







**Escola Superior de Tecnologia i Ciències Experimentals**

**Departament de Química Física i Analítica**

**Àrea de Química Física**

**DOCTORAL THESIS**

**European Joint Doctorate on Theoretical Chemistry and  
Computational Modelling**

*Computational Studies of the Mechanisms of Catalysis  
and Inhibition of Cysteine Proteases*

**Kemel Arafet Cruz**

**Castelló, September 2017**



**Dr. D. Vicent Moliner Ibáñez**, profesor Catedrático del Departament de Química Física i Analítica de la *Universitat Jaume I* de Castellón, y **Dra. D<sup>a</sup>. Silvia Ferrer Castillo**, Personal Investigador del Departament d'Enginyeria Mecànica i Construcció de la *Universitat Jaume I* de Castellón,

**CERTIFICAN:**

Que el trabajo titulado *Computational Studies of the Mechanisms of Catalysis and Inhibition of Cysteine Proteases* ha sido realizado por Don Kemel Arafet Cruz bajo nuestra dirección, para optar al grado de Doctor en Química.

Así, autorizan la presentación de este trabajo a efectos de seguir los trámites correspondientes de la *Universitat Jaume I* de Castellón.

Castellón de la Plana, a 12 de Junio de 2017.

**DIRECTORES**

Dr. D. Vicent Moliner Ibáñez      Dra. D<sup>a</sup>. Silvia Ferrer Castillo



*A mi esposa Grety*

*A mi hijo Thiago*

*A mis padres*

*A mi hermano*

*A mis abuelos Olga y Manolo*

*A mi familia*





## **Acknowledgements in Spanish**

Han pasado casi 7 años desde mi aterrizaje en esta tierra que me ha acogido con tanto cariño, tan así, que en el momento de escribir estas líneas estoy esperando el nacimiento de mi primogénito Thiago. Y entre correcciones y correcciones, nació Thiago.

Casi 7 años de haber iniciado esta difícil etapa en mi vida por la lejanía de mi familia, pero a la vez apasionante por lo que ha significado, que culmina con esta Tesis de Doctorado, y comienza otra etapa aún por descubrir.

Ahora toca agradecer a todas las personas que me han acompañado y han hecho posible que este sueño se haga realidad.

En primer lugar quiero agradecer a uno de mis directores: Vicent Moliner Ibáñez, por permitirme estar acá, sus conocimientos, su paciencia, sus críticas, su guía, su esfuerzo, y por su apoyo en todo momento. *Sencillamente estaré eternamente agradecido.*

A mi directora Silvia Ferrer Castillo, mi "jefa" como cariñosamente siempre le he dicho, gracias por tu apoyo, especialmente en los momentos más difíciles durante estos años. *Mi agradecimiento infinito.*

Al profesor Dr. Nino Russo, por ser mi tutor durante los 3 meses de estancia en el *Theoretical and Computational Chemistry Group* de la *Università della Calabria*, en Calabria, Italia. A las profesoras Dra. Emilia Zapata y Dra. Tiziana Marino. Agradecer especialmente a Paolo Piazzetta, a Mariagrazia Fortino, a Gloria Mazzone y a Marta Alberto, por todo su cariño y hacerme sentir como en casa. *Grazie a tutti.*

Al profesor Dr. Pedro Alexandrino Fernandes, por ser mi tutor durante los 3 meses de estancia en el *Theoretical and Computational Biochemistry Group* de la *Universidade do Porto*, en Oporto, Portugal. Agradecer a todos los miembros del grupo, especialmente a Rui Neves, por toda la ayuda tanto científica como administrativa. *Obrigado a todos.*

A Merche, por su cariño y por toda la ayuda administrativa que me ha brindado durante estos años.

A los profesores del Departamento de Química Física y Analítica que me han asesorado durante la docencia universitaria impartida: Sixte Safont, Mónica Oliva, Armando Beltrán, y Raquel Castillo.

Al profesor Juan Andrés por la Beca de investigación a cargo del proyecto “Caracterización de los mecanismos moleculares de las reacciones químicas catalizadas por enzimas. Un estudio basado en el análisis topológico de la función de localización electrónica y la Teoría de catástrofes (codi: 10I415.01)”. Financiado por la fundación Caixa Castelló- Bancaixa. Duración: 01/09/2011 a 30/06/2012.

A Kasia, por el artículo que tenemos publicado, gracias por permitirme trabajar contigo, ha sido una excelente experiencia, y espero sea el primero de muchos artículos.

A los miembros de los grupos de investigación *Computational Biochemistry at UJI* (BioComp), y *Química Teórica y Computacional* (QTC): Manuel, Kasia, Raquel, Sergio, Conchin, Pato, Marisa, Javi, Lourdes, Erica, Nacho, Maite e Isabel. Ustedes han sido parte de mi familia durante todos estos años y siempre ocuparán un lugar especial en mi corazón.

## *Acknowledgments in Spanish*

---

A los nuevos miembros del grupo BioComp: Daria, Natalia y Miquel, gracias por el cariño a pesar del poco tiempo que lleváis en el grupo.

A los miembros del grupo de Materiales Moleculares de la UJI dirigido por Rosa Llusar: Iván, Eva, David, Elena, Tomás y Carmina.

Agradecer a todos los profesores del Máster Europeo en Química Teórica y Modelización Computacional por todos los conocimientos transmitidos.

A los colegas del Máster Europeo en Química Teórica y Modelización Computacional por los momentos vividos, en especial a Frank, Rodrigo, Marcos, Rafa, Kirill, Ana, David, y Alfonso.

A todos los que habéis estado de estancia durante este tiempo (y aún están): André, Natalia, Natasha, Agnieszka, Felipe (y por supuesto a la carismática Milán), Joan, Amanda, Renan y Thiago.

A Cristina y a Quique, gracias por todo.

A mis amigos cubanos dispersos por el mundo, con los cuales he pasado momentos inolvidables: Deivy, Gerardo, Yonny, Marcos, Sulayma, Rafael, Evert, Gerzán, Estrada, Bergues, Michel, Guille, Miguelito, Ceila, Raulito e Idania, Georginita y Andrea.

A los que forman parte activa de mi vida, y se han convertido en parte de mi familia: Kiri, Othniel y mi sobri Mateo, gracias por estar ahí siempre. A Griselda, Mari y Arturo por toda su ayuda. A Jorgito, gracias por todo tu apoyo y cariño. A Arturito y Yoli, a Fiffe e Isa, a Juan José y Bea, a Amestoy y Karen, y a Daniel, que aunque estén lejos, siguen estando ahí. A Marilo, gracias por sacarme siempre una sonrisa. Gracias

a todos por los momentos alegres y su apoyo en los momentos más difíciles.

A mi familia por su amor incondicional, especialmente a mis padres y a mi hermano que son únicos, a mis abuelos Olga y Manolo (nunca los olvidaré), a Loida, a mis maravillosos suegros Niurka y Angel Luis, a mis abuelos Nonina y Chiqui, a mis abuelas Manuela y Esperanza, a mis cuñadas Grisel y Sahidy, a mis tías Cari y Juana, a mi “hermana” Ana Gloria y su familia por todo el cariño, a mis primos, a mis tíos y a Nuria.

A mi esposa Grety, gracias por tu amor incondicional, por tu compañía y apoyo, por hacer de cada día una aventura, por guiarme en este camino que iniciamos hace casi 10 años (y los que faltan ☺), todo lo que soy te lo debo a ti y todo lo que hago lo hago por ti, gracias mi *princesa*. Ahora iniciamos otra etapa con mucha ilusión y responsabilidad, con la llegada de nuestro hijo Thiago, una nueva aventura que disfrutaremos al máximo cada día, ya eres una *mami* maravillosa.

Para hacer realidad los sueños, se necesita esfuerzo, sacrificio, dedicación, pero también soporte económico, así que toca el turno de agradecer a las diferentes fundaciones y organismos que han financiado esta investigación:

- Banco Santander. Becas de ayuda de matrícula para cursar estudios de Máster durante los cursos académicos 2010/2011 y 2011/2012.
- Fundación Caixa Castelló- Bancaixa. Beca de investigación a cargo del proyecto “Diseño Computacional de Inhibidores Enzimáticos y Catalizadores Biológicos (codi: 08I437.01)”. Duración: 19/01/2011 a

## Acknowledgments in Spanish

---

19/07/2011.

- Universitat Jaume I-Fundación Bancaixa. Beca de investigación a cargo del proyecto “Estudios teóricos de inhibidores de enzimas quinasas: hacia un diseño computacional de fármacos (codi: B11B2011-23)”. Duración: 01/07/2012 a 30/04/2013.
- Departamento de Química Física y Analítica de la Universidad Jaume I. Contrato de Personal Investigador a tiempo parcial. Proyecto “PROGRAMA DE GRUPS D'INVESTIGACIÓ D'ALT RENDIMENT SEGONS EL PPF” codi: 13I053. Duración: 01/11/2013 a 15/12/2013.
- Ministerio Español de Economía y Competitividad. Contrato Predoctoral a cargo del Proyecto “Simulaciones Moleculares de Procesos Catalíticos. Desarrollos y aplicaciones en Biomedicina y Biotecnología”. Referencia del Proyecto: CTQ2012-36253-C03-01. Referencia de la ayuda: BES-2013-064450. Duración: 16/12/2013 a 15/12/2017.
- Ministerio Español de Economía y Competitividad. Ayudas a la movilidad predoctoral para la realización de estancias breves en Centros de I+D, convocatorias 2014 (Referencia de la ayuda: EEBB-I-15-10361) y 2015 (Referencia de la ayuda: EEBB-I-16-11425).



## **Abstract (English Version)**

Cysteine proteases (CP) are the most abundant proteases in parasitic protozoa and they are essential enzymes to the life cycle of several of them. For example, cruzain is crucial for the development and survival of the protozoan *Trypanosoma cruzi*, the etiologic agent of Chagas disease; and falcipain-2 (FP2) is expressed during the erythrocytic stage of the life cycle of the parasite causing Malaria. Hence, these enzymes have become attractive therapeutic targets for the development of new inhibitors. Irreversible inhibitors of cruzain and FP2, that covalently bind to the enzyme through a residue cysteine of the active site, have showed good inhibition activity.

Today, many aspects of both, inhibition mechanism and catalytic mechanism of these enzymes are unclear and still under debate. Thus, it is crucial to improve our understanding of how these enzymes work at molecular level for the design of new inhibitors. In the present Doctoral Thesis, different enzymatic mechanisms of the reaction catalyzed by cruzain, as well as several inhibition processes of cruzain and FP2, have been studied. Hybrid Quantum Mechanics/Molecular Mechanical (QM/MM) potentials have been employed to run Molecular Dynamics (MD) simulations that allow obtaining the Free Energy Surfaces (FES) in terms of Potential of Mean Force (PMF). These calculations have allowed obtaining a complete picture of the possible free energy reaction paths, including details of the reaction mechanism, their corresponding free energy barriers, average geometries and the interactions between the inhibitor and the protein. A deep analysis of the results has contributed to our understanding of the enzymatic mechanisms under study, which could be used for a rational design of new inhibitors with potential application in the treatment of severe human diseases.





## **Abstract (Spanish Version)**

Las cisteína proteasas son enzimas esenciales en el ciclo de vida de muchos protozoos patogénicos. Por ejemplo, el papel de la cruzaina es decisivo para el desarrollo del *Trypanosoma cruzi*, agente etiológico de la enfermedad de Chagas, y la falcipaina-2 se expresa durante la etapa eritrocítica del parásito causante de la malaria. Sin duda, éste es el motivo por el que estas enzimas se han convertido en objetivos terapéuticos para el desarrollo de nuevos inhibidores. Inhibidores irreversibles de la cruzaina o de la falcipaina-2, que se unen covalentemente a la enzima a través de un residuo de cisteína del sitio activo, han mostrado una buena actividad de inhibición.

Sin embargo, muchos aspectos tanto del mecanismo de inhibición como del mecanismo catalítico de las cisteína proteasas son todavía una incógnita. Por tanto, mejorar la comprensión de estos sistemas enzimáticos a nivel molecular es crucial para el diseño de nuevos inhibidores. En la presente Tesis Doctoral se han estudiado diferentes mecanismos de la reacción catalizada por la cruzaina así como diversos procesos de inhibición de la cruzaina y la falcipaina-2. Se han empleado potenciales híbridos Mecánica Cuántica/Mecánica Molecular para realizar simulaciones de dinámica molecular que permiten obtener las correspondientes superficies de energía libre en términos de potenciales de fuerza media. Estos cálculos nos han permitido obtener un panorama completo de los posibles caminos de reacción de energía libre, incluyendo detalles del mecanismo de reacción, las geometrías promedio de los estados estacionarios, además de las interacciones entre el inhibidor y la proteína. Un análisis profundo de los resultados ha contribuido a nuestra comprensión de estos procesos, lo cual podría ser utilizado para un diseño racional de nuevos inhibidores con potencial aplicación en el tratamiento de estas enfermedades.



The present Doctoral Thesis is based on the following publications:

**Paper I.** Kemel Arafet, Silvia Ferrer, Sergio Martí, and Vicent Moliner. *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Falcipain-2 Inhibition by the Epoxysuccinate E64*. *Biochemistry*. **2014**. 53 (20), 3336-3346. DOI: 10.1021/bi500060h.

**Paper II.** Kemel Arafet, Silvia Ferrer, and Vicent Moliner. *First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones*. *Biochemistry*. **2015**. 54 (21), 3381-3391. DOI: 10.1021/bi501551g.

**Paper III.** Kemel Arafet, Silvia Ferrer, Florenci V. González, and Vicent Moliner. *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Cysteine Protease Inhibition by Peptidyl-2,3-epoxyketones*. *Physical Chemistry Chemical Physics*. **2017**. 19. 12740-12748. DOI: 10.1039/C7CP01726J.

**Paper IV.** Kemel Arafet, Silvia Ferrer, and Vicent Moliner. *Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease*. *ACS Catalysis*. **2017**. 7, 1207-1215. DOI: 10.1021/acscatal.6b03096.

The following paper, not directly related with the present Doctoral Thesis, was also published:

**Paper V.** Katarzyna Świderek, Kemel Arafet, Amnon Kohen, and Vicent Moliner. *Benchmarking Quantum Mechanics/Molecular Mechanics (QM/MM) Methods on the Thymidylate Synthase-Catalyzed Hydride Transfer*. *Journal of Chemical Theory and Computation*. **2017**. 13, 1375-1388. DOI: 10.1021/acs.jctc.6b01032.

## **Contribution Report**

I have been responsible for the computations and to the writing of the Introduction and Computation Model sections of all the papers related with the Doctoral Thesis. I have contributed in the discussion, analyses, and corrections of the manuscripts of all the papers related with the Doctoral Thesis.

In the case of the paper V, I have performed some of the computations, and I have contributed in the corrections of the manuscripts.

