method, and the hybrid DFT functional, M06-2X, was employed as HL energy calculation. The standard 6-31+G(d,p) basis set was used.

3.2.3. Kinetic isotope effects (KIEs)

Kinetic isotope effects (KIE) can be computed for isotopic substitutions of key atoms, using the rate-limiting TSs and the reactant state (RS) localized at DFT and SE levels of theory. The definition of the free energy of a state can be expressed as shown in Eq. 3.37.

$$G_i = U_i - RTln\, Q_i + ZPE_i \quad \text{Eq. 3.37}$$

where U_i is the internal energy, Q_i is the total partition function and ZPE_i the zero point vibrational energy. Then, from this free energy definition by applying the TST,[51–53] the KIE can be expressed as the ratio between the rate constants corresponding to the light atom, L, and the heavier isotope, H.[39,54]

$$KIE = \frac{\left(\frac{Q_{TS}}{Q_{RS}}\right)_L}{\left(\frac{Q_{TS}}{Q_{RS}}\right)_H} e^{-\frac{1}{RT}(\Delta ZPE_L - \Delta ZPE_H)} \quad \text{Eq. 3.38}$$

The total partition function, Q, is computed as the product of the translational, rotational, and vibrational partition functions for the isotopologs in RS and TS in the protein active site. The Born−Oppenheimer, rigid-rotor, and harmonic oscillator approximations are considered to independently compute the different contributions, without the scaling of vibrational frequencies.[54,55] Since both involved states, RS and TS, are in a condensed media, the contribution of translation

and rotation to KIEs are negligible.[56] Thus, it is possible to assume that the 3N − 6 vibrational degrees of freedom are separable from the 6 translational and rotational degrees of freedom of the substrate.

KIEs were computed in this thesis at a high level, using the mentioned DFT functional M06-2X, with the standard basis set, 6-31+(d,p), while averaged KIEs were computed at low-level using the semiempirical Hamiltonian, AM1. Averages were obtained over different structures of RS and rate-limiting TSs. Structures for averaged values were obtained from snapshots generated in US-MD simulation that were further characterized at the SE level.

3.3. Statistical simulations: binding studies

To evaluate the interaction energy between the inhibitors and the protein different binding methods can be employed. The methods here explained are going to be introduced ranging from high-level to low-level of theory, considering the treatment of the system and the complexity of the calculations. Initially, the alchemical free energy perturbation method (FEP) is introduced, where the binding free energy of the inhibitor is obtained through computing an unnatural alchemical transformation in water and the protein, which later are related through a thermodynamic cycle. This method has been computed at a high level of theory treating the system with hybrid QM/MM potentials. The following method explained is the enhanced sampling steered molecular dynamics (SMD) approximation, where the unbinding process is simulated through a series of classical MD simulations that contain a guiding potential force that favors the unbinding of the inhibitor. Finally, the molecular docking method is introduced, it consists of the computation of the binding energy through a highly parametrized algorithm dependent on the docking program. Docking studies have two main goals, to obtain a geometry for the Michaelis complex and to evaluate the energy of this geometry.

3.3.1. Free energy perturbation method (FEP)

The alchemical free energy perturbation method (FEP)[57] can be used to express the binding free energy as a function of interaction energies between the receptor, inhibitor, and solvent.[58] Within this approach the inhibitor is treated as the QM subset of atoms and the rest of the system is described as the MM subset. The binding free energy is obtained by computing a series of non-physical overlapping states that are later related through a thermodynamic cycle.

To compute the binding energies, the free energy difference is obtained from FEP calculations of the inhibitor Michaelis complex with the protein and the inhibitor in water solution, as previously done in our laboratory.[57] For each system, a series of QM/MM MD simulations are performed introducing two parameters, λ, and γ, in the electrostatic and van der Waals interaction terms respectively, as shown in Eq. 3.39.

$$E(\lambda) = \langle \Psi | \hat{H}_0 | \Psi \rangle + \lambda \left(\sum_{J=1}^{L} \left\langle \Psi \left| \frac{q_J}{r_{iJ}} \right| \Psi \right\rangle + \sum_{A=1}^{M} \sum_{J}^{L} \frac{q_J Q_A}{R_{AJ}} \right) + \gamma E_{QM-MM}^{\text{vdW}} + E_{\text{QM-MM}}^{\text{bond}} + E_{\text{MM}} \quad \text{Eq. 3.39}$$

where J runs over the MM point charges, q_J , located at r_{iJ} distance from the QM electrons; and A runs over the QM nuclei, Q_A, located at R_{AJ} distance from the MM point charges. Then, changes are progressively introduced in λ and γ parameters from 1.0 to 0.0, to annihilate the contributions of the electrostatic and van der Waals terms respectively, by performing a set of QM/MM MD simulations for each type of interaction. The total free energy change is computed as the sum of all the contributions between the steps covering the transformation from the initial to the final states. Finally, the binding free energy can be obtained through a thermodynamic cycle, from the difference in free energy between the inhibitor in the water solvent, ΔG_{W}, and the inhibitor in the protein, ΔG_{E}.

$$\Delta\Delta G_{bind} = \Delta G_{\text{W}} - \Delta G_{\text{E}} \quad \text{Eq. 3.40}$$

In these studies, the level of theory of QM subset of atoms employed for FEP calculation was the semiempirical Hamiltonian, AM1. For the calculation of the free energy in the aqueous solution and the protein, a total of 100 windows were done to evaluate the electrostatic and van der Walls interaction terms respectively. The window spacing for λ and γ was set to 0.01. In each window, an equilibration step of 5 ps was done, followed by 100 ps of AM1/MM MD simulation at 310 K.

3.3.2. Steered molecular dynamic simulations (SMD)

The steered molecular dynamics approximation (SMD)[59–61] can be used to simulate the inhibitor unbinding by applying a guiding potential with constant force and velocity, along a reaction coordinate, in this case, the unbinding pathway.[62] With this technique, the work necessary to pull the inhibitor out of the active site can be obtained from the value of the exerted force applied to the system, as shown in Eq. 3.41.

$$F_{ex} = K(x_0 + v \cdot t - x) \quad \text{Eq. 3.41}$$

$$W_{[x(t)]} = \int_0^{x(t)} F_{ex}(t)dx(t) \quad \text{Eq. 3.42}$$

This force corresponds to the ligand being pulled by a harmonic spring of stiffness K with its end moving with velocity *v*. The resulting work applied in the system along the reaction coordinate, x, can be calculated by numerical integration using Eq. 3.42. Finally, if the Jarzynski equality[63] is applied, as shown in Eq. 3.43. Therefore, an approximation to the averaged PMF can be obtained as shown in Eq. 3.44.

$$e^{-\frac{\Delta G}{k_B T}} = \langle e^{-\frac{W}{k_B T}} \rangle \quad \text{Eq. 3.43}$$

$$\Delta G = \langle W \rangle - \frac{\sigma_W^2}{2k_B T} \quad \text{Eq. 3.44}$$

where $\sigma_W^2 = \langle W^2 \rangle - \langle W \rangle^2$, k_B is the Boltzmann constant and T is the simulation temperature.

The force constant (K) and the constant pulling velocity (*v*) are both selected through a calibration process, where the goal is to obtain the minimum value of these parameters that allow conserving the stiff-spring approximation. This is done by controlling that the expected value of the reaction coordinate corresponds to the real value along the simulation, therefore it can be assumed that the explored geometries correspond to those existing in the unbinding pathway and no other interactions are affecting the obtained result. In this work, this calibration process was designed as a two-step process. First, the K constant was selected, by performing SMD simulations at a given *v*, obtained from similar studies in the literature.[62,64] Then, once the K constant was selected, the *v* was optimized by performing SMD simulations at the selected K and different *v*. The parameters set for the computed SMD simulations in this thesis were a force constant of 15 kcal·mol^{-1}·A^{-2} and a pulling velocity of 0.0005 Å·ps.

SMD simulations were performed at the MM level of theory using the AMBER and the TIP3P force fields for treating the protein and the solvent molecules, respectively.

Since the final values obtained with this methodology are averaged values, it is necessary to set several replicas enough for considering the final averaged value representative of the simulated process. This can be done by computing the standard error of the mean (SEM) dependent on the number of replicas, considering enough replicas when the variation of the SEM tends to zero. In this work, the number of replicas for the selected values of K and *v* was set to 15.

3.3.3. Molecular docking

Molecular docking calculations are low-level classical mechanics-based calculations broadly used to assist drug discovery programs.[65] This type of calculation can be divided into two parts, the ligand docking, and the docking pose scoring. Usually, they produce good results for small molecules.[66] The docking calculation is considered as the positioning of the ligand in the rigid 3D structure of a protein binding site, and the scoring is the evaluation of the quality of docked ligands.[67] Since the protein is kept rigid these types of calculations neglect the possible conformational changes in solution or induced by the ligand binding.[68,69] Then the scoring function, usually based on simplified MM force fields, is used to calculate the binding energy, taking into account non-bonding interactions between the ligand and the protein. Scoring functions usually consist of a sum of terms corresponding to the different interaction types, i.e., hydrogen bond, electrostatic, or van der Waals terms; and other terms that can be relevant in the docking pose evaluation, such as ligand torsions, interatomic contacts, or desolvation.

The software employed for performing docking calculations used in this thesis is Glide.[70,71] This software approximates a systematic search of the conformation, orientation, and position of the docked ligand via Monte Carlo sampling. The starting point for the docking calculation goes through a torsionally flexible energy optimization using the OPLS-AA[72] nonbonded potential grid for the candidate docking poses. Then the scoring function evaluates the quality of the different resulting poses and can be employed

to evaluate the binding free energy through the different classical contributions, as shown in Eq. 3.45.

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{lr}) + C_{hbond} \sum g(\Delta r) h(\Delta \alpha) + C_{metal} \sum f(r_{lm}) + C_{rotb} H_{rotb} \qquad \text{Eq. 3.45}$$

where f, g, and h are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00-0.00) for distances or angles that lie outside those limits but inside larger threshold values.[70]

3.4. Model preparation and software

The X-ray structure of the *human* 20S proteasome in a complex with dihydroeponemycin (PDB ID: 5LF1)[73] was employed as the initial protein structure for the inhibition studies presented in this thesis. The setup procedure was generally the same in all cases, with just some differences in the preparation of the inhibitor starting structure. For the protein preparation, the pKa of titratable residues was predicted with the PropKa program ver. 3.5.[74,75] Then, missing hydrogen atoms were added using the tLEAP[76] module of the AmberTools program at pH equal to 7. The same module was used to add Na^+ counterions to neutralize the overall negative charge of the system. Finally, the full model was solvated in an orthorhombic box of TIP3P[77] water molecules with an approximated size of 17×17×20 nm^3. Missing force field parameters for the inhibitors, either free or covalently bound to the protein via reactive Thr1 were generated with the Antechamber software,[78] using GAFF force field[10] and charges computed with AM1-bcc method employing semiempirical quantum chemistry program (sqm).[79,80]

To obtain an equilibrated and more representative initial structure than the X-ray starting coordinates, classical MD simulations were performed with NAMD software.[81] The AMBER[8] force field was used to treat the inhibitor and protein while the solvent molecules were described with the TIP3P[77] force field. In these MD simulations, a preliminary equilibration of 100 ps was done, followed by progressive heating of the system to 310 K with a 0.001 K temperature increment. Then 100 ps of NPT equilibration were done keeping the full system free to move, followed by 100 ps of NVT equilibration with all residues beyond 40 Å from the inhibitor frozen. The length of the final MD production stage varies depending on the different inhibitors studied and if they were simulated in the Michaelis complex or the covalently bound state. All classical MD simulations were done employing a time step of 1 fs and periodic boundary conditions. The temperature was controlled using the Langevin thermostat[82] and the Nosé-Hoover Langevin piston[83] was employed for NPT control. In order to improve the time of simulations, cut-offs for nonbonding interactions were applied using a smooth switching function between 14.5 and 16 Å.

Analysis of all data obtained from MD simulations was performed by using the cpptraj program[84] for extracting key distances and root-mean-square-deviation (RMSD) of the protein backbone.

For reducing the cost of QM/MM calculations, the solvation box of water molecules was reduced to a 60 Å sphere centered on Thr1 residue of the β5-subunit., while the rest of the model was kept. The protein and solvent water molecules were represented by classical AMBER and TIP3P force fields, respectively, as implemented in the fDynamo library.[85,86]

All QM/MM calculations were carried out with the methods described before as implemented in the fDynamo library. During these calculations, all residues present beyond 20 Å from the inhibitor were frozen and cutoffs were applied for the non-bonding interactions, as explained in section 3.1.1. When performing high-level QM/MM calculations the Gaussia09 program, in combination with fDynamo, was employed.[87]

The VMD[88] and PyMOL[89] programs were employed for structure and trajectory visualization and image rendering.

SMD simulations were performed using the Gromacs[90] package as MD engine and the PLUMED library, version 2.6.4.[91]

Bibliography

(1) Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.

(2) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(3) Cramer, C. J. *Essentials of Computational Chemistry*, 2nd ed.; Wiley: Chichester West Sussex England; Hoboken NJ, 2004.

(4) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chemie - Int. Ed.* **2009**, *48*, 1198–1229.

(5) Weiner, S. J.; Kollman, P. A.; Singh, U. C.; Case, D. A.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(6) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230–252.

(7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(8) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(10) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(11) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(12) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations.

J. Comput. Chem. **1990**, *11*, 700–733.

(13) Tchougréeff, A. L. *Hybrid Methods of Molecular Modeling*; Springer: Dordrecht, 2008; Vol. 17.

(14) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Current Opinion in Structural Biology*. 2001, pp 236–242.

(15) Ren, P.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

(16) Qi, R.; Wang, L. P.; Wang, Q.; Pande, V. S.; Ren, P. United Polarizable Multipole Water Model for Molecular Mechanics Simulation. *J. Chem. Phys.* **2015**, *143*, 014504.

(17) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.

(18) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *J. Phys. Chem. B* **2004**, *108*, 13427–13437.

(19) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(20) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(21) Catlow, C. R. A.; van Santen, R. A.; Van Speybroeck, V. *Modelling and Simulation in the Science of Micro- and Meso-Porous Materials*; 2017.

(22) Gund, T. Molecular Modeling of Small Molecules. In *Guidebook on Molecular Modeling in Drug Design*; 1996; pp 55–92.

(23) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; 2001.

(24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(25) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New

Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(26) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; 2004; Vol. 42.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*.

(28) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, 1133–1138.

(29) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(30) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(31) The Royal Swedish Academy of Sciences. The Nobel Prize in Chemistry 2013 (Press Release). *Press. Press release* **2013**.

(32) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie - International Edition*. 2009.

(33) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chemical Reviews*. 2015, pp 5678–5796.

(34) Lodola, A.; De Vivo, M. The Increasing Role of QM/MM in Drug Discovery. In *Advances in Protein Chemistry and Structural Biology*; 2012; Vol. 87.

(35) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH3Cl + Cl− Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.

(36) Hestenes, M. R.; Stiefel, E. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Natl. Bur. Stand. (1934).* **1952**, *49*, 409–436.

(37) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208.

(38) Turner, A. J.; Moliner, V.; Williams, I. H. Transition-State Structural

Refinement with GRACE and CHARMM: Flexible QM/MM Modelling for Lactate Dehydrogenase. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323–1331.

(39) Martí, S.; Moliner, V.; Tuñón, I. Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy to Explore the Potential Energy Surface in Very Large Systems. *J. Chem. Theory Comput.* **2005**, *1*, 1008–1016.

(40) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385–395.

(41) Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.

(42) McDonald, I. R.; Singer, K. Machine Calculation of Thermodynamic Properties of a Simple Fluid at Supercritical Temperatures. *J. Chem. Phys.* **1967**, *47*, 4766–4772.

(43) McDonald, I. R.; Singer, K. Examination of the Adequacy of the 12-6 Potential for Liquid Argon by Means of Monte Carlo Calculations. *J. Chem. Phys.* **1969**, *50*, 2308–2315.

(44) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

(45) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(46) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(47) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(48) Chuang, Y.-Y.; Corchado, J. C.; Truhlar, D. G. Mapped Interpolation Scheme for Single-Point Energy Corrections in Reaction Rate Calculations and a Critical Evaluation of Dual-Level Reaction Path Dynamics Methods. *J. Phys. Chem. A* **1999**, *103*, 1140–1149.

(49) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes. *J. Phys. Chem. B* **2006**, *110*, 17663–17670.

(50) Ferrer, S.; Martí, S.; Moliner, V.; Tuñón, I.; Bertrán, J. Understanding the

Different Activities of Highly Promiscuous MbtI by Computational Methods. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3482–3489.

(51) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 63–71.

(52) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Trans. Faraday Soc.* **1935**, *31*, 875–894.

(53) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(54) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. QM/MM Calculations of Kinetic Isotope Effects in the Chorismate Mutase Active Site. *Org. Biomol. Chem.* **2003**, *1*, 483–487.

(55) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. Computing Kinetic Isotope Effects for Chorismate Mutase with High Accuracy. A New DFT/MM Strategy. *J. Phys. Chem. B* **2005**, *109*, 3707–3710.

(56) Ruggiero, G. D.; Guy, S. J.; Marti, S.; Moliner, V.; Williams, I. H. Vibrational Analysis of the Chorismate Rearrangement: Relaxed Force Constants, Isotope Effects and Activation Entropies Calculated for Reaction in Vacuum, Water and the Active Site of Chorismate Mutase. *J. Phys. Org. Chem.* **2004**, *17*, 592–601.

(57) Świderek, K.; Martí, S.; Moliner, V. Theoretical Studies of HIV-1 Reverse Transcriptase Inhibition. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12614–12624.

(58) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309–3319.

(59) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science.* **1996**, *271*, 997–999.

(60) Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered Molecular Dynamics Investigations of Protein Function. *J. Mol. Graph. Model.* **2001**, *19*, 13–25.

(61) Do, P. C.; Lee, E. H.; Le, L. Steered Molecular Dynamics Simulation in Rational Drug Design. *J. Chem. Inf. Model.* **2018**, *58*, 1473–1482.

(62) Capelli, A. M.; Costantino, G. Unbinding Pathways of VEGFR2 Inhibitors Revealed by Steered Molecular Dynamics. *J. Chem. Inf. Model.* **2014**, *54*, 3124–3136.

(63) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

(64) Capelli, A. M.; Bruno, A.; Entrena Guadix, A.; Costantino, G. Unbinding Pathways from the Glucocorticoid Receptor Shed Light on the Reduced Sensitivity of Glucocorticoid Ligands to a Naturally Occurring, Clinically Relevant Mutant Receptor. *J. Med. Chem.* **2013**, *56*, 7003–7014.

(65) Singh, N.; Warshel, A. Absolute Binding Free Energy Calculations: On the Accuracy of Computational Scoring of Protein-Ligand Interactions. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1705–1723.

(66) Merz, K. M. Limits of Free Energy Computation for Protein-Ligand Interactions. *J. Chem. Theory Comput.* **2010**, *6*, 1769–1776.

(67) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps Andrew. *J. Med. Chem.* **2006**, *49*, 5851–5855.

(68) Weill, N.; Therrien, E.; Campagna-Slater, V.; Moitessier, N. Methods for Docking Small Molecules to Macromolecules: A User's Perspective. 1. The Theory. *Curr. Pharm. Des.* **2014**, *20*, 3338–3359.

(69) Campagna-Slater, V.; Therrien, E.; Weill, N.; Moitessier, N. Methods for Docking Small Molecules to Macromolecules: A User's Perspective. 2. Applications. *Curr. Pharm. Des.* **2014**, *20*, 3360–3372.

(70) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

(71) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

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

(73) Schrader, J.; Henneberg, F.; Mata, R. A.; Tittmann, K.; Schneider, T. R.; Stark, H.; Bourenkov, G.; Chari, A. The Inhibition Mechanism of Human 20S Proteasomes Enables Next-Generation Inhibitor Design. *Science* **2016**, *353*, 594–598.

(74) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical

Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

(75) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(76) Schafmeister, C. E. A. .; Ross, W. S.; Romanovski, V. LEAP. University of California, San Francisco 1995.

(77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(78) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(79) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(80) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. Amber 2021; University of California: San Francisco, 2021.

(81) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(82) Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A* **1986**, *33*, 3628–3631.

(83) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(84) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for

Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(85) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The Dynamo Library for Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.* **2000**, *21*, 1088–1100.

(86) Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and Hydrophilic Binding Sites of HIV-1 RT. *J. Phys. Chem. B* **2015**, *119*, 917–927.

(87) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 09; Gaussian, Inc.: Wallingford, CT, 2009.

(88) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(89) DeLano, W. L. The PyMOL Molecular Graphics System, Version 2.3. *Schrödinger LLC*. 2020.

(90) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(91) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

Chapter 4. Results and discussion

The results of this thesis, derived from computational studies based on QM/MM methods, are presented in two parts according to the nature of the studied inhibitor. The first part corresponds to the study of the peptide-like inhibitor dihydroeponemicin, an α,β-epoxyketone that irreversibly inhibits the β5 active site of the 20S proteasome. The second part is devoted to the family of non-peptidic β-lactones-γ-lactam inhibitors, where salinosporamide A (SalA) is the main compound of the study due to its advanced position in clinical trials. The mode of action of SalA has been studied in deep, and the obtained results provided information to propose some variants that were later explored to understand how drug lead modification affects the compound activity.

4.1. Peptide like epoxyketone in 20S proteasome

It is generally accepted that α,β-epoxyketone inhibitors can react with $O^{\gamma Thr1}$ and N^{Thr1} by the nucleophilic attack to the ketone and the epoxide group of the warhead, respectively. This mode of action is verified by X-ray studies.[1,2] Nevertheless, there is an open dispute about the final product that these inhibitors form when bound to the 20S proteasome active site. Initially, a 1,4-morpholino ring (six-membered ring) was shown to be present in the epoxomicin yeast 20S proteasome X-ray structure.[1] More recently, a 1,4-oxazepane ring (seven-membered ring) was identified in the dihydroeponemicin by solving the X-ray structure of the *human* 20S proteasome. These contradictory results opened new questions about the inhibitory mechanism.

Along with the dihydroeponemicin X-ray structure presented by Schrader et al.,[2] a small QM-cluster model was used trying to clarify this unexpected final product. Although computationally obtained results suggested the seven-membered ring was kinetically favored, the deeper inspection of the reported results revealed that the provided energy barriers were computed relative to the first intermediate, which was ca. 10 kcal·mol^{-1} higher in energy than reactants. Therefore, the employed QM-cluster model provided meaningfully overestimated barriers compared to the expected values for the inhibitory process. Additionally, these calculations lack the precision necessary to explore the full inhibitory mechanism, since they do not account for the enzyme participation in the reaction and some questions remained unsolved, i.e. the role of the oxyanion hole in the binding process.

In light of the apparent limitations of the published study and the remaining open questions, exploring the inhibitory process of dihydroeponemicin by QM/MM methods and characterizing the free energy landscape seemed an urgent study to carry out in order to reveal the mode of action of this inhibitor in the active site of *human* 20S proteasome. Therefore, two mechanisms were considered for the formation of the two possible products. First, the generally accepted mechanism was explored, where the first step of the reaction consists of the nucleophilic attack of $O^{\gamma Thr1}$ to the warhead ketone electrophilic center, followed by the nucleophilic attack of N^{Thr1} to the epoxide electrophilic center. The alternative proposal considers that the first atom to attack is the N^{Thr1} concomitant with the opening of the epoxide ring, followed by ring closure through the nucleophilic attack of $O^{\gamma Thr1}$ to the ketone in the warhead. This proposal seems reasonable since the epoxide group should be more chemically reactive than the ketone. The results presented in this thesis show that the seven-membered ring is always more stable than the six-membered ring. In contrast to the mechanism proposed in the literature with a high energy barrier for the formation of both products; the second explored mechanism shows that only the seven-member ring formation is energetically favorable, with a free energy barrier of the rate-limiting step of 23.7 kcal·mol^{-1}. This computationally estimated barrier is in very good agreement with experimentally determined rate constants reported in the literature, 21-23 kcal·mol^{-1} (see U.S. Patent 7,642,369). Therefore, the obtained results in our studies provided a solid explanation to the observed experimentally seven-membered structure as a final product generated by this inhibitor in complex with the *human* 20S proteasome.

Finally, the localized RS and TS structures for the rate-limiting step were employed to compute KIEs. The predicted values show that KIEs can be applied as a tool capable to distinguish between the formation of both products and consequently between both proposed mechanisms, providing an additional magnitude to be compared with experiments.

A summary of the results is presented in Figure 4.1.

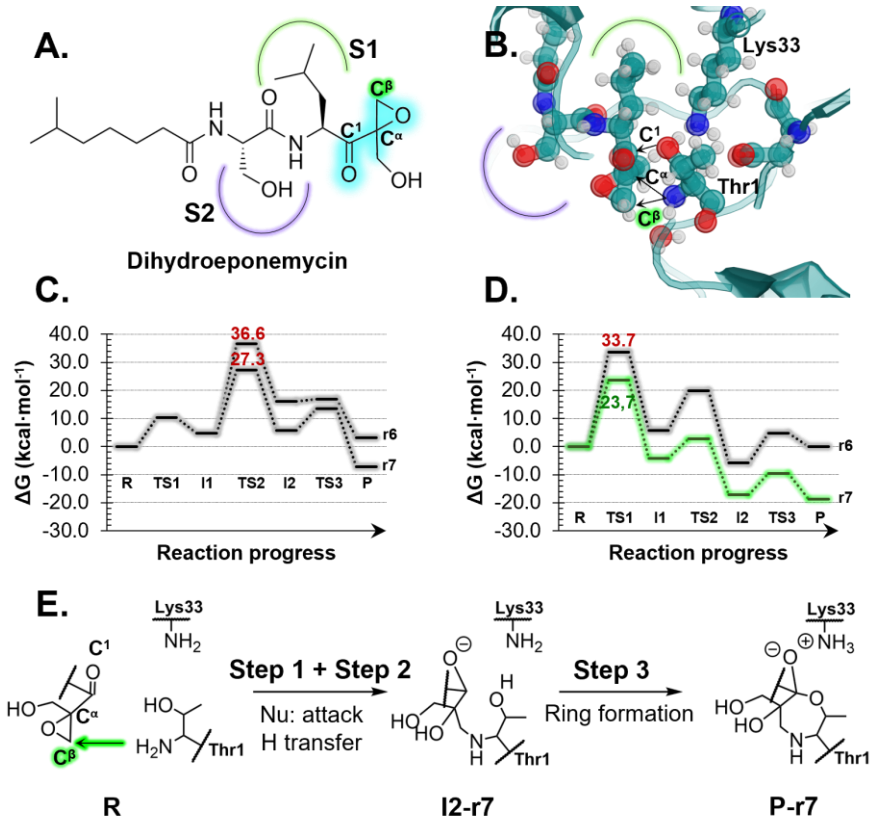


Figure 4.1. **A.** Schematic representation of dihydroeponemycin. **B.** Representation of the E•Inh complex. **C.** Free energy profiles for the proposed mechanism in literature computed at M06-2X:AM1/MM level of theory for the generation of the six- (r6) and seven- (r7) membered rings. **D.** Free energy profile for the new proposed mechanism computed at M06-2X:AM1/MM level of theory, showing in green the final reaction pathway for this inhibitor leading to the seven-membered ring product. **E.** Schematic representation of the final reaction mechanism followed by dihydroeponemycin leading to the seven-membered ring product.

These results were presented in the following paper:

Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, *164*, 399–407.
https://doi.org/10.1016/j.ejmech.2018.12.062

European Journal of Medicinal Chemistry 164 (2019) 399–407

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Research paper

Theoretical study of the inhibition mechanism of *human* 20S proteasome by dihydroeponemycin

Natalia Serrano-Aparicio, Katarzyna Świderek*, Vicent Moliner**

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain

ARTICLE INFO

Article history:
Received 16 October 2018
Received in revised form
12 December 2018
Accepted 24 December 2018
Available online 25 December 2018

Keywords:
Proteasome 20S
Dihydroeponemycin
QM/MM
Molecular dynamics
Inhibition mechanism
Free energy surfaces
Kinetic isotope effects
Epoxyketones
1,4-Oxazepane

ABSTRACT

Proteasome deregulation has been related with several human diseases and, consequently, a detailed knowledge of its inhibition is essential for the design of efficient and selective drugs. The present paper is focused on the inhibition mechanism of proteasome 20S on the β5-subunit by dihydroeponemycin, an epoxyketone. The presence of a dual electrophilic center in this α,β-epoxyketone allows its irreversible bind to the active site by formation of two strong covalent bonds with the N-terminal threonine residue. Free energy surfaces for all possible mechanisms have been generated in terms of potentials of mean force (PMFs) within hybrid QM/MM potentials, with the QM subset of atoms described at semiempirical (AM1) and DFT (M06-2X) level. Two alternative reaction pathways, differentiated by reversing the order of chemical steps in full catalytic process and the product species, were explored. The resulting activation free energy barriers ($\Delta G^{\ddagger}$) indicate that the most favourable mechanism is the one in which the reaction starts with epoxide-ring opening and finishing with 1,4-oxazepane product formation. This result is in agreement with the seven-membered product of inhibition recently determined by X-ray crystallography. Finally, calculations of primary kinetic isotope effects (1^{α}-KIEs) on Cα and Cβ of epoxide and secondary 2^{α}-KIE on C1 reveal their possible application in distinguishing between the formation of six- and seven-membered product and verifying the reaction mechanism proposed in the present work.

1. Introduction

Protein synthesis [1] and protein degradation [2] are two complementary processes, incessantly occurring in living cells. Proteasomes are large multiunit proteases that are found in the cytosol, both free and attached to the endoplasmic reticulum, and in the nucleus of eukaryotic cells [3–6]. Their ubiquitous presence and high abundance in these compartments reflect their central role in cellular protein turnover. In fact, proteasomes are responsible for the major non-lysosomal protein regulation that degrades regulatory as well as misfolded and potentially toxic proteins in eukaryotic cells [7]. The importance of proteolytic degradation inside cells and the role of ubiquitin in proteolytic pathways was acknowledged in the award of the 2004 Nobel Prize in Chemistry to Aaron Ciechanover, Avram Hershko and Irwin Rose [8,9]. The majority of intracellular protein degradation (up to 80–90%) is accomplished by the ubiquitin-proteasome pathway (UPP) [10]. After being considered a static garbage disposal unit for cellular waste, now proteasome is recognized as a multifaceted mediator of many essential cellular processes via proteolytic and nonproteolytic mechanisms [11]. The modular and dynamic composition of the proteasome and its multiple regulators allows proteasome subtypes to be adapted to a wide array of physiological roles [11]. Proteasomes are generally essential for viability of eukaryotic cells [12], and can have regulatory control over a variety of cellular processes, such as cell immunity [13], cell differentiation [14], cell signalling [15] and cell growth [16], among other physiological processes [17]. Therefore, consequences for cells can be severe in case of proteasome misfunction. Actually, proteasome deregulation has been related with cardiovascular diseases, developmental disorders, vision impairment, ageing, neurodegeneration, inflammatory diseases, renal failure and cancer [18]. Cancer cells produce proteins that promote both cell survival and proliferation, and/or inhibit mechanisms of cell death. This notion set the stage for preclinical testing of proteasome inhibitors as a mean to shift this fine equilibrium towards cell death [19].

The protein complex involved in Ubiquitin-Proteasome System (UPS) is called 26S proteasome and its individual subcomplexes are

* Corresponding author.
** Corresponding author.
E-mail addresses: swiderek@uji.es (K. Świderek), moliner@uji.es (V. Moliner).

https://doi.org/10.1016/j.ejmech.2018.12.062
0223-5234/

related with different tasks [11]. The catalytic core of this multi-subunit complex, the 20S proteasome core particle, is a new class of proteases discovered in 90's and classified as a new threonine protease catalytic type [20]. It is a cylindrical structure consisting of four heptameric rings formed by seven α or β subunits each (see Fig. 1). The two inner rings contain three proteolytically active sites, so called caspase-like (β1), tryptic-like (β2), and chymotryptic-like (β5) sites. The responsibility for recognition and binding of poly-ubiquitinated substrates belongs to 19S subunit, which binds to each end of 20S proteasome. These external subunits unfold substrates and introduce them into the 20S catalytic core, subsequently removing the attached ubiquitin molecules [11,21]. Both of these processes require the presence of ATP. Before the degradation of the protein is achieved, target substrates require labelling with ubiquitin, the small protein which is recognized by proteasome. However, as it was shown, neither ubiquitylation nor ubiquitylated protein association with proteasome will necessary lead to protein degradation [22].

Inhibitors of the 20S proteolytic core of the proteasome have been extensively studied. Three of them (bortezomib, ixazomib and carfilzomib), proposed as the result of previous drugs studies, are currently approved by FDA as a restrictive barrier for cancer cell growth in multiple myeloma or mantle-cell lymphoma [24–28]. Peptide boronates (including bortezomib and ixazomib) are the most common inhibitors that form covalent bond with threonine terminal residue in β-subunits, resulting in reversible inhibition. Other promising inhibitors are epoxyketones, which are characterized by a dual electrophilic α,β-epoxyketone. This is capable of forming two covalent bonds with threonine residue leading to irreversible binding [29], prolonging in such way the duration of proteasome inhibition.

In the present paper, a computational study of the inhibition mechanism of 20S proteasome through covalent binding of an epoxyketone — dihydroeponemycin — to the proteolytic active site of β5-subunit is presented. Although it is rather wildly accepted that in case of dual-electrophile inhibitors, one of the electrophilic centers will be attacked by oxygen Oγ of Thr1 and the second one by the N-terminal amino group of the same residue [30], the reaction mechanism has not been yet clarified. This limitation hampers the development of new efficient and specific drugs. Initially, as shown on Fig. 2, the N-terminal amino group of Thr1 was thought to react with Cα position of the epoxide, according to Baldwin's rules [31], resulting in a six-membered ring product [32]. However, recently obtained high-resolution X-ray structure of the human 20S proteasome inhibited by dihydroeponemycin revealed formation of a seven-membered ring product [30], as also illustrated in Fig. 2. Thus, the presence of 1,4-oxazepane (seven-membered) ring covalently bond to an active site threonine residue (Thr1) is in contradictory to the 1,4-morpholine ring structure proposed in previous studies. This result open new questions about reaction mechanism, kinetics and stability of both six- and seven-membered ring products.

The exploration of the mechanisms rendering the two possible inhibitor-proteasome covalently bound complexes — the six-membered ring product (P-r6) or the seven-membered ring product (P-r7) — has been carried out by generating the free energy surfaces (FESs) of the different steps with quantum mechanics combined with molecular mechanics (QM/MM) methods. Because previous computational studies were carried out based on a reduced cluster model in gas phase [30], we decided to revise the proposed mechanisms including the effects of the full protein environment. This study was partially motivated by the fact that the overall energy barriers for inhibition process predicted in previous gas phase calculations were rather high: ca. 29.3 and 25.0 kcal mol^{-1}, for six- and seven-ring product formation, respectively. These values are derived from the reported barriers, 18.8 and 14.5 kcal mol^{-1}, that were measured with respect to first intermediate, which was 10.5 kcal mol^{-1} higher in energy than reactants [30]. Additionally, some inconsistency was found between the final proposed mechanism ending in P-r7, presented in Fig. 2, and the computational results. In particular, the suggested importance of the so-called nucleophilic water (NUK, as depicted in Fig. 2) was never addressed and, in fact, no role was assigned to this water molecule. Thus, in agreement with Schrader and co-workers [30], new insights into the atomic details of the catalytic center of the human 20S proteasome has to be done in order to get clues for the design of a new class of proteasome inhibitors.

2. Computational methods

System setup. Structure of *human* proteasome complexed with dihydroeponemycin was adapted from X-Ray structure as available in Protein Data Base (PDB ID: 5LF1) [30]. Missing force field parameters for inhibitor covalently bound to Thr1 of protein were computed using Antechamber software [34]. (see Table S1 in Supporting Information) PropKa program ver. 3.5 [35,36] was used in order to predict shifts of the pKa of titratable residues and missing hydrogen atoms were added using the tLEAP [37] module of Amber

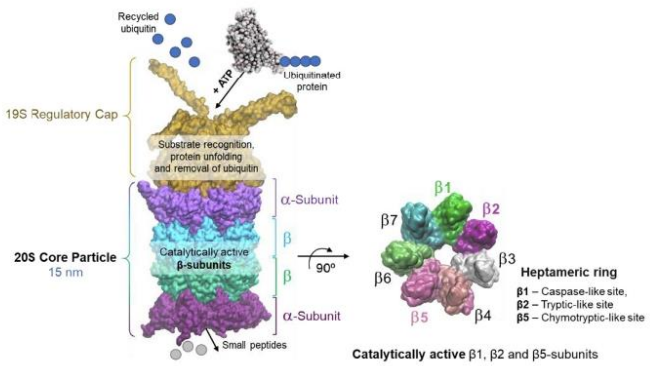


Fig. 1. The ubiquitin proteasome pathway (UPP) for protein degradation [23].

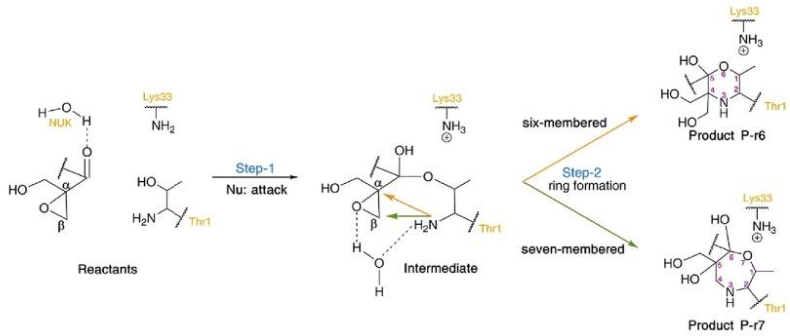


Fig. 2. Proposed inhibition mechanism of proteasome by epoxyketones rendering a six-membered ring covalent product (P-r6) [32,33]; or a seven-membered ring covalent product (P-r7) [30].

Tools program at pH 7. The obtained non-standard pKa values together with protonation states are listed in Supporting Information in Table S2. Neutralization of the system was achieved by adding 43 Na^+ counterions in the electrostatically most favourable positions. Finally, full protein together with counterions was solvated in an orthorhombic box of TIP3P [38] water with size of $17.0 \times 16.6 \times 19.4\,nm^3$ (532.081 atoms in total). Then, classical molecular dynamic (MD) simulation was performed, with time step of 1 fs, using AMBER [39] force field, with NAMD software [40]. After a preliminary equilibration of 100 ps, the system was heated to 310 K with 0.001 K temperature increment. Then, 100 ps of NPT equilibration was carried out with the full system free to move, followed by 100 ps of NVT equilibration with all residues beyond 40 Å from the inhibitor kept frozen. Finally, 5 ns of NVT MD simulation was performed keeping the same conditions than in the last NVT equilibration step. The temperature during the MD simulation was controlled using the Langevin thermostat [41]. In order to improve the time of simulations, cut-offs for nonbonding interactions were applied using a smooth switching function between 14.5 and 16 Å. Root-mean-square-deviation (RMSD) and analysis of key distances evolution along the MD simulations are provided in SI. According to the time-dependent evolution of the RMSD of those atoms belonging to the protein backbone, the system can be considered equilibrated (see Fig. S1).

QM/MM calculations. The solvation box of water molecules was reduced to a 60 Å sphere centred on Thr1 residue of β5-subunit (the total number of atoms is equal to 154348) for the QM/MM calculations. Part of the inhibitor, one water molecule and the sidechain of Thr-1 and Lys33, as presented in Fig. 3, were described by quantum mechanics, while the rest of the protein and solvent water molecules were represented by classical AMBER and TIP3P force fields, respectively, as implemented in fDynamo library [42,43]. Three link atoms [44] were inserted where the QM/MM boundary intersected covalent bonds: these were placed between the carbon-carbon bond after first peptide bond present in inhibitor, the Cα-Cβ bond of Lys33 and C-Cα bond of Thr2 (see Fig. 3). The QM region contained a total of 79 QM atoms, including the link atoms. The Austin Model 1 (AM1) [45] semiempirical Hamiltonian and the Minnesota Functional M06-2X [46], with the standard 6-31+G(d,p) basis set, were used to treat the QM sub-set of atoms. The atom positions of all residues presented beyond 20 Å from the inhibitor were frozen and the same cut-offs were applied for the non-bonding interactions.

Potential Energy Surfaces (PES). The potential energy of our scheme is derived from the standard additive scheme of QM/MM

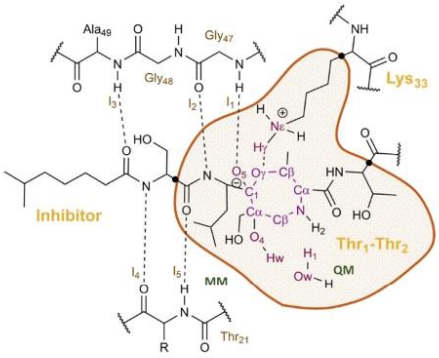


Fig. 3. Schematic representation of the active site with seven-membered ring product of the inhibition of the β5-subunit 20S proteasome core particle by Dihydroeponemycin.

formulation (see Supporting Information for details). PESs were obtained by grid scanning of appropriate distinguish reaction coordinates in each of the chemical step. A micro-macro iteration optimization algorithm [47,48] was used to localize, optimize, and characterize the TS structures using a Hessian matrix containing all the coordinates of the QM subsystem, whereas the gradient norm of the remaining movable atoms was maintained less than 0.01 kcal mol^{-1} $Å^{-1}$. Intrinsic reaction coordinates (IRCs) were traced down from located TSs to the connecting valleys in mass-weighted Cartesian coordinates.

Free Energy Surfaces (FES). In order to obtain the reaction energy profile in terms of free energies, FESs were computed in terms of Potentials of Mean Force, PMFs. All PMFs have been calculated using the weighted histogram analysis method (WHAM) [49], with a threshold density value of 10^{-2}, combined with the umbrella sampling (US) [50] approach as implemented in fDYNAMO. The number of iterations for each surface is reported in Table S3 of the Supporting Information. The procedure for the PMF calculation requires a series of hybrid QM/MM molecular dynamics simulations in which the distinguished reaction coordinate variable (or a couple of them) is constrained around particular values. (See Supporting Information for details). Because of the large number of structures that must be evaluated during free energy calculations,

QM/MM calculations are usually restricted to the use of semi-empirical Hamiltonians. In order to reduce the errors associated with the quantum level of theory employed in these simulations, a new energy function defined in terms of interpolated corrections was added into the US AM1/MM derived PMF [51–53]. See Supporting Information for details. The M06-2X density functional, with the standard 6-31+G(d,p) basis set, was selected for the high energy calculation. These calculations were carried out using the Gaussian09 program [54]. This methodology has been extensively employed in our laboratory and appears to be quite robust, according to recent benchmarking with different QM/MM methods [55] and the good agreement with experimental data, when available [56–58].

Kinetic Isotope Effects. Kinetic isotope effects (KIE) were computed for isotopic substitutions of key atoms from the rate limiting transition states (TSs) and the reactant state (RS) localized at M06-2X/6-31+G(d,p)//AMBER level of theory. Additionally, average KIEs at AM1/AMBER over 10 different structures of rate limiting TS and RS were computed and deposited in the Supporting Information, Table S4. According to the Transition State Theory (TST), the ratio between the rate constants corresponding to the light atom "L" and the heavier isotope "H" can be computed as:

$$KIE = \frac{\left(\frac{Q_{TS}}{Q_{RS}}\right)_L}{\left(\frac{Q_{TS}}{Q_{RS}}\right)_H} e^{-\frac{1}{RT}(\Delta ZPE_L - \Delta ZPE_H)} \quad (1)$$

In eq (1), the total partition function, Q, was computed as the product of the translational, rotational, and vibrational partition functions for the isotopologs in RS and TS in the active site of proteasome. The Born–Oppenheimer, rigid-rotor, and harmonic oscillator approximations were considered to independently compute the different contributions, without the scaling of vibrational frequencies, as explained and applied in previous papers [59,60]. Keeping in mind that reactants and TS are in a condensed media (the active site of a protein), contribution of translation and rotation to KIEs are negligible. Nevertheless, the full $3N \times 3N$ Hessians have been subjected to a projection procedure to eliminate translational and rotational components, which give rise to small nonzero frequencies, as previously described [61]. Thus, it has been assumed that the $3N - 6$ vibrational degrees of freedom are separable from the 6 translational and rotational degrees of freedom of the substrate.

3. Results and discussion

Inhibitor-protein interactions. As mentioned in the Introduction, one of the uncertainties in defining a reaction mechanism for the inhibition of proteasome is the protonation state of the oxyanion oxygen atom (O5 on Fig. 3) formed on dihydroeponemycin as a result of nucleophile attack of $O\gamma^{Thr1}$ to the C1. The knowledge about protonation state of atom O5 is crucial for starting the mechanistic studies at molecular level. It seems that there are enough evidences indicating that atom O5 should remain deprotonated along full reaction path. Analysis of X-ray structure of seven-membered product complexed with β5-subunit active site, together with analysis of our MD simulations reveals existence of very strong conserved interaction between oxygen atom O5 and nitrogen of the backbone of Gly47: the distance between O5 atom and N^{Gly47} was found to be 3.0 Å in crystallographic structure [30], and 2.95 ± 0.16 Å along the MD simulation of the covalently bound inhibitor-proteasome P-r7 (see Fig. 4). These results suggest that the role of H-bond donor can be assigned to N^{Gly47}, while the oxygen atom would play the role of the acceptor. This hypothesis is in agreement with previously proposed catalytic mechanisms [62,63],

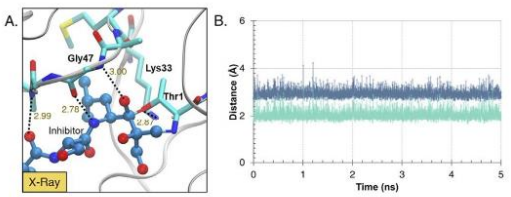


Fig. 4. A) Detail of the active site in X-Ray structure (PDB ID: 5LF1). Distances in Å. B) Time evolution of O5-N^{Gly47} (blue line) and O5-H^{Gly47} (green line) distances along the 5 ns classical MD simulations of the covalently bound inhibitor-proteasome 7-membered ring product. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

where the oxyanion was never considered to be protonated. According to these results, Gly47 would be part of the oxyanion hole stabilizing the negatively charged oxygen atom, as also observed in the active site of other proteases [64]. Consequently, atom O5 will be deprotonated in the presented work.

Mechanism of Inhibition. Mechanistic studies of β5-subunit inhibition by dihydroeponemycin began from final snapshot of the MD simulations on the structure of seven-membered ring inhibitor-proteasome complex, covalently bond through the active site Thr1, P-r7. The reaction was explored in a backward manner from P-r7 to the reactants state (RS). Two different mechanism, depending on the timing of the epoxide ring opening and the nucleophilic attack of Thr1 to the epoxiketone inhibitor, were explored for the formation of the six- and seven-membered ring inhibitor-proteasome complexes in the present study.

Mechanism A. In order to study the step from P-r7 to Intermediate 1, I1 (see Fig. 2), a PES was explored where symmetric combination of H^{wat}–O^{wat} and H^{Thr1}–N^{Thr1} distances (representing proton transfer from O4 of the epoxide to N-terminal of Thr1 through a water molecule) together with the antisymmetric combination of N^{Thr1}–Cβ and Cβ–O4 (corresponding to nucleophilic attack of nitrogen of Thr1 on Cβ of epoxide and epoxide ring breaking) were controlled. The step from I1 to the RS was explored by controlling the antisymmetric combination of distances between $N\varepsilon^{Lys33}$ and $H\gamma^{Thr1}$ and $H\gamma^{Thr1}$ and $O\gamma^{Thr1}$ (describing proton transfer between Lys33 and Thr1) together with the C1–$O\gamma^{Thr1}$ bond breaking. The resulting PESs, deposited in Fig. S2 of Supporting Information, show how the C1-$O\gamma^{Thr1}$ bond forming and the proton transfer from Thr1 to Lys33 take place in a concerted manner but, the transformation from I1 to P-r7 takes places in two steps through a stable zwitterion intermediate, I2-r7. The presence of this intermediate and the corresponding TSs was confirmed by optimizing the structures at DFT/MM level. According to this result, the FESs for the full mechanism were then explored by generating three surfaces with only anti-symmetric combination of interatomic distances, thus avoiding uncertainties in the position of atoms related if symmetric combination of distances were used. The resulting FES, computed in terms of DFT:AM1/MM 2D-PMFs, are shown in Fig. 5. Thus, the reaction takes place first by a nucleophilic attack of N^{Thr1} to the Cβ of epoxide concomitant with the epoxide ring-opening, as an S_N2 reaction, forming I2-r7. Afterwards a proton is transferred from N^{Thr1} thought water molecule to the O4, thus forming the products P-r7. The energy paths obtained at DFT:AM1/MM level were confirmed by localizing and characterizing the TSs at DFT/MM level. Thus, the position of the optimized TSs at DFT/MM level are located on the quadratic regions of the DFT:AM1/MM surfaces. This result validates the computational protocol employed in the present study.

The formation of the six-membered ring product, P-r6, requires

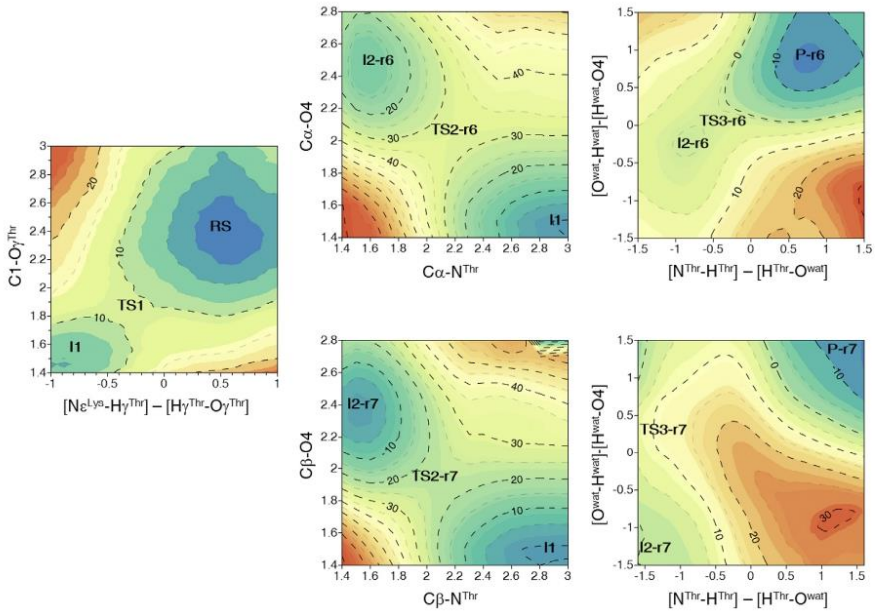


Fig. 5. M06-2X:AM1/MM free energy surfaces for the inhibition of proteasome rendering the six- and seven-membered ring product formation through Mechanism A. Distances in Å and values on iso-energetic lines in kcal·mol^{-1}.

the attack of N^{Thr1} to the Cα of the epoxide ring in I1, as illustrated in Fig. 2. The FES, generated from the corresponding PESs (see Fig. S2), are shown in Fig. 5. As in the case of the formation of the P-r7 mechanism, it appears that this reaction takes place in two steps: first the breaking of the epoxide ring and the forming of the Cα-N^{Thr1} bond, generating a zwitterionic intermediate I2-r6, followed by a step consisting on the water assisted proton transfer from N^{Thr1} to O4, to produce P-r6.

The full free energy profile, derived from the 2D-PMFs and depicted in Fig. 6, shows how the rate limiting step for the generation of both products, P-r6 and P-r7, is the second step, with activation free energies of 36.6 and 27.3 kcal·mol^{-1}, respectively. P-r7 appears as more stable than P-r6 by ca. 10 kcal mol^{-1}.

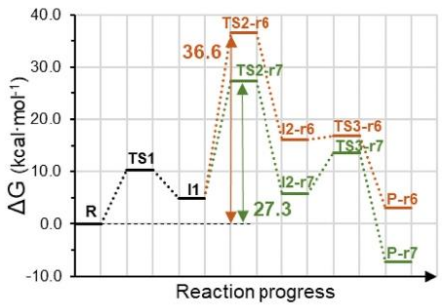


Fig. 6. M06-2X:AM1/MM free energy profile for the inhibition of proteasome rendering the six- and seven-membered ring product through Mechanism A. Results derived from the FESs of Fig. 5.

Thus, the formation of P-r7 would be either kinetically and thermodynamically favoured over the formation of P-r6, which is in agreement with the experimental observations [30]. Nevertheless, the predicted free energy barriers are clearly overestimated, by comparison with the experimental rate constant (k_{obs}/[dihydroeponemycin] = 721 ± 84 $M^{-1}s^{-1}$, when [dihydroeponemycin] = 0.25–5 μM, see U.S. Patent 7,642,369), that render values of the activation free energies ranging between 21 and 23 kcal·mol^{-1} at 25 °C.

Mechanism B. Due to the discrepancy between the experimental rate constants and the predicted activation free energies derived from mechanism A, a new mechanism was proposed. In this one, the N^{Thr1} would attack the epoxide center in the first step, instead of the carbonyl C1 atom of the inhibitor (see Fig. 7).

As depicted in Fig. 7, the first step for the formation of the P-r6 consists in the N^{Thr1} attack to the Cα of epoxide ring, opening the epoxide group of the inhibitor and forming a zwitterionic intermediate, I1-r6. The formation of the P-r7 starts with the nucleophilic attack of N^{Thr1} terminal to the Cβ of the epoxide ring with the following opening of the ring, giving the zwitterionic intermediate I1-r7. The following steps are equivalent for both paths as depicted in Fig. 7. Therefore, the second step consists on the stabilization of the zwitterionic intermediates by transferring a H atom from N^{Thr1} to O4 through water molecule. And finally, the last step of this mechanism consists on the nucleophilic attack of $O\gamma^{Thr1}$ atom to C1, carbonyl atom in inhibitor, and transfer of $H\gamma^{Thr1}$ atom to $N\varepsilon^{Lys33}$. As in the study of Mechanism A, FESs were computed in terms of 2D-PMFs from the previously generated PESs of every single chemical states. The FESs are shown in Fig. 8 while the PESs are deposited in the Supporting Information as Fig. S3.

As observed in Fig. 8, the computed FESs confirm that both reaction paths take place in three steps, as proposed in the

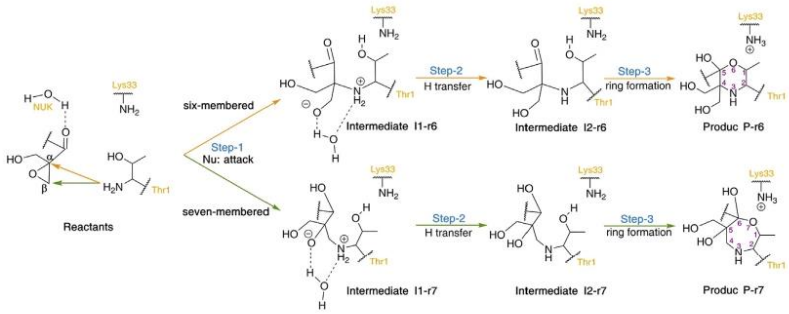


Fig. 7. Schematic representation of the inhibition of proteasome through Mechanism B.

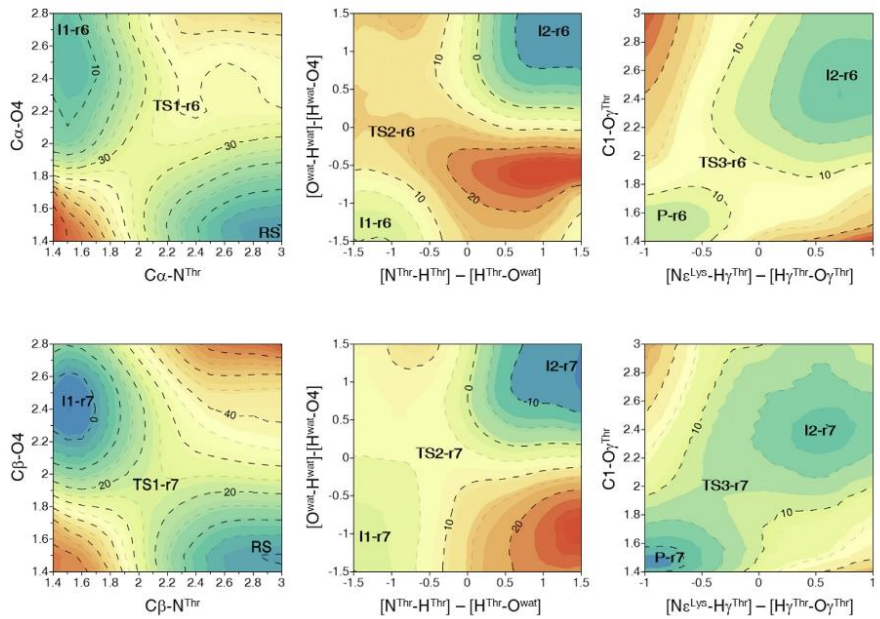


Fig. 8. M06-2X:AM1/MM free energy surfaces for the inhibition of proteasome rendering the six- and seven-membered ring product through Mechanism B. Distances in Å and values on iso-energetic lines in kcal·mol^{-1}.

mechanism depicted in Fig. 7. Same as in the study of Mechanism A, the optimized TSs at DFT/MM support the conclusions derived from the M06-2X:AM1/MM free energy surfaces. Interestingly, the resulting free energy profile derived from these 2D-PMFs, and presented in Fig. 9, clearly shows now that the rate limiting step for the formation of P-r6 and P-r7 corresponds to the nucleophilic attack of N^{Thr1} concomitant with the opening of the epoxide ring. The activation free energy values are 33.7 and 23.7 kcal·mol^{-1} for the P-r6 and P-r7, respectively. This, together with the fact that P-r7 is ca. 20 kcal·mol^{-1} more stable that P-r6, supports the preferential formation of P-r7, as experimentally evidenced by the X-ray crystallography structures. In addition, the free energy barrier for the formation of this proteasome-inhibitor covalent complex is now in very good agreement with the kinetic measurements.

Kinetic Isotope Effects. The study of the inhibition of the proteasome presented above indicates that the most favourable mechanism takes place through mechanism B, where the 1,4-oxazepane ring (seven-membered) is formed. As mentioned, this result is in agreement with experimentally determined crystal structures and experimental reaction rate constants. Calculation of KIEs can provide additional magnitudes that can be measured experimentally to confirm the proposed mechanism. KIE values for the rate-limiting TSs localized at DFT/MM level for Mechanism A and B for six- and seven-member product formation are presented in Table 1. AM1/MM KIEs, obtained as average of 10 different structures of TSs and RSs are reported in Table S4 of Supporting

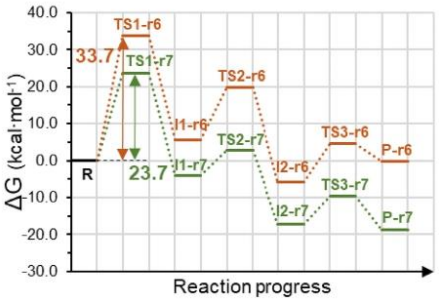


Fig. 9. M06-2X:AM1/MM free energy profile for the inhibition of proteasome rendering the six- and seven-membered ring product through Mechanism B. Results derived from the FESs of Fig. 8.

Table 1
Kinetic isotope effects computed at M06-2X/6-31+G(d,p)//AMBER level of theory for the rate limiting steps in mechanisms A and B, for the formation of six- and seven-member ring products.

	Mechanism A		Mechanism B	
	ring-6	ring-7	ring-6	ring-7
C1 $^{12}C/1-^{14}C$	1.027	1.016	0.997	0.990
Cα $^{12}C/\alpha-^{14}C$	1.114	0.999	1.119	0.992
Cβ $^{12}C/\beta-^{14}C$	0.992	1.121	0.994	1.110
O5 $^{16}O/5-^{18}O$	1.025	1.032	1.014	0.995

Information. Detail of the active site of the structures of the rate limiting TSs localized at M06-2X/6-31+G(d,p)/MM level for Mechanism A and B are depicted in Fig. 10.

The key distances that must be analyzed to recall the chemical step of the rate limiting steps are the ones that describe the S_N2 type reaction produced when the epoxide ring is attacked by N^{Thr1}, either in Cα or Cβ position: TS2 in Mechanism A and TS1 in Mechanism B. As observed in Fig. 10, when epoxide is attacked in Cα, the TS structures show a distance between Cα and N^{Thr1} of around 2 Å, meanwhile distance between Cβ and N^{Thr1} in this TS is never shorter than 2.6 Å (see Table S5 in Supporting Information). Similar trend is observed when epoxide is attacked in Cβ: TS structures show 2 Å distance between Cβ and N^{Thr1}, and Cα—N^{Thr1} distance is never shorter than 2.7 Å (see Table S5 in Supporting Information). These results means that the distinction between six- and seven-member ring formation can be directly related with carbon $^{12}C/^{14}C$ KIEs, in position Cα and Cβ of epoxide. Isotopic substitution of ^{12}C by ^{14}C in Cα renders normal KIEs (>1) only when the carbon plays role of electrophilic center (formation of six-member ring) either in mechanism A or B (1.114 and 1.119, respectively) while a slightly inverse value (<1) of 0.8% should be observed if the reaction would proceed to the formation of the seven-member ring: 0.999 and 0.992 in mechanism A and B, respectively. Taking into account the standard deviation of the $^{12}C/^{14}C$ KIEs on Cα obtained for 100 combinations of 10 structures of RS and TS localized at AM1/MM level (oscillating around value of 1, as reported in Table S4 of Supporting Information), this value could be unity, but clearly different from the normal KIEs obtained in the case of the formation of the six-member ring product. Similar experiment can be done if substitution is done on Cβ: normal $^{12}C/^{14}C$ KIEs would be obtained in case of formation of seven-member ring, either in mechanism A or B (1.121 and 1.110, respectively), while a slightly inverse or no effect should be observed if the reaction would proceed to the formation of the six-member ring: 0.992 and 0.994 in mechanism A and B, respectively.

Next, in order to distinguish between mechanism A and B, secondary $^{12}C/^{14}C$ KIE in position of C1 could be measured. According to our predictions, if inhibition proceed thought mechanism A, $^{12}C/^{14}C$ isotopically substitution on C1 would render normal 2° KIE of 2.7%, while in case of the mechanism B a slightly inverse 2° KIE of 1.0%, or unity, would be expected. Analysis of the TS structures of the different rate limiting steps explains this results: the C1-Oγ^{Thr1} bond is formed in the rate limiting step of mechanism A (bond distance of ca. 1.52 Å) meanwhile, in mechanism B the formation of the C1-Oγ^{Thr1} bond occurs in the last step. The C1-Oγ^{Thr1} bond isn't formed yet in the TS of the rate limiting step (bond distance of ca. 2.2 Å, close to the distances in the RS) thus providing negligible KIEs. Within the same arguments, $^{16}O/^{18}O$ substitution on O4 should provide complementary values of KIEs: significantly more normal in the case of the reaction proceeding by mechanism A than in B.

4. Conclusions

The present paper reports a theoretical study of the inhibition mechanism of proteasome 20S on the β5-subunit by an epoxyketone; dihydroeponemycin. The presence of a dual electrophilic center in this α,β-epoxyketone is responsible of the observed irreversible bind in the active site by formation of two strong covalent bonds with the N-terminal threonine residue. We have first explored the reaction mechanisms rendering a six-membered or a seven-membered ring covalent inhibitor-proteasome product (P-r6, and P-r7, respectively), as previously proposed [30,32,65]. The free energy surfaces generated in terms of potentials of mean force (PMFs) within hybrid QM/MM potentials, with the QM subset of atoms described at semiempirical (AM1) and DFT (M06-2X) level theory, predicts the formation of the seven-membered product as the most favourable one, in agreement with recently determined X-ray structures [30]. Nevertheless, the activation free energy barrier of the rate limiting step appears to be too high, with respect to kinetic experimental data. This result motivated us to propose a

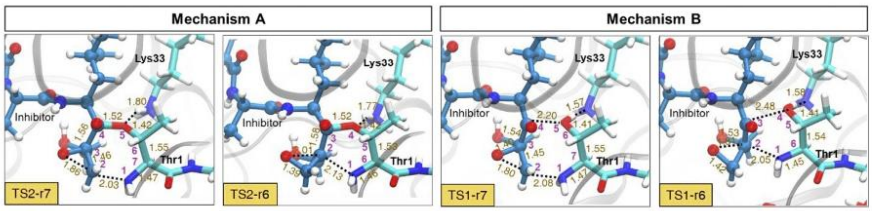


Fig. 10. Structures of rate limiting TSs localized at M06-2X/6-31+G(d,p)/MM level for Mechanism A and B. Distances in Å.

new mechanism, differentiated by reversing the order of chemical steps in the full catalytic process, starting the reaction with epoxide-ring opening and finishing with 1,4-oxazepane product formation. The resulting free energy profiles for this new mechanism predicts, as in the original mechanism, the formation of the seven-member ring product complex to be favoured either thermodynamically and kinetically with respect to the formation of the six-member ring product complex. However, in this case a lower rate limiting activation free energy barrier is obtained, in very good agreement with the range of the experimentally measured inhibition rate constants. Finally, theoretically predicted primary $^{12}C/^{14}C$ KIEs on Cα and Cβ of epoxide and secondary $^{12}C/^{14}C$ KIE on C1 reveal their possible application in distinguishing between the formation of six- and seven-membered product, which could be use in the future to verify the reaction mechanism proposed in the present work.

Acknowledgements

We would like to thank the Spanish Ministerio de Economía y Competitividad and FEDER funds (project CTQ2015-66223-C2) and Universitat Jaume I (project UJI•B2017-31). KŚ thanks the MINECO for a Juan de la Cierva – Incorporación (ref. IJCI-2016-27503) contract. NSA thanks the MINECO for a doctoral FPI grant (BES-2016-078029).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmech.2018.12.062.

References

[1] K. Świderek, I. Tuñón, S. Martí, V. Moliner, J. Bertran, Do zwitterionic species exist in the non-enzymatic peptide bond formation? Chem. Commun. 48 (2012) 11253–11255, https://doi.org/10.1039/c2cc35409h.
[2] A. Krzemińska, V. Moliner, K. Świderek, Dynamic and electrostatic effects on the reaction catalysed by HIV-1 protease, J. Am. Chem. Soc. 138 (2016) 16283–16298, https://doi.org/10.1021/jacs.6b06856.
[3] K.D. Wilkinson, M.K. Urban, A.L. Haas, Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes, J. Biol. Chem. 255 (1980) 7529–7532.
[4] J.D. Etlinger, A.L. Goldberg, A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 54–58.
[5] A. Hershko, A. Ciechanover, I.A. Rose, Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP, Proc. Natl. Acad. Sci. U.S.A. 76 (1979) 3107–3110.
[6] A. Ciechanover, S. Elias, H. Heller, S. Ferber, A. Hershko, Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes, J. Biol. Chem. 255 (1980) 7525–7528.
[7] E.E. Manasanch, N. Korde, A. Zingone, N. Tageja, C. Fernandez de Larrea, M. Bhutani, P. Wu, M. Roschewski, O. Landgren, The proteasome: mechanisms of biology and markers of activity and response to treatment in multiple myeloma, Leuk. Lymphoma 55 (2014) 1707–1714, https://doi.org/10.3109/10428194.2013.828351.
[8] A. Ciechanover, H. Heller, S. Elias, A.L. Haas, A. Hershko, ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 1365–1368, https://doi.org/10.1073/pnas.77.3.1365.
[9] A. Hershko, A. Ciechanover, H. Heller, A.L. Haas, I.A. Rose, Proposed role of ATP in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 1783–1786, https://doi.org/10.1073/pnas.77.4.1783.
[10] J. Neu, Gastroenterology and Nutrition, Elsevier Science, 2012, ISBN 978-1-4377-2603-9, pp. 3–12.
[11] G.N. DeMartino, T.G. Gillette, Proteasomes: machines for all reasons, Cell 129 (2007) 659–662, https://doi.org/10.1016/j.cell.2007.05.00.
[12] W. Heinemeyer, J.A. Kleinschmidt, J. Saidowsky, C. Escher, D.H. Wolf, Proteinase yscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for cell survival, EMBO J. 10 (1991) 555–562.
[13] J. Wang, M.A. Maldonado, The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases, Cell. Mol. Immunol. 3 (2006) 255–261.
[14] C. Naujokat, T. Šarić, Concise review: role and function of the ubiquitin-proteasome system in mammalian stem and progenitor cells, Stem Cell. 25 (2007) 2408–2418, https://doi.org/10.1634/stemcells.2007-0255.
[15] M.S. Willis, W.H.D. Townley-Tilson, E.Y. Kang, W. Jonathon, C. Patterson, Sent to destroy: the ubiquitin protrasome system regulates cell signaling and protein quality control in cardiovascular disease, Circ. Res. 106 (2011) 463–478, https://doi.org/10.1161/CIRCRESAHA.109.208801.
[16] J.A. Benanti, Coordination of cell growth and division by the ubiquitin-proteasome system, Semin. Cell Dev. Biol. 23 (2012) 492–498, https://doi.org/10.1016/j.semcdb.2012.04.005.
[17] S. Bhattacharyya, H. Yu, C. Mim, A. Matouschek, Regulated protein turnover: snapshots of the proteasome in action, Nat. Rev. Mol. Cell Biol. 15 (2014) 122–133, https://doi.org/10.1038/nrm3741.
[18] R. Mishra, A. Upadhyay, V.K. Prajapati, A. Mishra, Proteasome-mediated proteostasis: novel medicinal and pharmacological strategies for diseases, Med. Res. Rev. (2018) 1–58, https://doi.org/10.1002/med.21502.
[19] E.E. Manasanch, R.Z. Orlowski, Proteasome inhibitors in cancer therapy, Nat. Rev. Clin. Oncol. 14 (2017) 417–433, https://doi.org/10.1038/nrclinonc.2016.206.
[20] J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister, R. Huber, Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution, Science 268 (1995) 533–539, https://doi.org/10.1126/science.7725097.
[21] A.J. Marques, R. Palanimurugan, A.C. Mafias, P.C. Ramos, R.J. Dohmen, Catalytic mechanism and assembly of the proteasome, Chem. Rev. 109 (2009) 1509–1536, https://doi.org/10.1021/cr8004857.
[22] G.A. Collins, A.L. Goldberg, The logic of the 26S proteasome, Cell 169 (2017) 792–806, https://doi.org/10.1016/j.cell.2017.04.023.
[23] A. Scheweitzer, A. Aufderheide, T. Rudack, F. Beck, G. Pfeifer, J.M. Plitzko, E. Sakata, K. Schulten, F. Förster, W. Baumeister, Structure of the human 26S proteasome at a resolution of 3.9Å, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 7816–7821, https://doi.org/10.1073/pnas.160805.
[24] J. Adams, V.J. Palombella, E.A. Sausville, J. Johnson, A. Destree, D.D. Lazarus, J. Maas, C.S. Pien, S. Prakash, P.J. Elliott, Proteasome Inhibitors : a novel class of potent and effective antitumor agents, Cancer Res. 59 (1999) 2615–2622.
[25] E.E. Manasanch, R.Z. Orlowski, Proteasome inhibitors in cancer therapy, Nat. Rev. Clin. Oncol. 14 (7) (2017 Jul) 417–433, https://doi.org/10.1038/nrclinonc.2016.206.
[26] L.A. Raedler, Ninlaro (Ixazomib): first oral proteasome inhibitor approved for the treatment of patients with relapsed or refractory multiple myeloma, Am. Heal. Drug Benefits. 9 (2016) 102–105.
[27] B. Muz, R.N. Ghazarian, M. Ou, M.J. Luderer, H.D. Kusdono, A.K. Azab, Spotlight on ixazomib: potential in the treatment of multiple myeloma, Drug Des. Dev. Ther. 10 (2016) 217–226, https://doi.org/10.2147/DDDT.S93602.
[28] D.J. Kuhn, Q. Chen, P.M. Voorhees, J.S. Strader, K.D. Shenk, C.M. Sun, S.D. Demo, M.K. Bennett, F.W.B. van Leeuven, A.A. Chanan-Khan, R.Z. Orlowski, Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome pathway, against preclinical models of multiple myeloma, Blood J 110 (2007) 3281–3290, https://doi.org/10.1182/blood-2007-01-065888.
[29] L. Borissenko, M. Groll, 20S proteasome and its inhibitors: crystallographic knowledge for drug development, Chem. Rev. 107 (2007) 687–717, https://doi.org/10.1021/cr0502504.
[30] J. Schrader, F. Henneberg, R.A. Mata, K. Tittmann, T.R. Schneider, H. Stark, G. Bourenkov, A. Chari, The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design, Science 353 (2016) 594–598, https://doi.org/10.1126/science.aaf8993.
[31] J.E. Baldwin, Rules for ring closure, J. Chem. Soc., Chem. Commun. 0 (1976) 734–736, https://doi.org/10.1039/c39760000734.
[32] M. Groll, K.B. Kim, N. Kairies, R. Huber, C.M. Crews, Crystal structure of epoxomicin: 20S proteasome reveals a molecular basis for selectivity of α',β'-epoxyketone proteasome inhibitors, J. Am. Chem. Soc. 122 (2000) 1237–1238, https://doi.org/10.1021/ja993588m.
[33] D. Wei, B. Lei, M. Tang, C.G. Zhan, Fundamental reaction pathway and free energy profile for inhibition of proteasome by epoxomicin, J. Am. Chem. Soc. 134 (2012) 10436–10450, https://doi.org/10.1021/ja3006463.
[34] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graph. Model. 25 (2006) 247–260, https://doi.org/10.1016/j.jmgm.2005.12.005.
[35] C.R. Sondergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values, J. Chem. Theor. Comput. 7 (2011) 2284–2295, https://doi.org/10.1021/ct200133y.
[36] M.H.M. Olsson, C.R. Sondergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions, J. Chem. Theor. Comput. 7 (2011) 525–537, https://doi.org/10.1021/ct100578z.
[37] C.E.A.F. Schafmeister, W.S. Ross, V. Romanovski, LEAP, University of California, San Francisco, 1995.
[38] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935, https://doi.org/10.1063/1.445869.
[39] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W.E.I. Zhang, R. Yang, P. Cieplak, R.A.Y. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012, https://doi.org/10.1002/jcc.10349.