## Contents

<b>Acknowledgements in Spanish</b> .....	vii
<b>Abstract (English Version)</b> .....	xiii
<b>Abstract (Spanish Version)</b> .....	xv
<b>List of Acronyms and Abbreviations</b> .....	xxi
<b>1. Introduction</b> .....	<b>1</b>
1.1. Introduction (English Version).....	1
1.2. Introduction (Spanish Version).....	7
<b>2. Objectives</b> .....	<b>15</b>
<b>3. Enzyme Kinetics and Catalysis</b> .....	<b>17</b>
3.1. Enzyme Kinetics.....	17
3.2. Enzyme Catalysis.....	19
<b>4. Enzyme Inhibition</b> .....	<b>25</b>
4.1. Reversible Inhibitors.....	25
4.2. Irreversible Inhibitors.....	28
4.3. Transition State Analogues Inhibitors.....	29
4.4. Some Parameters to Report Inhibition Data.....	30
<b>5. Cysteine Proteases</b> .....	<b>33</b>
5.1. Catalytic Mechanism of the Cysteine Proteases.....	34
5.2. Cruzain and Chagas Disease.....	38
5.3. Falcipain-2 and Malaria.....	41
5.4. Inhibition Mechanism of Falcipain-2 by the Epoxysuccinate E64.....	44
5.5. Inhibition Mechanism of Cruzain by Halomethyl Ketones and Epoxyketones.....	47
<b>6. Theoretical Methods</b> .....	<b>53</b>
6.1. QM/MM Hybrid Methodology.....	53
6.1.1 QM Methods.....	60

---

6.2. Potential Energy Surfaces .....	62
6.3. Free Energy .....	67
6.3.1. Molecular Dynamics .....	68
6.3.2. Umbrella Sampling .....	73
6.3.3. Transition State Theory .....	77
<b>7. Papers .....</b>	<b>81</b>
7.1. Paper I .....	81
7.1.1. Supporting Information .....	93
7.2. Paper II .....	101
7.2.1. Supporting Information .....	113
7.3. Paper III .....	121
7.3.1 Supporting Information .....	131
7.4. Paper IV .....	140
7.4.1. Supporting Information .....	150
<b>8. Conclusions and Outlook .....</b>	<b>159</b>
8.1 Conclusions .....	159
8.2 Outlook .....	162
<b>References.....</b>	<b>165</b>
<b>Appendices .....</b>	<b>187</b>
Appendix A: Others Scientific Contributions Related with the Doctoral Thesis .....	187
Poster Communications.....	187
Appendix B: Scientific Contributions not Directly Related with the Doctoral Thesis .....	188
Paper V .....	188

## **List of Acronyms and Abbreviations**

- AMBER** Assisted Model Building and Energy Refinement
- AM1** Austin Model 1
- BFGS** Broyden-Fletcher-Goldfarb-Shanno
- CHARMM** Chemistry at Harvard Macromolecular Mechanics
- CP/CPs** Cysteine protease/Cysteine proteases
- DFT** Density Functional Theory
- DHAM** Dynamic Histogram Analysis Method
- E** Enzyme
- EI** Enzyme-inhibitor
- ES** Enzyme-substrate
- ESI** Enzyme-substrate-inhibitor
- EVB** Empirical valence bond
- E64** L-trans-epoxysuccinyl-leucyl-amido[4-guanidino]butane
- FP2** Falcipain-2/Falcipaina-2
- FE** Free energy
- FES** Free energy surface
- HF** Hartree-Fock
- Hph** Homophenylalanine
- L-BFGS** Limited-memory Broyden-Fletcher-Goldfarb-Shanno
- MC** Michaelis-Menten's complex
- MCS** Michaelis-Menten's complex in solution
- MD** Molecular Dynamics
- MM** Molecular Mechanics
- MNDO** Modified Neglect of Diatomic Overlap
- NDDO** Neglect of Differential Diatomic Overlap
- NMA** N-methyl-acetamide

## *List of Acronyms and Abbreviations*

---

**ONIOM** Our own N-layered Integrated Molecular Orbital and Molecular Mechanics

**OPLS** Optimized Potential for Liquid Simulations

**OPLS-AA** Optimized Potential for Liquid Simulations All Atoms

**P** Products

**PBC** Periodic boundary conditions

**PDB** Protein Data Bank

**PEK** Peptidyl-2,3-epoxyketones/ peptidyl epoxy ketones

**PES** Potential Energy Surfaces

**PCIK** Peptidyl chloromethyl ketones

**PFK** Peptidyl fluoromethyl ketones

**PHK** Peptidyl halomethyl ketones

**PMF** Potential of Mean Force

**PM3** Parametric Method 3

**QM** Quantum Mechanics

**QM/MM** Quantum Mechanics/Molecular Mechanics

**R** Reactants

**TIP3P** Transferable Intermolecular Potential 3P

**TMSI** Three-membered sulfonium intermediate

**TS** Transition state

**TSA** Transition state analogue

**TST** Transition State Theory

**WHAM** Weighted Histogram Analysis Method



# 1. Introduction

## 1.1. Introduction (English Version)

Enzymes are complex biological molecules that allow organisms to reduce the timescales of the chemical reactions, in some cases from millions of years to fractions of seconds. The extraordinary catalytic power of the enzymes, often far greater than that of synthetic or inorganic catalysts, is crucial for the life, without this ability, the life would be impossible. In addition, they function in aqueous solutions under very mild conditions of temperature and pH.<sup>1</sup> For example, in aqueous solution the hydrolysis of a peptide bond would take hundreds of years, while a protease can degrade as many as one million of peptide bonds per second.<sup>2</sup>

In 1878, Frederik W. Kühne used for the first time the word enzyme, and it means “*in yeast*”,<sup>3</sup> what intended to emphasize that the catalytic activity was the manifestation of an extract or secretion rather than of the whole organism. Years later, James B. Sumner isolated an enzyme in its crystal form for the first time, the enzyme urease from the jack bean *Canavalia ensiformis*.<sup>4</sup> In 1960, Stein, Moore and Hirs have determined the first amino acid sequence of an enzyme, corresponding to the complete sequence of 124 amino acid residues in oxidized bovine pancreatic ribonuclease through a structural study of the peptides liberated by enzymatic cleavage.<sup>5</sup> William H. Stein and Stanford Moore in 1972 shared the Nobel Prize in Chemistry “*for their contribution to the understanding of the connection between chemical structure and catalytic activity of the active centre of the ribonuclease molecule*”.<sup>6</sup> Phillips and co-workers in 1965 solved the first X-ray structure of an

enzyme, the hen egg-white lysozyme.<sup>7</sup> This work allowed an enzyme structure to be examined at atomic resolution, giving birth to several studies about of the structure-function of enzymes. Four years later, B. Gutte and R. B. Merrifield reported the first synthesis of an enzyme from the component amino acids, the ribonuclease A.<sup>8</sup> In 1971, was established the first open access digital resource for sharing three-dimensional protein structures, the Protein Data Bank (PDB).<sup>9</sup> Initially, the PDB only contained 7 protein structures, today, that number has grown exponentially to almost 130,000 entries of natural and designed macromolecules. Thus, in the following years, with the corresponding experimental and computational advances, several studies were carried out in order to understanding the behaviour of the enzymes, however, today many aspects are still under debate in the enzymology.

It is clear that the study of these macromolecules has a practical importance. A deficiency or a total absence of one or more enzymes could cause inheritable genetic disorders. The malfunction of just one type of enzyme in the human body can be fatal. The deficiency of CP, for example, may cause many diseases such as cancer<sup>10</sup> and Alzheimer's disease.<sup>11</sup> On the other hand, some diseases may be caused by excessive activity of an enzyme; for example, excess activity of both CPs cathepsin-B and cathepsin-L results in the degradation of muscle tissue and intracellular proteolysis.<sup>12</sup> Thus, **enzymes become attractive targets in medicinal chemistry**: from the known human proteases, almost 14% are under investigation as drug targets for several diseases,<sup>13</sup> and about 5-10% of all biological target for development of drugs are proteases.<sup>14</sup> The study of the enzymes is also important in chemical engineering, food technology, and agriculture.<sup>1</sup> For example, the protease applications in

## *1. Introduction*

---

industrial processes present several advantages compared to chemical processes, due to increasing hydrolysis specificity, product preservation and purity, and reducing environmental impact.<sup>15</sup> Proteases are also important tools of the biotechnological industry because of their usefulness as biochemical reagents or in the manufacture of numerous products.<sup>16</sup>

Regarding the particular implication of CPs in developing new drugs, it can be mentioned that for example, cruzain and FP2 are drug targets against Chagas disease and malaria.<sup>17,18</sup> The first reported case of Chagas disease was in 1835,<sup>19</sup> and currently affects about 6-7 million of people living mainly in endemic areas of 21 Latin American countries. There is not available vaccine to prevent this disease and the two drugs for treatments are toxic, with important contra-indications and also ineffective for the chronic stage of the disease.<sup>20,21</sup> Another of the most serious and neglected diseases in the world is malaria, although that the global malaria death rate declined by 60%, the disease continues to have a negative impact on the public health. The inhibition of FP2 has proven indispensable in order to completely block parasitic growth and proliferation.<sup>22,23</sup> Consequently, there is a need to develop efficient therapies against these diseases. One promising approach is the development of inhibitors of both CPs cruzain and FP2. Irreversible inhibitors that contain an electrophilic functional group, such as halomethyl ketones, epoxy ketones, or epoxysuccinates, covalently bind to the enzyme through a residue cysteine via nucleophilic attack of the active site, showing good inhibition activity.<sup>24</sup> Today, many aspects of both inhibition mechanism and catalytic mechanism of CPs are unclear and are still under debate. However, despite the importance of the

problem, only few computational studies of these mechanisms including the protein environment effects have been reported. Thus, it is crucial to improve our understanding of these enzymatic mechanisms at molecular level to design new inhibitors that allow an effective treatment of these diseases.

The complexity of the enzymes and their reactions make the experiments a limited tool for the study of the enzyme activity.<sup>25</sup> Computational chemistry has become an invaluable tool for studying the activity of the enzymes, through which we can understand their fundamental mechanisms from studying transition states, structures impossible to study with experimental techniques due to their short lifetimes.<sup>25,26</sup> All the improvements obtained in the last decades are applied in several fields such as medicine, biology, chemistry, material sciences or physics, highlighting an important feature of the computational and theoretical chemistry. On the other hand, it is important to emphasize that the development in computational chemistry depends especially on the increase in computer power.

Nowadays, a useful and powerful computational tool to study the enzymatic behavior is the hybrid QM/MM methodology.<sup>27-29</sup> This takes advantage of the accuracy of the QM methods and the low computational cost of the MM methods, being a good balanced choice between both factors. The idea of the QM/MM methodology was introduced by Karplus, Warshel and Levitt.<sup>27-29</sup> In the last decades, the number of papers related with QM/MM studies of enzymes has grown considerably thus showing the success and usefulness of this technique. The award of the Chemistry Nobel Prize in 2013 to Martin Karplus, Michael Levitt and

## ***1. Introduction***

---

Arieh Warshel “*for the development of multiscale models for complex chemical systems*”<sup>30</sup> recognizes the important role of the QM/MM methodology for the study and understanding of the biological systems.

In the present Doctoral Thesis, different CPs inhibition mechanisms have been studied by MD simulations using hybrid AM1d/MM and M06-2X/6-31+G(d,p):AM1d/MM potentials for describing the atoms of the QM region, while the rest of the protein and water molecules were described by OPLS-AA<sup>31</sup> and TIP3P<sup>32</sup> force fields. The first contribution of this Thesis was the computational study of the inhibition mechanism of FP2 by the epoxysuccinate E64 (see Paper I in Chapter 7). FESs generated in term of PMF were explored for the two possible mechanisms, the sulfur attack of conserved residue Cys42 on C2 or C3 atoms of the epoxide ring. An alternative mechanism consisting of a protonation of the epoxy O3 atom prior to the Cys42 attack and the epoxy ring opening has been also explored. Three possible sources of protons have been considered: a water molecule, protonated His174, and a direct intramolecular transfer of a proton from the carboxylic group of the inhibitor. In addition, an analysis of the averaged geometries, averaged Mulliken charges and the interactions between the FP2 and the E64 were carried out for the key stationary points.

The second contribution of the present Doctoral Thesis (see Paper II in Chapter 7) was the computational study of the inhibition mechanism of cruzain by the peptidyl halomethyl ketones (PHK) Bz-Tyr-Ala-CH<sub>2</sub>F (PFK) and Bz-Tyr-Ala-CH<sub>2</sub>Cl (PCIK). Three different mechanisms proposed for the Cys25 attack to both inhibitors have been initially explored by generating FESs in term of PMFs. Later an energetic comparison between the mechanisms proposed for the inhibitors PFK

and PCIK has been established, as well as the detailed analysis of the averaged geometries of the key stationary points, and interactions between the inhibitor (PFK or PCIK) and the cruzain.

The last study of the CP inhibition mechanism was focused on the inhibition of cruzain by the epoxyketone Cbz-Phe-Hph-(2S) (see Paper III in Chapter 7). FESs were explored for all the possible mechanism proposed in the literature: the stepwise mechanism where Cys25 attacks the epoxy ring before the activation of the ring by a proton transfer, and the concerted mechanism or the stepwise mechanism where the activation of the epoxy ring precedes the Cys25 attack. A protonated His159 and a water molecule have been considered as possible source of protons. For the proposed mechanism, the average values of interatomic distances and interactions in the key states located along the reaction of both attacks of Cys25 on the C2 and C3 atoms have been subject of a deep analysis.

As mentioned above, diverse aspects around the catalytic mechanism of CP are under discussion. Is the acylation stage a serine protease-like mechanism? Is the deacylation concerted, or is it a stepwise process? In order to answers all these questions, the catalytic mechanism of the CP cruzain (see Paper IV in Chapter 7) has been studied for the hydrolysis of a small peptide. Different FESs were explored for all the possible chemical steps corresponding to the two stages of the mechanism: acylation and deacylation. The activation and reaction free energies for all of the possible chemical steps have been analysis in order to propose the most favorable one for the acylation stage and for deacylation stage. As in previous studies, analysis of the averaged

## ***1. Introduction***

---

geometries, interactions distances and interaction energies between the peptide and the protein for the proposed stages has been carried out.

### **1.2. Introduction (Spanish Version)**

Las enzimas son moléculas biológicas complejas que permiten a los organismos reducir las escalas de tiempo de las reacciones químicas, en algunos casos desde millones de años hasta fracciones de segundos. El extraordinario poder catalítico de las enzimas, a menudo mucho mayor que el de los catalizadores sintéticos o inorgánicos, es crucial para la vida. Sin esta capacidad, la vida sería imposible.<sup>1</sup> Por ejemplo, en disolución acuosa la hidrólisis de un enlace peptídico llevaría cientos de años, sin embargo, una proteasa puede degradar hasta un millón de enlaces peptídicos por segundo.<sup>2</sup>

En 1878, Frederik W. Kühne usó por primera vez la palabra enzima, la cual significa "en levadura",<sup>3</sup> lo que pretendía enfatizar que la actividad catalítica era la manifestación de un extracto o secreción más que de todo el organismo. Años más tarde, James B. Sumner aisló por primera vez una enzima en su forma cristalina, la enzima ureasa de *Canavalia ensiformis*.<sup>4</sup> En 1960, Stein, Moore y Hirs determinaron la primera secuencia de aminoácidos de una enzima.<sup>5</sup> William H. Stein y Stanford Moore en 1972 compartieron el Premio Nobel de Química "por su contribución a la comprensión de la conexión entre la estructura química y la actividad catalítica del centro activo de la molécula de ribonucleasa".<sup>6</sup> Phillips y colaboradores, en 1965, resolvieron la primera estructura de rayos X de una enzima, la lisozima de la clara de huevo de la gallina.<sup>7</sup> Este trabajo permitió examinar la estructura de una enzima a

niveles atómicos, dando lugar a diversos estudios sobre la estructura-función de las enzimas. Cuatro años después, B. Gutte y R. B. Merrifield reportaron la primera síntesis de una enzima, la ribonucleasa A, a partir de sus aminoácidos.<sup>8</sup> En 1971, se estableció el primer recurso digital de acceso abierto para compartir estructuras tridimensionales de proteínas, el PDB (Banco de datos de proteínas, de sus siglas en inglés).<sup>9</sup> Inicialmente, el PDB sólo contenía 7 estructuras de proteínas. Hoy, ese número ha crecido exponencialmente a casi 130.000 entradas de macromoléculas naturales y de diseño. Así, en los años siguientes, con los correspondientes avances experimentales y computacionales, se realizaron varios estudios con el fin de comprender el comportamiento de las enzimas, sin embargo, hoy en día muchos aspectos están todavía bajo debate en el campo de la enzimología.

Es evidente que el estudio de estas macromoléculas tiene una gran importancia práctica. La deficiencia o la ausencia total de una o más enzimas puede causar trastornos genéticos hereditarios. El mal funcionamiento de un solo tipo de enzima en el cuerpo humano puede ser fatal. La deficiencia de CPs (cisteína-proteasas, de sus siglas en inglés). Por ejemplo, puede causar muchas enfermedades como el cáncer<sup>10</sup> y el Alzheimer.<sup>11</sup> Por otro lado, algunas enfermedades pueden ser causadas por la actividad excesiva de una enzima; por ejemplo, el exceso de actividad de ambas CPs catepsina-B y catepsina-L da lugar a la degradación del tejido muscular y a la proteólisis intracelular.<sup>12</sup> Es sin duda por ello que **las enzimas se han convertido en objetivos terapéuticos**: de las proteasas humanas conocidas, casi el 14% están bajo investigación como dianas de fármacos para el tratamiento de varias enfermedades<sup>13</sup> y alrededor del 5-10% de todas las dianas biológicas para



## *1. Introduction*

---

el desarrollo de fármacos son proteasas.<sup>14</sup> El estudio de las enzimas también es importante en ingeniería química, tecnología de alimentos y agricultura.<sup>1</sup> Por ejemplo, las aplicaciones en procesos industriales de las proteasas presentan varias ventajas en comparación con los procesos químicos, debido al aumento de la especificidad de la hidrólisis, la preservación y pureza del producto y la reducción del impacto ambiental.<sup>15</sup> Las proteasas son también importantes de la industria biotecnológica debido a su utilidad como reactivos bioquímicos o en la fabricación de numerosos productos.<sup>16</sup>

Retomando la implicación de las CPs en el desarrollo de nuevos fármacos podríamos mencionar, por ejemplo, la cruzaina y la FP2 (falcipaína-2, de sus siglas en inglés) son objetivos para el desarrollo de fármacos contra la enfermedad de Chagas y la malaria.<sup>17,18</sup> El primer caso que se conoce de la enfermedad de Chagas data de 1835,<sup>19</sup> y actualmente afecta a aproximadamente 6-7 millones de personas que viven principalmente en áreas endémicas de 21 países latinoamericanos. No existe una vacuna disponible para prevenir esta enfermedad y los dos fármacos existentes para el tratamiento, además de ser tóxicos, presentan contraindicaciones importantes y son ineficaces para la fase crónica de la enfermedad.<sup>20,21</sup> Otra de las enfermedades más graves y olvidadas en el mundo es la malaria, a pesar de que la tasa global de mortalidad por malaria disminuyó en un 60%, la enfermedad continúa teniendo un impacto negativo en la salud pública. La inhibición del FP2 ha demostrado ser indispensable para bloquear completamente el crecimiento y la proliferación del parásito causante de la malaria.<sup>22,23</sup> Consecuentemente, existe la necesidad de desarrollar terapias eficaces contra estas enfermedades. Un enfoque prometedor es el desarrollo de

inhibidores de las CPs cruzaña y FP2. Los inhibidores irreversibles que contienen un grupo funcional electrofílico, tales como halometil cetonas, epoxi cetonas o epoxisuccinatos, se unen covalentemente a la enzima a través de un residuo de cisteína mediante ataque nucleofílico del sitio activo, mostrando una buena actividad de inhibición.<sup>24</sup> Hoy en día, muchos aspectos tanto del mecanismo de inhibición como del mecanismo catalítico de las CPs no están claros y todavía están en discusión, sin embargo, se han realizado pocos estudios computacionales de estos mecanismos, incluyendo los efectos del ambiente proteico. Por lo tanto, mejorar nuestra comprensión de estos mecanismos enzimáticos a nivel molecular es crucial para el diseño de nuevos inhibidores que permitan el tratamiento eficaz de estas enfermedades.

La complejidad de las enzimas y sus reacciones hace de los experimentos una herramienta limitada para el estudio de su actividad.<sup>25</sup> La química computacional se ha convertido en una potente herramienta para estudiar la actividad de las enzimas, a través de la cual podemos lograr comprender a niveles moleculares los mecanismos fundamentales de la catálisis enzimática caracterizando los estados de transición de las reacciones que catalizan. Estos vienen determinados por estructuras imposibles de estudiar directamente con técnicas experimentales debido a su corto tiempo de vida medio.<sup>25,26</sup> Todos los avances obtenidos en las últimas décadas se aplican en varios campos como la medicina, biología, química, ciencias de los materiales o la física. Por otra parte, es importante enfatizar que el desarrollo de la química computacional depende especialmente del aumento de la potencia de los ordenadores.

## *1. Introduction*

---

Hoy en día, una herramienta computacional útil y potente para estudiar el comportamiento enzimático es la metodología híbrida QM/MM (Mecánica Cuántica/Mecánica Molecular, de sus siglas en inglés);<sup>27-29</sup> la cual aprovecha la precisión de los métodos QM (Mecánica Cuántica, de sus siglas en inglés) y el bajo coste computacional de los métodos MM (Mecánica Molecular, de sus siglas en inglés), logrando un compromiso adecuado entre el grado de precisión química deseado y el coste computacional. La idea de la metodología QM/MM fue introducida por Karplus, Warshel y Levitt.<sup>27-29</sup> En las últimas décadas, el número de artículos relacionados con las enzimas aplicando la metodología QM/MM ha crecido considerablemente lo cual demuestra su utilidad. El premio Nobel de Química en 2013 a Martin Karplus, Michael Levitt y Arieh Warshel "*por el desarrollo de modelos multiescalas para sistemas químicos complejos*"<sup>30</sup> reconoce el papel importante de esta metodología para el estudiar y comprender los sistemas biológicos.

En la presente Tesis Doctoral, se han estudiado diferentes mecanismos de inhibición mediante simulaciones de MD utilizando los potenciales híbridos AM1d/MM y M06-2X/6-31+G(d,p):AM1d/MM para describir los átomos de la región QM, mientras que el resto de la proteína y las moléculas de agua fueron descritas por los campos de fuerza OPLS-AA<sup>31</sup> y TIP3P.<sup>32</sup> La primera contribución de esta Tesis fue el estudio computacional del mecanismo de inhibición de la FP2 por el epoxisuccinato E64 (véase el Paper I en el Capítulo 7). Se exploraron las FES generadas en términos de los PMF para los dos posibles mecanismos, el ataque del residuo Cys42 sobre los átomos C2 o C3 del anillo epóxido. También se exploró un mecanismo alternativo consistente en una protonación del átomo O3 del anillo epóxido antes del ataque del

residuo Cys42 y la abertura del anillo epóxido. Se consideró tres posibles fuentes de protones: una molécula de agua, el residuo protonado His174, y una transferencia intramolecular directa de un protón del grupo carboxílico del inhibidor. Además, se realizó un análisis de las geometrías promediadas, las cargas promedio de Mulliken y las interacciones entre el FP2 y el E64 para los puntos estacionarios localizados sobre las FESs.

La segunda contribución de la presente Tesis Doctoral (véase el Paper II en el Capítulo 7) fue el estudio computacional del mecanismo de inhibición de la cruzaína por los inhibidores irreversibles PFK y PCIK. Tres mecanismos diferentes propuestos para el ataque de Cys25 a ambos inhibidores se han explorado inicialmente a través de las FES generadas en términos de los PMFs. Posteriormente se compararon los mecanismos propuestos para los inhibidores PFK y PCIK, sus energías libres, las geometrías promediadas de los puntos estacionarios y las interacciones entre el inhibidor (PFK o PCIK) y la cruzaína.

El último estudio del mecanismo de inhibición de una CP se centró en la inhibición de la cruzaína por la epoxi cetona Cbz-Phe-Hph-(2S) (véase el Paper III en el Capítulo 7). Las FESs se exploraron para todos los mecanismos posibles propuestos en la literatura: el mecanismo por etapas donde el residuo Cys25 ataca el anillo epóxido antes de la activación del anillo por una transferencia de protones, y el mecanismo concertado o el mecanismo por etapas donde la activación del anillo epoxi precede al ataque del residuo Cys25. El residuo His159 protonado y una molécula de agua se han considerado como posible fuente de protones. Para el mecanismo propuesto, se analizó los valores promedios

## ***1. Introduction***

---

de las distancias interatómicas y las interacciones en los estados estacionarios clave del mecanismo de reacción del ataque del residuo Cys25 sobre los átomos C2 y C3.

Tal como se ha indicado anteriormente, diversos aspectos relacionados con el mecanismo catalítico de la CP están siendo discutidos en la comunidad científica. ¿La etapa de acilación ocurre a través de un mecanismo similar al de las serina-proteasas? ¿La deacilación es una etapa concertada, u ocurre en varias etapas? Con el fin de responder a todas estas preguntas, el mecanismo catalítico de la cisteína proteasa cruzaña en la ruptura de un péptido se ha estudiado mediante simulaciones de MD aplicando los potenciales híbridos AM1d/MM y M06-2X/6-31+G(d,p):AM1d/MM (véase el Paper IV en el Capítulo 7). Se exploraron diferentes FESs para todas las posibles rutas químicas correspondientes a las dos etapas del mecanismo: acilación y deacilación. Las barreras de energía libre y la energía libre de reacción para todas las posibles etapas químicas han sido analizadas para proponer la etapa más favorable en la acilación y en la deacilación. Para las etapas propuestas se llevó a cabo un análisis de las geometrías promedios, las interacciones promedios y las energías de interacción entre el péptido y la cruzaña.



## 2. Objectives

The main purpose of the present Doctoral Thesis is to progress in the understanding of the mechanisms of catalysis and inhibition of two CPs, cruzain and FP2. Computational Chemistry techniques, based on hybrid QM/MM methods will be applied. FES, derived from hybrid MD simulations will provide a detailed description of the studied processes at molecular level. In particular, the objectives that have been raised and addressed are the following:

- to study the mechanism of FP2 inhibition by the epoxysuccinate E64.
- to study the mechanism of cruzain inhibition by the peptidyl halomethyl ketones Bz-Tyr-Ala-CH<sub>2</sub>F and Bz-Tyr-Ala-CH<sub>2</sub>Cl.
- to study the inhibition mechanism of cruzain by the peptidyl epoxyketone Cbz-Phe-Hph-(2S).
- to study the catalytic mechanism of the cruzain cysteine protease.

In all the aforementioned studied systems, a complete analysis of the role of each residue in the active site of the corresponding enzymes was performed with the final objective of understanding the molecular mechanism that could open the door to the design of more selective and potent inhibitors to be applied in biomedicine.





## 3. Enzyme Kinetics and Catalysis

### 3.1. Enzyme Kinetics

Since nineteenth century, scientists have tried to understand the catalytic power of the enzymes. In 1835, the chemist Jöns Jacob Berzelius (1779-1848) was the first that use the term *catalytic power*, referring to the *ability of substances to awaken affinities, which are asleep at a particular temperature, by their mere presence and not by their own affinity*.<sup>33</sup>

Later on, about 1850, Louis Pasteur (1822-1895) suggested that the fermentation process of sugar into alcohol by yeast is catalyzed by ferments that were inseparable from the structure of living yeast cell. However, in years later, Eduard Buchner (1860-1917) demonstrated that molecules which promote the fermentation continue their function after being removed from the living yeast cell.<sup>1</sup> Buchner was awarded the Nobel Prize in Chemistry in 1907 “*for his biochemical research and his discovery of cell-free fermentation*”.<sup>34</sup>

The first step in a reaction catalyzed by an enzyme is the manner in which the substrate to combine with the enzyme with the corresponding formation of the enzyme-substrate complex. In this sense, in 1894 Emil Fischer (1852-1919) proposed, the *lock and key model*,<sup>35</sup> in this model, the active site is a rigid and unchanging form structure and the substrate fits it. The rigidity of the active site is a limitation of this theory. Later, Daniel Edward Koshland Jr. (1920-2007) proposed a more adequate model, the *induced-fit model*,<sup>36</sup> since the enzyme is rather a flexible structure, the active site changes its conformation slightly by the

presence of the substrate due to the interactions between the amino acids of the active site and the substrate.

In 1913, Leonor Michaelis (1875-1949) and Maud Menten (1879-1960),<sup>37</sup> suggested a new theory about the kinetic of the reaction catalyzed by the enzymes, this theory was further developed by George Edward Briggs (1893-1985) and John Burdon Sanderson Haldane (1892-1964).<sup>38</sup>



**Scheme 3.1.1** Kinetic pattern for the enzymatic reaction mechanism proposed by Michaelis and Menten.

According to the Scheme 3.1.1, the enzyme (E) first bind to the substrate in a reversible way to form the enzyme-substrate (ES) complex, known as the Michaelis-Menten's complex (MC). Then, this ES complex breaks down in a slow step to give products (P), and the enzyme is released. The rate of the reaction taking into account the irreversible step can be expressed as:

$$v_{cat} = k_{cat}[ES] \quad (3.1.1)$$

Haldane and Briggs developed the steady-state approximation,<sup>38</sup> they assume that the concentration of the ES complex remains constant, due to its rate of formation is equal to its rate of breakdown, thus the equation 3.1.1 can be written as:<sup>39</sup>

### 3. Enzyme Kinetics and Catalysis

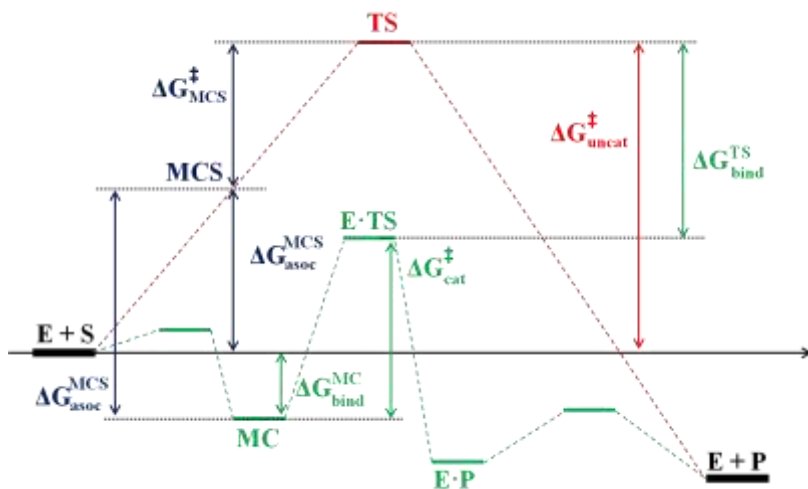
---

$$v_{\text{cat}} = \frac{v_{\text{max}} [S]}{K_M + [S]} \quad (3.1.2)$$

The equation 3.1.2 is known as the Michaelis-Menten equation and represents the reaction rate for a one-substrate reaction catalyzed by enzymes.  $K_M$  is the Michaelis-Menten constant, which characterizes the thermodynamics of the formation of the ES complex and represents the concentration of substrate at which the reaction rate is half of the maximum reaction rate, it is independent of the total amount of E and it gives a numerical value to the affinity of the enzyme. The maximum reaction rate ( $v_{\text{max}}$ ) occurs at a higher concentration of the substrate and when the enzyme is saturated with the substrate.