[40] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802, https://doi.org/10.1002/jcc.20289.
[41] G.S. Grest, K. Kremer, Molecular dynamics simulation for polymers in the presence of a heat bath, Phys. Rev. A 33 (1986) 3628–3631, https://doi.org/10.1103/PhysRevA.33.3628.
[42] M.J. Field, M. Albe, C. Bret, F. Proust-De Martin, A. Thomas, The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials, J. Comput. Chem. 21 (2000) 1088–1100, https://doi.org/10.1002/1096-987X(200009)21:12<1088::AID-JCC5>3.0.CO;2-8.
[43] A. Krzemińska, P. Paneth, V. Moliner, K. Świderek, Binding isotope effects as a tool for distinguishing hydrophobic and hydrophilic binding sites of HIV-1 RT, J. Phys. Chem. B 119 (2015) 917–927, https://doi.org/10.1021/jp506119h.
[44] M.J. Field, P.A. Bash, M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, J. Comput. Chem. 11 (1990) 700–733, https://doi.org/10.1002/jcc.540110605.
[45] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902–3909, https://doi.org/10.1021/ja00299a024.
[46] Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor. Chem. Acc. 120 (2008) 215–241, https://doi.org/10.1007/s00214-007-0310-x.
[47] A. Turner, V. Moliner, I.H. Williams, Transition-state structural refinement with GRACE and CHARMM: flexible QM/MM modelling for lactate dehydrogenase, Phys. Chem. Chem. Phys. 1 (1999) 1323–1331, https://doi.org/10.1039/A808364I.
[48] S. Martí, V. Moliner, I. Tuñón, Improving he QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems, J. Chem. Theor. Comput. 1 (2005) 1008–1016, https://doi.org/10.1021/ct0501396.
[49] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021, https://doi.org/10.1002/jcc.540130812.
[50] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation – umbrella sampling, J. Comput. Phys. 23 (1977) 187–199, https://doi.org/10.1016/0021-9991(77)90121-8.
[51] J.J. Ruiz-Pernía, E. Silla, I. Tuñón, S. Martí, Hybrid Quantum Mechanics/Molecular Mechanics simulations with two-dimensional interpolated corrections: application to enzymatic processes, J. Phys. Chem. B 110 (2006) 17663–17670, https://doi.org/10.1021/jp063520a.
[52] Y.Y. Chuang, J.C. Corchado, D.G. Truhlar, Mapped interpolation scheme for single-point energy corrections in reaction rate calculations and a critical evaluation of dual-level reaction path dynamics methods, J. Phys. Chem. A 103 (1999) 1140–1149, https://doi.org/10.1021/jp9842493.
[53] S. Ferrer, S. Martí, V. Moliner, I. Tuñón, J. Bertran, Understanding the different activities of highly promiscuous MbtI by computational methods, Phys. Chem. Chem. Phys. 14 (2012) 3482–3489, https://doi.org/10.1039/C2CP23149B.
[54] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Jr Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N.J. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Gaussian, Inc., Wallingford, CT, 2009.
[55] K. Świderek, K. Arafet, A. Kohen, V. Moliner, Benchmarking quantum mechanics/molecular mechanics (QM/MM) methods on the thymidylate synthase-catalyzed hydride transfer, J. Chem. Theor. Comput. 13 (2017) 1375–1388, https://doi.org/10.1021/acs.jctc.6b01032.
[56] K. Świderek, I. Tuñón, I.H. Williams, V. Moliner, Insights on the origin of catalysis on glycine N-methyltransferase from computational modelling, J. Am. Chem. Soc. 140 (2018) 4327–4334, https://doi.org/10.1021/jacs.7b13655.
[57] S.A. Kholodar, A.K. Ghosh, K. Świderek, V. Moliner, A. Kohen, Parallel reaction pathways and noncovalent intermediates in thymidylate synthase revealed by experimental and computational tools, Proc. Nat. Acad. Sci. U.S.A. 115 (2018) 10311–10314, https://doi.org/10.1073/pnas.1811059115.
[58] K. Świderek, S. Martí, V. Moliner, Theoretical study of primary reaction of Pseudozyma antarctica lipase B as the starting point to understand its promiscuity, ACS Catal. 4 (2014) 426–434, https://doi.org/10.1021/cs401047k.
[59] S. Martí, V. Moliner, I. Tuñón, I.H. Williams, Computing kinetic isotope effects for chorismate mutase with high accuracy. A new DFT/MM strategy, J. Phys. Chem. B 109 (2005) 3707–3710, https://doi.org/10.1021/jp044387u.
[60] S. Martí, V. Moliner, I. Tuñón, I.H. Williams, QM/MM calculations of kinetic isotope effects in the chorismate mutase active site, Org. Biomol. Chem. 1 (2003) 483–487, https://doi.org/10.1039/B210508J.
[61] G.D. Ruggiero, S.J. Guy, S. Martí, V. Moliner, I.H. Williams, Vibrational analysis of the chorismate rearrangement: relaxed force constants, isotope effects and activation entropies calculated for reaction in vacuum, water and the active site of chorismite mutase, J. Phys. Org. Chem. 17 (2004) 592–601, https://doi.org/10.1002/poc.781.
[62] E.M. Huber, W. Heinemeyer, X. Li, C.S. Arendt, M. Hochstrasser, M. Groll, A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome, Nat. Commun. 7 (2016) 1–10, https://doi.org/10.1038/ncomms10900.
[63] D. Wei, L. Fang, M. Tang, C.G. Zhan, Fundamental reaction pathway for peptide metabolism by proteasome: insights from first-principles quantum mechanical/molecular mechanical free energy calculations, J. Phys. Chem. B 117 (2013) 13418–13434, https://doi.org/10.1021/jjp405337v.
[64] K. Arafet, S. Ferrer, V. Moliner, Computational study of the catalytic mechanism of the cruzain cysteine protease, ACS Catal. 7 (2017) 1207–1215, https://doi.org/10.1021/acscatal.6b3096.
[65] D. Wei, B. Lei, M. Tang, C.G. Zhan, Fundamental reaction pathway and free energy profile for inhibition of proteasome by epoxomicin, J. Am. Chem. Soc. 134 (2012) 10436–10450, https://doi.org/10.1021/ja3006463.

4.2. γ-lactam-β-lactone inhibitors family in 20S proteasome

This section of the results is devoted to the family of γ-lactam-β-lactone inhibitors. SalA irreversibly inhibits the three active sites of the 20S proteasome with different potency.[3] Its small size and irreversibility mode of action makes this molecule an interesting drug lead. The first goal of our study was to clarify the mechanism in which the γ-lactam-β-lactone warhead forms the linkage with the 20S proteasome active site since the exact mechanism has not yet been computationally confirmed. Then the knowledge acquired on SalA mechanistic studies has been applied to its analog, homo-salinosporamide A (hSalA). And finally, the rationalization process of how different modifications can influence SalA as a drug lead compound was investigated through a systematic binding study, where different methodologies were applied to characterize the binding energies.

4.2.1. Salinosporamide A (SalA) inhibition mechanism

In this study, the full inhibition mechanism of SalA has been explored. First, the binding interactions of SalA in the three different active sites of the 20S proteasome were investigated. This was done in two manners: the stability of the starting structure was evaluated first by classical MD simulations; then, the QM/MM inhibitor-protein interaction energy was computed by residue followed by a qualitative estimation of the affinity of this inhibitor with the different active sites. These simulations lead to conclusive results in agreement with experimental measurements of IC_{50} and K_I. Additionally, different S* binding pockets were identified (see Figure 1.1), and their interactions with the inhibitor were computed. Then, the interaction by pocket was compared between the different binding sites (β1, β2, and β5). The results allowed establishing pocket interactions guidelines for evaluating this family of inhibitors through QM/MM interaction energies.

The inhibition mechanism of SalA was proposed previously in literature based on the available data from crystallographic studies.[4] Although the binding mode of action of SalA is clear since it appears bound to $O^{\gamma Thr1}$ in the X-ray structures, there is not enough information for deducing which residue is responsible for Thr1 activation i.e. which plays a role of the proton acceptor in the first step, initiating the reactivity. There are mainly two possible proton acceptors for this step, the most accepted proposal in the literature suggested that it is SalA β-lactone oxygen atom,[4] whereas in an alternative scenario this function could be assigned to lysine $N^{\zeta Lys33}$,

stabilized within the Lys33-Asp17 dyad.[5,6] In the present reactivity studies the free energy profile for both proposed mechanisms were explored, and the protein electrostatic effects were analyzed.

According to our results, in the first proposed mechanism of inhibition of 20S proteasome with SalA, as described above, the inhibitor is assumed to be the Thr1 activator. This mechanism was found to proceed in two steps, where the first step involves the proton transfer to the β-lactone oxygen atom concomitant with the nucleophilic attack of $O^{\gamma Thr1}$ to the carbonyl carbon of the β-lactone ring and the opening of the ring. The second step involves the internal cyclization step where the chlorine leaving group departs and the proton transfer from the O^{SalA} involved in the S_N2 process to the N^{Thr1} atom. The computation of the FES for this mechanism renders a too high free energy barrier, 36.9 kcal·mol^{-1}. The protein electrostatic potential computed on key atoms involved in the reactivity for this process revealed that the highly positive electrostatic potential in the crucial positions does not facilitate the positive charge transfer necessary for transferring the proton in both steps. Therefore, the protein showed to be de-stabilizing this process.

The second explored mechanism, where the Lys33 acts as the proton acceptor and initiates the reaction mechanism was demonstrated to take place in three steps. The first step involves the proton transfer to $N^{\zeta Lys33}$ concomitant with the nucleophilic attack of $O^{\gamma Thr1}$ to the carbonyl carbon in the β-lactone ring, creating a first tetrahedral reaction intermediate stabilized by the protein oxyanion hole (Gly47); the second step involves the opening of the β-lactone ring, and finally, in the third step the S_N2 internal cyclization process occurs, leading to the chlorine release. The free energy profile computed for this process renders a free energy barrier of 20.4 kcal·mol^{-1}, which is in very good agreement with the value that can be deduced from the experimentally measured rate constant,[3] 20.9 kcal·mol^{-1}. Additionally, protein electrostatic potential generated on $N^{\zeta Lys33}$ in the first TS is much less positive, and changes to negative on the same atom after the transfer of hydrogen is accomplished, showing that the protein can rearrange to stabilize the positive charge accumulated Lys33 during this process.

A summary of the results is presented in Figure 4.2.

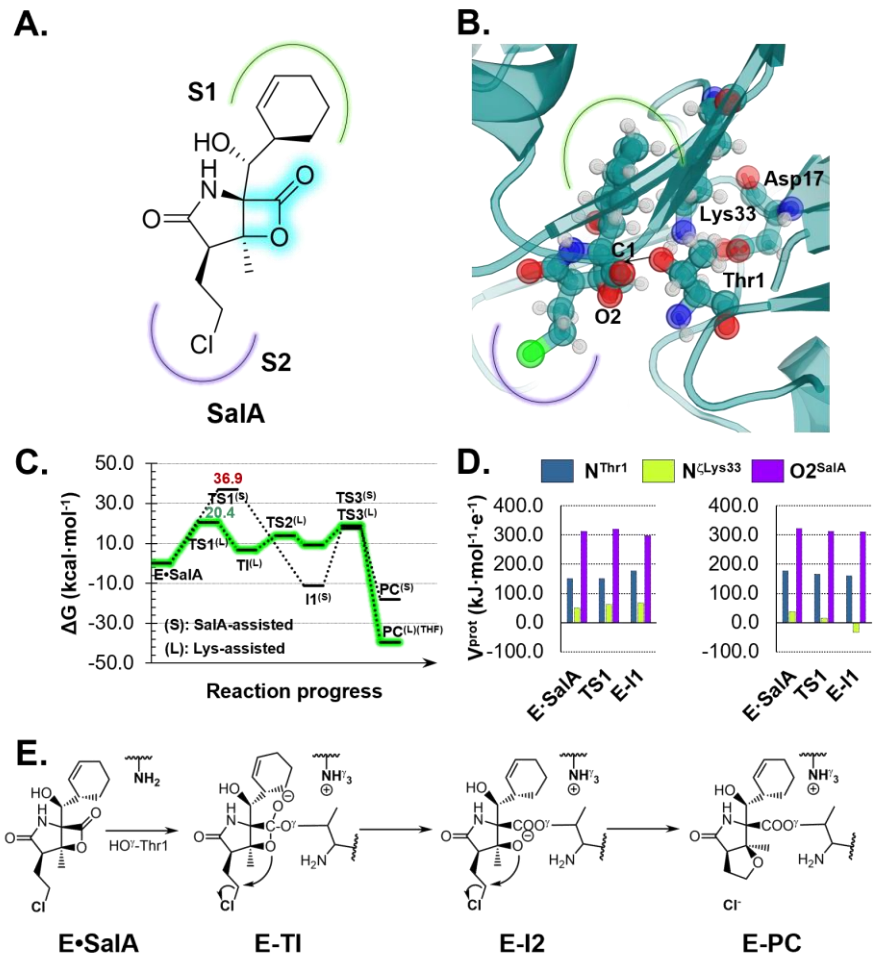


Figure 4.211. **A.** Schematic representation of SalA. **B.** Representation of the E•SalA complex. **C.** Free energy profiles for the inhibitor activated mechanism (in black) and the Lys33 activated mechanism (in green) computed at M06-2X:AM1/MM level of theory. **D.** Evolution of the electrostatic potential generated on key atoms of the active site by the 20S proteasome along the acylation step in inhibitor activated (left) and Lys33 activated (right) mechanisms. **E.** Schematic representation of the proposed reaction mechanism followed by SalA.

These results were presented in the following paper:

Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, *11*, 3575–3589. https://doi.org/10.1021/acscatal.0c05313.

ACS Catalysis

pubs.acs.org/acscatalysis Research Article

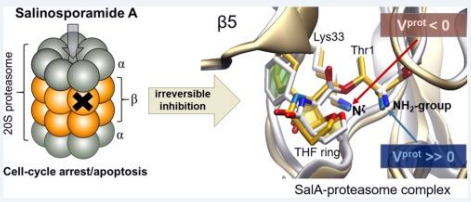
Nature of Irreversible Inhibition of *Human* 20S Proteasome by Salinosporamide A. The Critical Role of Lys−Asp Dyad Revealed from Electrostatic Effects Analysis

Natalia Serrano-Aparicio, Vicent Moliner,* and Katarzyna Świderek*

Cite This: *ACS Catal.* 2021, 11, 3575−3589 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: 20S proteasome is the enzymatic core engine involved in the processes of cellular protein regulation, thus becoming a promising target inter alia in the therapy of many diseases. In this work, the mechanism of the irreversible inhibition of the β5 subunit of human 20S proteasome by salinosporamide A (SalA) was studied by means of QM/MM molecular dynamics simulations. The free energy landscape computed at the M06-2X:AM1/AMBER level reveals the existence of a favorable pathway, where Lys33 actively participates in the inhibition process. This mechanism that presents a free energy barrier for the rate-limiting step in agreement with the experimentally measured rate constant is different from the widely accepted SalA-assisted mechanism. Analysis of the electrostatic features of 20S proteasome reveals the importance of the electrostatic preorganization/reorganization of the enzyme and the pivotal role of Asp17 in modulating pK_a of Lys33 and explains how a molecule, completely unlike the natural substrate of 20S proteasome, binds and inhibits its active site. Our results can contribute to the refinement of efficient inhibitors of a key player in the cell protein quality control system, with the consequent potential applications in medical treatments.



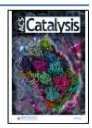
KEYWORDS: *20S proteasome, covalent inhibition, inhibition mechanism, salinosporamide A, QM/MM, free energy surfaces, electrostatic potential, anticancer natural products*

Degradation of proteins is required for the regulation of a broad array of cellular processes, such as cell cycle, cell growth and division, regulation of transcription factors, and assurance of the cellular quality control.[1,2] The majority of intracellular protein degradation (up to 90%) is accomplished by the ubiquitin-proteasome pathway.[3] Proteasome was first considered a garbage disposal unit for cellular waste, but now it is recognized as a multifaceted mediator of many essential cellular processes that take place *via* proteolytic and non-proteolytic mechanisms.[4] In fact, since the discovery of the lysosome[5,6] and later the ubiquitin-proteasome system (UPS),[7,8] the routes of protein regulation started to be a dominant area of research.[9]

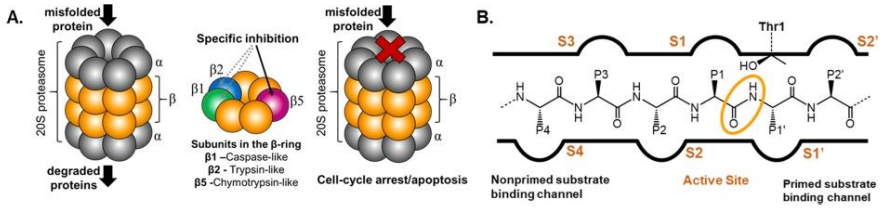
The UPS is the major machinery that regulates the protein degradation pathway in eukaryotes[10] and the 20S proteasome is its enzymatic core engine. This 20S core contains the catalytic centers where proteolysis takes place. Proteasome has a barrel-like structure consisting of four heptameric rings formed by seven α or β subunits each,[11−14] as shown in Scheme 1A. The two inner rings contain three active sites β1, β2, and β5, characterized by the presence of catalytically important *N*-terminal threonine (Thr1).[15] However, each of them has different substrate specificity determined by the presence of a characteristic S1 cavity near the active site (see Scheme 1B). The reason for the singularity of each S1 is related to the residue located at the bottom of the pocket. Thus, β1 binds polypeptides with a basic side chain in position P1 as a result of the presence of Arg45 in this S1, β2 subunit consists of Asp53 that favors binding of an acidic group of P1, and finally Met45 present in S1 of β5 generates its hydrophobic character.[16,17] Therefore, the proteolytic activities of each unit can be characterized by a similar function already observed in other proteins such as caspase-(β1), trypsin-(β2), or chymotrypsin-like (β5) structures.[18]

Because 20S proteasome plays a critical role in many cellular processes, its dysregulation at one or more stages is related to many cellular dysfunctions and consequently diseases.[19,20] In cancer cells, for instance, in which proliferation takes place at a greater rate than in healthy cells, protein synthesis, and degradation is enhanced.[21] Therefore, proteasome has shown

Received: December 4, 2020
Revised: January 29, 2021
Published: March 5, 2021



Scheme 1. (A): 20S Proteasome with Indicated Three Types of Proteolytic Sites Located in the Central Chamber of β-rings (in Orange) that can Serve as a Potential Candidate for its Inhibition, thus Producing the Cell-Cycle Arrest/Apoptosis and (B): Binding Channel of 20S Proteasome with the Bound Peptide in Catalytically Active β Subunits[a]



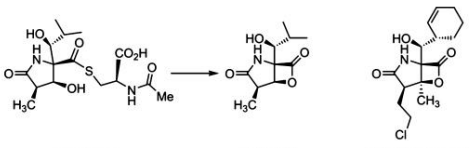
[a]Location of catalytically important Thr1 is highlighted. The scissile peptide bond is labeled by orange oval.

to be a good inhibition target, leading to the attenuate growth and survival of the patients.[22] Moreover, proteasome inhibitors have been proven to be effective against Chagas' disease caused by *Trypanosoma cruzi*[23] or against malaria, provoked by *Plasmodium falciparum*, as these parasites depend on a rapid protein turnover while dividing inside the host cells.[24−26] Also, the bacterial proteasome became an attractive target for drug discovery against *Mycobacterium tuberculosis*, especially important because of its growing resistance to antibiotics.[27] Abnormal levels of immunoproteasome have been also associated with rheumatoid arthritis, colitis, and neurodegenerative disorders such as Alzheimer's and Huntington's diseases that can be treated by proteasome-oriented inhibitors.[16,24,28,29] Different covalent inhibitors were proposed during the last years to reduce the activity of 20S proteasome. They can be classified based on the nature of the electrophilic warhead employed to interact with Thr1.[30] Therefore, using this rule, eight groups of covalent inhibitors can be distinguished such as aldehydes,[31−34] boronates,[31,35] epoxyketones,[36−38] α-ketoaldehydes,[39] vinyl sulfones,[40] syrbactins,[41,42] bacteria-specific oxariazol-2-ones,[43] and β-lactones.[44] However, inhibitory studies have led to just three peptide-like compounds approved by the Food and Drug Administration (FDA) for cancer treatment: two peptide boronate inhibitors, bortezomib (VELCADE) and ixazomib (NINLARO), and the peptide epoxyketone inhibitor carfilzomib (KYPROLIS). Among these the reduced number of possible effective inhibitors, it is important to point out that the borate pharmacophore is extremely toxic and it causes side effects,[45,46] and inherent and acquired resistance was found with bortezomib, the first approved drug.

Up to date, the design of promising anti-cancer drugs oriented to inhibit 20S proteasome was focused mostly on targeting β5 site because of its crucial role in protein degradation and cell permeability of hydrophobic peptides. Thus, although the β1 or β2 subunits also contribute to protein degradation,[47] drugs accepted by the FDA do not inhibit their active sites. In general, not much attention was put on β1 and β2 sites in the process of drug development by pharmaceutical companies.[48] As demonstrated in recent studies, the co-inhibition of β1 and β2 sites increases the cytotoxicity of β5-oriented inhibitors.[49−52] This observation is especially important in the treatment of patients who did not respond to original drugs.[22] In such cases, the implementation of combined therapies[53,54] or the search for new proteasome inhibitors must be applied.[22]

A promising solution has arrived with the discovery of nonpeptidic natural products such as β-lactone-γ-lactams. These compounds are much smaller than the previously developed peptide-like inhibitors. The first active molecule from this family used against the β5 subunit of the proteasome was lactacystin, isolated from *Streptomyces lactacystanaeus*[55,56] (see Scheme 2). It was found that lactacystin is transformed

Scheme 2. Structure of Discovered β-Lactone-γ-lactam Active Inhibitors Against 20S Proteasome



into clastro-lactacystin-β-lactone (omuralide)[55,56] during a spontaneous process in which the lactone ring is formed, by attack of the β-hydroxy of the γ-lactam ring, followed by the cleavage of the thioester bond and loss of *N*-acetylcysteine.[57,58] Later observations that omuralide is able to link covalently by ester bond formation to the threonine of the β5 subunit of the proteasome[59] established an important role of β-lactone-γ-lactams as a potent candidate for the inhibition of the proteasome. A search of this type of compounds was expanded with the discovery of structurally related secondary metabolites. Salinosporamide A (SalA), also known as Marizomib or NPI-0052, isolated from the *marine actinomycete Salinospora tropica*,[60] shares the same core structure with omuralide, but its substituents are modified, as shown in Scheme 2.

The substitution of the isopropyl by a cyclohexenyl group, and the methyl group by a chloroethyl side chain in the α position resulted in remarkable blocking potency of the active sites of the 20S proteasome.[57] The activity of SalA against rabbit and human 20S proteasome was confirmed for all three catalytic subunits β5, β2, and β1 and demonstrated by solving the crystal structure of the SalA complex with the yeast 20S proteasome.[61] *In vivo* studies of SalA in mice confirmed its promising inhibition activity and identified its mechanism as an irreversible process, in contrast to the reversible inhibition character observed for omuralide.[61] SalA exhibits enhanced potency and increased breadth of proteasome inhibitors compared to other β-lactones.[60,62,63] According to the National Institute of Health (USA), SalA has already completed Phase I

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

and in some cases Phase II trials in patients with advanced malignancies, nonsmall cell lung cancer, pancreatic cancer, melanoma, or lymphoma, and multiple myeloma.[64]

Overall, although the activity of SalA in inhibiting the 20S proteasome is well described, the mechanism of the formation of a covalent adduct between the inhibitor and its target is not yet clarified at the molecular level. Attempts to describe this process were mainly based on limited experimental evidence.[61,65,66] Hence, understanding the way SalA blocks the activity of human 20S proteasome can shed light on the important properties and interactions that should be considered in the process of designing reliable covalent inhibitors for this specific system. Generally, to understand the process of any covalent inhibition, the study usually requires the inclusion of two steps: recognition and inactivation. During the first step, the formation of the noncovalent adduct (in this case E•SalA) takes place, which is directly related to the affinity of the inhibitor to the binding pocket of the enzyme. In the second step the formation of a covalent linkage between the inhibitor and the enzyme (E-PC) occurs, as schematically presented in eq 1

$$\mathrm{E} + \mathrm{SalA} \overset{K_i}{\leftrightarrow} \mathrm{E{\bullet}SalA} \overset{K_{inact}}{\longleftrightarrow} \mathrm{E\text{-}PC} \quad (1)$$

SalA was demonstrated to be capable to block all three active sites of 20S proteasome by the formation of covalent E-PC adducts, although its binding affinity was found to be much smaller in $\beta1$ (K_i > 1000 nM) and $\beta2$ (K_i of 97 nM) subunits than in $\beta5$ (K_i of 26 nM).[65]

In the present work, the key interactions between SalA and residues of the binding sites of three different $\beta1$, $\beta2$, and $\beta5$ subunits of human proteasome are analyzed to understand the recognition mechanism of the inhibitor at the atomic level. Subsequently, a theoretical study oriented toward the description of the inhibition mechanism of the $\beta5$-subunit by SalA is carried out. Two mechanisms proposed in the literature are explored using multiscale QM/MM molecular dynamics (MD) approaches. The first mechanism, widely accepted and cited in more than 200 research articles and reviews,[61] was proposed based on the crystallographic studies for SalA in the complex with the yeast 20S proteasome.[63] The second one is inspired by the mechanism of proteolysis catalyzed by yeast 20S proteasome recently suggested by Groll and co-workers, in which Lys33 acts as the general base for $O^{\gamma Thr1}$ activation.[66,67] Experimental evidence showing reduced activity toward substrates and inhibitors observed in the $\beta5$-Lys33Ala variant of 20S proteasome has suggested that besides Thr1, the presence of Lys33 in the active site should be essential for catalysis and also for inhibition.[68,69] Thus, the deprotonated Lys33 could play a key role in activating the *N*-terminal threonine, serving as a general base.[66] Interestingly, Lys33 was also demonstrated to have a key role by yielding an intramolecular cross-linking in the inhibition of immunoproteasome by peptido sulfonylfluoride.[70] The detail description of both mechanisms is presented in the Results and Discussion section, followed by a deep analysis oriented to explain the inhibitory activity of SalA on the 20S proteasome enzyme.

■ COMPUTATIONAL METHODS

Model Set Up. The structure of 20S proteasome from *Homo sapiens* was obtained from the Protein Data Bank (PDB ID: 5LF1).[71] The peptide-like dihydroeponemycin molecule that is present in the active site of the $\beta5$ subunit of the X-ray structure was manually substituted by the covalently bound product of the inhibition by SalA. The orientation of the inhibitor was adapted from the crystal structure of 20S proteasome from *Saccharomyces cerevisiae* complexed with this molecule (PDB ID: 2FAK).[61] Chlorine atom, not present in the crystal structure of the final product, was manually added to the system. Missing force field parameters for the inhibitor and its covalent adduct were obtained using Antechamber software.[72] (See Supporting Information) The protonation state of the titratable residues at pH 7 was predicted as described in a previous work[73] using PropKa program ver. 3.1[74,75] and missing hydrogen atoms were added using the tLEAP[76] module of Amber Tools package. The neutralization of the system was achieved by adding 43 sodium counterions at the electrostatically most favorable positions. Finally, the full protein together with counterions was solvated in an orthorhombic box of TIP3P[77] water molecules of 17.2 × 16.9 × 19.7 nm^3, providing a model containing more than half a million of atoms (527,831 atoms in total). After preliminary minimization using AMBER[78] force field, with NAMD[79] software, the system was heated to 310 K with 0.001 K temperature increment. A 100 ps of *NPT* equilibration was carried out with nonbiased MD simulations. Then, positions of all residues located beyond 40 Å from the inhibitor were frozen and 100 ps of the additional equilibration under *NVT* conditions were applied. First MD simulations were done in the product complex state allowing the structure of the protein to adapt to the presence of the inhibitor. Because the chlorine atom is free in the final product, its position was fixed at 6 Å from the inhibitor, to avoid its diffusion from the active site to the bulk solvent during the simulations. Finally, 5 ns of nonaccelerated unbiased *NVT* MD simulations were performed. The final structure was then modified to prepare the Michaelis complex structure. This new reactant complex served as a starting geometry for 50 ns of nonaccelerated *NVT* MD simulations. In order to model SalA in $\beta2$ and $\beta1$ active sites, the reactant complex was modeled by overlapping the $\beta5$ reactant complex along with the conserved active site residues, Thr1-Thr2. Therefore, the simulation for these two active sites could be started from the Michaelis complex structure. The temperature during the simulations was controlled using a Langevin thermostat.[80] In order to improve the time of simulations, cut-offs for nonbonding interactions were applied using a smooth switching function between 14.5 and 16 Å. According to the time-dependent evolution of the RMSD and the analysis of the evolution of key distances along the MD simulations, the systems can be considered equilibrated (see Supporting Information).

QM/MM Calculations. An equilibrated active conformer structure was selected to explore the inhibition mechanisms. In order to reduce the size of the model for the QM/MM calculations, the system was reduced to a 60 Å sphere centered on the Thr1 residue of the $\beta5$-subunit thus decreasing the total number of atoms in the system to 152,390, as shown in Supporting Information. A small portion of the system, that is, the full inhibitor, Thr1, part of Thr2, and the sidechain of Lys33, highlighted in yellow as shown in Figure 1, is described by quantum mechanics, and the protein and solvent water molecules by AMBER and TIP3P force fields, as implemented in the fDynamo library.[81,82] Two link atoms[83] were inserted on the QM/MM boundary intersecting covalent bonds. These were placed between the Cα–Cβ bond of Lys33 and the C–Cα bond of Thr2 (see Figure 1). The QM region consists of

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

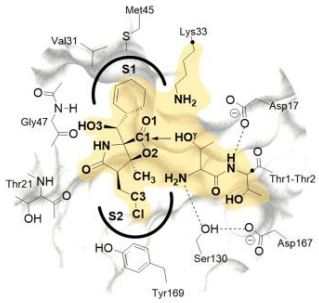


Figure 1. Structure of the active site of β5 subunit of human 20S proteasome with bound SalA. Atoms included into the yellow area were described at QM level of theory. Two black dots indicate positions of link atoms.

85 atoms. The AM1[84] semiempirical Hamiltonian and the M06-2X functional,[85] with the standard 6-31+G(d,p) basis set, were used to treat the QM sub-set of atoms. The positions of the atoms of residues beyond 20 Å from the inhibitor were frozen and the same cut-offs as in MD simulations were applied for the nonbonding interactions. Potential energy surfaces were obtained by the grid scanning of appropriate distinguished reaction coordinates in each of the chemical steps. Based on these surfaces, structures of stationary points of the reactant complex (E•SalA), intermediates (E-I), product complex (E-PC), and transition state (TS) structures were optimized using a micro−macro iteration method[86] at AM1/MM and M06-2X/MM levels, and characterized by computing the matrix of second energy derivatives. IRC paths were traced down from the M06-2X/MM TSs to further confirm that they connect the expected minima.

Free Energy Surfaces. Reaction free energy profiles were obtained based on free energy surfaces (FESs) explored in terms of potentials of mean force (PMFs). The PMF calculation requires a series of molecular dynamics simulations using a limited number of distinguished reaction coordinate variables that are constrained during the simulation. All PMFs were initially calculated at the AM1/MM level using the weighted histogram analysis method (WHAM)[87] combined with the umbrella sampling (US)[88] approach as implemented in fDYNAMO. The value of the force constant used for the harmonic umbrella sampling to generate the PMFs was 2500 kJ mol^{-1} Å^{-2}, and the simulation windows consisted of 5 ps of equilibration and 20 ps of production, with a time step of 1 fs. WHAM calculations were converged with a 10^{-3} density tolerance. Altogether, 11,020 US windows were used for the results presented in this work, as well as the same number of the corresponding single point calculations at a high M06-2X/MM level of theory.

Spline Corrections. To reduce the errors associated with the quantum level of theory employed in our simulations, a

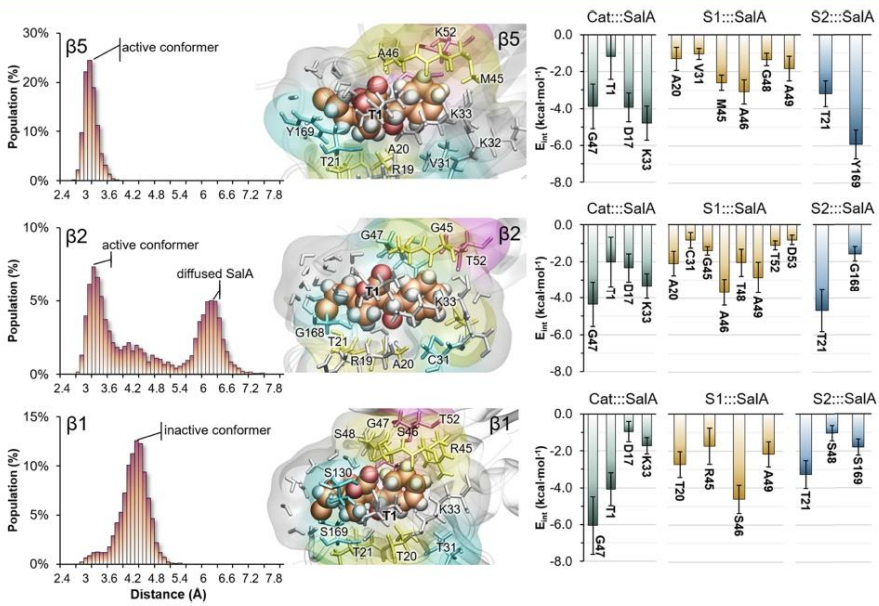


Figure 2. Left panel: distribution of the electrophile-nucleophile distance $d(C1^{SalA}–O^{\gamma Thr1})$ over 5000 structures generated along 50 ns of MM MD simulations in β5, β2, and β1 subunits. Center panel: structure of SalA bound in the active site of each β5, β2, and β1 subunit. Right panel: main average interaction energies (electrostatic plus Lennard−Jones, larger than 1 kcal·mol^{-1}) between SalA and the catalytic site (Cat::SalA), the S1-pocket (S1::SalA), and the S2-pocket (S2::SalA). Values obtained as an average over 10,000 structures generated during 100 ps QM/MM MD simulations.

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

Table 1. Total QM−MM Interaction Energy (in kcal·mol^{-1}), Together with a Contribution from Electrostatic (E_{elec}) and Lennard−Jones Interactions (E_{LJ}), of SalA with the Catalytic Site, Oxyanion Hole, S1-pocket, and S2-pocket[a]

	β5			β2			β1		
subunit	E_{int}	E_{elec}	E_{LJ}	E_{int}	E_{elec}	E_{LJ}	E_{int}	E_{elec}	E_{LJ}
catalytic site	−9.9	−2.3	−7.6	−7.7	−0.9	−6.8	−6.8	0.0	−6.8
oxyanion hole	−3.9	−1.3	−2.6	−4.3	−1.1	−3.2	−6.1	−3.1	−2.9
S1-pocket	−11.2	−0.3	−10.9	−14.9	−1.9	−13.0	−11.3	−1.0	−10.2
S2-pocket	−9.1	−2.1	−7.0	−6.2	−1.5	−4.7	−6.1	−0.6	−5.5
total	−34.1	−6.0	−28.1	−33.1	−5.4	−27.7	−30.2	−4.7	−25.5

[a]See Supporting Information for details.

new energy function defined in terms of interpolated corrections was used.[89−91]

$$E = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\xi_1, \xi_2)] \quad (2)$$

where S denotes a two-dimensional spline function, and its argument is a correction term evaluated from the single-point energy difference between a high-level (HL) and a low-level (LL) calculation of the QM subsystem. The AM1 semi-empirical Hamiltonian was used as the LL method, while the hybrid M06-2X functional with the standard 6-31+G(d,p) basis set was selected for the HL calculation. These calculations were carried out using the Gaussian09 program[92] in combination with fDynamo.

RESULTS AND DISCUSSION

Estimation of SalA-Proteasome Binding Affinities. Natural products targeting the proteasome challenge the classical concept of enzyme binding. Thus, a multiple analysis is required to identify the most potent motif that would ensure effective enzyme blocking.[93,94] Because of the lack of the peptide scaffold in the β-lactone-γ-lactams, the overall ligand architecture is as important as the reactive warhead structure to gain high binding affinity. Therefore, because SalA was demonstrated to possess different affinities toward the active sites located in β1, β2, and β5 subunits (i.e., K_i < 1000, 97 and 23 nM, respectively),[65] an analysis of specific interactions between this inhibitor and its different locations in the enzyme was performed to shed light on the recognition process. An interesting preliminary observation was made already at the stage of analysis of unbiased MM MD simulations (all data from MD analysis can be found in the Supporting Information). As shown in Figure 2, only in the case of the active site of the β5 subunit the strong contact between the electrophile (carbon C1 of the carbonyl group of SalA) and nucleophile (oxygen O$^{\gamma}$ of Thr1) was preserved along 50 ns of simulations. The average C1-O$^{\gamma}$ distance of 3.1 ± 0.1 Å corresponds to the 'active variant' of E•SalA, where the inhibitor adopts an orientation prepared for the chemical transformation. This trend is changing for SalA placed in the active site of the β2 subunit. The strength of the anchoring of SalA in this active site is reduced, as demonstrated by the existence of two conformers. The first one referred to the "active variant" is characterized similarly as in the β5 subunit, by a short C1-O$^{\gamma}$ distance of 3.2 ± 0.1 Å. In the second variant, the SalA diffuses out from the active site that results in the elongation of the C1-O$^{\gamma}$ distance to 6.2 ± 0.1 Å.

Finally, in the active site of β1, which according to experimental data provides the least favorable conditions for binding of inhibitor, the diffusion of SalA is not observed but, as shown in Figure 2, an "active variant" was neither clearly identified. The dominating conformer observed for the β1-SalA complex among the overall population of structures generated during MD simulations is characterized by a weaker contact between the electrophile and nucleophile, according to the C1-O$^{\gamma}$ distance (4.3 ± 0.1 Å), which can be considered as the "inactive" enzyme-inhibitor variant. This result is related to the reorientation of the SalA in the active site of β1 and can explain the poor binding affinity of the inhibitor to this subunit, as well as the unsuccessful efforts that have been done to measure the rate constant of inhibition for this particular active site.[65]

The next stage in the analysis of SalA recognition by all three catalytically active subunits of the proteasome was to compute the energy of interactions formed between this inhibitor and each of the residues of the active site. For this purpose, the residues were classified into three groups depending on their location in the cavity, that is, catalytic site (Cat::SalA), S1-pocket (S1::SalA), and S2-pocket (S2::SalA).

Catalytic Site. As shown in Figure 2, the composition of the catalytic site is conserved among all subunits, and it is characterized by the presence of Thr1, Asp17, Gly47, and Lys33 residues. This similar architecture suggests that similar catalytic mechanisms must be employed in the three β5, β2, and β1 subunits of human 20S Proteasome. It is important to note that because the oxyanion hole is formed by just Gly47, the interactions with this residue will be discussed separately. Based on the results, it can be concluded that the interactions between the warhead of SalA and the residues of the catalytic site depend on the particular enzymatic subunit. For instance, while a strong interaction was found with Thr1 in β1, this is the weakest interaction measured in β5. On the contrary, the most favorable contact in the β5 subunit was created between inhibitor and Lys33. Analysis of the total interaction energies, as listed in Table 1, shows how the strongest overall interactions between SalA and the catalytic site were registered for β5 (−9.9 kcal·mol^{-1}), followed by β2 (−7.7 kcal·mol^{-1}) and β1 subunits (−6.8 kcal·mol^{-1}).

Oxyanion Hole. Surprisingly, the strongest interaction between SalA and the oxyanion hole (Gln47 residue) was found in β1 (−6.1 kcal·mol^{-1}). This is a rather unexpected result considering the strength of the hydrogen bond formed between the carbonyl oxygen of β-lactone and the hydrogen from the backbone (NH-group) of Gly47. In the case of β1, this distance is longer (2.73 ± 0.34 Å) than in β5 (2.63 ± 0.41 Å) and β2 (2.39 ± 0.30 Å). However, as shown in Figure 2, the active site of β1 is the one where the strongest electrostatic interaction between Gln47 and SalA was found. This result can be explained based on the additional interaction formed between SalA and this active site. After the reorientation during MD simulation, the inhibitor has established an additional weak H-bond between the hydrogen of the amino

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

group from γ-lactam of SalA and the backbone oxygen atom (−C═O) of Gly47 (average distance of 2.42 ± 0.29 Å). Such contact is not established either in β5 or in β2 subunits (4.00 ± 0.71 Å, and 3.05 ± 0.40 Å, respectively). As will be discussed below, stronger interactions with the oxyanion hole are expected in the states where $O1^{SalA}$ become negatively charged, which will be associated with a change in the hybridization of the carbonyl $C1^{SalA}$ atom from sp^2 to sp^3.

S1-Pocket. Multiple hydrophobic interactions established between the cyclohexenyl ring of SalA and the S1-pocket that are believed to be crucial for recognition were further analyzed. As indicated in the introduction, all three catalytically active subunits of the proteasome are characterized by different architectures of the S1-pocket. Therefore, the direct comparison of interactions between SalA and each pocket is complicated. Nevertheless, the largest number of favorable contacts for the P1-site was found in the S1-pocket of β2 (8 favorable interactions), which results in the strongest energy of interaction (−14.9 kcal·mol^{-1}), with the highest contribution from three alanine residues, that is, Ala20, Ala46, and Ala49. The interaction with Asp53, a typical residue in the trypsin-like structures, is weak but still beneficial. In the β5 subunit, six contacts were identified including, as in the previous case, key interactions from Ala20, Ala46, and Ala49. The meaningful contribution beneficial for S1::SalA interactions was also provided by Met45, the residue that characterizes the chymotrypsin-like cavity. Nevertheless, the interaction in the S1-pocket of the β5 subunit are weaker than in β2 (−11.2 kcal·mol^{-1}). Virtually the same value of energy interaction was computed for the cyclohexenyl ring bound in the S1-pocket of β1 even though only four favorable contacts were identified (−11.3 kcal·mol^{-1}).

S2-Pocket. The Thr21 residue that is preserved in each S2-pocket among all catalytically active subunits is believed to play a vital role in the selective binding of omuralide and SalA inhibitors over vibralactone because of the strong hydrogen bond interaction with the hydroxy group in P1, which was found in the crystal structure of *yeast* 20S proteasome in the complex with the first two inhibitors.[94] Nevertheless, results of the MD simulations done for our model revealed a lack of strong contacts between Thr21 and SalA, demonstrating that the contribution to the favorable interaction of the inhibitor with this residue is because of beneficial Lennard−Jones interactions. The distance between the oxygen atom $O3^{SalA}$ and the hydrogen of the NH-group of the Thr21 backbone (4.03 ± 0.33 Å, 3.11 ± 0.26 Å, and 3.12 ± 0.28 Å in β5, β2, and β1, respectively) is much longer than the distances of the standard hydrogen bond interactions. Finally, the presence of Tyr169 in the S2-pocket of β5 appears to be vital to guarantee the best complementarity between the inhibitor and the binding pocket. The strong π···Cl interaction created between the chloroethyl group of SalA and protein seem to compensate for weaker interactions settle with the S1-pocket. As shown in Figure 2, the presence of the large aromatic ring on the bottom of the S2-pocket, found only in the β5 subunit, can limit the possible reorientation of the SalA. This does not happen in the wide solvent accessible S2-pockets located in the other subunits. Therefore, it can be concluded that the presence of Tyr169 ensures preserved "active" pose of SalA in the active site of β5.

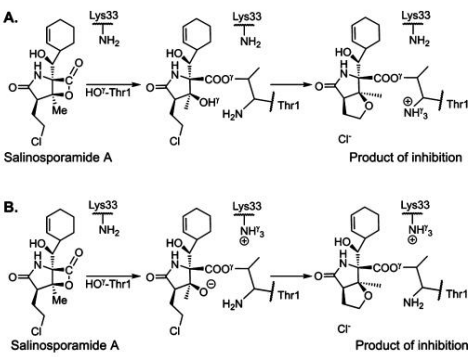
In summary, as shown in Table 1, the favorable interactions between SalA and its binding sites are dominated by van der Waals interactions regardless of the type of the beta subunit (82% contribution in β5, and 84% in β2 and β1). Our results confirmed the previous statement that the electrophilic lactone requires the extensive support of a precisely suited P1 anchor residue to fulfil the minimal requirements to act as a proteasome inhibitor.[93] Nevertheless, based on the obtained results it has to be pointed out that good interaction between SalA and S1-pockets are not sufficient to ensure the best binding affinity. Thus, as revealed by the results obtained on the β5 subunit, the S2-pocket can have a strategically important role in the inhibitor recognition.

Reaction Mechanism. The mechanism of covalent inhibition of 20S proteasome by SalA has never been explored using computational methods. Thus, the main goal of the present study was to clarify the reaction path at the molecular level. Based on experimental evidence showing that values of measured k_{cat} are very similar for the reaction occurring in active sites of β2 (21.6 kcal·mol^{-1}) and β5 (20.6 kcal·mol^{-1}),[65] and through the structural analysis of each of the active sites observed during the MD simulation (revealing equivalent architectures composed by the same active site residues, as described in the previous section), it can be concluded that the reaction must take place by means of the same mechanism in the three catalytic subunits. Therefore, in this work, the mechanism of inhibition of 20S proteasome by SalA is explored in the β5 active site, which should be representative for all three binding sites.

Generally, based on crystallographic and kinetic studies, it has been proposed that the formation of the covalent ester bond between Thr1 and the β-lactone ring of SalA precedes the departure of the chlorine atom and the cyclic ether product formation.[61,66] The remaining question is how the catalytic Thr1 residue is activated. Thus, two contrasting proposals can be found in the literature. In the first one, the SalA-assisted mechanism (Scheme 3A), the catalytic Thr1 is activated by the inhibitor,[61] while in the second one, Lys-assisted mechanism (Scheme 3B), this role is assigned to the neutral Lys33 located in the active site.[66] Thus, herein both proposed mechanisms were explored.

SalA-Assisted Mechanism. The exploration of the mechanism of the formation of the covalent adduct between

Scheme 3. Proposed Reaction Mechanisms of Inhibition of the β5-Active Site of 20S Proteasome by SalA (A); Thr1 is Activated by the SalA,[61,63] and (B); Thr1 is Activated by Lys33[66]



https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

the β5 subunit of the human 20S proteasome and SalA was initiated based on an earlier proposal, originally formulated based on the X-ray structure of SalA in the complex with the enzyme at the product state.[61] This assumes an analogy between the structure of the active site of the proteasome and previously described mechanism of hydrolysis of SalA taking place in aqueous solution.[63] Thus, as shown in Scheme 3A, the inhibition proceeds *via* a two-step reaction. In the first step, the $O^{\gamma Thr1}$ atom of *N*-termini threonine attacks $C1^{SalA}$ of the β-lactone ring causing the cleavage of the β-lactone ring and resulting in the SalA-proteasome adduct formation. SalA is then covalently attached to $O^{\gamma Thr1}$ through an ester linkage, as previously observed for other molecules.[95] In the second step, the deprotonation of the tertiary hydroxy group of SalA by the NH_2 group of Thr1 is concomitant with the nucleophilic displacement of the side-chain chlorine atom by the alcoholate leading to the formation of the tetrahydrofuran (THF) ring. The obtained FESs corresponding to the exploration of the SalA-assisted mechanism are shown in Figure 3A. The

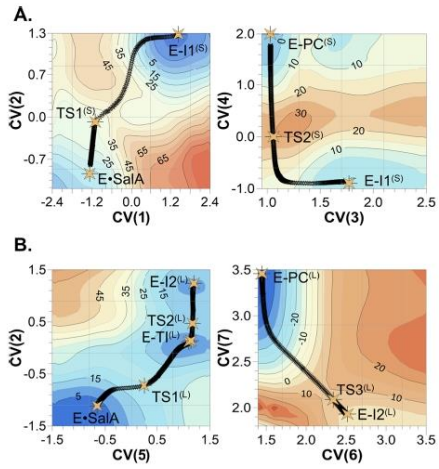


Figure 3. 2D-PMF surfaces computed at M06-2X/AM1/MM level of theory at 37 °C for the 20S proteasome inhibition with SalA thorough (A) SalA-assisted mechanism and (B) Lys-assisted mechanism. Yellow stars on the surfaces indicate positions of M06-2X/MM-optimized structures, which are connected by MEP generated along the IRC calculations, presented as black stars. Distances on the axes are given in Å and iso-energetic lines in kcal·mol^{-1}. Applied collective variables are defined as follow CV(1) = $d(O^{\gamma Thr1}–H^{\gamma Thr1}) – d(H^{\gamma Thr1}–O2^{SalA})$, CV(2) = $d(O2^{SalA}–C1^{SalA}) – d(O^{\gamma Thr1}–C1^{SalA})$, CV(3) = $d(H^{\gamma Thr1}–N^{Thr1})$, CV(4) = $d(O2^{SalA}–C3^{SalA}) – (C3^{SalA}–Cl)$, CV(5) = $d(O^{\gamma Thr1}–H^{\gamma Thr1}) – d(H^{\gamma Thr1}–N^{\zeta Lys33})$, CV(6) = $d(O2^{SalA}–C3^{SalA})$, and CV(7) = $d(C3^{SalA}–Cl)$.

mechanism that can be derived from the FESs agrees with the originally proposed reaction pathway. The process occurs in two steps that are characterized by the presence of two transition states, $TS1^{(S)}$ and $TS2^{(S)}$, whose optimized structures at the M06-2X/MM level are shown in Figure 4. Interestingly, at first glance and according to the key interatomic distances, as reported in Figure 4, the $TS1^{(S)}$ structure could suggest that proton $H^{\gamma Thr1}$ is transferred to the

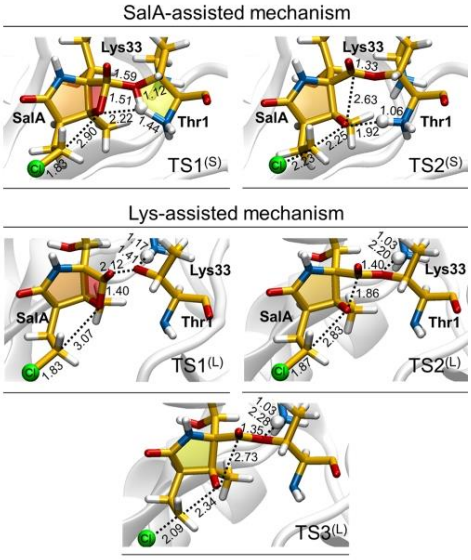


Figure 4. M06-2X/MM-optimized transition state structures for the SalA-assisted mechanism, and the Lys33-assisted mechanism. The geometrical coordinates of these structures can be found in the Supporting Information. Values of key distances are given in Å.

NH_2 terminal group of Thr1 rather than to $O2^{SalA}$ of the lactone ring. Nevertheless, detailed analysis of the evolution of distances along the explored minimum energy pathway (MEP) along the IRC calculations shows that the $H^{\gamma Thr1}$ transfer to $O2^{SalA}$ is assisted by the NH_2 group of Thr1 and only after the lactone ring-opening ($C1^{SalA}–O2^{SalA}$ distance increasing up to 2.6 Å), the proton is finally transferred to the $O2^{SalA}$ oxygen atom. Plots presenting the evolution of key distances and the Bürgi–Dunitz angle obtained in the IRC calculations are provided in the Supporting Information. Nevertheless, according to the high value of the resulting activation free energy of the rate determining step (see Figure 5), this widely accepted reaction mechanism[61,96] appears to be clearly unfavorable. The rate-limiting step of this process, determined

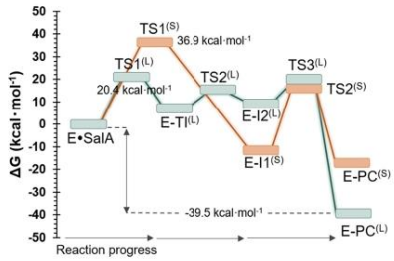


Figure 5. Free energy profile for the irreversible inhibition of 20S proteasome by SalA through SalA-assisted (orange line) and Lys-assisted mechanism (green line).

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575–3589

by the free energy barrier of 36.9 kcal·mol^{-1} of the first step, is much higher than the value that can be deduced from the experimentally measured rate constant, a k_{inact} of 0.015 s^{-1} (20.9 kcal·mol^{-1} at T = 37 °C, according to the Transition State Theory).[65] The exploration of this mechanism using a different protonation state of Lys33 (its standard protonation state) also confirms that this proposal cannot be considered as realistic. The new FESs (deposited in the Supporting Information) show that the high activation free energy of the first step of the SalA-assisted mechanism does not depend on the protonation state of Lys33: $\Delta G^{\ddagger}$ = 35.2 kcal·mol^{-1} for the protonated Lys33 variant of the active site. Therefore, our computational results allow discarding this mechanism and motivated us to explore an alternative course for this reaction.

Lys-Assisted Mechanism. In contrast to the previous inhibitor-assisted mechanism, the proton from Thr1 is now abstracted by Lys33 and the nucleophilic attack of $O^{\gamma Thr1}$ on the carbonyl carbon ($C1^{SalA}$) of β-lactone is accompanied by ring opening. The resulting ionic pair intermediate would be then directly transformed into the product without the participation of any more residues of the active site (see Scheme 3B). The possible role of Lys33 was already indicated in experimental studies on the inhibition of 20S proteasome[66−69] and, in the case of human 20S proteasome inhibited by epoxyketone,[71] this role was confirmed by our previous computational studies.[73] Participation of the nearby lysine residue in the proteolysis mechanism was also considered in a recent theoretical study by Warshel and co-workers, assisting the proton transfer from the hydroxyl group to the amino group of Thr1.[97] Thus, herein the Lys33-assisted mechanism is explored by means of two new FESs, as shown in Figure 3B. The adequate orientation of Lys33 and Asp17 residues (as discussed below) is also observed in the equilibrated structures of β1 and β2, thus supporting the viability of this mechanism in the three active sites.

As can be observed in Figure 3B, the first step of the reaction does not correspond to the concerted process as it was originally assumed. It proceeds in a stepwise manner indicating the existence of a stable tetrahedral intermediate (E-TI) that precedes the opening of the β-lactone ring. This course of the reaction was confirmed by the localization of all three TS structures at the M06-2X/MM level (see Figure 4), the subsequent calculation of the MEP and the evolution of key distances and the Bürgi−Dunitz angle obtained in the IRC calculations are provided in the Supporting Information. Comparative analysis of the evolution of this angle in each mechanism, together with the rehybridization of the carbonyl carbon atom, stresses the lack of correlation between these parameters and the activation free energy because of the different character of the TS1 in both mechanisms (see Figure 4 and Supporting Information). The existence of E-TI is especially interesting because the formation of a tetrahedral intermediate was previously excluded based on experimental studies on β-lactone inhibitors in serine proteases.[98] Additionally, Kim *et al.* suggested that the tetrahedral TSs generated in the acylation reaction may not be stabilized, and thus, the relief of high ring strain energy associated with the lactone ring cleavage may constitute the driving force for the acylation reaction. Our results do not support this statement, showing that the formation of the tetrahedral intermediate is a prelude to the ring opening. The full free energy profile of the Lys33-assisted process (see Figure 5) shows that the rate-limiting step corresponds to the first step with an overall barrier equal to 20.4 kcal·mol^{-1}, a value that is in excellent agreement with the barrier that can be deduced from a rate constant of 0.015 s^{-1} (20.9 kcal·mol^{-1}), experimentally measured at the same temperature. The two remaining steps, the β-lactone ring opening and the THF formation with chloride atom departure, are energetically less demanding, with the corresponding TS energies at 13.8 and 18.8 kcal·mol^{-1} with respect to the E•SalA complex. The highly exergonic character of the last step of the reaction (ΔG_r equal to −39.5 kcal·mol^{-1}) ensures the formation of a very stable covalent adduct of the product in the active site of β5 (E-PC), thereby confirming the irreversible character of this process.

This computational finding is supported by the fact that the structure of the *yeast* 20S proteasome in the complex with SalA corresponds to the E−PC complex.[61] The good overlapping of the X-ray structure of Groll *et al.* and the E-PC$^{(L)}$ complex generated with our theoretical approach (see Figure 6) provides additional argument to support our proposed mechanism, which is summarized in Scheme 4.

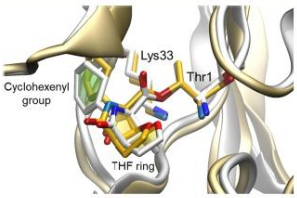


Figure 6. Overlay of the crystal structure of 20S proteasome in the complex with SalA with PDB code 2FAK, (in grey) and the product complex obtained from our QM/MM MD studies of the Lys33-assisted reaction pathway (in orange).

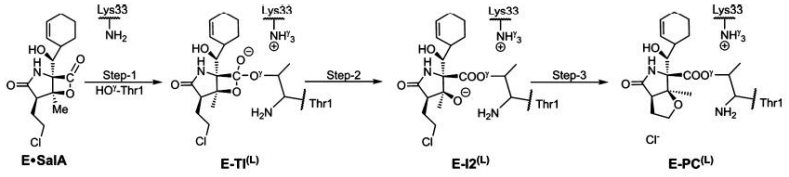
Finally, KIE values were computed for key atoms involved in the rate-limiting step of both mechanisms in order to predict whether they could be used to confirm the theoretical predictions. As presented in the Supporting Information, isotopic labeling of key atoms of SalA resulted in very similar normal KIEs (>1) for the two mechanisms. The differences of KIE values are so subtle that it would be difficult to use them as a decisive tool. Moreover, although, the isotopic substitution of $O^{\gamma Thr1}$ could provide slightly inverse KIEs for the SalA-assisted mechanism (0.997 ± 0.010) and normal KIEs for the Lys-assisted mechanism (1.042 ± 0.011), the computed standard deviation of 1% is too high to provide a robust conclusion.

Electrostatic Effects. Despite that proteasome evolved to cleavage the peptide bond of polypeptides, and most of its inhibitors are peptide-like molecules, it is quite enigmatic how SalA, a molecule completely unlike its natural substrate, efficiently binds in its active site. Therefore, we will try to identify the features of the enzyme that can assist the chemical transformation occurring in the active site. This analysis may shed light on the understating of the inhibition process and, consequently, it can provide new ideas for designing reliable alternative blockers for the proteasome in the future.

One of the properties that has been shown to control the enzyme-catalyzed processes is the electrostatic potential generated by the protein (V^{prot}) on the atoms that are key players in the chemical reaction steps.[99−101] So far, it has been demonstrated that electrostatic effects in enzymes can be

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

Scheme 4. Proposed Lys-Assisted Molecular Mechanism of Irreversible Inhibition of 20S Proteasome with SalA



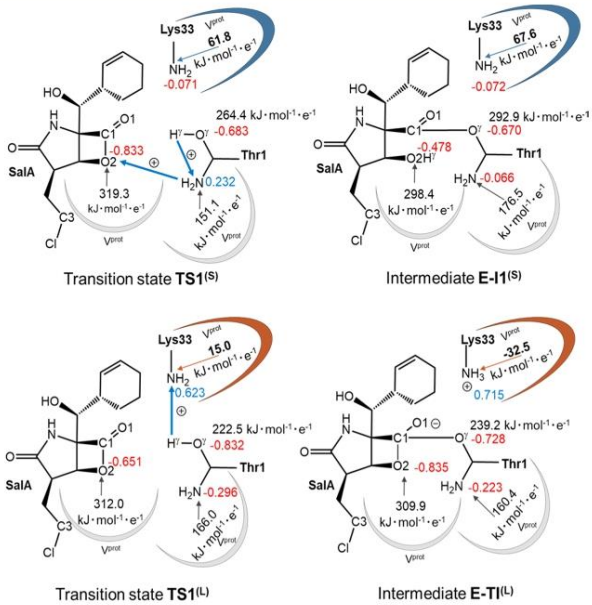


Figure 7. Averaged values of the electrostatic potential (V^{prot}, in black), charges (in red and blue), and schematic representation of the flow of charges taking place in the first step of the reaction in both SalA-assisted and Lys-assisted mechanisms illustrated as blue arrows in the TSs.

crucial, especially in processes where significant charge transfer takes place.[102−105] Therefore, the inhibition of proteasome by SalA is an excellent example to study the role of electrostatic properties of this huge biological machinery, considering the charge flow that takes place in its active site along the reaction.

As shown above, the main divergence between the rate determining step of the SalA-assisted mechanism and the Lys-assisted mechanism is associated with the transfer of the positive charge accompanied by the $H^{\gamma Thr1}$ proton. In the first scenario, the proton is moved to the oxygen $O2^{SalA}$ of the β-lactone ring with important participation of the termini amino group of Thr1, while in the second case, it is directly transferred to the nitrogen N^{ζ} of Lys33, as illustrated in Figure 7.

The computed values of the electrostatic potential generated by 20S proteasome (V^{prot}) on the key atoms of the acylation step in E•SalA, $TS1^{(S)}$, $TS1^{(L)}$, $E\text{-}I1^{(S)}$, and $E\text{-}TI^{(L)}$ are shown in Figure 8, while charges on the key atoms are listed in Table 2. As observed, the V^{prot} values in E•SalA complex are positive in the oxygen $O2^{SalA}$ and in the termini amino group of Thr1 atoms. Moreover, as expected, they are very weakly dependent on the mechanism that is taking place. Hence, it can be concluded that the accumulation of the positive charge on any of these atoms located in the active site would be energetically unfavored. The high value of V^{prot} on the terminal NH_2 group of Thr1 of 150.8 ± 20.1 kJ·mol^{-1}·e^{-1} should be responsible for the meaningful increase of the cost of the acylation step in the SalA-assisted mechanism. As observed, a significant increase of the positive charge is observed on this moiety from E•SalA to $TS1^{(S)}$ (from −0.280 ± 0.019 to 0.232 ± 0.059 a.u.), basically because of the significant positive charge associated with the transferring $H^{\gamma Thr1}$. Alternatively, the proton transfer to the oxygen $O2^{SalA}$ atom appears to be less unfavorable because, despite the loss of the negative charge on $O2^{SalA}$ from the E•SalA complex (−0.623 ± 0.078 a.u.) to $E\text{-}I1^{(S)}$ intermediate (−0.478 ± 0.074 a.u.), it remains negatively charged and the positive V^{prot} generated on this oxygen atom (298.4 ± 16.1 kJ·mol^{-1}·e^{-1}) stabilizes this structure, that is, the product of the first step. Nonetheless, once the β-lactone ring is fully open in $E\text{-}I1^{(S)}$, the NH_2 group becomes almost neutral (−0.066 ±

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

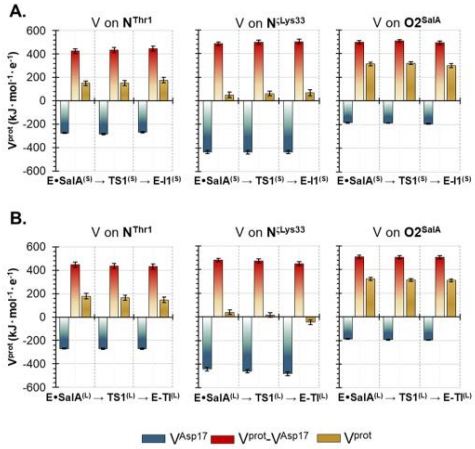


Figure 8. Evolution of the electrostatic potential generated on key atoms of the active site by the 20S proteasome (V^{prot}), by just Asp17 (V^{Asp17}), and by 20S proteasome excluding the contribution of Asp17 ($V^{prot}-V^{Asp17}$) along the acylation step in: (A) SalA-assisted and (B) Lys-assisted mechanisms.

0.045). The achieved electronic charge on this group facilitates the final proton transfer that takes place in the next step of the reaction, which is revealed by a lower activation free energy in the second step of this mechanism. Thus, the NH_2 group accumulates the positive charge at the final E-PC$^{(S)}$ product complex (0.625 ± 0.016).

A significantly different situation is observed in the Lys-assisted mechanism, where the positive charge is stored on N$^{\zeta}$ of Lys33 in the first step of the reaction (Figure 7 and Table 2). The V^{prot} on this atom, 38.8 ± 19.5 kJ·mol^{-1}·e^{-1} in the E•SalA complex, is dramatically smaller than those computed in the proton acceptor atoms of the SalA-assisted mechanism (H$^{\gamma Thr1}$ and O2^{SalA}) and, consequently, the penalty cost of transferring a proton to these atoms must be lower. This result is in agreement with the lower activation energy compared with the SalA-assisted mechanism.

It is widely accepted that enzymes are preorganized to stabilize the TS, and thus meaningful changes of the V^{prot} generated by the protein along the reaction progress are not expected. The almost invariant computed values of the electrostatic potential generated on selected atoms in all stationary points identified along reaction progress, despite being on an inhibition process with a nonnatural substrate, agrees with this statement. Nevertheless, as shown in Figure 8B, an exceptional behavior is observed for V^{prot} generated on Lys33. In this case, the significant reduction of the positive electrostatic potential in the chemical transformation from E•SalA into TS1$^{(L)}$ was observed up to complete change of its sign in E-TI$^{(L)}$ intermediate in the Lys-assisted mechanism (38.8 ± 19.5, 15.0 ± 20.1 and −32.5 ± 22.5 kJ·mol^{-1}·e^{-1}, respectively). This finding suggests that, despite its preorganization, the enzyme is also flexible enough to adapt to the changes occurring in the active site. The existence of this crucial feature was already observed in other not related enzymes.[104] Deeper insights reveal that the significant change of the sign of the electrostatic potential generated on Lys33 is because of the presence of the Asp17 residue, and thus, it is related to the changes occurring in the local environment of the proton acceptor rather than the global effect generated by the protein.

As shown in Figure 8B, the deprotonated Asp17 (confirmed by the calculation of the pK_a values, in the three subunits) generates a highly negative electrostatic potential that meaningfully influence and modulate the value of the total effective V^{prot} observed in the active site. In the SalA-assisted mechanism, the V^{Asp17} generated on N$^{\zeta}$ of Lys33 is rather constant and kept around −433 kJ·mol^{-1}·e^{-1} along the reaction pathway, while in the Lys-assisted mechanism the role of aspartic acid increases and an important shift is observed for V^{Asp17} and V^{prot}. In this case, V^{Asp17} changes from −442.3 to −481.7 kJ·mol^{-1}·e^{-1} in transformation from E•SalA into E-TI$^{(L)}$. Although all atoms of the active site are affected by the presence of Asp17, its main influence would be the modulation of the pK_a value of Lys33. Subsequently, depending on the course of the reaction, the closest surrounding of the Lys33 can be coupled to its electronic density changes provoked by the proton transfer. That includes the minor but significant reorganization mostly related with the shortening of the distance between O$^{\delta1}$ of Asp17 and H$^{\zeta2}$ of Lys33 from 1.91 ± 0.11 Å in the E•SalA complex to 1.78 ± 0.08 Å in the E-TI$^{(L)}$ intermediate in the Lys-assisted process.

Table 2. Averaged Charge Values (e$^-$) on Key Atoms Involved in the Chemical Transformation in SalA- and Lys-Assisted Mechanisms Computed at the M06-2X/MM Level Using the ChelpG Method[a]

		SalA-assisted mechanism			
atoms	E•SalA$^{(S)}$	TS1$^{(S)}$	E-I1$^{(S)}$	TS2$^{(S)}$	E-PC$^{(S)}$
O2^{SalA}	−0.623 ± 0.078	−0.833 ± 0.086	−0.478 ± 0.074	−1.056 ± 0.027	−0.834 ± 0.025
NH$^{Thr1}_2$	−0.280 ± 0.019	0.232 ± 0.059	−0.066 ± 0.045	0.610 ± 0.018	0.625 ± 0.016
NH$_2^{\zeta Lys33}$	0.070 ± 0.079	−0.071 ± 0.044	−0.072 ± 0.037	−0.054 ± 0.029	−0.067 ± 0.039
O$^{\gamma Thr1}$	−0.411 ± 0.013	−0.683 ± 0.020	−0.670 ± 0.064	−0.711 ± 0.041	−0.729 ± 0.050
		Lys-assisted mechanism			
atoms	**E•SalA$^{(L)}$**	**TS1$^{(L)}$**	**E-TI$^{(L)}$**	**TS3$^{(L)}$**	**E-PC$^{(L)}$**
O2^{SalA}	−0.548 ± 0.070	−0.651 ± 0.049	−0.835 ± 0.067	−1.023 ± 0.015	−0.804 ± 0.044
NH$^{Thr1}_2$	−0.297 ± 0.003	−0.296 ± 0.011	−0.223 ± 0.014	−0.151 ± 0.021	−0.155 ± 0.020
NH$_2^{\zeta Lys33}$	0.122 ± 0.020	0.623 ± 0.010	0.715 ± 0.032	0.689 ± 0.008	0.688 ± 0.010
O$^{\gamma Thr1}$	−0.420 ± 0.012	−0.832 ± 0.035	−0.728 ± 0.071	−0.464 ± 0.049	−0.466 ± 0.020

[a]Charges of covalently bound hydrogens are summed into overall charge of heavy atoms.

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

On the contrary, this effect is not observed along the SalA-assisted mechanism, where this distance is virtually unchanged, 2.02 ± 0.14 Å in the E•SalA complex and 1.98 ± 0.13 Å in the E-I1$^{(S)}$ intermediate.

Finally, the existence of a tetrahedral intermediate (E-TI) with the negative charge accumulated on carbonyl oxygen O1^{SalA} of the β-lactone ring, that was neglected in previous studies, can be also explained based on electrostatic effects. The positive value of V^{prot} (424.3 ± 18.1 kJ·mol^{-1}·e^{-1}) computed on O1^{SalA} in E-TI indicates an important role of the protein in stabilizing this intermediate. Thus, despite the active site of the β5-subunit is characterized by the absence of a proper oxyanion hole (only one hydrogen bond interaction between O1^{SalA} and the backbone of Gly47 was identified), the distribution of charges in the structure of 20S proteasome ensures the electrostatic stabilization of this structure.

Interestingly, the fact that the mechanism through the Lys33-assisted mechanism renders a product conformer dramatically more stable than the one obtained through the SalA-assisted mechanism can also be rationalized in terms of electrostatic arguments. Thus, while both the charge and V^{prot} on O2^{SalA} are similar in E•SalA$^{(S)}$ and E•SalA$^{(L)}$, the charge on N^{Thr1} and N$^{\zeta Lys33}$ on the product complex strongly depends on the mechanism employed to reach the product state (0.625 ± 0.016 and −0.067 ± 0.039 a.u. in E-PC$^{(S)}$ and −0.155 ± 0.020 and 0.688 ± 0.010 a.u. in E-PC$^{(L)}$), obviously because of the different protonation states of Thr1 and Lys33 and different polarization effects exerted by different protein conformations. This different charge distribution on the substrate, combined with different V^{prot} on N^{Thr1} and on N$^{\zeta Lys33}$ in the two different product states (104.1 ± 22.0 and 25.2 ± 18.2 kJ·mol^{-1}·e^{-1} in E-PC$^{(S)}$ and 118.1 ± 23.1 and −66.3 ± 21.6 kJ·mol^{-1}·e^{-1} in E-PC$^{(L)}$, respectively), is responsible for a more exergonic reaction when proceeding through the Lys-assisted mechanism. As discussed above, this a priori unexpected different V^{prot} obtained when comparing the protein, in the same state but obtained from different mechanisms, is basically because of the different behaviors of Asp17 in the close vicinity of the former.

In fact, the possibility that a nearby lysine residue acts as a proton acceptor has been largely dismissed given the assumption that at neutral pH it will likely remain positively charged.[106] Our results have proven that the electrostatic potential generated by Asp17 on N$^{\zeta}$ of Lys33 can have a direct impact on the lysine protonation state and, as a consequence, in the energy barrier of the rate-limiting nucleophilic first step of the inhibition process.

■ CONCLUSIONS

In this work, the mechanism of 20S proteasome inhibition by SalA was described at the atomic level using QM/MM MD simulations. The interaction of the inhibitor with three catalytically active subunits (β1, β2, and β5) were first examined. The stabilization of the reactive pose of the inhibitor in the active site as well as the interactions with individual residues from the close vicinity of each active site were determined, demonstrating that the three active sites present similar architectures. Our results show that the favorable interaction with the S1-pocket is not sufficient to ensure the best binding affinity. The binding energies reveal how the S2-pocket can have strategically important role in the inhibitor recognition, especially in the β5 subunit that presents the most favorable total binding energy.

Based on the exploration of the full free energy landscape of the 20S proteasome inhibition by SalA, we demonstrate that the widely accepted proposed mechanism, where SalA is assumed to activate Thr1, is questionable. The high free energy barrier of 36.9 kcal·mol^{-1} obtained for the rate-limiting step is inconsistent with the available experimentally measured rate constant. Thus, an alternate reaction pathway was explored in the β5 subunit, where the lysine residue from the close vicinity of the nucleophile, that is, Thr1, is involved in the chemical transformation. We can assume that this mechanism can also take place in the other subunits, considering the similar architecture of the active sites. The participation of Lys33 in the process leading to the formation of the covalent adduct between the active site and SalA resulted in a mechanism with a significantly lower free energy barrier of 20.4 kcal·mol^{-1}, which is in excellent agreement with the 20.9 kcal·mol^{-1} that can be deduced from the measured rate constant of 0.015 s^{-1} at T = 37 °C.[65] Other ways of activation of the N-terminus of Thr1 *via*, for instance, Ser129, Ser169, and Asp166 can be excluded in this particular case of the inhibition mechanism by SalA because of the pK_a value computationally predicted for Asp166 (6.36) in the reactant complex. Nevertheless, as it has been recently shown by Pavlova *et al.* on a different class of protease,[107] the protonation states of residues can strongly depend on the substrate. Moreover, our analysis of the electrostatic effects reveals the very high positive electrostatic potential created by the protein on Asp166, indicating that the participation of Asp166 in the first step by acting as a proton acceptor can also be excluded. Interestingly, this Lys-assisted mechanism shows similarities with that proposed for the fatty acid amide hydrolase (FAAH), in which (for both substrates and covalent inhibitors) a neutral lysine residue acts to activate a nucleophile (in that case, a serine pair) and, again, the stabilization of the protonated lysine appears to be important (as shown by analysis of reactive and nonreactive conformations).[108,109] These results suggest a possible generality of such effects in proteins.

Additionally, the highly exergonic last step of the reaction, determined from our calculations, confirmed the irreversible character of proteasome inhibition by SalA. The Lys-assisted mechanism suggests the presence of a stable tetrahedral intermediate along the inhibition pathway. The existence of this intermediate contradicts the experimentally supported proposal that its formation is impossible in the case of β-lactone derivatives.[97]

Analysis of the electrostatic potential generated in the active site of the β5-subunit revealed the origin of the dramatic difference observed in free energy barriers between SalA- and Lys-assisted mechanisms. The vital catalytic role was assigned to Asp17 in the close vicinity of Lys33, showing that even small changes in its position can modulate the role of the lysine as a proton acceptor and thus affecting its pK_a. Thus, the presence of Asp17 nearby the active site allows to compensate for the high positive electrostatic potential generated by the remaining residues of the protein, and consequently, it helps to stabilize the positive charge accumulated on lysine during the reaction pathway. Interestingly, it appears that the most favorable reaction path is associated with certain flexibility of the 20S proteasome along the reaction that facilitates the electrostatic stabilization of not only the rate-limiting TS but the conformation of the covalent product complex. Thus, it appears that the structure of this complex biocatalyst is preorganized in such a way that it electrostatically stabilizes the

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

TS of the rate-limiting step of the reaction with a molecule completely unlike its natural substrate. Nevertheless, at the same time, it is flexible enough to reorganize itself during the reaction path also stabilizing different states appearing along the reaction path. These two features of 20S proteasome explain the capabilities of SalA to act as a covalent inhibitor.

Finally, the proposed mechanism of inhibition opens a new question related to the hydrolysis of an ester bond formed between Thr1 and the inhibitor. As previously suggested based on the SalA-assisted inhibition mechanism, the protonated variant of the neutral form of the amino-termini group of Thr1 generated in the product complex (E-PC$^{(S)}$) would prevent its participation in water activation and possible hydrolysis of the enzyme-inhibitor adduct.[58] However, in the herein proposed Lys-assisted inhibition mechanism, the amino group of Thr1 is deprotonated in the product complex, and therefore, it could be easily involved in the hydrolysis process. Nevertheless, assuming that proton would be transferred from the water molecule to NH_2 group, the positive charge would be accumulated on the nitrogen atom. Consequently, this process would be expected to be energetically unfavorable considering the high positive electrostatic potential generated by the protein in this position and could explain the irreversible character of inhibition by SalA. Nevertheless, future studies should be conducted to shed more light onto this problem.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.0c05313.

MM parameters for SalA; details of the MD simulations; free energy surfaces; evolution of key distances along the reaction paths; averaged ChelpG charges on key atoms; results of study of the SalA-assisted mechanism with the protonated state of Lys33; list of the electrostatic potential on key atoms and kinetic isotope effects; and cartesian coordinates of transition structures optimized at M06-2X/MM level (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Vicent Moliner − *Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain;* orcid.org/0000-0002-3665-3391; Email: moliner@uji.es

Katarzyna Świderek − *Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain;* orcid.org/0000-0002-7528-1551; Email: swiderek@uji.es

Author

Natalia Serrano-Aparicio − *Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.0c05313

Funding

This work was supported by the Spanish Ministerio de Ciencia e Innovación (grant PGC2018-094852-B-C21 and PID2019-107098RJ-I00), the Generalitat Valenciana (grant AICO/2019/195 and SEJI/2020/007) and Universitat Jaume I (grant UJI-B2017-31 and UJI-A2019-04).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank the Spanish Ministerio de Ciencia e Innovación (grant PGC2018-094852-B-C21 and PID2019-107098RJ-I00), the Generalitat Valenciana (grant AICO/2019/195 and SEJIGENT/2020/007) and Universitat Jaume I (grant UJI-B2020-03 and UJI-A2019-04). N.S.A. thanks the MINECO for doctoral FPI grant (BES-2016-078029). The authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I. The authors thankfully acknowledge the computer resources at Pirineus and the technical support provided by Consorci de Serveis Universitaris de Catalunya (RES-QSB-2020-3-0012).

■ REFERENCES

(1) Eisenberg-Lerner, A.; Benyair, R.; Hizkiahou, N.; Nudel, N.; Maor, R.; Kramer, M. P.; Shmueli, M. D.; Zigdon, I.; Cherniavsky Lev, M.; Ulman, A.; Sagiv, J. Y.; Dayan, M.; Dassa, B.; Rosenwald, M.; Shachar, I.; Li, J.; Wang, Y.; Dezorella, N.; Khan, S.; Porat, Z.; Shimoni, E.; Avinoam, O.; Merbl, Y. Golgi Organization Is Regulated by Proteasomal Degradation. *Nat. Commun.* **2020**, *11*, 409−423.