### 3.2. Enzyme Catalysis

In 1946, Linus Carl Pauling (1901-1994) proposed that the enzymes catalyze the reactions by binding to the transition state (TS). An enzyme may bind to the TS of the reaction that catalyzes with greater affinity than to its substrates or products. Thus, the active site of the enzyme is complementary to the TS, and not to the substrate in its fundamental state.<sup>40</sup> This prediction was satisfied by the first enzyme structure solved by Phillips and co-workers in 1965,<sup>7</sup> and as Warshel and Levitt later demonstrated in a theoretical study of the reaction of lysozyme.<sup>28</sup> The structure showed the transition state for glycoside cleavage to be stabilized by the enzyme: the strong interaction between the carboxylate of the residue Asp52 (lysozyme numbering) and the developing positive charge on the carbonium ion is about  $9 \text{ kcal}\cdot\text{mol}^{-1}$  more favorable than in the ground state.



**Figure 3.2.1** Free energy diagram of a reaction catalyzed by an enzyme (dashed line in green color) compared to the non-catalyzed reaction (dashed line in red color). Figure adapted from reference 41.

A comparison of the energetic profile of both enzyme-catalyzed and non-catalyzed reactions is shown in Figure 3.2.1;<sup>41</sup> the former would correspond to the kinetic pattern represented in the Scheme 3.1. The non-catalyzed reaction occurs in just one step, with an activation free energy of  $\Delta G_{\text{uncat}}^{\ddagger}$ . In contrast, the reaction catalyzed by the enzyme, consists of two steps: first, the MC is formed, losing binding energy,  $\Delta G_{\text{bind}}^{\text{MC}}$ . Later, the chemical reaction takes place itself, with an energetic cost of  $\Delta G_{\text{cat}}^{\ddagger}$ . This value is smaller than  $\Delta G_{\text{uncat}}^{\ddagger}$  and it can be related to the reaction rate,  $k_{\text{cat}}$ , by the Transition State Theory (TST, the section 6.3.3 is devoted to this theory in more detail), as showed in the equation 3.2.1.

### 3. Enzyme Kinetics and Catalysis

---

$$k_{\text{cat}}(T) = \frac{k_B T}{h} e^{\left(\frac{-\Delta G(T)_{\text{cat}}^{\ddagger}}{RT}\right)} \quad (3.2.1)$$

The catalytic power of an enzyme is usually expressed by the ratio  $k_{\text{cat}}/k_{\text{uncat}}$ , this means, it is proportional to the difference between the activation free energy of the non-catalyzed reaction and of the catalyzed one. According to Figure 3.2.1:

$$-\Delta G_{\text{bind}}^{\text{TS}} + \Delta G_{\text{cat}}^{\ddagger} = -\Delta G_{\text{bind}}^{\text{MC}} + \Delta G_{\text{uncat}}^{\ddagger} \quad (3.2.2)$$

$$\Delta G_{\text{uncat}}^{\ddagger} - \Delta G_{\text{cat}}^{\ddagger} = \Delta G_{\text{bind}}^{\text{MC}} - \Delta G_{\text{bind}}^{\text{TS}} \quad (3.2.3)$$

If the enzyme accelerates the rate of the reaction, the left term of equation 3.2.3 should be positive, so  $\Delta G_{\text{bind}}^{\text{TS}}$  should be greater, in its absolute value, than  $\Delta G_{\text{bind}}^{\text{MC}}$ . This means, the enzyme show a higher affinity for the TS than for the MC complex, as proposed Pauling in 1946,<sup>40</sup> and as it has been demonstrated by Warshel and co-workers in several theoretical studies about different enzymatic mechanism,<sup>28,42-44</sup> thus, the catalytic power of the enzymes lies in the stabilization of the TS rather than the MC.

On the other hand, a second interpretation is focusing in the MC. The term  $\Delta G_{\text{bind}}^{\text{MC}}$  may contain an important energy contribution based on a change in the substrate geometry when moving from one reactant structure in which the reacting groups are separated and fully solvated, to another reactant structure where the reacting groups are in close contact and are in a proper orientation to react. If it is define a similar structure to the MC, but in solution, MCS, the binding free energy of the MC can be

expressed as a contribution of two terms, the free energy needed to arrive from reactants in solution to the MCS structure,  $\Delta G_{\text{asoc}}^{\text{MCS}}$ , and the binding free energy of the MCS complex,  $\Delta G_{\text{bind}}^{\text{MCS}}$ , obtaining this equation:

$$\Delta G_{\text{bind}}^{\text{MC}} = \Delta G_{\text{asoc}}^{\text{MCS}} + \Delta G_{\text{bind}}^{\text{MCS}} \quad (3.2.4)$$

Replacing the equation 3.2.4 in equation 3.2.3:

$$\Delta G_{\text{uncat}}^{\ddagger} - \Delta G_{\text{cat}}^{\ddagger} = \Delta G_{\text{asoc}}^{\text{MCS}} + \Delta G_{\text{bind}}^{\text{MCS}} - \Delta G_{\text{bind}}^{\text{TS}} \quad (3.2.5)$$

Since  $\Delta G_{\text{asoc}}^{\text{MCS}}$  is always positive, a high catalytic power can be obtained even when the enzyme shows greater affinity to the MC than to the TS.

Apart of both of the main theories proposed to explain the origin of enzyme catalysis, the TS-theories and the MC-theories, several additional factors are derived from the MC-theories, as for example the concept of the near-attack conformation<sup>45-48</sup> and the entropic factors (entropic trap).<sup>49-53</sup> Page and Jencks introduced the concept of entropic trap,<sup>49</sup> a loss of entropy upon substrate binding decreases the activation entropy for the rate-limiting step of the catalysis. In the last years, in the scientific community debates have raised about if the non-thermodynamic factors contribute to the catalytic power of enzymes, for example, the quantum tunneling<sup>54-59</sup> and the dynamics effects.<sup>60-65</sup> The contributions of the tunneling effect in catalysis have been studied by Klinman and co-workers in several enzymes;<sup>54,66-70</sup> they proposed that this nuclear quantum effect contribute significantly to the catalytic power of enzymes. The tunneling effect is a phenomenon in which a light particle tunnels through an energetic barrier that classically it could not

### *3. Enzyme Kinetics and Catalysis*

---

overcome. Probably, the dynamical contribution to the catalytic power of enzymes is the subject most debated currently in enzyme catalysis.<sup>71-76</sup> This idea, introduced by several authors<sup>77-80</sup> is based that the dynamical effects require deviations from the TST, and the deviation of the rate constant is obtained by means of a transmission coefficient. Although several studies support the idea that dynamical effects contribute to the enzyme catalysis,<sup>59,63,81-84</sup> Warshel and co-workers in several reviews about the role of dynamics in enzyme catalysis<sup>71,73,76</sup> concluded that no dynamical effect has ever been experimentally shown to contribute to catalysis and from theoretical studies, the contributions are small or negligible.





### 4. Enzyme Inhibition

Some enzymes are involved in the development of diseases. In such cases, we are interested in the inhibition of their enzyme activity. The substrate that is responsible for decreasing or stopping the enzymatic reactions is known as the *enzyme inhibitor*. In the pharmaceutical industry, many of the drugs are enzyme inhibitors. Also, the study of these inhibitors provides important information about the catalytic mechanisms of enzymes.

The inhibitors of the enzyme can be grouped into two general categories: reversible or irreversible.<sup>1</sup> Irreversible inhibitors bind covalently or form a stable noncovalent association with the enzyme through strong interactions.<sup>1,85</sup> A difference of the irreversible inhibition, the reversible inhibitor is characterized by a rapid dissociation of the enzyme-inhibitor complex. For the pharmaceutical industry, the view is the inhibitors acting as a drug should be reversible, due to concerns about the antigenicity of the enzymes covalently modified, and the consequences of suboptimal specificity, especially during long periods of therapy. However, in the case of CPs, in which the irreversible inhibitors are more effective than the reversible inhibitors, the irreversible inhibitors can be used therapeutically.<sup>14</sup>

#### 4.1. Reversible Inhibitors

In general, three classes of reversible inhibitors can be found: competitive, uncompetitive and mixed or noncompetitive inhibitors.<sup>1,86</sup> A *competitive inhibitor* competes with the substrate (S) for the active site of an enzyme (E), and it binds reversibly to the free enzyme to form an

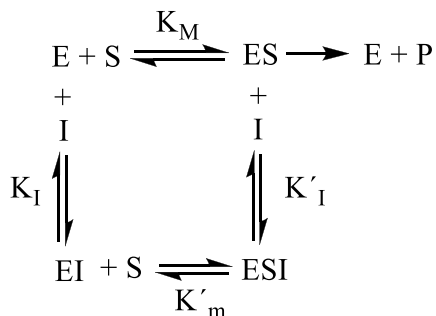
enzyme-inhibitor (EI) complex. A competitive inhibitor decreases the catalytic power of the enzyme due to the reduction the proportion of the enzyme molecules bound to a substrate. This class of inhibition can be overcome by increasing the concentration of the substrate.

Unlike the competitive inhibitors, the *uncompetitive inhibitor* binds only to the ES complex but not to free enzyme, and then producing an inactive enzyme-substrate-inhibitor (ESI) complex. This type of inhibition cannot be overcome by increasing the substrate concentration and is observed only for enzymes with two or more substrates.

The third class of reversible inhibitor can bind to both E and ES complex (see Scheme 4.1.1); the inhibitor binds to a site different to the active site, so that there is no competition with the substrate. And, like for uncompetitive inhibitor, this inhibition is observed only for enzymes with two or more substrates. When the dissociation constant of S from ES ( $K_M$ ) is the same as the dissociation constant of S from ESI complex ( $K'_m$ ) the inhibition is purely *noncompetitive*. Commonly, both dissociation constants are different. When E has a higher affinity for S does EI complex, the inhibition is a mixture of partial competitive and pure noncompetitive and is called *mixed inhibition*.

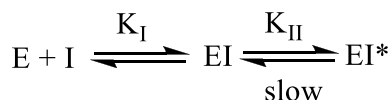
#### 4. Enzyme Inhibition

---



**Scheme 4.1.1** Non competitive/mixed inhibition. Scheme adapted from the reference 86.

Some inhibitors reach slowly the equilibrium between E, I and EI complex (see Scheme 4.1.2), these reversible inhibitors are known as *slow-binding inhibitors*.<sup>86</sup> This inhibition occurs as the result of a conformational change, from EI complex to another complex (EI\* complex) where the enzyme is forming a more stable complex with I than in the EI complex.<sup>87,88</sup> In some cases, the inhibitor reaches the equilibrium between E, I and EI complex when the concentrations of I and E are almost the same, in this case, the inhibitor is known as *slow, tight-binding inhibitor*.<sup>86</sup>



**Scheme 4.1.2** Slow-binding and slow, tight-binding inhibition. Scheme adapted from the reference 86.

## 4.2. Irreversible Inhibitors

A class of irreversible inhibitor is named as *affinity labels inhibitors*. These inhibitors have a high affinity for the active site of the enzyme and have a group chemically highly reactive toward a functional group of the enzyme. Initially, these inhibitors form a reversible EI complex and then reacts with the EI complex to form the covalent adduct enzyme-inhibitor.<sup>86</sup>

An important class of irreversible inhibitors is the *mechanism-based inhibitors*, also called *suicide inhibitors* or *enzyme-activated inhibitors*. These inhibitors have a structural similarity to the substrate or product for the target enzyme and provide the most specific means to modify the active site of the enzyme. The inhibitor is non-reactive compounds until they bind to the enzyme through a covalent modification, forming a chemically reactive intermediate that inactivates the enzyme.

From a kinetic point of view, the inhibition by an irreversible inhibitor usually proceeds by the rapid formation of a reversible EI complex (see Scheme 4.2.1). In a limiting chemical step, is formed the enzyme-inhibitor intermediate (E-I) through a covalent bond. The first-order rate constant ( $k_2$ ) is defined as the maximum or limiting inhibition rate if the enzyme is saturated with the inhibitor. The ratio  $k_2/K_I$  is a second-order inhibition rate ( $k_{2nd}$ ), and is one of the most used parameters to report inhibition data and is a measure of the preference of an enzyme by different inhibitors.<sup>89</sup>



**Scheme 4.2.1** Irreversible inhibition kinetics. EI is a reversible complex, and E-I is a covalent intermediate. Where  $K_I$  is the dissociation constant of the EI complex, and  $k_2$  is the first-order rate constant. Scheme adapted from reference 89.

### 4.3. Transition State Analogues Inhibitors

The most accepted theory for explain the enzymes catalysis is related with the idea proposed by Pauling in 1946,<sup>40</sup> the high affinity between the active site of the enzyme and the TS of the reaction catalyzed by the enzyme. In this sense, Bernhard and Orgel raised that an inhibitor structurally similar to the TS would bind to the enzyme much more strongly than the substrate.<sup>90</sup> This type of inhibitor is known as *transition state analogue* (TSA) inhibitor. Richard Wolfenden demonstrated that the affinity of the TSA inhibitor is proportional to the catalytic rate improvement imposed by the enzyme.<sup>91,92</sup> The TSAs are noncovalent inhibitors that mimic the charge distribution and the shape of the TS of the enzymatic reaction. Lipscomb and Christianson introduced a type of TSA inhibitor called *reaction-coordinate analog*, which interact covalently and reversibly with the enzyme.<sup>93</sup>

The binding of TSA inhibitors is the opposite in catalytic time scale to the formation of the TS, while the TS in an enzyme has lifetimes as short as  $10^{-15}$  s, TSA can bind tightly to the enzyme with release rates greater than  $10^3$  s. Compared to the weak associations found in the MC, TSA inhibitor involve strong interactions related to those in the TS. When a TSA inhibitor is bound to the enzyme, all the interactions is

stabilized, due to the dynamic motions characteristic of the catalytic site during the catalysis are less dynamic during this binding.<sup>94</sup>

The advances in the measuring of the kinetic isotope effects and modern methods of computational chemistry allow us to analyze in detail the TS structures and move forward in the design of TSAs as powerful inhibitors of the enzymes.<sup>95,96</sup> The tight binding affinity and the high specificity, together with the confirmed effectiveness of the TSA inhibitors,<sup>97-101</sup> becomes them in a promise for the development of drugs.

### 4.4. Some Parameters to Report Inhibition Data

Other parameter that is used in some cases to compare the activity of enzyme inhibitors is the  $IC_{50}$ .<sup>102</sup> This value is the concentration of inhibitor required to reduce the activity of the enzyme by 50 % under the conditions of the enzyme assay. The  $IC_{50}$  for the irreversible inhibitors depends on the time during which the enzyme is incubated with the inhibitor.<sup>89</sup> In this sense,  $IC_{50}$  values are largely arbitrary and can be even misleading, the incubation for a different period of time would give a different value.<sup>103</sup> Thus, a common recommendation is to utilize the values of  $k_{2nd}$  in preference to the  $IC_{50}$  values for comparing the potency of these inhibitors.<sup>104</sup> There are many cases in which the comparison of  $k_{2nd}$  values allowed a better differentiation between compounds structurally very close and with almost similar  $IC_{50}$  values.<sup>104-107</sup>

One problem in the development of effective inhibitors is the drug resistance, as for example occurs in malaria, the virus by mutating the enzyme used as a drug target can acquire resistance to the drug. In such

#### ***4. Enzyme Inhibition***

---

case, one strategy it is based in design inhibitors, whose binding to the target enzyme and cannot be reduced by mutations, maintaining a good values of  $k_{cat}/K_M$ .<sup>108</sup> This strategy is based into calculate the vitality value,  $\gamma$ , this term was introduced by Erickson and co-workers<sup>109</sup> and validated by Warshel and co-workers.<sup>110</sup> The vitality value is calculated according to:

$$\gamma = K_I \frac{k_{cat}}{K_M} \quad (4.4.1)$$





### 5. Cysteine Proteases

Proteases are an important class of enzymes involved in the hydrolysis of peptides and proteins; based on their catalytic mechanism have been separated into different groups: serine proteases, aspartic proteases, CPs, glutamic proteases, threonine proteases, or even metalloproteases. The CPs, as their name implies, are characterized principally by the presence of a cysteine residue in the active site. These enzymes are present in a variety of living organisms, from fungi to humans.<sup>111</sup> The CPs are divided into clans, different clans do not share the sequence or the structural identity; only share the use of a cysteine to catalyze the hydrolysis of peptide bonds, for example, clan CA utilize catalytic Cys, His, and Asn residues that are invariably in this order in the primary sequence.<sup>112</sup> Clans are divided into families depending on a number of characteristics including sequence similarity, possession of inserted peptide loops, and biochemical specificity to small peptide substrates.<sup>2</sup> The MEROPS database of proteolytic enzymes counts more than 10 clans of CPs.<sup>113</sup>

The best known family of the CPs is the papain family (family C1), belonging to clan CA. Papain was the first cysteine protease purified and characterized from *Carica papaya*, the papaya fruit; was also the first cysteine protease to be solved.<sup>2</sup> The CPs of the papain family are involved in many diseases, as it is shown in Table 5.1, making them attractive targets for developing new drugs.

**Table 5.1** Examples of some CPs of the papain family implicated in several diseases.

Cysteine Protease	Disease
Cruzain	Chagas
Falcipain-2	Malaria
Cathepsin K	Osteoporosis
Cathepsin L	Atherosclerosis
Rhodesain	Sleeping sickness

### 5.1. Catalytic Mechanism of the Cysteine Proteases

The catalytic mechanism of the CPs depends on two residues that are located in the active site, cysteine and histidine. It has been proposed that the imidazole group of histidine polarizes the SH group of the cysteine and enables deprotonation of SH group, thus forming a highly nucleophilic  $\text{CysS}^-/\text{HisH}^+$  ion pair.<sup>114</sup> This ion pair explains the unusually high reactivity of these proteases toward electrophilic reagents in comparison with the nucleophilic power of the sulfur of cysteine, especially in slightly acidic environments.<sup>115</sup> The existence of the ion pair was experimentally proven by different studies,<sup>116-118</sup> and it has been studied through the application of computational tools.<sup>119-123</sup>

Shafer and co-workers, in a potentiometric determination of ionization states at the active site of papain, concluded that the active site thiol group exists mainly as a thiol anion in the physiological pH range.<sup>116</sup> In 1997, Hillier and co-workers, in a QM/MM study of the catalytic mechanism of the enzyme papain showed that in the absence of the enzymatic environment the pair of neutral residues is strongly favoured; however, with the presence of the enzyme, the potential energy

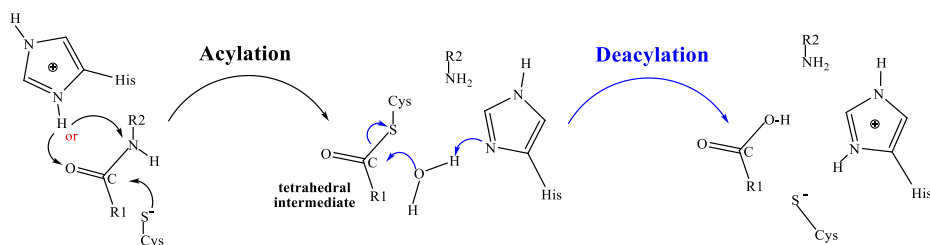
## 5. Cysteine Proteases

---

surface (PES) for proton transfer was very flat, with the ion pair more stabilized than the neutral species.<sup>119</sup> This was further supported by Suhai and co-workers that, in a QM/MM study of the active site of free papain and the NMA-Papain complex, concluded that the ion pair becomes more stable than its neutral form in the protein environment.<sup>120</sup> In 2008, Engels and co-workers concluded that ion pair was about  $7.2 \text{ kcal}\cdot\text{mol}^{-1}$  more stable than the neutral form based on MD simulations using QM/MM potentials on the CP cathepsin B, a CP that belongs to the papain family.<sup>121</sup> Moreover, according to their calculations, the enzyme environment would stabilize the ion pair by more than  $23.8 \text{ kcal}\cdot\text{mol}^{-1}$  compared with the gas phase, and by more than  $9.6 \text{ kcal}\cdot\text{mol}^{-1}$  compared with the ion pair in aqueous solution. More recently, Grazioso and co-workers based on exploration of QM/MM PES through ONIOM hybrid calculations of the mechanism for the FP2 inhibition by  $\alpha,\beta$ -unsaturated benzo[1,4]diazepin-2-one methyl ester, showed that the activation of Cys42 (FP2 numbering) by His174 is not the rate-limiting step and occur without important change in the relative energy of the system.<sup>122</sup>

In general, it is accepted that the catalytic mechanism of cysteine protease consists of two stages: acylation and deacylation. The acylation stage basically consists in the nucleophilic attack on the carbonyl carbon of peptide bond by the thiolate anion from cysteine producing a tetrahedral intermediate (see Scheme 5.1.1). Then, the deacylation occurs via a general base-catalyzed reaction, whereby the imidazole nitrogen from histidine activates a water molecule that attacks the carbonyl carbon of peptide bond to produce the product and regeneration of the  $\text{CysS}^-/\text{HisH}^+$  ion pair. Several experimental<sup>124-131</sup> and

computational<sup>119,120,123,132-141</sup> studies have been carried out in order to explain the manner in which the cysteine proteases catalyze the breaking of peptide bonds.



**Scheme 5.1.1** General mechanism of peptide hydrolysis catalyzed by CPs.

In the acylation step, it is discussed the manner in which the formation of the tetrahedral intermediate occurs. Thus, Drenth and co-workers, based on X-ray diffraction studies, proposed that this stage occurs as in the mechanism of serine proteases:<sup>127,128</sup> the thiolate anion attacks the carbonyl carbon of peptide bond to produce a thiohemiketal intermediate. This intermediate would be stabilized by a hydrogen bond interaction with the backbone NH group of cysteine and to the NH<sub>2</sub> group of the glutamine side chain that create an oxyanion hole similar to the one proposed in serine proteases. Then a proton is transferred from histidine to the nitrogen of peptide bond, concomitant with the C-N peptide bond cleavage, to form the tetrahedral intermediate. However, the stability of the thiohemiketal intermediate has been questioned by several authors.<sup>120,123,132,133,135,136</sup> Studies based on MM and semiempirical QM methods with reduced models of the papain-catalyzed reaction carried out by Kollman and co-workers, suggested that the

## 5. Cysteine Proteases

---

protonation of O or N atoms must occur prior or in a concerted manner with the nucleophilic attack of the thiolate anion.<sup>132,133</sup> Later, Hillier and co-workers by exploring QM/MM potential energy pathways with NMA as substrate, found that the attack of the thiolate anion occur in a concerted manner with the proton transfer from the histidine to the N atom of a peptide bond, with a potential energy barrier of 20.1 kcal·mol<sup>-1</sup>.<sup>119</sup> Kolandaivel and co-workers explored the PES of the inhibition of haloketones<sup>138</sup> and diketones<sup>139</sup> in gas phase at DFT level, concluding that the intermediate obtained from the attack of the thiolate anion and the proton transfer from the histidine to the O atom of a peptide bond, would be stable but, in general, less stable than the reactant. Albeck and co-workers in a QM/self consistent reaction field (virtual solvent) study of the inhibition mechanism of by peptidyl aldehydes proposed that the final product of the acylation step is the thiohemiketal protonated.<sup>140</sup> Recently, Komáromi and Fekete have been studied a proteolytic reaction of papain with a substrate NMA using dispersion corrected DFT methods in ONIOM-type hybrid QM/MM calculations, they concluded that the acylation step occur in two steps through the formation of a zwitterionic tetrahedral intermediate, being the formation of this intermediate the rate limiting step of the acylation.<sup>141</sup>

After acylation, deacylation occurs when a water molecule attacks the acyl-enzyme complex at the carbonyl carbon to produce the product and regeneration of the CysS<sup>-</sup>/HisH<sup>+</sup> ion pair. The formation of the ion pair also makes the thiol a good leaving group during the deacylation due its low pKa value.<sup>111</sup> Welsh and co-workers, in a discussion of the catalytic pathway of cysteine proteases based on AM1 calculations,

showed that the protonation of histidine by a water molecule and the formation of the C-O<sub>water</sub> bond occur simultaneously and with an activation barrier of 39.0 kcal·mol<sup>-1</sup>.<sup>134</sup> Years later, Gao and co-workers, in a study of the catalytic pathway of a human cathepsin K (belongs to the papain family), based on QM/MM MD simulations, published that the proton transfer from the water molecule to the imidazole ring of the catalytic histidine was fully concerted with the nucleophilic attack from O<sub>water</sub> atom to the carbonyl carbon and presenting a free energy of activation of 16.7 kcal·mol<sup>-1</sup>.<sup>137</sup> The computational study of Komáromi and Fekete,<sup>141</sup> yielded to the same conclusion that the computational studies conducted separately by Gao and Welsh. However, Warshel and co-workers, in an *ab initio* study of several reference reactions in solution (the histidine-assisted thiomethanolysis of formamide and the hydrolysis of the resulting thioester), had concluded that the deacylation has stepwise character through a thiohemiketal intermediate that would be formed prior to the cleavage of the S-C bond, being the thiohemiketal formation the rate-limiting step.<sup>136</sup> In 2013, Zhan and co-workers<sup>123</sup> in a pseudobond first-principles QM/MM FE study of the reaction pathway for papain-catalyzed hydrolysis of N-acetyl-Phe-Gly 4-nitroanilide arrived to the same conclusions than Warshel and co-workers. Thus, to date, many aspects around the catalytic mechanism of CPs are unclear and under discussion.

### 5.2. Cruzain and Chagas Disease

Cruzain, a CP of the papain family, was initially discovered from the parasite cell-free extracts and subsequently heterologously expressed

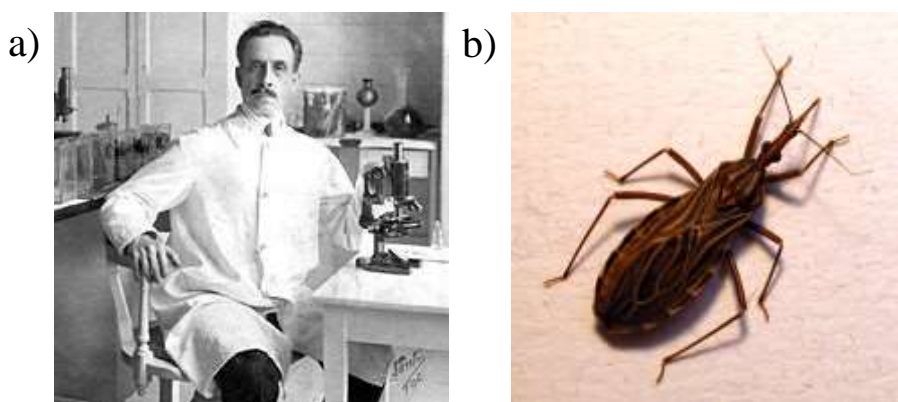
## 5. Cysteine Proteases

---

in *Escherichia coli*.<sup>142,143</sup> Cruzain is a key protease in *Trypanosoma cruzi* (*T. cruzi*), the protozoan parasite responsible for the Chagas disease. This CP is important to the development and survival of the parasite; and is expressed in all life cycle stages of the parasite.<sup>144-147</sup>

Chagas disease (American trypanosomiasis) was discovered in 1909 by the Brazilian doctor Carlos Chagas (1879-1934, see Figure 5.2.1a).<sup>148</sup> The first reported case of Chagas disease was in 1835; Charles Darwin possibly contracted the disease during his expedition to South America, as suggested by his description of contact with the triatomine bug and by the symptoms described by him in later years.<sup>19</sup> It is estimated that the illness affects about 6 million to 7 million people living mainly in endemic areas of 21 Latin American countries, with the exception that the disease has not occurred in the Caribbean isles. The parasite is transmitted to humans mostly by contact with faeces or urine of triatomine bugs (see Figure 5.2.1b). More than 9000 years ago, *T. cruzi* only affected wild animals, but it later extended to domestic animals and people.<sup>149</sup> The movement of Chagas disease to areas previously considered non-endemic, resulting from increasing population mobility between Latin America and the rest of the world, represents a serious public health challenge,<sup>150</sup> being Spain the most affected country in Europe.<sup>151</sup> There is not any available vaccine to prevent Chagas disease and the two available drugs for treatment, benznidazole and nifurtimox, are not only toxic and with important contra-indications (pregnancy, renal or hepatic failure, psychiatric and neuronal disorders), but also ineffective for the chronic stage of the disease.<sup>20,21</sup> Thus, to date, the control of the vector is the most effective method of prevention in Latin American.<sup>152</sup> Clearly, there is an urgent need for developing an

effective therapy against Chagas disease. One approach consists of developing inhibitors of cruzain, the primary CP expressed by *T. cruzi* during infection. [17,18,153](#)



**Figure 5.2.1** a) Carlos Justiniano Ribeiro Chagas, in his laboratory at the Instituto Oswaldo Cruz. b) *Rhodnius prolixus* is the second most important triatomine vector of the Chagas parasite.