(2) Rousseau, A.; Bertolotti, A. Regulation of Proteasome Assembly and Activity in Health and Disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 697−712.

(3) Neu, J. *Gastroenterology and Nutrition*; Elsevier Science, 2012; pp 3−12. ISBN: 978-1-4377-2603-9.

(4) DeMartino, G. N.; Gillette, T. G. Proteasomes: Machines for All Reasons. *Cell* **2007**, *129*, 659−662.

(5) Duve, C. D.; Gianetto, R.; Appelmans, F.; Wattiaux, R. Enzymic Content of the Mitochondria Fraction. *Nature* **1953**, *172*, 1143−1144.

(6) Gianetto, R.; De Duve, C. Tissue Fractionation Studies. 4. Comparative Study of the Binding of Acid Phosphatase, β-Glucuronidase and Cathepsin by Rat-Liver Particles. *Biochem. J.* **1955**, *59*, 433−438.

(7) Hershko, A.; Ciechanover, A.; Rose, I. A. Resolution of the ATP-Dependent Proteolytic System from Reticulocytes: A Component That Interacts with ATP. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3107−3110.

(8) Ciechanover, A.; Elias, S.; Heller, H.; Ferber, S.; Hershko, A. Characterization of the Heatstable Polypeptide of the ATP-Dependent Proteolytic System from Reticulocytes. *J. Biol. Chem.* **1980**, *255*, 7525−7528.

(9) Ciechanover, A. Intracellular Protein Degradation: From a Vague Idea through the Lysosome and the Ubiquitin-Proteasome System, and onto Human Diseases and Drug Targeting (Nobel Lecture). *Angew. Chem., Int. Ed.* **2005**, *44*, 5944−5967.

(10) Manasanch, E. E.; Korde, N.; Zingone, A.; Tageja, N.; Fernandez De Larrea, C.; Bhutani, M.; Wu, P.; Roschewski, M.; Landgren, O. The Proteasome: Mechanisms of Biology and Markers of Activity and Response to Treatment in Multiple Myeloma. *Leuk. Lymphoma* **2014**, *55*, 1707−1714.

(11) Tanaka, K. The Proteasome: Overview of Structure and Functions. *Proc. Jpn. Acad. B* **2009**, *85*, 12−36.

(12) Coux, O.; Tanaka, K.; Goldberg, A. L. Structure and Functions of the 20S and 26S Proteasomes. *Annu. Rev. Biochem.* **1996**, *65*, 801−847.

(13) Lupas, A.; Zwickl, P.; Wenzel, T.; Seemüller, E.; Baumeister, W. Structure and function of the 20S proteasome and of its regulatory complexes. *Cold Spring Harbor Symp. Quant. Biol.* **1995**, *60*, 515−524.

(14) Kunjappu, M. J.; Hochstrasser, M. Assembly of the 20S Proteasome. *Biochim. Biophys. Acta Mol. Cell Res.* **2014**, *1843*, 2−12.

(15) Lowe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal Structure of the 20S Proteasome from the Archaeon T. Acidophilum at 3.4 A Resolution. *Science* **1995**, *268*, 533−539.

(16) Huber, E. M.; Groll, M. Inhibitors for the Immuno- and Constitutive Proteasome: Current and Future Trends in Drug Development. *Angew. Chem., Int. Ed.* **2012**, *51*, 8708−8720.
(17) Marques, A. J.; Palanimurugan, R.; Mafias, A. C.; Ramos, P. C.; Dohmen, R. J. Catalytic Mechanism and Assembly of the Proteasome. *Chem. Rev.* **2009**, *109*, 1509−1536.
(18) Orlowski, M.; Cardozo, C.; Michaud, C. Evidence for the Presence of Five Distinct Proteolytic Components in the Pituitary Multicatalytic Proteinase Complex. Properties of Two Components Cleaving Bonds on the Carboxyl Side of Branched Chain and Small Neutral Amino Acids. *Biochemistry* **1993**, *32*, 1563−1572.
(19) Mishra, R.; Upadhyay, A.; Prajapati, V. K.; Mishra, A. Proteasome-Mediated Proteostasis: Novel Medicinal and Pharmacological Strategies for Diseases. *Med. Res. Rev.* **2018**, *38*, 1916−1973.
(20) Sengupta, S.; Mehta, G. Non-Peptidic Natural Products as Ubiquitin-Proteasome Inhibitors. *Tetrahedron* **2019**, *75*, 817−853.
(21) Potts, B. C. M.; Albitar, M. X.; Anderson, K. C.; Baritaki, S.; Berkers, C.; Bonavida, B.; Chandra, J.; Chauhan, D.; Cusack, J. C., Jr.; Fenical, W.; Ghobrial, I. M.; Groll, M.; Jensen, P. R.; Lam, K. S.; Lloyd, G. K.; McBride, W.; McConkey, D. J.; Miller, C. P.; Neuteboom, S. T. C.; Oki, Y.; Ovaa, H.; Pajonk, F.; Richardson, P. G.; Roccaro, A. M.; Sloss, C. M.; Spear, M. A.; Valashi, E.; Younes, A.; Palladino, M. A. Marizomib, a Proteasome Inhibitor for All Seasons: Preclinical Profile and a Framework for Clinical Trials. *Curr. Cancer Drug Targets* **2011**, *11*, 254−284.
(22) Manasanch, E. E.; Orlowski, R. Z. Proteasome Inhibitors in Cancer Therapy. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 417−433.
(23) Zmuda, F.; Sastry, L.; Shepherd, S. M.; Jones, D.; Scott, A.; Craggs, P. D.; Cortes, A.; Gray, D. W.; Torrie, L. S.; Rycker, M. De. Identification of Novel Trypanosoma Cruzi Proteasome Inhibitors Using a Luminiscence-Based High-Throughput Screening Assay. *Antimicrob. Agents Chemother.* **2019**, *63*, No. e00309.
(24) Cromm, P. M.; Crews, C. M. The Proteasome in Modern Drug Discovery: Second Life of a Highly Valuable Drug Target. *ACS Cent. Sci.* **2017**, *3*, 830−838.
(25) Aminake, M. N.; Arndt, H.-D.; Pradel, G. The Proteasome of Malaria Parasites: A Multi-Stage Drug Target for Chemotherapeutic Intervention? *Int. J. Parasitol.: Drugs Drug Resist.* **2012**, *2*, 1−10.
(26) Le Chapelain, C.; Groll, M. Rational Design of Proteasome Inhibitors as Antimalarial Drugs. *Angew. Chem., Int. Ed.* **2016**, *55*, 6370−6372.
(27) Gandotra, S.; Schnappinger, D.; Monteleone, M.; Hillen, W.; Ehrt, S. *In Vivo* Gene Silencing Identifies the Mycobacterium Tuberculosis Proteasome as Essential for the Bacteria to Persist in Mice. *Nat. Med.* **2007**, *13*, 1515−1520.
(28) Kim, B. K.; Crews, C. M. Chemical Genetics: Exploring the Role of the Proteasome in Cell Biology Using Natural Products and Other Small Molecule Proteasome Inhibitors. *J. Med. Chem.* **2008**, *51*, 2600−2605.
(29) Lupoli, T. J.; Vaubourgeix, J.; Burns-Huang, K.; Gold, B. Targeting the Proteostasis Network for Mycobacterial Drug Discovery. *ACS Infect. Dis.* **2018**, *4*, 478−498.
(30) Kisselev, A. F.; van der Linden, W. A.; Overkleeft, H. S. Proteasome Inhibitors: An Expanding Army Attacking a Unique Target. *Chem. Biol.* **2012**, *19*, 99−115.
(31) Adams, J.; Behnke, M.; Chen, S.; Cruickshank, A. A.; Dick, L. R.; Grenier, L.; Klunder, J. M.; Ma, Y.-T.; Plamondon, L.; Stein, R. L. Potent and Selective Inhibitors of the Proteasome: Dipeptidyl Boronic Acids. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 333−338.
(32) Palombella, V. J.; Rando, O. J.; Goldberg, A. L.; Maniatis, T. The Ubiquitinproteasome Pathway Is Required for Processing the NF-κB1 Precursor Protein and the Activation of NF-κB. *Cell* **1994**, *78*, 773−785.
(33) Tsubuki, S.; Kawasaki, H.; Saito, Y.; Miyashita, N.; Inomata, M.; Kawashima, S. Purification and characterization of a Z-Leu-Leu-Leu-MCA degrading protease expected to regulate neurite formation: a novel catalytic activity in proteasome. *Biochem. Biophys. Res. Commun.* **1993**, *196*, 1195−1201.
(34) Figueiredo-Pereira, M. E.; Berg, K. A.; Wilk, S. A New Inhibitor of the Chymotrypsin-Like Activity of the Multicatalytic Proteinase Complex (20S Proteasome) Induces Accumulation of Ubiquitin-Protein Conjugates in a Neuronal Cell. *J. Neurochem.* **1994**, *63*, 1578−1581.
(35) Fishovitz, J.; Li, M.; Frase, H.; Hudak, J.; Craig, S.; Ko, K.; Berdis, A. J.; Suzuki, C. K.; Lee, I. Active-Site-Directed Chemical Tools for Profiling Mitochondrial Lon Protease. *ACS Chem. Biol.* **2011**, *6*, 781−788.
(36) Meng, L.; Mohan, R.; Kwok, B. H. B.; Elofsson, M.; Sin, N.; Crews, C. M. Epoxomicin, a Potent and Selective Proteasome Inhibitor, Exhibits in Vivo Antiinflammatory Activity. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10403−10408.
(37) Meng, L.; Kwok, B. H. B. B.; Sin, N.; Crews, C. M. Eponemycin Exerts Its Antitumor Effect through the Inhibition of Proteasome Function. *Cancer Res.* **1999**, *59*, 2798−2801.
(38) Demo, S. D.; Kirk, C. J.; Aujay, M. A.; Buchholz, T. J.; Dajee, M.; Ho, M. N.; Jiang, J.; Laidig, G. J.; Lewis, E. R.; Parlati, F.; Shenk, K. D.; Smyth, M. S.; Sun, C. M.; Vallone, M. K.; Woo, T. M.; Molineaux, C. J.; Bennett, M. K. Antitumor Activity of PR-171, a Novel Irreversible Inhibitor of the Proteasome. *Cancer Res.* **2007**, *67*, 6383−6391.
(39) Lynas, J. F.; Harriott, P.; Healy, A.; McKervey, M. A.; Walker, B. Inhibitors of the Chymotrypsin-like Activity of Proteasome Based on Di- and Tri-Peptidyl α-Keto Aldehyde (Glyoxals). *Bioorg. Med. Chem. Lett* **1998**, *8*, 373−378.
(40) Bogyo, M.; McMaster, J. S.; Gaczynska, M.; Tortorella, D.; Goldberg, A. L.; Ploegh, H. Covalent Modification of the Active Site Threonine of Proteasomal β Subunits and the Escherichia Coli Homolog HslV by a New Class of Inhibitors. *Proc. Natl. Acad. Sci.* **1997**, *94*, 6629−6634.
(41) Groll, M.; Schellenberg, B.; Bachmann, A. S.; Archer, C. R.; Huber, R.; Powell, T. K.; Lindow, S.; Kaiser, M.; Dudler, R. A Plant Pathogen Virulence Factor Inhibits the Eukaryotic Proteasome by a Novel Mechanism. *Nature* **2008**, *452*, 755−758.
(42) Oka, M.; Yaginuma, K.; Numata, K.; Konishi, M.; Oki, T.; Kawaguchi, H. Glidobactins A, B and C, New Antitumor Antibiotics. II. Structure Elucidation. *J. Antibiot.* **1988**, *41*, 1338−1350.
(43) Lin, G.; Li, D.; De Carvalho, L. P. S.; Deng, H.; Tao, H.; Vogt, G.; Wu, K.; Schneider, J.; Chidawanyika, T.; Warren, J. D.; Li, H.; Nathan, C. Inhibitors Selective for Mycobacterial versus Human Proteasomes. *Nature* **2009**, *461*, 621−626.
(44) Craiu, A.; Gaczynska, M.; Akopian, T.; Gramm, C. F.; Fenteany, G.; Goldberg, A. L.; Rock, K. L. Lactacystin and Clasto-Lactacystin β-Lactone Modify Multiple Proteasome β-Subunits and Inhibit Intracellular Protein Degradation and Major Histocompatibility Complex Class I Antigen Presentation. *J. Biol. Chem.* **1997**, *272*, 13437−13445.
(45) Richardson, P. G.; Briemberg, H.; Jagannath, S.; Wen, P. Y.; Barlogie, B.; Berenson, J.; Singhal, S.; Siegel, D. S.; Irwin, D.; Schuster, M.; Srkalovic, G.; Alexanian, R.; Rajkumar, S. V.; Limentani, S.; Alsina, M.; Orlowski, R. Z.; Najarian, K.; Esseltine, D.; Anderson, K. C.; Amato, A. A. Frequency, Characteristics, and Reversibility of Peripheral Neuropathy during Treatment of Advanced Multiple Myeloma with Bortezomib. *J. Clin. Oncol.* **2006**, *24*, 3113−3120.
(46) Lonial, S.; Waller, E. K.; Richardson, P. G.; Jagannath, S.; Orlowski, R. Z.; Giver, C. R.; Jaye, D. L.; Francis, D.; Giusti, S.; Torre, C.; Barlogie, B.; Berenson, J. R.; Singhal, S.; Schenkein, D. P.; Esseltine, D.-L. W.; Anderson, J.; Xiao, H.; Heffner, L. T.; Anderson, K. C. Risk Factors and Kinetics of Thrombocytopenia Associated with Bortezomib for Relapsed, Refractory Multiple Myeloma. *Blood* **2005**, *106*, 3777−3784.
(47) 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.
(48) Goldberg, A. L. Development of Proteasome Inhibitors as Research Tools and Cancer Drugs. *J. Cell Biol.* **2012**, *199*, 583−588.

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

(49) Britton, M.; Lucas, M. M.; Downey, S. L.; Screen, M.; Pletnev, A. A.; Verdoes, M.; Tokhunts, R. A.; Amir, O.; Goddard, A. L.; Pelphrey, P. M.; Wright, D. L.; Overkleeft, H. S.; Kisselev, A. F. Selective Inhibitor of Proteasome's Caspase-like Sites Sensitizes Cells to Specific Inhibition of Chymotrypsin-like Sites. *Chem. Biol.* **2009**, *16*, 1278−1289.
(50) Kraus, M.; Bader, J.; Geurink, P. P.; Weyburne, E. S.; Mirabella, A. C.; Silzle, T.; Shabaneh, T. B.; van der Linden, W. A.; de Bruin, G.; Haile, S. R.; van Rooden, E.; Appenzeller, C.; Li, N.; Kisselev, A. F.; Overkleeft, H.; Driessen, C. The Novel B2-Selective Proteasome Inhibitor LU-102 Synergizes with Bortezomib and Carfilzomib to Overcome Proteasome Inhibitor Resistance of Myeloma Cells. *Haematologica* **2015**, *100*, 1350−1360.
(51) Mirabella, A. C.; Pletnev, A. A.; Downey, S. L.; Florea, B. I.; Shabaneh, T. B.; Britton, M.; Verdoes, M.; Filippov, D. V.; Overkleeft, H. S.; Kisselev, A. F. Specific Cell-Permeable Inhibitor of Proteasome Trypsin-like Sites Selectively Sensitizes Myeloma Cells to Bortezomib and Carfilzomib. *Chem. Biol.* **2011**, *18*, 608−618.
(52) Weyburne, E. S.; Wilkins, O. M.; Sha, Z.; Williams, D. A.; Pletnev, A. A.; de Bruin, G.; Overkleeft, H. S.; Goldberg, A. L.; Cole, M. D.; Kisselev, A. F. Inhibition of the Proteasome B2 Site Sensitizes Triple-Negative Breast Cancer Cells to B5 Inhibitors and Suppresses Nrf1 Activation. *Cell Chem. Biol.* **2017**, *24*, 218−230.
(53) Rinnerthaler, G.; Gampenrieder, S. P.; Petzer, A.; Burgstaller, S.; Fuchs, D.; Rossmann, D.; Balic, M.; Egle, D.; Rumpold, H.; Singer, C. F.; Bartsch, R.; Petru, E.; Melchardt, T.; Ulmer, H.; Mlineritsch, B.; Greil, R. Ixazomib in Combination with Carboplatin in Pretreated Women with Advanced Triple-Negative Breast Cancer, a Phase I/II Trial of the AGMT (AGMT MBC-10 Trial). *BMC Canc.* **2018**, *18*, 1074.
(54) Auner, H. W.; Yong, K. L. More Convenient Proteasome Inhibition for Improved Outcomes. *Lancet Oncol.* **2018**, *19*, 856−858.
(55) Omura, S.; Fujimoto, T.; Otoguro, K.; Matsuzaki, K.; Moriguchi, R.; Tanaka, H.; Sasaki, Y. Structure of Lactacystin, a New Microbial Metabolite Which Induces Differentiation of Neuroblastoma Cells. *J. Antibiot.* **1991**, *44*, 113−116.
(56) Tomoda, H.; Omura, S. Lactacystin, a Proteasome Inhibitor: Discovery and Its Application in Cell Biology. *Yakugaku Zasshi* **2000**, *120*, 935−949.
(57) Fenteany, G.; Standaert, R.; Lane, W.; Choi, S.; Corey, E.; Schreiber, S. Inhibition of Proteasome Activities and Subunit-Specific Amino-Terminal Threonine Modification by Lactacystin. *Science* **1995**, *268*, 726−731.
(58) Gulder, T. A. M.; Moore, B. S. Salinosporamide Natural Products: Potent 20 S Proteasome Inhibitors as Promising Cancer Chemotherapeutics. *Angew. Chem., Int. Ed.* **2010**, *49*, 9346−9367.
(59) Groll, M.; Berkers, C. R.; Ploegh, H. L.; Ovaa, H. Crystal Structure of the Boronic Acid-Based Proteasome Inhibitor Bortezomib in Complex with the Yeast 20S Proteasome. *Structure* **2006**, *14*, 451−456.
(60) Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. Salinosporamide A: A Highly Cytotoxic Proteasome Inhibitor from a Novel Microbial Source, a Marine Bacterium of the New Genus Salinospora. *Angew. Chem., Int. Ed.* **2003**, *42*, 355−357.
(61) Groll, M.; Huber, R.; Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of -Lactone Ring Opening and a Mechanism for Irreversible Binding. *J. Am. Chem. Soc.* **2006**, *128*, 5136−5141.
(62) Groll, M.; Potts, B. C. Proteasome Structure, Function, and Lessons Learned from Beta-Lactone Inhibitors. *Curr. Top. Med. Chem.* **2011**, *11*, 2850−2878.
(63) Macherla, V. R.; Mitchell, S. S.; Manam, R. R.; Reed, K. A.; Chao, T.-H.; Nicholson, B.; Deyanat-yazdi, G.; Mai, B.; Jensen, P. R.; Fenical, W. F.; Neuteboom, S. T. C.; Lam, K. S.; Palladino, M. A.; Potts, B. C. M. Structure - Activity Relationship Studies of Salinosporamide A (NPI-0052), a Novel Marine Derived Proteasome Inhibitor. *J. Med. Chem.* **2005**, *48*, 3684−3687.
(64) ClinicalTrials.gov., 2020. U.S. National Library of Medicine, U.S. National Institutes of Health, U.S. Department of Health and Human Services, https://clinicaltrials.gov/ (accessed Jan 29, 2020).
(65) Manam, R. R.; McArthur, K. A.; Chao, T.-H.; Weiss, J.; Ali, J. A.; Palombella, V. J.; Groll, M.; Lloyd, G. K.; Palladino, M. A.; Neuteboom, S. T. C.; Macherla, V. R.; Potts, B. C. M. Leaving Groups Prolong the Duration of 20S Proteasome Inhibition and Enhance the Potency of Salinosporamides. *J. Med. Chem.* **2008**, *51*, 6711−6724.
(66) Groll, M.; Nguyen, H.; Vellalath, S.; Romo, D. (−)-Homosalinosporamide A and Its Mode of Proteasome Inhibition: An X-Ray Crystallographic Study. *Mar. Drugs* **2018**, *16*, 240−249.
(67) Huber, E. M.; Heinemeyer, W.; Li, X.; Arendt, C. S.; Hochstrasser, M.; Groll, M. A Unified Mechanism for Proteolysis and Autocatalytic Activation in the 20S Proteasome. *Nat. Commun.* **2016**, *7*, 10900.
(68) Groll, M.; Heinemeyer, W.; Jäger, S.; Ullrich, T.; Bochtler, M.; Wolf, D. H.; Huber, R. The Catalytic Sites of 20S Proteasomes and Their Role in Subunit Maturation: A Mutational and Crystallographic Study. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10976−10983.
(69) Arendt, C. S.; Hochstrasser, M. Eukaryotic 20S Proteasome Catalytic Subunit Propeptides Prevent Active Site Inactivation by N-Terminal Acetylation and Promote Particle Assembly. *EMBO J.* **1999**, *18*, 3575−3585.
(70) Dubiella, C.; Cui, H.; Gersch, M.; Brouwer, A. J.; Sieber, S. A.; Krüger, A.; Liskamp, R. M. J.; Groll, M. Selective Inhibition of the Immunoproteasome by Ligand-Induced Crosslinking of the Ac-tive Site. *Angew. Chem., Int. Ed.* **2014**, *53*, 11969−11973.
(71) Schrader, J.; Henneberg, F.; Mata, R. A.; Tittmann, K.; Schneider, T. R.; Stark, H.; Bourenkov, G.; Chari, A. The Inhibition Mechanism of Human 20S Proteasomes Enables Next-Generation Inhibitor Design. *Science* **2016**, *353*, 594−598.
(72) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(73) Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, *164*, 399−407.
(74) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(75) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(76) Schafmeister, C. E. A.; Ross, W. S.; Romanovski, V. *LEAP*; University of California: San Francisco, 1995.
(77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(78) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(79) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(80) Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3628−3631.
(81) Field, M. J.; Albe, M.; Bret, C. l.; Proust-De Martin, F.; Thomas, A. The Dynamo Library for Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.* **2000**, *21*, 1088−1100.
(82) Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

Hydrophilic Binding Sites of HIV-1 RT. *J. Phys. Chem. B* **2015**, *119*, 917−927.
(83) Field, M. J.; Bash, P. A.; Karplus, M. Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(84) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(85) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(86) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385−395.
(87) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(88) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(89) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes. *J. Phys. Chem. B* **2006**, *110*, 17663−17670.
(90) Chuang, Y.-Y.; Corchado, J. C.; Truhlar, D. G. Mapped Interpolation Scheme for Single-Point Energy Corrections in Reaction Rate Calculations and a Critical Evaluation of Dual-Level Reaction Path Dynamics Methods. *J. Phys. Chem. A* **1999**, *103*, 1140−1149.
(91) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V. Protein conformational landscapes and catalysis. Influence of active site conformations in the reaction catalyzed by L-lactate dehydrogenase. *ACS Catal.* **2015**, *5*, 1172−1185.
(92) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(93) Åberg, E.; Lund, B.; Pflug, A.; Gani, O. A. B. S. M.; Rothweiler, U.; De Oliveira, T. M.; Engh, R. A. Structural Origins of AGC Protein Kinase Inhibitor Selectivities: PKA as a Drug Discovery Tool. *Biol. Chem.* **2012**, *393*, 1121−1129.
(94) List, A.; Zeiler, E.; Gallastegui, N.; Rusch, M.; Hedberg, C.; Sieber, S. A.; Groll, M. Omuralide and Vibralactone: Differences in the Proteasome-β-Lactone-γ-Lactam Binding Scaffold Alter Target Preferences. *Angew. Chem., Int. Ed.* **2014**, *53*, 571−574.
(95) Borissenko, L.; Groll, M. 20S Proteasome and Its Inhibitors: Crystallographic Knowledge for Drug Development. *Chem. Rev.* **2007**, *107*, 687−717.
(96) Gräwert, M. A.; Groll, M. Exploiting Nature's Rich Source of Proteasome Inhibitors as Starting Points in Drug Development. *Chem. Commun.* **2012**, *48*, 1364−1378.
(97) Saha, A.; Oanca, G.; Mondal, D.; Warshel, A. Exploring the Proteolysis Mechanism of the Proteasomes. *J. Phys. Chem. B* **2020**, *124*, 5626−5635.
(98) Kim, D. H.; Park, J.-i.; Chung, S. J.; Park, J. D.; Park, N.-K.; Han, J. H. Cleavage of β-Lactone Ring by Serine Protease. Mechanistic Implications. *Bioorg. Med. Chem.* **2002**, *10*, 2553−2560.
(99) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motion. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14115−14120.
(100) Moliner, V. "Eppur si muove" (yet it moves). *Proc. Acad. Nat. Sci. U.S.A.* **2011**, *108*, 15013−15014.
(101) Krzemińska, A.; Moliner, V.; Świderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J. Am. Chem. Soc.* **2016**, *138*, 16283−16298.
(102) Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J. Protein flexibility and preorganization in the design of enzymes. The kemp elimination catalyzed by HG3.17. *ACS Catal.* **2015**, *5*, 2587−2595.
(103) Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J. Revealing the origin of the efficiency of the de novo designed Kemp eliminase HG-3.17 by comparison with the former developed HG-3. *Chem. —Eur. J.* **2017**, *23*, 7582−7589.
(104) Świderek, K.; Tuñón, I.; Williams, I. H.; Moliner, V. Insights on the origin of catalysis on glycine N-methyltransferase from computational modelling. *J. Am. Chem. Soc.* **2018**, *140*, 4327−4334.
(105) Martí, S.; Bastida, A.; Świderek, K. Theoretical Studies on Mechanism of Inactivation of Kanamycin A by 4'-O-Nucleotidyl-transferase. *Front. Chem.* **2019**, *6*, 660.
(106) Marques, A. J.; Palanimurugan, R.; Matias, A. C.; Ramos, P. C.; Dohmen, R. J. Catalytic mechanism and assembly of the proteasome. *Chem. Rev.* **2009**, *109*, 1509−1536.
(107) Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.; Chipot, C.; Kneller, D. W.; Kovalevsky, A.; Coates, L.; Golosov, A. A.; Dickson, C. J.; Velez-Vega, C.; Duca, J. S.; Vermaas, J. V.; Pang, Y. T.; Acharya, A.; Parks, J. M.; Smith, J. C.; Gumbart, J. C. Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. **2020**, bioRxiv 2020.09.07.28634. preprint.
(108) Lodola, A.; Mor, M.; Rivara, S.; Christov, C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. Identification of productive inhibitor binding orientation in fatty acid amide hydrolase (FAAH) by QM/MM mechanistic modelling. *Chem. Commun.* **2008**, 214−216.
(109) Lodola, A.; Mor, M.; Hermann, J. C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. QM/MM modelling of oleamidehydrolysis in fatty acid amide hydrolase (FAAH) reveals a new mechanism of nucleophile activation. *Chem. Commun.* **2005**, 4399−4401.

https://dx.doi.org/10.1021/acscatal.0c05313
ACS Catal. 2021, 11, 3575−3589

4.2.2. Rationalization of the homo-salinosporamide A (hSalA) reversible inhibition character

The great inhibition power of SalA reported in the literature[3,7,8] encouraged the exploration of new inhibitors in the same family. Homo-salinosporamide A (hSalA) is an analog of SalA whose difference relies on the C3-γ-lactam ring substitution. In the case of SalA, there is a chloro-ethyl side chain, while on hSalA there is a longer chain, a chloro-propyl, that according to the X-ray studies does not form the expected cyclic tetrahydropyran (THP) in the final product. Interestingly, both inhibitors show the same IC_{50} values.[5] Based on the results presented in the previous section (section 4.2.1), the inhibitory mechanism for hSalA in the β5 active site of the proteasome was explored. The results showed that the two initial steps for the inhibition mechanism are in very good agreement with the results obtained with SalA, and the rate-limiting step of the reaction process corresponds also to the first step, 20.3 kcal·mol^{-1}. Therefore, it can be assumed that the agreement on the IC_{50} values corresponds to both inhibitors having the same kinetic properties in terms of the rate liming step. Additionally, the results showed that the chloro-propyl side chain of hSalA has larger degrees of freedom, and it needs an additional intermediate rearrangement once the β-lactone ring is open. This makes the final cyclization step too energetically demanding to occur. Therefore, a new reaction step was necessary to explain the stability of the product that is detected in crystallographic studies. This was a new proposed step based on proton rearrangement required to neutralize the charges of the active site. This final step was explored for both inhibitors, and showed to be a competitive process in the case of SalA, where the final cyclization product is thermodynamically favored (-39.5 kcal·mol^{-1}), while in hSalA this is the only kinetically available process, leading to a stable neutral product (-8.5 kcal·mol^{-1}) with the final structure in agreement with the X-ray determined structure.[5]

The nature of the reversibility of hSalA can be caused by two possible reasons: the regress of the reaction process and dissociation of the inhibitor or the ester-linkage slow hydrolysis carried out by a water molecule. Therefore, the hydrolysis process was studied, showing that when the ester-linkage was attacked by a water molecule (E-I2 and E-PC$^{(OH)}$) the process

rendered a high energy barrier (33.6 and 67.7 kcal·mol^{-1}, respectively). This discards the possibility of hSalA acting as a slow substrate and suggests that this inhibitor can be classified as a covalent reversible inhibitor (as explained in section 1.2.1).

All nucleophilic attack processes explored in this section were analyzed in terms of the geometrical evolution of the obtained IRC paths at the M06-2X/AMBER level of theory. For this analysis, two typical angles used in organic chemistry for the rationalization of this type of reaction process were applied. The evolution of the Bürgi−Dunitz[9–11] (α_{BD}) and Flippin−Lodge[12] (α_{FL}) angles were monitored along the IRC paths. This analysis showed an optimal orientation in the case of the nucleophilic attack by $O^{\gamma Thr1}$ in the first inhibition step, with values ca. 100º for α_{BD} and ca. 5º for α_{FL} for both inhibitors, SalA and hSalA, respectively. Nevertheless, the results showed that the water nucleophilic attack trajectory for the explored hydrolysis paths was not favorable, as derived from the α_{FL} angle, which showed values further from the optimal perpendicular attack (ca. 10º in the case of the higher barrier). This structural analysis showed that the optimal attack position corresponds to the space occupied by the β-lactone O2 oxygen, which is responsible for the reversibility by regressing the reaction process and impeding access to the water molecule for ester linkage hydrolysis. The employed analysis has been proven to be useful in drawing key conclusions, i.e., giving a geometrical explanation for the size of the computed free energy barriers obtained, and showing that it can be further applied to other mechanisms involving a nucleophilic attack to a planar-carbonyl center.

A summary of the results is presented in Figure 4.3.

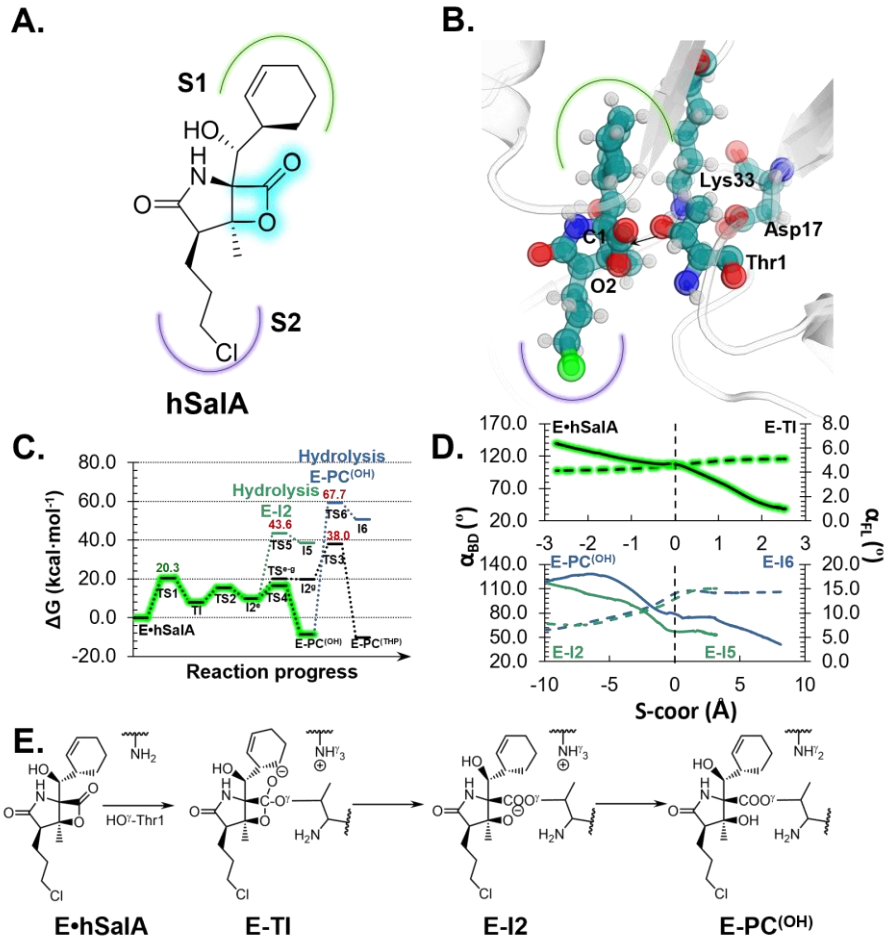


Figure 4.3. **A.** Schematic representation of hSalA. **B.** Representation of the E•hSalA complex. **C.** Free energy profiles for the inhibition and hydrolysis mechanisms computed at M06-2X:AM1/MM level of theory showing, highlighted in green, the favored inhibition mechanism. **D.** Evolution of Bürgi−Dunitz (α_{BD}) and Flippin−Lodge (α_{FL}) angles for the nucleophilic attack steps: Thr1 attack to the carbonyl carbon of the lactone ring (upper), water attack to the carbonyl carbon during ester bond hydrolysis in E−PC$^{(OH)}$ and E−I2 complexes (lower). **E.** Schematic representation of the proposed reaction mechanism followed by hSalA.

These results were presented in the following paper:

Serrano-Aparicio, N.; Moliner, V.; Świderek, K. On the Origin of the Different Reversible Characters of Salinosporamide A and Homosalinosporamide A in the Covalent Inhibition of the Human 20S Proteasome. *ACS Catal.* **2021**, 11806–11819. https://doi.org/10.1021/acscatal.1c02614.

ACS Catalysis

pubs.acs.org/acscatalysis Research Article

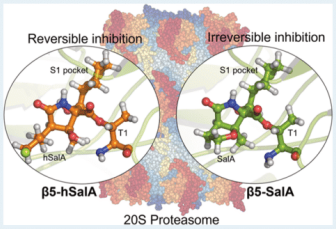
On the Origin of the Different Reversible Characters of Salinosporamide A and Homosalinosporamide A in the Covalent Inhibition of the *Human* 20S Proteasome

Natalia Serrano-Aparicio, Vicent Moliner,* and Katarzyna Świderek*

Cite This: *ACS Catal.* 2021, 11, 11806−11819 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Covalent inhibition of the 20S proteasome is one of the strategies used in the fight against cancer, Chagas' disease, malaria, tuberculosis, or neurodegenerative disorders. In this work, studies based on quantum mechanics/molecular mechanics (M06-2X/AMBER) molecular dynamics simulations reveal, in agreement with experiments, that homosalinosporamide A (hSalA), contrary to the natural analogue, salinosporamide A (SalA), inhibits the β5 subunit in a reversible manner. The computed free energy landscapes indicate that either SalA or hSalA blocks the active site of the enzyme by a three-step process that includes two common chemical transformations. The last step of the inhibition with SalA, which is an intramolecular cyclization step that defines its irreversible character, was found to be energetically prohibited in the case of hSalA. In contrast, an additional chemical step was found that ensures neutralization of the negative charge accumulated during the β-lactone ring opening, thus providing a stable product of inhibition that agrees with X-ray diffraction studies. The weaker stabilization of this final product explains its reversible character that was further confirmed by exploring possible paths of hydrolysis of the covalent adduct. Finally, the rate-limiting step in both inhibitors corresponds to the nucleophilic attack of Thr1 on C1 carbon of the β-lactone. The identical activation free energies computed for SalA and hSalA agree with the undistinguishable experimentally measured values of IC_{50}.



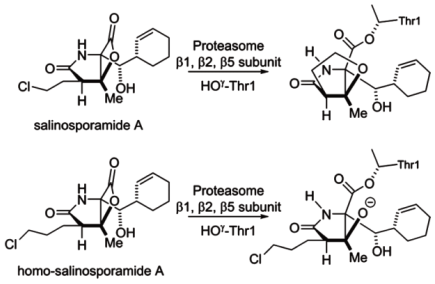
KEYWORDS: *20S proteasome, salinosporamide A, homo-salinosporamide A, QM/MM, inhibition mechanism*

The interest in covalent inhibitors (CIs) in drug discovery increases rapidly and is considered a viable option for searching new enzyme inhibitors.[1] Last year alone, a plethora of different variants of CIs against the main protease of Covid19 (M^{pro}) have been proposed, as a result of the global scientific effort put in the fight against the pandemic expansion.[2−8] CIs have been already used in the successful design of several drugs against cancer and gastrointestinal, central nervous system, cardiovascular, and inflammatory diseases,[9] including the group of well-known medications such as penicillin or aspirin. The choice of CIs as potential drug candidates is dictated by their advantages, despite that many of them present some disadvantages, as summarized elsewhere.[8−24] Generally, the main advantage from the formation of a covalent bond between an inhibitor and the target enzyme is the improvement in the stability of the formed complex, which can have therapeutic benefits such as longer duration of action that leads to less-frequent dosing.[25] Additionally, the use of CIs enhances the possibility of blocking the so-called "undruggable" shallow binding pockets and/or reduces the sensitiveness to resistant mutations.[26] Despite the indicated risk of non-specific reactivity with off-target proteins, DNA, or small molecules leading to toxicological events,[27−29] many of the approved CIs by the Food and Drug Administration have excellent safety records.[14] Interestingly, these drugs are often derived from natural products that are capable of forming a covalent bond with the target enzyme.[17,30]

An excellent example of this type of inhibitors is salinosporamide A (SalA) (also known as Marizomib or NPI-0052).[31] This γ-lactam-β-lactone compound (see Scheme 1), which was isolated from marine actinomycete *Salinispora tropica*, exhibits potent cancer cell cytotoxicity through inhibition of the 20S proteasome.[32,33] The 20S proteasome is the enzymatic core engine of the ubiquitin−proteasome system that regulates the protein degradation pathway in eukaryotes.[34,35] Because it is involved in many cellular processes, its dysregulation is related to many cellular dysfunctions and consequently diseases.[36,37] Hence, proteasome has become a good inhibition target that can lead to the recovery of the patients not only attacked by cancer[38,39] but also suffering from

Received: June 10, 2021
Revised: August 16, 2021
Published: September 8, 2021

Scheme 1. Structure of SalA and Its Derivative hSalA with Products of Inhibition as Observed in the Crystal Structures[30]



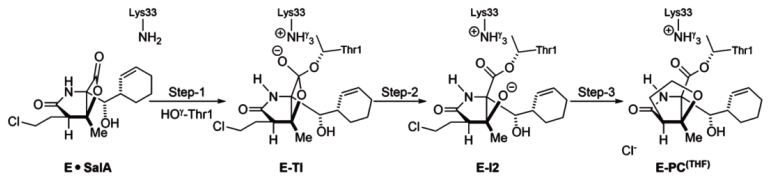
Chagas' disease,[40] malaria,[41−43] tuberculosis,[44] or neurodegenerative disorders such as Alzheimer's and Huntington's disease.[45,46] Proteasome adopts the structure of a barrel-like shape consisting of four heptameric rings formed by seven α or β subunits each.[47−50] The two inner rings contain six active subunits including β1, β2, and β5 types characterized by the presence of catalytical N-terminal threonine in their active sites.[51] Interestingly, although the architecture of the catalytically important residues is preserved in the three active sites in each type of β subunit, the character of the S1 pockets located in their vicinity is meaningfully different. The reason for the peculiarity of each S1 is related to the residue located at the bottom of the pocket. Thus, the presence of Met45, Asp53, or Arg45 in the S1 pocket determines their chymotrypsin (β5)-, trypsin (β2)-, and caspase-like (β1) characters.[52] The specific interactions that can be formed between the molecule and this pocket are the main reason for substrate selectivity demonstrated by each of the active sites. Nevertheless, structural features of SalA have allowed overcoming this obstacle that resulted in remarkable blocking potency of all three different subunits of the 20S proteasome, as demonstrated by several experimental studies.[53−55] These include crystallographic studies that revealed the existence of an irreversible covalent complex formed between each of the active sites of yeast 20S proteasome and SalA.[56] Due to these properties, SalA is currently undergoing clinical trials for the treatment of different cancer types. Thus, since 2018, it is in phase 3 of European Organisation for Research and Treatment of Cancer (EORTC) and Celgene and Canadian Cancer Trials Group (CCTG) studies on newly diagnosed glioblastoma and from 2020 in phase 2 of National Cancer Institute (NCI), National Institutes of Health Clinical Center (CC) studies on ependymoma.[57−60]

The success of discovering a non-peptide-like efficient irreversible inhibitor for the 20S proteasome encouraged researchers to test derivatives of SalA for their possible applications in the fight against pathogenic pathways. In this regard, an unnatural analogue of SalA, homosalinosporamide A (hSalA) (Scheme 1), has been revealed to possess very similar blocking properties of the β5 subunit.[30] Surprisingly, despite the only difference between SalA and hSalA molecules being the length of the chloro moiety, where hSalA has a chloro propyl chain instead of a chloro ethyl chain emerging from the lactam ring of SalA, the presence of the final product of the inhibition process in the latter, that is, tetrahydropyran (THP) ring, was not detected in crystallographic studies,[30] as illustrated in Scheme 1. This finding was rather unexpected, considering the comparable inhibition potency of both SalA and hSalA represented by IC_{50} values measured for the chymotrypsin-like active site. The similar experimentally obtained IC_{50} values (0.8 ± 0.08 and 0.7 ± 0.04 nM for SalA and hSalA, respectively) indicate that the process of inhibition should follow a similar pathway.[61] Nevertheless, it was found that hSalA, in contrast to SalA, blocks the proteasome by a covalent reversible mode of action.[30]

Therefore, herein, our main goal was to understand this enigmatic case by employing theoretical tools within the frame of QM/MM methods. These tools that allow for a detailed tracking of the course of chemical transformations occurring in the active site of the enzyme at the molecular level can be used to explain the origin of differences in the inhibition process observed for SalA and hSalA.

Recently, the mechanism of irreversible inhibition of the β5 subunit of 20S proteasome by SalA was revised in our group by computational studies using QM/MM schemes,[62] proving that this molecule is a suicidal inhibitor that exploits the enzyme's machinery, forms a covalent adduct with the protein, and becomes trapped in a stable, inactive state, thus operating exactly as an "mechanism-based inhibitor".[63] It was shown that this reaction occurs in three steps (as presented in Scheme 2) that involve the formation of a metastable tetrahedral intermediate (E−TI), the β-lactone ring opening (E−I2), and the generation of the final product complex involving a THF ring formation accompanied by chloride atom departure. The first step was found to be rate-limiting with the free energy barrier of 20.4 kcal mol^{-1} that is in excellent agreement with the experimentally determined rate constant (k_{cat} = 0.015 s^{-1} measured at 37 °C).[64] Additionally, it was shown that the formation of THF at the last step of the reaction pathway leads to a thermodynamically very stable product of inhibition. This result confirmed that SalA irreversibly blocks this active site and works according to the classical irreversible inhibition scheme, as shown in Table 1.

Scheme 2. Mechanism of Irreversible Inhibition of 20S Proteasome with SalA as Proposed in Our Previous Computational Studies[61]



https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

Table 1. Covalent Inhibition Strategy for SalA and hSalA

	class of inhibitor	scheme[a]
irreversible inhibition	classical	$E + I \overset{K_i}{\leftrightarrow} E{\bullet}I \xrightarrow{k_{inact}} E - I$
reversible inhibition	covalent reversible	$E + I \overset{K_i}{\leftrightarrow} E{\bullet}I \rightleftarrows E - I$
	slow substrate	$E + I \overset{K_i}{\leftrightarrow} E{\bullet}\vec{I}E - I \xrightarrow[slow]{} E + P$

[a]Here, K_i is the inhibition constant that measures binding affinity and k_{inact} is the inactivation rate constant that measures chemical reactivity.

Thus, assuming the correctness of the previously proposed inhibition mechanism by SalA, herein, the same scenario for the reaction mechanism involving hSalA is explored using an identical computational approach. The results of these studies should allow us to understand the unexpected different outcomes from crystallographic studies obtained for both γ-lactam-β-lactone compounds.

Inhibitors herein explored may have a different mechanism with respect to the substrate, other ligands, or even themselves due to their molecular architecture or/and the reaction environment provided by the protein. In this work, we will try to elucidate the role of each of these factors. Hence, in a broader perspective, the knowledge of this work may contribute not only to design efficient inhibitors but also to provide information that can be applied to develop specific biocatalysts.

Finally, taking into account that by definition, reversible inhibition of enzymes can be achieved according to different schemes (as shown in Table 1),[25] we also explored the most possible scenario for 20S proteasome inhibition with hSalA. This includes studies of the mechanism in which hSalA plays the role of a slow substrate, instead of an inhibitor. In such a case, the hydrolysis of the covalent adduct formed in the final product of inhibition is explored, to shed light on the existence of possible alternative reaction pathways.

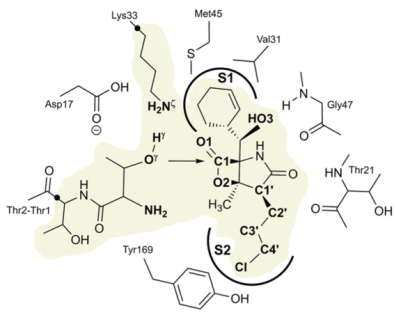
■ COMPUTATIONAL METHODS

System Setup. The model for studying the inhibitory pathway of hSalA in *Homo sapiens* 20S proteasome was built by overlapping the structure of the covalent adduct formed between SalA and the active site of the β5 subunit of the proteasome (that corresponds to the E−I2 intermediate structure from our previous studies,[61] as depicted in Scheme 2) with the 20S proteasome crystal structure from the *Saccharomyces cerevisiae* with hSalA (PDB ID: 6GOP).[30] To prepare such a constructed molecular system for simulations, missing force field parameters for hSalA and its covalent adduct were obtained using Antechamber software,[65] as provided in the Supporting Information. Then, PropKa program ver. 3.1[66,67] was used to predict the pK_a of all titratable residues. As expected, the obtained results revealed the same protonation state of titratable residues of the closest vicinity of the binding site of hSalA as for SalA at pH 7.0. The full list of protonation variants of titratable residues can be found elsewhere.[68]

Subsequently, the tLEAP module of the AmberTools[69] package was used to add all missing hydrogen atoms and to neutralize the system by the addition of 43 sodium counterions places at grid points where the energies are the lowest. An orthorhombic box of TIP3P[70] water molecules was built to solvate the full neutral system, reaching a size of 17.2 × 17.4 × 19.7 nm^3, resulting in a model with more than half a million atoms (534,380 atoms in total). The AMBER[71] and TIP3P force fields were employed to describe the protein and waters, respectively, and NAMD[72] software was used as the molecular dynamics (MD) engine. The system was minimized and gradually heated to 310 K with a 0.001 K temperature increment. Then, non-biased *NPT* equilibration for 100 ps was done followed by 100 ps of *NVT* equilibration and 5 ns of MD production, all simulations in the covalent adduct form. From the final structure, the reactant state of hSalA was prepared and equilibrated by a non-accelerated *NVT* MD simulation for 50 ns. In all MD simulations, the time step was equal to 1 fs. The temperature of the system was controlled using the Langevin thermostat,[73] and additionally, under NPT conditions, the Nosé−Hoover Langevin piston[74] pressure control was used. In all MD simulations, periodic boundary conditions were set. For reducing the simulation time, all atoms beyond 40 Å from the inhibitor were kept frozen in the production stages. A smooth switching function between 14.5 and 16 Å was used as a cutoff for non-bonding interactions. Finally, according to the evolution of the root-mean-square deviation during the simulation and the analysis of key distances, the systems could be considered equilibrated (see the Supporting Information).

QM/MM Methods. Exploration of the inhibitory mechanism started based on the representative structure for the active conformer of hSalA extracted from the MD trajectory. This conformer was chosen based on population analysis of key distances established between hSalA and the active site. Similar to previous work,[61] the size of the system was reduced for QM/MM calculations. Thus, the box of waters was cut into a water sphere with a radius of 60 Å with a geometrical center located on the Thr1 residue of the β5 active site. Therefore, the number of atoms in the system was reduced to 150,725. The protein, solvent water molecules, and counterions were described by AMBER and TIP3P force fields, as implemented in the fDynamo library.[75,76] Only the essential part for the study of the reactivity portion of a model was selected as the QM part, including in total 88 atoms. As illustrated in Scheme 3, this involves the inhibitor, terminus threonine (Thr1), part of Thr2, and the side chain of Lys33. Two link atoms[77] were inserted on the QM/MM boundaries, placed between the $C^{\alpha}-C^{\beta}$ atoms of Lys33 and $C^{\alpha}-C$ of Thr2. The QM subset of atoms was treated with the AM1[78] semi-empirical Hamiltonian and the M06-2X functional,[79] with

Scheme 3. Representation of the Active Site of the β5 Subunit of 20S Proteasome with hSalA[a]



[a]The green region contains the QM atoms. Black dots indicate the position of the quantum link atoms.

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

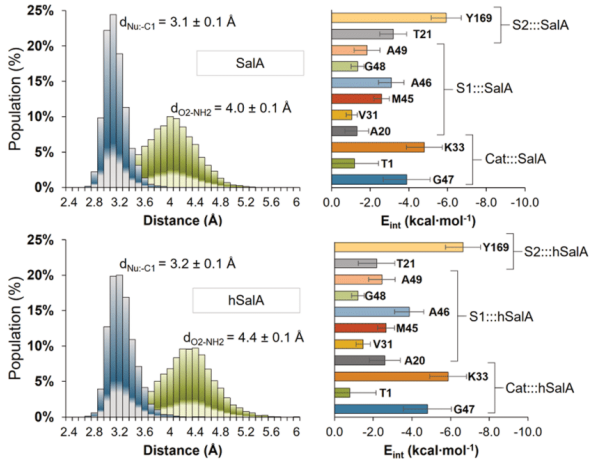


Figure 1. Key distances and interactions between SalA/hSalA and the active site of the β5 subunit from 20S proteasome from 50 ns of MM MD simulations and 100 ps of QM/MM MD simulations.

the standard 6-31+G(d,p) basis set. The cutoff scheme was conserved from the MD simulations for the non-bonding interactions, and the positions of atoms of residues beyond 20 Å from the inhibitor were frozen.

PES. The inhibition mechanism of hSalA was explored by selecting the appropriate distinguished reaction coordinates in each chemical step, by grid scanning; hence, the potential energy surfaces (PESs) were produced. Then, all stationary points found on these surfaces were optimized using a micro−macro[80] -iteration method at AM1/MM and M06-2X/MM levels. The matrix of second energy derivatives was computed to characterize these points. Moreover, to confirm that all transition state structures (TSs) optimized at the M06-2X/MM level connect with the expected minima, the IRC paths were traced down.

US. The umbrella sampling (US) approach[81] as implemented in fDYNAMO was used to produce a series of MD simulations where the distinguished reaction coordinate was constrained. The force constant used along these simulations was 2,500 kJ·mol^{-1} Å^{-2}, where an initial equilibration of 5 ps with 20 ps of production was done. Based on the obtained PES, in terms of potentials of mean force (PMF), the free energy surfaces (FESs) were computed initially at the AM1/MM level of theory. The weighted histogram analysis method[82] was used for the integration of the produced PMFs.

Spline Corrections. To correct the possible inaccuracy of the FESs computed at the AM1/MM level, a new energy function was employed in terms of interpolated corrections, to reduce the errors associated with the initial level of theory of our simulations.[83−85] The value of the corrected energy was computed according to eq 1

$$E = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\xi_1, \xi_2)] \quad (1)$$

where the final energy is obtained using a correction term evaluated from the single-point energy difference between high-level (HL) and low-level (LL) calculations of the QM subsystem. This is the two-dimensional spline correction, denoted by S. The selected LL method was the AM1 semi-empirical Hamiltonian, while the HL was the hybrid M06-2X functional with the standard 6-31+G(d,p) basis set. The Gaussian09[86] program in combination with fDynamo was used for these calculations. To produce the results presented in this work, 31,952 US windows and the respective single-point calculations at a high M06-2X/MM level of theory were computed.

RESULTS AND DISCUSSION

Reaction Mechanism. Based on the results of MD simulations, it was confirmed that the elongation of the spacer bearing the chlorine atom does not affect the inhibitor binding mode. The overall interactions created between hSalA and the active site of the β5 subunit are similar, as shown by computed values of interaction energy created between the protein and natural product (−54.1 ± 3.1 kcal mol^{-1}) and its analogue hSalA (−52.5 ± 3.1 kcal mol^{-1}). This result is in agreement with X-ray structures, revealing similar binding poses of both inhibitors. Thus, as shown in Figure 1, both inhibitors keep unchanged orientations in the E•I complex during the time of simulations that is characterized by a short and similar distance established between the nucleophile, O$^{\gamma Thr1}$ atom, and carbon C1 of the β-lactone ring (d_{Nu-C1} of 3.1 ± 0.1 and 3.2 ± 0.1 Å for SalA and hSalA, respectively) and between O2 atom of a lactone and N-terminus group of Thr1 (d_{NH2-O2} of 4.0 ± 0.1 and 4.4 ± 0.1 Å for SalA and hSalA, respectively). Both inhibitors create a similar pattern of electrostatic and van der Waals contacts within the binding pocket that is confirmed by values of interaction energies (E_{int}) computed for each inhibitor and amino acid residue of the catalytic site (Cat:::Inh) and the S1 (S1:::Inh) and S2 (S2:::Inh) pockets, as shown in Figure 1.

The mechanism of inhibition with hSalA was investigated applying a previously established computational procedure used in the study of the inhibition process with SalA.[61] The most favorable mechanism that involves the active participation of the deprotonated Lys33 residue, as demonstrated by previous computational studies published by Warshel and co-workers[87] and by us,[61,67] was herein considered. Hence, two FESs were

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

ACS Catalysis pubs.acs.org/acscatalysis Research Article

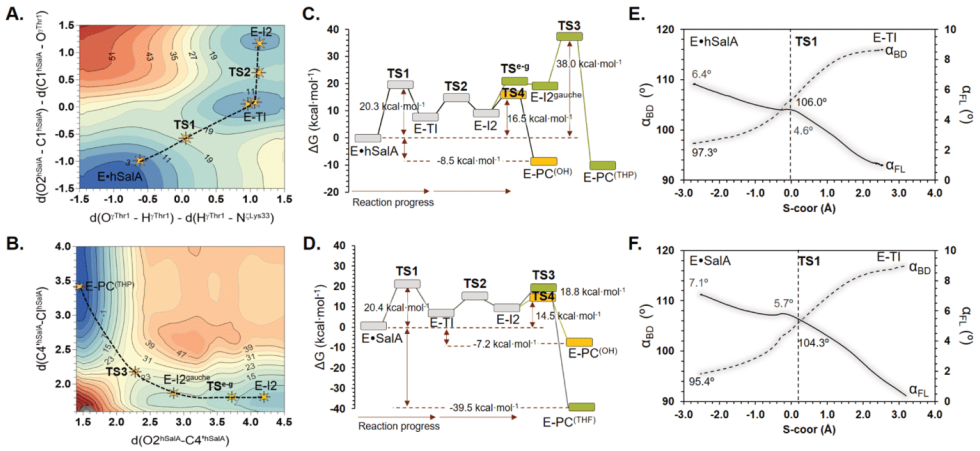


Figure 2. FESs computed at the M06-2X:AM1/MM level for the inhibition mechanism of the β5 subunit of 20S proteasome with hSalA: (A) formation of the tetrahedral intermediate E-TI and lactone ring opening E−I2 and (B) THP ring closing and E−PC$^{(THP)}$ adduct formation (on the bottom). Values of distances are given in Å, and isoenergetic lines are given in kcal mol^{-1}. The yellow stars indicate positions of the stationary structure localized at the M06-2X/AMBER level on the potential energy surface. Free energy profiles computed at the M06-2X:AM1/AMBER level for inhibition of the β5 subunit of 20S proteasome with (C) hSalA and (D) SalA. Data to generate gray and green paths on panel D are derived.[61] Evolution of Bürgi−Dunitz (α_{BD}) and Flippin−Lodge (α_{FL}) angles for nucleophilic attack of Thr1 onto carbonyl carbon of the lactone ring of (E) hSalA and (F) SalA along the first step of the inhibition reaction pathway.

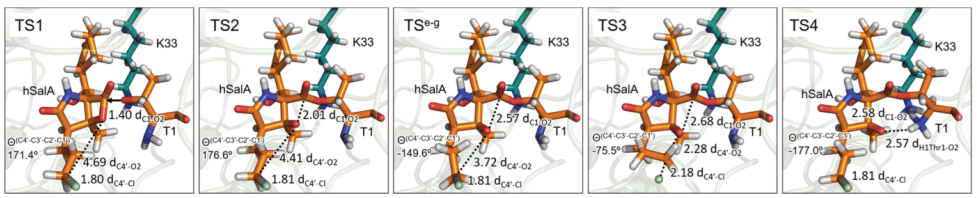


Figure 3. TS structures localized along the reaction pathway explored for inhibition of the β5 subunit of 20S proteasome with hSalA computed at the M06-2X/AMBER level of theory.

explored (as shown in Figure 2A,B), identifying a three-step mechanism. In the first FES, two antisymmetric combinations of distances were controlled, the first one corresponding to the transfer of the H$^{\gamma Thr1}$ proton from O$^{\gamma Thr1}$ to N$^{\zeta Lys33}$ and the second being related to the β-lactone ring breaking C1^{hSalA}−O2^{hSalA} and the nucleophilic attack C1^{hSalA}−O$^{\gamma Thr1}$. In the second FES distances, the formation of the O2^{hSalA}−C4$'^{hSalA}$ bond and the breaking of the C4$'^{hSalA}$−Cl^{hSalA} leading to the closing of the THP ring and departure of the chloride atom were controlled. Thus, it was confirmed that in the first step, the reaction proceeds in a stepwise manner via a metastable tetrahedral intermediate (E−TI) whose formation precedes the opening of the β-lactone ring. Subsequently, the β-lactone ring is opened, and the THP ring is formed, accompanied by the chloride atom departure. Interestingly, deep insights into the FES computed for the last step of inhibition with hSalA (shown in Figure 2B) revealed the existence of an additional event not observed in the case of SalA. This step that occurs before ring closure is related to the conformational change experienced by the chloro-*n*-propyl group of hSalA, in which the initial extended position of the ligand side chain found in E−I2 is reoriented toward the reactive position, that is, gauche conformer (E−I2^{gauche}).

All TS structures optimized at the M06-2X/AMBER level for each step of reaction (i.e., TS1, TS2, TS^{e-g}, and TS3 provided in Figure 3) confirmed the course of chemical transformations described by the FESs. Additionally, the position of the optimized TSs on the quadratic regions of the corrected M06-2X:AM1/MM FESs supports our predictions (Figure 2A,B).

Despite that the mechanism of inhibition is almost identical to the one previously computed for SalA (see Scheme 2), the obtained free energy profile, presented in Figure 2C,D, exposed meaningful differences in the last stage of the process that corresponds to the intramolecular cyclization step (TS3). While in the case of SalA, the free energy barrier that allows for THF formation is barely 18.8 kcal mol^{-1}, in the case of hSalA, the cyclization step is much more energetically demanding with a computed free energy barrier of 38.0 kcal mol^{-1}. This high value of $\Delta G^{\ddagger}$ with respect to the non-CI−enzyme complex (E•hSalA) is an important obstacle that precludes the system to reach a more stable covalent product, E−P$^{(THP)}$, and thus, it explains the absence of E−P$^{(THP)}$ in the crystal structure of the proteasome−

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

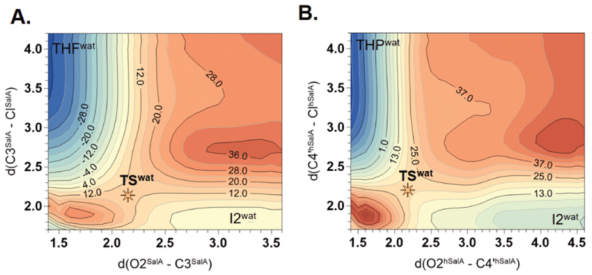


Figure 4. FESs for the intramolecular cyclization step for (A) SalA and (B) hSalA in the aqueous solution.

hSalA complex. Nevertheless, the obtained result raises more questions. For instance, (i) is the formation of the THP energetically less favorable than the formation of THF per se? (ii) how is it possible that the final E−I2 product complex that corresponds to the one observed in the crystal structure is not thermodynamically stable as deduced from FESs? and (iii) is the reversible character of the hSalA inhibitor controlled by reversible dissociation of the inhibitor from the active site or is it dictated by an additional step involving slow hydrolysis of the ester linkage? We will answer all these questions based on results from our computational studies.

Analysis of the Reaction Progress. As deduced from the perfect overlapping of the FESs computed for the first and second step of the reaction, both hSalA and SalA succumb to chemical transformations that must be mechanistically identical. This is confirmed by geometrical analysis of optimized structures of TS1 and TS2, as presented in Figure 3, and by the evolution of key distances along the reaction pathway (see the Supporting Information) that were compared and found to be identical to the results from our previous studies.[61] Thus, in both cases, Lys33 that serves as a proton acceptor in the first step activates O^{γ} of Thr1 that consequently attacks carbonyl carbon, C1 of lactone ring, and the covalent tetrahedral intermediate E−TI is produced. In the second step, the lactone ring is opening because of the C1−O2 bond breaking, and the E−I2 intermediate is formed.

The similar behavior of hSalA and SalA in the first, rate-limiting step of the reaction was also elucidated by additional analysis of the nucleophilic attack trajectory, as presented in Figure 2E,F. Attack of O^{γ} of Thr1 toward the electrophilic sp^2 carbonyl center was determined by following the evolution of the Bürgi−Dunitz[88−90] (α_{BD}) and Flippin−Lodge[91] (α_{FL}) angles. Interestingly, α_{BD}, defined as an angle between the nucleophile, carbonyl atom, and carbonyl oxygen ($\angle O^{\gamma Thr1}$−$C1^{SalA/hSalA}$−$O1^{SalA/hSalA}$) in the non-covalent E•I complex differs from the value of 105 ± 5° determined by Bürgi et al. for small-molecule substrates as the angle that ensures a reliable position for the nucleophile to attack.[88] In fact, in the Michaelis complex of 20S proteasome with a bound inhibitor, the value of α_{BD} was found to be 97.3 and 95.4° for hSalA and SalA, respectively. This outcome is very similar to the results of the analysis done by Radisky and Koshland[92] for the approach of the oxygen atom of catalytic serine toward reactive carbonyl in some proteases. In their case, the values of α_{BD} were found to be within the range of 88.3−90.8°. Nevertheless, as can be seen in Figure 2E,F, the relative position of carbonyl carbon *versus* the nucleophile changes once the reaction progresses. Although the α_{BD} is almost straight in the non-covalent E•I complex, an obtuse angle is achieved in the TS1, with values of 106 and 104° for hSalA and SalA, respectively. Thus, the rearrangement of the nucleophile and electrophile observed along reaction progress reflected by an increase in the α_{BD} value by 8.7 and 8.9° for hSalA and SalA, respectively, can contribute, to some extent, to rise of the free energy barrier computed for this step.

A similar analysis can be done for the α_{FL} angle that quantifies the deviation from the normal plane to the electrophile plane that is experienced by the nucleophile during the first step of the reaction. In the ideal case, this angle is expected to be 0°, but because the value of α_{FL} strongly depends on the size and electronic characteristic of the substituents of the carbonyl group, the expected range of α_{FL} cannot be defined. Thus, values of α_{FL} are always system-dependent. Our results show that in the studied active site, the values of α_{FL} in the non-covalent E•I complex are small, that is, 6.4 and 7.1° for hSalA and SalA, respectively, and decrease as the reaction progresses. In structures of TS1, the α_{FL} was found to reach 4.6 and 5.7° for hSalA and SalA, respectively.

As mentioned above, the main difference in the free energy profile was found in the third step that corresponds to the intramolecular alkylation of the alkoxide onto the chloro moiety side chain. The proposed hypothesis explaining the absence of the THP in the crystal structure suggests that a longer chloro-*n*-propyl group of hSalA adopts an extended conformation in the active site of proteasome that can prevent the ring closure.[30] However, in the same studies, it was shown that this group is highly flexible, which was confirmed by experimentally measured high values of B-factor for this fragment. Therefore, it is expected that rotation from an extended to a closed conformer should not be expensive. Hence, we will try to explain this phenomenon.

As determined by the computed FES (Figure 2B), the third step of the reaction follows the nucleophilic substitution via a S_N2 mechanism that is accompanied by chloride ion departure. The origin of a very high barrier of 38 kcal mol^{-1} computed for THP formation in the case of hSalA, in comparison to 18.8 kcal mol^{-1} obtained for the same step and resulted in THF ring formation for SalA, is not so obvious. In fact, we have double-checked this result by exploring a new FES exceptionally focusing on the conformational change process. To obtain this surface (provided in the Supporting Information), the C4′−O2 distance was controlled together with the dihedral angle formed in C4′−C3′−C2′−C1′ that defines extended (ca. −175°) and gauche (ca. −100°) conformers. Our results confirmed that the originally chosen coordinate (C4′−O2 distance) was sufficient

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

to correctly describe the process of the conformational evolvement of the chloro moiety side chain and, hence, that the originally computed high barrier of 38 kcal mol^{-1} is correct.

Additionally, we have decided to explore the THP and THF ring formation steps using simplified models in the aqueous phase, to exclude the possibility that the difference in the computed free energy barriers for this step of the reaction taking place in the proteasome is related to the specific environment created by the active site of the protein. Structures and specific computational procedures used for this test are provided in the Supporting Information. Two FESs for the cyclization step in SalA and hSalA in the aqueous environment were explored controlling the same distances as in the enzymatic reaction, as shown in Figure 4. The obtained free energy barriers present a very similar trend, as shown in Table 2, as those computed in the presence of the enzyme. Thus, the energy barriers for the THF formation in SalA are lower than those required for the THP formation in hSalA, either in the enzyme or in water. Interestingly, it can be noticed that the presence of the protein slightly facilitates the formation of THF, decreasing the resulting barrier by 3.4 kcal mol^{-1}. This reduction in the barrier can be simply explained based on the improved stabilization of the final product in the enzyme, as formulated in Hammond's postulate.[93]

Table 2. Free Energy Computed for THF and THP Formation (PC) in the Aqueous Solution and the Active Site of the 20S Proteasome[a]

	SalA		hSalA	
	water	enzyme	water	enzyme
E−I2^{extended}			−5.2	−9.7
E−I2/E−I2^{gauche}	0.0	0.0	0.0	0.0
TS	13.1	9.7	16.8	18.3
PC	−45.3	−48.6	−37.2	−30.0

[a]Values are given in kcal mol^{-1}.

Our results indicated that to create THP in hSalA, 28.0 and 22.0 kcal mol^{-1} are required to cross the energetic barrier in the enzyme and water, respectively. In this case, the barriers depend on both the cost of chemical transformation and the conformational change that is experienced by the chloro moiety side chain that is necessary to achieve the reactive conformation of carbon C4. As expected, the estimated cost of the conformational change of the chloro-*n*-propyl from the extended to gauche orientation is lower in the aqueous solution (5.2 kcal mol^{-1}) than in the enzyme (9.7 kcal mol^{-1}). This difference may be related to the degree of flexibility of the environment in which the reaction takes place. Because the plasticity of the enzyme is limited, compared to the mobility of the molecules in the aqueous solution, it is, therefore, more expensive to change the orientation of the chloro moiety from extended to gauche in the former. Moreover, it was found that the gauche conformer is a metastable intermediate in the enzyme, which confirms the prediction from experimental studies that assumes a short half-time for this inhibitor side chain orientation.[30]

The same trends in free energy barriers preserved under different reaction conditions herein explored indicate that the formation of THP is always energetically more expensive than the formation of a smaller THF ring. The obtained difference in energy would not be so dramatic if not for the fact that the intermediate E−I2 that is formed before the cyclization step is metastable and consequently increases the final barrier by 10 kcal mol^{-1}. Thus, it can be concluded that the reversible character of hSalA in the inhibition process of 20S proteasome is achieved by combining the structural features of the molecule itself and the stabilization of the product of the preceding reaction step achieved in the active site. Hence, based on the obtained results, it is confirmed that tuning of the effectiveness and reversibility of the inhibitors for 20S proteasome is experimentally very challenging because it not only involves knowledge of the chemical features of the used molecule but also requires detailed information about the reaction path that can be only delivered by theoretical studies.

Stabilization of the Final Inhibition Product. Interestingly, the experimental outcome indicating[30] that the process of *β*5 subunit inhibition with hSalA finishes at E−I2 adduct formation (as shown in Scheme 1) is not supported by the values of computed free energies. As deduced from this study, 20S proteasome with bound hSalA in the E−I2 state is 10 kcal mol^{-1} less stable than the non-covalent enzyme−inhibitor complex (E•hSalA). Thus, the occurrence of an additional process is expected, that is, a chemical step which could ensure the stabilization of the final product. Considering that experimental evidence is based on the result of crystallographic studies that employ the traditional X-ray diffraction technique, we can assume that this stabilization could be achieved by the reorganization of hydrogens within the active site. This possible rearrangement of hydrogens must be explored computationally since it is impossible to determine the positions of hydrogen atoms based on the crystal structure of the 20S proteasome solved with the resolution of 2.90 Å. Therefore, it was proposed that the proton that originally was transferred from Thr1 to Lys33 in the first step of the reaction will now be shifted toward the negatively charged O2 atom of the open lactone ring. This

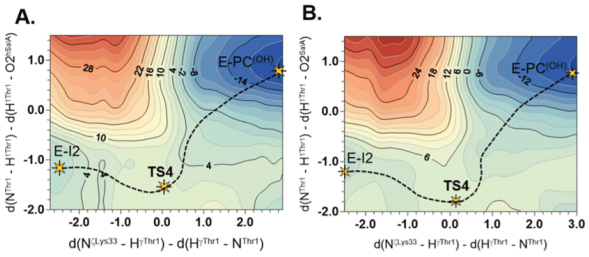


Figure 5. FESs for the additional double-proton transfer step explored for (A) hSalA and (B) SalA.

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

proposal was inspired by results from our previous studies where an identical stable intermediate in the SalA-assisted mechanism was observed.[61] Anyway, the direct transfer of the proton from Lys33 to hSalA cannot be achieved because the distance between the donor N^{z} of Lys33 and the O2 acceptor of the proton is too long (d_{Nz-O2} = 5.56 Å). Thus, the existence of an alternative pathway was considered. Hence, the participation of the amino group (NH_2 group) of Thr1 in the proton transfer was proposed. In other words, the NH_2 group was assumed to play the role of a proton shuttle. To investigate this possibility, a FES was explored where the transfer of two protons was controlled separately, that is, transfer of the $H^{\gamma Thr1}$ atom from N^{zLys33} to the NH_2 group of Thr1 and transfer of one proton from the NH_2 group ($H^{NH2-Thr1}$) to the oxygen of the opened lactone ring (O2), as shown in Figure 5A. Results show that the process in which the originally negatively charged oxygen is neutralized by forming a new product, E−$PC^{(OH)}$ complex, is achieved in a concerted manner, which means that the amino group simultaneously donates and accepts a proton. The existence of this step was confirmed by the localization of the TS, TS4, at the M06-2X/MM level, whose structure is shown in Figure 3. Finally, the computed free energy barrier for double-proton transfer was found to be much lower (16.5 kcal mol^{-1}) than the one obtained for the THP formation step (38.0 kcal mol^{-1}). However, the final covalent products of the proton transfer E−$PC^{(OH)}$ and cyclization step E−$PC^{(THP)}$ were equally stable, as presented in the free energy profile in Figure 2C. As proven by computational studies, the reason that the E−$PC^{(OH)}$ covalent adduct between proteasome and hSalA was observed in the crystal structure must originate from the existence of the additional proton transfer step. This computational result was further confirmed by a perfect overlap between the experimentally determined structure and the M06-2X/MM product structure delivered by this work. Both structures corresponding to the final inhibition product of the chymotrypsin-like active site by hSalA are shown in Figure 6.

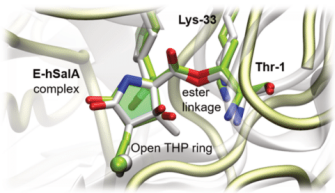


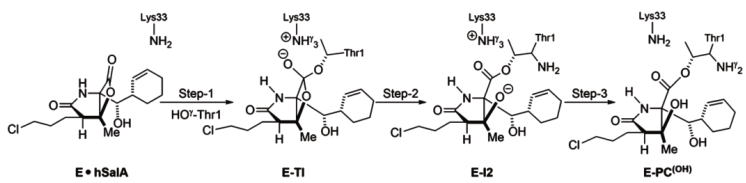
Figure 6. Overlay of the crystal structure of the final product of inhibition of the β5 subunit of yeast 20S proteasome (in gray) by hSalA with a human variant delivered by the M06-2X/MM optimization E−$PC^{(OH)}$ (in green).

The relatively low energy barrier computed for the alternative to the cyclization step reaction pathway raises additional concerns about the possibility that the same process can compete with irreversible product formation in the case of SalA. Therefore, this double-proton transfer step was explored once again but now starting from the E−$I2^{SalA}$ adduct. The computed FES for this step is provided in Figure 5B. The obtained free energy barrier is slightly lower (14.5 kcal mol^{-1}) than the one that leads to the final E−$PC^{(THF)}$ product (18.8 kcal mol^{-1}), as shown in Figure 2D, but the difference is sufficient to kinetically disfavor the formation of the THF ring. However, the formed E−$PC^{(OH)}$ product in the case of SalA is much less stable (by 32 kcal mol^{-1}) than the obtained E−$PC^{(THF)}$ complex after the cyclization step. Thus, it can be concluded that the formation of the irreversible product in the case of SalA is clearly thermodynamically favored, rather than due to kinetic arguments. Moreover, based on the obtained free energy profile, it can be hypothesized that in the case of SalA, the formation of the final E−$PC^{(THF)}$ complex identified in the crystal structure competes with the formation of the less-stable E−$PC^{(OH)}$ product. The optimized and characterized structure of TS4 for SalA is provided in the Supporting Information.

Finally, the existence of the additional step in the inhibition mechanism with hSalA, as summarized in Scheme 4, does not influence the overall kinetics of the inhibition process with hSalA in comparison to that with SalA. The rate-limiting step remains unchanged, and it is the one that is reflected in the measurements of IC_{50}. Thus, similar interactions formed between hSalA and the active site to those observed for SalA and the identical free energy barrier computed for the rate-limiting step (20.3 and 20.4 kcal mol^{-1} for hSalA and SalA, respectively)[61] explain why measurements of IC_{50} were not able to distinguish the reversible character of hSalA from its irreversible competitor, that is, SalA.

Mechanism of Ester Linkage Hydrolysis. The proposed mechanism of reversible inhibition opens a new question related to the hydrolysis of an ester bond formed between Thr1 and the molecule of the blocker. Being considered a reversible inhibitor, hSalA can be recovered from the active site according to two different mechanisms, that is, via a reversible process to covalent product formation or due to the hydrolysis of the ester bond. According to the computed free energy profile for the reversible process to covalent product formation, hSalA needs to pass chemical transformations to dissociate from the active site that require reaching a high barrier of 28.8 kcal mol^{-1}. Nevertheless, this is a significantly energetically cheaper process than the recovery of SalA, the irreversible inhibitor, that requires surmounting the energetical barrier of ca. 60 kcal mol^{-1} to cleave the ester linkage in the backward process. The high barrier for the deacylation step for SalA (without the participation of a

Scheme 4. Full Mechanism of the Reversible Inhibition of the β-Subunit of 20S Proteasome with hSalA Revealed by Computational Studies

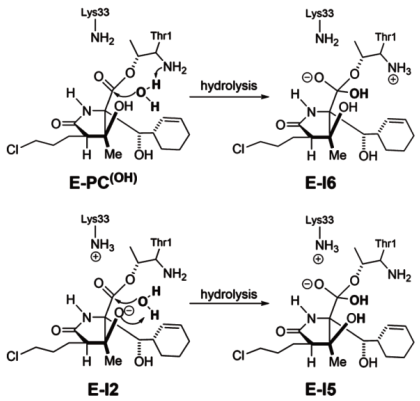


https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

water molecule) is dictated by the formation of a much more stable final inhibition product involving THF ring formation than the product of inhibition with hSalA. Thus, the value of a free energy barrier for hSalA deduced from computational studies permits very slow recuperation of enzyme activity.