Addition of a cruzain inhibitor to cultures of mammalian cells exposed to trypomastigotes or to mammalian cells already infected with *T. cruzi* amastigotes blocks replication and differentiation of the parasite, thus interrupting the parasite life cycle. [154-160](#) In this sense, several groups have demonstrated that irreversible inhibition of cruzain by small molecules eradicates infection of the parasite in cell culture and animal models. [156,161-169](#) McKerrow and co-workers reported the first successful treatment disease in mice with fluoromethyl ketone-derivatized pseudopeptides inhibitors. [156](#)

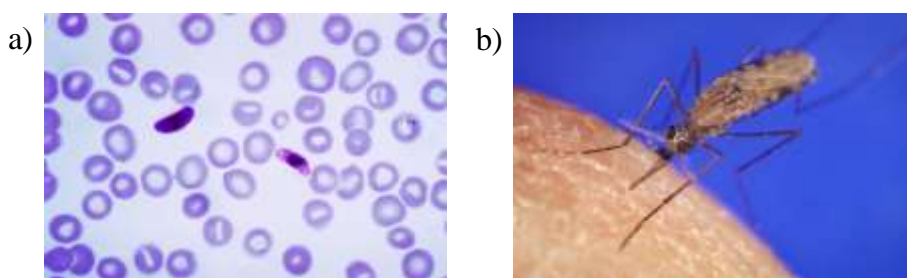


### 5.3. Falcipain-2 and Malaria

Malaria is one of the most serious infectious diseases in the world. In humans, is caused by five species of parasites of the genus *Plasmodiums*: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. Malaria due to *P. falciparum* (see Figure 5.3.1a) is the most deadly and prolific form and *P. vivax* is the most extended; the other three species are found much less frequently. The malaria parasite was discovered by the French Alphonse Laveran in 1878,<sup>170</sup> he was awarded in 1907 with the Nobel Prize in Medicine for this and other discoveries related with the role played by protozoa in causing diseases.<sup>171</sup> Later, Carlos Juan Finlay (1833-1915), a Cuban doctor and scientist that researched about the yellow fever in Havana, provided strong evidence that mosquitoes were transmitting disease to and from humans.<sup>172</sup> In 1898, the British Sir Ronald Ross (1857-1932) demonstrated malaria was transmitted by mosquitoes; he was awarded in 1902 with the Nobel Prize in Medicine "*for his work on malaria, by which he has shown how it enters the organism and thereby has laid the foundation for successful research on this disease and methods of combating it*".<sup>173</sup>

Malaria is transmitted to humans by the bite of infected female mosquitoes of more than 30 *Anopheles* species (see Figure 5.3.1b). Globally, an estimated 3.2 billion people in 91 countries and territories are at risk of being infected with malaria and developing disease, and 1.2 billion are at high risk.<sup>174</sup> In the last years, the rate of new malaria cases reduced globally by ~40%, and the global malaria death rate declined by 60%, due to mainly to the wide-scale deployment of malaria control interventions. However, malaria continues to have a negative

impact on the public health, in 2015, 429 000 people died from the disease, most of them children under 5 years in Africa. The prevention strategies of the disease are based in an effective vector control (use of insecticide-treated mosquito nets and indoor residual spraying), the chemoprevention for pregnant women, or treatment with artemisinin-based combination therapy for uncomplicated cases of the disease (particularly for *P. falciparum* malaria).<sup>175</sup> In the beginning of 2018, a pilot implementation of a vaccine will begin, the RTS,S/AS01 malaria vaccine, also known as Mosquirix™.<sup>176</sup> This is the most advanced vaccine today, and aims to activate the immune system to defend against the first stages when the *P. falciparum* malaria parasite enters the human host's bloodstream through a mosquito bite and infects liver cells. Despite having gone through the different phases of study with excellent results,<sup>177-180</sup> the vaccine provides partial protection only in young children.



**Figure 5.3.1** a) *Plasmodium falciparum* in the blood. b) The *Anopheles gambiae* mosquito, one of the best known because it transmits the specie *P. falciparum*.

The hydrolysis of hemoglobin is the best characterized function of *P. Falciparum* (see Figure 5.3.2),<sup>181</sup> which occurs in an acidic digestive vacuole. As hemoglobin is processed, its heme component is converted

## 5. Cysteine Proteases

into the pigment hemozoin, and globin is hydrolyzed into its amino acids. This process is necessary to provide amino acids for parasite protein synthesis<sup>182,183</sup> or to keep the osmotic stability of the parasite.<sup>184</sup> Hemoglobin hydrolysis appears to be the result of a cooperative process involving proteases of different catalytic classes: cysteine, aspartic (different plasmepsins), and metallo proteases (falcilysin).<sup>185-190</sup> CPs belonging to the papain family are involved in this metabolic process, known as falcipains, which work in an ordered manner, and with different specificities.<sup>191</sup> FP2, the most abundant and best-studied of the falcipains, is located in the food vacuole and expressed during the erythrocytic stage of the life cycle of the parasite.<sup>192</sup> The inhibition of FP2 has proven indispensable in order to completely block parasitic growth and proliferation,<sup>22,23</sup> therefore, FP2 has emerged as a promising target for the development of novel antimalarial drugs.<sup>22,193,194</sup>

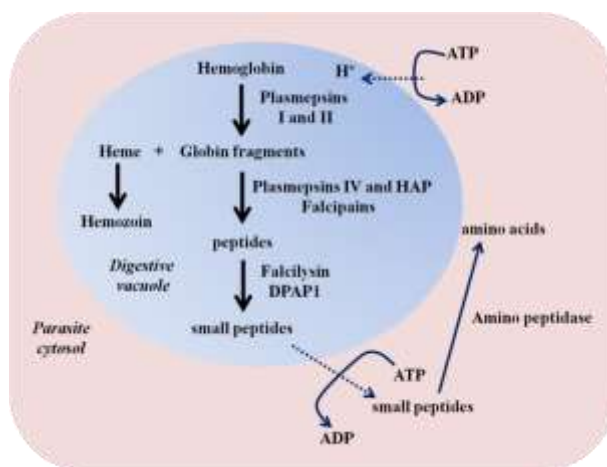


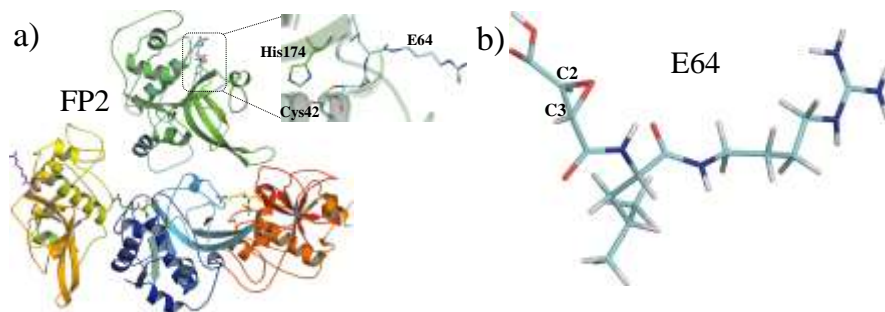
Figure 5.3.2 Hydrolysis of hemoglobin in Plasmodium food vacuole. Figure adapted from reference 181. HAP is the histo-aspartic protease and DPAP1 is the dipeptidyl aminopeptidase 1.

## 5.4. Inhibition Mechanism of Falcipain-2 by the Epoxysuccinate E64

Several types of inhibitors of FP2 have been reported in the last years to be capable to inactivate the enzyme;<sup>195-200</sup> however, only few structures reported in the PDB correspond to the crystal structure of FP2 with its inhibitors.<sup>195,201,202</sup> One of them is the structure of FP2 complexed with E64 (see Figure 5.4.1),<sup>195</sup> the first epoxysuccinyl peptide discovered and a natural potent and specific irreversible inhibitor of CP. One advantage of this type of inhibitors is their stability under physiological conditions toward simple thiols. In addition, although they have limited selectivity toward different CPs, their reactivity toward CPs and not other proteases, along with their chemical unreactivity, makes this class of inhibitors useful as pharmaceutical agents.<sup>203</sup> E64 was isolated from a culture of *Asperigillus japonicas*<sup>204</sup> and its structure was determined by Hanada and co-workers in 1978.<sup>205</sup> E64 contains a *trans* (2S,3S) configured epoxide ring, whereas the amino acid residues of the peptidyl part of the inhibitor have the L-configuration. The substituents at C2 and C3 on the epoxide ring are in *trans* position to one another,<sup>206,207</sup> which appears to be crucial since *cis* configuration leads to total loss of inhibition activity.

## 5. Cysteine Proteases

---



**Figure 5.4.1** a) Full tetramer showing four chains that compose the crystal structure of FP2 inhibited from *P. Falciparum*, 3BPF PDB code. Detail of the active site where the inhibitor (E64) binds to the enzyme through the residue Cys42. E64, Cys42 and His174 are represented as thick sticks. b) Inhibitor E64.

E64 inhibits CPs by S-alkylation of the active site cysteine, which results in the opening of the epoxide ring and stabilization of the protein-inhibitor complex.<sup>208,209</sup> Nevertheless, there are some open questions related with the molecular mechanism, since it appears that the attack can take place at either C2 or C3 depending on the orientation of the epoxysuccinate in the active site.<sup>89</sup> On one hand, it has been demonstrated by X-ray investigations, that the attack of the alkylation step with epoxide takes place at C2 when having an acid at C2 substituent on this carbon.<sup>195,210-216</sup> For example, the crystal structure of papain-E64c complex reported by Ishida and co-workers (the E64c is a synthetic inhibitor developed from the natural inhibitor E64), in which the inhibitor (E64c) is bound to the residue Cys25 (papain numbering) of papain through a covalent bond formed between the epoxy C2 atom of E64c and the Cys25 residue.<sup>212</sup> The regiospecificity and inhibition potency of epoxide-based inhibitors were studied by theoretical methods by Engels and co-workers who, based on exploration of PES obtained

with QM/MM potentials, proposed the attack at C2 as the more favourable.<sup>121</sup> Bihosky, in an experimental study of the rate and regioselectivity of reactions between 2,3-epoxy carbonyl compounds and methanethiolate in solution, observed that for 2,3-epoxy amides the C3 attack is preferred, while for 2,3-epoxy esters and carboxylic acids the cysteine can attack both C2 or C3 atoms in a similar way.<sup>217</sup>

An insight into the molecular mechanism of CP inhibition by epoxy inhibitors shows that an important feature is that the stereochemistry of the enzyme–inhibitor adduct suffered an inversion of configuration at the reaction site due to a nucleophilic attack by the active site thiolate in an  $S_N2$  reaction.<sup>89</sup> For instance, E64, which has the 2S,3S configuration before the nucleophilic attack, becomes 2R,3R after the covalent bond between the cysteine residue and C2 is formed.<sup>89</sup>

Another questions of debate is related with the nature of the catalytic acid on the active site and also if the attack of the residue cysteine occurs after or before the protonation of the epoxy ring. It was initially postulated that when E64 inhibits papain, the oxirane ring would be protonated by a histidine.<sup>218</sup> Varughese and co-workers<sup>210,213</sup> rejected this suggestion, based on both crystal structures of papain inhibited by E64 and actinidin inhibited by E64. They suggest that the epoxide was more likely to be protonated by a water molecule because of the large distance between the resulting hydroxyl group and the histidine. Meara and co-workers in a mechanistic studies based on experimental determination of rate constants and pH dependence of inhibition confirmed that the histidine is not necessary for protonation of the epoxide and that water is the predominant source of protons; these results

## *5. Cysteine Proteases*

---

are likely to be general for all epoxysuccinyl inhibitors with a free carboxylate bind to C2 atom.<sup>219</sup> Kim and co-workers, in a mechanistic study of the inhibition of papain with 2-benzyl-3,4-epoxybutanoic acid esters,<sup>220</sup> concluded that the sulfur attack of Cys25 to the oxirane ring can occur without a prior protonation of the oxirane. Engels and co-workers, in several theoretical studies,<sup>121,221-223</sup> proposed that the protonation of the oxirane ring takes place far behind the transition state.

### **5.5. Inhibition Mechanism of Cruzain by Halomethyl Ketones and Epoxyketones**

As mentioned before, only irreversible inhibitors of cruzain have successfully cured parasitic infection, irreversible inhibitors that contain an electrophilic functional group, such as PHK or PEK, covalently bind to cruzain via nucleophilic attack of the active site cysteine,<sup>24</sup> showing good inhibition activity.<sup>11,111,206,224-226</sup>

Shoellmann and Shaw developed in 1962 the first PHK, L-1-tosyl-amido-2-phenylethyl chloromethyl ketone, as specific inhibitors for the serine protease chymotrypsin.<sup>227</sup> The major disadvantage of this inhibitor was their lack of selectivity due the great chemical reactivity of the chloroketone functional group. The development of chloromethyl ketone inhibitors leads to the investigation of analogous inhibitor structures with different halo leaving groups replacing the chlorine atom. Bromomethyl and iodomethyl ketones have been synthesized and are typically more reactive but less stable in aqueous solutions. In the 1980s were reported the first peptide

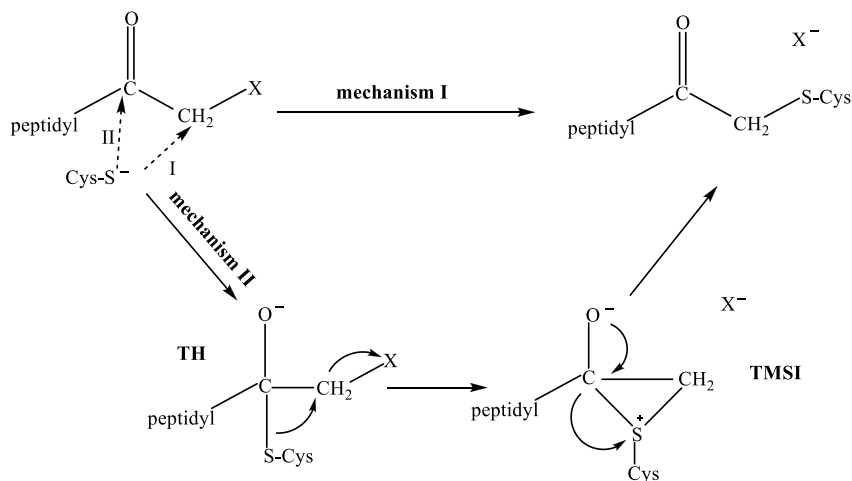
fluoromethyl ketones inhibitors in the literature.<sup>228,229</sup> These peptide fluoromethyl ketones were shown to be highly reactive and selective irreversible inhibitors for cysteine proteases. They are poor irreversible inactivators for serine proteases and are not reactive toward bionucleophiles. Inhibition of cruzain activity with fluoromethyl ketone-based inhibitors seems to be correlated with the loss of feasibility of parasites in both *ex-vivo* tissue culture and *in vivo* mouse models.<sup>154,230-232</sup>

Powers and co-workers proposed two possible mechanisms of irreversible inhibition by PHK based on different crystal structures of cysteine proteases (see Scheme 5.5.1).<sup>89</sup> The mechanism I is the direct displacement of the halide group by the thiolate anion. The mechanism II involves formation of a tetrahedral intermediate named thiohemiketal (TH), the formation of this intermediate is equivalent to the presence of a tetrahedral intermediate in the catalytic mechanism of serine proteases;<sup>111</sup> a three-membered sulfonium intermediate (TMSI) is then formed, which is rearranged to give the final thioether adduct. Nevertheless, mechanism II has not been supported by theoretical studies due to principally to the instability of the TH intermediated, as it mentioned before in section 5.1. Kollman and co-workers, showed that the attack of sulfur on a carbonyl carbon does not involve a stable anionic tetrahedral structure.<sup>132,133</sup> Later, Suhai and co-workers in a QM/MM study of the active site of free papain and of the NMA-Papain complex did not obtain a stable tetrahedral NMA-papain complex.<sup>120</sup> Gao and Byun, in a combined QM/MM study of the nucleophilic addition reaction of methanethiolate and NMA, concluded that there is no stable tetrahedral intermediate in going from the reactant to the tetrahedral adduct.<sup>135</sup>



## 5. Cysteine Proteases

Moreover, Warshel and co-workers shown that the anionic tetrahedral intermediate for the acylation reaction was found to be unstable in aqueous solution and to collapse immediately into the neutral form, which is the only intermediate on the reaction pathway.<sup>136</sup>



**Scheme 5.5.1** Reaction mechanisms of inhibition of cysteine proteases by peptidyl halomethyl ketones (X: F, Cl). TH refers to thiohemiketal intermediate and TMSI to the three-membered sulfonium intermediate. Scheme adapted from reference 89.

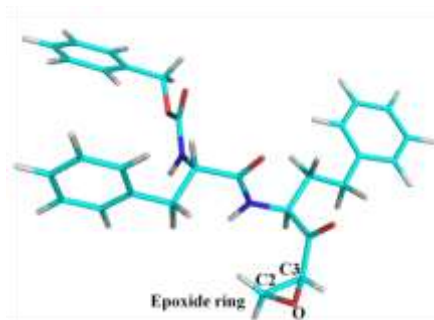
The PEK are irreversible inhibitors of cruzain that belong to the same family that the epoxysuccinyl peptide E64 (mentioned in section 5.4), the family of epoxides. Several studies have been done to establish an order of reactivity of different substituent in the C2 atom of the epoxy ring.<sup>217,219,222,233</sup> In 1992, Ron Bihovsky in an experimental study of the reactions of 2,3-epoxy carbonyl compounds in solution with methanethiolate, observed for this non-enzymatic thiolate that the reactivity decreases in the order  $\text{COCH}_3 > \text{CO}_2\text{CH}_3 > \text{CONH}_2 > \text{H} > \text{CO}_2\text{H}$ .<sup>217</sup> However, the following year, Bihovsky and co-workers

synthesized and tested several epoxysuccinates as inhibitors of papain, and observed the inverse trend for the inhibition rate constants:  $\text{CO}_2\text{H} > \text{CONH}_2 > \text{CO}_2\text{R} > \text{COCH}_3 > \text{H}$ ,<sup>233</sup> they conclude that the electrostatic attraction between the protonated His159 (cruzain numbering) and the carboxylate of the inhibitor facilitates the docking of the inhibitor in the active site of papain, thus the rate-limiting step in the enzyme inhibition is not the reactivity of the epoxide, but rather appears to be the rate with which the inhibitor docks in the active site prior to formation of the covalent complex. Later, Engels and co-workers, in a theoretical study about the influence of different ring substituents on the nucleophilic ring opening of three-membered heterocyclic, concluded for epoxides that acidic substituent increase the reaction barrier of the alkylation step.<sup>222</sup>

In 1998, Roush and co-workers designed and synthesized a new class of cruzain inhibitor,<sup>206</sup> a PEK represented in the Figure 5.5.1 Cbz-Phe-Hph-(2S), where Hph is homophenylalanine. This potent and irreversible inhibitor was designed by combining a portion of the epoxide moiety of the known epoxysuccinyl inhibitor E64c<sup>205</sup> and the peptide sequence of the dipeptidyl fluoromethyl ketone, Cbz-Phe-AIa-CH<sub>2</sub>F.<sup>229</sup> In fact, the second-order rate constant ( $k_{2\text{nd}}$ ) for the inhibition of cruzain by Cbz-Phe-Hph-(2S) epoxyketone proved to be 4.5-fold greater than that by E64c ( $333000 \text{ M}^{-1}\cdot\text{s}^{-1}$  and  $70600 \text{ M}^{-1}\cdot\text{s}^{-1}$  for Cbz-Phe-Hph-(2S)<sup>206</sup> and E64c,<sup>234</sup> respectively).

## 5. Cysteine Proteases

---



**Figure 5.5.1** Dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph-(2S).

The inhibition mechanism of CPs by PEK is not clear yet. Albeck and co-workers proposed that is based on an enzyme-catalyzed alkylation of the active-site cysteine from a Michaelis complex between the enzyme and the inhibitor.<sup>235-237</sup> They suggested in correspondence with other studies<sup>132,133</sup> that the protonation of the oxirane ring by the residue histidine occurs first or in a concerted manner with the nucleophilic attack on the amide scissile bond. Albeck and Kliper proposed that the inhibition mechanism of cysteine proteases by the epoxysuccinyl E64 is different than by PEK;<sup>237</sup> whereas the PEK are simple epoxides which interact directly with the thiolate, the E64 is activated epoxide and interacts with the enzyme through an initial attack on the carbonyl carbon, in analogy with the inhibition mechanism by peptidyl diazomethanes.<sup>237</sup>

Nevertheless, to date, the inhibition mechanism of cysteine protease cruzain by PHK and PEK has not yet been studied by computational tools including the protein environment effects.

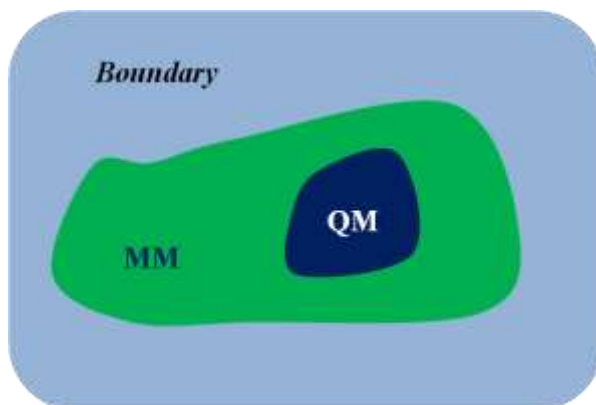


## 6. Theoretical Methods

### 6.1. QM/MM Hybrid Methodology

The inclusion of the environment effects in the computational simulations is basic for the study of the enzymatic reactions. The enzyme has an influence on the whole reaction mechanism, generating specific interactions which can modify the process. An important method to study the details of the chemical reaction and the effects of the enzyme and solvent environment in the chemical reaction is the QM/MM methods introduced by Karplus, Warshel, and Levitt.<sup>27-29</sup> Another notable method in this field is the empirical valence bond (EVB) method developed by Warshel and co-workers.<sup>238,239</sup> The EVB method is an empirical version of the QM/MM methods and the reactions are modelled by empirical potential functions. An alternative method to the QM/MM methods for calculating potential energy barriers and reaction energies in enzyme reactions is the quantum chemical cluster approach when a large molecular model is used.<sup>240,241</sup> This method, nevertheless, can not reproduce long range interactions and they are very dependent on the initial X-ray structure.

In order to use the QM/MM hybrid methodology, the system is divided into different regions (see Figure 6.1.1). The QM region should contain the atoms which participate in the chemical reaction and it is treated by the QM methods (see section 6.1.1). The MM region contains the rest of the atoms in the system and it is described by a force field. Finally, the boundary is a collection of restrictions applied to the system due to the fact that it is impossible to simulate an infinite system.



**Figure 6.1.1** Schematic representation of the partitioning of the system to be studied applying the QM/MM hybrid methodology. Figure adapted from reference 28.

There are two general types of QM/MM scheme taking into consideration the way the coupling between the QM and the MM part is performed: additive and subtractive methods. In the subtractive QM/MM scheme, different parts of the system are subjected to independent calculations at different levels of theory, as used in the ONIOM method developed by Morokuma and co-workers.<sup>242,243</sup> This scheme requires an MM calculation of the entire system, a QM calculation on the inner subsystem and an MM calculation on the inner subsystem; thus, no explicit QM-MM coupling terms are needed due to the coupling between subsystems is handled at the MM level of theory. The main advantage of this scheme is its simplicity; however, the scheme is not adequate for cases in which the electronic structure of the QM region is perturbed by interaction with the environment.

The additive scheme is probably the most widely adopted approach to QM/MM calculations and its principal advantage is that the energy

## 6. Theoretical Methods

---

calculation of the QM region can be directly executed in the presence of the classical environment in such a way that the electron density of the QM region is optimized in the external electrostatic field of the surroundings.<sup>244</sup> In the additive QM/MM scheme, the effective Hamiltonian,  $\hat{H}_{\text{eff}}$ ,<sup>29</sup> consists of:

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} + \hat{H}_{\text{BC}} \quad (6.1.1)$$

In the equation 6.1.1 the term  $\hat{H}_{\text{QM}}$  describes the electrons, nuclei and interactions among the atoms of the QM region and corresponds to an electronic Hamiltonian of the subsystem in vacuum:

$$\hat{H}_{\text{QM}} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{i\alpha} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \quad (6.1.2)$$

The equation 6.1.2 is the sum of the kinetic energy of the electrons, the electron-electron repulsion, the nucleus-electron attraction and the nucleus-nucleus repulsion; where  $i$  and  $j$  are the electronic coordinates,  $\alpha$  and  $\beta$  are the nuclear coordinates,  $r$  is electron-electron or nucleus-electron distance,  $R$  is the nucleus-nucleus distance,  $Z$  is the nuclear charge, and  $\nabla_i^2$  is the kinetic energy operator or Laplacian operator. The term  $\hat{H}_{\text{QM}}$  depends on the quantum method selected to describe the QM region.

In the previous equation 6.1.1, the term  $\hat{H}_{\text{MM}}$  describes the interactions between the atoms of the MM region. These atoms are the rest of the enzymes and surrounding solvent, and are described by a mechanical force field. There are several parameterized force fields to describe the enzyme's atom of the MM region; the most popular choices are AMBER,<sup>245-247</sup> CHARMM<sup>248-252</sup> or OPLS.<sup>31,253,254</sup> In the present

Doctoral Thesis, the force field employed was the OPLS-AA.<sup>31,253,254</sup> In the case of the solvent environment, generally water molecules, there are two general methods to model these effects, methods that include implicit solvent or explicit solvent. In the first case, the models do not include individual water molecules, and the effect is modelled through a dielectric constant, the disadvantage of such methods is that the hydrogen bonding interactions to water molecules cannot be modelled. In the explicit models the water molecules are represented explicitly. One of the most used force field for the description of the water molecules is the TIP3P model.<sup>32</sup>

The term  $\hat{H}_{QM/MM}$  in the equation 6.1.1 describes the interactions between the QM and MM atoms. Due to the MM atoms are represented by punctual charges and van der Waals parameters, this Hamiltonian is expressed as:

$$\hat{H}_{QM/MM} = - \sum_{iM} \frac{q_M}{r_{iM}} + \sum_{\alpha M} \frac{Z_{\alpha} q_M}{R_{\alpha M}} + \sum_{\alpha M} \left\{ \frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^6} \right\} \quad (6.1.3)$$

In the equation 6.1.3,  $i$  is the electrons of the QM atoms,  $\alpha$  is the nuclei of the QM atoms,  $M$  refers to the MM atoms,  $q_M$  is the charge of the MM atoms, and  $Z_{\alpha}$  is the nuclear charge of the QM atoms. The first two terms represent the electrostatic interactions, which consist of two Coulombic terms, one between the electrons and the MM atoms and one between the MM atoms and the nuclei of the QM atoms. The last term describes the Lennard-Jones interactions between the MM and QM atoms. In this equation, just the first term is an operator as it contains the electrons' coordinates, so, it must be incorporated in the *self-consistent*

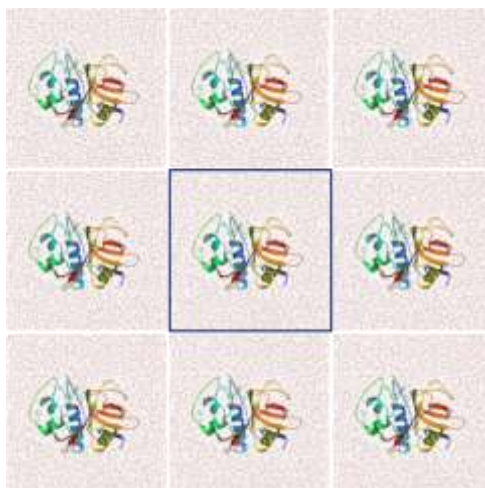


## 6. Theoretical Methods

---

*field* procedure. The remaining terms are constant for a given set of atomic coordinates, like the MM energy.

Returning to the equation 6.1.1, the term  $\hat{H}_{BC}$  appears due to the attempt to define something that is infinite using a system that can only be defined in an explicit way, with a finite number of molecules. Nevertheless, the term  $\hat{H}_{BC}$  is not always a real or computable term and it will depend on the chosen model, as it is explained next. The most commonly used representation is Periodic Boundary Conditions (PBC, see Figure 6.1.2).<sup>255</sup>

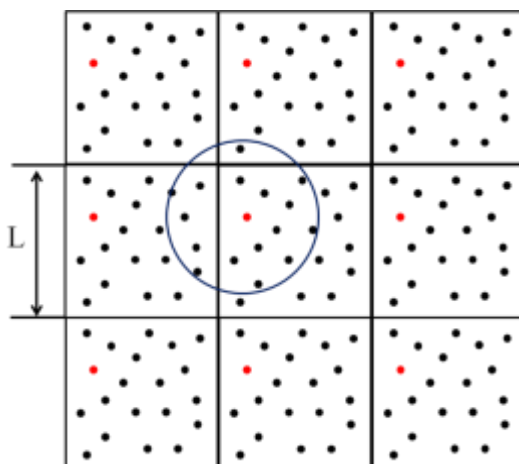


**Figure 6.1.2** PBC in 2D representations. The central square cell, highlighted in blue, is replicated in all directions of the space. Figure adapted from reference 255.

In PBC, the system is introduced in a solvent box which is replicated in an infinite way in the three dimensions of the space. The finite system should have a regular shape in order to fill the space when it is replicated, what is achieved using very regular boxes such as cubic or orthorhombic ones. In this approximation, the equivalent atoms in each

of the copies behave identically and they do not need to be treated differently during the simulation.<sup>255</sup>

To calculate the non-bonding energy, we should avoid that each particle could interact with its nearest copy in the system, this is the *minimum image convention* (see Figure 6.1.3).<sup>255</sup> The *minimum image convention* is a cheap method and easy to implement, but it means that the cutoff distance for truncation of the non-bonding interactions must be less than or equal to half the length of the side of the periodic box.



**Figure 6.1.3** Schematic representation of the *minimum image convention*. The red particle will interact with all the particles within the circle. It is assumed a circular truncation of the interaction between particles. The circle has a radius of  $L/2$ . Figure adapted from reference 255.

### Cutoff methods for the calculation of non-bonding interactions.

An important problem with the calculation of the non-bonding energy is the long-range electrostatic interaction, which decrease as the reciprocal of the distance between the atoms. The non-bonding energy

## 6. Theoretical Methods

---

should be evaluated over all the atom pairs of the system, but this would be computationally inaccessible. Then, different methods have been proposed to solve this problem based on functions that avoid the estimation of atom pairs located at long distance. In brief, these are “truncated functions” or a “smoothing functions”.<sup>255</sup> In the former, all the interactions beyond a cutoff distance ( $r_c$ ) are neglected. In the later the interaction is progressively tapped to zero at given distance. In turn, there are two types of smoothing functions.