Alternatively, the catalytic function of the proteasome can be regained by employing its deacylation step involving the participation of a water molecule that is used by the enzyme in the recovery process during proteolysis. In such a case, the water molecule is expected to attack the trigonal planar carbonyl carbon atom of the ester, which is formed in two occasions during the reaction progress, in the intermediate E−I2 and in the product E−PC$^{(OH)}$ complex. Hence, both possible hydrolysis mechanisms, as shown in Scheme 5, were explored. Although an

Scheme 5. Mechanism of the First Step of Ester Linkage Hydrolysis in the E−PC$^{(OH)}$ and E−I2 Complex



attack of the water molecule can occur from the Re or Si face, the latter orientation was never considered. This choice was dictated by a detailed geometrical analysis of the structure of the active site combined with the results from MD simulations. Therefore, the possible role of Lys33 as an activator of the water was immediately excluded based on its position in the binding cavity. As illustrated in the Supporting Information, access to this residue is hindered by its localization at the bottom of the active site. Moreover, the only entrance channel that could allow water to attack from the Si face of carbonyl carbon through the S1 pocket, a highly hydrophobic region, is entirely occupied by the cyclohexenyl group of the inhibitor. Hence, it seems unlikely that water would cross through this region. This analysis is supported by the computational results. The presence of the water molecule was never registered in this surrounding of the active site during MD simulations, which is monitored by the analysis of established H-bond contacts involving Lys33. A list of these contacts is provided in the Supporting Information. Thus, only the Re position of water was chosen for further calculations.

Assuming this orientation of the water molecule, relative to the carbonyl carbon C1, the only possible candidate found for its activation in the E−PC$^{(OH)}$ complex was the NH_2 group of terminus Thr1. Nevertheless, the employment of this group resulted in a highly unfavorable process with an estimated barrier of 67.7 kcal mol^{-1}, according to the computed FES (Figure 7A). This FES was explored by transferring the H1^{wat} from the water molecule to N^{Thr1} of the NH_2 group of terminus Thr1 and by controlling the distance between O^{wat} and C1, defining nucleophilic attack of water, to the carbonyl carbon of hSalA. In fact, the existence of an unfavorable process was already qualitatively predicted in our previous work[61] based on electrostatic effect analysis,[94−100] where it was shown that the highly positive electrostatic potential generated by the protein on the N^{Thr1} atom should disfavor accumulation of the positive charge on this group. Hence, as expected, the metastable product of the water attack (E−I6) was found to be 59.1 kcal mol^{-1} higher in energy than the starting E−PC$^{(OH)}$ complex. Thus, at this stage, it can be concluded that hydrolysis in the inhibition product complex cannot take place, and unbinding of the hSalA from the active site can occur only according to the backward transformation toward the formation of E−I2, as shown in Figure 7C. Nevertheless, the ester bond can be also hydrolyzed in the intermediate state (E−I2), and alternatively, in this case, the hydrolysis can be initiated by a negatively charged alcoholate, that is, O2 oxygen of the open lactone ring that can serve as a proton acceptor. Thus, a new FES for this mechanism (Figure 7B) was explored where the antisymmetric combination for the H1^{wat} transfer from O^{wat} of water to O2^{hSalA} of the hSalA was controlled together with the nucleophilic attack distance, O^{wat}−C1^{hSalA}, provided as an output a lower free energy barrier than that computed for product hydrolysis. Nevertheless, the free energy barrier of 33.6 kcal mol^{-1} (see Figure 7D) is still too high for this process to compete kinetically with the reversible transformation toward E•hSalA. Thus, based on the obtained results, it can be stated that hSalA behaves as a covalent reversible inhibitor and not as a slow substrate.

The computed high free energy barrier for ester linkage in the hydrolysis process can be somehow surprising at first because, as mentioned before, the attack of the water is an indispensable step in the deacylation process that occurs in the primary reaction catalyzed by the proteasome. However, in this case, the architecture of the active site of the proteasome only permits the attack of the water from the side from which the leaving group departs. This is an important obstacle that can prevent hydrolysis of the enzyme complexed with γ-lactam-β-lactone because in this case, the "leaving group" is still attached to the inhibitor, regardless of the fact that the C1−O2 bond is already cleaved. Thus, it is believed that the presence of the alcoholate or hydroxyl group is responsible for the displacement of the nucleophilic water from its ideal position for the attack.[30] This is confirmed by analysis of the nucleophilic attack trajectory. As shown in Figure 7E,F, the original position of the water molecule is far from favorable, as indicated by values of α_{BD} that are around 60° in both intermediate E−I2 and product E−PC$^{(OH)}$ complexes. On the contrary, the same α_{BD} angle of 103.9° computed for O2 toward the carbonyl carbon C1 process is much more reactive. In fact, this ideal orientation for the attack position is reflected by the low value of the free energy barrier (5.3 kcal mol^{-1}) computed for the transformation of E−I2 to E−I1. The same behavior of the water molecule was observed during the hydrolysis of the inhibition product with SalA. An unfavorable α_{BD} of 77.5° indicates that closure of the THF ring does not facilitate access of water to the carbonyl carbon. This observation agrees with the previously formulated hypothesis based on experimental evidence[53,63] and is reflected by a high free energy barrier (53.4 kcal mol^{-1}), as shown in the Supporting Information. The high barrier for hydrolysis of the inhibition product with SalA together with the high barrier for the reverse process of the inhibition reflects experimental findings that show

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

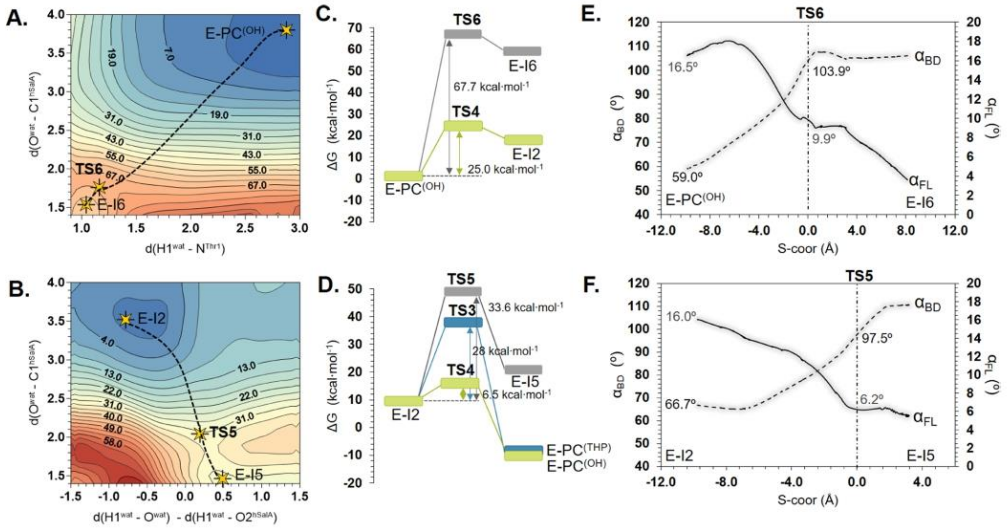


Figure 7. (A) FESs computed for the first step of hydrolysis in (A) E−PC$^{(OH)}$ and (B) E−I2 complexes with the provided position of localized stationary structures at the M06-2X/MM level (yellow stars). Values of the axes are given in Å, and isoenergic lines are given in kcal mol^{-1}. Free energy profiles for the first step of hydrolysis initiated in (C) E−PC$^{(OH)}$ and (D) E−I2 complexes including competitive processes. Evolution of Bürgi−Dunitz (α_{BD}) and Flippin−Lodge (α_{FL}) angles for nucleophilic attack of water onto carbonyl carbon of the lactone ring of hSalA along the first step of the ester bond hydrolysis in (E) E−PC$^{(OH)}$ and (F) E−I2 complexes.

a much longer duration of proteasome inhibition with this compound (no recovery of activity after 24 h of dialysis) than with similar β-lactone-γ-lactam-bearing substituents without a good leaving group (partial recovery after 5 h and complete recovery after 12 h of dialysis) that can be assumed to work according to the mechanism determined herein for hSalA.[63]

■ CONCLUSIONS

This work describes the mechanism of 20S proteasome inhibition with hSalA using a theoretical approach based on multiscale QM/MM methods. The results revealed, in agreement with experiments, that hSalA reversibly inhibits the β5 subunit, in contrast to its natural analogue, SalA, that behaves as an irreversible inhibitor. The most probable reaction mechanism based on the computed free energy profiles indicates that hSalA blocks the active site of the enzyme by a process that includes three chemical transformations. The first two steps of the reaction were found to be identical to those described previously for SalA,[61] as shown in Schemes 2 and 4. The last step, however, corresponds to the double-proton transfer from Lys33 to O2 with the participation of the NH_2 group of Thr1 that plays the role of a proton shuttle. This transfer that is for the first time proposed ensures the stabilization of the final product of inhibition by neutralizing the negative charge accumulated on O2 of the open lactone ring, as presented in Scheme 4. The low barrier of 18.2 kcal mol^{-1} computed for this process does not affect the overall rate-limiting step that in both SalA and hSalA corresponds to the nucleophilic attack of Thr1 on C1 carbon of the inhibitors, thus explaining the identical values of IC_{50} measured experimentally. The weaker stabilization of the final E−PC$^{(OH)}$ ensures the reversible character of the hSalA, which was further confirmed by the studies of possible paths of hydrolysis of the covalent adduct formed between the active site and inhibitor. It was demonstrated that hSalA works according to the covalent reversible scheme due to the presence of the alcoholate or hydroxyl group, and it cannot serve as a slow substrate. Hence, the activity of the enzyme can be regained only in the reverse process to the inhibition path. The alternative reaction pathway for the 20S proteasome inhibition with hSalA that imitates the last step of the inhibition process observed for SalA (i.e., THF formation) was also considered. However, the possibility of generation of an alternative inhibition product (THP) was excluded based on a high energy barrier of 38.0 kcal mol^{-1} computed for the intramolecular cyclization step. This high value, with respect to the non-CI−enzyme complex (E•hSalA), is an important obstacle that precludes the system to reach a more stable covalent product, E−P$^{(THP)}$, and thus, it explains the absence of E−P$^{(THP)}$ in the crystal structure of the proteasome−hSalA complex. It was found that this high barrier consists of three energetically crucial stages, that is, formation of the metastable intermediate E−I2, the conformational change required to obtain the reactive position of atoms involved in the chemical transformation (C4′ and O2), and finally the S_N2 reaction.

The results of this research confirm the significant role of computational studies in providing information that cannot be easily delivered by experimental measurements. Thus, computational techniques can be considered a powerful tool to assist the design of inhibitors and, in a broader perspective, new biocatalysts. As revealed, the tuning of the effectiveness and reversibility of the inhibitors for 20S proteasome must be experimentally very challenging because the process of inhibitor design, in this case, not only involves knowledge of the chemical features of the used molecule but also requires detailed

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

information about the reaction path that can be only described at the atomistic level with theoretical studies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.1c02614.

MM parameters for hSalA; details of the MD simulations; definition of analyzed Bürgi−Dunitz (α_{BD}) and Flippin−Lodge (α_{FL}) angles; FESs; evolution of key distances along the reaction paths; averaged ChelpG charges on key atoms; results of study of the cyclization step for SalA and hSalA in aqueous solution; results of study of the double-proton transfer step on SalA; results of study of the ester bond hydrolysis mechanism of hSalA; results of study of the ester bond hydrolysis mechanism of SalA; and cartesian coordinates of transition structures optimized at the M06-2X/MM level (PDF)

hSalA inhibition mechanism from structures obtained along M06-2X/MM IRC paths (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Vicent Moliner − *Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain*; orcid.org/0000-0002-3665-3391; Email: moliner@uji.es

Katarzyna Świderek − *Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain*; orcid.org/0000-0002-7528-1551; Email: swiderek@uji.es

Author

Natalia Serrano-Aparicio − *Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain*; orcid.org/0000-0002-4429-8940

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.1c02614

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank the Spanish Ministerio de Ciencia e Innovación (grant PGC2018-094852-B-C21 and PID2019-107098RJ-I00), the Generalitat Valenciana (grant AICO/2019/195 and SEJI/2020/007), and Universitat Jaume I (grant UJI-B2020-03 and UJI-A2019-04). N.S.A. thanks the MINECO for the doctoral FPI grant (BES-2016-078029). Authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I. The authors thankfully acknowledge the computer resources at Pirineus and the technical support provided by the Consorci de Serveis Universitaris de Catalunya (RES-QSB-2020-3-0012).

■ REFERENCES

(1) Strelow, J. M. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS Discov. Adv. Sci. Drug Discov.* **2016**, *22*, 3−20.

(2) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289−293.

(3) Erlanson, D. A. Many Small Steps towards a COVID-19 Drug. *Nat. Commun.* **2020**, *11*, 5048−5052.

(4) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331−1335.

(5) Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y.; Lockner, J. W.; Lou, J.; Marra, M. T.; Mitchell, L. J.; Murray, B. W.; Nieman, J. A.; Noell, S.; Planken, S. P.; Rowe, T.; Ryan, K.; Smith, G. J.; Solowiej, J. E.; Steppan, C. M.; Taggart, B. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63*, 12725−12747.

(6) Weglarz-Tomczak, E.; Tomczak, J. M.; Talma, M.; Burda-Grabowska, M.; Giurg, M.; Brul, S. Identification of Ebselen and Its Analogues as Potent Covalent Inhibitors of Papain-like Protease from SARS-CoV-2. *Sci. Rep.* **2021**, *11*, 3640.

(7) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Świderek, K.; Moliner, V. Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021**, *12*, 1433−1444.

(8) Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α-Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochemistry* **2020**, *59*, 4601−4608.

(9) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discovery* **2011**, *10*, 307−317.

(10) De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A. K.; Moitessier, N. Covalent Inhibitors Design and Discovery. *Eur. J. Med. Chem.* **2017**, *138*, 96−114.

(11) Krantz, A. A classification of enzyme inhibitors. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 1327−1334.

(12) Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009**, *52*, 225−233.

(13) Potashman, M. H.; Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *J. Med. Chem.* **2009**, *52*, 1231−1246.

(14) Johnson, D. S.; Weerapana, E.; Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem.* **2010**, *2*, 949−964.

(15) Kalgutkar, A. S.; Dalvie, D. K. Drug Discovery for a New Generation of Covalent Drugs. *Expet Opin. Drug Discov.* **2012**, *7*, 561−581.

(16) Niphakis, M. J.; Cravatt, B. F. Enzyme Inhibitor Discovery by Activity-Based Protein Profiling. *Annu. Rev. Biochem.* **2014**, *83*, 341−377.

(17) Mah, R.; Thomas, J. R.; Shafer, C. M. Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 33−39.

(18) Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discov. Today* **2015**, *20*, 1061−1073.

(19) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chem. Int. Ed.* **2016**, *55*, 13408−13421.

(20) González-Bello, C. Designing Irreversible Inhibitors-Worth the Effort? *ChemMedChem* **2016**, *11*, 22−30.

(21) Bandyopadhyay, A.; Gao, J. Targeting Biomolecules with Reversible Covalent Chemistry. *Curr. Opin. Chem. Biol.* **2016**, *34*, 110−116.

(22) Lu, S.; Zhang, J. Designed Covalent Allosteric Modulators: An Emerging Paradigm in Drug Discovery. *Drug Discov. Today* **2017**, *22*, 447−453.
(23) Jackson, P. A.; Widen, J. C.; Harki, D. A.; Brummond, K. M. Covalent Modifiers: A Chemical Perspective on the Reactivity of α,β-Unsaturated Carbonyls with Thiols via Hetero-Michael Addition Reactions. *J. Med. Chem.* **2017**, *60*, 839−885.
(24) Visscher, M.; Arkin, M. R.; Dansen, T. B. Covalent Targeting of Acquired Cysteines in Cancer. *Curr. Opin. Chem. Biol.* **2016**, *30*, 61−67.
(25) Cully, M. Novel chemistry for covalent inhibitors. *Nat. Rev. Drug Discovery* **2020**, *19*, 754.
(26) Tuley, A.; Fast, W. The Taxonomy of Covalent Inhibitors. *Biochemistry* **2018**, *57*, 3326−3337.
(27) Evans, D. C.; Watt, A. P.; Nicoll-Griffith, D. A.; Baillie, T. A. Drug−Protein Adducts: An Industry Perspective on Minimizing the Potential for Drug Bioactivation in Drug Discovery and Development. *Chem. Res. Toxicol.* **2004**, *17*, 3−16.
(28) Liebler, D. C.; Guengerich, F. P. Elucidating Mechanisms of Drug-Induced Toxicity. *Nat. Rev. Drug Discovery* **2005**, *4*, 410−420.
(29) Williams, D. P. Toxicophores: Investigations in drug safety. *Toxicology* **2006**, *226*, 1−11.
(30) Drahl, C.; Cravatt, B. F.; Sorensen, E. J. Protein-Reactive Natural Products. *Angew. Chem. Int. Ed.* **2005**, *44*, 5788−5809.
(31) Groll, M.; Nguyen, H.; Vellalath, S.; Romo, D. (−)-Homosalinosporamide A and Its Mode of Proteasome Inhibition: An X-ray Crystallographic Study. *Mar. Drugs* **2018**, *16*, 240−249.
(32) Mincer, T. J.; Jensen, P. R.; Kauffman, C. A.; Fenical, W. Widespread and Persistent Populations of a Major New Marine Actinomycete Taxon in Ocean Sediments. *Appl. Environ. Microbiol.* **2002**, *68*, 5005.
(33) Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. Salinosporamide A: A Highly Cytotoxic Proteasome Inhibitor from a Novel Microbial Source, a Marine Bacterium of the New Genus Salinospora. *Angew. Chem. Int. Ed.* **2003**, *42*, 355−357.
(34) Manasanch, E. E.; Korde, N.; Zingone, A.; Tageja, N.; Fernandez de Larrea, C.; Bhutani, M.; Wu, P.; Roschewski, M.; Landgren, O. The Proteasome: Mechanisms of Biology and Markers of Activity and Response to Treatment in Multiple Myeloma. *Leuk. Lymphoma* **2014**, *55*, 1707−1714.
(35) Eisenberg-Lerner, A.; Benyair, R.; Hizkiahou, N.; Nudel, N.; Maor, R.; Kramer, M. P.; Shmueli, M. D.; Zigdon, I.; Cherniavsky Lev, M.; Ulman, A.; Sagiv, J. Y.; Dayan, M.; Dassa, B.; Rosenwald, M.; Shachar, I.; Li, J.; Wang, Y.; Dezorella, N.; Khan, S.; Porat, Z.; Shimoni, E.; Avinoam, O.; Merbl, Y. Golgi Organization Is Regulated by Proteasomal Degradation. *Nat. Commun.* **2020**, *11*, 409.
(36) Mishra, R.; Upadhyay, A.; Prajapati, V. K.; Mishra, A. Proteasome-Mediated Proteostasis: Novel Medicinal and Pharmacological Strategies for Diseases. *Med. Res. Rev.* **2018**, *38*, 1916−1973.
(37) Sengupta, S.; Mehta, G. Non-Peptidic Natural Products as Ubiquitin-Proteasome Inhibitors. *Tetrahedron* **2019**, *75*, 817−853.
(38) Manasanch, E. E.; Orlowski, R. Z. Proteasome Inhibitors in Cancer Therapy. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 417−433.
(39) Gulder, T. A. M.; Moore, B. S. Salinosporamide Natural Products: Potent 20 S Proteasome Inhibitors as Promising Cancer Chemotherapeutics. *Angew. Chem. Int. Ed.* **2010**, *49*, 9346−9367.
(40) Zmuda, F.; Sastry, L.; Shepherd, S. M.; Jones, D.; Scott, A.; Craggs, P. D.; Cortes, A.; Gray, D. W.; Torrie, L. S.; De Rycker, M. Identification of Novel Trypanosoma cruzi Proteasome Inhibitors Using a Luminescence-Based High-Throughput Screening Assay. *Antimicrob. Agents Chemother.* **2019**, *63*, 1−15.
(41) Cromm, P. M.; Crews, C. M. The Proteasome in Modern Drug Discovery: Second Life of a Highly Valuable Drug Target. *ACS Cent. Sci.* **2017**, *3*, 830−838.
(42) Aminake, M. N.; Arndt, H.-D.; Pradel, G. The Proteasome of Malaria Parasites: A Multi-Stage Drug Target for Chemotherapeutic Intervention? *Int. J. Parasitol.: Drugs Drug Resist.* **2012**, *2*, 1−10.
(43) Le Chapelain, C.; Groll, M. Rational Design of Proteasome Inhibitors as Antimalarial Drugs. *Angew. Chem., Int. Ed.* **2016**, *55*, 6370−6372.
(44) Gandotra, S.; Schnappinger, D.; Monteleone, M.; Hillen, W.; Ehrt, S. In Vivo Gene Silencing Identifies the Mycobacterium Tuberculosis Proteasome as Essential for the Bacteria to Persist in Mice. *Nat. Med.* **2007**, *13*, 1515−1520.
(45) Mishto, M.; Bellavista, E.; Santoro, A.; Stolzing, A.; Ligorio, C.; Nacmias, B.; Spazzafumo, L.; Chiappelli, M.; Licastro, F.; Sorbi, S.; Pession, A.; Ohm, T.; Grune, T.; Franceschi, C. Immunoproteasome and LMP2 polymorphism in aged and Alzheimer's disease brains. *Neurobiol. Aging* **2006**, *27*, 54−66.
(46) Díaz-Hernández, M.; Hernández, F.; Martín-Aparicio, E.; Gómez-Ramos, P.; Morán, M. A.; Castaño, J. G.; Ferrer, I.; Avila, J.; Lucas, J. J. Neuronal Induction of the Immunoproteasome in Huntington's Disease. *J. Neurosci.* **2003**, *23*, 11653−11661.
(47) Tanaka, K. The Proteasome: Overview of Structure and Functions. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2009**, *85*, 12−36.
(48) Coux, O.; Tanaka, K.; Goldberg, A. L. Structure and Functions of the 20S and 26S Proteasomes. *Annu. Rev. Biochem.* **1996**, *65*, 801−847.
(49) Lupas, A.; Zwickl, P.; Wenzel, T.; Seemüller, E.; Baumeister, W. Structure and function of the 20S proteasome and of its regulatory complexes. *Cold Spring Harbor Symp. Quant. Biol.* **1995**, *60*, 515−524.
(50) Kunjappu, M. J.; Hochstrasser, M. Assembly of the 20S Proteasome. *Biochim. Biophys. Acta, Mol. Cell Res.* **2014**, *1843*, 2−12.
(51) Lowe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal Structure of the 20S Proteasome from the Archaeon T. Acidophilum at 3.4 A Resolution. *Science* **1995**, *268*, 533−539.
(52) Orlowski, M.; Cardozo, C.; Michaud, C. Evidence for the Presence of Five Distinct Proteolytic Components in the Pituitary Multicatalytic Proteinase Complex. Properties of Two Components Cleaving Bonds on the Carboxyl Side of Branched Chain and Small Neutral Amino Acids. *Biochemistry* **1993**, *32*, 1563−1572.
(53) Fenteany, G.; Standaert, R.; Lane, W.; Choi, S.; Corey, E.; Schreiber, S. Inhibition of Proteasome Activities and Subunit-Specific Amino-Terminal Threonine Modification by Lactacystin. *Science* **1995**, *268*, 726−731.
(54) Groll, M.; Potts, B. C. Proteasome Structure, Function, and Lessons Learned from Beta-Lactone Inhibitors. *Curr. Top. Med. Chem.* **2011**, *11*, 2850−2878.
(55) Macherla, V. R.; Mitchell, S. S.; Manam, R. R.; Reed, K. A.; Chao, T.-H.; Nicholson, B.; Deyanat-yazdi, G.; Mai, B.; Jensen, P. R.; Fenical, W. F.; Neuteboom, S. T. C.; Lam, K. S.; Palladino, M. A.; Potts, B. C. M. Structure−Activity Relationship Studies of Salinosporamide A (NPI-0052), a Novel Marine Derived Proteasome Inhibitor. *J. Med. Chem.* **2005**, *48*, 3684−3687.
(56) Groll, M.; Huber, R.; Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of β-Lactone Ring Opening and a Mechanism for Irreversible Binding. *J. Am. Chem. Soc.* **2006**, *128*, 5136−5141.
(57) Marizomib Salinosporamide, A.: Proteasome Inhibitor: MedChemExpress. 2021, https://www.medchemexpress.com/Marizomib.html (accessed May 20, 2021).
(58) Di, K.; Lloyd, G. K.; Abraham, V.; MacLaren, A.; Burrows, F. J.; Desjardins, A.; Trikha, M.; Bota, D. A. Marizomib Activity as a Single Agent in Malignant Gliomas: Ability to Cross the Blood-Brain Barrier. *Neuro Oncol.* **2016**, *18*, 840−848.
(59) Kale, A. J.; Moore, B. S. Molecular Mechanisms of Acquired Proteasome Inhibitor Resistance. *J. Med. Chem.* **2012**, *55*, 10317−10327.
(60) Singh, A. V.; Palladino, M. A.; Lloyd, G. K.; Potts, B. C.; Chauhan, D.; Anderson, K. C. Pharmacodynamic and Efficacy Studies of the Novel Proteasome Inhibitor NPI-0052 (Marizomib) in a Human Plasmacytoma Xenograft Murine Model. *Br. J. Haematol.* **2010**, *149*, 550−559.
(61) Nguyen, H.; Ma, G.; Gladysheva, T.; Fremgen, T.; Romo, D. Bioinspired Total Synthesis and Human Proteasome Inhibitory Activity of (−)-Salinosporamide A, (−)-Homosalinosporamide A,

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

and Derivatives Obtained via Organonucleophile Promoted Bis-cyclizations. *J. Org. Chem.* **2011**, *76*, 2−12.
(62) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, *11*, 3575−3589.
(63) Ray, S.; Murkin, A. S. New Electrophiles and Strategies for Mechanism-Based and Targeted Covalent Inhibitor Design. *Biochemistry* **2019**, *58*, 5234−5244.
(64) Manam, R. R.; McArthur, K. A.; Chao, T.-H.; Weiss, J.; Ali, J. A.; Palombella, V. J.; Groll, M.; Lloyd, G. K.; Palladino, M. A.; Neuteboom, S. T. C.; Macherla, V. R.; Potts, B. C. M. Leaving Groups Prolong the Duration of 20S Proteasome Inhibition and Enhance the Potency of Salinosporamides. *J. Med. Chem.* **2008**, *51*, 6711−6724.
(65) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(66) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(67) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(68) Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, *164*, 399−407.
(69) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2021*; University of California: San Francisco, 2021.
(70) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(71) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(72) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(73) Grest, G. S.; Kremer, K. Molecular Dynamics Simulation for Polymers in the Presence of a Heat Bath. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3628−3631.
(74) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.
(75) Field, M. J.; Albe, M.; Bret, C. l.; Proust-De Martin, F.; Thomas, A. The Dynamo Library for Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.* **2000**, *21*, 1088−1100.
(76) Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. Binding Isotope Effects as a Tool for Distinguishing Hydrophobic and Hydrophilic Binding Sites of HIV-1 RT. *J. Phys. Chem. B* **2015**, *119*, 917−927.
(77) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(78) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(79) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(80) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385−395.
(81) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(82) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(83) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes. *J. Phys. Chem. B* **2006**, *110*, 17663−17670.
(84) Chuang, Y.-Y.; Corchado, J. C.; Truhlar, D. G. Mapped Interpolation Scheme for Single-Point Energy Corrections in Reaction Rate Calculations and a Critical Evaluation of Dual-Level Reaction Path Dynamics Methods. *J. Phys. Chem. A* **1999**, *103*, 1140−1149.
(85) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V. Protein conformational landscapes and catalysis. Influence of active site conformations in the reaction catalyzed by L-lactate dehydrogenase. *ACS Catal.* **2015**, *5*, 1172−1185.
(86) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(87) Saha, A.; Oanca, G.; Mondal, D.; Warshel, A. Exploring the Proteolysis Mechanism of the Proteasomes. *J. Phys. Chem. B* **2020**, *124*, 5626−5635.
(88) Bürgi, H. B.; Dunitz, J. D.; Shefter, E. Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.* **1973**, *95*, 5065−5067.
(89) Bürgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G. Stereochemistry of Reaction Paths at Carbonyl Centres. *Tetrahedron* **1974**, *30*, 1563−1572.
(90) Fleming, I. *Molecular Orbitals and Organic Chemical Reactions*; Wiley: Chichester, 2011; pp 214−215.
(91) Heathcock, C. H. Understanding and controlling diastereofacial selectivity in carbon-carbon bond-forming reactions. *Aldrichim Acta* **1990**, *23*, 94−111.
(92) Radisky, E. S.; Koshland, D. E. A Clogged Gutter Mechanism for Protease Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10316.
(93) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(94) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by Dihydrofolate Reductase and Other Enzymes Arises from Electrostatic Preorganization, Not Conformational Motions. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14115−14120.

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

(95) Moliner, V. "Eppur Si Muove" (yet It Moves). *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15013−15014.
(96) Krzemińska, A.; Moliner, V.; Świderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J. Am. Chem. Soc.* **2016**, *138*, 16283−16298.
(97) Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J. Protein Flexibility and Preorganization in the Design of Enzymes. The Kemp Elimination Catalyzed by HG3.17. *ACS Catal.* **2015**, *5*, 2587−2595.
(98) Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J. Revealing the Origin of the Efficiency of the De Novo Designed Kemp Eliminase HG-3.17 by Comparison with the Former Developed HG-3. *Chem.—Eur. J.* **2017**, *23*, 7582−7589.
(99) Świderek, K.; Tuñón, I.; Williams, I. H.; Moliner, V. Insights on the Origin of Catalysis on Glycine N-Methyltransferase from Computational Modeling. *J. Am. Chem. Soc.* **2018**, *140*, 4327−4334.
(100) Martí, S.; Bastida, A.; Świderek, K. Theoretical Studies on Mechanism of Inactivation of Kanamycin A by 4′-O-Nucleotidyltransferase. *Front. Chem.* **2019**, *6*, 660−671.

https://doi.org/10.1021/acscatal.1c02614
ACS Catal. 2021, 11, 11806−11819

4.2.3. Binding and reactivity of C5-phenyl-SalA (pSalA)

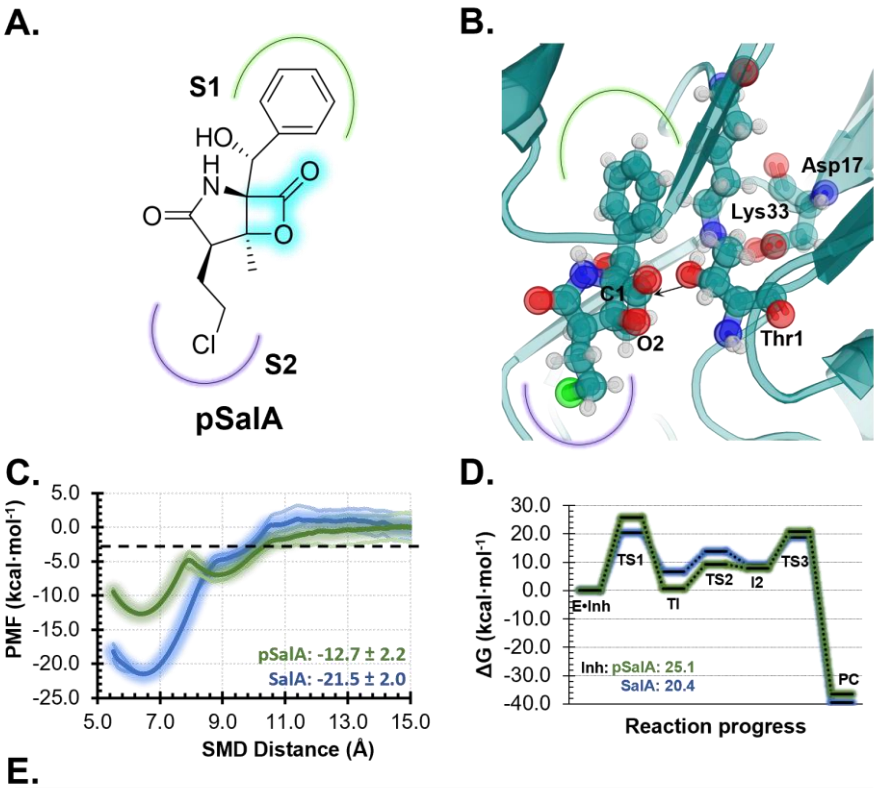
This last section of the Results chapter is focused on understanding the binding process in the γ-lactam-β-lactone family of inhibitors, and how modifications in the P1 substituent can affect the E•I formation. Additionally, we are testing docking protocols, since docking is one of the most widely used methods in the computational aided drug design process. The suitability of this tool for this type of modification in a small drug has been evaluated. This has been done by computing the binding affinity of an available data set of compounds in the β5 active site with experimentally characterized IC_{50} values.[13] Therefore, the first employed tool was molecular docking, done using Glide software,[14,15] where the full data set was studied and the compounds were ranked in terms of gscore, the Glide internal scoring function for evaluating the docking poses. The results obtained with this docking assay show a very low regression coefficient for the representation of the gscore vs. pIC50 (lower than 0.1). Moreover, the compound showing the lowest affinity for the β5 active site according to the experimentally measured pIC50, pSalA, was ranked as the best compound according to the computed gscore. Therefore, although docking protocols usually produce good results for small organic molecules such as SalA and its derivatives, and docking was able to reproduce the expected E•I geometry, it was not able to produce a good answer in terms of rating the compounds compared to their experimental activity for this set of compounds, not allowing to distinguish the analogs in the data set with better binding affinities. Based on this unexpected bad result from docking studies for pSalA, the next step in our study was devoted to finding a systematic way of computing binding energies with SalA and pSalA. The level of theory was progressively increased, starting at the MM level with SMD simulations. The SMD simulations were computed to produce an approximated PMF unbinding profile. To do so, it was necessary to follow a calibration protocol to select the optimal parameters for the simulation, i.e., the spring constant (K) and the velocity of unbinding (*v*). Two rounds of calibration were done. First, an initial velocity was selected according to similar calculations reported in literature[16,17] and the K parameter was tuned. Then different values of *v* were tested for the selected K. Finally, the values of K and *v* were selected as the smallest values that allowed to keep the stiff-spring approximation. Once the parameters were known, the number of replicas necessary to obtain an averaged value was determined by

computing the SEM by increasing the number of replicas until the variation was close to zero. Finally, FEP calculations were done at a higher level of theory with hybrid QM/MM potentials. A series of non-physical QM/MM simulations, where the inhibitor electrostatic and van der Waals interactions are progressively turned down in aqueous solution and in the active site of the protein. These simulations were performed to obtain the binding free energy based on a thermodynamic cycle.

The results obtained with both MD techniques render better binding affinity for SalA than for pSalA. Furthermore, the binding affinity of both compounds was tested considering different protonation states of Lys33 in the active site. The results obtained with both MD techniques showed that the binding is stronger when Lys33 is in its neutral state. This is an interesting finding which shows that neutral Lys33 is not only required for the kinetic step of inhibition but additionally favors the formation of the non-covalent inhibitor-enzyme adduct. Additionally, the computed results with both techniques show overestimated values with respect to the experimental binding energies deduced from IC_{50} values. This effect can be related to the two approximations considered, i.e., the deduction of the binding energies from IC_{50} values, and the approximations taken in the computational methods employed. Nevertheless, both MD techniques render binding free energies sharing the same trend observed in the experimental values.

Finally, the inhibition reaction path was computed for pSalA, based on the inhibition mechanism deduced in previous studies (section 4.2.1). The computed FES showed a higher barrier for the rate-limiting step with this inhibitor, 25.1 kcal·mol^{-1}, than with the SalA, 20.4 kcal·mol^{-1}. This result combined with the lower binding affinity of pSalA suggests that the experimental activity of this compound, reflected in IC_{50} values, is affected by both main processes of inhibition, binding, and reactivity.

A summary of the results is presented in Figure 4.4.



E.

	pIC_{50}	$gscore_{docking}$ (kcal·mol^{-1})	ΔG_{SMD} (kcal·mol^{-1})	$\Delta\Delta G_{FEP}$ (kcal·mol^{-1})
SalA	8.7	-6.64	-21.5 ± 2.0	-23.9 ± 0.3
pSalA	6.0	-6.87	-12.7 ± 2.2	-18.1 ± 0.3

Figure 4.4. **A.** Schematic representation of pSalA. **B.** Representation of the E•pSalA complex. **C.** Averaged free energy binding from SMD simulations for SalA and pSalA. **D.** Free energy profiles for the inhibition mechanism computed at M06-2X:AM1/MM level of theory for SalA (blue) and pSalA (green). **E.** Summary table of the computed binding energies from the different methods employed in this work.

These results are presented on a manuscript that is in preparation for submission:

Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog.

Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analogue

Natalia Serrano-Aparicio,[a] Laura Scalvini,[b,] Alessio Lodola,[b*,] Vicent Moliner,[a*,] Katarzyna Świderek[a*]

Abstract

Understanding the relationship of ligand binding (MC formation) and reaction mechanism and their role in the inhibitory activity of covalent inhibitors may be key on certain systems. A set of γ-lactam-β-lactone inhibitors distinguished by the P1 side chain has been considered to address the S1 pocket specificity. We show that docking studies fail in distinguishing the weakest inhibitor from the data set (pSalA) from the strongest (SalA). Therefore, we present a systematic approach to understand the key points of 20S Proteasome inhibition by these two covalent inhibitors (SalA and pSalA). We have focused on different computational techniques at different levels of theory, to see which ones allow to distinguish and reproduce the experimental inhibitory activity of both molecules. Addressing both major points of protein inhibition by covalent inhibitors, binding and reactivity.

KEYWORDS: Molecular dynamics simulation, Steered molecular dynamic, Drug design, Drug-receptor interactions, Ligand−protein complex, QM/MM

[a] *Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain*
[b] *Department of Food and Drug, University of Parma, Parma, Italy.*
E-mail:alessio.lodola@unipr.it, moliner@uji.es; swiderek@uji.es
† Electronic Supplementary Information (ESI) available: detailed representation of the obtained docking poses, SMD calibration protocol, electrostatic interactions obtained in 100ps of QM/MM MD simulation for both inhibitors, obtained free energy surfaces for pSalA at AM1/MM level.

Introduction

The use of covalent inhibitors (CI) has increased in the past years[1] since CIs increase the residence time in the active site by forming an stable covalent adduct with the drug-target and they have proved to be safe if some key aspects are fulfilled. A molecule is considered safe CI if it is capable to react in low doses and it is very selective to its target,[2,3] avoiding off-target effects that can lead to immune responses. Furthermore, CIs add key advantages to the possibilities to use the inhibitor as a drug. It is possible to use CI drugs with high potency at a low dose[3] and the possibility of stablishing non-equilibrium binding kinetics in the case of irreversible inhibitors.[4]

Computational tools have been widely applied in the study of enzyme activity and inhibition. Moreover, they have been demonstrated to be useful in drug discovery tasks, as seen in the increment of research literature in this field during the last years.[5] Once the protein target is well defined and the binding pocket is known, one key question is to evaluate the affinity ($\Delta G_{binding}$), and in the case of covalent inhibitors, to determine the reaction rate efficiency (ΔG_{inact}).

To assist drug discovery, numerous structure-based screening methods, such as docking, have been developed,[6] to deal with large compound libraries and prospectively identify promising hit compounds[7] and rank the potency of different classes. In this regard, an active research field has been protein-ligand docking and scoring, producing usually good results for small molecules.[8] Here, the docking calculation is considered as the positioning of the ligand in a protein binding site, and the scoring is the evaluation of the quality of docked ligands.[9] However, when addressing problems with similar compounds, the scoring of docked ligands can lead to errors, and it can be key to evaluate different types of methods and increase the precision of the calculations. Therefore, computing absolute binding free energies is a challenging task.[9,10]

Hence, improving the methodology by increasing the amount of sampling through MD-based techniques is required. The first option is to use unbiased MD for exploring spontaneous binding events, although this type of dynamic docking requires long MD simulations and in binding processes with barriers higher than 4-5 kcal·mol^{-1} it can be necessary to use enhanced sampling MD.[11] For this purpose, the steered molecular dynamics approximation (SMD)[12–14] can be used. SMD is a simple approach that forces ligand unbinding by applying a guiding potential with constant force and velocity, along a reaction coordinate, in this case the unbinding pathway.[15] With this technique, the averaged work necessary to pull the inhibitor out of the active site can be obtained. Furthermore, if the Jarzynski equality[16] is applied, an approximation to the averaged PMF is obtained.

Additionally, multiscale methods can be used to increase further the level of theory. Here, the unbinding can be obtained by applying several enhanced techniques such as PMF[17] or metadynamics[18]. But also alchemical simulations can be used, expressing the binding free energy as a function of interaction energies between the receptor, ligand and solvent.[19] The free energy perturbation method (FEP)[20] therefore can be applied *taking advantage of* the multiscale approach to treat the inhibitor and the rest of the system as the QM and the MM level of theory respectively. The binding free energy in this case is obtained by computing a series of nonphysical overlapping states, which then are related through a thermodynamic cycle.

The enzyme inhibition process by covalent compounds can be generally divided in two major steps, as shown in equation 1. Fist is the recognition step, where the non-covalent intermediate, E•I, is formed. Second is the chemical reaction step, until the formation of

the final covalently bound product, E-I.[21] Although these two processes can be studied by separated from a computational point of view, in some cases is important to study both to have a clearer description of the full inhibition process.

$$E + I \overset{K_i}{\leftrightarrow} E{\bullet}I \xrightarrow{k_{inact}} E - \mathrm{I} \qquad (1)$$

In the case of computational aided drug discovery, it is key to evaluate the obtained results with the available experimental data, especially in the stage of validation of the methods for a specific problem. In this regard, the best scenario is to have kinetic assays for the molecule and protein under study, but it is not the usual situation, since thorough enzymatic kinetic assays are not always available in literature. Usually, the first performed assays are are IC_{50} evaluations, where the inhibitor activity is tested as the averaged concentration of inhibitor necessary to reduce to half the activity of the protein.[22] The problem with this magnitude is that it typically differs from the K_i values, and sometimes can reflect artefacts.[23,24] Resulting, IC_{50} values are not the ideal magnitude to compare binding energies,[25,26] but an approximation that is usually considered when there is no other data available. Although it has been observed that changes in IC_{50} values sometimes can be related with differences in the recognition step for sets of ligands differing in an R position far from the reactive centre.[27–29]

The proteasome is a highly sophisticated nanomachine responsible for the major non-lysosomal protein degradation through the ubiquitin-proteasome system, that has evolved to a great selectivity and specificity through the unique architecture of the particle.[30–32] The 20S proteasome consists of two types of protein subunits arranged in four stacked heptameric rings ($\alpha_7\beta_7\beta_7\alpha_7$). Within the β-type subunits, the eucaryotic proteasome contains three different types of active sites, characterized by its selectivity. Namely being chymotrypsin-like (β5), trypsin-like (β2) and caspase-like (β1). All three active sites contain an N-terminal threonine residue that acts as nucleophile in peptide bond cleavage.[33–35]

As key protein regulatory pathway, the proteasome system has an important role in the regulation of cellular metabolism and homeostasis.[32,36] First proteasome inhibitors were used as tools to investigate different aspects of proteasome functions in cellular systems and tissues.[36] This type of research lead to clinical development of proteasome inhibitors for the treatment of inflammatory and neoplastic diseases,[37,38] where multiple myeloma cells are known as the most protein secretors of all cell types. Therefore, proteasome inhibition leads to rapid apoptosis.[39] With three proteasome inhibitors already approved by FDA,[40–42] proteasome inhibition is a very active area of research.[36,43]

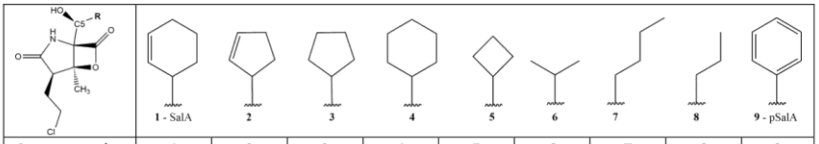
Many lead compounds are found as natural products,[1] such as omuralide and salinosporamide A (SalA),[44–46] both belonging to the γ-lactam-β-lactone inhibitor family. These inhibitors are densely functionalized small molecules[44] and among them SalA is currently in clinical trials for the treatment of several types of cancer.[47] SalA was isolated from marine actinomycete *salinospora tropica* and has IC_{50} activities in the low nanomolar range,[44] with different potency for the three proteasome active sites. As SalA is a small molecule, the selectivity for the different active sites relays on the C5 residue, which interacts within the S1 binding pocket in the proteasome.[44] Therefore, alterations at this position may lead to distinct selectivity, as shown by Nett et al.[48]

In this study, SalA and a set of congeneric analogues bearing substituents of different shape and size at position 5 (see Table 1) have been docked in the active site to rationalize the experimentally reported differences in pIC50 values. While computed scores parallel

potency values for most of the compounds present in the data set, our docking protocol fails to correctly rank pSal, the less potent inhibitor of the series. On the other hand, the high structural similarity between SaA and pSalA, also reflected in highly similar physico-chemical property (being their molecular weight 313.78 and 309.74 $g \cdot mol^{-1}$ and their logP 1.7 and 1.5 for SalA an pSalA, respectively), make difficult to understand while pSalA is 500-times less potent than SalA. The condition, in which a pair of similar compounds active against on the same target display a large difference in potency, is unknown as activity cliffs and represent a challenging scenario for any predictive computational method.

For this reason, binding and reactivity properties SalA and pSalA are compared in further detail with a wide arsenal of state-of-the-art computational approaches. Binding energies have been computed with molecular mechanics potentials by SMD simulations, and with QM/MM potentials by FEP method, to obtain this magnitude at different levels of theory. Finally, the inhibition mechanism of pSalA has been explored according to the previously explored mechanism for SalA[49] (see Scheme 1) The goal of this comparison is to gain deeper understanding on how affected is IC_{50} by binding and reactivity, and see whether it is possible to consider only one of them or not for obtaining conclusions for these irreversible covalent inhibitors of the 20S Proteasome.

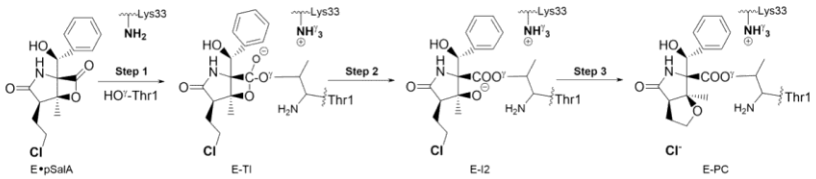
Table 1. Initial dataset of C5-SalA substitutions along with their experimental IC50 values for the chymotrypsin-like active site in yeast 20S Proteasome, data from ref [48]. Within the data set three types of compounds can be distinguished, alicyclic (compounds: 1, 10, 5, 4, 6), aliphatic (compounds: 3, 9, 8) and aromatic (compound: 7).



Compound	1 SalA	2	3	4	5	6	7	8	9 pSalA
CT-L YP IC_{50} [nM] [REF]	1.9 ± 0.2	2.2 ± 0.1	9.3 ± 1.6	27.5 ± 3.7	93.4 ± 4.3	101 ± 15	132 ± 19	245 ± 38	1029 ± 419
Molecular Weight ($g \cdot mol^{-1}$)	313.78	299.75	301.77	315.79	287.74	275.73	289.76	275.73	309.74
VdW Volume*	272.49	255.69	261.88	278.68	245.08	238.64	255.65	238.85	260.1
Consensus $logP_{o/w}$ **	1.71	1.41	1.65	2.01	1.32	1.2	1.62	1.24	1.5

* Data obtained from refs. [50–52]. ** Data obtained from ref. [53]

Scheme 1. Mechanism of inhibition of 20S Proteasome with pSalA explored as proposed for SalA in our previous studies.[49]



Additionally, having a reliable protocol to properly compute binding energies may be useful in the further development of SalA-C5 analogues, by tuning the selectivity towards the different S1 pockets in the proteasome.

Computational methods

Molecular Docking. The *Saccharomyces cerevisiae* crystal structure of the 20S proteasome complexed with SalA was used as starting structure (PDB ID: 2FAK).[46] The structure was cut 20Å apart from the inhibitor molecule in the β5 active site containing the covalent inhibition product. The reactant complex was built from the X-ray structure according to the known Michaelis complex characterized in our previous studies.[49] Two models varying the protonation state of Lys33 were prepared i.e. model (A) with its standard protonated form and the second model (B) with Lys33 in its neutral state. This is because the protonation state of this residue in the Michaelis complex is still a question of debate. However as shown in our previous studies, the most accepted mechanism at the moment is the one that assumes a direct role of Lys33 in Thr1 activation that takes place in the first step of the inhibition mechanism of the β5 active site by SalA and thus propose its neutral form in the Michaelis complex.[49,54] The protein preparation wizard was used to build the protein for the docking simulations using Maestro software.[55] Missing hydrogen atoms were added and the pKa of titratable residues was computed at pH 7.0. Restrained energy minimization was applied to the full model using the OPLS3e[56] force field.

The remaining inhibitors in the data set[48] were built by substituting the cyclohexene side chain in SalA (1) by the proposed functional groups such as cyclopentene (2), cyclopentane (3), cyclohexane (4), cyclobutane (5), isopropyl (6), n-butyl (7), ethyl (8), phenyl (9), as shown in Table 1. All newly build ligands were optimized in the gas phase employing the MacroModel[57] program applying the OPLS3e force field with a tolerance energy gradient of 0.01 kcal·mol^{-1}·Å^{-1}. The SalA model in the reactant complex was selected as the grid-defining ligand to generate the receptor grids. Dimensions of enclosing and bounding boxes were set to 32 and 14 Å on each side, respectively, centered on the inhibitor molecule. Default van der Waals radii scaling parameters were used in a way that a scaling factor and partial charge cut-off were 1 and 0.25, respectively. First, a round of free docking was done. Additionally, a second round of restricted docking was done assuming that the formation of a hydrogen bond between Gly47:H and the carbonyl oxygen atom of the β-lactone ring in the ligand was required and specific sampling conditions for the conformational search of the docked ligands were set (i.e., enhanced sampling for ligand conformations, no sampling for ring conformations, not canonicalize of the initial structure). Docking studies with all the compounds of the data set were performed with Glide software.[58–60]

Steered Molecular Dynamics. A new model was prepared for the SMD simulations, using the same crystal structure as used for docking studies. To decrease the simulation time, the system was reduced to three subunits surrounding the β5 active site. Chains K (containing the active site), L and W were selected. As in the previous method, two models varying the protonation state of Lys33 were prepared i.e. model (A) with its standard protonated form and the second model (B) with Lys33 in its neutral state. Hydrogen atoms were added to the protein preparation wizard matching a pH of 7.0. The system was solvated with a TIP3P[61] water box of 83x80x114 Å^3 and the system charge was neutralized by the addition of 11 sodium atoms. This model contains 5,130 heavy atoms in protein and a total of 74,216 atoms after solvation.

The best docking pose was selected as the initial structure for both inhibitors, SalA and pSalA. Missing parameters for the pSalA were obtained based on GAFF[62] using the Antechamber[63] software and charges were computed with Prime.[64–66] All simulations were performed using the AMBER[67] force field and GROMACS[68,69] package.

The system was minimized using 5000 steps of steepest descent minimization followed by 50000 steps of conjugate gradient, with 50 kcal·mol^{-1}·Å^{-2} positional restraints applied to all heavy atoms in the protein and ligand. The system was then heated at constant volume keeping the previous step positional restraints, varying the target temperature linearly from 10 to 300 K over 3ns using Berendsen thermostat.[70] Positional restraints were gradually reduced in two stages. The first stage was performed at constant temperature reducing first the restraint to 25 kcal·mol^{-1}·Å^{-2} during 1ns and then reducing the restrained atoms to the heavy atoms in the protein backbone and the heavy atoms of the bicyclic ring for 1ns with a 1 fs time step. In the second stage, the restraining force in the γ-lactam-β-lactone core heavy atoms was further reduced to 5 kcal·mol^{-1}·Å^{-2} at constant temperature and pressure during 8ns with 2fs time step. Parrinello-Rahman[71,72] pressure coupling was used for pressure control and Nosé-Hoover thermostat[73,74] for temperature control. In the final 2ns of this NPT equilibration, only the heavy atoms in the backbone of the protein were restrained and the ligand molecule restrain was released keeping the hydrogen bonds present in the docking pose between the Gly47 backbone and the β-lactone carbonyl oxygen and γ-lactam amide hydrogen atoms. Finally, 5ns of production at 2fs time step and constant temperature were performed releasing the restraints on the active site and keeping 5 kcal·mol^{-1}·Å^{-2} restraint on the heavy atoms of the backbone 21Å away from the ligand, for properly conserving the external structure of the system since it does not contain the full 20S Proteasome.

All SMD simulations were performed using GROMACS package as MD engine and PLUMED[75,76] for controlling the steering path. The collective variable was selected as the distance between two geometrical centers, the geometrical center of residues Thr1-Lys33-Ser131 and the geometrical center of the position of the heavy atoms of the inhibitor. Thus, the controlled distance is between Trh1 and the center of the inhibitor avoiding the application of forces in punctual atoms. SMD simulations were performed at constant temperature of 300K using 2fs time step. In all of them, the force constant (*k*) is applied gradually during the first 5ns of simulation with a starting value for the CV of 5Å. Once the target *k* is reached the elongation of the CV starts until reaching 15Å with the selected velocity (*v*). Then calibration procedure for choosing the velocity and force constant was applied Different magnitudes of *k* varying between 5 and 25 kcal·mol^{-1}·A^{-2} and *v* from 0.005 and 0.000008 Å·ps^{-1} were used as described in detail in Supplementary results.

Alchemical Free Energy Perturbation (FEP) For performing the QM/MM studies a new model for E•pSalA within the *Human* 20S proteasome was built containing the full protein complex. As done in previous studies,[49,54,77] the *Homo sapiens* 20S proteasome crystal structure containing the covalent product of dihydroeponemycin inhibition was used as starting structure for the model (PDB ID: 5LF1).[78] Missing H atoms for the full protein system were added, considering the obtained pKa by propKa for titratable residues. Protonation states for the active site residues were selected as in previous mechanistic studies for SalA in the β5 Proteasome active site.[49] The structure of pSalA was obtained from the best docking pose of this compound and placed into the *human* proteasome β5 active site. Parameters mentioned above were used. The system was neutralized by the addition of 43 sodium counterions and was solvated with a box of TIP3P waters of 172x174x197 A^3, forming a model with more than half million atoms (534,952 atoms in total). Additionally, two models for SalA and pSalA fully solvated in water were prepared, with TIP3P water molecules as solvent. Each inhibitor molecule was solvated in a box of water molecules of 31x32x36 A^3 using the tleap module of Ambertools, containing 12,442 and 12,436 atoms, for SalA and pSalA respectively.

Molecular dynamics simulations were performed with AMBER force field using NAMD software to equilibrate the system. Initially, minimization and gradual heating to 310K with 0.001 K temperature increment was carried out. Then 100ps of NPT equilibration was performed with 1fs time step. Finally, 50ns of MD simulation production was done

with 1fs time step, keeping all atoms 40Å away from the inhibitor frozen. From this simulation, according to the evolution of the RMSD and analysis of key distances, the system can be considered equilibrated.
The protein bound state model for SalA was obtained from the previous studies[49] and pSalA was prepared as mentioned above. From classical molecular dynamics, a reactive snapshot was selected for this simulation. The QM subset of atoms contained only the inhibitor molecule (SalA or pSalA), and the water molecules or the solvated protein were treated at the MM level. The positions of the atoms of residues beyond 20 Å from the inhibitor were frozen and cut-offs for the nonbonding interactions were applied using a smooth switching function between 14.5 and 16 Å. A series of QM/MM NVT MD simulations were performed with 5 ps equilibration and 100ps production at 310K, with 0.01fs of the time step. As done in previous studies,[20] two parameters, λ, and γ, are introduced in the electrostatic and van der Waals terms respectively and are progressively changed from 1.0 to 0.0, to annihilate both contributions. Therefore, the FEP QM/MM simulation series were done in the protein and water, so the final binding free energy can be related through a thermodynamic cycle, as shown in Figure 1.

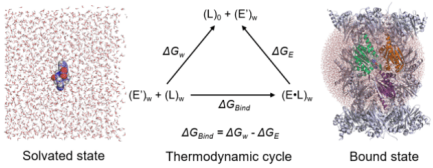


Figure 1. Thermodynamic cycle used to compute enzyme-inhibitor binding free energies with free energy perturbation method.

Reactivity calculations. The same E•pSalA model as for SalA was used to compute the free energy pathway assuming the same mechanism as obtained for SalA in previous studies.[49] A new QM subset was selected to include the active site residues participating in the inhibition reaction, consisting of 79 atoms. The inhibitor molecule, as well as Thr1 residue were fully included in the QM part, as well as the side chain of Lys33 and part of Thr2. Two link atoms were inserted between Lys33:C^{β}-C^{α} and Thr2:C^{α}-C. The full QM subset is depicted in Figure 2. The same cut-off scheme as in FEP simulations was employed and all residues beyond 20 Å from the inhibitor were frozen. The QM subsystem of atoms was treated with the AM1[79] semiempirical Hamiltonian and the M06-2X[80] functional, with the standard 6-31+G(d,p) basis set.[81,82]

To explore the inhibition mechanism of pSalA the appropriate distinguished reaction coordinates in each chemical step were explored by grid scanning, generating the potential energy surfaces (PESs) at AM1/MM level of theory. Then all stationary points of reactant complex (E•pSalA), intermediates (E-I), product complex (E-PC), and transition state (TS) structures were localized and characterized using a micro-macro iteration method[83] and by computing the matrix of second energy derivatives, at AM1/MM and M06-2X/MM levels. To confirm that each of the characterized TS structures was connected to the appropriate corresponding minima, IRC paths were traced down.

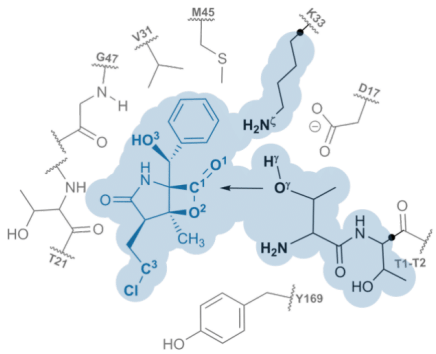


Figure 2. Schematic representation of the active site. The selected atoms for the QM part are within the blue area and the black dots show the position of link atoms.

The free energy profile was computed in terms of the potential of mean force (PMF) by computing a series of umbrella sampling[84] (US) AM1/AMBER MD simulations as implemented in fDYNAMO.[85,86] Based on the obtained PES, these simulations were computed by applying a force constant of 2500 kJ·mol^{-1}·Å^{-2} over the distinguished coordinate. An initial equilibration of 5 ps was done followed by 20 ps of production, with a time step of 1 fs. The weighted histogram analysis method[87] (WHAM) was used for the integration of the produced PMFs.

The level of theory was improved to reduce the errors by using the spline on the line approximation.[88–90] The corrected energy is calculated with a new energy function in terms of interpolated corrections, computed according to equation 2.

$$E = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\xi_1, \xi_2)] \qquad (2)$$

where S denotes a two-dimensional spline function, and the correction term is evaluated from the single-point energy difference between the high-level (HL) and the low-level (LL) calculation for the QM subsystem. The LL employed is the AM1 semiempirical Hamiltonian and the HL employed is the hybrid M06-2X functional, with the standard 6-31+G(d,p) basis set. These calculations were done using the Gaussian09 program[91] in combination with fDynamo.

Results and discussion

As explained in the methods section, all the compounds of the initial data set were built starting from SalA structure from the X-ray (2FAK). The E•SalA complex was built and, on this structure, the "C5 side chain" was modified in order to construct the rest of the compounds.

Since it has been shown that the protonation state of Lys33 is key for the reactivity, the binding assays were performed testing this property of the active site too, and both possible protonation states were tested. Therefore, a first docking model (A) was prepared setting Lys33 as protonated ($N^{\zeta}H_3^+$) and then a second docking model (B) was set with Lys33 deprotonated ($N^{\zeta}H_2$). The same binding assays were done for both models independently.

As learned from previous results,[49,54] the γ-lactam-β-lactone scaffold has some key requisites to consider that the ligand is in a reactive conformation. In order to lead to the nucleophilic attack by Thr1, a strong contact between the electrophile (carbon C1 of the carbonyl group of the β-lactone) and the nucleophile (oxygen $O^{\gamma Thr1}$) should be in the range of 3.0 – 3.3 Å. Additionally, this short distance is favoured if it exists a H bond contact with the O hole present in the proteasome active sites, the backbone of Gly47, ranging in between 1.0-1.2 Å. Finally, the

substituent located in the C5 of the SalA scaffold, must be interacting within the S1 pocket of the β5 active site, characterized by its hydrophobicity.

Binding energy calculations

Molecular docking calculations on the β5 20S Proteasome active site.

As commented before, the main goal of a docking assay can be divided in two main parts.[9] The ability if the program to produce a good geometry that will describe accurately the binding mode of the ligand within the active site. And the precision of scoring a different set of related ligands.

An initial blind docking assay was conducted to test if the program was able to obtain the expected reactive pose, as previously described. The result was that glide was not able to predict the expected geometry for all the compounds in none of the tested protonation stares for Lys33. ($Lys33NH_3^+$: 3 out of 9, $Lys33NH_2$: 6 out of 9 compounds produced the expected docking pose).

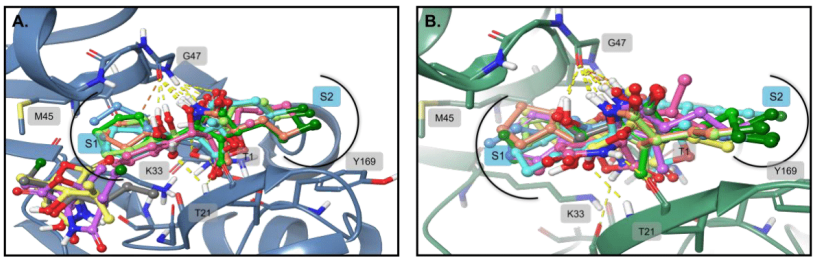


Figure 3. Docking pose obtained for all the data set with free docking for both protonation states, A and B respectively for $LysNH_3^+$ and $LysNH_2$.

As can be seen in Figure *3*, for $LysNH_3^+$ (A) there were 3 compounds laying outside of the active site and only 3 compounds were in a reactive conformation. Whereas for $LysNH_2$ (B) protonation state the docking poses were all within the active site but only 6 out of 9 compounds were in a reactive conformation.

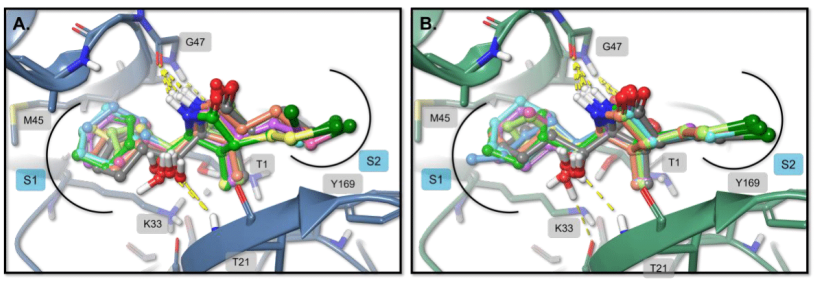


Figure 4. Docking pose obtained for all the data set with constrained docking for both protonation states, A and B respectively for $LysNH_3^+$ and $LysNH_2$.

Then a second docking experiment was designed to improve this result, since for scoring the different compounds the reactive docking pose needs to be evaluated. As explained in the methods section, several conditions were selected for this restrained docking assay, which lead

to the expected docking pose for all the compounds and bringing the possibility to compare the scoring values for each of them.

Given the restraints for obtaining the reactive docking pose, all the compounds share similar structure when binding, all of them fit in the S1 pocket and fulfil the prerequisite of having close electrophile-nucleophile contact and a H bond interaction with N^{Gly47}.

Additional figure in the SI can be found showing the individual docking pose for all the compounds in the data set of this last restrained docking assay.

Table 2. Experimental pIC50 values obtained from ref [48] and their corresponding experimental binding free energies, along with the results of Glide gscore scoring function for both protonation models, A and B.

Compound	pIC_{50}	$\Delta G_{bind.-exp}$ (kcal·mol^{-1})*	Gscore for Model A (Lys33:NH_3^+)	Gscore for Model B (Lys33:NH_2)
1	8.7	-12.4	-6.639	-6.644
2	8.7	-12.3	-6.301	-6.370
3	8.0	-11.4	-6.412	-6.235
4	7.6	-10.7	-6.788	-6.525
5	7.0	-10.0	-6.018	-5.904
6	7.0	-9.9	-5.734	-5.822
7	6.9	-9.8	-5.590	-5.756
8	6.6	-9.4	-5.617	-5.534
9	6.0	-8.5	-7.025	-6.872

*$\Delta G_{bind-exp}$ obtained as $\Delta G_{bind.}=RT\ln(IC_{50})$.[25,26]

Once the docking pose fulfils the reactive conformation requisites the glide scores were compared in between compounds to see which ones rank better according to the scoring function. For comparing this result with the experimental IC_{50}, the pIC_{50} was plotted against the obtained glide score (see Figure *5*).

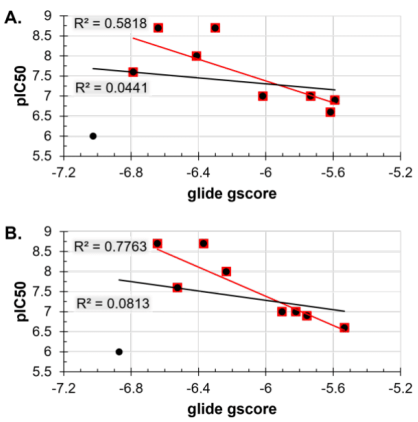


Figure 5. Representation of the obtained glide gscore (in kcal·mol^{-1}) vs. pIC50 for A. where Lys33 is protonated (NH_3^+) and B. where Lys33 is deprotonated (NH_2). The black line corresponds to

the trend line of the full data set, black dots. The red line corresponds to the trend line of all the data set but compound 7, red squares.

As it can be seen in Figure *5*, the regression coefficient for the full dataset gscore vs. pIC_{50} is lower than 0.1 (black line), meaning that the docking algorithm is ranking the low affinity compounds much higher than it should. As it can be seen, for both protonation states the program is ranking the lowest affinity compound as the best in the gscore scale. Then, in both cases if the lowest affinity compound, 7, is not considered the regression coefficient improves (red lines), but it never arrives to 0.8 value in any of the models.

For the Model A (Figure *5*.A, Lys33:NH_3^+) the best scored compound is 7. There are two compound groups visible according to their gscore. The first one corresponds to 4,1,5,10 (ordered by their gscore) which belong all of them to the group with alicyclic side chains. Then compound 6, the remaining alicyclic compound is scored. Finally, the three aliphatic groups correspond to the lowest ranked by docking.

For the Model B (Figure *5*.B, Lys33:NH_2) the best scored compound is 7. There are two compound groups visible according to the gscore. The first one corresponds to 1,4,10,5 (ordered by their gscore), which belong to the alicyclic side chains group. Then the second group is 6,3,9 the last alicyclic group and two aliphatic compounds. And finally, the lowest ranked is the aliphatic group in 8.

We have found a case where docking is still not able to produce a good answer with a set of related compounds, which shows that current docking scores although they are a useful tool must be considered carefully.

Since the docking calculations render a confusing result, ranking the least reactive compound as the best in both cases of study, the following techniques will focus on SalA and pSalA to evaluate if more precise methods can obtain a difference between them.

Steered molecular dynamics (SMD)

The SMD simulations were employed to have an enhanced sampling molecular dynamics method to quantify the binding energy with a classical molecular mechanics potential. The Jarzynski's rule[16] can be used to transform the obtained averaged work into potential of mean force, since this rule states that the explored path with minimum energy path will be the most likely to be unbinding minimum energy path. Therefore, the approximated PMF can be obtained as the exponential average of the work of all the replicas, where the replica with the lowest energy values will have the highest contribution.

The SMD energies were computed for both protonation models of Lys33. The resulting PMF for LysNH_3^+ (Figure *6*.A) shows that SalA has lower binding energy than pShal, as well as in the case of LysNH_2 (Figure *6*.B). Therefore, more work is required to pull SalA out of the active site, or that pSalA is easier to pull out of the active site. This result shows that SalA has larger binding energy than pSalA according to the SMD calculations. This result is qualitatively in agreement with the experimental IC_{50} mentioned above in the text. Additionally, it can be observed that when Lys33 is in its protonated state, both inhibitors have lower binding energy than when it is deprotonated. The obtained PMF absolute value must be considered as a qualitative result, since with different force constant and velocity conditions the absolute value can vary.

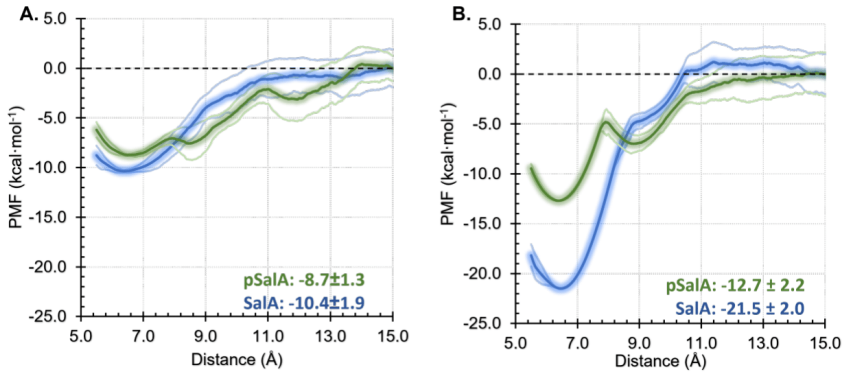


Figure 6. SMD results, PMF calculated by applying Jarzynski's equality and the standard deviation for SalA (blue lines) and pSalA (green lines). A. Model A: Lys33:NH_3^+, B. Model B: Lys33:NH_2.

Free energy perturbation (FEP)

The alchemical free energy perturbation method (FEP) was employed at a higher level of theory for obtaining the binding energies of both inhibitors. In this case the QM/MM approach was used to treat the inhibitor as the QM subset of atoms and the rest of the system at the MM level. As in the previous methods exposed above for estimating the binding free energy, both protonation models, A and B, were employed, to assess the difference of Lys33 being protonated or deprotonated (see Table 1). In water solution bot inhibitors show very close result for the electrostatic term (ΔG_{elec}) and the van der Waals term (ΔG_{vdW}). This shows how similar are these properties for both inhibitors in solution with a polar solvent. When the considered within the protein the β5 active site of the 20S proteasome is recognized by its hydrophobic character. Since in the FEP calculations the electrostatic and van der Waals contributions ($\Delta\Delta G_{elec}$ and $\Delta\Delta G_{vdW}$) can be separately obtained, it can be observed that for both inhibitors the major contribution to the binding free energy corresponds to the van der Walls term, being the electrostatic contribution much smaller. In model A, the contribution to the total binding energy from the electrostatic term is small for both inhibitors, -2.9 ± 0.3 and -1.0 ± 0.3 for SalA and pSalA respectively. Therefore, the total energy comes from the van der Waals contribution mainly, -17.8 ± 0.1 and -15.8 ± 0.1 for SalA and pSalA respectively. In both cases, SalA has larger contributions to the energy and the total binding free energy result is therefore better for SalA. In model B, the van der Waals contribution to the binding energies is almost the same as the obtained with the previous model, but the electrostatic contribution shows better interaction for both inhibitors. In the case of SalA this contribution increases in 3.2 kcal·mol^{-1}, whereas in the case of pSalA this contribution increases in 2.0 kcal·mol^{-1}. This shows that when Lys33 is neutral the binding free energy is improved.

Nevertheless, for both models, the results for SalA are better than the results obtained for pSalA, for both separated contribution and in the case of the total binding free energy. As observed in SMD results, binding free energy is improved in the case of model B, when Lys33 is neutral. Additionally, FEP calculations allow to distinguish that this improvement is due to the electrostatic contribution to the binding free energy, despite being the van der Waals contribution a major part of the resulting total free energy of binding. In Table 4 the results for the computed binding energies with the different employed methodologies is presented, along with the estimated experimental binding free energy deduced from IC_{50} values.[25,26]

Table 3. FEP results computed at AM1/MM level of theory for both protonation states models of the β5 active site and both inhibitors, SalA and pSalA. Electrostatic and van der Waals contributions to the binding energy are shown both in water and protein simulation. All energies are reported in kcal·mol^{-1}.

	Water		**Model A-Lys33:NH_3^+**		**Model B-Lys33:NH_2**	
	SalA	**pSalA**	**SalA**	**pSalA**	**SalA**	**pSalA**
ΔG_{elec}	-9.2 ± 0.2	-9.7 ± 0.2	-12.1 ± 0.2	-10.2 ± 0.1	-15.3 ± 0.2	-12.2 ± 0.2
ΔG_{vdW}	-23.8 ± 0.1	-23.6 ± 0.1	-41.6 ± 0.1	-39.5 ± 0.1	-41.6 ± 0.1	-39.2 ± 0.1
$\Delta\Delta G_{elec}$			-2.9 ± 0.3	-1.0 ± 0.3	-6.1 ± 0.3	-3.0 ± 0.3
$\Delta\Delta G_{vdW}$			-17.8 ± 0.1	-15.8 ± 0.1	-17.8 ± 0.1	-15.4 ± 0.1
$\Delta\Delta G_{FEP}$			**-20.7 ± 0.3**	**-16.8 ± 0.3**	**-23.9 ± 0.3**	**-18.4 ± 0.3**

Table 4. Summary of the results obtained for binding free energy with the different methods employed in this study, compared along with the experimental binding free energy (ΔG_{exp}) deduced from IC_{50} values. [25,26] All energies are reported in kcal·mol^{-1}.

	pIC_{50}	**ΔG_{exp}**	**$gscore_{docking}$**	**ΔG_{SMD}**	**$\Delta\Delta G_{FEP}$**
SalA	8.7	-12.4	-6.64	-21.5 ± 2.0	-23.9 ± 0.3
pSalA	6.0	-8.5	-6.87	-12.7 ± 2.2	-18.1 ± 0.3

From these results, it can be observed that docking is the only tool that does not reproduce the binding difference observed experimentally, rendering a better gscore value for pSalA (-6.87 kcal·mol^{-1}) than for SalA (-6.64 kcal·mol^{-1}). The SMD and FEP methodologies can reproduce the same trend observed from the deduced experimental values, although they differ in the absolute value of binding free energy. This can be caused by two reasons, an error of the methodology approximations or an error in the assumption that binding free energies can be directly deduced from IC_{50} values. Therefore, to completely understand the mode of action of pSalA, the inhibition mechanism is explored.

Reactivity calculations

As SalA reaction mechanism within the β5-active site of the 20S proteasome was characterized in our previous work,[49] the reaction mechanism of pSalA was explored following the same published mechanism. Since the reaction mechanism requires Lys33 to act as the proton acceptor in the first step, the reaction mechanism was explored only for the Model B-Lys33:NH_2. Since the binding step has been explored in the results presented above, the reactivity exploration for pSalA was initiated from a reactive structure obtained from an equilibrated classical MD simulation. Moreover, the overall interactions computed from 100ps of QM/MM MD simulations at AM1/MM level show that, in agreement with the previous results, the

interaction energy created between the protein and SalA (−54.1 ± 3.1 kcal·mol^{-1}) are much stronger than those created by pSalA (−37.7 ± 2.2 kcal·mol^{-1}). See supporting information for detailed contributions to the interaction energies for the residues in the β5 active site.

The inhibition reaction starts by activation of $O^{\gamma Thr1}$ by $N^{\zeta Lys33}$ with the consequent nucleophilic attack to the β-lactone carbonyl centre. This is followed by the opening of the β-lactone ring and finally an internal cyclization S_N2 reaction occurs leading to the formation of the tetrahydrofuran (THF) ring and release of chloride ion. To study this path two FESs were proposed, as done in previous studies.[49] The first and second steps were explored in one surface, where the proton transfer from $O^{\gamma Thr1}$ to $N^{\zeta Lys33}$ was controlled as the antisymmetric combination of distances in the X axis, while in the Y axis the antisymmetric combination of the nucleophilic attack, $O^{\gamma Thr1}$-C1 distance, and the β-lactone ring opening, C1-O2 distance, is controlled. Then the third step was explored in one surface where the nucleophilic attack of the O2 to the C3 is controlled in the X axis and the Cl-C3 bond breaking is controlled in the Y axis, as shown in Figure 7. All TS structures were optimized at the M06-2X/AMBER level of theory for each step of reaction (i.e., TS1, TS2, TS3 shown in Figure 8) and the course of the chemical transformations was confirmed by the positions of the optimized TSs on the quadratic regions of the corrected M06-2X:AM1/MM FESs (see Figure 7).

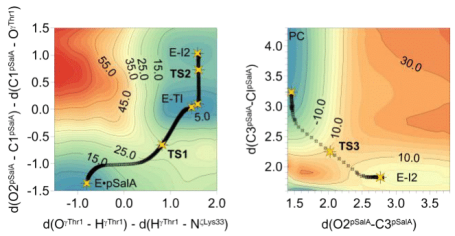


Figure 7. 2D-PMF surfaces computed at M06-2X/AM1/MM level of theory at 37 °C for the 20S proteasome inhibition with pSalA. Yellow stars on the surfaces indicate positions of M06-2X/MM-optimized structures, which are connected by MEP generated along the IRC calculations, presented as black stars. Distances on the axes are given in Å and iso-energetic lines in kcal·mol^{-1}.

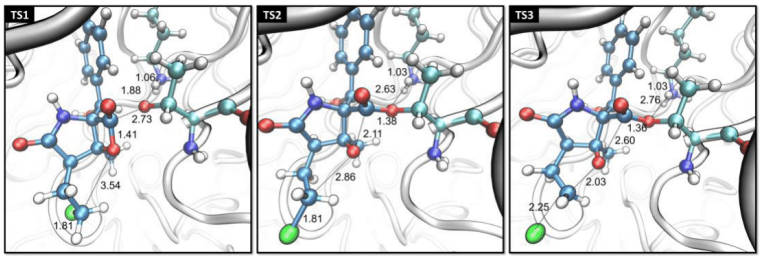


Figure 8. TS structures localized along the reaction pathway explored for inhibition ofthe β5 subunit of20S proteasome with pSalA computed at the M06-2X/AMBER level of theory.

The obtained results confirm that the inhibitory mechanism for pSalA is the same than the previously explored for SalA (see Scheme 1). The rate limiting step is in both cases the first step, that proceeds through a tetrahedral intermediate towards the second step, the opening of the β-lactone ring. Then the THP ring is formed concomitant to the release of the chloride atom. The obtained barriers, presented in Figure 9, show that the rate limiting step in the case of pSalA is

higher (25.1 kcal·mol^{-1}) than the obtained for SalA (20.4 kcal·mol^{-1}), showing that this inhibitor should overcome a more energetically demanding step to initiate the reactivity, added to the less stable binding free energy obtained in the previous section.

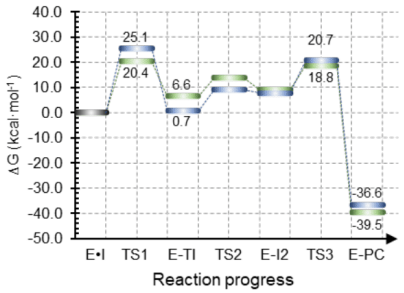


Figure 9. Free energy profile computed at the M06-2X:AM1/AMBER level for inhibition of the β5 subunit of 20S proteasome with SalA[49] (green line) and pSalA (blue line).

Surprisingly, the tetrahedral intermediate (E-TI) obtained for pSalA appears to be more stable (0.7 kcal·mol^{-1}) than in the case of SalA (6.6 kcal·mol^{-1}). While the rest of the computed free energy path is very close for both inhibitors.

Conclusions

The present work describes the full inhibition process of γ-lactam-β-lactone inhibitors in the 20S proteasome β5 active site through the computation of binding free energies with different computational methods, i.e., docking, SEM and FEP. Initially, docking was used to compute binding energies for a small data set of SalA analogues, which is characterized by the small differences between the studied molecules. The obtained results with docking show that, despite good results were expected for such small compounds, this tool is not able to reproduce the trend in binding energies observed experimentally. Therefore, despite docking being a good tool for virtual screening assays, when large libraries of very different compounds are tested, this tool is not adequate for addressing problems with such slight differences within the data set. Since from this result can be observed that the evaluation of punctual structures is not enough to evaluate the binding free energy differences of this type of compounds, enhanced sampling techniques were employed to improve the description of the binding process. The SMD method was employed at the MM level of theory to improve the previously obtained results, taking as a case of study the comparison between the best and the worst compound in the data set, SalA and pSalA. The obtained results with this methodology show that SalA has a more energetically demanding unbinding pathway than pSalA, reproducing the expected trend observed in experimental IC_{50} values. Although this tool provides only an approximate value for the unbinding PMF. Then, the FEP method was employed with QM/MM hybrid potentials to determine the binding free energies through a thermodynamic cycle. This tool gives results in agreement with SMD and the experimental binding energies, showing better binding free energy for SalA than for pSalA. The results obtained with the molecular dynamics employed methodologies show that for studying this type of inhibitors it is necessary to explore more than single structures and that the population of the binding process through molecular dynamics is necessary.
Finally, the inhibition mechanism for pSalA in the β5 active site of the 20S proteasome has been explored using multiscale QM/MM methods, to understand how IC_{50} values can be affected by both steps of the inhibition process, binding and reactivity. The obtained

results show that pSalA follows the same inhibition mechanism than the previously explored for SalA. This inhibitor shows that the first step is the rate limiting step, as it occurs with SalA, but in this case the free energy barrier obtained is higher, showing that this inhibitor must go through a more energetically demanding process.
The obtained results show that the effect of binding and reactivity on the inhibitory activity cannot be uncoupled for this type of covalent inhibitors, and free energy calculations and molecular dynamics are necessary for understanding the full inhibition process.
The results obtained in this work are an example of how computational studies can provide detailed information on the inhibition process of covalent inhibitors. Additionally, the knowledge gained from the analysis of the inhibition process can be essential for optimizing the inhibition of the 20S proteasome with inhibitors of the γ-lactam-β-lactone family and lead to new compounds in the drug design process.

Acknowledgements

We would like to thank the Spanish Ministerio de Ciencia e Innovación (grant PGC2018-094852-B-C21 and PID2019- 107098RJ-I00), the Generalitat Valenciana (grant AICO/2019/195 and SEJI/2020/007), and Universitat Jaume I (grant UJI-B2020-03 and UJI-A2019-04). N.S.A. thanks the MINECO for the doctoral FPI grant (BES-2016-078029). Authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I. The authors thankfully acknowledge the computer resources at Pirineus and the technical support provided by the Consorci de Serveis Universitaris de Catalunya (RES-QSB-2020-3-0012). This research also benefited from the HPC (High Performance Computing) facility of the University of Parma, Italy (http://www.hpc.unipr.it).

References

1 J. M. Strelow, *SLAS Discov. Adv. Sci. Drug Discov.*, 2016, **22**, 3–20.
2 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discov.*, 2011, **10**, 307–317.
3 T. A. Baillie, *Angew. Chemie - Int. Ed.*, 2016, **55**, 13408–13421.
4 D. C. Swinney, *Curr. Opin. Drug Discov. Devel.*, 2009, **12**, 31–39.
5 M. Castelli, S. A. Serapian, F. Marchetti, A. Triveri, V. Pirota, L. Torielli, S. Collina, F. Doria, M. Freccero and G. Colombo, *RSC Med. Chem.*, 2021, **12**, 1491–1502.
6 N. Singh and A. Warshel, *Proteins Struct. Funct. Bioinforma.*, 2010, **78**, 1705–1723.
7 A. Ivetac and J. Andrew McCammon, *Curr. Pharm. Des.*, 2011, **17**, 1663–1671.
8 K. M. Merz, *J. Chem. Theory Comput.*, 2010, **6**, 1769–1776.
9 A. R. Leach, B. K. Shoichet and C. E. Peishoff, *J. Med. Chem.*, 2006, **49**, 5851–5855.
10 C. Y. Yang, H. Sun, J. Chen, Z. Nikolovska-Coleska and S. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 13709–13721.
11 M. De Vivo and A. Cavalli, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2017, **7**, e1320.
12 H. Grubmüller, B. Heymann and P. Tavan, *Science*, 1996, **271**, 997–999.
13 B. Israelewitz, J. Baudry, J. Gullingsrud, D. Kosztin and K. Schulten, *J. Mol. Graph. Model.*, 2001, **19**, 13–25.
14 P. C. Do, E. H. Lee and L. Le, *J. Chem. Inf. Model.*, 2018, **58**, 1473–1482.
15 A. M. Capelli and G. Costantino, *J. Chem. Inf. Model.*, 2014, **54**, 3124–3136.
16 C. Jarzynski, *Phys. Rev. Lett.*, 1997, **78**, 2690–2693.
17 H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan and T. Hou, *Sci. Rep.*, 2015, **5**, 8457–8465.
18 R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni and M. Parrinello, *J. Am. Chem. Soc.*, 2017, **139**, 4780–4788.
19 J. Z. Wu, S. Azimi, S. Khuttan, N. Deng and E. Gallicchio, *J. Chem. Theory Comput.*, 2021, **17**, 3309–3319.
20 K. Świderek, S. Martí and V. Moliner, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12614–12624.
21 A. Tuley and W. Fast, *Biochemistry*, 2018, **57**, 3326–3337.
22 D. C. Swinney, in *Annual Reports in Medicinal Chemistry*, 2011, vol. 46.
23 S. L. McGovern, E. Caselli, N. Grigorieff and B. K. Shoichet, *J. Med. Chem.*, 2002, **45**, 1712–1722.
24 B. K. Shoichet, *Drug Discov. Today*, 2006, **11**, 607–615.
25 C. J. Woods, M. Malaisree, S. Hannongbua and A. J. Mulholland, *J. Chem. Phys.*, 2011, **134**, 054114.
26 R. C. Rizzo, M. Udier-Blagović, D.-P. P. Wang, E. K. Watkins, M. B. Kroeger Smith, R. H. Smith, J. Tirado-Rives, W. L. Jorgensen, Smith Richard H., J. Tirado-Rives and W. L. Jorgensen, *J. Med. Chem.*, 2002, **45**, 2970–2987.
27 M. Mor, A. Lodola, S. Rivara, F. Vacondio, A. Duranti, A. Tontini, S. Sanchini, G. Piersanti, J. R. Clapper, A. R. King, G. Tarzia and D. Piomelli, *J. Med. Chem.*, 2008, **51**, 3487–3498.
28 J. S. Patel, A. Berteotti, S. Ronsisvalle, W. Rocchia and A. Cavalli, *J. Chem. Inf. Model.*, 2014, **54**, 470–480.
29 Y. Sheng, H. Watanabe, K. Maruyama, C. Watanabe, Y. Okiyama, T. Honma, K. Fukuzawa and S. Tanaka, *Comput. Struct. Biotechnol. J.*, 2018, **16**, 421–434.
30 D. H. Wolf and W. Hilt, *Biochim. Biophys. Acta - Mol. Cell Res.*, 2004, **1695**, 19–31.
31 E. E. Manasanch, N. Korde, A. Zingone, N. Tageja, C. Fernandez de Larrea, M. Bhutani, P. Wu, M. Roschewski and O. Landgren, *Leuk. Lymphoma*, 2014, **55**, 1707–1714.
32 A. Eisenberg-Lerner, R. Benyair, N. Hizkiahou, N. Nudel, R. Maor, M. P. Kramer, M. D. Shmueli, I. Zigdon, M. Cherniavsky Lev, A. Ulman, J. Y. Sagiv, M. Dayan, B. Dassa, M. Rosenwald, I. Shachar, J. Li, Y. Wang, N. Dezorella, S. Khan, Z. Porat, E. Shimoni, O. Avinoam and Y. Merbl, *Nat. Commun.*, 2020, **11**, 409–423.
33 J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister and R. Huber, *Science*, 1995, **268**, 533–539.
34 M. Groll, L. Ditzel, J. Löwe, D. Stock, M. Bochtler, H. D. Bartunik and R. Huber, *Nature*, 1997, **386**, 463–471.
35 E. Seemüller, A. Lupas, D. Stock, J. Löwe, R. Huber and W. Baumeister, *Science*, 1995, **268**, 579–582.
36 T. A. Thibaudeau and D. M. Smith, *Pharmacol. Rev.*, 2019, **71**, 170–197.
37 V. J. Palombella, O. J. Rando, A. L. Goldberg and T. Maniatis, *Cell*, 1994, **78**, 773–785.
38 A. L. Goldberg, *Biochem. Soc. Trans.*, 2007, **35**, 12–17.
39 A. F. Kisselev, W. A. Van Der Linden and H. S. Overkleeft, *Chem. Biol.*, 2012, **19**, 99–115.
40 Q. P. Dou and R. H. Goldfarb, *IDrugs*, 2002, **5**, 828–34.
41 P. Efentakis, G. Kremastiotis, A. Varela, P. E. Nikolaou, E. D. Papanagnou, C. H. Davos, M. Tsoumani, G. Agrogiannis, A. Konstantinidou, E. Kastritis, Z. Kanaki, E. K. Iliodromitis, A. Klinakis, M. A. Dimopoulos, I. P. Trougakos, I. Andreadou and E. Terpos, *Blood*, 2019, **133**, 710–723.
42 L. A. Raedler, *Am. Heal. Drug Benefits*, 2016, **9**, 102–105.
43 A. Besse, L. Besse, M. Kraus, M. Mendez-Lopez, J. Bader, B. T. Xin, G. de Bruin, E. Maurits, H. S. Overkleeft and C. Driessen, *Cell Chem. Biol.*, 2019, **26**, 340-351.e3.
44 T. A. M. Gulder and B. S. Moore, *Angew. Chemie Int. Ed.*, 2010, **49**, 9346–9367.
45 H. Tomoda and S. Omura, *Yakugaku Zasshi*, 2000, **120**, 935–949.
46 M. Groll, R. Huber and B. C. M. Potts, *J. Am. Chem. Soc.*, 2006, **128**, 5136–5141.

47 B. C. M. Potts, M. X. Albitar, K. C. Anderson, S. Baritaki, C. Berkers, B. Bonavida, J. Chandra, D. Chauhan, J. C. Cusack Jr., W. Fenical, I. M. Ghobrial, M. Groll, P. R. Jensen, K. S. Lam, G. K. Lloyd, W. McBride, D. J. McConkey, C. P. Miller, S. T. C. Neuteboom, Y. Oki, H. Ovaa, F. Pajonk, P. G. Richardson, A. M. Roccaro, C. M. Sloss, M. A. Spear, E. Valashi, A. Younes and M. A. Palladino, *Curr. Cancer Drug Targets*, 2011, **11**, 254–284.
48 M. Nett, T. A. M. Gulder, A. J. Kale, C. C. Hughes and B. S. Moore, *J. Med. Chem.*, 2009, **52**, 6163–6167.
49 N. Serrano-Aparicio, V. Moliner and K. Świderek, *ACS Catal.*, 2021, **11**, 3575–3589.
50 P. Ertl, B. Rohde and P. Selzer, *J. Med. Chem.*, 2000, **43**, 3714–3717.
51 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2012, 64.
52 D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615–2623.
53 A. Daina, O. Michielin and V. Zoete, *Sci. Rep.*, 2017, **7**, 1–13.
54 N. Serrano-Aparicio, V. Moliner and K. Świderek, *ACS Catal.*, 2021, 11806–11819.
55 Schrödinger Release 2019-3: Maestro, Schrödinger,.
56 E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.*, 2016, **12**, 281–296.
57 Schrödinger Release 2019-3: MacroModel, Schrödinger.
58 Schrödinger Release 2019-3: Glide, Schrödinger, LL.
59 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177–6196.
60 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J. Med. Chem.*, 2004, **47**, 1750–1759.
61 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
62 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
63 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.
64 Schrödinger Release 2019-3: Prime, Schrödinger, LL.
65 M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw and R. A. Friesner, *Proteins Struct. Funct. Genet.*, 2004, **55**, 351–367.
66 M. P. Jacobson, R. A. Friesner, Z. Xiang and B. Honig, *J. Mol. Biol.*, 2002, **320**, 597–608.
67 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. A. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.
68 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
69 E. Lindahl, B. Hess and D. van der Spoel, *J. Mol. Model.*, 2001, 7, 306–317.
70 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
71 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
72 S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.
73 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
74 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
75 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, A. White and T. P. consortium, *Nat. Methods*, 2019, **16**, 670–673.
76 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
77 N. Serrano-Aparicio, K. Świderek and V. Moliner, *Eur. J. Med. Chem.*, 2019, **164**, 399–407.
78 J. Schrader, F. Henneberg, R. A. Mata, K. Tittmann, T. R. Schneider, H. Stark, G. Bourenkov and A. Chari, *Science*, 2016, **353**, 594–598.
79 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.
80 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
81 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
82 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.

83 J. Baker, *J. Comput. Chem.*, 1986, **7**, 385–395.

84 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

85 M. J. Field, M. Albe, C. Bret, F. Proust-De Martin and A. Thomas, *J. Comput. Chem.*, 2000, **21**, 1088–1100.

86 A. Krzemińska, P. Paneth, V. Moliner and K. Świderek, *J. Phys. Chem. B*, 2015, **119**, 917–927.

87 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

88 J. J. Ruiz-Pernía, E. Silla, I. Tuñón and S. Martí, *J. Phys. Chem. B*, 2006, **110**, 17663–17670.

89 Y.-Y. Chuang, J. C. Corchado and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 1140–1149.

90 K. Świderek, I. Tuñón, S. Martí and V. Moliner, *ACS Catal.*, 2015, **5**, 1172–1185.

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

Bibliography

(1) Groll, M.; Kim, K. B.; Kairies, N.; Huber, R.; Crews, C. M. Crystal Structure of Epoxomicin:20S Proteasome Reveals a Molecular Basis for Selectivity of α',β'-Epoxyketone Proteasome Inhibitors. *J. Am. Chem. Soc.* **2000**, *122*, 1237–1238.

(2) Schrader, J.; Henneberg, F.; Mata, R. A.; Tittmann, K.; Schneider, T. R.; Stark, H.; Bourenkov, G.; Chari, A. The Inhibition Mechanism of Human 20S Proteasomes Enables Next-Generation Inhibitor Design. *Science* **2016**, *353*, 594–598.

(3) Manam, R. R.; McArthur, K. A.; Chao, T.-H.; Weiss, J.; Ali, J. A.; Palombella, V. J.; Groll, M.; Lloyd, G. K.; Palladino, M. A.; Neuteboom, S. T. C.; Macherla, V. R.; Potts, B. C. M. Leaving Groups Prolong the Duration of 20S Proteasome Inhibition and Enhance the Potency of Salinosporamides. *J. Med. Chem.* **2008**, *51*, 6711–6724.

(4) Groll, M.; Huber, R.; Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of β-Lactone Ring Opening and a Mechanism for Irreversible Binding. *J. Am. Chem. Soc.* **2006**, *128*, 5136–5141.

(5) Groll, M.; Nguyen, H.; Vellalath, S.; Romo, D. (−)-Homosalinosporamide A and Its Mode of Proteasome Inhibition: An X-Ray Crystallographic Study. *Mar. Drugs* **2018**, *16*, 240–249.

(6) Huber, E. M.; Heinemeyer, W.; Li, X.; Arendt, C. S.; Hochstrasser, M.; Groll, M. A Unified Mechanism for Proteolysis and Autocatalytic Activation in the 20S Proteasome. *Nat. Commun.* **2016**, *7*, 10900.

(7) Potts, B. C. M.; Albitar, M. X.; Anderson, K. C.; Baritaki, S.; Berkers, C.; Bonavida, B.; Chandra, J.; Chauhan, D.; Cusack Jr., J. C.; Fenical, W.; Ghobrial, I. M.; Groll, M.; Jensen, P. R.; Lam, K. S.; Lloyd, G. K.; McBride, W.; McConkey, D. J.; Miller, C. P.; Neuteboom, S. T. C.; et al. Marizomib, a Proteasome Inhibitor for All Seasons: Preclinical Profile and a Framework for Clinical Trials. *Curr. Cancer Drug Targets* **2011**, *11*, 254–284.

(8) Groll, M.; Potts, B. C. Proteasome Structure , Function , and Lessons Learned from Beta-Lactone Inhibitors. *Curr. Top. Med. Chem.* **2011**, *11*, 2850–2878.

(9) Bürgi, H. B.; Dunitz, J. D.; Shefter, E. Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.* **1973**, *95*, 5065–5067.

(10) Bürgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G. Stereochemistry of

Reaction Paths at Carbonyl Centres. *Tetrahedron* **1974**, *30*, 1563–1572.

(11) Fleming, I. *Molecular Orbitals and Organic Chemical Reactions*; Wiley Online Books; Wiley, 2010.

(12) Heathcock, C. H. Understanding and Controlling Diastereofacial Selectivity in Carbon-Carbon Bond-Forming Reactions. *Aldrichimica Acta* **1990**, *23*, 94–111.

(13) Nett, M.; Gulder, T. A. M.; Kale, A. J.; Hughes, C. C.; Moore, B. S. Function-Oriented Biosynthesis of β-Lactone Proteasome Inhibitors in Salinispora Tropica. *J. Med. Chem.* **2009**, *52*, 6163–6167.

(14) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

(15) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(16) Capelli, A. M.; Bruno, A.; Entrena Guadix, A.; Costantino, G. Unbinding Pathways from the Glucocorticoid Receptor Shed Light on the Reduced Sensitivity of Glucocorticoid Ligands to a Naturally Occurring, Clinically Relevant Mutant Receptor. *J. Med. Chem.* **2013**, *56*, 7003–7014.

(17) Capelli, A. M.; Costantino, G. Unbinding Pathways of VEGFR2 Inhibitors Revealed by Steered Molecular Dynamics. *J. Chem. Inf. Model.* **2014**, *54*, 3124–3136.

Chapter 5. Conclusions and future perspectives

In the present thesis, the inhibition mechanism of the 20S proteasome with two main types of covalent inhibitors has been explored at the molecular level via multiscale QM/MM MD simulations. The 20S proteasome is approved as a valuable target for inhibition in the treatment of several types of diseases, with cancer among them. In particular, the tested inhibitors selected represent two of the main types of pharmacophores employed to target this specific macromolecular complex enzyme, intending to gain an understanding of how the inhibitory mechanism works at the molecular level for each type. Then all the obtained knowledge can be applied to the drug design process and help the optimization of lead compounds in drug discovery.

From the biological point of view, the studies of the inhibition processes of the 20S proteasome performed in this thesis have led to the following main conclusions:

- The studies of the inhibition of 20S proteasome with a peptide-like α,β-epoxyketone inhibitor, dihydroeponemicin, show a new mechanism responsible for the inhibition process, contrary to the proposed mechanism previously in the literature. This mechanism explains the most recent X-ray structure for this type of warhead, showing that the reaction starts with an epoxide-ring opening and finishes with 1,4-oxazepane product formation. The computed free energy barrier for this process is in very good agreement with the experimental rate constant found in literature and shows that the seven-membered ring product is both kinetically and thermodynamically favored compared to the formation of the six-membered ring product. Finally, the computed primary and secondary KIEs give an additional magnitude that could allow distinguishing between the formation of both possible products.
- The first study on the non-peptidic γ-lactam-β-lactone inhibitors was focused on SalA. Through interaction analysis and MD simulations it was shown that in addition to the S1 binding pocket, which architecture is responsible for the character of the active site, the S2 pocket can lead to crucial interactions that stabilize and orient the initial binding structure in this type of inhibitors. The most accepted mechanism where SalA acts as the proton acceptor in the first step renders too high energy barrier. Contrarily, the

computed free energy barrier for the inhibition mechanism where Lys33 acts as the initial proton acceptor is in very good agreement with experimental kinetic measurements and shows that the inhibition process goes through a three steps mechanism where in the first step the protein can stabilize a tetrahedral intermediate and explains the irreversible character of SalA with the obtained high exergonic final product. This result was supported by the analysis of the computed protein electrostatic potential that leads to the validation of the Lys33-Asp17 dyad as responsible for the activation of Thr1 and the reactivity for this type of inhibitor. The simulations show that the protein has certain flexibility that facilitates the electrostatic stabilization of the reaction TSs and intermediates to the final product.

- The study focused on the reversible SalA analog, hSalA, which was designed to explain its observed alternative behavior despite the identical measured experimentally IC_{50} values. The exploration of the FES shows that this inhibitor follows the same two first steps and has the first step as rate-limiting, as it was observed for SalA. Nevertheless, the results of calculations indicated that the final cyclization step is too energetically demanding in hSalA due to the inevitable conformational change in the chloro-propyl moiety that precedes the THP ring closure. This inhibitor endures an additional double proton transfer step necessary for neutralizing the charges in the structure of the final product. Then the reversible character was examined through the exploration of the ester-linkage hydrolysis process, which rendered a high energy barrier. These results suggest that the nature of the reversibility of hSalA originates in the possibility of regressing the inhibition steps. Such mode of action allows classifying this inhibitor as a covalent reversible inhibitor. Finally, these results show that the identical experimental IC_{50} values for SalA and hSalA are affected by both inhibitors having close binding properties and rate-limiting step barriers.
- The study of the binding process of SalA and pSalA into the β5 active site of the 20S proteasome intends to explain the differences in binding free energy of these two close analogs with very different inhibitory activity. Molecular docking showed an unexpected result for these two small molecules since the gscore scoring function was

not able to rank these compounds in the same way as the observed in experimental assays. Therefore, binding free energies were computed by employing two additional tools. By the enhanced sampling technique SMD, that provided results in good agreement with the experimental inhibitory activities, showing that the unbinding SalA from the active site is more energetically demanding than pSalA. Then binding free energies were computed using QM/MM potentials by the FEP method. These calculations showed that SalA has stronger binding energy than pSalA. Since both techniques are based on molecular dynamic simulations, these results show that computing binding free energies with techniques that allow exploring enzyme conformation is crucial in cases where the compared inhibitors are very similar. Finally, the inhibition mechanism of pSalA has been explored based on the mechanism validated for SalA. The results show a free energy barrier for the rate-limiting step higher than the one obtained for the inhibition with SalA. The weaker binding of the pSalA reflected by results from binding free energy calculations together with slower chemical transformation than those observed for SalA show certainly that both the binding and the chemical reaction step must be considered when comparing computational results with experimental IC_{50} values in the case of covalent inhibitors, as it was shown for the case of SalA and hSalA.

Computational techniques have proven to have a crucial role in providing information of the explored processes at the atomistic level, not accessible with experimental assays. The selected methods and models have allowed monitoring the evolution of the structures along the full inhibition paths, and to obtain the complete free-energy landscape. All the obtained data allows the analysis and rationalization of the inhibition process in detail, comparable with available experimental kinetic and thermodynamic data. From the technical point of view, a broad set of computational techniques have been applied and an adequate combination of them can be used in future studies, including the refinement of new compounds to control the action of the 20S proteasome. The main technical conclusions derived from this thesis comprise the following:

- Protein electrostatic potential has proven to have a key role in the enzyme process. Therefore, computation of this magnitude allows explaining how the enzyme is predisposed to carry certain tasks, as well as to test its possible inhibition.
- All structural analysis was done along the full inhibition paths over structures computed at a high level of theory (M06-2X/6-31G+(d,p)//AMBER). A deep understanding of the inhibition process is gained from the evolution of key distances and angles along the reaction pathway.
- The computation of KIE has shown to be a very valuable analysis tool that allows distinguishing between different mechanisms and can be further compared with experimental data. Thus, both types of experiments can be linked and experimental KIEs results can be understood from the provided computational atomistic details.
- The computation of binding energies represents a challenging task that can be faced with very different types of techniques. Fast binding energy results can be obtained with molecular docking tools when scanning large data sets, but when more precise descriptions are necessary, they can lead to errors in ranking the binding energies of congeneric compounds. Moreover, the evaluation of binding energies from punctual structures is not enough to obtain definite results and conformational changes in the inhibitor within the active site that can be explored along MD simulations must be considered. Therefore, enhanced sampling techniques such as SMD simulations are shown to lead to better qualitative results. Nevertheless, the absolute values of the binding free energy depend on the selected simulation parameters and the number of computing replicas generated to obtain an averaged value. On the other hand, employing FEP methods with hybrid QM/MM potentials gives more precise results, with the additional advantage of not requiring a preceding calibration process.

The results obtained in this thesis open the path to future studies on 20S proteasome inhibition, where the exploration of the inhibitory mechanism of other types of inhibitors containing different warheads, such as boronic acids, ketoaldehydes, or vinyl-sulfones, can provide alternative routes for the design of new inhibitors. The understanding of the mode of action of

different pharmacophores can help in the design of new drugs with potential medical applications. On the other hand, the study of the inhibition of the other catalytically active sites of the 20S proteasome can increase the possibilities of selectively modulating the activity of this macromolecular protein complex.

APPENDIX. List of abbreviations

ACE – Angiotensin converting enzyme

ADME – Absorption, distribution, metabolism, and elimination properties

AM1 – Austin Model 1

ATP – Adenosine triphosphate

BCL2 – B-cell lymphoma 2

CC – Coupled cluster

cCP – constitutive proteasome core particle

CI – Configuration interaction

CNS – Central nervous system

COPD – Chronic obstructive pulmonary disease

CP – Core particle

DFT – Density functional theory

EC – Enzyme nomenclature commission

EGF – Epidermal growth factor

EVB – Empirical valence bond

FDA – Food and drug administration

FEP – Free energy perturbation

FES – Free energy surface

GGA – Generalized gradient approximation

HF – Hartree-Fock

HIV – Human immunodeficiency virus

HK – Hohenberg-Kohn

HL – High-level

hSalA – homo-salinosporamide A

I – Intermediate

iCP – immunoproteasome proteasome core particle

IRC – Intrinsic reaction coordinate

KIE – Kinetic isotope effects

KS – Kohn-Sham

LDA – Local density approximation

LL – Low-level

LSDA – Local spin-density approximation

MC – Monte Carlo

MD – Molecular dynamics

MEP – Minimum energy path

MNDO – Modified neglect of diatomic overlap

MM – Molecular mechanics

MMM – Multiple myeloma

MP – Møller and Plesset

NDDO – Neglect of diatomic differential overlap

NF$\varkappa$B – Nuclear factor kappa-light-chain-enhancer of activated B cells

P – Product

PDB – Protein data bank

PES – Potential energy surface

PMF – Potential of mean force

pSalA – C5-phenyl-salinosporamide A

QCI – Quadratic configuration interaction

QM – Quantum mechanics

QM/MM – Quantum mechanics / molecular mechanics

QSAR – Quantitative structure activity relationship

R – Reactant

RMSD – Root mean square deviation

RNA – Ribonucleic acid

RP – Regulatory particle

RRMM – Relapsed and refractory multiple myeloma

RS – Reactant state

SalA – Salinosporamide A

SCF – Self-consistent field

SE – Semiempirical

SEM – Standard error of the mean

SMD – Steered molecular dynamics

TACE – Tumor necrosis factor Alpha activating enzyme

tCP – thymoproteasome proteasome core particle

TS – Transition state

TST – Transition state theory

UPS – Ubiquitin-proteasome system

US – Umbrella sampling

WHAM – Weighted histogram analysis method

ZPE – Zero point energy

ANNEX 1. Co-author's consent documents

ANNEX 1

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Castellón de la Plana, 1st December 2021

I, Katarzyna Świderek, hereby authorise Natalia Serrano Aparicio to include the publications listed below in his doctoral thesis. In addition, I waive the right to use those articles as part of any other doctoral thesis.

List of articles:

(1) Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, *164*, 399–407.

(2) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, *11*, 3575–3589.

(3) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. On the Origin of the Different Reversible Characters of Salinosporamide A and Homosalinosporamide A in the Covalent Inhibition of the Human 20S Proteasome. *ACS Catal.* **2021**, 11806–11819.

(4) Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog. Under preparation for submission.

Signed,

KATARZYNA PATRYCJA| SWIDEREK

Elektronicznie podpisany przez KATARZYNA PATRYCJA|SWIDEREK
Data: 2021.12.08 10:31:25 +01'00'

In accordance with article 28 of the Regulations on doctoral studies of the Universitat Jaume I in Castelló, *regulated by RD 99/2011, at the Universitat Jaume I (*Approved by the Governing Council at its meeting no. 8/2020 held on 2 October 2020*):*
"(...)
4. In the case of joint publications, all the co-authors must explicitly state their approval that the doctoral student presented the work as part of her/his thesis and the express waiver of presenting this same work as part of another doctoral thesis. This authorisation must be attached as documentation when the evaluation of the thesis begins."

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Castellón de la Plana, 1st December 2021

I, Vicent Moliner, hereby authorise Natalia Serrano Aparicio to include the publications listed below in his doctoral thesis. In addition, I waive the right to use those articles as part of any other doctoral thesis.

List of articles:

(1) Serrano-Aparicio, N.; Świderek, K.; Moliner, V. Theoretical Study of the Inhibition Mechanism of Human 20S Proteasome by Dihydroeponemycin. *Eur. J. Med. Chem.* **2019**, *164*, 399–407.

(2) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis. *ACS Catal.* **2021**, *11*, 3575–3589.

(3) Serrano-Aparicio, N.; Moliner, V.; Świderek, K. On the Origin of the Different Reversible Characters of Salinosporamide A and Homosalinosporamide A in the Covalent Inhibition of the Human 20S Proteasome. *ACS Catal.* **2021**, 11806–11819.

(4) Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog. Under preparation for submission.

Signed,

VICENTE MOLINER IBAÑEZ - NIF:25388050Y

Firmado digitalmente por VICENTE MOLINER IBAÑEZ - NIF:25388050Y
Fecha: 2021.12.08 10:36:10 +01'00'

In accordance with article 28 of the Regulations on doctoral studies of the Universitat Jaume I in Castelló*, regulated by RD 99/2011, at the Universitat Jaume I (*Approved by the Governing Council at its meeting no. 8/2020 held on 2 October 2020*):*
"(...)
4. In the case of joint publications, all the co-authors must explicitly state their approval that the doctoral student presented the work as part of her/his thesis and the express waiver of presenting this same work as part of another doctoral thesis. This authorisation must be attached as documentation when the evaluation of the thesis begins."

ANNEX 1

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Parma, 1st December 2021

I, Laura Scalvini, hereby authorise Natalia Serrano Aparicio to include the publications listed below in his doctoral thesis. In addition, I waive the right to use those articles as part of any other doctoral thesis.

List of articles:

(1) Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog. Under preparation for submission.

Signed,

Laura Scalvini

In accordance with article 28 of the Regulations on doctoral studies of the Universitat Jaume I in Castelló*, regulated by RD 99/2011, at the Universitat Jaume I (*Approved by the Governing Council at its meeting no. 8/2020 held on 2 October 2020*):*
"(...)
4. In the case of joint publications, all the co-authors must explicitly state their approval that the doctoral student presented the work as part of her/his thesis and the express waiver of presenting this same work as part of another doctoral thesis. This authorisation must be attached as documentation when the evaluation of the thesis begins."

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Parma, 1st December 2021

I, Alessio Lodola, hereby authorise Natalia Serrano Aparicio to include the publications listed below in his doctoral thesis. In addition, I waive the right to use those articles as part of any other doctoral thesis.

List of articles:

(1) Serrano-Aparicio, N.; Scalvini, L.; Lodola, A.; Moliner, V.; Świderek, K. Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analog. Under preparation for submission.

Signed,

In accordance with article 28 of the Regulations on doctoral studies of the Universitat Jaume I in Castelló, *regulated by RD 99/2011, at the Universitat Jaume I (*Approved by the Governing Council at its meeting no. 8/2020 held on 2 October 2020*):*
"(...)
4. In the case of joint publications, all the co-authors must explicitly state their approval that the doctoral student presented the work as part of her/his thesis and the express waiver of presenting this same work as part of another doctoral thesis. This authorisation must be attached as documentation when the evaluation of the thesis begins."

ANNEX 2. Supporting Information to the papers

4.1 Peptide like epoxyketone in 20S proteasome

Supporting Information

Theoretical Study of the Inhibition Mechanism of *Human* 20S Proteasome by Dihydroeponemycin

Natalia Serrano-Aparicio, Katarzyna Świderek,* Vicent Moliner*

Departamento de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain.

Corresponding authors:

Katarzyna Świderek: swiderek@uji.es

Vicent Moliner: moliner@uji.es

Computational Details

The potential energy of our scheme is derived from the standard additive scheme of QM/MM formulation:

$$E_{QM-MM} = \langle \Psi | \widehat{H_o} | \Psi \rangle + \left(\sum \left\langle \Psi \left| \frac{q_{MM}}{r_{e,MM}} \right| \Psi \right\rangle + \sum \sum \frac{Z_{QM} q_{MM}}{r_{QM,MM}} \right) + E_{QM-MM}^{vdW} + E_{MM} \quad \text{(S1)}$$

$$E_{QM-MM} = E_{vac} + E_{QM-MM}^{elect} + E_{QM-MM}^{vdW} + E_{MM} \quad \text{(S2)}$$

where E_{MM} is the energy of the MM subsystem term, $E_{QM-MM}^{vdW}$the van der Waals interaction energy between the QM and MM subsystems and E_{QM-MM}^{elect} includes both the Coulombic interaction of the QM nuclei (Z_{QM}) and the electrostatic interaction of the polarized electronic wave function (Ψ) with the charges of the protein (q_{MM}). According to this scheme, the interaction energy between the substrate and the environment is evaluated as the difference between the QM/MM energy and the energies of the separated, non-interacting, QM and MM subsystems with the same geometry. Considering that the MM part is described using a non-polarizable potential, the interaction energy contribution of each residue of the protein is given by the following expression:

$$E_{QM-MM}^{Int} = \sum \left\langle \Psi \left| \frac{q_{MM}}{r_{e,MM}} \right| \Psi \right\rangle + \sum \sum \frac{Z_{QM} q_{MM}}{r_{QM,MM}} + E_{QM-MM}^{vdW} \quad \text{(S3)}$$

This interaction energy can be exactly decomposed in a sum over residues provided that the polarized wave function (Ψ) is employed to evaluate this energy contribution. The global polarization effect can be obtained from the gas phase energy difference between the polarized, Ψ, and non-polarized, Ψ_0, wave functions.

Potential of Mean Force. The procedure for the PMF calculation is straightforward and requires a series of molecular dynamics simulations in which the distinguished reaction coordinate variable, ξ, is constrained around particular values. The values of the variables sampled during the simulations are then pieced together to construct a distribution function from which the PMF is obtained as a function of the distinguished reaction coordinate (W(ξ)).The PMF is related to the normalized probability of finding the system at a particular value of the chosen coordinate by eq S4:

$$W(\xi) = C - kT ln \int \rho(r^N) \delta(\xi(r^N) - \xi) dr^{N-1} \quad \text{(S4)}$$

The activation free energy can be then expressed as:

$$\Delta G^{\ddagger}(\xi) = W(\xi^{\ddagger}) - [W(\xi^R) + G_{\xi}(\xi^R)] \quad \text{(S5)}$$

where the superscripts indicate the value of the reaction coordinate at the reactants (R), and at the TS (‡), and G_ξ (ξ^R) is the free energy associated with setting the reaction coordinate to a specific value at the reactant state. Normally this last term makes a small contribution and the activation free energy is directly estimated from the PMF change between the maximum of the profile and the reactant's minimum:

$$\Delta G^{\ddagger}(\xi) \approx W(\xi^{\ddagger}) - W(\xi^{R}) = \Delta W^{\ddagger}(\xi) \quad \text{(S6)}$$

The selection of the reaction coordinate is usually trivial when the mechanism can be driven by a single internal coordinate or a simple combination (as the antisymmetric combination of two interatomic distances). However this is not the case for all possible steps of the reaction subject of study in this paper where many coordinates are participating. Instead we were compelled to obtain a much more computationally demanding 2D-PMF using two coordinates: ξ_1 and ξ_2. The 2D-PMF is related to the probability of finding the system at particular values of these two coordinates:

$$W(\xi) = C' - kTln \int \rho(r^N)\delta(\xi_1(r^N) - \xi_1)\delta(\xi_2(r^N) - \xi_2)dr^{N-2} \quad \text{(S7)}$$

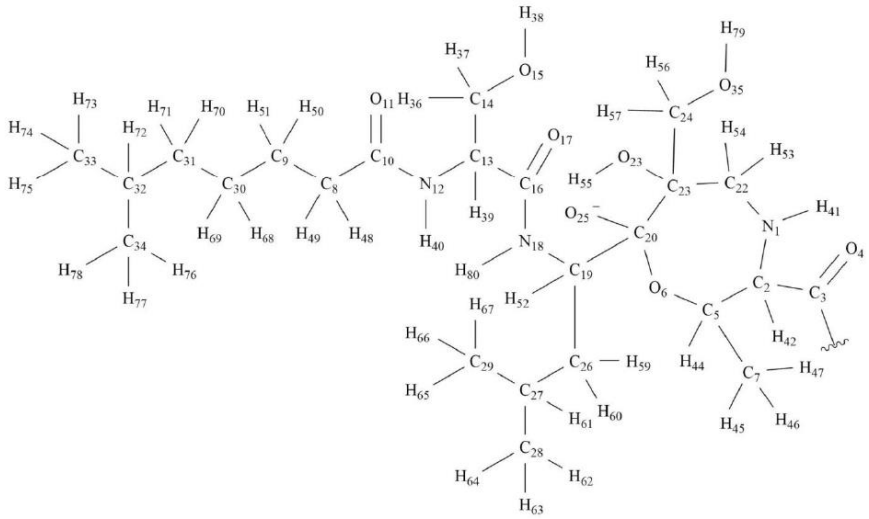
To estimate the activation free energy from this quantity, we recovered one-dimensional PMF changes tracing a maximum probability reaction path on the 2D-PMF surface and integrating over the perpendicular coordinate.

Spline corrections. Because of the large number of structures that must be evaluated during free energy calculations, QM/MM calculations are usually restricted to the use of semiempirical Hamiltonians. In order to reduce the errors associated with the quantum level of theory employed in our simulations, a new energy function defined in terms of interpolated corrections was used:[1,2,3]

$$E = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\xi_1, \xi_2)] \quad \text{(S8)}$$

where S denotes a two-dimensional spline function, and its argument is a correction term evaluated from the single-point energy difference between a high-level (HL) and a low-level (LL) calculation of the QM subsystem. The AM1 semiempirical Hamiltonian was used as LL method, while a density functional theory (DFT)-based method was selected for the HL energy calculation. In particular, HL energy calculations were performed by means of the hybrid M06-2X functional using the standard 6-31+G(d,p) basis set. These calculations were carried out using the Gaussian09 program.[4]

Table S1. Atom types, charges and parameters obtained for inhibitor covalently bound to Thr1 of protein were computed using Antechamber software.[5]



Atom name	Atom type	Charge	Atom name	Atom type	Charge
N1	NT	-0.822120	O11	O	-0.648826
C22	CT	0.132651	H40	H	0.334125
C21	CT	0.053740	C14	CT	0.112274
C20	CT	0.551681	O15	OH	-0.614488
O6	OS	-0.523186	H38	HO	0.415534
C5	CT	0.156924	H36	H1	0.070121
C7	CT	-0.098210	H37	H1	0.070121
H45	HC	0.027669	H39	H1	0.081608
H46	HC	0.027669	O17	O	-0.662842
H47	HC	0.027669	H80	H	0.331128
H44	H1	0.087602	C26	CT	-0.068477
C19	CT	0.055638	C27	CT	-0.066775
N18	N	-0.489448	C28	CT	-0.090201
C16	C	0.643378	H62	HC	0.026503
C13	CT	-0.009310	H63	HC	0.026503
N12	N	-0.547513	H64	HC	0.026503
C10	C	0.664355	C29	CT	-0.090201
C8	CT	-0.151570	H65	HC	0.026503
C9	CT	-0.079489	H66	HC	0.026503
C30	CT	-0.077487	H67	HC	0.026503
C31	CT	-0.078488	H61	HC	0.043651
C32	CT	-0.067776	H59	HC	0.060133
C33	CT	-0.089700	H60	HC	0.060133
H73	HC	0.031831	H52	H1	0.058634
H74	HC	0.031831	O25	O	-0.820418
H75	HC	0.031831	O23	OH	-0.608481
C34	CT	-0.089700	H55	HO	0.386566
H76	HC	0.031831	C24	CT	0.108279

H77	HC	0.031831	O35	OH	-0.636513
H78	HC	0.031831	H79	HO	0.404546
H72	HC	0.043651	H56	H1	0.058634
H70	HC	0.043651	H57	H1	0.058634
H71	HC	0.043651	H53	H1	0.046148
H68	HC	0.039156	H54	H1	0.046148
H69	HC	0.039156	H41	H	0.355401
H50	HC	0.052641	C2	CT	0.032464
H51	HC	0.052641	H42	H1	0.071620
H48	HC	0.066625	C3	C	0.563268
H49	HC	0.066625	O4	O	-0.564732

Parameters:

MASS

NT	14.010	0.530	same as n3
CT	12.010	0.878	same as c3
OS	16.000	0.465	same as os
HC	1.008	0.135	same as hc
H1	1.008	0.135	same as hc
N	14.010	0.530	same as n
C	12.010	0.616	same as c
O	16.000	0.434	same as o
H	1.008	0.161	same as hn
OH	16.000	0.465	same as oh
HO	1.008	0.135	same as ho

BOND

NT-CT	325.90	1.465	same as c3-n3
NT-H	392.40	1.019	same as hn-n3
CT-CT	300.90	1.538	same as c3-c3
CT-H1	330.60	1.097	same as c3-hc
CT-OH	316.70	1.423	same as c3-oh
CT-OS	308.60	1.432	same as c3-os
CT-O	449.90	1.317	same as c3-o
CT-HC	330.60	1.097	same as c3-hc
CT-N	328.70	1.462	same as c3-n
N -C	427.60	1.379	same as c -n
N -H	403.20	1.013	same as hn-n
C -CT	313.00	1.524	same as c -c3
C -O	637.70	1.218	same as c -o
OH-HO	371.40	0.973	same as ho-oh

ANGLE

NT-CT-CT	66.020	111.040	same as c3-c3-n3
NT-CT-H1	49.550	109.800	same as hc-c3-n3
NT-CT-C	66.320	111.140	same as c -c3-n3
CT-NT-H	47.420	109.290	same as c3-n3-hn
CT-NT-CT	63.820	112.350	same as c3-n3-c3
CT-CT-CT	62.860	111.510	same as c3-c3-c3
CT-CT-OH	67.470	110.190	same as c3-c3-oh
CT-CT-H1	46.340	109.800	same as c3-c3-hc
CT-CT-OS	68.000	107.970	same as c3-c3-os
CT-CT-O	68.500	113.010	same as c3-c3-o
CT-OH-HO	47.380	107.260	same as c3-oh-ho
CT-OS-CT	62.700	112.480	same as c3-os-c3
CT-CT-N	65.910	111.610	same as c3-c3-n
OS-CT-O	73.189	115.295	Calculated with empirical approach
OS-CT-H1	51.050	108.700	same as hc-c3-os

ANGLE

CT-CT-HC	46.340	109.800	same as c3-c3-hc
CT-CT-C	63.270	111.040	same as c -c3-c3
HC-CT-HC	39.400	107.580	same as hc-c3-hc
CT-N -C	63.390	120.690	same as c -n -c3
CT-N -H	45.800	117.680	same as c3-n -hn
N -CT-H1	49.690	109.500	same as hc-c3-n
N -C -CT	66.790	115.180	same as c3-c -n
N -C -O	74.220	123.050	same as n -c -o
C -N -H	48.330	117.550	same as c -n -hn
C -CT-N	67.000	109.060	same as c -c3-n
C -CT-H1	46.930	108.770	same as c -c3-hc
CT-C -O	67.400	123.200	same as c3-c -o
C -CT-HC	46.930	108.770	same as c -c3-hc
OH-CT-H1	51.110	109.500	same as hc-c3-oh
H1-CT-H1	39.400	107.580	same as hc-c3-hc

DIHEDRALS

NT-CT-CT-CT	1	0.156	0.000	3.000	same as X -c3-c3-X
NT-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
NT-CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
NT-CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
NT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
CT-NT-CT-CT	1	0.300	0.000	-3.000	same as c3-c3-n3-c3
CT-NT-CT-CT	1	0.480	180.000	2.000	same as c3-c3-n3-c3
CT-NT-CT-H1	1	0.300	0.000	3.000	same as X -c3-n3-X
CT-NT-CT-C	1	0.300	0.000	3.000	same as X -c3-n3-X
CT-CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-CT	1	0.180	0.000	-3.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.250	180.000	-2.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.200	180.000	1.000	same as c3-c3-c3-c3
CT-CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-OH-HO	1	0.160	0.000	-3.000	same as ho-oh-c3-c3
CT-CT-OH-HO	1	0.250	0.000	1.000	same as ho-oh-c3-c3
CT-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-H1	1	0.160	0.000	3.000	same as hc-c3-c3-c3
CT-CT-NT-H	1	0.300	0.000	3.000	same as X -c3-n3-X
CT-CT-OS-CT	1	0.383	0.000	-3.000	same as c3-c3-os-c3

CT-CT-OS-CT	1	0.100	180.000	2.000	same as c3-c3-os-c3
CT-CT-CT-N	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-OS-CT-H1	1	0.383	0.000	3.000	same as X -c3-os-X
CT-CT-N -C	1	0.500	180.000	-4.000	same as c3-c3-n -c
CT-CT-N -C	1	0.150	180.000	-3.000	same as c3-c3-n -c
CT-CT-N -C	1	0.000	0.000	-2.000	same as c3-c3-n -c
CT-CT-N -C	1	0.530	0.000	1.000	same as c3-c3-n -c
CT-CT-N -H	1	0.000	0.000	2.000	same as X -c3-n -X
CT-CT-CT-HC	1	0.160	0.000	3.000	same as hc-c3-c3-c3
OS-CT-CT-OH	1	0.144	0.000	-3.000	same as os-c3-c3-oh
OS-CT-CT-OH	1	1.175	0.000	2.000	same as os-c3-c3-oh
OS-CT-CT-N	1	0.156	0.000	3.000	same as X -c3-c3-X
OS-CT-CT-H1	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-H1	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-HC	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-HC	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-OS-CT-O	1	0.383	0.000	3.000	same as X -c3-os-X
CT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
HC-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
H1-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
H1-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-N -C -CT	1	0.000	0.000	-2.000	same as c3-c -n -c3
CT-N -C -CT	1	1.500	180.000	1.000	same as c3-c -n -c3
CT-N -C -O	1	2.500	180.000	2.000	same as X -c -n -X
N -CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
N -C -CT-N	1	1.700	180.000	-1.000	same as n -c3-c -n
N -C -CT-N	1	2.000	180.000	2.000	same as n -c3-c -n
N -C -CT-CT	1	0.100	0.000	-4.000	same as c3-c3-c -n
N -C -CT-CT	1	0.070	0.000	2.000	same as c3-c3-c -n
N -C -CT-H1	1	0.000	180.000	2.000	same as X -c -c3-X
C -N -CT-H1	1	0.000	0.000	2.000	same as X -c3-n -X
C -CT-N -C	1	0.850	180.000	-2.000	same as c -n -c3-c
C -CT-N -C	1	0.800	0.000	1.000	same as c -n -c3-c
C -CT-N -H	1	0.000	0.000	2.000	same as X -c3-n -X
C -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-C -N -H	1	2.500	180.000	2.000	same as X -c -n -X
N -CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
N -C -CT-HC	1	0.000	180.000	2.000	same as X -c -c3-X
C -CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
HC-CT-CT-HC	1	0.150	0.000	3.000	same as hc-c3-c3-hc
HC-CT-C -O	1	0.800	0.000	-1.000	same as hc-c3-c -o
HC-CT-C -O	1	0.000	0.000	-2.000	same as hc-c3-c -o
HC-CT-C -O	1	0.080	180.000	3.000	same as hc-c3-c -o
O -C -N -H	1	2.500	180.000	-2.000	same as hn-n -c -o
O -C -N -H	1	2.000	0.000	1.000	same as hn-n -c -o
H -N -CT-H1	1	0.000	0.000	2.000	same as X -c3-n -X
OH-CT-CT-H1	1	0.000	0.000	-3.000	same as hc-c3-c3-oh
OH-CT-CT-H1	1	0.250	0.000	1.000	same as hc-c3-c3-oh
HO-OH-CT-H1	1	0.167	0.000	3.000	same as X -c3-oh-X
H1-CT-C -O	1	0.800	0.000	-1.000	same as hc-c3-c -o
H1-CT-C -O	1	0.000	0.000	-2.000	same as hc-c3-c -o
H1-CT-C -O	1	0.080	180.000	3.000	same as hc-c3-c -o
H1-CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
O -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
OH-CT-CT-OH	1	0.144	0.000	-3.000	same as oh-c3-c3-oh
OH-CT-CT-OH	1	1.175	0.000	2.000	same as oh-c3-c3-oh
H1-CT-NT-H	1	0.300	0.000	3.000	same as X -c3-n3-X
H -NT-CT-C	1	0.300	0.000	3.000	same as X -c3-n3-X

IMPROPER

C -CT-N -H	1.1	180.0	2.0	Using default value
CT-N -C -O	1.1	180.0	2.0	Using default value

NONBON

NT	1.8240	0.1700	same as n3
CT	1.9080	0.1094	same as c3
OS	1.6837	0.1700	same as os
HC	1.4870	0.0157	same as hc
H1	1.4870	0.0157	same as hc
N	1.8240	0.1700	same as n

C	1.9080	0.0860	same as c
O	1.6612	0.2100	same as o
H	0.6000	0.0157	same as hn
OH	1.7210	0.2104	same as oh
HO	0.0000	0.0000	same as ho

Table S2. The obtained non-standard pKa values together with protonation states of titratable residues at pH 7, as derived from the PropKa results.[6,7]

Chain	ResID	Residue	pKa	model-pKa	Protonated Atom
I	176	ASP	7.89	3.8	OD2
J	119	ASP	7.02	3.8	OD2
M	126	ASP	8.05	3.8	OD2
V	53	ASP	7.51	3.8	OD2
V	104	ASP	7.64	3.8	OD2
W	176	ASP	7.82	3.8	OD2
AA	126	ASP	8.12	3.8	OD2
D	118	GLP	7.84	4.5	OE2
I	106	GLP	8.8	4.5	OE2
J	166	GLP	11.14	4.5	OE2
K	56	GLP	8.79	4.5	OE2
R	118	GLP	9.95	4.5	OE2
W	106	GLP	8.48	4.5	OE2
X	166	GLP	11.33	4.5	OE2
Y	56	GLP	8.76	4.5	OE2
J	99	HIS	7.07	6.5	Nδ, Nε
J	189	HIS	7.08	6.5	Nδ, Nε
X	189	HIS	7.08	6.5	Nδ, Nε
A	71	HIS	4.26	6.5	Nε
A	88	HIS	4.45	6.5	Nε
B	30	HIS	6.3	6.5	Nε
B	53	HIS	5.81	6.5	Nε
B	142	HIS	4.85	6.5	Nε
D	65	HIS	6.2	6.5	Nε
D	91	HIS	6.06	6.5	Nε
D	178	HIS	6.35	6.5	Nε
D	219	HIS	6.2	6.5	Nε
E	43	HIS	5.75	6.5	Nε
E	69	HIS	4.32	6.5	Nε
E	143	HIS	6.93	6.5	Nε
E	190	HIS	4.76	6.5	Nε
F	67	HIS	5.02	6.5	Nε
F	105	HIS	2.7	6.5	Nε
F	115	HIS	5.72	6.5	Nε
F	196	HIS	6.69	6.5	Nε
G	238	HIS	5.6	6.5	Nε
H	35	HIS	3.98	6.5	Nε
I	162	HIS	6.67	6.5	Nε

I	188	HIS	3.72	6.5	Nε
J	32	HIS	1.97	6.5	Nε
J	132	HIS	3.66	6.5	Nε
K	10	HIS	5.87	6.5	Nε
K	178	HIS	4.73	6.5	Nε
K	196	HIS	5.14	6.5	Nε
L	36	HIS	6.36	6.5	Nε
L	58	HIS	3.26	6.5	Nε
M	81	HIS	5.41	6.5	Nε
M	89	HIS	4.5	6.5	Nε
M	213	HIS	5.81	6.5	Nε
N	38	HIS	5.14	6.5	Nε
O	71	HIS	4.53	6.5	Nε
P	30	HIS	6.33	6.5	Nε
P	142	HIS	3.72	6.5	Nε
P	240	HIS	6.02	6.5	Nε
Q	154	HIS	3.64	6.5	Nε
R	65	HIS	6.24	6.5	Nε
R	91	HIS	5.84	6.5	Nε
R	147	HIS	1.52	6.5	Nε
S	43	HIS	5.27	6.5	Nε
S	65	HIS	4.2	6.5	Nε
S	69	HIS	4.83	6.5	Nε
S	143	HIS	6.85	6.5	Nε
S	190	HIS	3.16	6.5	Nε
T	68	HIS	5.36	6.5	Nε
T	106	HIS	6.82	6.5	Nε
T	116	HIS	5.69	6.5	Nε
T	220	HIS	3.32	6.5	Nε
U	68	HIS	4.33	6.5	Nε
V	35	HIS	3.91	6.5	Nε
V	66	HIS	3.34	6.5	Nε
X	32	HIS	1.83	6.5	Nε
X	110	HIS	6.86	6.5	Nε
Y	10	HIS	5.91	6.5	Nε
Y	178	HIS	4.5	6.5	Nε
Y	196	HIS	5.11	6.5	Nε
Z	36	HIS	6.36	6.5	Nε
Z	58	HIS	3.27	6.5	Nε
Z	77	HIS	0.93	6.5	Nε
Z	163	HIS	6.14	6.5	Nε
A	63	HIS	2.84	6.5	Nδ
A	189	HIS	5.83	6.5	Nδ
B	240	HIS	5.94	6.5	Nδ
C	15	HIS	6.68	6.5	Nδ
C	94	HIS	3.99	6.5	Nδ

C	154	HIS	4.2	6.5	Nδ
D	147	HIS	1.56	6.5	Nδ
E	20	HIS	3.38	6.5	Nδ
E	59	HIS	1.8	6.5	Nδ
E	65	HIS	3.71	6.5	Nδ
E	175	HIS	6.15	6.5	Nδ
F	219	HIS	3.3	6.5	Nδ
G	12	HIS	4.3	6.5	Nδ
G	68	HIS	3.68	6.5	Nδ
H	66	HIS	3.22	6.5	Nδ
H	109	HIS	3.8	6.5	Nδ
H	116	HIS	5.82	6.5	Nδ
J	110	HIS	6.94	6.5	Nδ
L	77	HIS	1.11	6.5	Nδ
L	163	HIS	6.58	6.5	Nδ
N	66	HIS	2.58	6.5	Nδ
N	77	HIS	4.11	6.5	Nδ
O	63	HIS	4.67	6.5	Nδ
O	88	HIS	4.88	6.5	Nδ
O	189	HIS	5.85	6.5	Nδ
P	53	HIS	5.64	6.5	Nδ
Q	15	HIS	6.51	6.5	Nδ
Q	94	HIS	3.83	6.5	Nδ
R	178	HIS	6.77	6.5	Nδ
R	219	HIS	6.21	6.5	Nδ
S	20	HIS	4.54	6.5	Nδ
S	59	HIS	5.19	6.5	Nδ
S	175	HIS	6.23	6.5	Nδ
T	197	HIS	6.79	6.5	Nδ
U	12	HIS	3.61	6.5	Nδ
U	238	HIS	6.18	6.5	Nδ
V	109	HIS	3.34	6.5	Nδ
V	116	HIS	5.86	6.5	Nδ
W	162	HIS	5.62	6.5	Nδ
W	188	HIS	3.82	6.5	Nδ
X	99	HIS	5.63	6.5	Nδ
X	132	HIS	2.99	6.5	Nδ
AA	81	HIS	5.49	6.5	Nδ
AA	89	HIS	3.21	6.5	Nδ
AA	213	HIS	5.77	6.5	Nδ
BB	38	HIS	5.29	6.5	Nδ
BB	66	HIS	3.11	6.5	Nδ
BB	77	HIS	4.43	6.5	Nδ

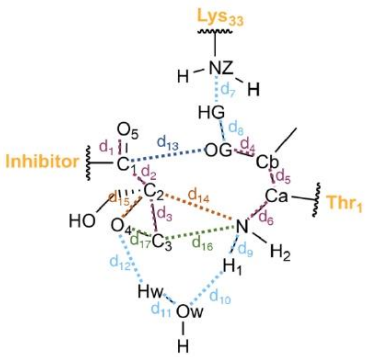
Table S3. Number of iterations during WHAM calculation performed to reach the density threshold value for each 2D-PMF computed in the study.

	Mechanism A		Mechanism B	
	Cycle 7	Cycle 6	Cycle 7	Cycle 6
Step 1	1377		1170	914
Step 2	996	1143	3120	3128
Step 3	3404	3134	1224	1286

Table S4. Quasiclassical KIEs, computed for isotopic substitutions of key atoms as average from 10 couples of stationary structures (rate limiting TS and RS), computed at AM1/AMBER level.

		Mechanism A		Mechanism B	
		Cycle 7	Cycle 6	Cycle 7	Cycle 6
TS1	[1-^{14}C]	1.018 ± 0.003		1.002 ± 0.001	1.011 ± 0.003
	[2-^{14}C]	1.002 ±0.004		0.999 ± 0.003	1.099 ± 0.009
	[3-^{14}C]	1.001 ± 0.008		1.137 ± 0.007	0.994 ± 0.007
	[4-^{18}O]	0.992 ± 0.008		1.018 ± 0.006	1.021 ± 0.007
	[5-^{18}O]	1.000 ± 0.002		0.998 ± 0.002	0.997 ± 0.003
TS2	[1-^{14}C]	1.018 ± 0.003	1.002 ± 0.001	1.012 ± 0.001	1.008 ± 0.003
	[2-^{14}C]	0.992 ± 0.003	0.999 ± 0.003	0.997 ± 0.004	1.025 ± 0.004
	[3-^{14}C]	1.131 ± 0.007	1.137 ± 0.007	0.963 ± 0.006	0.962 ± 0.011
	[4-^{18}O]	1.015 ± 0.006	1.018 ± 0.006	1.004 ± 0.007	0.993 ± 0.007
	[5-^{18}O]	1.001 ± 0.003	0.998 ± 0.002	1.004 ± 0.003	1.001 ± 0.002

Table S5. Averaged values of key distances (Å) along reaction path for Mechanism A and B. Obtained during 20 ps QM/MM MD simulations at 310K with 5ps of equilibration at AM1/MM level of theory. For TS structures, US approach was used in order to sample the saddle point area.



Mechanism A							
				Cycle 7			
Dist	RC	TS1	INT1	TS2	INT2	TS3	PC
d1	1.24 ± 0.02	1.25 ± 0.02	1.31 ± 0.02	1.31 ± 0.02	1.31 ± 0.02	1.32 ± 0.02	1.31 ± 0.02
d2	1.50 ± 0.03	1.53 ± 0.03	1.56 ± 0.03	1.57 ± 0.03	1.58 ± 0.04	1.58 ± 0.04	1.59 ± 0.03
d3	1.50 ± 0.02	1.50 ± 0.03	1.49 ± 0.02	1.49 ± 0.03	1.59 ± 0.03	1.59 ± 0.03	1.56 ± 0.03
d4	1.42 ± 0.03	1.41 ± 0.02	1.42 ± 0.03	1.42 ± 0.03	1.40 ± 0.02	1.41 ± 0.02	1.42 ± 0.02
d5	1.56 ± 0.03	1.57 ± 0.03	1.57 ± 0.03	1.57 ± 0.03	1.57 ± 0.03	1.59 ± 0.03	1.56 ± 0.03
d6	1.44 ± 0.03	1.44 ± 0.02	1.44 ± 0.02	1.47 ± 0.03	1.50 ± 0.03	1.47 ± 0.03	1.45 ± 0.03
d7	2.73 ± 0.37	1.34 ± 0.02	1.04 ± 0.03	1.03 ± 0.03	1.03 ± 0.03	1.04 ± 0.03	1.03 ± 0.03
d8	0.97 ± 0.02	1.21 ± 0.02	3.30 ± 0.30	2.15 ± 0.24	2.06 ± 0.12	3.34 ± 0.25	2.05 ± 0.12
d9	1.01 ± 0.03	1.01 ± 0.03	1.00 ± 0.03	1.01 ± 0.03	1.03 ± 0.03	1.54 ± 0.02	4.67 ± 0.64
d10	5.15 ± 0.33	6.85 ± 0.55	5.51 ± 0.38	5.22 ± 0.41	4.96 ± 0.28	1.19 ± 0.02	0.97 ± 0.03
d11	0.97 ± 0.03	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.03	0.97 ± 0.02	1.08 ± 0.02	3.09 ± 0.36
d12	3.58 ± 0.46	4.12 ± 0.70	4.24 ± 0.55	3.02 ± 0.63	2.93 ± 0.58	1.96 ± 0.02	0.97 ± 0.03
d13	3.10 ± 0.26	1.89 ± 0.02	1.48 ± 0.04	1.48 ± 0.03	1.52 ± 0.04	1.51 ± 0.04	1.47 ± 0.03
d14	3.29 ± 0.17	3.18 ± 0.12	3.17 ± 0.16	2.91 ± 0.08	2.68 ± 0.05	2.55 ± 0.04	2.55 ± 0.05
d15	1.45 ± 0.03	1.46 ± 0.03	1.45 ± 0.03	1.38 ± 0.03	1.34 ± 0.02	1.35 ± 0.02	1.43 ± 0.03
d16	2.94 ± 0.14	3.07 ± 0.18	3.04 ± 0.18	1.99 ± 0.06	1.49 ± 0.03	1.45 ± 0.03	1.44 ± 0.03
d17	1.44 ± 0.03	1.44 ± 0.02	1.44 ± 0.02	1.93 ± 0.06	2.34 ± 0.05	2.33 ± 0.04	2.38 ± 0.05

	Cycle 6			
Dist	TS2	INT2	TS3	PC
d1	1.30 ± 0.02	1.30 ± 0.02	1.31 ± 0.02	1.31 ± 0.02
d2	1.56 ± 0.03	1.61 ± 0.04	1.62 ± 0.04	1.62 ± 0.04
d3	1.51 ± 0.03	1.60 ± 0.04	1.58 ± 0.03	1.55 ± 0.03
d4	1.42 ± 0.03	1.42 ± 0.03	1.42 ± 0.03	1.42 ± 0.02
d5	1.56 ± 0.03	1.56 ± 0.03	1.57 ± 0.03	1.55 ± 0.03

d6	1.46 ± 0.03	1.49 ± 0.03	1.48 ± 0.03	1.46 ± 0.03
d7	1.03 ± 0.03	1.03 ± 0.03	1.03 ± 0.03	1.03 ± 0.03
d8	2.07 ± 0.12	2.08 ± 0.14	2.05 ± 0.12	2.10 ± 0.18
d9	1.01 ± 0.03	1.04 ± 0.03	1.37 ± 0.02	2.75 ± 0.32
d10	5.26 ± 0.23	2.09 ± 0.14	1.19 ± 0.02	0.97 ± 0.03
d11	0.97 ± 0.03	0.96 ± 0.03	1.03 ± 0.02	2.25 ± 0.24
d12	2.90 ± 0.51	2.20 ± 0.19	1.62 ± 0.02	0.97 ± 0.03
d13	1.49 ± 0.04	1.49 ± 0.04	1.49 ± 0.04	1.48 ± 0.03
d14	2.12 ± 0.06	1.53 ± 0.03	1.51 ± 0.03	1.49 ± 0.03
d15	2.24 ± 0.05	2.48 ± 0.05	2.49 ± 0.05	2.51 ± 0.05
d16	2.61 ± 0.02	2.53 ± 0.05	2.51 ± 0.05	2.48 ± 0.05
d17	1.35 ± 0.02	1.33 ± 0.02	1.36 ± 0.02	1.42 ± 0.02

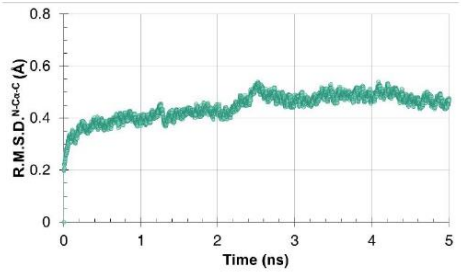


Figure S1. Time-dependent evolution of the RMSD of the atoms belonging to the protein backbone during the 5 ns MD trajectory.

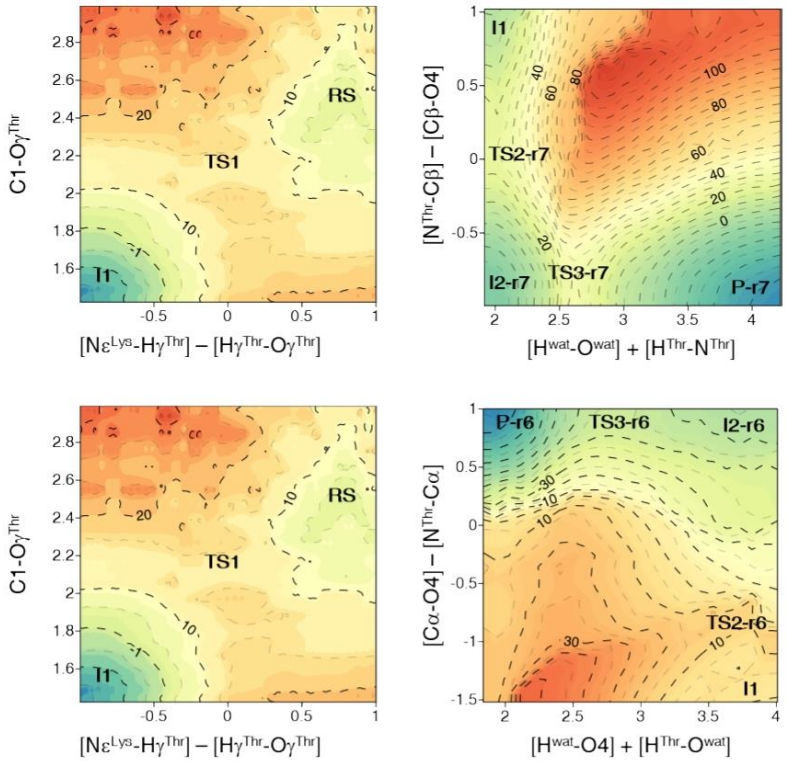


Figure S2. AM1/MM PESs corresponding to the inhibition mechanism rendering the seven-membered ring product formation above and the six-membered ring product formation below through Mechanism A. Distances in Å and values on iso-energetic lines in kcal·mol^{-1}.

$^{-1}$.

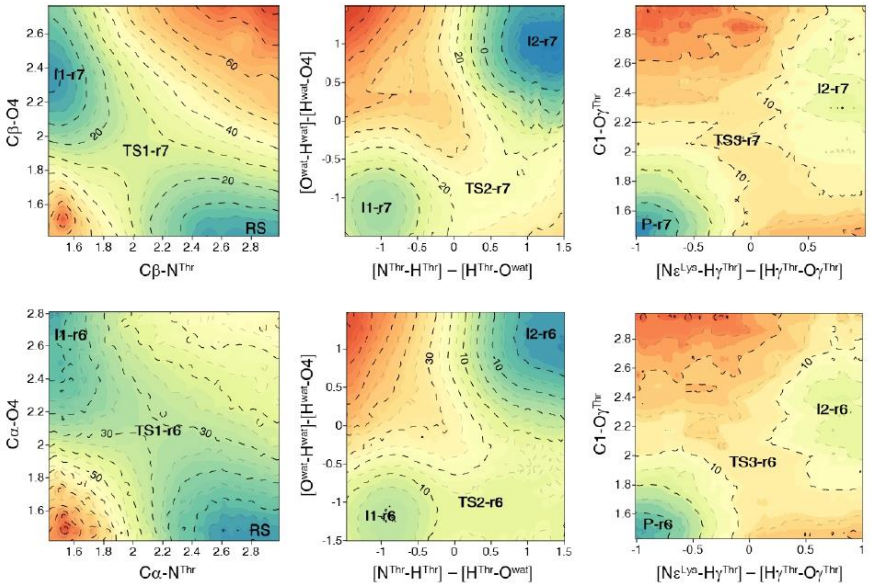


Figure S3. AM1/MM PESs corresponding to the inhibition mechanism rendering the seven-membered ring product formation above and six-membered ring product formation below through Mechanism B. Distances in Å and values on iso-energetic lines in kcal·mol^{-1}.

References

[1] J.J. Ruiz-Pernía, E. Silla, I. Tuñón, S. Martí, Hybrid Quantum Mechanics/Molecuar Mechanics simulations with two-dimensional interpolated corrections: appliation to enzymatic processes, J. Phys. Chem. B, 110 (2006) 17663 – 17670. DOI: 10.1021/jp063520a.

[2] Y.Y. Chuang, J.C. Corchado, D.G. Truhlar, Mapped interpolation scheme for single-point energy corrections in reaction rate calculations and a critical evaluation of dual-level reaction path dynamics methods, J. Phys. Chem. A, 103 (1999) 1140 – 1149. DOI: 10.1021/jp9842493.

[3] S. Ferrer, S. Martí, V. Moliner, I. Tuñón, J. Bertran, Understanding the different activities of highly promiscuous MbtI by computational methods, Phys. Chem. Chem. Phys., 14 (2012) 3482 – 3489. DOI: 10.1039/C2CP23149B.

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

[5] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graph. Model. , 25 (2006) 247 – 260. DOI: 10.1016/j.jmgm.2005.12.005.

[6] C.R. Sondergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values, J. Chem. Theory Comput., 7 (2011) 2284 – 2295. DOI: 10.1021/ct200133y.

[7] M.H.M Olsson, C.R. Sondergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions, J. Chem. Theory Comput., 7 (2011) 525 – 537. DOI: 10.1021/ct100578z.

4.2 γ-lactam-β-lactone inhibitors family in 20S proteasome

4.2.1 Salinosporamide A (SalA) inhibition mechanism

Supporting Information

Nature of Irreversible Inhibition of Human 20S Proteasome by Salinosporamide A. The Critical Role of Lys-Asp Dyad Revealed from Electrostatic Effects Analysis

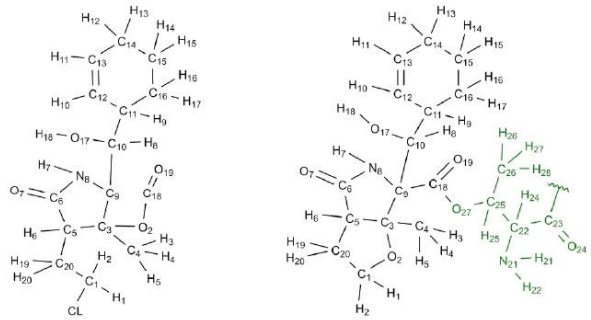
Natalia Serrano-Aparicio, Vicent Moliner*, Katarzyna Świderek*

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castellón, Spain.

*V.M.: moliner@uji.es *K.Ś.: swiderek@uji.es

Table of contents:

1. MM Parameters for SalA



Scheme S1. Structure and numbering of **A.** SalA inhibitor and **B.** SalA in covalent complex with Thr1 from active site of the 20S proteasome.

Table S1. Atom types, charges and parameters for bonding and non-bonding interactions obtained for SalA inhibitor in its free form computed using Antechamber software.

Atom name	Atom type	Charge	Atom name	Atom type	Charge
C1	CT	0.024000	C12	CM	-0.184200
C20	CT	-0.067400	C13	CM	-0.149200
C5	CT	-0.141700	C14	CT	-0.053200
C6	C	0.699500	C15	CT	-0.073400
O7	O	-0.594500	C16	CT	-0.079400
N8	N	-0.572900	H16	HC	0.053200
H7	H	0.342500	H17	HC	0.053200
H6	HC	0.100700	H14	HC	0.045700
H19	HC	0.072200	H15	HC	0.045700
H20	HC	0.072200	H12	HC	0.053200
H1	H1	0.079700	H13	HC	0.053200
H2	H1	0.079700	H11	HA	0.126000
O2	OS	-0.408900	H10	HA	0.137000
C3	CT	0.145800	H9	HC	0.054700
C4	CT	-0.095100	O17	OH	-0.588800
H3	HC	0.066367	H18	HO	0.410000
H4	HC	0.066367	H8	H1	0.097700
H5	HC	0.066367	C18	C	0.692600
C9	CT	-0.017600	O19	O	-0.525500
C10	CT	0.160100	CL	Cl	-0.187400
C11	CT	-0.059500			

Parameters:							
MASS				**BOND**			
CT	12.010	0.878	same as c3	CT-CT	300.90	1.538	same as c3-c3
C	12.010	0.616	same as c	CT-H1	330.60	1.097	same as c3-hc
O	16.000	0.434	same as o	CT-OS	308.60	1.432	same as c3-os
N	14.010	0.530	same as n	CT-HC	330.60	1.097	same as c3-hc
H	1.008	0.161	same as hn	CT-C	313.00	1.524	same as c -c3
HC	1.008	0.135	same as hc	CT-Cl	266.30	1.804	same as c3-cl

H1	1.008	0.135	same as hc	C -O	637.70	1.218	same as c -o
OS	16.000	0.465	same as os	C -N	427.60	1.379	same as c -n
CM	12.010	0.360	same as c2	N -H	403.20	1.013	same as hn-n
HA	1.008	0.135	same as hc	N -CT	328.70	1.462	same as c3-n
OH	16.000	0.465	same as oh	CT-OH	316.70	1.423	same as c3-oh
HO	1.008	0.135	same as ho	CT-CM	326.80	1.510	same as c2-c3
				CM-CM	419.80	1.428	same as cc-cc
				CM-HA	344.30	1.087	same as c2-hc
				OH-HO	371.40	0.973	same as ho-oh
				C -OS	390.80	1.358	same as c -os
ANGLE				**ANGLE**			
CT-CT-CT	62.860	111.510	same as c3-c3-c3	H1-CT-H1	39.400	107.580	Same as hc-c3-hc
CT-CT-HC	46.340	109.800	same as c3-c3-hc	CT-CT-OH	67.470	110.190	same as c3-c3-oh
CT-CT-H1	46.340	109.800	same as c3-c3-hc	CT-C -OS	68.890	110.720	same as c3-c -os
CT-CT-OS	68.000	107.970	same as c3-c3-os	CT-CT-CM	63.410	111.560	same as c2-c3-c3
CT-CT-C	63.270	111.040	same as c -c3-c3	CT-OH-HO	47.380	107.260	same as c3-oh-ho
CT-C -O	67.400	123.200	same as c3-c -o	CT-CM-CM	64.060	123.630	same as c2-c2-c3
CT-C -N	66.790	115.180	same as c3-c -n	CT-CM-HA	45.110	120.000	same as c3-c2-hc
C -CT-HC	46.930	108.770	same as c -c3-hc	CM-CT-HC	46.990	110.360	same as c2-c3-hc
C -N -H	48.330	117.550	same as c -n -hn	CM-CM-HA	50.010	119.700	same as c2-c2-hc
C -N -CT	63.390	120.690	same as c -n -c3	OH-CT-H1	51.110	109.500	same as hc-c3-oh
O -C -N	74.220	123.050	same as n -c -o	C -OS-CT	63.280	115.980	same as c -os-c3
N -CT-CT	65.910	111.610	same as c3-c3-n	O -C -OS	75.320	123.250	same as o -c -os
N -CT-C	67.000	109.060	same as c -c3-n	H1-CT-Cl	39.900	107.650	same as cl-c3-hc
H -N -CT	45.800	117.680	same as c3-n -hn	Cl-CT-CT	57.390	110.410	same as c3-c3-cl
HC-CT-HC	39.400	107.580	same as hc-c3-hc				

DIHEDRALS					
CT-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-CT	1	0.180	0.000	-3.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.250	180.000	-2.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.200	180.000	1.000	same as c3-c3-c3-c3
CT-CT-CT-HC	1	0.160	0.000	3.000	same as hc-c3-c3-c3
H1-CT-CT-CT	1	0.160	0.000	3.000	same as hc-c3-c3-c3
H1-CT-CT-HC	1	0.150	0.000	3.000	same as hc-c3-c3-hc
Cl-CT-CT-CT	1	0.156	0.000	3.000	same as X -c3-c3-X
Cl-CT-CT-HC	1	0.000	0.000	-3.000	same as hc-c3-c3-cl
Cl-CT-CT-HC	1	0.250	0.000	1.000	same as hc-c3-c3-cl
CT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-C -N	1	0.100	0.000	-4.000	same as c3-c3-c -n
CT-CT-C -N	1	0.070	0.000	2.000	same as c3-c3-c -n
CT-CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-C -N -H	1	2.500	180.000	2.000	same as X -c -n -X
CT-C -N -CT	1	0.000	0.000	-2.000	same as c3-c -n -c3
CT-C -N -CT	1	1.500	180.000	1.000	same as c3-c -n -c3
CT-CT-CT-N	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-OS-C	1	0.383	0.000	-3.000	same as c3-c3-os-c
CT-CT-OS-C	1	0.800	180.000	1.000	same as c3-c3-os-c
C -CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
C -CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
C -N -CT-CT	1	0.500	180.000	-4.000	same as c3-c3-n -c
C -N -CT-CT	1	0.150	180.000	-3.000	same as c3-c3-n -c
C -N -CT-CT	1	0.000	0.000	-2.000	same as c3-c3-n -c
C -N -CT-CT	1	0.530	0.000	1.000	same as c3-c3-n -c
C -N -CT-C	1	0.850	180.000	-2.000	same as c -n -c3-c
C -N -CT-C	1	0.800	0.000	1.000	same as c -n -c3-c
O -C -CT-HC	1	0.800	0.000	-1.000	same as hc-c3-c -o
O -C -CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c -o
O -C -CT-HC	1	0.080	180.000	3.000	same as hc-c3-c -o
O -C -N -H	1	2.500	180.000	-2.000	same as hn-n -c -o
O -C -N -H	1	2.000	0.000	1.000	same as hn-n -c -o
O -C -N -CT	1	2.500	180.000	2.000	same as X -c -n -X
N -C -CT-HC	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X

N -CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
H -N -CT-CT	1	0.000	0.000	2.000	same as X -c3-n -X
H -N -CT-C	1	0.000	0.000	2.000	same as X -c3-n -X
CT-C -OS-CT	1	2.700	180.000	-2.000	same as c3-c -os-c3
CT-C -OS-CT	1	0.000	0.000	-1.000	same as c3-c -os-c3
CT-C -OS-CT	1	1.150	0.000	3.000	same as c3-c -os-c3
CT-CT-CT-CM	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-OH-HO	1	0.160	0.000	-3.000	same as ho-oh-c3-c3
CT-CT-OH-HO	1	0.250	0.000	1.000	same as ho-oh-c3-c3
CT-CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-OS-C -O	1	2.700	180.000	-2.000	same as o -c -os-c3
CT-OS-C -O	1	1.400	180.000	1.000	same as o -c -os-c3
OS-CT-CT-HC	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-HC	1	0.250	0.000	1.000	same as hc-c3-c3-os
C -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
C -CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CM-CM	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CM-CM-CT	1	6.650	180.000	-2.000	same as c3-c2-c2-c3
CT-CM-CM-CT	1	1.900	180.000	1.000	same as c3-c2-c2-c3
CT-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
CM-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CM-CT-HC	1	0.380	180.000	-3.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	1.150	0.000	1.000	same as hc-c3-c2-c2
HC-CT-CT-HC	1	0.150	0.000	3.000	same as hc-c3-c3-hc
HC-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
HA-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
HC-CT-CT-OH	1	0.000	0.000	-3.000	same as hc-c3-c3-oh
HC-CT-CT-OH	1	0.250	0.000	1.000	same as hc-c3-c3-oh
HO-OH-CT-H1	1	0.167	0.000	3.000	same as X -c3-oh-X

IMPROPER				
CT-N -C -O	1.1	180.0	2.0	Using default value
C -CT-N -H	1.1	180.0	2.0	Using default value
CM-CT-CM-HA	1.1	180.0	2.0	Using default value
CT-O -C -OS	1.1	180.0	2.0	Using default value

NONBON			
CT	1.9080	0.1094	same as c3
C	1.9080	0.0860	same as c
O	1.6612	0.2100	same as o
N	1.8240	0.1700	same as n
H	0.6000	0.0157	same as hn
HC	1.4870	0.0157	same as hc
H1	1.4870	0.0157	same as hc
OS	1.6837	0.1700	same as os
CM	1.9080	0.0860	same as cc
HA	1.4870	0.0157	same as hc
OH	1.7210	0.2104	same as oh
HO	0.0000	0.0000	same as ho
Cl	1.9480	0.2650	same as cl

Table S2. Atom types, charges and parameters for bonding and non-bonding interactions obtained for SalA inhibitor covalently bound to Thr1 residue from the active site of 20S proteasome computed using Antechamber software.

Atom name	Atom type	Charge	Atom name	Atom type	Charge
C1	CT	0.123435	H17	HC	0.051214
C20	CT	-0.098372	H14	HC	0.044713
C5	CT	-0.166653	H15	HC	0.044713
C6	C	0.694697	H12	HC	0.050214
O7	O	-0.596331	H13	HC	0.050214
N8	N	-0.553743	H11	HA	0.125035
H7	H	0.335595	H10	HA	0.137039
H6	HC	0.109731	H9	HC	0.064718
H19	HC	0.074721	O17	OH	-0.595631
H20	HC	0.074721	H18	HO	0.411116
H1	H1	0.059217	H8	H1	0.103729
H2	H1	0.059217	C18	C	0.646283
O2	OS	-0.441475	O19	O	-0.536848
C3	CT	0.165847	O27	OS	-0.432777
C4	CT	-0.101071	C25	CT	0.110131
H3	HC	0.062718	C26	CT	-0.118067
H4	HC	0.062718	H26	HC	0.051715
H5	HC	0.062718	H27	HC	0.051715
C9	CT	0.026407	H28	HC	0.051715
C10	CT	0.160145	H25	H1	0.091726
C11	CT	-0.058483	C22	CT	0.053515
C12	CM	-0.177150	N21	NT	-0.891547
C13	CM	-0.158155	H21	H	0.367904
C14	CT	-0.051186	H22	H	0.367904
C15	CT	-0.073379	H24	H1	0.078722
C16	CT	-0.087375	C23	C	0.573062
H16	HC	0.051214	O24	O	-0.511955

Parameters:

MASS				**BOND**			
CT	12.010	0.878	same as c3	CT-CT	300.90	1.538	same as c3-c3
C	12.010	0.616	same as c	CT-H1	330.60	1.097	same as c3-hc
O	16.000	0.434	same as o	CT-OS	308.60	1.432	same as c3-os
N	14.010	0.530	same as n	CT-HC	330.60	1.097	same as c3-hc
H	1.008	0.161	same as hn	CT-C	313.00	1.524	same as c -c3
HC	1.008	0.135	same as hc	C -O	637.70	1.218	same as c -o
H1	1.008	0.135	same as hc	C -N	427.60	1.379	same as c -n
OS	16.000	0.465	same as os	N -H	403.20	1.013	same as hn-n
CM	12.010	0.360	same as c2	N -CT	328.70	1.462	same as c3-n
HA	1.008	0.135	same as hc	CT-OH	316.70	1.423	same as c3-oh
OH	16.000	0.465	same as oh	CT-CM	326.80	1.510	same as c2-c3
HO	1.008	0.135	same as ho	CM-CM	419.80	1.428	same as cc-cc
NT	14.010	0.530	same as n3	CM-HA	344.30	1.087	same as c2-hc
				OH-HO	371.40	0.973	same as ho-oh
				C -OS	390.80	1.358	same as c -os
				CT-NT	325.90	1.465	same as c3-n3
				NT-H	392.40	1.019	same as hn-n3
ANGLE				**ANGLE**			
CT-CT-CT	62.860	111.510	same as c3-c3-c3	CT-CT-OH	67.470	110.190	same as c3-c3-oh
CT-CT-HC	46.340	109.800	same as c3-c3-hc	CT-C -OS	68.890	110.720	same as c3-c -os
CT-OS-CT	62.700	112.480	same as c3-os-c3	CT-CT-CM	63.410	111.560	same as c2-c3-c3

CT-CT-H1	46.340	109.800	same as c3-c3-hc	CT-OH-HO	47.380	107.260	same as c3-oh-ho
CT-CT-OS	68.000	107.970	same as c3-c3-os	CT-CM-CM	64.060	123.630	same as c2-c2-c3
CT-CT-C	63.270	111.040	same as c -c3-c3	CT-CM-HA	45.110	120.000	same as c3-c2-hc
CT-C -O	67.400	123.200	same as c3-c -o	CM-CT-HC	46.990	110.360	same as c2-c3-hc
CT-C -N	66.790	115.180	same as c3-c -n	CM-CM-HA	50.010	119.700	same as c2-c2-hc
C -CT-HC	46.930	108.770	same as c -c3-hc	OH-CT-H1	51.110	109.500	same as hc-c3-oh
C -N -H	48.330	117.550	same as c -n -hn	C -OS-CT	63.280	115.980	same as c -os-c3
C -N -CT	63.390	120.690	same as c -n -c3	O -C -OS	75.320	123.250	same as o -c -os
O -C -N	74.220	123.050	same as n -c -o	CT-CT-NT	66.020	111.040	same as c3-c3-n3
N -CT-CT	65.910	111.610	same as c3-c3-n	CT-NT-H	47.420	109.290	same as c3-n3-hn
N -CT-C	67.000	109.060	same as c -c3-n	NT-CT-H1	49.550	109.800	same as hc-c3-n3
H -N -CT	45.800	117.680	same as c3-n -hn	NT-CT-C	66.320	111.140	same as c -c3-n3
HC-CT-HC	39.400	107.580	same as hc-c3-hc	H -NT-H	41.400	106.400	same as hn-n3-hn
H1-CT-H1	39.400	107.580	same as hc-c3-hc	H1-CT-C	46.930	108.770	same as c -c3-hc
H1-CT-OS	51.050	108.700	same as hc-c3-os				

DIHEDRALS

CT-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-HC	1	0.160	0.000	3.000	same as hc-c3-c3-c3
CT-CT-CT-CT	1	0.180	0.000	-3.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.250	180.000	-2.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.200	180.000	1.000	same as c3-c3-c3-c3
CT-OS-CT-CT	1	0.383	0.000	-3.000	same as c3-c3-os-c3
CT-OS-CT-CT	1	0.100	180.000	2.000	same as c3-c3-os-c3
CT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-C -N	1	0.100	0.000	-4.000	same as c3-c3-c -n
CT-CT-C -N	1	0.070	0.000	2.000	same as c3-c3-c -n
CT-CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-H1	1	0.160	0.000	3.000	same as hc-c3-c3-c3
CT-C -N -H	1	2.500	180.000	2.000	same as X -c -n -X
CT-C -N -CT	1	0.000	0.000	-2.000	same as c3-c -n -c3
CT-C -N -CT	1	1.500	180.000	1.000	same as c3-c -n -c3
CT-CT-CT-N	1	0.156	0.000	3.000	same as X -c3-c3-X
C -CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
C -CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
C -N -CT-CT	1	0.500	180.000	-4.000	same as c3-c3-n -c
C -N -CT-CT	1	0.150	180.000	-3.000	same as c3-c3-n -c
C -N -CT-CT	1	0.000	0.000	-2.000	same as c3-c3-n -c
C -N -CT-CT	1	0.530	0.000	1.000	same as c3-c3-n -c
C -N -CT-C	1	0.850	180.000	-2.000	same as c -n -c3-c
C -N -CT-C	1	0.800	0.000	1.000	same as c -n -c3-c
O -C -CT-HC	1	0.800	0.000	-1.000	same as hc-c3-c -o
O -C -CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c -o
O -C -CT-HC	1	0.080	180.000	3.000	same as hc-c3-c -o
O -C -N -H	1	2.500	180.000	-2.000	same as hn-n -c -o
O -C -N -H	1	2.000	0.000	1.000	same as hn-n -c -o
O -C -N -CT	1	2.500	180.000	2.000	same as X -c -n -X
N -C -CT-HC	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-CT-OS	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
H -N -CT-CT	1	0.000	0.000	2.000	same as X -c3-n -X
H -N -CT-C	1	0.000	0.000	2.000	same as X -c3-n -X
HC-CT-CT-HC	1	0.150	0.000	3.000	same as hc-c3-c3-hc
HC-CT-CT-OS	1	0.000	0.000	-3.000	same as hc-c3-c3-os
HC-CT-CT-OS	1	0.250	0.000	1.000	same as hc-c3-c3-os
HC-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
H1-CT-OS-CT	1	0.383	0.000	3.000	same as X -c3-os-X
CT-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-CT-CM	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-OH-HO	1	0.160	0.000	-3.000	same as ho-oh-c3-c3

CT-CT-OH-HO	1	0.250	0.000	1.000	same as ho-oh-c3-c3
CT-C -OS-CT	1	2.700	180.000	-2.000	same as c3-c -os-c3
CT-C -OS-CT	1	0.000	0.000	-1.000	same as c3-c -os-c3
CT-C -OS-CT	1	1.150	0.000	3.000	same as c3-c -os-c3
CT-CT-CM-CM	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CM-CM-CT	1	6.650	180.000	-2.000	same as c3-c2-c2-c3
CT-CM-CM-CT	1	1.900	180.000	1.000	same as c3-c2-c2-c3
CT-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
CM-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CM-CT-HC	1	0.380	180.000	-3.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	1.150	0.000	1.000	same as hc-c3-c2-c2
HC-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
HA-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
HC-CT-CT-OH	1	0.000	0.000	-3.000	same as hc-c3-c3-oh
HC-CT-CT-OH	1	0.250	0.000	1.000	same as hc-c3-c3-oh
OH-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
HO-OH-CT-H1	1	0.167	0.000	3.000	same as X -c3-oh-X
H1-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
C -OS-CT-CT	1	0.383	0.000	-3.000	same as c3-c3-os-c
C -OS-CT-CT	1	0.800	180.000	1.000	same as c3-c3-os-c
C -OS-CT-H1	1	0.383	0.000	3.000	same as X -c3-os-X
O -C -OS-CT	1	2.700	180.000	-2.000	same as o -c -os-c3
O -C -OS-CT	1	1.400	180.000	1.000	same as o -c -os-c3
OS-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
OS-CT-CT-H1	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-H1	1	0.250	0.000	1.000	same as hc-c3-c3-os
CT-CT-NT-H	1	0.300	0.000	3.000	same as X -c3-n3-X
CT-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
H1-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
H1-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
NT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
H -NT-CT-H1	1	0.300	0.000	3.000	same as X -c3-n3-X
H -NT-CT-C	1	0.300	0.000	3.000	same as X -c3-n3-X
H1-CT-C -O	1	0.800	0.000	-1.000	same as hc-c3-c -o
H1-CT-C -O	1	0.000	0.000	-2.000	same as hc-c3-c -o
H1-CT-C -O	1	0.080	180.000	3.000	same as hc-c3-c -o
IMPROPER					
CT-N -C -O	1.1	180.0	2.0	Using default value	
C -CT-N -H	1.1	180.0	2.0	Using default value	
CM-CT-CM-HA	1.1	180.0	2.0	Using default value	
CT-O -C -OS	1.1	180.0	2.0	Using default value	
NONBON					
CT	1.9080	0.1094	same as c3		
C	1.9080	0.0860	same as c		
O	1.6612	0.2100	same as o		
N	1.8240	0.1700	same as n		
H	0.6000	0.0157	same as hn		
HC	1.4870	0.0157	same as hc		
H1	1.4870	0.0157	same as hc		
OS	1.6837	0.1700	same as os		
CM	1.9080	0.0860	same as cc		
HA	1.4870	0.0157	same as hc		
OH	1.7210	0.2104	same as oh		
HO	0.0000	0.0000	same as ho		
NT	1.8240	0.1700	same as n3		

2. Results of MD simulations

2.1. Energy, Temperature and RMSD evolution along 5ns MD simulation in β5.

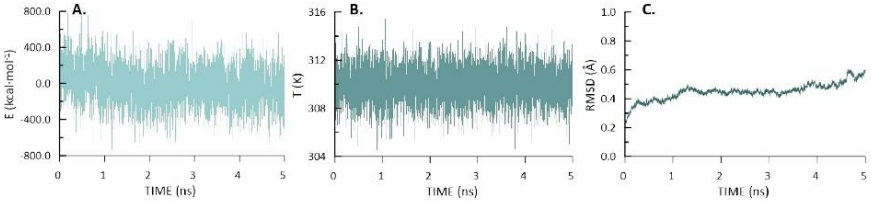


Figure S1. Time-dependent evolution of **A.** the total energy, **B.** temperature, and **C.** RMSD of the N, Cα and C atoms of the protein backbone along the 5 ns NVT-MD simulations of 20S proteasome with the covalent adduct formed between SalA and Thr1 in the active site of β5 subunit.

2.2. Energy, Temperature and RMSD evolution along 50ns MD simulation in β5

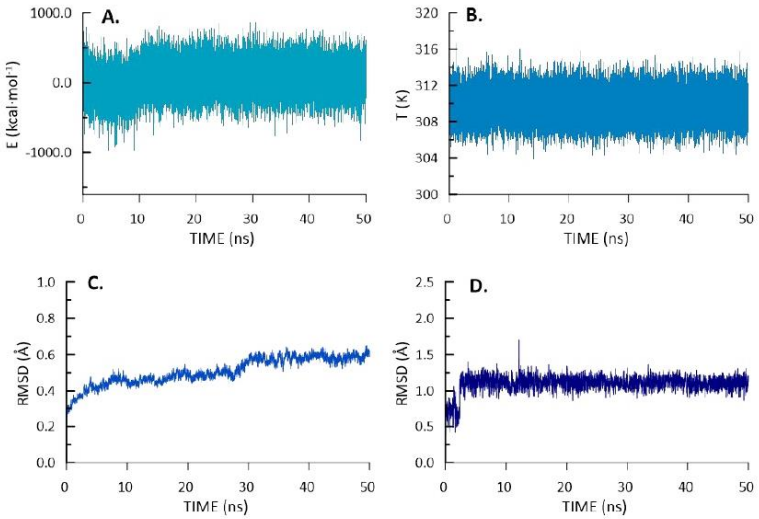


Figure S2. Time-dependent evolution of **A.** the total energy, **B.** temperature, **C.** RMSD of the N, Cα and C atoms of the protein backbone and **D.** RMSD of the atoms of SalA along the 50 ns NVT-MD simulations of 20 proteasome in the Michaelis complex form with SalA bound in the active site of β5 subunit.

2.3. Energy, Temperature and RMSD evolution along 50ns MD simulation in β2

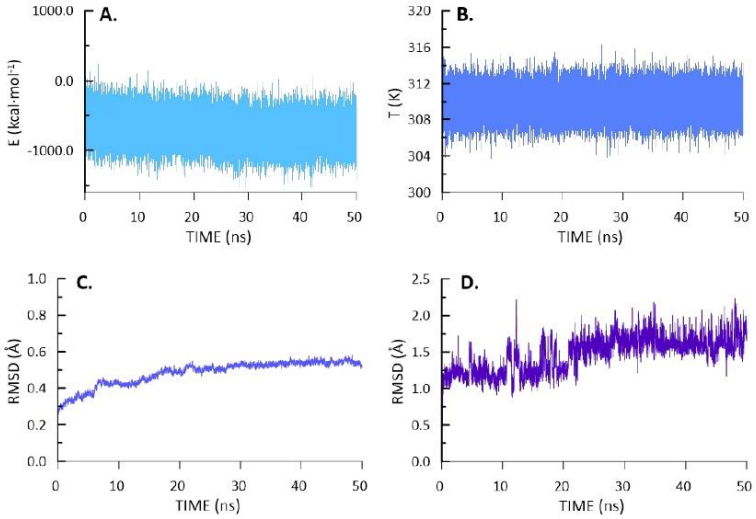


Figure S3. Time-dependent evolution of **A.** the total energy, **B.** temperature, **C.** RMSD of the N, Cα and C atoms of the protein backbone and **D.** RMSD of the atoms of SalA along the 50 ns NVT-MD simulations of 20 proteasome in the Michaelis complex form with SalA bound in the active site of β2 subunit.

2.4. Energy, Temperature and RMSD evolution along 50ns MD simulation in β1

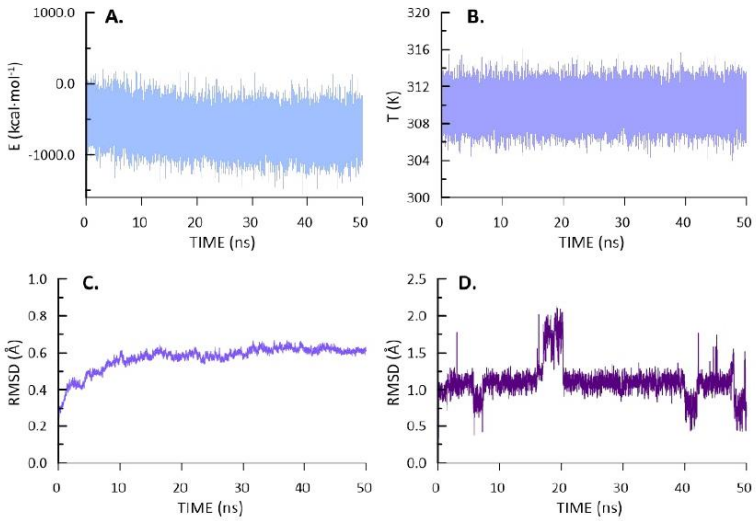


Figure S4. Time-dependent evolution of **A.** the total energy, **B.** temperature, **C.** RMSD of the N, Cα and C atoms of the protein backbone and **D.** RMSD of the atoms of SalA along the 50 ns NVT-MD simulations of 20 proteasome in the Michaelis complex form with SalA bound in the active site of β1 subunit.

2.5. Key distances analysis in β5, β2 and β1active sites after MD simulations

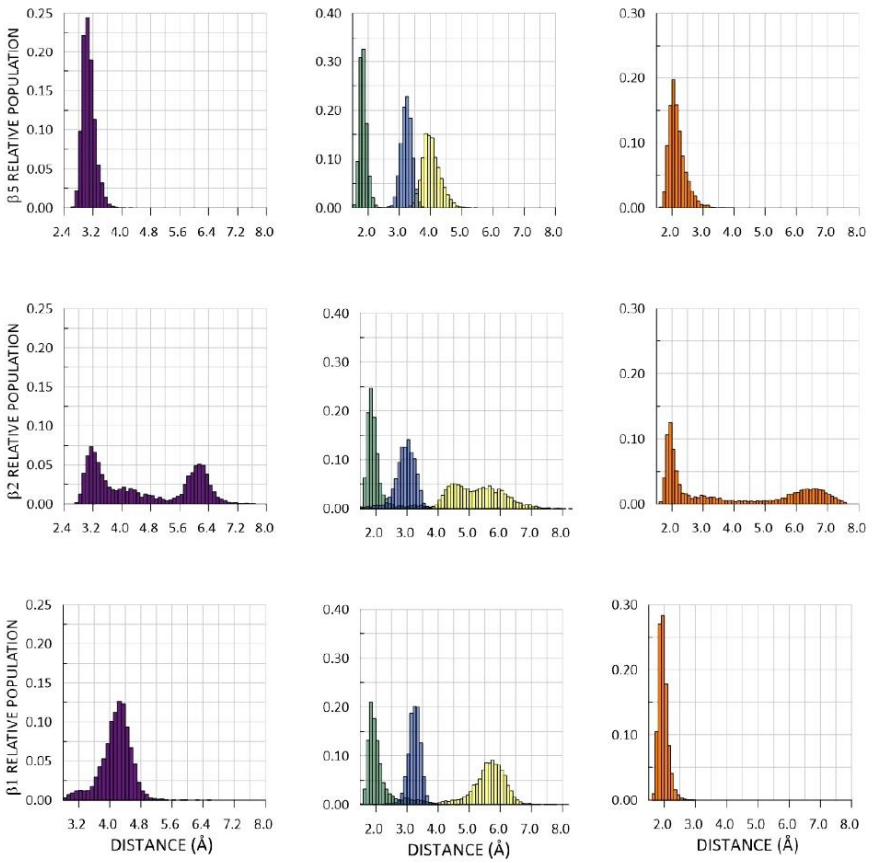


Figure S5. Analysis of the population of different E•SalA conformers based on 5000 structures produced during 50ns of NVT-MD simulation for the five chosen distances: $O^{\gamma Thr1}$-$C1^{SalA}$ (violet), $H^{\gamma Thr1}$-$N^{\zeta Lys33}$ (green), $H^{\gamma Thr1}$-$O2^{SalA}$ (yellow), $H^{\gamma Thr1}$-N^{Thr1} (blue) and H^{Gly47} – $O1^{SalA}$ (orange) distance for Michaelis complex of β5, β2 and β1 subunits of proteasome.

2.6. Model set up for QM/MM calculations

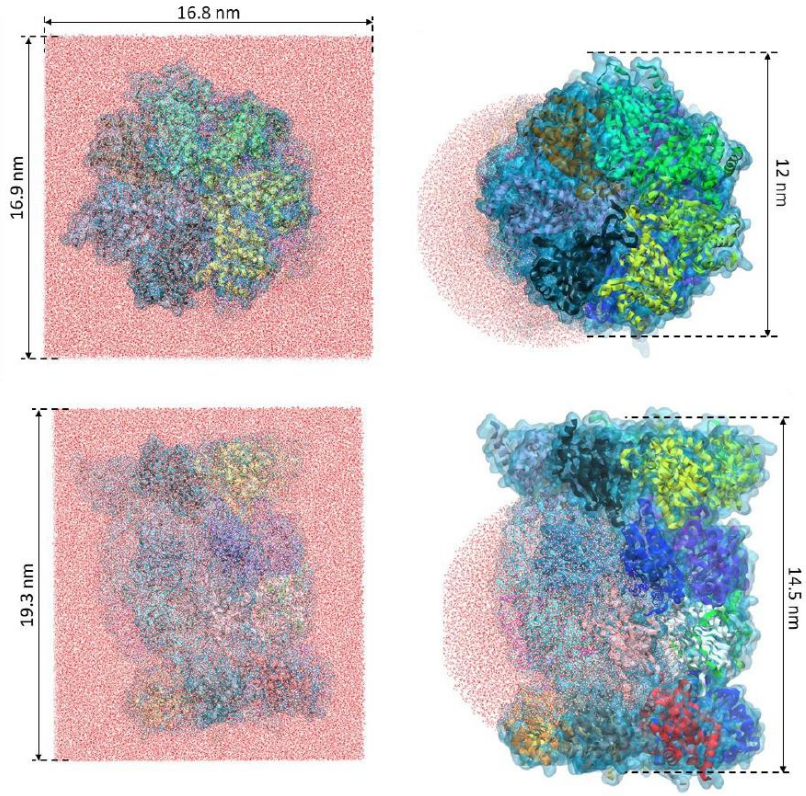


Figure S6. Reduction of the system from a water box (left panels) to a 60 Å water sphere centred on Thr1 residue of the catalytic subunit (right panels).

2.7. Interactions analysis between SalA and active site of β5, β2 and β1

As In this work, an additive hybrid QM/MM scheme was employed for the construction of the total Hamiltonian where the total energy is obtained from the sum of each contribution to the energy.

$$E_{QM/MM} = E_{QM} + E_{QM/MM}^{elect} + E_{QM/MM}^{vdW} + E_{MM} \quad \text{(S1)}$$

Here, E_{QM} describes the atoms in the QM region, $E_{QM/MM}$ defines the interaction between the QM and MM region and E_{MM} describes the rest of the MM region. Contribution of QM-MM interaction energies between SalA (treated quantum mechanically) with catalytic site, oxyanion hole, S1-pocket and S2-pocket of the protein (included in the MM region) was computed using the following expression:

$$E_{QM/MM}^{Int} = \sum \left\langle \Psi \left| \frac{q_{MM}}{r_{e,MM}} \right| \Psi \right\rangle + \sum\sum \frac{Z_{QM} q_{MM}}{r_{QM,MM}} + E_{QM/MM}^{vdW} \quad \text{(S2)}$$

This interaction energy can be exactly decomposed in a sum over residues (or group of residues) provided that the polarized wave function (Ψ) is employed to evaluate this energy contribution. The global polarization effect can be obtained from the gas phase energy difference between the polarized, Ψ, and non-polarized, Ψ_0, wave functions.

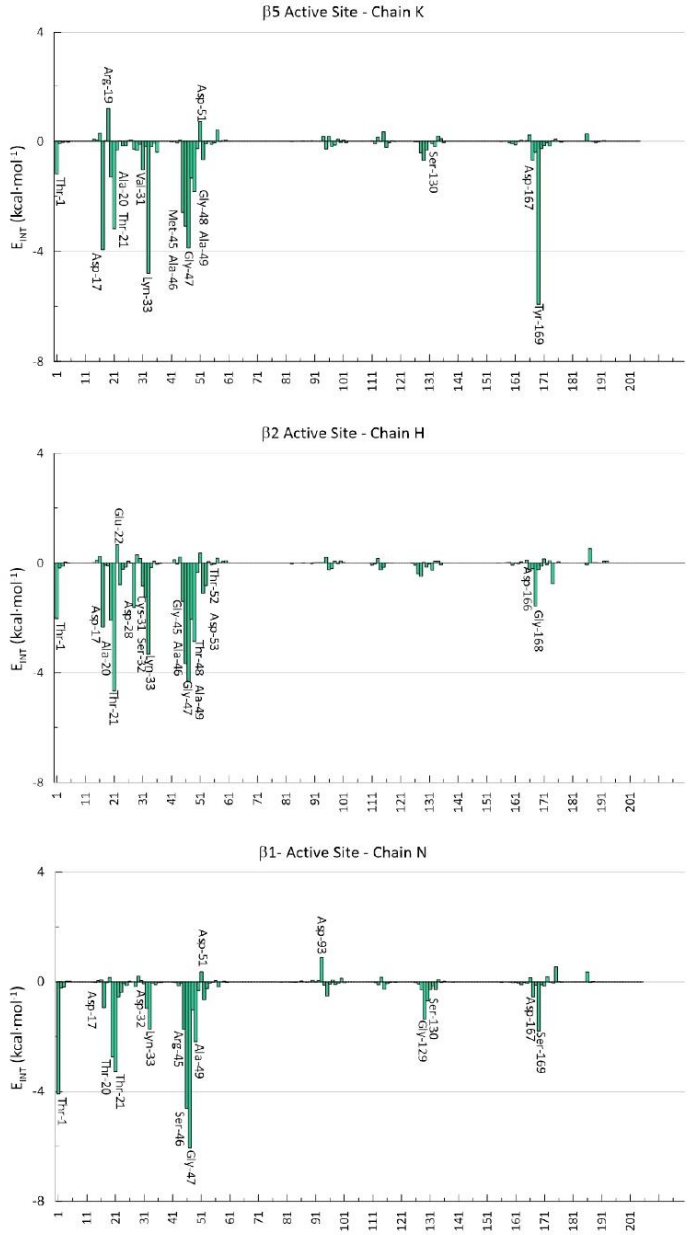


Figure S7. Key non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between SalA and the active site of β5, β2 and β1 subunits computed for 10000 structures at AM1/MM level.

Table S3. Non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between SalA and the key residues of β5 active site, averaged over 10000 structures at AM1/MM level.

Major Role	Residue	TOTAL	Electrostatic	Lennard Jones
Oxyanion Hole	Gly47	-3.9 ± 1.2	-1.3 ± 1.1	-2.6 ± 0.6
Catalytic Residues	Thr1	-1.2 ± 1.2	2.2 ± 1.0	-3.4 ± 0.6
	Asp17	-3.9 ± 0.8	-3.7 ± 0.8	-0.3 ± 0.0
	Lyn33	-4.8 ± 0.9	-0.9 ± 0.5	-3.9 ± 0.6
S1-Pocket	Ala20	-1.3 ± 0.6	0.5 ± 0.4	-1.8 ± 0.5
	Val31	-1.0 ± 0.3	0.2 ± 0.1	-1.2 ± 0.3
	Met45	-2.6 ± 0.4	-0.3 ± 0.2	-2.3 ± 0.4
	Ala46	-3.1 ± 0.7	-1.3 ± 0.4	-1.8 ± 0.4
	Gly48	-1.3 ± 0.3	0.4 ± 0.3	-1.7 ± 0.4
	Ala49	-1.8 ± 0.7	0.3 ± 0.2	-2.1 ± 0.6
S2-Pocket	Thr21	-3.2 ± 0.7	-0.5 ± 0.7	-2.7 ± 0.5
	Tyr169	-5.9 ± 0.8	-1.6 ± 0.4	-4.3 ± 0.7

Table S4. Non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between SalA and the key residues of β2 active site, averaged over 10000 structures at AM1/MM level.

Major Role	Residue	TOTAL	Electrostatic	Lennard Jones
Oxyanion Hole	Gly47	-4.3 ± 1.2	-1.1 ± 1.2	-3.2 ± 0.6
Catalytic Residues	Thr1	-2.0 ± 1.3	1.6 ± 1.0	-3.7 ± 0.8
	Asp17	-2.3 ± 0.8	-2.1 ± 0.8	-0.2 ± 0.0
	Lyn33	-3.3 ± 0.7	-0.4 ± 0.3	-3.0 ± 0.5
S1-Pocket	Ala20	-2.1 ± 0.7	-0.1 ± 0.5	-2.0 ± 0.4
	Cys31	-0.8 ± 0.4	0.0 ± 0.3	-0.8 ± 0.3
	Gly45	-1.4 ± 0.2	-0.5 ± 0.2	-0.9 ± 0.1
	Ala46	-3.7 ± 0.7	-1.7 ± 0.4	-1.9 ± 0.4
	Thr48	-2.0 ± 0.7	1.0 ± 0.5	-3.1 ± 0.6
	Ala49	-2.9 ± 0.8	0.0 ± 0.4	-2.9 ± 0.7
	Thr52	-1.1 ± 0.2	0.1 ± 0.2	-1.2 ± 0.2
	Asp 53	-0.8 ± 0.3	-0.7 ± 0.2	-0.1 ± 0.0
S2-Pocket	Thr21	-4.7 ± 1.2	-0.9 ± 0.9	-3.7 ± 0.7
	Gly168	-1.6 ± 0.4	-0.6 ± 0.3	-1.0 ± 0.2

Table S5. Non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between SalA and the key residues of β1 active site, averaged over 10000 structures at AM1/MM level.

Major Role	Residue	TOTAL	Electrostatic	Lennard Jones
Oxyanion Hole	Gly47	-6.1 ± 1.6	-3.1 ± 1.4	-2.9 ± 0.4
Catalytic Residues	Thr1	-4.1 ± 0.9	0.7 ± 0.7	-4.8 ± 0.6
	Asp17	-0.9 ± 0.6	-0.8 ± 0.6	-0.1 ± 0.0
	Lyn33	-1.7 ± 0.4	0.1 ± 0.1	-1.9 ± 0.4
S1-Pocket	Thr20	-2.7 ± 0.7	-0.5 ± 0.5	-2.2 ± 0.4
	Arg45	-1.7 ± 1.0	1.6 ± 0.8	-3.3 ± 0.5
	Ser46	-4.6 ± 0.8	-2.2 ± 0.6	-2.4 ± 0.4
	Ala49	-2.2 ± 0.7	0.2 ± 0.4	-2.4 ± 0.5
S2-Pocket	Thr21	-3.3 ± 0.7	-0.4 ± 0.5	-2.9 ± 0.5
	Ser48	-1.0 ± 0.4	0.0 ± 0.4	-1.1 ± 0.2
	Ser169	-1.8 ± 0.4	-0.2 ± 0.4	-1.6 ± 0.4

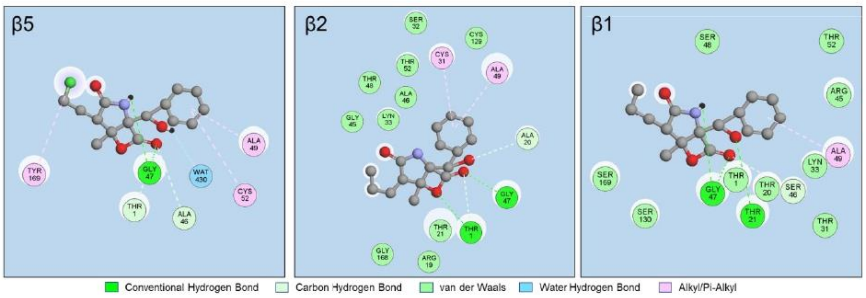


Figure S8. Schematic representation of key non-bonding interactions between SalA and the active site of β5, β2 and β1 subunits.

3. Results of SalA-assisted mechanism studies

3.1. Free energy surfaces computed at AM1/MM level

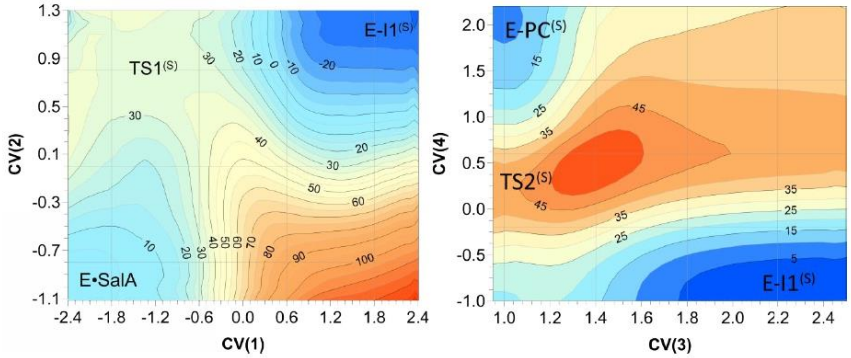


Figure S9. 2D-PMF surfaces computed at AM1/MM level of theory for SalA-assisted mechanism. The values of distances are given in Å and isolines are in kcal·mol^{-1}. Used collective variables are defined as follow: CV(1) = d($O^{\gamma Thr1}$-$H^{\gamma Thr1}$) - d($H^{\gamma Thr1}$-$O2^{SalA}$), CV(2) = d($O2^{SalA}$-$C1^{SalA}$) – d($C1^{SalA}$- $O^{\gamma Thr1}$), CV(3) = d($H^{\gamma Thr1}$- N^{Thr1}) and CV(4) = d($O2^{SalA}$-$C3^{SalA}$) – ($C3^{SalA}$-Cl).

3.2. Evolution of key distances along reaction pathway.

Table S6. Key distances (in Å) for located at M06-2X/MM stationary structures for SalA-assisted mechanism.

Distance	E•SalA$^{(S)}$	TS1$^{(S)}$	E-I1(S)	TS2$^{(S)}$	E-PC$^{(S)}$
$N^{\zeta Lys33}$ – $H^{\gamma Thr1}$	3.71	3.74	4.59	4.36	4.39
$H^{\gamma Thr1}$ – $O^{\gamma Thr1}$	0.98	1.12	2.45	2.36	2.35
$O^{\gamma Thr1}$ – $C1^{SalA}$	2.25	1.59	1.31	1.33	1.33
$C1^{SalA}$ – $O1^{SalA}$	1.19	1.23	1.22	1.21	1.21
$O1^{SalA}$ – H^{Gly47}	1.84	1.98	1.87	1.94	1.92
$C1^{SalA}$ – $O2^{SalA}$	1.38	1.52	2.60	2.63	2.63
$O2^{SalA}$ – $C3^{SalA}$	2.89	2.90	2.81	2.25	1.46
$C3^{SalA}$ – Cl^{SalA}	1.83	1.83	1.84	2.23	3.48
$O2^{SalA}$ – N^{Thr1}	3.70	3.42	2.82	2.97	3.12
$H^{\gamma Thr1}$ – $O2^{SalA}$	2.25	2.22	1.01	1.92	2.09
$H^{\gamma Thr1}$ – N^{Thr1}	1.89	1.44	1.82	1.06	1.03

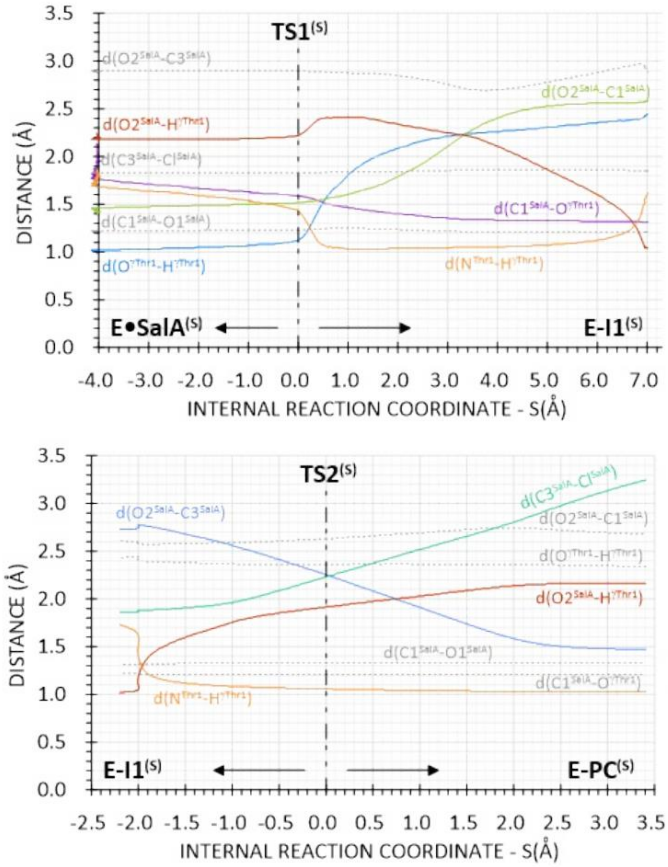


Figure S10. Evolution of key distances on minimum potential energy pathway along reaction progress registered in IRC calculations at M06-2X/MM level of theory, for SalA-assisted mechanism.

3.3. ChelpG charges on key atoms.

Table S7. Averaged ChelpG charges (in a.u.) computed on key atoms along SalA-assisted mechanism. Charges of hydrogens are summed into overall values of charges of heavy atoms.

Atom	E•SalA$^{(S)}$	TS1$^{(S)}$	E-I1$^{(S)}$	TS2$^{(S)}$	E-PC$^{(S)}$
$C3^{SalA}$	0.271 ± 0.08	0.288 ± 0.014	0.286 ± 0.010	0.462 ± 0.015	0.437 ± 0.016
$O2^{SalA}$	-0.623 ± 0.078	-0.833 ± 0.086	-0.478 ± 0.074	-1.056 ± 0.027	-0.834 ± 0.025
$C1^{SalA}$	0.930 ± 0.140	1.306 ± 0.176	1.023 ± 0.144	1.072 ± 0.113	1.138 ± 0.106
$O1^{SalA}$	0.577 ± 0.044	-0.799 ± 0.050	-0.673 ± 0.035	-0.680 ± 0.015	-0.665 ± 0.018
$O^{\gamma Thr1}$	-0.411 ± 0.013	-0.683 ± 0.020	-0.670 ± 0.064	-0.711 ± 0.041	-0.729 ± 0.050
N^{Thr1}	-0.280 ± 0.019	0.232 ± 0.059	-0.066 ± 0.045	0.610 ± 0.018	0.625 ± 0.016
$N^{\zeta Lys33}$	0.070 ± 0.079	-0.071 ± 0.044	-0.072 ± 0.037	-0.054 ± 0.029	-0.067 ± 0.039
Cl^{SalA}	-0.337 ± 0.003	-0.343 ± 0.004	-0.37 ± 0.018	-0.701 ± 0.024	-1.002 ± 0.014

3.4. Electrostatic potential generated by protein with and without Asp17 contribution

Table S8. Averaged electrostatic potential generated by protein (V^{prot}) on key atoms of SalA and residues involved in the covalent inhibition process via SalA-assisted mechanism. Values are given in kJ·mol^{-1}·e^{-1}.

Atom	E•SalA$^{(S)}$	TS1$^{(S)}$	E-I1$^{(S)}$	TS2$^{(S)}$	E-PC$^{(S)}$
$C3^{SalA}$	337.0 ± 13.3	339.8 ± 12.4	346.8 ± 11.8	331.7 ± 14.3	315.9 ± 13.3
$O2^{SalA}$	312.7 ± 14.0	319.3 ± 12.5	298.4 ± 16.1	276.0 ± 14.4	280.8 ± 13.2
$C1^{SalA}$	348.0 ± 15.1	349.3 ± 13.2	374.1 ± 22.6	336.1 ± 13.1	331.5 ± 12.3
$O1^{SalA}$	418.3 ± 19.3	434.8 ± 20.6	474.2 ± 28.2	428.5 ± 18.5	428.4 ± 14.4
$O^{\gamma Thr1}$	251.7 ± 14.4	264.4 ± 13.2	292.9 ± 20.7	258.3 ± 13.0	248.0 ± 12.4
N^{Thr1}	150.8 ± 20.1	151.1 ± 20.7	176.5 ± 24.7	109.6 ± 20.3	104.1 ± 22.0
$N^{\zeta Lys33}$	50.5 ± 22.8	61.8 ± 20.7	67.6 ± 24.6	29.7 ± 21.0	25.2 ± 18.2
Cl^{SalA}	372.5 ± 13.4	368.5 ± 12.4	383.2 ± 11.0	379.0 ± 13.4	389.5 ± 12.8

Table S9. Averaged electrostatic potential generated by protein without contribution of Asp17 (V^{prot}-V^{Asp17}) on key atoms of SalA and residues involved in the covalent inhibition process via SalA-assisted mechanism. Values are given in kJ·mol^{-1}·e^{-1}.

Atom	E•SalA$^{(S)}$	TS1$^{(S)}$	E-I1$^{(S)}$	TS2$^{(S)}$	E-PC$^{(S)}$
$C3^{SalA}$	477.3 ± 13.1	480.5 ± 12.5	489.5 ± 11.5	479.8 ± 14.2	476.8 ± 12.9
$O2^{SalA}$	497.9 ± 13.6	506.9 ± 12.5	492.8 ± 15.4	467.8 ± 14.3	471.5 ± 12.8
$C1^{SalA}$	535.5 ± 14.7	549.1 ± 13.2	579.3 ± 21.4	542.7 ± 13.1	538.3 ± 11.8
$O1^{SalA}$	592.5 ± 19.1	617.0 ± 20.7	655.7 ± 27.1	611.7 ± 18.4	612.1 ± 14.3
$O^{\gamma Thr1}$	513.8 ± 13.0	519.6 ± 13.0	545.3 ± 18.9	512.9 ± 12.6	503.0 ± 11.7
N^{Thr1}	425.9 ± 19.3	434.4 ± 20.3	445.7 ± 23.2	385.4 ± 20.1	380.6 ± 21.2
$N^{\zeta Lys33}$	483.9 ± 16.0	495.8 ± 17.5	501.8 ± 20.2	477.6 ± 15.5	463.8 ± 13.7
Cl^{SalA}	491.5 ± 13.2	487.5 ± 12.4	503.1 ± 10.8	498.7 ± 13.3	504.4± 12.7

4. Results of Lys-assisted mechanism studies

4.1. Free energy surfaces computed at AM1/MM level

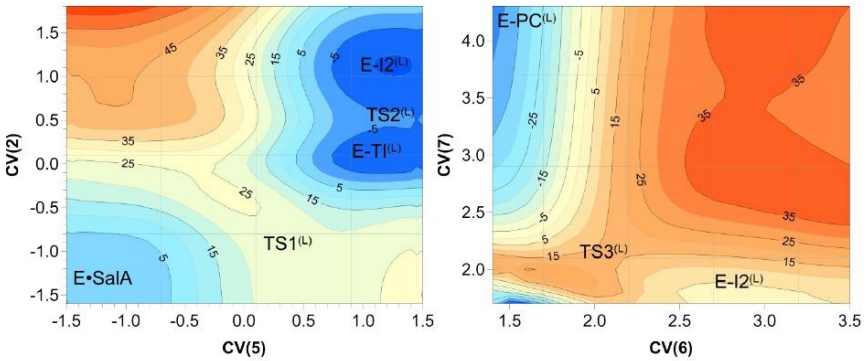


Figure S11. 2D-PMF surfaces computed at AM1/MM level of theory for Lys-assisted mechanism. Values of energies are given in kcal·mol^{-1}. The values of distances are given in Å and isolines are in kcal·mol^{-1}. Used collective variables are defined as follow: CV(2) = d(O2^{SalA}-C1^{SalA}) – d(C1^{SalA}-O$^{\gamma Thr1}$), CV(5) = d(O$^{\gamma Thr1}$-H$^{\gamma Thr1}$) – d(H$^{\gamma Thr1}$-N$^{\zeta Lys33}$), CV(6) = d(O2^{SalA}-C3^{SalA}), CV(7) = d(C3^{SalA}-Cl).

4.2. Evolution of key distances along reaction pathway

Table S10. Key distances (in Å) for located at M06-2X/MM stationary structures for Lys-assisted mechanism.

Distance	E•SalA$^{(L)}$	TS1$^{(L)}$	E-TI$^{(L)}$	TS2$^{(L)}$	E-I2$^{(L)}$	TS3$^{(L)}$	E-PC$^{(L)}$
N$^{\zeta Lys33}$ – H$^{\gamma Thr1}$	1.67	1.17	1.03	1.03	1.03	1.03	1.03
H$^{\gamma Thr1}$ – O$^{\gamma Thr1}$	1.02	1.41	2.14	2.20	2.23	2.28	2.29
O$^{\gamma Thr1}$ – C1^{SalA}	2.48	2.12	1.45	1.40	1.34	1.35	1.33
C^{SalA} – O1^{SalA}	1.20	1.21	1.26	1.23	1.21	1.21	1.21
O^{SalA} – H^{Gly47}	1.87	1.90	1.84	1.86	1.89	1.93	1.92
C^{SalA} – O2^{SalA}	1.36	1.40	1.57	1.86	2.58	2.73	2.69
O2^{SalA} – C3^{SalA}	2.98	3.07	3.02	2.83	2.51	2.34	1.45
C3^{SalA} – Cl^{SalA}	1.83	1.83	1.84	1.87	1.92	2.09	3.46
O2^{SalA} – N^{Thr1}	3.58	3.48	3.41	3.52	3.61	3.92	3.62
H$^{\gamma Thr1}$ – O2^{SalA}	3.79	3.91	4.20	4.38	4.77	5.04	4.88
H$^{\gamma Thr1}$ – N^{Thr1}	3.34	3.55	3.77	3.79	3.78	3.78	3.83

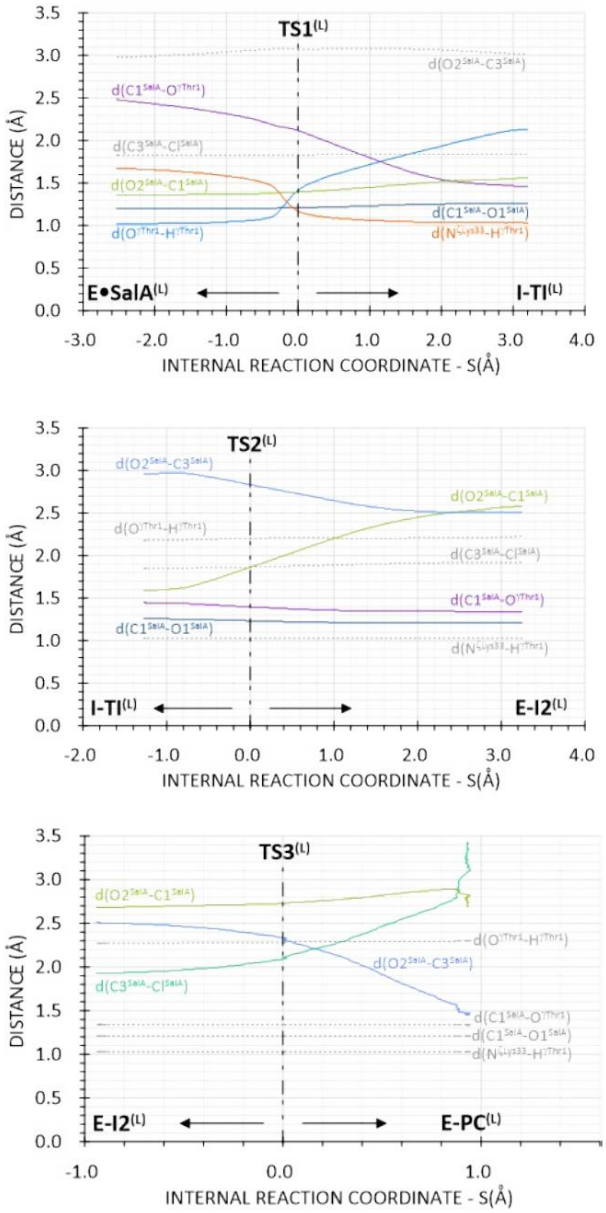


Figure S12. Evolution of key distances on minimum potential energy pathway along reaction progress registered in IRC calculations at M06-2X/MM level of theory, for Lys-assisted mechanism.

4.3. Averaged ChelpG charges on key atoms

Table S11. Averaged ChelpG charges (in a.u.) computed on key atoms along Lys-assisted mechanism. Charges of hydrogens are summed into overall values of charges of heavy atoms.

Atom	E•SalA$^{(L)}$	TS1$^{(L)}$	E-TI$^{(L)}$	TS2$^{(L)}$	E-I2$^{(L)}$	TS3$^{(L)}$	E-PC$^{(L)}$
C3^{SalA}	0.261 ± 0.001	0.287 ± 0.003	0.300 ± 0.011	0.304 ± 0.003	0.367 ± 0.010	0.492 ± 0.029	0.493 ± 0.019
O2^{SalA}	-0.548 ± 0.070	-0.651 ± 0.049	-0.835 ± 0.067	-0.887 ± 0.041	-1.010 ± 0.038	-1.023 ± 0.015	-0.804 ± 0.044
C1^{SalA}	0.829 ± 0.135	1.001 ± 0.127	1.355 ± 0.120	1.091 ± 0.208	0.875 ± 0.100	0.941 ± 0.035	1.009 ± 0.066
O1^{SalA}	-0.580 ± 0.037	-0.649 ± 0.050	-0.927 ± 0.024	-0.765 ± 0.063	-0.662 ± 0.016	-0.659 ± 0.005	-0.653 ± 0.013
O$^{\gamma Thr1}$	-0.420 ± 0.012	-0.832 ± 0.035	-0.728 ± 0.071	-0.476 ± 0.093	-0.372 ± 0.033	-0.464 ± 0.049	-0.466 ± 0.020
N^{Thr1}	-0.297 ± 0.003	-0.296 ± 0.011	-0.223 ± 0.014	-0.199 ± 0.015	-0.167 ± 0.015	-0.151 ± 0.021	-0.155 ± 0.020
N$^{\zeta Lys33}$	0.122 ± 0.020	0.623 ± 0.010	0.715 ± 0.032	0.677 ± 0.017	0.670 ± 0.013	0.689 ± 0.008	0.688 ± 0.010
Cl^{SalA}	-0.333 ± 0.008	-0.348 ± 0.007	-0.367 ± 0.012	-0.396 ± 0.015	-0.478 ± 0.032	-0.639 ± 0.031	-1.006 ± 0.019

4.4. Electrostatic potential generated by protein with and without Asp17 contribution

Table S12. Averaged electrostatic potential generated by protein (V^{prot}) on key atoms of SalA and residues involved in the covalent inhibition process via Lys-assisted mechanism. Values are given in kJ•mol^{-1}•e^{-1}.

Atom	E•SalA$^{(L)}$	TS1$^{(L)}$	E-TI$^{(L)}$	TS2$^{(L)}$	E-I2$^{(L)}$	TS3$^{(L)}$	E-PC$^{(L)}$
C3^{SalA}	340.2 ± 11.7	343.3 ± 12.8	341.2 ± 14.5	337.4 ± 16.0	345.8 ± 12.2	343.8 ± 11.5	309.5 ± 13.3
O2^{SalA}	322.5 ± 13.1	312.0 ± 12.5	309.9 ± 14.3	311.3 ± 16.0	315.4 ± 12.3	309.6 ± 11.7	282.2 ± 14.6
C1^{SalA}	357.0 ± 18.0	337.7 ± 13.0	327.1 ± 14.1	331.3 ± 14.3	333.3 ± 13.4	321.8 ± 13.1	294.3 ± 19.0
O1^{SalA}	430.1 ± 31.7	411.0 ± 18.6	424.3 ± 18.1	428.4 ± 18.6	427.6 ± 16.4	413.6 ± 18.3	379.3 ± 28.7
O$^{\gamma Thr1}$	232.4 ± 15.7	222.5 ± 13.9	239.2 ± 16.1	243.4 ± 15.6	247.7 ± 13.9	235.8 ± 13.5	206.3 ± 16.5
N^{Thr1}	177.5 ± 23.1	166.0 ± 21.2	160.4 ± 24.7	155.5 ± 25.2	169.1 ± 20.5	147.4 ± 22.4	118.1 ± 23.1
N$^{\zeta Lys33}$	38.8 ± 19.5	15.0 ± 20.1	-32.5 ± 22.5	-26.1 ± 21.9	-25.2 ± 20.8	-44.9 ± 19.8	-66.3 ± 21.6
Cl^{SalA}	374.3 ± 12.7	381.1 ± 12.4	378.1 ± 13.8	371.7 ± 15.6	376.6 ± 12.0	381.8 ± 11.3	387.4 ± 10.5

Table S13. Averaged electrostatic potential generated by protein without contribution of Asp17 (V^{prot}-V^{Asp17}) on key atoms of SalA and residues involved in the covalent inhibition process via Lys-assisted mechanism. Values are given in kJ·mol^{-1}·e^{-1}.

Atom	E•SalA$^{(L)}$	TS1$^{(L)}$	E-TI$^{(L)}$	TS2$^{(L)}$	E-I2$^{(L)}$	TS3$^{(L)}$	E-PC$^{(L)}$
C3^{SalA}	480.1 ± 11.4	485.1 ± 12.9	484.6 ± 14.4	480.3 ± 15.7	488.1 ± 12.3	484.9 ± 11.4	466.6 ± 13.0
O2^{SalA}	509.2 ± 12.7	505.0 ± 12.7	504.5 ± 14.2	502.5 ± 15.6	500.9 ± 12.4	488.6 ± 11.6	466.4 ± 14.1
C1^{SalA}	548.0 ± 17.8	538.8 ± 13.1	539.9 ± 13.8	545.4 ± 13.9	552.1 ± 12.9	540.0 ± 13.0	511.8 ± 17.9
O1^{SalA}	612.2 ± 31.7	598.9 ± 18.7	615.2 ± 18.1	620.1 ± 18.6	621.7 ± 16.4	607.4 ± 18.6	571.9 ± 27.6
O$^{\gamma Thr1}$	517.6 ± 14.6	508.9 ± 14.0	509.5 ± 15.7	512.9 ± 14.9	520.6 ± 12.8	507.3 ± 13.1	476.7 ± 15.4
N^{Thr1}	446.6 ± 22.2	436.3 ± 21.3	431.0 ± 24.7	425.6 ± 23.7	444.1 ± 20.7	421.9 ± 21.9	388.4 ± 22.4
N$^{\zeta Lys33}$	481.1 ± 15.3	473.6 ± 18.7	449.2 ± 17.8	454.0 ± 18.2	455.2 ± 15.7	443.1 ± 15.2	418.6 ± 16.8
Cl^{SalA}	492.7 ± 12.4	500.8 ± 12.5	498.7 ± 13.8	491.8 ± 15.4	495.8 ± 12.1	498.8 ± 11.1	501.1± 10.3

5. Comparison between SalA-assisted and Lys-assisted mechanism

5.1 Evolution of the Bürgi-Dunitz angle in the nuclephilic attack step.

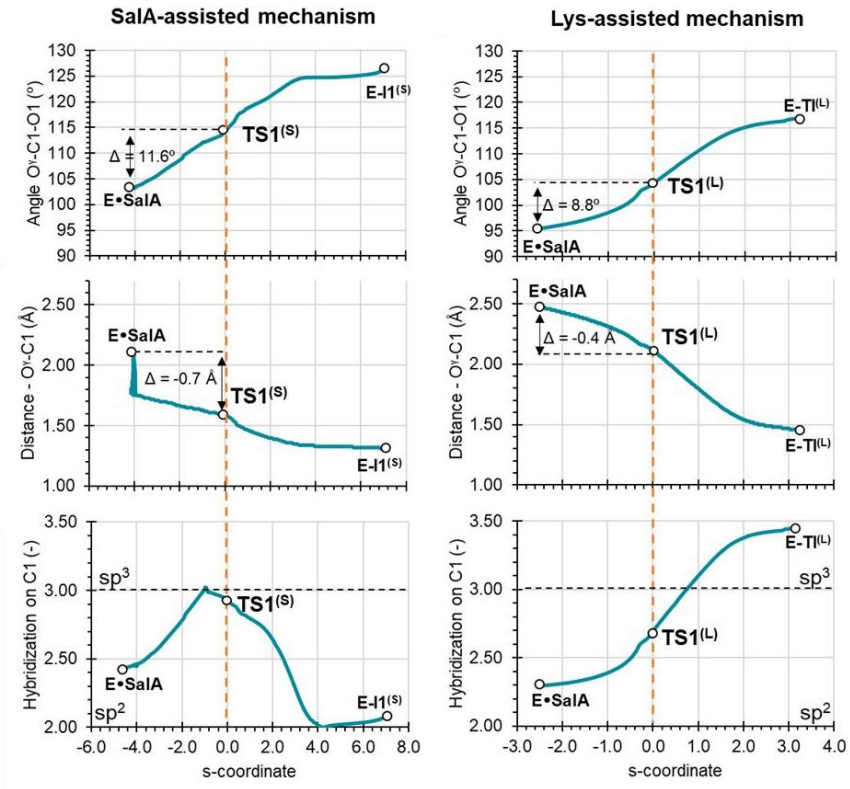


Figure S13. Evolution the Bürgi-Dunitz angle, $O\gamma^{Thr1}$-$C1^{SalA}$-$O1^{SalA}$ (top panels), distance of nucleophilic attack, $O\gamma^{Thr1}$-$C1^{SalA}$ (middle panels) and the hybridization of $C1^{SalA}$ (bottom panels) from E•SalA to first intermediate, obtained in the IRC calculations on the SalA-assisted mechanism (left column) and Lys-assisted mechanism (right column).

5.2 Kinetic isotope effects

Kinetic isotope effects (KIE) were computed for isotopic substitutions of key atoms from the rate limiting transition states (TSs) and the reactant state (RS) localized at M06-2X/6-31+(d,p)//AMBER level of theory. According to the Transition State Theory (TST), the ratio between the rate constants corresponding to the light atom "L" and the heavier isotope "H" can be computed as:

$$KIE = \frac{\left(\frac{Q_{TS}}{Q_{RS}}\right)_L}{\left(\frac{Q_{TS}}{Q_{RS}}\right)_H} e^{-\frac{1}{RT}(\Delta ZPE_L - \Delta ZPE_H)} \quad (1)$$

In eq 1, the total partition function, Q, was computed as the product of the translational, rotational, and vibrational partition functions for the isotopologs in RS and TS in the active site of Proteasome. The Born−Oppenheimer, rigid-rotor, and harmonic oscillator approximations were considered to independently compute the different contributions, without the scaling of vibrational frequencies, as explained and applied in previous papers.[1,2] Keeping in mind that reactants and TS are in a condensed media (the active site of a protein), contribution of translation and rotation to KIEs are negligible. Nevertheless, the full 3N × 3N Hessians have been subjected to a projection procedure to eliminate translational and rotational components, which give rise to small nonzero frequencies, as previously described.[3] Thus, it has been assumed that the 3N − 6 vibrational degrees of freedom are separable from the 6 translational and rotational degrees of freedom of the substrate.

Table S14. Kinetic isotope effects computed at M06-2X/6-31+G(d,p)//AMBER level of theory for the rate limiting step $TS1^{(S)}$ and $TS1^{(L)}$ determined for SalA-assisted and Lys-assisted mechanism.

Atoms	SalA-assisted	Lys-assisted
[1-^{18}O]-SalA	1.023 ± 0.016	1.020 ± 0.016
[1-^{14}C]-SalA	1.070 ± 0.004	1.053 ± 0.004
[2-^{18}O]-SalA	1.015 ± 0.003	1.007 ± 0.003
[γ-^{18}O]-Thr1	0.997 ± 0.010	1.042 ± 0.011
[^{15}N]-Thr1	1.000 ± 0.006	1.006 ± 0.006

6. SalA-assisted mechanism with standard protonation state of Lys33

6.1 MD simulations

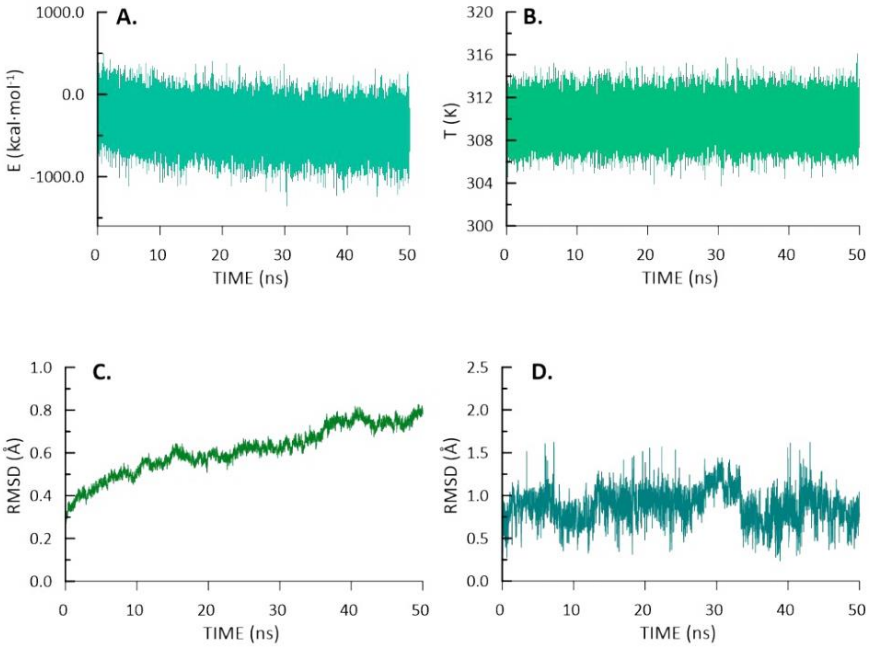


Figure S14. Time-dependent evolution of **A.** the total energy, **B.** temperature, **C.** RMSD of the N, Cα and C atoms of the protein backbone and **D.** RMSD of the atoms of SalA along the 50 ns NVT-MD simulations of 20 proteasome in the Michaelis complex form with SalA bound in the active site of β5 subunit.

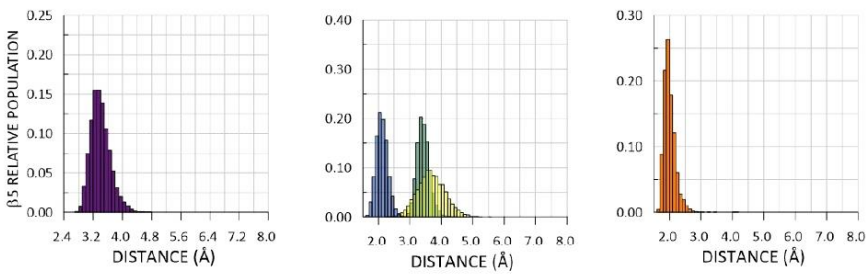


Figure S15. Analysis of the population of different E•SalA conformers based on 5000 structures produced during 50ns of NVT-MD simulation for the five chosen distances: $O^{\gamma Thr1}$-$C1^{SalA}$ (violet), $H^{\gamma Thr1}$-$N^{\zeta Lys33}$ (green), $H^{\gamma Thr1}$-$O2^{SalA}$ (yellow), $H^{\gamma Thr1}$-N^{Thr1} (blue) and H^{Gly47} – $O1^{SalA}$ (orange) for Michaelis complex of β5 subunit of proteasome.

6.2 Free energy surfaces computed at AM1/MM and M06-2X:AM1/MM level

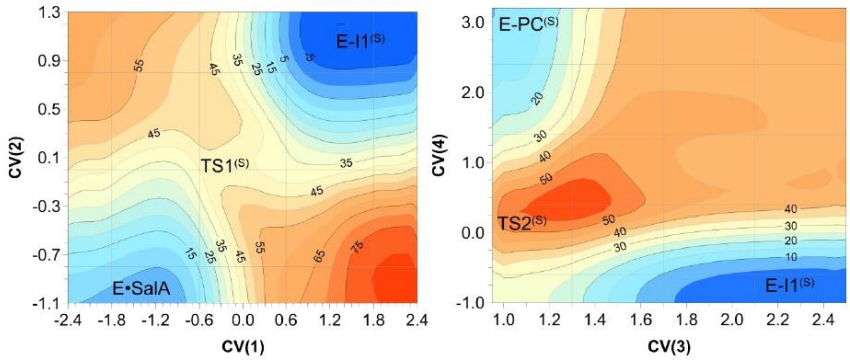


Figure S16. 2D-PMF surfaces computed at AM1/MM level of theory for SalA-assisted mechanism. The values of distances are given in Å and isolines are in kcal·mol^{-1}. Used collective variables are defined as follow: CV(1) = d($O^{\gamma Thr1}$-$H^{\gamma Thr1}$) - d($H^{\gamma Thr1}$-$O2^{SalA}$), CV(2) = d($O2^{SalA}$-$C1^{SalA}$) – d($O^{\gamma Thr1}$-$C1^{SalA}$), CV(3) = d($H^{\gamma Thr1}$- N^{Thr1}) and CV(4) = d($O2^{SalA}$-$C3^{SalA}$) – ($C3^{SalA}$-Cl).

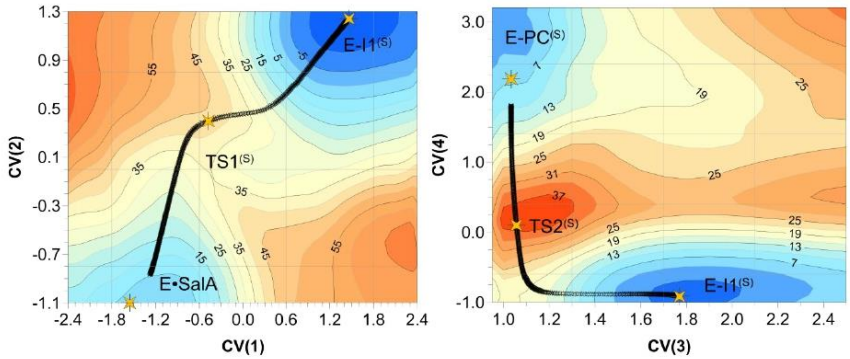


Figure S17. 2D-PMF surfaces corrected at M062X:AM1/MM level of theory for SalA-assisted mechanism. The values of distances are given in Å and isolines are in kcal·mol^{-1}. Used collective variables are defined as follow: CV(1) = d($O^{\gamma Thr1}$-$H^{\gamma Thr1}$) - d($H^{\gamma Thr1}$-$O2^{SalA}$), CV(2) = d($O2^{SalA}$-$C1^{SalA}$) – d($O^{\gamma Thr1}$-$C1^{SalA}$), CV(3) = d($H^{\gamma Thr1}$- N^{Thr1}) and CV(4) = d($O2^{SalA}$-$C3^{SalA}$) – ($C3^{SalA}$-Cl).

6.3 Structure of localized TSs at M06-2X/MM level

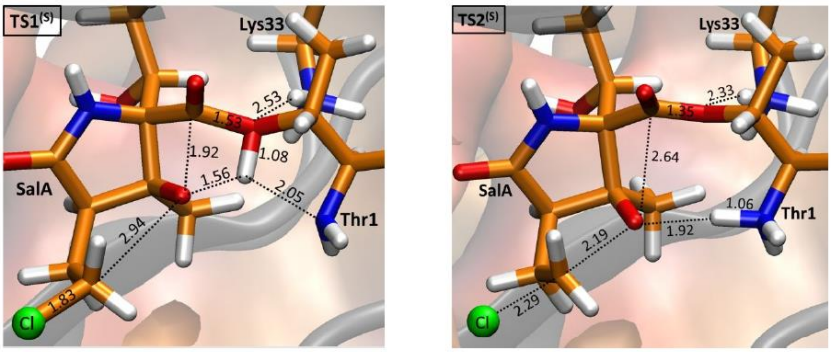


Figure S18. Transition state structures localized at M06-2X/6-31+(d,p)//AMBER level of theory. The key distances are given in Å.

6.4 Evolution of key distances along reaction pathway

Table S15. Key distances (in Å) for located at M06-2X/MM stationary structures for SalA-assisted mechanism.

Distance	E•SalA(S)	TS1(S)	E-I1(S)	TS2(S)	E-PC(S)
$N^{\zeta Lys33}$ – $H^{\gamma Thr1}$	3.22	4.05	4.73	4.56	4.66
$H^{\gamma Thr1}$ – $O^{\gamma Thr1}$	0.98	1.08	2.46	2.44	2.45
$O^{\gamma Thr1}$ – $C1^{SalA}$	2.57	1.53	1.34	1.35	1.35
$C1^{SalA}$ – $O1^{SalA}$	1.19	1.20	1.21	1.21	1.20
$O1^{SalA}$ – H^{Gly47}	1.95	1.96	1.96	1.97	2.00
$C1^{SalA}$ – $O2^{SalA}$	1.37	1.92	2.58	2.64	2.65
$O2^{SalA}$ – $C3^{SalA}$	3.04	2.94	2.81	2.19	1.47
$C3^{SalA}$ – Cl^{SalA}	1.82	1.83	1.83	2.29	3.65
$O2^{SalA}$ – N^{Thr1}	3.67	3.25	2.95	2.97	3.19
$H^{\gamma Thr1}$ – $O2^{SalA}$	2.53	1.56	1.00	1.92	2.16
$H^{\gamma Thr1}$ – N^{Thr1}	2.09	2.05	1.97	1.06	1.03

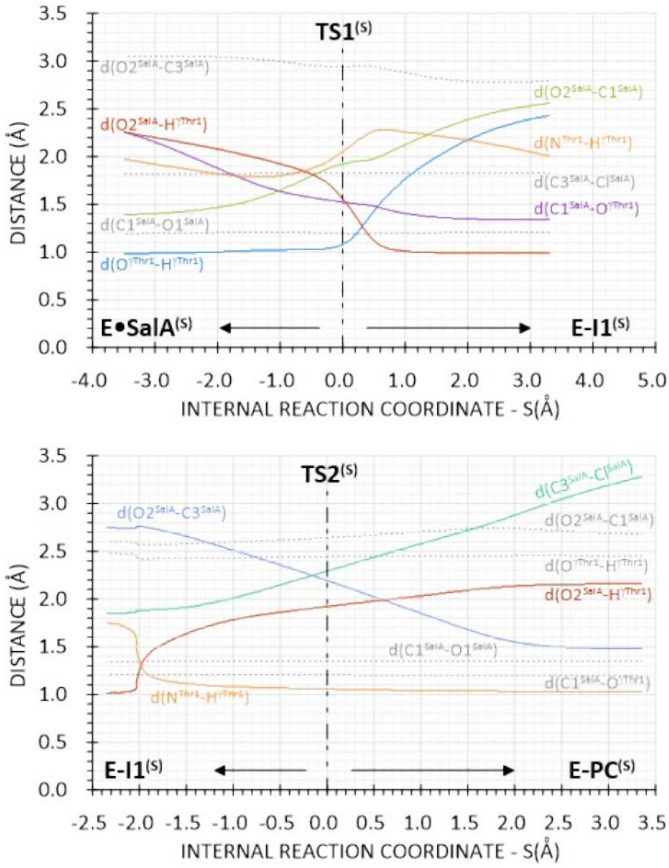


Figure S19. Evolution of key distances on minimum potential energy pathway along reaction progress registered in IRC calculations at M06-2X/MM level of theory, for SalA-assisted mechanism.

7. Summary of free energy profile calculations

Table S16. Free energy (ΔG) computed for SalA-assisted mechanism with (-NH_2) deprotonated and protonated (-NH_3) variant of Lys33, and for Lys-assisted mechanism at M06-2X/MM level. Values are given in kcal·mol^{-1}.

	Sal-assisted			Lys-assisted
	-NH_2 variant	-NH_3 variant		
E•SalA(S)	0.0	0.0	**E•SalA(L)**	0.0
TS(S)	36.9	35.2	**TS1(L)**	20.4
E-I1(S)	-11.3	-13.4	**E-TI(L)**	6.6
TS2(S)	17.5	16.7	**TS2(L)**	13.8
E-PC(S)	-18.1	-17.7	**E-I2(L)**	9.1
			TS3(L)	18.8
			E-PC(L)	-39.5

8. Cartesian coordinates of QM sub-set of atoms of the TS structures optimized at M06-2X/MM level

Table S17. Cartesian coordinates (in Å) of QM atoms for Transition State structures for SalA-assisted mechanism localized at M06-2X/6-31+G(d,p)/MM level.

	TS1(S)				TS2(S)		
ATOM	X	Y	Z	ATOM	X	Y	Z
C	-20.87223	-15.93573	-10.65489	C	-21.16647	-16.32651	-10.71200
C	-21.95706	-15.12460	-11.35483	C	-22.17512	-15.34643	-11.24068
C	-22.39084	-15.68811	-12.70520	C	-22.60088	-15.75150	-12.64397
C	-21.29397	-15.72836	-13.75398	C	-21.44409	-15.70994	-13.62663
O	-20.34879	-14.92482	-13.83641	O	-20.62167	-14.78364	-13.73073
N	-21.53000	-16.73392	-14.59535	N	-21.45437	-16.82161	-14.37061
H	-20.84951	-16.95809	-15.32693	H	-20.76009	-16.96479	-15.10748
H	-23.15774	-15.01171	-13.10637	H	-23.36528	-15.06456	-13.02289
H	-22.82319	-15.01264	-10.69775	H	-23.04218	-15.31344	-10.57768
H	-21.57660	-14.11791	-11.54977	H	-21.74960	-14.34228	-11.28014
H	-21.24291	-16.60947	-9.88411	H	-21.32378	-16.81795	-9.76208
H	-20.27805	-16.51023	-11.36147	H	-20.47889	-16.80410	-11.39428
O	-22.03814	-18.06518	-12.23869	O	-22.45113	-17.91653	-11.65825
C	-22.96782	-17.09371	-12.78559	C	-23.09963	-17.21805	-12.66475
C	-24.38738	-17.18360	-12.26930	C	-24.60877	-17.31262	-12.48378
H	-24.80622	-18.18653	-12.29517	H	-24.96340	-18.33989	-12.59500
H	-24.44327	-16.80538	-11.24456	H	-24.84720	-16.97376	-11.47107
H	-25.01668	-16.57971	-12.91866	H	-25.13963	-16.71185	-13.21387
C	-22.65481	-17.59982	-14.21844	C	-22.60153	-17.68407	-14.12956
C	-23.77886	-17.56055	-15.25543	C	-23.68272	-17.48810	-15.23891
C	-23.27660	-17.77135	-16.68472	C	-23.16587	-17.68316	-16.67664
C	-24.40377	-17.68408	-17.68667	C	-24.31574	-17.52109	-17.64955
C	-24.39625	-18.35674	-18.84089	C	-24.42527	-18.20987	-18.78934
C	-23.28520	-19.28511	-19.26066	C	-23.41006	-19.21852	-19.25859
C	-22.05554	-19.15082	-18.35973	C	-22.11919	-19.12386	-18.44677
C	-22.47474	-19.05421	-16.89203	C	-22.41979	-18.99097	-16.95289
H	-21.60261	-19.06904	-16.22960	H	-21.48144	-19.00504	-16.39402
H	-23.09500	-19.91991	-16.62474	H	-23.02572	-19.84648	-16.61976
H	-21.37800	-19.99333	-18.51834	H	-21.48919	-19.99815	-18.62973
H	-21.49698	-18.24281	-18.62287	H	-21.54449	-18.24445	-18.76514
H	-23.65550	-20.32038	-19.21057	H	-23.83389	-20.22958	-19.15116
H	-23.02606	-19.08973	-20.30782	H	-23.21200	-19.06652	-20.32612
H	-25.24061	-18.25757	-19.52047	H	-25.29953	-18.05608	-19.41946
H	-25.25035	-17.05248	-17.42266	H	-25.10512	-16.83117	-17.35713
H	-22.59278	-16.92508	-16.87211	H	-22.44556	-16.86243	-16.83592
O	-24.40139	-16.28007	-15.15795	O	-24.23892	-16.18972	-15.12924
H	-23.85727	-15.59693	-15.59554	H	-23.71773	-15.51274	-15.60522
H	-24.54055	-18.30549	-14.97895	H	-24.49795	-18.19471	-15.02961
C	-22.03564	-18.86542	-13.52520	C	-22.19517	-19.13420	-13.97664
O	-21.03938	-19.50438	-13.86784	O	-21.07976	-19.60056	-14.06492
O	-23.30294	-19.76262	-13.17865	O	-23.29344	-19.82976	-13.69965
C	-23.27964	-21.21091	-13.16092	C	-23.19319	-21.18783	-13.25287
C	-23.57625	-21.82631	-14.50185	C	-23.38281	-22.14190	-14.41635
H	-23.69764	-22.90741	-14.38319	H	-23.19281	-23.17224	-14.10615
H	-22.76845	-21.64191	-15.21174	H	-22.69396	-21.88146	-15.22254
H	-24.51490	-21.41131	-14.87721	H	-24.40805	-22.07746	-14.79162
H	-22.30577	-21.51759	-12.75697	H	-22.21559	-21.31610	-12.77208
C	-24.39905	-21.46468	-12.11070	C	-24.31216	-21.30066	-12.21074
N	-24.15601	-20.47132	-11.05782	N	-24.02372	-20.36420	-11.07800
H	-23.61669	-20.89094	-10.29580	H	-23.53085	-20.89176	-10.33966
H	-25.03358	-20.09459	-10.69292	H	-24.90935	-19.98970	-10.70084
H	-25.35814	-21.24804	-12.58781	H	-25.25493	-21.00647	-12.68354
C	-24.32992	-22.90248	-11.64663	C	-24.30791	-22.73704	-11.69364
O	-23.42919	-23.26260	-10.86595	O	-23.42636	-23.05608	-10.87207
N	-25.22341	-23.67639	-12.24150	N	-25.16501	-23.53363	-12.30206
H	-25.98135	-23.23663	-12.78780	H	-25.95632	-23.12410	-12.82790
C	-25.15786	-25.12219	-12.31452	C	-25.10842	-24.99115	-12.31162
H	-24.17838	-25.41959	-11.91929	H	-24.12959	-25.27748	-11.90753
C	-26.23981	-25.73837	-11.42567	C	-26.19736	-25.54389	-11.38718
H	-27.22180	-25.45555	-11.83403	H	-27.17108	-25.21900	-11.78184

C	-26.12222	-27.25307	-11.38989	**C**	-26.14802	-27.06081	-11.32857
H	-26.92673	-27.69009	-10.79115	**H**	-26.97085	-27.44825	-10.72195
H	-26.15835	-27.66150	-12.40264	**H**	-26.20622	-27.48454	-12.33323
H	-25.16509	-27.53827	-10.93834	**H**	-25.20479	-27.38020	-10.87057
O	-26.13478	-25.29171	-10.08073	**O**	-26.04451	-25.08935	-10.04964
H	-26.10926	-24.31856	-10.04282	**H**	-26.05257	-24.11464	-10.00375
C	-26.63502	-21.94488	-18.91640	**C**	-26.62803	-21.98960	-18.95034
H	-25.60281	-22.23470	-18.68612	**H**	-25.61194	-22.33955	-18.73030
H	-27.26865	-22.79517	-18.64169	**H**	-27.30618	-22.81023	-18.68994
C	-27.03047	-20.72577	-18.06757	**C**	-26.95120	-20.76829	-18.07941
H	-28.11953	-20.61242	-18.04233	**H**	-28.02821	-20.56576	-18.09280
H	-26.62520	-19.80682	-18.50556	**H**	-26.45449	-19.87742	-18.48144
C	-26.50532	-20.85611	-16.64189	**C**	-26.50637	-20.96404	-16.63076
H	-25.40973	-20.76859	-16.66265	**H**	-25.40748	-20.98843	-16.58420
H	-26.74463	-21.84717	-16.23282	**H**	-26.87080	-21.92552	-16.24470
C	-27.07437	-19.78564	-15.72177	**C**	-27.02219	-19.83362	-15.74960
H	-28.17366	-19.86149	-15.74741	**H**	-28.11932	-19.79828	-15.85200
H	-26.81654	-18.79373	-16.11686	**H**	-26.64079	-18.87995	-16.13994
N	-26.52786	-19.89020	-14.36772	**N**	-26.58618	-19.95776	-14.35735
H	-26.83018	-19.05435	-13.86236	**H**	-26.86143	-19.09098	-13.88957
H	-26.97522	-20.68211	-13.88835	**H**	-27.12078	-20.70262	-13.89332
H	-23.56785	-19.66744	-12.09525	**H**	-23.45137	-19.52266	-11.36069
Cl	-19.71134	-14.80617	-9.79660	**Cl**	-19.58149	-14.98393	-9.89774

Table S18. Cartesian coordinates (in Å) of QM atoms for Transition State structures for Lys-assisted mechanism localized at M06-2X/6-31+G(d,p)/MM level.

	TS1(L)				**TS2(L)**		
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
C	-20.79506	-15.97071	-10.67012	**C**	-20.80895	-15.98045	-10.7189
C	-21.94395	-15.18743	-11.2957	**C**	-21.90159	-15.16708	-11.39939
C	-22.37957	-15.70255	-12.66307	**C**	-22.37667	-15.75421	-12.72525
C	-21.30978	-15.59538	-13.74075	**C**	-21.30578	-15.77149	-13.79824
O	-20.41255	-14.735	-13.79533	**O**	-20.41449	-14.91144	-13.94955
N	-21.52325	-16.55459	-14.64555	**N**	-21.49251	-16.82639	-14.58729
H	-20.84561	-16.7223	-15.39535	**H**	-20.81317	-17.05123	-15.31752
H	-23.19124	-15.04155	-13.00192	**H**	-23.17456	-15.09864	-13.10162
H	-22.79659	-15.1623	-10.61361	**H**	-22.74612	-15.03991	-10.71708
H	-21.63092	-14.15023	-11.44572	**H**	-21.51834	-14.16609	-11.62164
H	-21.09856	-16.64186	-9.86836	**H**	-21.15329	-16.6198	-9.90827
H	-20.23991	-16.54138	-11.41373	**H**	-20.2365	-16.58478	-11.42095
O	-22.03205	-18.17934	-12.41466	**O**	-21.97852	-18.09657	-12.19565
C	-22.94515	-17.11362	-12.84171	**C**	-22.89544	-17.20043	-12.75059
C	-24.33616	-17.30154	-12.27909	**C**	-24.30548	-17.32474	-12.18969
H	-24.50707	-18.35865	-12.0795	**H**	-24.50674	-18.36452	-11.93274
H	-24.44057	-16.74394	-11.34227	**H**	-24.37668	-16.74442	-11.26228
H	-25.0778	-16.93032	-12.98897	**H**	-25.05123	-16.94067	-12.89043
C	-22.62263	-17.48164	-14.31904	**C**	-22.64585	-17.65081	-14.22322
C	-23.74078	-17.40637	-15.34798	**C**	-23.7502	-17.48426	-15.27469
C	-23.24859	-17.60587	-16.77992	**C**	-23.25087	-17.64569	-16.71553
C	-24.37634	-17.44268	-17.77136	**C**	-24.37397	-17.44004	-17.70437
C	-24.44994	-18.13506	-18.91151	**C**	-24.45355	-18.09013	-18.86942
C	-23.4262	-19.15715	-19.33307	**C**	-23.43463	-19.10122	-19.33032
C	-22.1619	-19.08724	-18.47429	**C**	-22.17061	-19.06796	-18.46923
C	-22.51198	-18.92893	-16.99468	**C**	-22.51747	-18.96076	-16.98476
H	-21.60784	-18.9592	-16.37859	**H**	-21.61328	-19.0209	-16.37417
H	-23.14284	-19.76232	-16.65571	**H**	-23.14999	-19.80949	-16.68487
H	-21.55136	-19.98005	-18.62937	**H**	-21.56243	-19.9568	-18.65478
H	-21.55448	-18.22667	-18.78463	**H**	-21.5583	-18.19989	-18.74498
H	-23.86964	-20.1611	-19.2402	**H**	-23.88132	-20.10735	-19.27514
H	-23.18889	-19.01142	-20.39312	**H**	-23.19827	-18.91975	-20.38528
H	-25.29662	-17.97608	-19.57736	**H**	-25.29525	-17.8942	-19.53185
H	-25.16351	-16.73716	-17.50784	**H**	-25.14845	-16.72747	-17.42236
H	-22.52421	-16.79237	-16.94665	**H**	-22.52223	-16.83033	-16.85055
O	-24.37849	-16.14554	-15.19598	**O**	-24.37691	-16.22147	-15.11209
H	-23.85852	-15.42393	-15.61085	**H**	-23.84684	-15.50255	-15.51764
H	-24.48438	-18.15924	-15.08368	**H**	-24.53068	-18.2243	-15.06774
C	-21.96934	-18.72782	-13.69633	**C**	-22.19388	-19.06564	-13.77219
O	-21.12097	-19.52039	-14.03638	**O**	-21.15404	-19.63697	-14.11081
O	-23.76572	-19.84728	-13.7128	**O**	-23.35128	-19.80084	-13.51107

C	-23.44789	-21.10976	-13.18548
C	-23.50771	-22.17378	-14.28378
H	-23.12741	-23.14354	-13.94611
H	-22.90859	-21.83053	-15.13257
H	-24.5354	-22.33789	-14.62887
H	-22.41306	-21.0927	-12.79685
C	-24.34904	-21.42457	-11.95098
N	-23.93964	-20.60322	-10.81304
H	-23.38952	-21.18731	-10.18084
H	-24.7665	-20.34236	-10.28604
H	-25.38279	-21.17638	-12.21017
C	-24.30355	-22.9184	-11.62884
O	-23.41362	-23.36875	-10.8808
N	-25.21826	-23.66478	-12.25082
H	-25.96644	-23.19777	-12.77908
C	-25.20766	-25.12162	-12.29819
H	-24.23523	-25.43576	-11.89808
C	-26.29661	-25.66025	-11.36643
H	-27.27237	-25.33437	-11.75902
C	-26.25314	-27.17689	-11.30639
H	-27.09648	-27.5714	-10.73239
H	-26.26917	-27.59228	-12.31645
H	-25.32581	-27.49266	-10.81472
O	-26.14454	-25.20083	-10.03115
H	-26.07048	-24.22847	-10.00179
C	-26.59258	-21.95038	-18.88321
H	-25.58426	-22.33413	-18.68602
H	-27.29325	-22.73752	-18.58401
C	-26.83485	-20.70692	-18.00602
H	-27.90245	-20.46204	-17.97888
H	-26.31858	-19.83694	-18.42699
C	-26.33546	-20.9277	-16.57031
H	-25.24083	-21.02254	-16.56957
H	-26.74491	-21.86136	-16.16483
C	-26.75348	-19.76878	-15.67357
H	-27.84406	-19.68861	-15.65619
H	-26.3608	-18.81991	-16.05392
N	-26.28795	-19.91505	-14.26506
H	-26.68369	-19.12707	-13.73411
H	-26.6993	-20.77582	-13.85079
H	-25.12925	-19.89466	-14.08409
Cl	-19.59687	-14.81352	-9.9026

C	-23.16693	-21.14059	-13.01358
C	-23.26262	-22.10277	-14.19094
H	-23.00484	-23.12636	-13.90482
H	-22.59119	-21.76827	-14.98312
H	-24.28545	-22.13811	-14.58671
H	-22.17872	-21.19763	-12.5457
C	-24.2403	-21.39159	-11.93158
N	-24.00164	-20.58768	-10.74844
H	-23.36856	-21.09342	-10.12574
H	-24.87618	-20.50201	-10.23863
H	-25.21856	-21.11923	-12.34412
C	-24.25733	-22.89179	-11.62766
O	-23.39055	-23.36755	-10.87354
N	-25.18257	-23.60594	-12.26461
H	-25.91438	-23.13403	-12.81093
C	-25.17994	-25.06111	-12.32777
H	-24.22077	-25.39015	-11.91092
C	-26.29933	-25.60351	-11.43509
H	-27.25955	-25.24702	-11.84026
C	-26.29203	-27.12129	-11.40862
H	-27.15585	-27.50405	-10.85721
H	-26.29746	-27.51619	-12.42653
H	-25.38314	-27.4708	-10.90661
O	-26.16621	-25.18341	-10.08463
H	-26.09065	-24.21204	-10.02358
C	-26.57758	-21.91583	-18.91486
H	-25.55864	-22.26679	-18.71286
H	-27.25425	-22.72062	-18.60718
C	-26.8629	-20.67101	-18.05334
H	-27.93483	-20.4456	-18.05108
H	-26.35216	-19.79387	-18.465
C	-26.39718	-20.87669	-16.60306
H	-25.29917	-20.91904	-16.57453
H	-26.77617	-21.82597	-16.20678
C	-26.89363	-19.73159	-15.73215
H	-27.98436	-19.69914	-15.72226
H	-26.52167	-18.76515	-16.08518
N	-26.47081	-19.85557	-14.29672
H	-26.8046	-19.01792	-13.78808
H	-26.89512	-20.69639	-13.84254
H	-25.44585	-19.89753	-14.18361
Cl	-19.60246	-14.80912	-9.9112

TS3$^{(L)}$

ATOM	X	Y	Z
C	-20.81207	-15.64586	-10.88353
C	-21.86184	-14.89374	-11.65958
C	-22.28749	-15.63234	-12.92731
C	-21.15643	-15.85199	-13.91827
O	-20.20347	-15.08273	-14.1442
N	-21.33863	-17.01201	-14.55699
H	-20.68001	-17.29664	-15.28382
H	-23.06161	-15.0419	-13.42625
H	-22.73224	-14.71314	-11.02314
H	-21.46477	-13.92031	-11.96076
H	-21.07546	-16.18153	-9.98206
H	-20.04396	-16.10928	-11.4791
O	-21.77801	-17.56964	-11.79494
C	-22.72243	-17.09247	-12.6369
C	-24.14732	-17.25676	-12.11655
H	-24.28287	-18.28848	-11.78224
H	-24.26469	-16.61346	-11.23734
H	-24.91208	-16.98555	-12.84913
C	-22.52108	-17.74203	-14.12572
C	-23.6764	-17.5182	-15.15162
C	-23.23891	-17.68176	-16.62228
C	-24.41112	-17.43184	-17.54677
C	-24.56605	-18.03832	-18.7283
C	-23.59288	-19.04616	-19.28275
C	-22.28634	-19.06237	-18.48771
C	-22.54761	-18.99963	-16.98086
H	-21.60008	-19.08745	-16.44455
H	-23.17223	-19.85387	-16.67788

H	-21.70497	-19.95452	-18.73219
H	-21.67454	-18.1955	-18.76614
H	-24.05423	-20.04628	-19.24517
H	-23.40531	-18.82842	-20.34042
H	-25.43978	-17.80385	-19.33436
H	-25.15776	-16.71811	-17.20244
H	-22.50086	-16.87892	-16.77975
O	-24.27359	-16.24686	-14.99913
H	-23.77214	-15.54776	-15.46748
H	-24.47168	-18.23665	-14.92206
C	-22.23767	-19.22134	-13.91517
O	-21.16979	-19.76681	-14.09389
O	-23.34387	-19.86482	-13.50646
C	-23.20577	-21.2438	-13.06028
C	-23.35993	-22.16745	-14.26067
H	-23.12704	-23.2012	-13.99102
H	-22.69671	-21.85791	-15.07142
H	-24.39298	-22.1699	-14.62941
H	-22.2145	-21.33564	-12.6068
C	-24.27705	-21.4598	-11.97342
N	-24.02489	-20.62783	-10.81668
H	-23.3674	-21.10167	-10.19506
H	-24.8909	-20.53504	-10.29363
H	-25.25579	-21.19623	-12.39072
C	-24.29007	-22.95454	-11.64497
O	-23.40397	-23.42092	-10.90959
N	-25.23624	-23.67015	-12.24519
H	-25.96243	-23.20103	-12.8019
C	-25.21483	-25.12244	-12.32434
H	-24.25588	-25.44897	-11.90643
C	-26.33985	-25.70543	-11.46991
H	-27.29994	-25.37896	-11.90047
C	-26.28179	-27.22321	-11.46454
H	-27.13687	-27.64465	-10.92853
H	-26.26584	-27.60208	-12.48907
H	-25.36602	-27.55426	-10.96356
O	-26.2553	-25.28599	-10.11608
H	-26.16173	-24.31726	-10.06333
C	-26.59687	-21.92841	-18.92153
H	-25.5752	-22.26593	-18.71162
H	-27.26661	-22.74317	-18.62593
C	-26.91208	-20.68774	-18.07099
H	-27.98884	-20.48822	-18.06938
H	-26.42442	-19.80064	-18.48836
C	-26.43974	-20.88032	-16.6248
H	-25.34067	-20.89829	-16.60309
H	-26.79649	-21.83829	-16.22883
C	-26.95908	-19.74478	-15.75896
H	-28.05012	-19.73536	-15.75225
H	-26.60292	-18.7733	-16.11312
N	-26.53499	-19.86147	-14.32456
H	-26.85338	-19.01495	-13.82018
H	-26.96573	-20.69247	-13.85889
H	-25.51415	-19.92007	-14.21744
Cl	-19.63876	-14.15606	-10.01046

Table S19. Cartesian coordinates (in Å) of QM atoms for Transition State structures for SalA-assisted mechanism and standard protonation for Lys33, localized at M06-2X/6-31+G(d,p)/MM level.

	TS1				TS2		
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
H	-23.24254	-19.31878	-12.26995	**H**	-23.35763	-19.5652	-11.26219
C	-20.93608	-15.97433	-10.57632	**C**	-21.13933	-16.39941	-10.76747
C	-22.01649	-15.18932	-11.32002	**C**	-22.12404	-15.38206	-11.2626
C	-22.40074	-15.74898	-12.68648	**C**	-22.58351	-15.77175	-12.65895
C	-21.27682	-15.77341	-13.70636	**C**	-21.4301	-15.76186	-13.64859
O	-20.362	-14.94948	-13.81525	**O**	-20.60071	-14.85476	-13.77331
N	-21.47319	-16.81257	-14.53769	**N**	-21.45363	-16.89632	-14.36951
H	-20.81323	-16.99179	-15.29984	**H**	-20.7705	-17.04729	-15.1163
H	-23.14669	-15.06241	-13.11159	**H**	-23.33643	-15.06831	-13.02803
H	-22.90846	-15.10105	-10.69449	**H**	-22.97266	-15.3191	-10.57854
H	-21.66486	-14.16872	-11.49409	**H**	-21.6612	-14.39549	-11.31001
H	-21.31324	-16.53344	-9.72117	**H**	-21.25887	-16.86488	-9.79903
H	-20.39486	-16.6618	-11.22412	**H**	-20.43371	-16.84944	-11.44999
O	-22.24964	-18.13574	-12.07984	**O**	-22.4419	-17.92315	-11.65885
C	-23.00364	-17.15688	-12.77248	**C**	-23.10049	-17.22887	-12.66183
C	-24.46707	-17.1633	-12.34999	**C**	-24.60649	-17.31169	-12.44182
H	-24.87341	-18.17088	-12.2529	**H**	-24.93545	-18.35408	-12.42288
H	-24.55267	-16.70714	-11.35947	**H**	-24.82578	-16.88956	-11.45669
H	-25.07192	-16.61432	-13.06562	**H**	-25.1655	-16.78043	-13.20609
C	-22.63698	-17.62934	-14.22583	**C**	-22.62703	-17.71695	-14.12916
C	-23.71849	-17.51748	-15.32122	**C**	-23.68693	-17.48476	-15.25198
C	-23.19689	-17.6763	-16.75218	**C**	-23.14868	-17.65127	-16.68418
C	-24.33298	-17.51968	-17.7386	**C**	-24.28778	-17.47295	-17.66617
C	-24.39348	-18.18616	-18.89523	**C**	-24.39477	-18.15795	-18.80953
C	-23.34064	-19.15941	-19.35554	**C**	-23.38376	-19.17195	-19.27555
C	-22.07707	-19.07658	-18.49901	**C**	-22.09565	-19.08854	-18.45814
C	-22.42429	-18.96964	-17.01359	**C**	-22.39663	-18.95599	-16.96429
H	-21.50942	-18.9883	-16.41562	**H**	-21.45759	-18.97033	-16.40654
H	-23.0294	-19.83848	-16.71231	**H**	-22.99524	-19.81986	-16.63411
H	-21.43702	-19.94302	-18.67842	**H**	-21.47067	-19.96588	-18.63935
H	-21.49611	-18.18999	-18.77911	**H**	-21.51344	-18.21365	-18.77109
H	-23.75097	-20.17983	-19.30338	**H**	-23.81322	-20.18228	-19.18109
H	-23.1125	-18.96455	-20.40901	**H**	-23.1822	-19.01309	-20.34085
H	-25.24189	-18.0222	-19.55725	**H**	-25.24945	-17.97332	-19.458
H	-25.12805	-16.82822	-17.46526	**H**	-25.05628	-16.75109	-17.39523
H	-22.49365	-16.83958	-16.88715	**H**	-22.42878	-16.82764	-16.81373
O	-24.37289	-16.27582	-15.1722	**O**	-24.27689	-16.21471	-15.10667
H	-23.86634	-15.53602	-15.57959	**H**	-23.77353	-15.49535	-15.55391
H	-24.48273	-18.28357	-15.12276	**H**	-24.49929	-18.20658	-15.09309
C	-22.15142	-19.02399	-13.78212	**C**	-22.27761	-19.17953	-13.97847
O	-21.18998	-19.66161	-14.09661	**O**	-21.18895	-19.68961	-14.06502
O	-23.35691	-19.78426	-13.24046	**O**	-23.40458	-19.87067	-13.6874
C	-23.275	-21.24264	-13.01141	**C**	-23.26095	-21.22079	-13.20198
C	-23.50875	-22.00045	-14.29577	**C**	-23.48622	-22.20947	-14.332
H	-23.33712	-23.06519	-14.11318	**H**	-23.2368	-23.2255	-14.01747
H	-22.83197	-21.67168	-15.0875	**H**	-22.85485	-21.94018	-15.18186
H	-24.54707	-21.91371	-14.63396	**H**	-24.53288	-22.2262	-14.65529
H	-22.28016	-21.42866	-12.5949	**H**	-22.25385	-21.32165	-12.78308
C	-24.33537	-21.44757	-11.91301	**C**	-24.2871	-21.32373	-12.06716
N	-24.01511	-20.55075	-10.81868	**N**	-23.89521	-20.41869	-10.94715
H	-23.50525	-21.05586	-10.08943	**H**	-23.3314	-20.95653	-10.26295
H	-24.87098	-20.1853	-10.40787	**H**	-24.7448	-20.06546	-10.48046
H	-25.32339	-21.20041	-12.31788	**H**	-25.28321	-21.03013	-12.40585
C	-24.30135	-22.93045	-11.56447	**C**	-24.28591	-22.77419	-11.59149
O	-23.38822	-23.34675	-10.83615	**O**	-23.38519	-23.12166	-10.81402
N	-25.20911	-23.66411	-12.19105	**N**	-25.177	-23.53288	-12.20041
H	-25.96279	-23.21161	-12.72686	**H**	-25.93269	-23.10065	-12.75214
C	-25.1566	-25.11565	-12.28476	**C**	-25.11181	-24.98894	-12.28421
H	-24.17808	-25.42298	-11.89804	**H**	-24.1328	-25.28645	-11.89024
C	-26.24192	-25.71792	-11.38914	**C**	-26.19812	-25.57406	-11.3749
H	-27.22288	-25.42421	-11.79411	**H**	-27.17756	-25.25894	-11.76678
C	-26.1423	-27.23261	-11.35203	**C**	-26.13356	-27.08943	-11.32752
H	-26.96015	-27.65941	-10.76543	**H**	-26.95399	-27.48609	-10.72442
H	-26.16804	-27.64144	-12.36477	**H**	-26.18654	-27.51035	-12.33354
H	-25.19596	-27.52684	-10.88521	**H**	-25.18941	-27.40219	-10.86876
O	-26.12461	-25.26916	-10.0474	**O**	-26.05081	-25.12564	-10.03559

H	-26.1113	-24.29663	-10.00313	**H**	-26.07192	-24.15447	-9.98143
C	-26.62472	-21.93187	-18.89926	**C**	-26.62971	-21.92613	-18.93399
H	-25.60459	-22.26732	-18.67818	**H**	-25.61234	-22.26855	-18.7104
H	-27.29638	-22.74495	-18.60592	**H**	-27.30796	-22.73351	-18.63859
C	-26.95131	-20.68487	-18.06507	**C**	-26.947	-20.67325	-18.10893
H	-28.02878	-20.49264	-18.07572	**H**	-28.0242	-20.47821	-18.11472
H	-26.46481	-19.79833	-18.48533	**H**	-26.46162	-19.79134	-18.53964
C	-26.49332	-20.85354	-16.61186	**C**	-26.48193	-20.83065	-16.65824
H	-25.39581	-20.78358	-16.57077	**H**	-25.38463	-20.75044	-16.61636
H	-26.7825	-21.83761	-16.22412	**H**	-26.76354	-21.81484	-16.265
C	-27.12032	-19.7653	-15.75922	**C**	-27.11893	-19.74324	-15.81447
H	-28.20717	-19.85741	-15.76266	**H**	-28.20427	-19.85045	-15.80794
H	-26.84837	-18.76626	-16.11147	**H**	-26.86106	-18.74464	-16.17789
N	-26.69784	-19.8392	-14.31877	**N**	-26.67705	-19.80357	-14.38039
H	-26.99863	-18.97012	-13.84253	**H**	-27.00116	-18.94556	-13.89968
H	-27.13648	-20.64599	-13.81595	**H**	-27.08036	-20.62723	-13.87909
H	-25.68073	-19.91673	-14.22398	**H**	-25.6536	-19.83709	-14.3109
Cl	-19.69622	-14.8201	-9.87989	**Cl**	-19.49064	-15.02225	-9.96639

REFERENCES

(1) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. Computing kinetic isotope effects for chorismite mutase with high accuracy. A new DFT/MM strategy, *J. Phys. Chem. B* **2005**, *109*, 3707 − 3710. https://doi.org/10.1021/jp044387u.

(2) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. QM/MM calculations of kinetic isotope effects in the chorismite mutase active site. *Org. Biomol. Chem.* **2003**, *1*, 483 − 487. https://doi.org/10.1039/B210508J.

(3) Ruggiero, G. D.; Guy, S. J.; Martí, S.; Moliner, V.; Williams, I. H. Vibrational analysis of the chorismite rearrangement: relaxed force constants, isotope effects and activation entropies calculated for reaction in vacuum, water and the active site of chorismite mutase. *J. Phys. Org. Chem.* **2004**, *17*, 592 − 601. https://doi.org/10.1002/poc.781.

4.2.2 Rationalization of the homo-salinosporamide A (hSalA) reversible inhibition character

Supporting Information

On the origin of the different reversible character of Salinosporamide A and Homo-salinosporamide A in the covalent inhibition of the Human 20S Proteasome

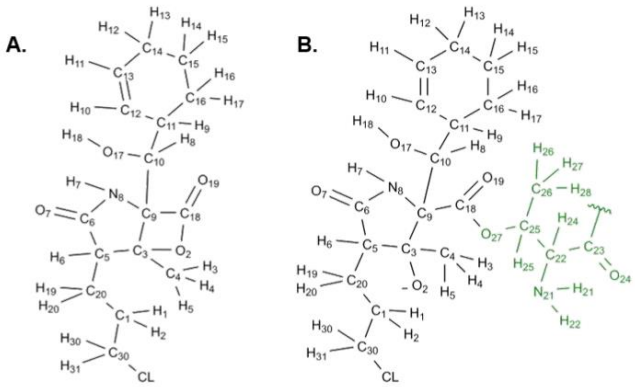
Natalia Serrano-Aparicio, Vicent Moliner*, Katarzyna Świderek*

BioComp group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain.

*V.M. moliner@uji.es; *K.Ś. swiderek@uji.es

Table of contents:

1. MM Parameters for hSalA



Scheme S1. Structure of **A**. hSalA inhibitor and **B**. hSalA in covalent complex with Thr1 from active site of the 20S proteasome.

Table S1. Atom types, charges and parameters for bonding and non-bonding interactions obtained for hSalA inhibitor molecule computed using Antechamber software.

Atom name	Atom type	Charge	Atom name	Atom type	Charge
C1	CT	-0.1034	C13	CM	-0.1492
C20	CT	-0.0564	C14	CT	-0.0532
C5	CT	-0.1397	C15	CT	-0.0734
C6	C	0.7035	C16	CT	-0.0794
O7	O	-0.5955	H16	HC	0.0537
N8	N	-0.5719	H17	HC	0.0537
H7	H	0.3415	H14	HC	0.0457
H6	HC	0.0967	H15	HC	0.0457
H19	HC	0.0587	H12	HC	0.0532
H20	HC	0.0587	H13	HC	0.0532
H1	HC	0.0707	H11	HA	0.126
H2	HC	0.0707	H10	HA	0.136
O2	OS	-0.4079	H9	HC	0.0547
C3	CT	0.1448	O17	OH	-0.5888
C4	CT	-0.0941	H18	HO	0.41
H3	HC	0.065033	H8	H1	0.0977
H4	HC	0.065033	C18	C	0.6936
H5	HC	0.065033	O19	O	-0.5255
C9	CT	-0.0176	C30	CT	0.041
C10	CT	0.1601	H30	H1	0.0642
C11	CT	-0.0595	H31	H1	0.0642
C12	CM	-0.1842	CL	Cl	-0.1914

Parameters:

MASS				**BOND**			
CT	12.010	0.878	same as c3	CT-HC	330.60	1.097	same as c3-hc
HC	1.008	0.135	same as hc	CT-CT	300.90	1.538	same as c3-c3
H1	1.008	0.135	same as hc	CT-H1	330.60	1.097	same as c3-hc
Cl	35.450	1.910	same as cl	CT-Cl	266.30	1.804	same as c3-cl
C	12.010	0.616	same as c	CT-C	313.00	1.524	same as c -c3
O	16.000	0.434	same as o	C -O	637.70	1.218	same as c -o
N	14.010	0.530	same as n	C -N	427.60	1.379	same as c -n
H	1.008	0.161	same as hn	N -H	403.20	1.013	same as hn-n
CM	12.010	0.360	same as c2	N -CT	328.70	1.462	same as c3-n
HA	1.008	0.135	same as hc	CT-O	449.90	1.317	same as c3-o
OH	16.000	0.465	same as oh	CT-OH	316.70	1.423	same as c3-oh
HO	1.008	0.135	same as ho	CT-CM	326.80	1.510	same as c2-c3
OS	16.000	0.465	same as os	CM-CM	419.80	1.428	same as cc-cc
				CM-HA	344.30	1.087	same as c2-hc
				OH-HO	371.40	0.973	same as ho-oh
				C -OS	390.80	1.358	same as c -os

				OS-CT	308.60	1.432	same as c3-osNT
ANGLE				**ANGLE**			
CT-CT-H1	46.340	109.800	same as c3-c3-hc	N -CT-C	67.000	109.060	same as c -c3-n
CT-CT-Cl	57.390	110.410	same as c3-c3-cl	H -N -CT	45.800	117.680	same as c3-n -hn
CT-CT-HC	46.340	109.800	same as c3-c3-hc	CT-CT-OH	67.470	110.190	same as c3-c3-oh
CT-CT-CT	62.860	111.510	same as c3-c3-c3	CT-C -OS	68.890	110.720	same as c3-c -os
HC-CT-HC	39.400	107.580	same as hc-c3-hc	CT-CT-CM	63.410	111.560	same as c2-c3-c3
H1-CT-H1	39.400	107.580	same as hc-c3-hc	CT-OH-HO	47.380	107.260	same as c3-oh-ho
H1-CT-Cl	39.900	107.650	same as cl-c3-hc	CT-CM-CM	64.060	123.630	same as c2-c2-c3
CT-CT-C	63.270	111.040	same as c -c3-c3	CT-CM-HA	45.110	120.000	same as c3-c2-hc
CT-C -O	67.400	123.200	same as c3-c -o	CM-CT-HC	46.990	110.360	same as c2-c3-hc
CT-C -N	66.790	115.180	same as c3-c -n	CM-CM-HA	50.010	119.700	same as c2-c2-hc
CT-CT-O	68.500	113.010	same as c3-c3-o	OH-CT-H1	51.110	109.500	same as hc-c3-oh
C -CT-HC	46.930	108.770	same as c -c3-hc	C -OS-CT	63.280	115.980	same as c -os-c3
C -N -H	48.330	117.550	same as c -n -hn	O -C -OS	75.320	123.250	same as o -c -os
C -N -CT	63.390	120.690	same as c -n -c3	OS-CT-CT	68.000	107.970	same as c3-c3-os
O -C -N	74.220	123.050	same as n -c -o	OS-CT-H1	51.050	108.700	same as hc-c3-os
N -CT-CT	65.910	111.610	same as c3-c3-n	H1-CT-C	46.930	108.770	same as c -c3-hc
DIHEDRALS							
CT-CT-CT-C	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-CT-CT-HC	1		0.160	0.000	3.000		same as hc-c3-c3-c3
CT-CT-CT-CT	1		0.180	0.000	-3.000		same as c3-c3-c3-c3
CT-CT-CT-CT	1		0.250	180.000	-2.000		same as c3-c3-c3-c3
CT-CT-CT-CT	1		0.200	180.000	1.000		same as c3-c3-c3-c3
HC-CT-CT-H1	1		0.150	0.000	3.000		same as hc-c3-c3-hc
HC-CT-CT-Cl	1		0.000	0.000	-3.000		same as hc-c3-c3-cl
HC-CT-CT-Cl	1		0.250	0.000	1.000		same as hc-c3-c3-cl
HC-CT-CT-HC	1		0.150	0.000	3.000		same as hc-c3-c3-hc
H1-CT-CT-CT	1		0.160	0.000	3.000		same as hc-c3-c3-c3
Cl-CT-CT-CT	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-CT-C -O	1		0.000	180.000	2.000		same as X -c -c3-X
CT-CT-C -N	1		0.100	0.000	-4.000		same as c3-c3-c -n
CT-CT-C -N	1		0.070	0.000	2.000		same as c3-c3-c -n
CT-CT-CT-O	1		0.156	0.000	3.000		same as X -c3-c3-X
HC-CT-CT-C	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-C -N -H	1		2.500	180.000	2.000		same as X -c -n -X
CT-C -N -CT	1		0.000	0.000	-2.000		same as c3-c -n -c3
CT-C -N -CT	1		1.500	180.000	1.000		same as c3-c -n -c3
CT-CT-CT-N	1		0.156	0.000	3.000		same as X -c3-c3-X
C -CT-CT-O	1		0.156	0.000	3.000		same as X -c3-c3-X
C -N -CT-CT	1		0.500	180.000	-4.000		same as c3-c3-n -c
C -N -CT-CT	1		0.150	180.000	-3.000		same as c3-c3-n -c
C -N -CT-CT	1		0.000	0.000	-2.000		same as c3-c3-n -c
C -N -CT-CT	1		0.530	0.000	1.000		same as c3-c3-n -c
C -N -CT-C	1		0.850	180.000	-2.000		same as c -n -c3-c
C -N -CT-C	1		0.800	0.000	1.000		same as c -n -c3-c
O -C -CT-HC	1		0.800	0.000	-1.000		same as hc-c3-c -o
O -C -CT-HC	1		0.000	0.000	-2.000		same as hc-c3-c -o
O -C -CT-HC	1		0.080	180.000	3.000		same as hc-c3-c -o
O -C -N -H	1		2.500	180.000	-2.000		same as hn-n -c -o
O -C -N -H	1		2.000	0.000	1.000		same as hn-n -c -o
O -C -N -CT	1		2.500	180.000	2.000		same as X -c -n -X
N -C -CT-HC	1		0.000	180.000	2.000		same as X -c -c3-X
N -CT-CT-O	1		0.156	0.000	3.000		same as X -c3-c3-X
N -CT-CT-OH	1		0.156	0.000	3.000		same as X -c3-c3-X
N -CT-CT-H1	1		0.156	0.000	3.000		same as X -c3-c3-X
N -CT-C -O	1		0.000	180.000	2.000		same as X -c -c3-X
N -CT-C -OS	1		0.000	180.000	2.000		same as X -c -c3-X
H -N -CT-CT	1		0.000	0.000	2.000		same as X -c3-n -X
H -N -CT-C	1		0.000	0.000	2.000		same as X -c3-n -X
HC-CT-CT-O	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-CT-CT-OH	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-CT-C -OS	1		0.000	180.000	2.000		same as X -c -c3-X
CT-CT-CT-CM	1		0.156	0.000	3.000		same as X -c3-c3-X
CT-CT-OH-HO	1		0.160	0.000	-3.000		same as ho-oh-c3-c3
CT-CT-OH-HO	1		0.250	0.000	1.000		same as ho-oh-c3-c3
CT-C -OS-CT	1		2.700	180.000	-2.000		same as c3-c -os-c3
CT-C -OS-CT	1		0.000	0.000	-1.000		same as c3-c -os-c3
CT-C -OS-CT	1		1.150	0.000	3.000		same as c3-c -os-c3
CT-CT-CM-CM	1		0.000	0.000	2.000		same as X -c2-c3-X
CT-CT-CM-HA	1		0.000	0.000	2.000		same as X -c2-c3-X
CT-CM-CM-CT	1		6.650	180.000	-2.000		same as c3-c2-c2-c3
CT-CM-CM-CT	1		1.900	180.000	1.000		same as c3-c2-c2-c3
CT-CM-CM-HA	1		6.650	180.000	2.000		same as X -c2-c2-X
CM-CT-CT-OH	1		0.156	0.000	3.000		same as X -c3-c3-X
CM-CT-CT-H1	1		0.156	0.000	3.000		same as X -c3-c3-X
CM-CT-CT-HC	1		0.156	0.000	3.000		same as X -c3-c3-X
CM-CM-CT-HC	1		0.380	180.000	-3.000		same as hc-c3-c2-c2

CM-CM-CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	1.150	0.000	1.000	same as hc-c3-c2-c2
HC-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
HA-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
HC-CT-CT-OH	1	0.000	0.000	-3.000	same as hc-c3-c3-oh
HC-CT-CT-OH	1	0.250	0.000	1.000	same as hc-c3-c3-oh
OH-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
HO-OH-CT-H1	1	0.167	0.000	3.000	same as X -c3-oh-X
H1-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
C -OS-CT-CT	1	0.383	0.000	-3.000	same as c3-c3-os-c
C -OS-CT-CT	1	0.800	180.000	1.000	same as c3-c3-os-c
C -OS-CT-H1	1	0.383	0.000	3.000	same as X -c3-os-X
O -C -OS-CT	1	2.700	180.000	-2.000	same as o -c -os-c3
O -C -OS-CT	1	1.400	180.000	1.000	same as o -c -os-c3
OS-CT-CT-HC	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-HC	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-H1	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-H1	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
H1-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
H1-CT-C -O	1	0.800	0.000	-1.000	same as hc-c3-c -o
H1-CT-C -O	1	0.000	0.000	-2.000	same as hc-c3-c -o
H1-CT-C -O	1	0.080	180.000	3.000	same as hc-c3-c -o
IMPROPER					
CT-N -C -O	1.1	180.0	2.0	Using default value	
C -CT-N -H	1.1	180.0	2.0	Using default value	
CM-CT-CM-HA	1.1	180.0	2.0	Using default value	
CT-O -C -OS	1.1	180.0	2.0	Using default value	
NONBON					
CT	1.9080	0.1094	same as c3		
HC	1.4870	0.0157	same as hc		
H1	1.4870	0.0157	same as hc		
Cl	1.9480	0.2650	same as cl		
C	1.9080	0.0860	same as c		
O	1.6612	0.2100	same as o		
N	1.8240	0.1700	same as n		
H	0.6000	0.0157	same as hn		
CM	1.9080	0.0860	same as cc		
HA	1.4870	0.0157	same as hc		
OH	1.7210	0.2104	same as oh		
HO	0.0000	0.0000	same as ho		
OS	1.6837	0.1700	same as os		

Table S2. Atom types, charges and parameters for bonding and non-bonding interactions obtained for hSalA inhibitor covalently bound to Thr1 of protein computed using Antechamber software.

Atom name	Atom type	Charge	Atom name	Atom type	Charge
C1	CT	-0.110564	C16	CT	-0.063494
H1	HC	0.080580	H16	HC	0.050624
H2	HC	0.080580	H17	HC	0.050624
C30	CT	0.054918	H14	HC	0.028657
H30	H1	0.055617	H15	HC	0.028657
H31	H1	0.055617	H12	HC	0.029156
Cl	Cl	-0.228740	H13	HC	0.029156
C20	CT	-0.049474	H11	HA	0.102847
H19	HC	0.043635	H10	HA	0.129806
H20	HC	0.043635	H9	HC	0.052621
C5	CT	-0.190984	O17	OH	-0.633743
C6	C	0.683480	H18	HO	0.388420
O7	O	-0.711559	H8	H1	0.076586
N8	N	-0.411512	C18	C	0.671099
H7	H	0.285075	O19	O	-0.721073
H6	HC	0.130505	O27	OS	-0.470600
C3	CT	0.421870	C25	CT	0.126911
O2	O	-0.499243	C26	CT	-0.102252
C4	CT	-0.175361	H26	HC	0.036313
H3	HC	0.081246	H27	HC	0.036313
H4	HC	0.081246	H28	HC	0.036313
H5	HC	0.081246	H25	H1	0.077584
C9	CT	-0.488327	C22	CT	0.036446
C10	CT	0.263706	N21	NT	-0.888121
C11	CT	-0.066599	H21	H	0.363757
C12	CM	-0.131395	H22	H	0.363757

C13	CM	-0.214519	H24	H1	0.075587		
C14	CT	-0.034251	C23	C	0.576040		
C15	CT	-0.075512	O24	O	-0.542908		

Parameters:

MASS				**BOND**			
CT	12.010	0.878	same as c3	CT-HC	330.60	1.097	same as c3-hc
HC	1.008	0.135	same as hc	CT-CT	300.90	1.538	same as c3-c3
H1	1.008	0.135	same as hc	CT-H1	330.60	1.097	same as c3-hc
Cl	35.450	1.910	same as cl	CT-Cl	266.30	1.804	same as c3-cl
C	12.010	0.616	same as c	CT-C	313.00	1.524	same as c -c3
O	16.000	0.434	same as o	C -O	637.70	1.218	same as c -o
N	14.010	0.530	same as n	C -N	427.60	1.379	same as c -n
H	1.008	0.161	same as hn	N -H	403.20	1.013	same as hn-n
CM	12.010	0.360	same as c2	N -CT	328.70	1.462	same as c3-n
HA	1.008	0.135	same as hc	CT-O	449.90	1.317	same as c3-o
OH	16.000	0.465	same as oh	CT-OH	316.70	1.423	same as c3-oh
HO	1.008	0.135	same as ho	CT-CM	326.80	1.510	same as c2-c3
OS	16.000	0.465	same as os	CM-CM	419.80	1.428	same as cc-cc
NT	14.010	0.530	same as n3	CM-HA	344.30	1.087	same as c2-hc
				OH-HO	371.40	0.973	same as ho-oh
				C -OS	390.80	1.358	same as c -os
				OS-CT	308.60	1.432	same as c3-os
				CT-NT	325.90	1.465	same as c3-n3
				NT-H	392.40	1.019	same as hn-n3
ANGLE				**ANGLE**			
CT-CT-H1	46.340	109.800	same as c3-c3-hc	CT-C -OS	68.890	110.720	same as c3-c -os
CT-CT-Cl	57.390	110.410	same as c3-c3-cl	CT-CT-CM	63.410	111.560	same as c2-c3-c3
CT-CT-HC	46.340	109.800	same as c3-c3-hc	CT-OH-HO	47.380	107.260	same as c3-oh-ho
CT-CT-CT	62.860	111.510	same as c3-c3-c3	CT-CM-CM	64.060	123.630	same as c2-c2-c3
HC-CT-HC	39.400	107.580	same as hc-c3-hc	CT-CM-HA	45.110	120.000	same as c3-c2-hc
H1-CT-H1	39.400	107.580	same as hc-c3-hc	CM-CT-HC	46.990	110.360	same as c2-c3-hc
H1-CT-Cl	39.900	107.650	same as cl-c3-hc	CM-CM-HA	50.010	119.700	same as c2-c2-hc
CT-CT-C	63.270	111.040	same as c -c3-c3	OH-CT-H1	51.110	109.500	same as hc-c3-oh
CT-C -O	67.400	123.200	same as c3-c -o	C -OS-CT	63.280	115.980	same as c -os-c3
CT-C -N	66.790	115.180	same as c3-c -n	O -C -OS	75.320	123.250	same as o -c -os
CT-CT-O	68.500	113.010	same as c3-c3-o	OS-CT-CT	68.000	107.970	same as c3-c3-os
C -CT-HC	46.930	108.770	same as c -c3-hc	OS-CT-H1	51.050	108.700	same as hc-c3-os
C -N -H	48.330	117.550	same as c -n -hn	CT-CT-NT	66.020	111.040	same as c3-c3-n3
C -N -CT	63.390	120.690	same as c -n -c3	CT-NT-H	47.420	109.290	same as c3-n3-hn
O -C -N	74.220	123.050	same as n -c -o	NT-CT-H1	49.550	109.800	same as hc-c3-n3
N -CT-CT	65.910	111.610	same as c3-c3-n	NT-CT-C	66.320	111.140	same as c -c3-n3
N -CT-C	67.000	109.060	same as c -c3-n	H -NT-H	41.400	106.400	same as hn-n3-hn
H -N -CT	45.800	117.680	same as c3-n -hn	H1-CT-C	46.930	108.770	same as c -c3-hc
CT-CT-OH	67.470	110.190	same as c3-c3-oh				

DIHEDRALS					
CT-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-HC	1	0.160	0.000	3.000	same as hc-c3-c3-c3
CT-CT-CT-CT	1	0.180	0.000	-3.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.250	180.000	-2.000	same as c3-c3-c3-c3
CT-CT-CT-CT	1	0.200	180.000	1.000	same as c3-c3-c3-c3
HC-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
HC-CT-CT-Cl	1	0.000	0.000	-3.000	same as hc-c3-c3-cl
HC-CT-CT-Cl	1	0.250	0.000	1.000	same as hc-c3-c3-cl
HC-CT-CT-HC	1	0.150	0.000	3.000	same as hc-c3-c3-hc
H1-CT-CT-CT	1	0.160	0.000	3.000	same as hc-c3-c3-c3
Cl-CT-CT-CT	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-C -N	1	0.100	0.000	-4.000	same as c3-c3-c -n
CT-CT-C -N	1	0.070	0.000	2.000	same as c3-c3-c -n
CT-CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
HC-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-C -N -H	1	2.500	180.000	2.000	same as X -c -n -X
CT-C -N -CT	1	0.000	0.000	-2.000	same as c3-c -n -c3
CT-C -N -CT	1	1.500	180.000	1.000	same as c3-c -n -c3
CT-CT-CT-N	1	0.156	0.000	3.000	same as X -c3-c3-X
C -CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
C -N -CT-CT	1	0.500	180.000	-4.000	same as c3-c3-n -c
C -N -CT-CT	1	0.150	180.000	-3.000	same as c3-c3-n -c
C -N -CT-CT	1	0.000	0.000	-2.000	same as c3-c3-n -c
C -N -CT-CT	1	0.530	0.000	1.000	same as c3-c3-n -c
C -N -CT-C	1	0.850	180.000	-2.000	same as c -n -c3-c
C -N -CT-C	1	0.800	0.000	1.000	same as c -n -c3-c
O -C -CT-HC	1	0.800	0.000	-1.000	same as hc-c3-c -o
O -C -CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c -o
O -C -CT-HC	1	0.080	180.000	3.000	same as hc-c3-c -o
O -C -N -H	1	2.500	180.000	-2.000	same as hn-n -c -o
O -C -N -H	1	2.000	0.000	1.000	same as hn-n -c -o

O -C -N -CT	1	2.500	180.000	2.000	same as X -c -n -X
N -C -CT-HC	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
N -CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
N -CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
H -N -CT-CT	1	0.000	0.000	2.000	same as X -c3-n -X
H -N -CT-C	1	0.000	0.000	2.000	same as X -c3-n -X
HC-CT-CT-O	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-C -OS	1	0.000	180.000	2.000	same as X -c -c3-X
CT-CT-CT-CM	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-OH-HO	1	0.160	0.000	-3.000	same as ho-oh-c3-c3
CT-CT-OH-HO	1	0.250	0.000	1.000	same as ho-oh-c3-c3
CT-C -OS-CT	1	2.700	180.000	-2.000	same as c3-c -os-c3
CT-C -OS-CT	1	0.000	0.000	-1.000	same as c3-c -os-c3
CT-C -OS-CT	1	1.150	0.000	3.000	same as c3-c -os-c3
CT-CT-CM-CM	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
CT-CM-CM-CT	1	6.650	180.000	-2.000	same as c3-c2-c2-c3
CT-CM-CM-CT	1	1.900	180.000	1.000	same as c3-c2-c2-c3
CT-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
CM-CT-CT-OH	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-H1	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CT-CT-HC	1	0.156	0.000	3.000	same as X -c3-c3-X
CM-CM-CT-HC	1	0.380	180.000	-3.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	0.000	0.000	-2.000	same as hc-c3-c2-c2
CM-CM-CT-HC	1	1.150	0.000	1.000	same as hc-c3-c2-c2
HC-CT-CM-HA	1	0.000	0.000	2.000	same as X -c2-c3-X
HA-CM-CM-HA	1	6.650	180.000	2.000	same as X -c2-c2-X
HC-CT-CT-OH	1	0.000	0.000	-3.000	same as hc-c3-c3-oh
HC-CT-CT-OH	1	0.250	0.000	1.000	same as hc-c3-c3-oh
OH-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
HO-OH-CT-H1	1	0.167	0.000	3.000	same as X -c3-oh-X
H1-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
C -OS-CT-CT	1	0.383	0.000	-3.000	same as c3-c3-os-c
C -OS-CT-CT	1	0.800	180.000	1.000	same as c3-c3-os-c
C -OS-CT-H1	1	0.383	0.000	3.000	same as X -c3-os-X
O -C -OS-CT	1	2.700	180.000	-2.000	same as o -c -os-c3
O -C -OS-CT	1	1.400	180.000	1.000	same as o -c -os-c3
OS-CT-CT-HC	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-HC	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
OS-CT-CT-H1	1	0.000	0.000	-3.000	same as hc-c3-c3-os
OS-CT-CT-H1	1	0.250	0.000	1.000	same as hc-c3-c3-os
OS-CT-CT-C	1	0.156	0.000	3.000	same as X -c3-c3-X
CT-CT-NT-H	1	0.300	0.000	3.000	same as X -c3-n3-X
CT-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
H1-CT-CT-NT	1	0.156	0.000	3.000	same as X -c3-c3-X
H1-CT-CT-H1	1	0.150	0.000	3.000	same as hc-c3-c3-hc
NT-CT-C -O	1	0.000	180.000	2.000	same as X -c -c3-X
H -NT-CT-H1	1	0.300	0.000	3.000	same as X -c3-n3-X
H -NT-CT-C	1	0.300	0.000	3.000	same as X -c3-n3-X
H1-CT-C -O	1	0.800	0.000	-1.000	same as hc-c3-c -o
H1-CT-C -O	1	0.000	0.000	-2.000	same as hc-c3-c -o
H1-CT-C -O	1	0.080	180.000	3.000	same as hc-c3-c -o

IMPROPER				
CT-N -C -O	1.1	180.0	2.0	Using default value
C -CT-N -H	1.1	180.0	2.0	Using default value
CM-CT-CM-HA	1.1	180.0	2.0	Using default value
CT-O -C -OS	1.1	180.0	2.0	Using default value

NONBON			
CT	1.9080	0.1094	same as c3
HC	1.4870	0.0157	same as hc
H1	1.4870	0.0157	same as hc
Cl	1.9480	0.2650	same as cl
C	1.9080	0.0860	same as c
O	1.6612	0.2100	same as o
N	1.8240	0.1700	same as n
H	0.6000	0.0157	same as hn
CM	1.9080	0.0860	same as cc
HA	1.4870	0.0157	same as hc
OH	1.7210	0.2104	same as oh
HO	0.0000	0.0000	same as ho
OS	1.6837	0.1700	same as os
NT	1.8240	0.1700	same as n3

2. Results of MM MD simulations

2.1. Energy, temperature and RMSD evolution along 5ns MD simulation for the covalent adduct of the active site of β5 subunit with hSalA (E-I2)

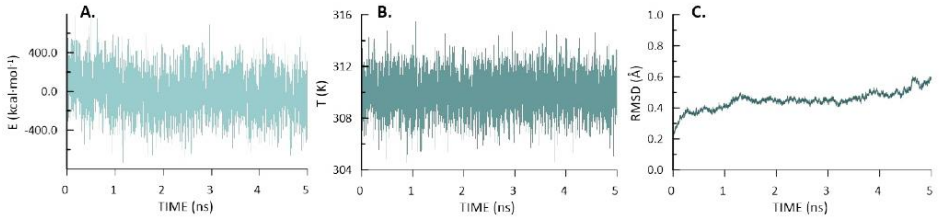


Figure S1. Time-dependent evolution of **A**. the total energy, **B**. temperature and **C**. RMSD of the N, C^{α} and C atoms of the protein backbone along the 5 ns NVT-MD simulations of 20S proteasome in the covalently bond form with hSalA in the active site of β5 subunit.

2.2. Energy, temperature and RMSD evolution along 50ns MD simulation for the Michaelis Complex (E•hSalA)

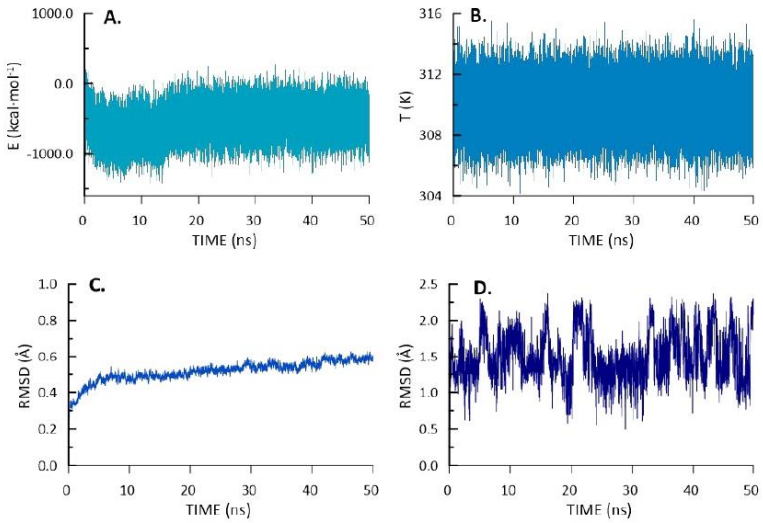


Figure S2. Time-dependent evolution of **A**. the total energy, **B**. temperature, **C**. RMSD of the N, C^{α} and C atoms of the protein backbone and **D**. RMSD of the atoms of hSalA along the 50 ns NVT-MD simulations of 20S proteasome in the Michaelis complex form with hSalA bound in the active site of β5 subunit.

2.3. Population analysis of key distances in β5 active site during MD simulations

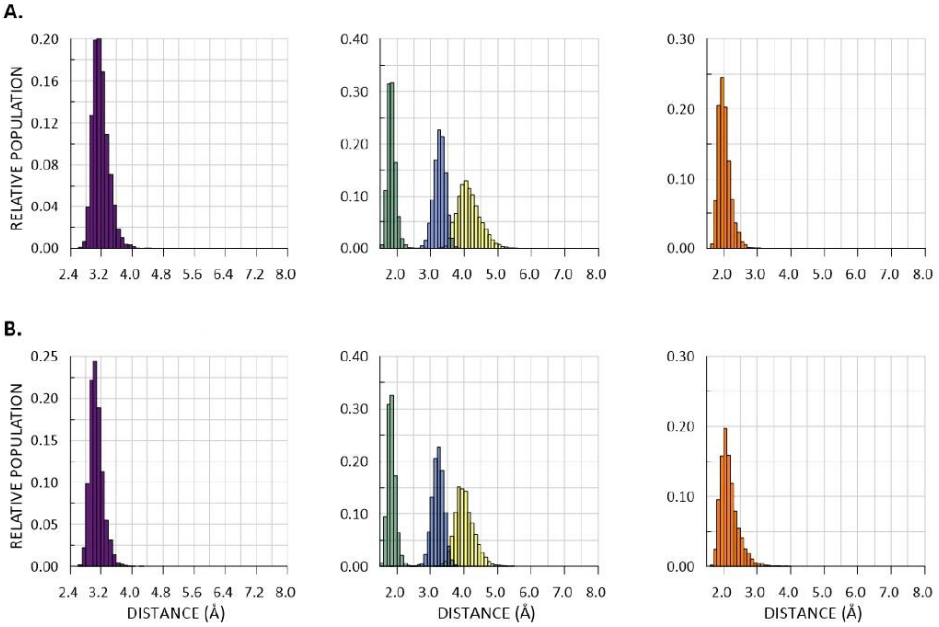


Figure S3. Analysis of the population of different **A.** E•hSalA and **B.** E•SalA complexes based on 5000 structures produced during 50ns of NVT-MD simulation for the five chosen distances: $O^{\gamma Thr1}$-$C1^{SalA}$ (violet), $H^{\gamma Thr1}$-$N^{\zeta Lys33}$ (green), $H^{\gamma Thr1}$-$O2^{SalA}$ (yellow), $H^{\gamma Thr1}$-N^{Thr1} (blue) and H^{Gly47} – $O1^{SalA}$ (orange) distance for Michaelis complex of β5 subunit of 20S proteasome.

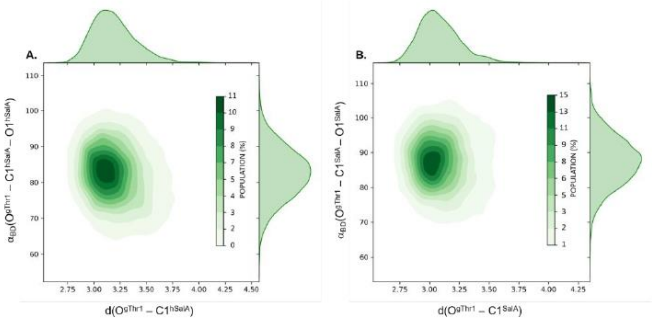


Figure S4. Analysis of the population of the Bürgi-Dunitz angle and the $O^{\gamma Thr1}$-$C1^{Inh}$ distance for **A.** E•hSalA and **B.** E•SalA complexes based on 5000 structures produced during 50ns of NVT-MD simulation.

2.4. Interactions energies between hSalA and SalA and active site of β5

In this work, an additive hybrid QM/MM scheme was employed for the construction of the total Hamiltonian where the total energy is obtained from the sum of each contribution to the energy.

$$E_{QM/MM} = E_{QM} + E_{QM/MM}^{elect} + E_{QM/MM}^{vdW} + E_{MM} \quad \text{(S1)}$$

Here, E_{QM} describes the atoms in the QM region, $E_{QM/MM}$ defines the interaction between the QM and MM region and E_{MM} describes the rest of the MM region. Contribution of QM-MM interaction energies between SalA (treated quantum mechanically) with catalytic site, oxyanion hole, S1-pocket and S2-pocket of the protein (included in the MM region) was computed using the following expression:

$$E_{QM/MM}^{Int} = \sum\left\langle \Psi \left| \frac{q_{MM}}{r_{e,MM}} \right| \Psi \right\rangle + \sum\sum \frac{Z_Q q_{MM}}{r_{QM,MM}} + E_{QM/MM}^{vdW} \quad \text{(S2)}$$

This interaction energy can be exactly decomposed in a sum over residues (or group of residues) provided that the polarized wave function (Ψ) is employed to evaluate this energy contribution. The global polarization effect can be obtained from the gas phase energy difference between the polarized, Ψ, and non-polarized, Ψ_0, wave functions.

Table S3. Non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between **hSalA** and the key residues of β5 active site, averaged over 10000 structures derived from MD simulations at AM1/MM level.

Major Role	Residue	TOTAL	Electrostatic	Lennard Jones
Oxyanion Hole	Gly47	-4.8 ± 1.2	-2.6 ± 1.0	-2.2 ± 0.8
Catalytic Residues	Thr1	-0.8 ± 1.4	2.5 ± 1.2	-3.3 ± 0.7
	Lyn33	-5.9 ± 1.0	-1.6 ± 0.7	-4.2 ± 0.5
S1-Pocket	Ala20	-2.6 ± 0.8	-0.7 ± 0.6	-1.9 ± 0.5
	Val31	-1.5 ± 0.4	-0.1 ± 0.2	-1.4 ± 0.3
	Met45	-2.7 ± 0.4	-0.3 ± 0.2	-2.4 ± 0.4
	Ala46	-3.9 ± 0.8	-1.9 ± 0.5	-2.0 ± 0.4
	Gly48	-1.2 ± 0.3	-0.1 ± 0.3	-1.1 ± 0.3
	Ala49	-2.4 ± 0.7	-0.3 ± 0.3	-2.1 ± 0.6
S2-Pocket	Thr21	-2.2 ± 0.9	0.2 ± 1.2	-2.4 ± 0.5
	Tyr169	-6.6 ± 0.9	-2.3 ± 0.5	-4.3 ± 0.8

Table S4. Non-bonding (electrostatic and Lennard-Jones) interactions in (kcal·mol^{-1}) between **SalA** and the key residues of β5 active site, averaged over 10000 structures derived from MD simulations at AM1/MM level.

Major Role	Residue	TOTAL	Electrostatic	Lennard Jones
Oxyanion Hole	Gly47	-3.9 ± 1.2	-1.3 ± 1.1	-2.6 ± 0.6
Catalytic Residues	Thr1	-1.2 ± 1.2	2.2 ± 1.0	-3.4 ± 0.6
	Lyn33	-4.8 ± 0.9	-0.9 ± 0.5	-3.9 ± 0.6
S1-Pocket	Ala20	-1.3 ± 0.6	0.5 ± 0.4	-1.8 ± 0.5
	Val31	-1.0 ± 0.3	0.2 ± 0.1	-1.2 ± 0.3
	Met45	-2.6 ± 0.4	-0.3 ± 0.2	-2.3 ± 0.4
	Ala46	-3.1 ± 0.7	-1.3 ± 0.4	-1.8 ± 0.4
	Gly48	-1.3 ± 0.3	0.4 ± 0.3	-1.7 ± 0.4
	Ala49	-1.8 ± 0.7	0.3 ± 0.2	-2.1 ± 0.6
S2-Pocket	Thr21	-3.2 ± 0.7	-0.5 ± 0.7	-2.7 ± 0.5
	Tyr169	-5.9 ± 0.8	-1.6 ± 0.4	-4.3 ± 0.7

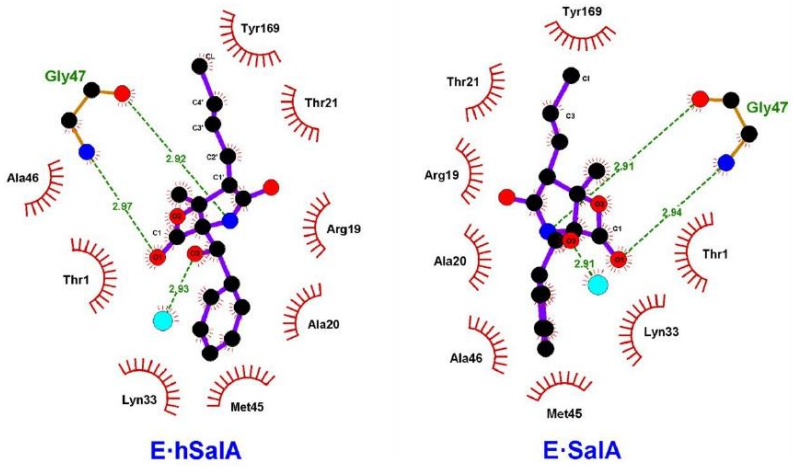


Figure S5. Representation of β5 active site of 20S Proteasome representation in the Michaelis complex form with hSalA and SalA inhibitors.

3. Graphical definition of Bürgi–Dunitz (α_{BD}) and Flippin–Lodge (α_{FL}) angles

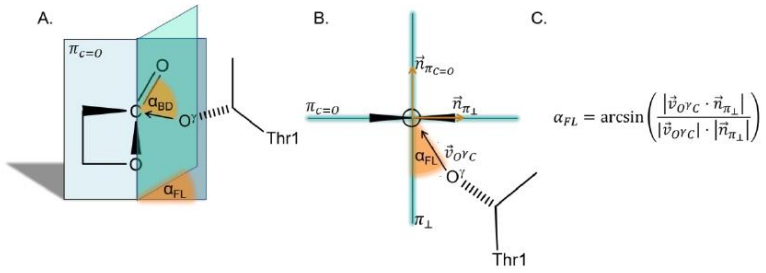


Figure S6. Representation of measured Bürgi–Dunitz (α_{BD}) and Flippin–Lodge (α_{FL}) angles. A. 3D representation showing the position of both angles. B. Elevated plane representation showing the key vectors necessary for obtaining the Flippin–Lodge (α_{FL}) angle. C. Equation used for obtaining αFL once the key vectors are known.

4. Results of reversible inhibition mechanism studies

4.1. Free energy surfaces computed at AM1/MM level

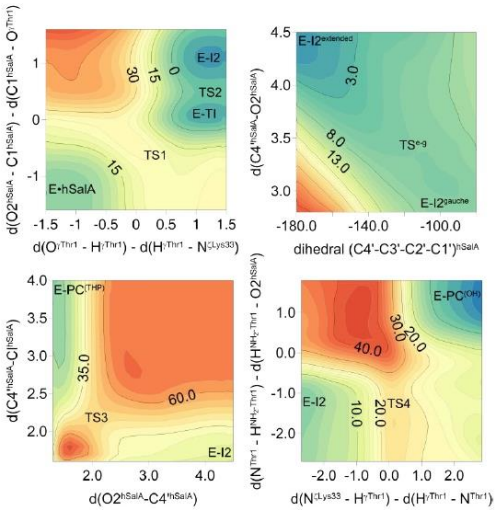


Figure S7. 2D-PMF surfaces computed at AM1/MM level of theory for the inhibition mechanism of 20S proteasome by hSalA.

4.2. Free energy surfaces computed at M06-2X:AM1/MM level

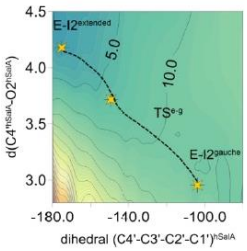


Figure S8. 2D-PMF surface computed at M06-2X:AM1/MM level of theory for the inhibition mechanism of 20S proteasome by hSalA.

4.3. Evolution of key distances along reaction pathway.

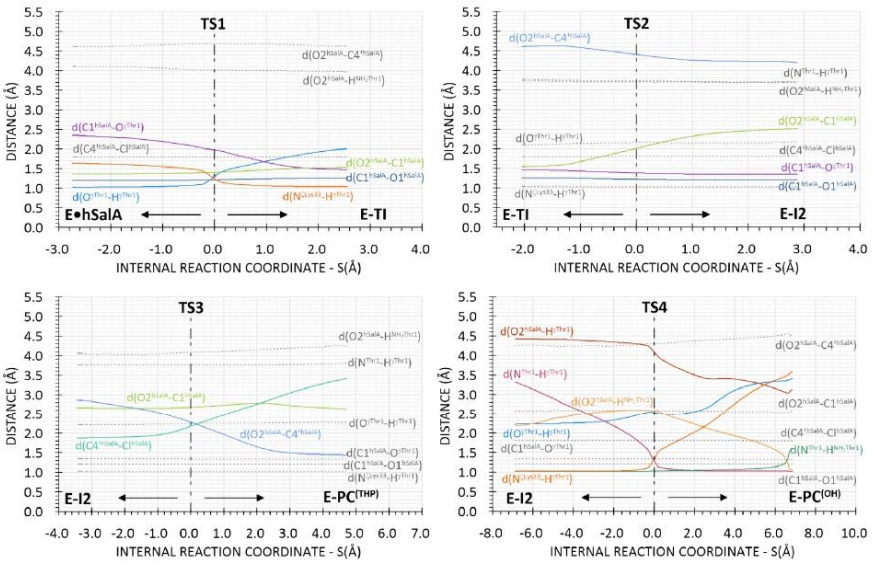


Figure S9. Evolution of key distances on minimum potential energy pathway along reaction progress registered from IRC calculations at M06-2X/MM level of theory, for the inhibition mechanism of 20S proteasome by hSalA.

Table S5. Key distances for located at M06-2X/MM stationary structures for inhibition mechanism with **hSalA**.

	E•hSalA	TS1	E-TI	TS2	E-I2^{ext}	TS$^{e\text{-}g}$	E-I2^{gauche}	TS3	E-PC$^{(THP)}$	TS4	E-PC$^{(OH)}$
N$^{\zeta Lys33}$ – H^{gThr1}	1.64	1.24	1.04	1.03	1.03	1.03	1.03	1.03	1.03	1.33	3.83
H^{gThr1} – O^{gThr1}	1.02	1.30	2.00	2.16	2.16	2.18	2.22	2.25	2.27	2.52	3.40
O^{gThr1} – C1^{hSalA}	2.35	1.98	1.47	1.38	1.35	1.35	1.35	1.34	1.34	1.34	1.31
C1^{hSalA} – O1^{hSalA}	1.20	1.21	1.26	1.23	1.21	1.21	1.21	1.21	1.21	1.21	1.21
O1^{hSalA} – H^{Gly47}	1.83	1.87	1.85	1.88	1.91	1.94	1.95	1.97	1.98	1.91	1.85
C1^{hSalA} – O2^{hSalA}	1.36	1.40	1.53	2.01	2.52	2.57	2.66	2.68	2.63	2.58	2.57
O2^{hSalA} – C4'^{hSalA}	4.60	4.69	4.62	4.41	4.19	3.72	2.85	2.28	1.44	4.30	4.37
C4'^{hSalA} – Cl^{hSalA}	1.80	1.80	1.80	1.81	1.81	1.81	1.87	2.18	3.42	1.81	1.80
O2^{hSalA} – N^{Thr1}	3.48	3.37	3.33	3.55	3.65	3.73	3.81	3.79	3.83	3.56	2.80
H^{gThr1} – O2^{hSalA}	3.70	3.72	4.09	4.42	4.71	4.73	4.91	4.92	4.99	4.09	3.21
H^{gThr1} – N^{Thr1}	3.38	3.53	3.77	3.71	3.69	3.72	3.76	3.76	3.81	1.29	1.02
N$^{\zeta Lys33}$ – N^{Thr1}	4.36	4.24	4.34	4.30	4.28	4.31	4.34	4.35	4.40	2.60	4.06
H1^{Thr1} – N^{Thr1}	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.79
H1^{Thr1} – O2^{hSalA}	4.10	4.02	3.97	3.74	3.72	3.92	4.06	4.08	4.24	2.57	1.01
(C4'-C3'-C2'-C1')^{hSalA}	171.45	170.18	172.38	176.60	-175.19	-149.56	-100.49	-75.46	-51.68	-177.03	172.65
(Cl-C4'-C3'-C2')^{hSalA}	178.48	176.66	-179.37	175.91	172.40	-174.78	-168.67	-152.23	-164.80	177.45	173.39

4.4. ChelpG charges on key atoms

Table S6. ChelpG charges on key atoms for located M06-2X/MM stationary structures for inhibition mechanism with **hSalA**.

	E•hSalA	TS1	E-TI	TS2	E-I2^{ext}	TS$^{e\text{-}g}$	E-I2^{gauche}	TS3	E-PC$^{(THP)}$	TS4	E-PC$^{(OH)}$
C3$^{\gamma hSalA}$	0.181	0.203	0.241	0.206	0.053	0.112	-0.391	-0.370	-0.228	0.184	0.179
O2^{hSalA}	-0.788	-0.883	-0.950	-1.006	-1.089	-1.033	-0.988	-1.137	-0.925	-1.066	-0.937
C1^{hSalA}	1.126	1.337	1.445	1.336	0.951	0.921	0.927	0.993	1.107	1.054	1.150
O1^{hSalA}	-0.666	-0.776	-0.952	-0.806	-0.665	-0.654	-0.645	-0.654	-0.654	-0.671	-0.707
O^{gThr1}	-0.577	-0.644	-0.540	-0.550	-0.325	-0.364	-0.375	-0.424	-0.447	-0.522	-0.678
N^{Thr1}	-1.057	-1.034	-0.946	-0.884	-0.851	-0.901	-0.943	-0.963	-1.098	-0.612	-0.737
C4$^{\gamma hSalA}$	-0.084	-0.116	-0.057	-0.023	-0.025	0.049	0.391	0.443	0.502	0.024	0.008
N^{zLys33}	-0.387	-0.256	-0.222	-0.202	-0.129	-0.155	-0.086	-0.117	-0.167	-0.523	-0.650
Cl^{hSalA}	-0.254	-0.252	-0.274	-0.295	-0.305	-0.328	-0.495	-0.732	-1.060	-0.306	-0.294

Table S7. ChelpG charges on key atoms with hydrogens summed into heavy atoms, for located M06-2X/MM stationary structures for inhibition mechanism with **hSalA**.

	E•hSalA	TS1	E-TI	TS2	E-I2^{ext}	TS$^{e\text{-}g}$	E-I2^{gauche}	TS3	E-PC$^{(THP)}$	TS4	E-PC$^{(OH)}$
C3$^{\gamma hSalA}$	0.204	0.228	0.237	0.219	0.163	0.070	-0.256	-0.206	0.007	0.164	0.174
O2^{hSalA}	-0.788	-0.883	-0.950	-1.006	-1.089	-1.033	-0.988	-1.137	-0.925	-1.066	-0.505
C1^{hSalA}	1.126	1.337	1.445	1.336	0.951	-0.654	0.927	0.993	1.107	1.054	1.150
O1^{hSalA}	-0.666	-0.776	-0.952	-0.806	-0.665	-0.364	-0.645	-0.654	-0.654	-0.671	-0.707
O^{gThr1}	-0.457	-0.644	-0.540	-0.550	-0.325	0.297	-0.375	-0.424	-0.447	-0.521	-0.678
N^{Thr1}	-0.262	-0.256	-0.204	-0.166	-0.143	-0.173	-0.196	-0.204	-0.260	0.351	-0.075
C4$^{\gamma hSalA}$	0.106	0.082	0.095	0.115	0.131	0.207	0.534	0.668	0.553	0.156	0.158
N^{zLys33}	0.217	0.534	0.654	0.721	0.670	0.685	0.694	0.684	0.677	0.148	0.007
Cl^{hSalA}	-0.254	-0.252	-0.274	-0.295	-0.305	-0.328	-0.495	-0.732	-1.060	-0.306	-0.294

5. Cyclization step studies for SalA and hSala in aqueous solution

Model preparation: The model for the study of the final cyclization step in the aqueous phase was prepared by extracting from the optimized E-I2 structure the minimum set of atoms necessary, consisting of the inhibitor atoms (SalA and hSalA, respectively), the $O^{\gamma Thr1}$ covalently bond to C1 and the $C^{\beta Thr1}$ covalently bond to O^{γ}Thr1 and saturated with three hydrogen atoms. Parameters for this small model were obtained with Antechamber. The model was solvated and 1 Na^+ atom was added to neutralize the system by using the tLeap module of AmberTools. As done previously, classical MD simulations were done to equilibrate the system with NAMD software. The system was minimized and heated up to 310 K. Non-biased 100ps of NPT equilibration were done and finally 10 ns of NVT production were performed. Once the system was equilibrated QM/MM exploration of the final step was done. The QM part consisted of the full I2 model (46 atoms).

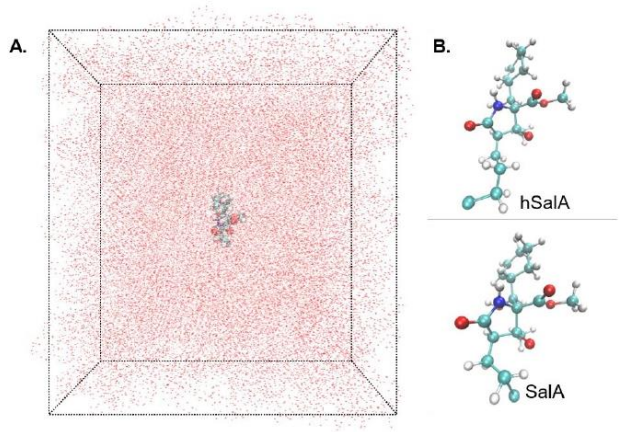


Figure S10. Representation of the model built in aqueous solution for the exploration of the intramolecular cyclization step in hSalA and SalA. **A.** Representation of the small model solvated in a box of explicit water molecules. **B.** representation of the QM subset of atoms.

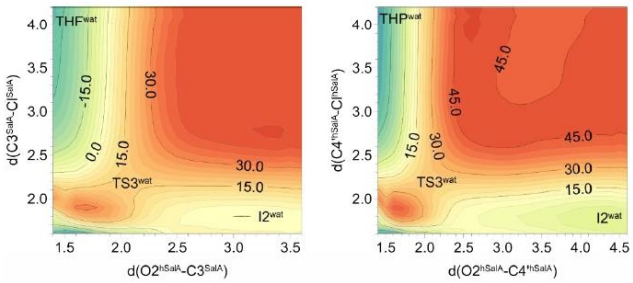


Figure S11. 2D-PMF for intramolecular cyclization step for SalA and hSalA in the aqueous solution computed at AM1/MM level of theory.

6. New double proton transfer step exploration on SalA

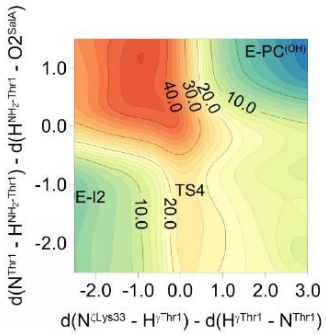


Figure S12. 2D-PMF surface computed at AM1/MM level of theory for the additional double proton transfer in 20S proteasome inhibition by SalA.

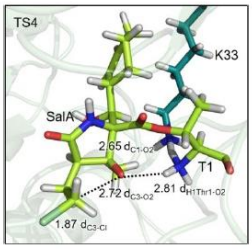


Figure S13. Transition state structure (TS4) localized for the double proton transfer step explored in the inhibition mechanism of 20S proteasome with SalA, computed at M06-2X/MM level of theory.

Table S8. Key distances for structures optimized at M06-2X/MM for additional double proton transfer step with **SalA**.

	E-I2	TS4	E-PC(OH)
$N^{\zeta Lys33}$ – $H^{\gamma Thr1}$	1.03	1.38	5.02
$H^{\gamma Thr1}$ – $O^{\gamma Thr1}$	2.43	2.56	3.58
$O^{\gamma Thr1}$ – $C1^{SalA}$	1.34	1.34	1.31
$C1^{SalA}$ – $O1^{SalA}$	1.21	1.21	1.22
$O1^{SalA}$ – H^{Gly47}	1.91	1.90	1.86
$C1^{SalA}$ – $O2^{SalA}$	2.66	2.65	2.62
$O2^{SalA}$ – $C3^{SalA}$	2.64	2.72	2.88
$C3^{SalA}$ – Cl^{SalA}	1.89	1.87	1.84
$O2^{SalA}$ – N^{Thr1}	3.79	3.79	2.75
$H^{\gamma Thr1}$ – $O2^{SalA}$	4.60	4.25	3.26
$H^{\gamma Thr1}$ – N^{Thr1}	2.19	1.25	1.02
$N^{\zeta Lys33}$ – N^{Thr1}	2.98	2.60	4.07
$H1^{Thr1}$ – N^{Thr1}	1.02	1.02	1.74
$H1^{Thr1}$ – $O2^{SalA}$	2.98	2.81	1.02

7. Results of ester bond hydrolysis mechanism studies in hSalA

7.1. Representation of the water access to the Lys33 in the active site

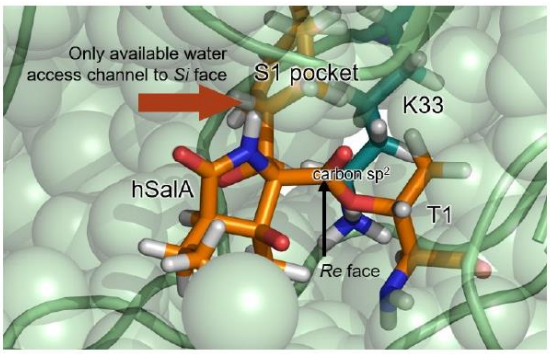


Figure S14. Representation of the active site illustrating Lys33 location in the active site. The only available water access channel to the *Si* face is indicated by red arrow.

Table S9. Analysis of hydrogen bond contacts of the $N^{\zeta Lys33}$ in β5 active site in E•hSalA complex, computed on 5000 structures produced along 50ns of NVT-MD simulation.

Acceptor	DonorH	Donor	Frames	Fraction (%)	AvgDist	AvgAng
$N^{\zeta Lys33}$	$H^{\gamma Thr1}$	$O^{\gamma Thr1}$	4993	99.86	2.7815	166.2605
$O^{\delta 1 Asp17}$	$H^{\zeta 2 Lys33}$	$N^{\zeta Lys33}$	4709	94.18	3.2777	157.6741
O^{Arg19}	$H^{\zeta 3 Lys33}$	$N^{\zeta Lys33}$	3133	62.66	3.0098	146.7563
$O^{\delta 2 Asp17}$	$H^{\zeta 2 Lys33}$	$N^{\zeta Lys33}$	2933	58.66	3.0496	149.5141
$N^{\zeta Lys33}$	$H3^{hSalA}$	$O3^{hSalA}$	138	2.76	3.8861	142.4833
N^{Arg19}	$H^{\zeta 3 Lys33}$	$N^{\zeta Lys33}$	131	2.62	3.5817	139.5899
O^{Val31}	H^{Lys33}	N^{Lys33}	90	1.8	3.1056	139.7599
$N^{\zeta Lys33}$	H^{Arg19}	N^{Arg19}	48	0.96	3.5895	138.4898
O^{Asp17}	H^{Lys33}	N^{Lys33}	42	0.84	3.8172	147.4168
$O^{\delta 2 Asp17}$	$H^{\zeta 3 Lys33}$	$N^{\zeta Lys33}$	11	0.22	2.964	145.1895
N^{Ser18}	$H^{\zeta 2 Lys33}$	$N^{\zeta Lys33}$	8	0.16	3.9216	139.1877
$N^{\zeta Lys33}$	$H2^{Thr1}$	N^{Thr1}	5	0.1	3.7271	139.5889
$O3^{hSalA}$	$H^{\zeta 3 Lys33}$	$N^{\zeta Lys33}$	1	0.02	3.8595	137.1862

7.2. Free energy surfaces computed at AM1/MM level

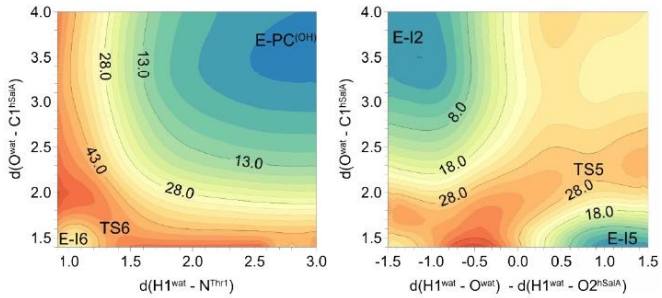


Figure S15. 2D-PMF computed at AM1/MM level of theory for the first step of ester bond hydrolysis in E-PC$^{(OH)}$ and E-I2 complex during hSalA inhibition of 20S proteasome.

7.3. Evolution of key distances along the hydrolysis pathway

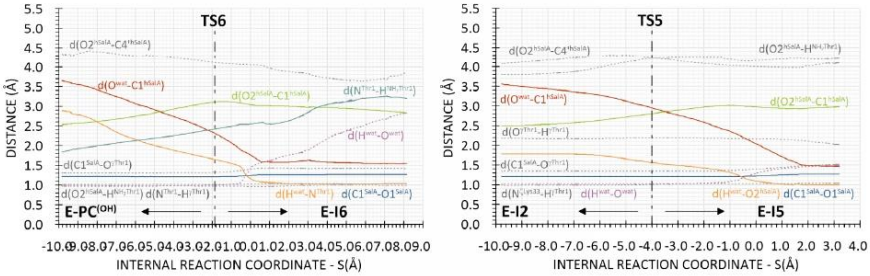


Figure S16. Evolution of key distances on the minimum potential energy pathway along the reaction progress registered from IRC calculations at M06-2X/MM level of theory, for the ester bond hydrolysis of hSalA in the final product complex E-PC$^{(OH)}$ (on the left) and intermediate E-I2 (on the right).

Table S10. Key distances for optimized structures at M06-2X/MM for the ester bond hydrolysis of hSalA in the final product complex and intermediate, E-PC$^{(OH)}$ and E-I2, respectively.

	E-PC$^{(OH)}$	TS6^{hyd}	E-I6	E-I2^{ext}	TS5^{hyd}	E-I5
$N^{\zeta Lys33}$ – H^{gThr1}	3.78	3.84	3.87	1.03	1.03	1.04
H^{gThr1} – O^{gThr1}	3.41	3.18	3.34	2.16	2.16	2.02
O^{gThr1} – $C1^{hSalA}$	1.31	1.40	1.42	1.35	1.39	1.51
$C1^{hSalA}$ – $O1^{hSalA}$	1.22	1.24	1.27	1.21	1.22	1.27
$O1^{hSalA}$ – H^{Gly47}	1.91	1.83	1.78	1.94	1.86	1.82
$C1^{hSalA}$ – $O2^{hSalA}$	2.55	3.04	2.82	2.51	2.98	2.98
$O2^{hSalA}$ – $C4'^{hSalA}$	4.38	4.05	4.10	4.14	4.01	4.08
$C4'^{hSalA}$ – Cl^{hSalA}	1.82	1.82	1.82	1.83	1.82	1.81
$O2^{hSalA}$ – N^{Thr1}	2.82	3.40	3.07	3.12	3.66	3.65
H^{gThr1} – $O2^{hSalA}$	3.21	3.28	3.38	4.47	4.91	4.96
H^{gThr1} – N^{Thr1}	1.03	1.03	1.03	3.83	3.78	3.74
$N^{\zeta Lys33}$ – N^{Thr1}	4.07	4.18	4.09	4.50	4.41	4.39
$H1^{Thr1}$ – N^{Thr1}	1.81	2.55	3.23	1.01	1.02	1.02
$H1^{Thr1}$ – $O2^{hSalA}$	1.01	0.96	0.98	3.79	4.16	4.21
O^{wat} – $C1^{hSalA}$	3.80	1.76	1.54	3.52	2.04	1.46
O^{wat} – $H1^{wat}$	0.97	1.39	2.78	0.97	0.97	1.01
O^{wat} – $H2^{wat}$	0.97	0.96	0.96	0.99	1.28	1.48
N^{Thr1} – $H1^{wat}$	2.88	1.16	1.04	2.32	2.03	1.67
$O2^{hSalA}$ – $H2^{wat}$	2.07	1.94	3.32	1.77	1.09	1.00
(C4'-C3'-C2'-C1')^{hSalA}	175.78	178.80	-174.45	-176.60	-177.70	176.17
(Cl-C4'-C3'-C2')^{hSalA}	-154.41	-165.88	-155.64	-77.54	-158.10	-165.32

8. Results of ester bond hydrolysis mechanism studies in SalA

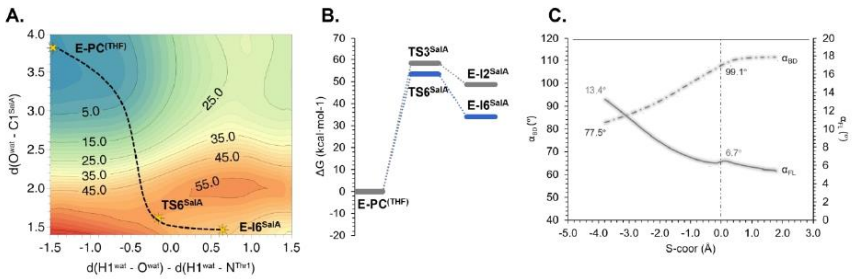


Figure S17. **A.** Free energy surfaces computed for the first step of hydrolysis of the final inhibition product of 20S proteasome with SalA (E-PC$^{(THF)}$). The position of localized stationary structures at M06-2X/MM level are marked as yellow stars. Values of the axes are given in Å and isoenergic lines in kcal·mol^{-1}. **B.** Free energy profiles for the first step of hydrolysis initiated in E-PC$^{(THF)}$ including competitive processes. **C.** Evolution of Bürgi–Dunitz (α_{BD}) and Flippin–Lodge (α_{FL}) angles for nucleophilic attack of water onto carbonyl carbon of the lactone ring of SalA along the first step of the ester bond hydrolysis.

9. Summary of free energy profiles computed at M06-2X:AM1/MM level

Table S11. Relative free energies (ΔG) computed for hSalA and SalA inhibition mechanism at M06-2X:AM1/AMBER level. Values are given in kcal·mol^{-1}.Free energy results for mechanism of **intramolecular cyclization**and **O2 neutralization** by double proton transfer.

SalA				hSalA			
Intramolecular cyclization		O2 neutralization		Intramolecular cyclization		O2 neutralization	
E•SalA	0	**E•SalA**	0	**E•hSalA**	0	**E•hSalA**	0
TS1	20.4	**TS1**	20.4	**TS1**	20.3	**TS1**	20.3
E-TI	6.6	**E-TI**	6.6	**E-TI**	7.9	**E-TI**	7.9
TS2	13.8	**TS2**	13.8	**TS2**	15.3	**TS2**	15.3
E-I2	9.1	**E-I2**	9.1	**E-I2^{extended}**	10.0	**E-I2^{extended}**	10.0
TS3	18.8	**TS4**	14.5	**TS^{e-g}**	20.2	**TS4**	16.5
E-PC$^{(THF)}$	-39.5	**E-PC$^{(OH)}$**	-7.2	**E-I2^{gauche}**	19.7	**E-PC$^{(OH)}$**	-8.5
				TS3	38		
				E-PC$^{(THP)}$	-10.3		

Table S12. Relative free energies (ΔG) computed for hSalA ester-linkage hydrolysis mechanism at M06-2X:AM1/AMBER level. Values are given in kcal·mol^{-1}.

hSalA			
E-PC$^{(OH)}$ hydrolysis		**E-I2 hydrolysis**	
E-PC$^{(OH)}$	0	**E-I2**	0
TS6	67.7	**TS5**	33.6
E-I6	59.1	**E-I5**	28.6

Table S13. Relative free energies (ΔG) computed for SalA final inhibition product ester-linkage hydrolysis mechanism at M06-2X:AM1/AMBER level. Values are given in kcal·mol^{-1}.

SalA	
E-PC$^{(THF)}$ hydrolysis	
E-PC$^{(THF)}$	0
TS6	53.4
E-I6	34.0

10. Cartesian coordinates of QM sub-set of atoms of the TS structures optimized at M06-2X/MM level

Table S14. Cartesian coordinates (in Å) of QM atoms for Transition State structures localized at M06-2X/6-31+G(d,p)/MM level for inhibition mechanism of 20S proteasome with hSalA.

	TS1				TS2		
ATOM	X	Y	Z	ATOM	X	Y	Z
C	-20.74093	-16.02046	-10.61327	C	-20.75534	-15.75497	-10.72016
C	-21.77256	-15.21961	-11.40969	C	-21.85583	-15.05172	-11.51643
C	-22.20854	-15.77078	-12.76083	C	-22.26551	-15.71753	-12.8246
C	-21.15426	-15.76809	-13.84759	C	-21.17684	-15.82178	-13.8701
O	-20.16589	-15.02215	-13.92207	O	-20.21261	-15.04955	-14.03179
N	-21.49301	-16.66799	-14.77402	N	-21.44118	-16.86381	-14.66622
H	-20.83899	-16.8761	-15.53584	H	-20.78522	-17.10321	-15.41165
H	-22.98039	-15.08172	-13.13998	H	-23.04198	-15.08293	-13.27973
H	-22.66424	-15.06366	-10.79316	H	-22.75068	-14.95486	-10.88907
H	-21.36713	-14.21731	-11.6088	H	-21.53697	-14.02471	-11.7537
H	-21.16167	-16.96264	-10.25123	H	-21.08493	-16.75677	-10.43414
H	-19.88372	-16.27268	-11.2476	H	-19.86144	-15.86904	-11.34412
O	-21.97829	-18.2558	-12.52987	O	-21.9481	-18.05422	-12.19138
C	-22.85302	-17.14983	-12.91011	C	-22.81857	-17.16458	-12.79028
C	-24.22706	-17.26303	-12.29243	C	-24.24285	-17.22787	-12.24134
H	-24.46674	-18.31003	-12.12323	H	-24.47725	-18.25742	-11.96521
H	-24.24987	-16.73913	-11.33134	H	-24.29899	-16.62927	-11.32436
H	-24.96707	-16.81808	-12.95916	H	-24.97566	-16.83401	-12.95155
C	-22.61952	-17.53228	-14.40291	C	-22.62091	-17.63433	-14.27791
C	-23.76136	-17.42312	-15.39886	C	-23.73961	-17.44599	-15.3202
C	-23.30555	-17.65014	-16.84034	C	-23.26289	-17.61556	-16.77073
C	-24.46101	-17.51562	-17.80292	C	-24.40999	-17.44616	-17.73872
C	-24.56299	-18.23061	-18.92688	C	-24.48707	-18.08947	-18.90858
C	-23.54569	-19.2565	-19.35501	C	-23.44203	-19.05945	-19.39804
C	-22.2559	-19.14668	-18.54138	C	-22.17433	-19.01424	-18.54223
C	-22.5604	-18.96851	-17.05482	C	-22.50948	-18.91619	-17.0543
H	-21.6357	-18.97906	-16.47175	H	-21.59464	-18.95892	-16.45961
H	-23.16565	-19.80575	-16.67936	H	-23.12219	-19.77828	-16.74919
H	-21.6314	-20.02791	-18.69908	H	-21.55704	-19.89618	-18.73411
H	-21.67792	-18.27985	-18.88783	H	-21.57225	-18.13893	-18.81738
H	-23.9732	-20.26292	-19.21754	H	-23.86264	-20.07832	-19.36862
H	-23.3427	-19.13633	-20.42502	H	-23.21308	-18.84459	-20.44875
H	-25.43031	-18.09056	-19.57033	H	-25.34852	-17.92349	-19.55443
H	-25.24915	-16.81547	-17.52645	H	-25.20981	-16.77279	-17.43025
H	-22.59319	-16.83458	-17.04175	H	-22.55402	-16.78637	-16.92466
O	-24.36246	-16.1443	-15.25322	O	-24.36056	-16.17838	-15.17129
H	-23.794	-15.43946	-15.63346	H	-23.81352	-15.46244	-15.55061
H	-24.52924	-18.14375	-15.11773	H	-24.52872	-18.17537	-15.1027
C	-22.00598	-18.8179	-13.81499	C	-22.2436	-19.08471	-13.89282
O	-21.16481	-19.60551	-14.19658	O	-21.21633	-19.67151	-14.21327
O	-23.70231	-19.82881	-13.72855	O	-23.39238	-19.77754	-13.56464
C	-23.40317	-21.08843	-13.15926	C	-23.2101	-21.1217	-13.06547
C	-23.49037	-22.16935	-14.23613	C	-23.31617	-22.08479	-14.2403
H	-23.12464	-23.13928	-13.88582	H	-23.03407	-23.1012	-13.95302
H	-22.88343	-21.84935	-15.08765	H	-22.66647	-21.74534	-15.04853
H	-24.51914	-22.32648	-14.57835	H	-24.34707	-22.13843	-14.61172
H	-22.37098	-21.05992	-12.77848	H	-22.21679	-21.17907	-12.60948
C	-24.29641	-21.31377	-11.90185	C	-24.27125	-21.37828	-11.97431
N	-23.82334	-20.48078	-10.8006	N	-24.05036	-20.55384	-10.80509
H	-23.43936	-21.0673	-10.05794	H	-23.29391	-20.95158	-10.24579
H	-24.6055	-19.96329	-10.40932	H	-24.87879	-20.61758	-10.21821
H	-25.32573	-21.03783	-12.1501	H	-25.2592	-21.13368	-12.38205
C	-24.29125	-22.79917	-11.57206	C	-24.26394	-22.87958	-11.65493
O	-23.42571	-23.27204	-10.81361	O	-23.39846	-23.33824	-10.89201
C	-20.29882	-15.15118	-9.44572	C	-20.42312	-14.9125	-9.50397
H	-21.14769	-14.8615	-8.82584	H	-21.30217	-14.744	-8.88472
H	-19.78588	-14.25133	-9.79368	H	-19.98974	-13.94464	-9.77162
N	-25.19708	-23.52406	-12.22986	N	-25.18559	-23.61385	-12.27987
H	-25.91883	-23.04288	-12.78	H	-25.92551	-23.15322	-12.82333
C	-25.15636	-24.97136	-12.37797	C	-25.18208	-25.06966	-12.30217
H	-24.18619	-25.29394	-11.9778	H	-24.2126	-25.38457	-11.89671
C	-26.25033	-25.57393	-11.48307	C	-26.27875	-25.59592	-11.37286

H	-27.22814	-25.28056	-11.8972	**H**	-27.25008	-25.25801	-11.76827
C	-26.15786	-27.08828	-11.41156	**C**	-26.25442	-27.11321	-11.31329
H	-26.97092	-27.49058	-10.79918	**H**	-27.10876	-27.49682	-10.74793
H	-26.19607	-27.52947	-12.40994	**H**	-26.2654	-27.53009	-12.32288
H	-25.20833	-27.36818	-10.94106	**H**	-25.33656	-27.44071	-10.81219
O	-26.15817	-25.12852	-10.13796	**O**	-26.12537	-25.13986	-10.03611
H	-26.05756	-24.1589	-10.0844	**H**	-26.02782	-24.16963	-10.00588
C	-26.6861	-21.93344	-18.87169	**C**	-26.61138	-21.9262	-18.86775
H	-25.70102	-22.36682	-18.65907	**H**	-25.60088	-22.29683	-18.65654
H	-27.42675	-22.69814	-18.61498	**H**	-27.30828	-22.71781	-18.57181
C	-26.89279	-20.70609	-17.96447	**C**	-26.87822	-20.67377	-18.01702
H	-27.95823	-20.45822	-17.90346	**H**	-27.94708	-20.43167	-18.02121
H	-26.38612	-19.83336	-18.3877	**H**	-26.34971	-19.80921	-18.43382
C	-26.34699	-20.9366	-16.5471	**C**	-26.41672	-20.87517	-16.56691
H	-25.25414	-21.04484	-16.58562	**H**	-25.31882	-20.92482	-16.5376
H	-26.74945	-21.86329	-16.11989	**H**	-26.80328	-21.81948	-16.16527
C	-26.71843	-19.76271	-15.64918	**C**	-26.90755	-19.71843	-15.71098
H	-27.80854	-19.68609	-15.5822	**H**	-27.99845	-19.68061	-15.70406
H	-26.35049	-18.82334	-16.07502	**H**	-26.53145	-18.75864	-16.07791
N	-26.1817	-19.87611	-14.26733	**N**	-26.48652	-19.82429	-14.27549
H	-26.56188	-19.0795	-13.73769	**H**	-26.81184	-18.9752	-13.78153
H	-26.58555	-20.72765	-13.82985	**H**	-26.91714	-20.6552	-13.81077
H	-24.95014	-19.85296	-14.08744	**H**	-25.46149	-19.86725	-14.1636
Cl	-19.14713	-15.98816	-8.34606	**Cl**	-19.21279	-15.72047	-8.43386

	TS^{+g}				TS3		
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
C	-20.5004	-15.44136	-10.9483	**C**	-20.59913	-15.19358	-11.01631
C	-21.76252	-14.91292	-11.65557	**C**	-21.89878	-14.75635	-11.70487
C	-22.19391	-15.62642	-12.94138	**C**	-22.32943	-15.54885	-12.93185
C	-21.10332	-15.86445	-13.9601	**C**	-21.22519	-15.75433	-13.95092
O	-20.11679	-15.1439	-14.2062	**O**	-20.31182	-14.96172	-14.23836
N	-21.36105	-16.98713	-14.64285	**N**	-21.38425	-16.93757	-14.55033
H	-20.71926	-17.26279	-15.38768	**H**	-20.72829	-17.22104	-15.27947
H	-22.95221	-15.00405	-13.43487	**H**	-23.13027	-14.99178	-13.43145
H	-22.60959	-15.00117	-10.96268	**H**	-22.72078	-14.75623	-10.9792
H	-21.64915	-13.83965	-11.85573	**H**	-21.75876	-13.71359	-12.01325
H	-20.43702	-16.51727	-11.12886	**H**	-19.80626	-15.34129	-11.75796
H	-19.59022	-14.95358	-11.31863	**H**	-20.3012	-14.37956	-10.34852
O	-21.92157	-17.70806	-11.88398	**O**	-21.92361	-17.54911	-11.76082
C	-22.7489	-17.0736	-12.7068	**C**	-22.79623	-16.99948	-12.63925
C	-24.20616	-17.12773	-12.23536	**C**	-24.24612	-17.1119	-12.16724
H	-24.41521	-18.1508	-11.91003	**H**	-24.41567	-18.13299	-11.81733
H	-24.29776	-16.49183	-11.34665	**H**	-24.37923	-16.44842	-11.30405
H	-24.93123	-16.79649	-12.98186	**H**	-24.97692	-16.83857	-12.93192
C	-22.55811	-17.69489	-14.22372	**C**	-22.56902	-17.65965	-14.12049
C	-23.69682	-17.47511	-15.25342	**C**	-23.70302	-17.46494	-15.17746
C	-23.25329	-17.66428	-16.7182	**C**	-23.2224	-17.65971	-16.63163
C	-24.42679	-17.46457	-17.65038	**C**	-24.37684	-17.49753	-17.59622
C	-24.57319	-18.11363	-18.81019	**C**	-24.45929	-18.14122	-18.76566
C	-23.58081	-19.1212	-19.33158	**C**	-23.41699	-19.11006	-19.26233
C	-22.27251	-19.0816	-18.54071	**C**	-22.14184	-19.04633	-18.42107
C	-22.53515	-18.97765	-17.03754	**C**	-22.46393	-18.95599	-16.92854
H	-21.58809	-19.02819	-16.4964	**H**	-21.53555	-18.9884	-16.35635
H	-23.14121	-19.83571	-16.70738	**H**	-23.06434	-19.82724	-16.62378
H	-21.67086	-19.96714	-18.75822	**H**	-21.51267	-19.91747	-18.61993
H	-21.68195	-18.20992	-18.84855	**H**	-21.55552	-18.16177	-18.69815
H	-24.01916	-20.12983	-19.25503	**H**	-23.82927	-20.13145	-19.21969
H	-23.39944	-18.93523	-20.39632	**H**	-23.20117	-18.90273	-20.31723
H	-25.45381	-17.92038	-19.42132	**H**	-25.3226	-17.97028	-19.4076
H	-25.18987	-16.76127	-17.31889	**H**	-25.17254	-16.81914	-17.29085
H	-22.52952	-16.8518	-16.89063	**H**	-22.51701	-16.82964	-16.79926
O	-24.29443	-16.1991	-15.1233	**O**	-24.32159	-16.19736	-15.07847
H	-23.75513	-15.49423	-15.53464	**H**	-23.79539	-15.49601	-15.51577
H	-24.49962	-18.18477	-15.01971	**H**	-24.49851	-18.18447	-14.95149
C	-22.27607	-19.16781	-13.97469	**C**	-22.28298	-19.13213	-13.88809
O	-21.21342	-19.72212	-14.15594	**O**	-21.21625	-19.67965	-14.05791
O	-23.37515	-19.80837	-13.52814	**O**	-23.39187	-19.76999	-13.47423
C	-23.19307	-21.1625	-13.0297	**C**	-23.23856	-21.13964	-13.01486
C	-23.28561	-22.12665	-14.20468	**C**	-23.36426	-22.07196	-14.21103

H	-23.01757	-23.14362	-13.90657	H	-23.11737	-23.10066	-13.93633
H	-22.61803	-21.80708	-15.00646	H	-22.69395	-21.75155	-15.01147
H	-24.30948	-22.17669	-14.59375	H	-24.39221	-22.09258	-14.59311
H	-22.20635	-21.19766	-12.55661	H	-22.25127	-21.2155	-12.55096
C	-24.27144	-21.39767	-11.954	C	-24.31257	-21.35932	-11.93403
N	-24.04937	-20.56725	-10.79272	N	-24.0767	-20.51336	-10.7857
H	-23.31427	-20.97332	-10.21031	H	-23.43567	-20.97803	-10.14124
H	-24.89219	-20.58382	-10.22462	H	-24.95248	-20.39121	-10.28702
H	-25.24949	-21.14363	-12.37714	H	-25.29351	-21.11333	-12.35634
C	-24.27939	-22.89733	-11.64117	C	-24.29889	-22.85325	-11.60801
O	-23.40927	-23.36951	-10.89228	O	-23.41689	-23.30651	-10.86116
C	-20.64436	-15.19741	-9.4592	C	-20.69828	-16.44576	-10.18814
H	-21.42608	-15.83193	-9.04604	H	-20.2492	-17.36259	-10.52036
H	-20.84036	-14.15268	-9.20921	H	-21.50798	-16.54196	-9.47318
N	-25.21079	-23.61346	-12.26967	N	-25.21457	-23.5825	-12.23852
H	-25.9366	-23.13914	-12.81955	H	-25.94355	-23.11998	-12.79689
C	-25.19345	-25.06534	-12.35038	C	-25.18244	-25.03352	-12.32422
H	-24.22463	-25.386	-11.94797	H	-24.22309	-25.35372	-11.90095
C	-26.28945	-25.64527	-11.45305	C	-26.30719	-25.602	-11.45353
H	-27.26512	-25.33265	-11.85859	H	-27.26659	-25.26513	-11.87916
C	-26.218	-27.16239	-11.41512	C	-26.27437	-27.11943	-11.41661
H	-27.04645	-27.57537	-10.83198	H	-27.13955	-27.51087	-10.8732
H	-26.24122	-27.5712	-12.42766	H	-26.26036	-27.52723	-12.42956
H	-25.27959	-27.4702	-10.94112	H	-25.36791	-27.44935	-10.89803
O	-26.17245	-25.20933	-10.10679	O	-26.20885	-25.1739	-10.1026
H	-26.0795	-24.23956	-10.05727	H	-26.11006	-24.20422	-10.04706
C	-26.66176	-21.91706	-18.86627	C	-26.63051	-21.93012	-18.88969
H	-25.6583	-22.29951	-18.64418	H	-25.61737	-22.29043	-18.6757
H	-27.36784	-22.70564	-18.58478	H	-27.31969	-22.72993	-18.59947
C	-26.92827	-20.67132	-18.0059	C	-26.91734	-20.67983	-18.04434
H	-27.99872	-20.44015	-17.99378	H	-27.99066	-20.46144	-18.04198
H	-26.41673	-19.79977	-18.42615	H	-26.41377	-19.8065	-18.47163
C	-26.44106	-20.87273	-16.56326	C	-26.43902	-20.86342	-16.59862
H	-25.34247	-20.91822	-16.5548	H	-25.33984	-20.87517	-16.58073
H	-26.81501	-21.81867	-16.15445	H	-26.78934	-21.81992	-16.19342
C	-26.9195	-19.71543	-15.70076	C	-26.96148	-19.72227	-15.7422
H	-28.00975	-19.68109	-15.67466	H	-28.0527	-19.71735	-15.73604
H	-26.55112	-18.75617	-16.07478	H	-26.61004	-18.75255	-16.10602
N	-26.47806	-19.81629	-14.27097	N	-26.5375	-19.8213	-14.30691
H	-26.79208	-18.96426	-13.7737	H	-26.85898	-18.97018	-13.8125
H	-26.90657	-20.64465	-13.79983	H	-26.96051	-20.65208	-13.83451
H	-25.45444	-19.86836	-14.17055	H	-25.5155	-19.86806	-14.19884
Cl	-19.14318	-15.64744	-8.55166	Cl	-19.12632	-16.0701	-8.72823

TS4

ATOM	X	Y	Z	ATOM	X	Y	Z
C	-20.62722	-15.67606	-10.73956	H	-22.67616	-23.19167	-14.09413
C	-21.78738	-15.05296	-11.51651	H	-22.27759	-21.87951	-15.23457
C	-22.17233	-15.73624	-12.82446	H	-23.94241	-22.42064	-15.04286
C	-21.08028	-15.881	-13.85474	H	-22.37386	-21.21008	-12.69045
O	-20.13031	-15.10552	-14.07908	C	-24.52009	-21.49138	-12.70422
N	-21.31909	-16.96152	-14.60142	N	-24.99287	-20.27667	-12.03996
H	-20.69209	-17.17417	-15.38036	H	-24.22862	-19.65654	-11.7422
H	-22.90633	-15.08222	-13.3169	H	-25.60462	-20.45054	-11.24127
H	-22.67816	-15.04446	-10.87635	H	-25.26674	-21.69322	-13.47861
H	-21.55694	-13.99913	-11.73382	C	-24.37362	-22.79154	-11.90732
H	-20.85562	-16.72712	-10.54422	O	-23.44265	-22.98972	-11.11811
H	-19.70601	-15.61624	-11.33432	C	-20.44234	-14.90445	-9.44718
O	-22.2754	-17.98046	-11.82432	H	-21.34516	-14.92966	-8.84112
C	-22.83597	-17.16235	-12.72191	H	-20.15425	-13.86376	-9.61687
C	-24.33504	-16.98417	-12.44301	N	-25.2753	-23.68165	-12.31276
H	-24.86102	-17.93664	-12.48441	H	-26.06149	-23.33629	-12.87668
H	-24.42284	-16.61156	-11.4166	C	-25.14426	-25.12291	-12.31064
H	-24.81581	-16.30616	-13.14131	H	-24.14056	-25.36119	-11.93545
C	-22.52313	-17.69271	-14.25525	C	-26.17502	-25.73126	-11.35812
C	-23.64486	-17.43824	-15.28995	H	-27.17863	-25.49549	-11.74587
C	-23.21287	-17.65882	-16.7496	C	-26.01493	-27.23798	-11.25111
C	-24.3987	-17.45469	-17.66594	H	-26.78136	-27.66242	-10.59627
C	-24.56623	-18.11259	-18.81665	H	-26.0839	-27.70077	-12.2386
C	-23.58152	-19.12165	-19.34823	H	-25.03258	-27.47286	-10.82554

C	-22.26702	-19.09314	-18.56802	**O**	-26.04438	-25.21735	-10.04144
C	-22.52301	-18.98785	-17.06501	**H**	-26.02477	-24.2414	-10.04632
H	-21.57169	-19.05317	-16.53146	**C**	-26.6506	-21.9712	-18.65763
H	-23.14784	-19.82998	-16.73233	**H**	-25.67455	-22.42807	-18.45283
H	-21.67579	-19.9853	-18.78739	**H**	-27.40665	-22.71671	-18.3877
H	-21.66961	-18.22679	-18.87833	**C**	-26.82216	-20.73724	-17.73799
H	-24.0235	-20.12775	-19.26581	**H**	-27.86177	-20.39285	-17.77304
H	-23.40865	-18.93576	-20.414	**H**	-26.20353	-19.91134	-18.10609
H	-25.46778	-17.94063	-19.40266	**C**	-26.44372	-20.98796	-16.25438
H	-25.16465	-16.76522	-17.31345	**H**	-25.37645	-21.23415	-16.1854
H	-22.47347	-16.86295	-16.93901	**H**	-27.00391	-21.83559	-15.84447
O	-24.1807	-16.13252	-15.17907	**C**	-26.74042	-19.71222	-15.457
H	-23.63589	-15.45281	-15.62161	**H**	-27.75901	-19.39054	-15.70406
H	-24.46857	-18.10779	-15.026	**H**	-26.07827	-18.91197	-15.79674
C	-22.2095	-19.16972	-14.1078	**N**	-26.65824	-19.75183	-13.96026
O	-21.10702	-19.67065	-14.15276	**H**	-27.01192	-18.82622	-13.68205
O	-23.33282	-19.87649	-13.93122	**H**	-27.35239	-20.42489	-13.58794
C	-23.19998	-21.21717	-13.40946	**H**	-25.69433	-19.86475	-13.04341
C	-23.00446	-22.23699	-14.51415	**Cl**	-19.14465	-15.61789	-8.41414

Table S15. Cartesian coordinates (in Å) of QM atoms for Transition State structure localized at M06-2X/6-31+G(d,p)/MM level in the intramolecular cyclization step for SalA and hSalA in the aqueous solution.

	SalA TS^{wat}				hSalA TS^{wat}		
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
C	-4.67776	0.24424	1.0051	**C**	4.29246	-0.8217	1.12546
C	-4.293	-1.11427	0.44973	**C**	3.67502	-1.59964	-0.03876
C	-2.77109	-1.25867	0.2931	**C**	2.17895	-1.38982	-0.22965
C	-2.31423	-1.41315	-1.14125	**C**	1.37727	-1.73482	1.00207
O	-2.88194	-2.0827	-2.02602	**O**	1.5829	-2.68237	1.77442
N	-1.17206	-0.74504	-1.3222	**N**	0.34855	-0.87335	1.12881
H	-0.76087	-0.61755	-2.25264	**H**	-0.37195	-1.03388	1.8309
H	-2.41867	-2.14347	0.82828	**H**	1.81945	-2.06364	-1.01893
H	-4.66271	-1.89416	1.11507	**H**	4.19334	-1.34623	-0.97165
H	-4.79918	-1.25655	-0.51016	**H**	3.84967	-2.66635	0.13449
H	-4.2661	0.6181	1.92079	**H**	3.65387	-0.87534	2.01048
H	-5.13781	0.96133	0.34735	**H**	5.23481	-1.3024	1.39029
O	-2.80865	1.08466	0.34933	**O**	2.49536	0.96041	0.16402
C	-2.03812	0.02885	0.75471	**C**	1.76357	0.06971	-0.55195
C	-1.7389	0.06863	2.25953	**C**	1.77846	0.36539	-2.05196
H	-1.27981	1.03029	2.50614	**H**	1.34252	1.34876	-2.24077
H	-2.67583	-0.00612	2.81672	**H**	2.82466	0.4049	-2.36351
H	-1.09677	-0.74201	2.60234	**H**	1.26235	-0.39288	-2.64937
C	-0.71164	-0.03333	-0.14442	**C**	0.25175	0.10484	0.06703
C	0.4868	-0.77062	0.53001	**C**	-0.91438	-0.25766	-0.88458
C	1.70481	-0.9775	-0.40031	**C**	-2.29927	-0.35769	-0.21005
C	2.65123	-2.015	0.16762	**C**	-3.36253	-0.7686	-1.20581
C	3.97933	-1.87305	0.2156	**C**	-4.63567	-0.36172	-1.13695
C	4.71024	-0.67642	-0.32615	**C**	-5.16346	0.56045	-0.06994
C	3.82054	0.12619	-1.2718	**C**	-4.17374	0.70743	1.08456
C	2.43796	0.34841	-0.66145	**C**	-2.74248	0.88262	0.57395
H	1.85026	0.984	-1.32798	**H**	-2.07844	1.04053	1.42833
H	2.54918	0.88416	0.29138	**H**	-2.67837	1.76914	-0.07369
H	4.27849	1.09245	-1.50329	**H**	-4.45819	1.55401	1.71861
H	3.72038	-0.42004	-2.21749	**H**	-4.21043	-0.18973	1.71638
H	5.04092	-0.04843	0.51178	**H**	-5.37874	1.54084	-0.51576
H	5.61831	-1.0048	-0.83944	**H**	-6.12128	0.18227	0.3002
H	4.57648	-2.66465	0.66635	**H**	-5.34586	-0.72038	-1.88012
H	2.20771	-2.92789	0.5572	**H**	-3.06177	-1.46293	-1.98691
H	1.3189	-1.37255	-1.35489	**H**	-2.21195	-1.18913	0.50579
O	0.05136	-2.005	1.08737	**O**	-0.6353	-1.51931	-1.49883
H	-0.0604	-2.70822	0.39124	**H**	-0.79105	-2.21344	-0.83849
H	0.79887	-0.16496	1.38805	**H**	-0.9379	0.48293	-1.69064
C	-0.39968	1.40273	-0.51421	**C**	0.15165	1.5257	0.57857
O	-0.52577	1.88552	-1.62315	**O**	0.21777	1.88	1.74981
O	0.04941	2.0887	0.53499	**O**	-0.06678	2.34418	-0.42994
C	0.24772	3.49309	0.32051	**C**	-0.01549	3.75907	-0.20442
H	1.09241	3.6582	-0.34637	**H**	-0.66417	4.04688	0.62369
Cl	-6.47629	-0.25982	2.04614	**C**	4.54945	0.63022	0.82162
H	0.47264	3.90883	1.29865	**H**	4.17277	1.41512	1.45601

H	-0.66046	3.93751	-0.08451	**H**	4.82135	0.91504	-0.18592

Table S16. Cartesian coordinates (in Å) of QM atoms for Transition State structures localized at M06-2X/6-31+G(d,p)/MM level for ester bond hydrolysis in hSalA-proteasome complex.

E-I2 hydrolysis: TS5^{hyd}				E-PC$^{(OH)}$ hydrolysis: TS6^{hyd}			
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
C	-20.57874	-15.23742	-10.86563	**C**	-20.86152	-15.27248	-10.92077
C	-21.91294	-14.87629	-11.52531	**C**	-22.17316	-14.90712	-11.62568
C	-22.2731	-15.64598	-12.8061	**C**	-22.4679	-15.6834	-12.91881
C	-21.11326	-15.87808	-13.76041	**C**	-21.24244	-15.89592	-13.79053
O	-20.22052	-15.06281	-14.05911	**O**	-20.36035	-15.04552	-14.00935
N	-21.19405	-17.10911	-14.28235	**N**	-21.24924	-17.13356	-14.28711
H	-20.56995	-17.37461	-15.04383	**H**	-20.57298	-17.40428	-15.00319
H	-23.00734	-15.05116	-13.35767	**H**	-23.16626	-15.0973	-13.52022
H	-22.72557	-15.03842	-10.80738	**H**	-23.01708	-15.0474	-10.94168
H	-21.90739	-13.80343	-11.75887	**H**	-22.14396	-13.84101	-11.87947
H	-20.55933	-16.30393	-10.64362	**H**	-20.85348	-16.32788	-10.64614
H	-19.7582	-15.0291	-11.55898	**H**	-20.02994	-15.10197	-11.61234
O	-22.07866	-17.54044	-11.47988	**O**	-22.49269	-17.53816	-11.51527
C	-22.81003	-17.08576	-12.55979	**C**	-23.05479	-17.10887	-12.75608
C	-24.29648	-17.1964	-12.22206	**C**	-24.56496	-17.21406	-12.62196
H	-24.49183	-18.23079	-11.9267	**H**	-24.86766	-18.26814	-12.63431
H	-24.49673	-16.57149	-11.34538	**H**	-24.87237	-16.76603	-11.67114
H	-24.96435	-16.89918	-13.02971	**H**	-25.08079	-16.72918	-13.44318
C	-22.41169	-17.8308	-13.92891	**C**	-22.46943	-17.89602	-14.02427
C	-23.48207	-17.58324	-15.07566	**C**	-23.46764	-17.70093	-15.25023
C	-23.03108	-17.70592	-16.55954	**C**	-22.9609	-17.87225	-16.70996
C	-24.24718	-17.44941	-17.43473	**C**	-24.15471	-17.64549	-17.62976
C	-24.47409	-18.04912	-18.60891	**C**	-24.29291	-18.21112	-18.83318
C	-23.53485	-19.04474	-19.23628	**C**	-23.2755	-19.13377	-19.44729
C	-22.17652	-19.01238	-18.54184	**C**	-21.9729	-19.1091	-18.6509
C	-22.32684	-18.98474	-17.02057	**C**	-22.23032	-19.14208	-17.14286
H	-21.34549	-19.04353	-16.55965	**H**	-21.28619	-19.20349	-16.61263
H	-22.88343	-19.87133	-16.68664	**H**	-22.81265	-20.03251	-16.87495
H	-21.5763	-19.87847	-18.83409	**H**	-21.33267	-19.94893	-18.93473
H	-21.62354	-18.11915	-18.85765	**H**	-21.41393	-18.19482	-18.88967
H	-23.966	-20.05518	-19.14559	**H**	-23.68339	-20.15681	-19.45033
H	-23.43947	-18.83666	-20.30838	**H**	-23.10917	-18.85462	-20.49472
H	-25.3869	-17.81082	-19.15405	**H**	-25.19823	-18.01588	-19.4067
H	-24.9835	-16.74729	-17.04769	**H**	-24.95617	-17.01669	-17.24743
H	-22.31589	-16.87695	-16.70492	**H**	-22.24237	-17.0435	-16.85629
O	-24.05442	-16.28883	-14.9435	**O**	-23.97172	-16.36119	-15.20304
H	-23.58819	-15.62275	-15.47968	**H**	-23.46438	-15.77055	-15.78088
H	-24.30645	-18.28346	-14.89976	**H**	-24.32253	-18.36433	-15.06875
C	-22.1495	-19.36855	-13.8281	**C**	-22.11225	-19.43403	-13.86616
O	-21.30479	-19.9124	-14.5255	**O**	-21.15452	-19.86252	-14.53335
O	-23.35803	-19.98328	-13.52425	**O**	-23.35207	-20.07543	-13.8661
C	-23.29564	-21.37999	-13.17626	**C**	-23.37115	-21.45651	-13.50212
C	-23.64832	-22.24994	-14.37159	**C**	-23.82295	-22.32443	-14.65984
H	-23.46008	-23.30087	-14.1389	**H**	-23.71574	-23.38191	-14.40893
H	-23.03834	-21.9628	-15.2316	**H**	-23.21042	-22.10345	-15.53792
H	-24.71006	-22.17994	-14.63638	**H**	-24.87315	-22.13757	-14.89104
H	-22.26888	-21.58673	-12.86043	**H**	-22.36117	-21.74637	-13.19141
C	-24.24267	-21.54053	-11.95932	**C**	-24.33575	-21.53859	-12.29245
N	-23.78539	-20.72062	-10.85277	**N**	-23.85173	-20.63511	-11.20585
H	-23.30287	-21.2858	-10.15009	**H**	-23.07162	-18.21045	-11.13972
H	-24.57405	-20.29189	-10.37904	**H**	-23.61521	-21.20665	-10.38372
H	-25.24523	-21.21722	-12.26351	**H**	-25.32137	-21.20802	-12.63455
C	-24.30729	-23.02199	-11.60864	**C**	-24.35211	-22.9592	-11.76014
O	-23.44717	-23.51184	-10.85904	**O**	-23.47819	-23.31351	-10.94325
C	-20.38028	-14.44401	-9.58622	**C**	-20.65167	-14.43166	-9.67531
H	-20.88293	-14.90561	-8.73569	**H**	-21.2806	-14.75099	-8.84495
H	-20.69339	-13.40117	-9.67443	**H**	-20.80328	-13.36625	-9.85802
N	-25.26055	-23.72024	-12.22248	**N**	-25.26274	-23.73806	-12.31443
H	-25.97324	-23.23128	-12.77794	**H**	-26.0328	-23.30746	-12.85179
C	-25.24659	-25.17131	-12.31589	**C**	-25.1644	-25.18388	-12.38344
H	-24.28318	-25.49795	-11.90674	**H**	-24.17415	-25.45531	-11.99846
C	-26.35467	-25.74869	-11.43474	**C**	-26.22063	-25.83461	-11.49703
H	-27.32472	-25.45377	-11.86599	**H**	-27.21239	-25.57634	-11.89808

C	-26.26286	-27.2635	-11.37804	C	-26.05427	-27.34477	-11.48953
H	-27.1055	-27.68758	-10.82432	H	-26.82794	-27.8207	-10.87972
H	-26.2399	-27.6776	-12.38896	H	-26.10667	-27.7302	-12.51111
H	-25.33778	-27.55208	-10.86772	H	-25.07448	-27.60994	-11.07693
O	-26.26864	-25.28327	-10.09565	O	-26.11701	-25.3861	-10.1511
H	-26.13549	-24.31894	-10.07519	H	-26.0587	-24.41608	-10.13933
C	-26.64415	-21.922	-18.86823	C	-26.69173	-21.89121	-18.67764
H	-25.63724	-22.29926	-18.65324	H	-25.68355	-22.24052	-18.42675
H	-27.34675	-22.71094	-18.58034	H	-27.38977	-22.68687	-18.39943
C	-26.90865	-20.67092	-18.01502	C	-27.01229	-20.61911	-17.88605
H	-27.9797	-20.4429	-17.99636	H	-28.08878	-20.4142	-17.91394
H	-26.40198	-19.80053	-18.44523	H	-26.50645	-19.75438	-18.33393
C	-26.40726	-20.87284	-16.57909	C	-26.55916	-20.76474	-16.44126
H	-25.3091	-20.91357	-16.58336	H	-25.46172	-20.74874	-16.40524
H	-26.77532	-21.8246	-16.17741	H	-26.88777	-21.73451	-16.04354
C	-26.88162	-19.72863	-15.69825	C	-27.09868	-19.6542	-15.5603
H	-27.97175	-19.68505	-15.67622	H	-28.19577	-19.62357	-15.68198
H	-26.50088	-18.76638	-16.05141	H	-26.72127	-18.68906	-15.92497
N	-26.44644	-19.8614	-14.2676	N	-26.68028	-19.80818	-14.16801
H	-26.76069	-19.01792	-13.7592	H	-26.91026	-18.92885	-13.70356
H	-26.88193	-20.69497	-13.81135	H	-27.25603	-20.52408	-13.70937
H	-25.42114	-19.92563	-14.1671	H	-24.6021	-19.98322	-10.94714
Cl	-18.61033	-14.37209	-9.17136	Cl	-18.932	-14.59235	-9.11706
O	-21.25426	-19.69299	-12.02233	O	-21.67199	-19.81241	-12.2061
H	-21.96543	-20.06292	-11.46784	H	-22.8615	-20.15343	-11.57467
H	-21.6144	-18.50843	-11.68127	H	-21.2743	-19.00968	-11.85048

Table S17. Cartesian coordinates (in Å) of QM atoms for Transition State structure localized at M06-2X/6-31+G(d,p)/MM level for the ester bond hydrolysis in final inhibition product with SalA-proteasome complex.

TS6^{SalA}							
ATOM	X	Y	Z	ATOM	X	Y	Z
C	-21.31641	-16.17602	-10.96092	H	-23.60492	-23.30939	-14.20543
C	-21.93594	-14.93493	-11.5661	H	-23.21018	-21.98707	-15.32498
C	-22.41609	-15.48817	-12.9022	H	-24.86371	-22.18952	-14.67783
C	-21.24009	-15.76203	-13.83711	H	-22.3622	-21.60776	-13.00468
O	-20.37013	-14.93816	-14.14354	C	-24.29174	-21.50294	-12.02697
N	-21.26708	-17.04245	-14.23721	N	-23.75299	-20.64791	-10.95153
H	-20.60728	-17.37737	-14.94079	H	-23.41513	-21.22645	-10.17233
H	-23.14598	-14.85032	-13.39618	H	-24.48865	-20.03729	-10.59065
H	-22.7819	-14.58983	-10.96834	H	-25.30957	-21.18357	-12.26934
H	-21.2173	-14.1218	-11.6703	C	-24.31193	-22.96262	-11.60937
H	-21.28908	-16.20173	-9.87205	O	-23.4205	-23.39141	-10.85867
H	-20.31044	-16.34733	-11.35846	N	-25.25384	-23.68662	-12.20041
O	-22.17422	-17.24493	-11.41081	H	-25.97998	-23.21533	-12.75631
C	-22.93814	-16.90605	-12.60734	C	-25.17407	-25.12933	-12.35732
C	-24.39192	-17.09806	-12.22814	H	-24.19098	-25.42781	-11.97251
H	-24.54702	-18.14674	-11.96256	C	-26.2389	-25.80201	-11.48934
H	-24.62331	-16.49001	-11.34926	H	-27.2308	-25.56063	-11.90251
H	-25.05986	-16.8424	-13.047	C	-26.05578	-27.31037	-11.46354
C	-22.48082	-17.76597	-13.86303	H	-26.82553	-27.78044	-10.84483
C	-23.5424	-17.58645	-15.03358	H	-26.10156	-27.72014	-12.47565
C	-23.06481	-17.76268	-16.49789	H	-25.07607	-27.55503	-11.03809
C	-24.25851	-17.526	-17.41041	O	-26.16321	-25.35855	-10.14291
C	-24.4495	-18.15495	-18.57562	H	-26.11073	-24.38813	-10.10754
C	-23.49384	-19.16852	-19.14664	C	-26.67699	-21.94165	-18.91672
C	-22.16025	-19.12648	-18.40414	H	-25.66954	-22.31376	-18.69555
C	-22.36772	-19.06556	-16.89008	H	-27.3773	-22.73594	-18.63789
H	-21.41307	-19.1332	-16.37691	C	-26.95953	-20.69472	-18.06711
H	-22.96125	-19.92807	-16.55576	H	-28.03364	-20.48397	-18.04621
H	-21.55315	-19.99802	-18.66107	H	-26.46962	-19.81643	-18.49942
H	-21.59345	-18.24027	-18.71741	C	-26.45193	-20.88501	-16.63422
H	-23.93222	-20.17543	-19.05263	H	-25.35272	-20.87487	-16.63616
H	-23.36008	-18.98179	-20.21871	H	-26.77904	-21.85421	-16.23843
H	-25.33985	-17.92105	-19.15801	C	-26.98156	-19.76701	-15.75438
H	-25.00056	-16.80529	-17.07315	H	-28.07199	-19.78905	-15.72027
H	-22.33824	-16.94706	-16.6599	H	-26.65633	-18.78746	-16.11482
O	-24.1292	-16.29456	-14.95161	N	-26.51657	-19.87342	-14.33104
H	-23.65937	-15.62856	-15.49251	H	-26.84316	-19.0338	-13.82406
H	-24.35966	-18.28562	-14.84092	H	-26.92288	-20.71132	-13.85714

C	-22.15032	-19.31143	-13.61798	**H**	-25.48798	-19.91404	-14.25834
O	-21.22013	-19.80643	-14.29104	**Cl**	-19.03942	-13.71051	-10.06035
O	-23.44681	-19.98376	-13.66228	**O**	-21.80293	-19.53861	-12.05841
C	-23.39412	-21.36595	-13.2844	**H**	-22.71726	-20.00015	-11.48223
C	-23.79605	-22.26076	-14.44401	**H**	-21.73694	-18.62632	-11.649

Table S18. Cartesian coordinates (in Å) of QM atoms for Transition State structure localized at M06-2X/6-31+G(d,p)/MM level for the alternative product E-PC$^{(OH)}$ formation of 20S proteasome with SalA.

			TS4				
ATOM	**X**	**Y**	**Z**	**ATOM**	**X**	**Y**	**Z**
C	-20.73892	-15.77982	-10.79755	**C**	-22.94712	-22.23702	-14.49691
C	-21.85024	-15.04864	-11.52756	**H**	-22.77248	-23.23288	-14.07845
C	-22.24706	-15.70718	-12.84742	**H**	-22.09607	-21.93777	-15.09381
C	-21.11464	-15.8695	-13.83516	**H**	-23.81751	-22.30098	-15.15304
O	-20.15607	-15.09203	-14.01569	**H**	-22.38218	-21.24694	-12.63836
N	-21.30412	-16.97024	-14.55977	**C**	-24.52877	-21.53051	-12.73289
H	-20.64885	-17.19759	-15.31153	**N**	-25.04909	-20.33672	-12.06078
H	-22.96735	-15.04423	-13.34179	**H**	-24.31189	-19.71281	-11.71871
H	-22.72711	-14.96074	-10.88009	**H**	-25.68146	-20.54301	-11.28538
H	-21.51233	-14.03325	-11.76028	**H**	-25.24757	-21.73045	-13.53453
H	-21.05635	-16.4897	-10.03976	**C**	-24.40817	-22.82186	-11.92238
H	-20.0625	-16.25153	-11.50245	**O**	-23.46789	-23.03107	-11.1459
O	-22.20929	-17.84897	-11.78665	**N**	-25.34658	-23.68911	-12.28625
C	-22.85272	-17.14153	-12.72585	**H**	-26.111	-23.34464	-12.88343
C	-24.35854	-17.05813	-12.45515	**C**	-25.19018	-25.12829	-12.32585
H	-24.83702	-18.02726	-12.57426	**H**	-24.18099	-25.35803	-11.95905
H	-24.47807	-16.75977	-11.40948	**C**	-26.20398	-25.79158	-11.39685
H	-24.86783	-16.36291	-13.11669	**H**	-27.21354	-25.57833	-11.7811
C	-22.50717	-17.70995	-14.22966	**C**	-25.99636	-27.29502	-11.32709
C	-23.61555	-17.45458	-15.27957	**H**	-26.73738	-27.75517	-10.66724
C	-23.181	-17.68429	-16.73639	**H**	-26.07096	-27.73879	-12.32299
C	-24.35613	-17.42216	-17.65365	**H**	-24.99878	-27.51134	-10.92753
C	-24.55923	-18.07309	-18.80229	**O**	-26.08841	-25.30261	-10.07001
C	-23.62261	-19.12304	-19.33878	**H**	-26.08538	-24.32747	-10.05879
C	-22.30361	-19.14419	-18.56627	**C**	-26.68765	-21.98307	-18.6638
C	-22.54596	-19.039	-17.05956	**H**	-25.71558	-22.44468	-18.45157
H	-21.59419	-19.14571	-16.53199	**H**	-27.45114	-22.71935	-18.39119
H	-23.20207	-19.85783	-16.72931	**C**	-26.85606	-20.73908	-17.75846
H	-21.74387	-20.05356	-18.79411	**H**	-27.89361	-20.38996	-17.79788
H	-21.68062	-18.29597	-18.87732	**H**	-26.23433	-19.91722	-18.13191
H	-24.10432	-20.1102	-19.25383	**C**	-26.48403	-20.98483	-16.27493
H	-23.45007	-18.94424	-20.40572	**H**	-25.41962	-21.24405	-16.20148
H	-25.45256	-17.85555	-19.38482	**H**	-27.05333	-21.82515	-15.86344
H	-25.08485	-16.69138	-17.30757	**C**	-26.76958	-19.70597	-15.48086
H	-22.40527	-16.92158	-16.9149	**H**	-27.78195	-19.37037	-15.73435
O	-24.11366	-16.13326	-15.16409	**H**	-26.09485	-18.91255	-15.81437
H	-23.58685	-15.48431	-15.66798	**N**	-26.6909	-19.76251	-13.98765
H	-24.45403	-18.10897	-15.022	**H**	-27.04781	-18.84245	-13.69663
C	-22.18468	-19.18329	-14.07262	**H**	-27.38433	-20.44249	-13.62708
O	-21.08007	-19.68048	-14.12003	**H**	-25.71366	-19.90605	-13.02186
O	-23.30528	-19.88698	-13.88102	**Cl**	-19.66149	-14.52997	-9.90829
C	-23.18481	-21.23843	-13.38311				

4.2.3 Binding and reactivity of C5-phenyl-SalA (pSalA)

SUPORTING INFORMATION

Activity Cliff in 20S Proteasome β-lactone inhibitors: the case of salinosporamide A and its phenyl analogue

Natalia Serrano-Aparicio,[a] Laura Scalvini,[b] Alessio Lodola,[b] Vicent Moliner,[a] Katarzyna Świderek[a]

[a]Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castellón, Spain

[b]Department of Food and Drug, University of Parma, Parma, Italy

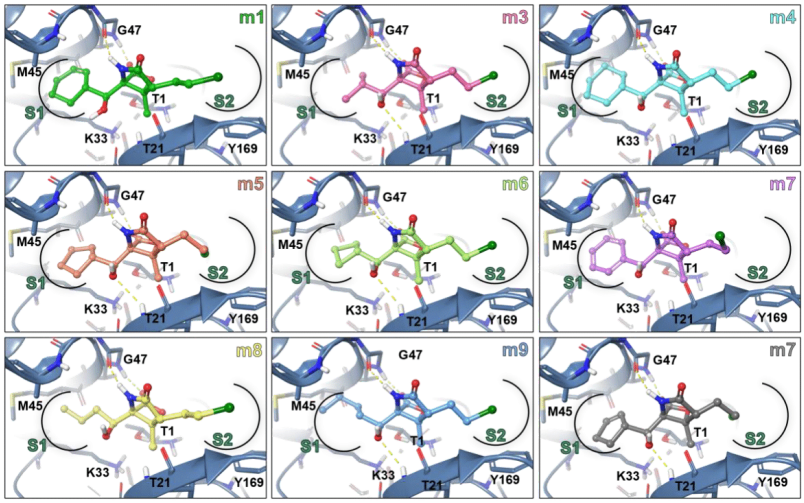


Figure S1. Representation of all docking poses for the compounds in the data set obtained with restrained docking and in the Model A-LysNH_3^+.

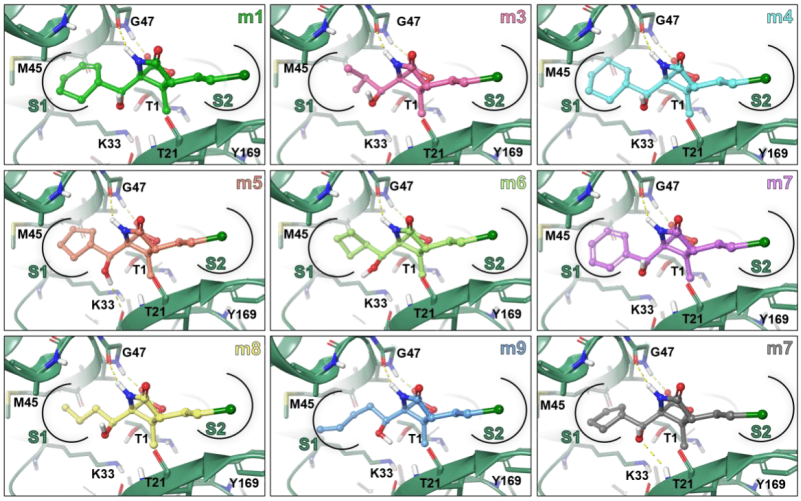


Figure S2. Representation of all docking poses for the compounds in the data set obtained with restrained docking and in the Model B-LysNH_2.

Calibration of the SMD simulation parameters

Both, the force constant (*k*) and the constant velocity (*v*) were selected using the following steps. Firstly, SMD simulations were conducted for both inhibitor molecules at different values of *k* (5,10,15,20,25 kcal·mol^{-1}·A^{-2}) and constant velocity (0.0005 Å·ps^{-1}). Secondly, SMD simulations were conducted setting different velocities (0.005, 0.001, 0.0005, 0.00025, 0.000125, 0.000008 Å·ps^{-1}) and keeping *k* fixed as selected in the previous step. The controlled distance was compared to the reference, and the lowest force constant that allowed the conservation of the stiff-spring approximation was selected. Finally, the same criteria was used for choosing the velocity, being the final parameters for the SMD simulation the following: *k*=15 kcal·mol^{-1}·A^{-2} and *v*=0.0005 Å·ps^{-1}.

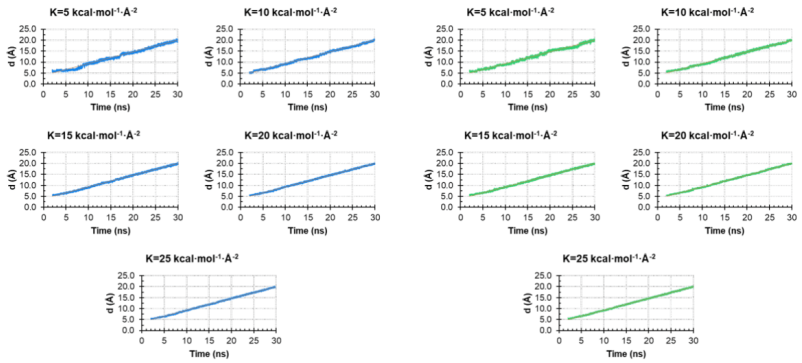


Figure S3. SMD calibration of the force constant (k) by using the same velocity (*v*=0.0005 Å·ps^{-1}) at different values of k, for each inhibitor. Blue lines correspond to SalA k calibration and green lines correspond to pSalA k calibration.

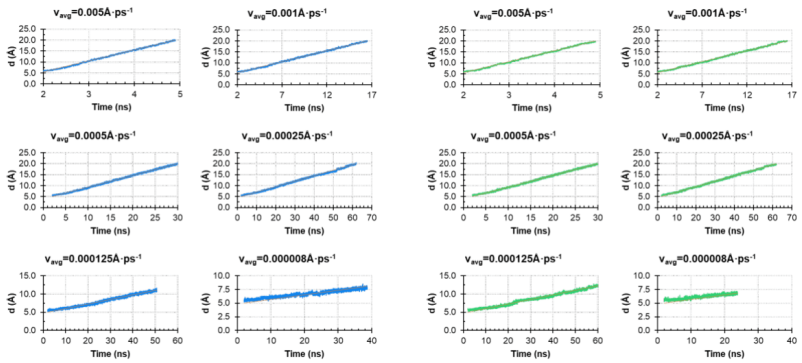


Figure S4. SMD calibration of the velocity (*v*) at the selected force constant (k=15 kcal·mol^{-1}·A^{-2}) at different values of k, for each inhibitor. Blue lines correspond to SalA *v* calibration and green lines correspond to pSalA *v* calibration.

Since the result of the different replicas can vary depending on the initial MD simulation seed and the path that the inhibitor follows while being pulled out of the active site, it is necessary to do a certain number of replicas. Once the parameters for the SMD calculations where set, the number of replicas was stablished by controlling the evolution of the standard error of the mean (SEM) when more replicas were added. By observing a reduction in the SEM until it can be considered nearly to constant, the proper exploration of the unbinding path is ensured. Therefore, the SEM was plotted along the number of replicas, and more replicas were added until the final values variation was close to zero (see Figure S5). SEM is plotted when the inhibitors are considered unbound (15 Å), since this is the point of larger error due to the increased degrees of freedom. In this plot we can see how for few replicas the variability of SEM is large. When more replicas are added it starts to decrease slowly, at 10 replicas, the value goes lower than 1.5. Finally, 15 replicas are enough to ensure the proper exploration. In the plot the regression line for the final 4 points has a slope close to 0, which means that the parameter with the addition of the last 4 replicas has varied very little. This is the chosen criteria for deciding the number of replicas. But it is necessary to remember that this result will depend on *k* and v, and for different conditions of these two parameters the number of replicas should be evaluated.

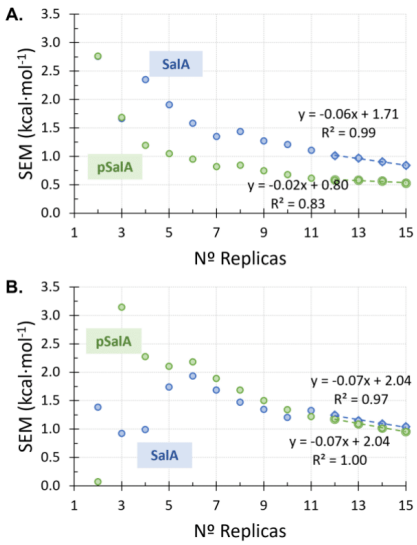


Figure S5. Evolution of SEM with the addition of replicas at d=15Å, k=15 kcal·mol^{-1}·A-2 v=0.0005 Å·ps^{-1}. A. SEM evolution in Model A, with Lys33:NH_3^+. B. SEM evolution in Model B, with Lys33:NH_2.

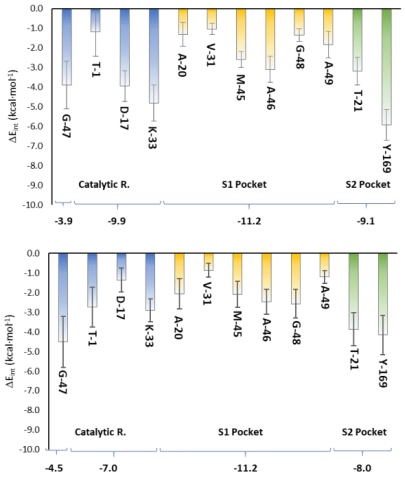


Figure S6. Interactions obtained at AM1/MM level along 100ps of MD simulation for both inhibitors.

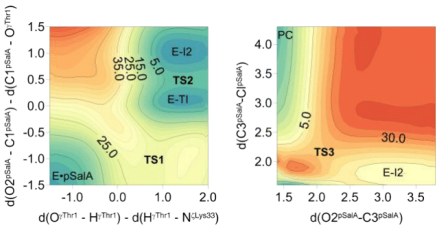


Figure S7. 2D-PMF surfaces computed at AM1/MM level of theory at 37 °C for the 20S proteasome inhibition with pSalA.