### **Frontier bonds: link atom method.**

In the study of enzymatic systems applying QM/MM schemes, a covalent bond has to be cut when dividing the MM and QM part in many occasions (generally a covalent bond between two residues of the enzyme or between two atoms in a residue with a large side-chain). When a covalent bond between two atoms is cut by dividing the system (QM and MM regions), the valence of the QM atom is incomplete, therefore, the valence of this QM atom needs to be satisfied. There are several methods to perform this partition: the *link atom* based method<sup>29,256</sup> and the *frontier region* based methods: the *Local Self-Consistent Field*<sup>257-261</sup> and the *Generalized Hybrid Orbital* method.<sup>260-262</sup> In the present Doctoral Thesis all the calculations were conducted using the link atom method, the more straightforward treatment of the frontier bonds.

The *link atom* method is the most widely used method to treat the frontier bond between the QM and MM regions. It is based on replacing the MM part of the bond by an atom, usually a hydrogen atom, due to the

fact that the electronegativities of the C and H atoms are similar. This link atom is invisible to the MM region and it is used to saturate the valences of the truncated bond. Usually, the position of the link atom do not cut across any polar bonds, so, the most common option to place the link atom is in the C<sub>sp3</sub>-C<sub>sp3</sub> bond.

### 6.1.1 QM Methods

The use of QM methods is necessary for the study of the electronic rearrangements which take place during the breaking and forming bond in chemical reactions. It is crucial to use a QM method that represents accurately the system under consideration, thus, it is required a balanced compromise between the QM method, the size of the simulation system, and the nature of the calculations.

There is a great variety of QM methods, from fastest semiempirical methods, to more accurate and more computationally expensive, DFT and *ab initio* methods. In general, the most used methods for the study of enzymatic reactions are the semiempirical and DFT methods. The Hartree-Fock (HF) and post-HF methods are rarely used. In the case of HF methods are less accurate due to the lack of electron correlation; despite the inclusion of certain degree of electron correlation in the post-HF method, there are computationally very expensive.<sup>263</sup>

The semiempirical methods are based on the HF formalism, but they make many approximations and obtain some parameters from empirical data. Of which, methods based on the NDDO integral approximation, as for example AM1,<sup>264</sup> PM3<sup>265</sup> or MNDO,<sup>266,267</sup> are the

## 6. Theoretical Methods

---

most used for the QM/MM MD studies. For reactions involving elements as phosphorus and sulfur, NDDO methods that explicitly include *d* orbitals, such as AM1d<sup>268</sup> and MNDO(d)<sup>269</sup> methods have been developed. In the present Doctoral Thesis, it is used the AM1d method due to the participation of a sulfur atom in the enzymatic mechanisms studied (S from Cys residue).

The DFT methods, based on the theoretical foundations of Hohenberg, Kohn, and Sham,<sup>270,271</sup> uses the electron density instead of a wave function to determinate the energy of a molecule. Of these, the time-dependent DFT (TD-DFT)<sup>272</sup> has become in a powerful tool for evaluating the medium effects with the help of the QM/MM methods.<sup>27-29</sup> The B3LYP hybrid functional<sup>273,274</sup> is the most popular functional in chemistry, however, in the last years, other hybrid functionals relatively broadly accurate are being extensively used, as for example the Minnesota M06 suite of functionals density, M06<sup>275</sup> and M06-2X.<sup>275</sup> The hybrid meta-GGA M06-2X (the functional used in the present Doctoral Thesis) is a high nonlocality functional with a 54% of HF exchange, and it is parametrized only for non-metals. It has been recommended by Truhlar and co-workers for applications involving the main-group thermochemistry (should predict accurate structures, energies, and vibrational frequencies for compounds containing only main-group elements), chemical reaction barriers, noncovalent interactions, and S<sub>N</sub>2 reactions.<sup>275-279</sup>

Another important aspect to take into account when performing DFT calculations (or in *ab initio* calculations) is the correct choice of the basis set. The basis set is the group of mathematical functions from

which the wave function is constructed. There are two types of basis functions commonly used in electronic structure calculations, the Slater-type orbitals<sup>280</sup> and Gaussian-type orbitals.<sup>281</sup> The basis sets may be improved by including certain features such as polarization and diffuse functions. In the present Doctoral Thesis have been used the 6-31G+(d,p) basis set,<sup>282</sup> an accurate basis set for the study of chemical reaction barriers.<sup>283</sup>

### 6.2. Potential Energy Surfaces

The PES are based on the Born-Oppenheimer approximation<sup>284</sup> and is an important concept when carrying out the computational study of a chemical reaction. The exploration of the PES allows obtaining the information of the reaction mechanism. A PES is a representation of the potential energy (electronic and nuclear) depending on the several coordinates of the chemical system; these coordinates depend on the nature of the chemical process that is being studied. The PES has  $3N-6$  (or  $3N-5$ ) coordinate dimensions, being  $N$  the number of atoms. This dimensionality derives from the three-dimensional nature of Cartesian space.

An important concept is necessary to describe prior the study of the PES: the gradient. The gradient ( $\vec{g}_i$ ) is a vector composed of the first derivatives of the potential energy with respect to the nuclear coordinates:

$$\vec{g}_i = \frac{\partial U}{\partial \vec{r}_i} \quad \forall i=1, \dots, N \quad (6.2.1)$$

being  $N$  the number of atoms.

## 6. Theoretical Methods

---

Another crucial concept related with the PES is the Hessian or force constants. The Hessian is a symmetric square matrix. The matrix contains topological information about the system, and its elements are the second derivatives of the energy with respect to the displacement of two coordinates for all the atoms.

$$H_{ij} = \frac{\partial^2 U}{\partial \vec{r}_i \partial \vec{r}_j} \quad (6.2.2)$$

The H matrix is a  $3N \times 3N$  matrix which is symmetric if the particles are the same,  $i$  is equal to  $j$ . When the Hessian is diagonalized,  $|\mathbf{H} - \lambda \mathbf{I}| = 0$ , the eigenvalues equation can be solved:

$$F \vec{u}_i = \lambda \vec{u}_i \quad (6.2.3)$$

In the equation 6.2.3 the term  $F$  is the diagonalized Hessian matrix,  $\vec{u}_i$  is the vector that represents the curvature principal axes and represents the normal modes, and  $\lambda$  represents the eigenvalues. There will be  $3N$  eigenvalues and eigenvectors, being  $N$  the number of atoms in the system.

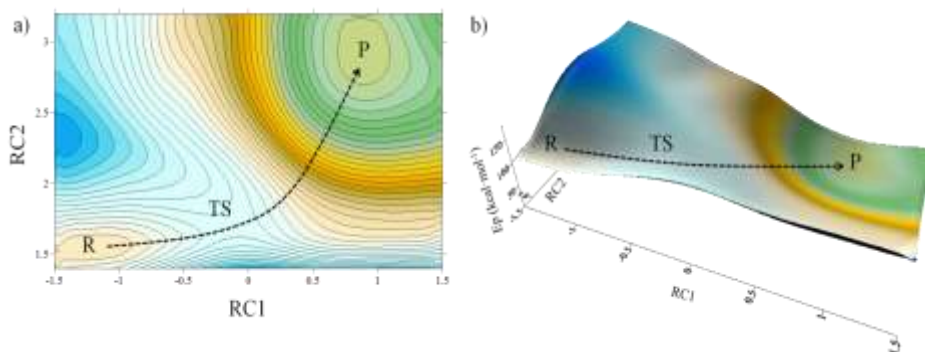
In order to calculate the vibrational frequencies, the Hessian matrix is first mass-weighted:

$$H_{i,j}^m = \frac{H_{ij}}{\sqrt{M_i M_j}} \quad (6.2.4)$$

A PES is a function of the positions of all the atoms in the system. In a system with two atoms, it is easy to visualize that the PES will depend on just one geometrical variable, the bond distance between the two atoms. However, if we consider a system composed of three atoms, there will be at least three independent geometrical parameters. So, this

will lead to a lot more of calculations compared to the bidimensional PES, with which is necessary to choose *distinguished reaction coordinates*,  $\xi_i$ . The  $\xi_i$  are usually geometrical parameters that participate during the chemical reaction (for example bond distances, or bond angles).

On a PES there are stationary points with one important characteristic: the gradient is equal to zero ( $\vec{g}_i = 0$ ). The two more important stationary points of a PES are the minimum and the saddle point (see Figure 6.2.1). The minimum is located when after the diagonalization of the Hessian matrix, all the eigenvalues obtained are positive. This means that an infinitesimal displacement of the geometry of the system along the direction defined by any of its eigenvectors will increase the energy. Examples of the minimum located on the PES are the reactants (R in Figure 6.2.1) and products (P in Figure 6.2.1) of a chemical reaction.



**Figure 6.2.1** PES for a hypothetical reaction. (a) Contour plot, (b) three-dimensional plot. RC1 is the antisymmetric combination of the bond-breaking and bond-forming distances; and RC2 is the bond-breaking distance.



## 6. Theoretical Methods

---

Another important stationary point of a PES is the saddle point (TS in Figure 6.2.1). When the Hessian matrix is diagonalized, all the eigenvectors obtained are positive except for one of them, in this case, the saddle point is first-order and it is a TS of the reaction. The TS is characterized by one imaginary frequency and its corresponding normal mode of vibration is associated with the motion of the atoms along the reaction coordinate. The eigenvector associated to the negative eigenvalue is known as the transition vector and indicates the direction from the transition structure to reactants and products.

There are several methods to locate a minimum of a PES. These methods are classified by the type of derivatives that they use. Some of them are, for example, the first derivatives or first-order methods like the *steepest descent*<sup>255</sup> or the *conjugate gradient*<sup>285</sup> methods. The algorithms that make use the energy values, its first (gradient) and second (Hessian) derivatives as *the Newton-Raphson methods*.<sup>286,287</sup> Or the more efficient quasi-Newton methods as for example the BFGS method<sup>288-290</sup> and the L-BFGS method.<sup>291</sup>

The location of transition states in enzymes is a difficult task, the standard method make use of the Hessian matrix, the algorithms used in small systems are computationally demanding due to they have to calculate and manipulate a complete matrix of secondary derivatives, becomes impractical for large systems due to the large amount of degrees of freedom. Some methods often used in the localization of TS are the adiabatic mapping method,<sup>292</sup> and the conjugate peak refinement algorithm proposed by Karplus and Fisher for locating TS for systems with many atoms.<sup>293</sup> One successful method to localize and characterize

TS in large system described with QM/MM methods is the micro/macro-iteration method.<sup>294-299</sup>

In the micro/macro-iteration method, the system is divided into the control space and into the complementary space. In the control space appears all the geometrical variables with a high contribution to the reaction coordinate, being the Hessian matrix calculated and used to guide the localization of the stationary structure. In the complementary space, the rest of the coordinates appears and at each step of the search this space is minimized using only gradients. Once this partitioning has been done, the saddle point location is carried out in the degrees of freedom of the core, under a fully relaxed environment. Thus, before each gradient or Hessian guided step for the core, the degrees of freedom of the environment are optimized in order to maintain an approximately zero gradient and to minimize the potential energy.

Martí et al.<sup>300</sup> proposed an alternative method for searching stationary structures, a dual level strategy combined with a micro/macroiiteration method. This method is based on a combination of a high level description for the gas-phase energy of the QM subsystem and a low level description for the interaction energy with the MM subsystem. Martí et al. in a computational study of the kinetic isotope effects for the enzyme Chorismate Mutase showed that this method provide a better descriptions than standard QM/MM calculations at the semiempirical level, with a reasonable computational cost.<sup>301</sup>

### 6.3. Free Energy

The calculation of free energy differences is one of the most important magnitudes to be estimated in biomolecular simulations. Free energy calculations using MD provide a direct connection between the microscopic structure and fluctuations of a system and its most important equilibrium thermodynamic properties. In order to obtain reliable free energy estimates from simulations, it is necessary carried out sufficient conformational sampling of, not only the starting and final states, but also of the intermediate states and transition states.

The free energy can be expressed as the Helmholtz function,  $A$ , or the Gibbs function,  $G$ . The Helmholtz free energy is the observable to systems with constant number of particles, temperature and volume (conditions used in our experiments), whereas the Gibbs free energy is for constant number of particles, temperature and pressure. The Helmholtz free energy is defined as:

$$A = -k_B T \ln Z_{NVT} \quad 6.3.1$$

where  $Z_{NVT}$  is the partition function for a system formed by  $N$  indistinguishable particles at a constant volume and temperature (it is the probability to find the system in a certain state), and its define as:

$$Z_{NVT} = \frac{1}{h^{3N} N!} \int dP \int dR e^{-\frac{H(P,R)}{k_B T}} \quad 6.3.2$$

In the equation 6.3.2, the term  $H$  is the Hamiltonian for the system,  $N!$  is a factorial term that takes into account the indistinguishable nature of the  $N$  particles and the term  $h^{3N}$  is the minimum size of the microstate

and is related with the uncertainty principle. For the calculation of the free energy from a simulation, it is necessary rewrite the equation 6.3.1 to obtain the partition function for a system as a function of an average:

$$Z_{NVT} \propto \frac{1}{\langle e^{\frac{v}{k_B T}} \rangle} \quad 6.3.3$$

From the above equation, it might be thought that if we evaluate the average  $\langle e^{\frac{v}{k_B T}} \rangle$  along a MD trajectory, we will obtain the free energy. But this is not the case, because it is very difficult to obtain reliable averages due to the fact that the simulation will preferentially explore the configurations with a large negative potential energies due to these configurations have higher probabilities. Thus, the problem lies in the inability to explore all the configurations leaving the system to evolve freely, the simulation would not be able to access to the structures with a higher potential energy as for example the structures of TS. There are several methods for overcoming the sampling problem like thermodynamic integration,<sup>302</sup> or *umbrella sampling*.<sup>303</sup> The *umbrella sampling* method proposed by J. P. Valleau and G. M Torrie, was the one used in this Doctoral Thesis and it will be analyzed in the section 6.3.2.

### 6.3.1. Molecular Dynamics

A chemical process in condensed media involves a huge number of degrees of freedom and it is necessary to take into account all the conformations to which the system can be accessed at a given temperature. In a reaction that takes place in a condensed media, unlike when it occurs in a vacuum, there are several conformations

## 6. Theoretical Methods

---

corresponding to the reactant or transition structures. The cause is the effect of the environment, there is a wide range of accessible structures which are minimum or saddle points in the PES, with almost similar geometries and energies. Therefore, the properties of the transition state, for example, will be the average properties of all the transition structures. Thus, the evaluation of as many configurations as possible is necessary to obtain average properties of the system, which can be compared with the experimental data.

Thus, it is necessary to run statistical simulations by means of *MD*<sup>304</sup> or *Monte Carlo* simulations.<sup>305</sup> In the present Doctoral Thesis, the method employed to evaluate the contributions of all the conformations to the properties of the system was based on MD simulations.

MD simulations are based in the exploration of the dynamic of the nucleic atomics that form the chemical system. Since there are several formulations for the classical dynamics study of a system, we will adopt the description that begins by defining the classical Hamiltonian of the system ( $\hat{H}$ ),<sup>255</sup>

$$\hat{H}(\vec{p}_i, \vec{r}_i) = \sum_{i=1}^N \frac{1}{2m_i} \vec{p}_i^2 + V(\vec{r}_i) \quad (6.3.1.1)$$

The  $\hat{H}$  is the sum of the kinetic and potential energy terms and it is a function of the 3N particle momenta and the 3N particle positions; the term  $\vec{p}_i$  is the momentum of a particle  $i$ ; and the term  $V$  is the effective potential or in our case the potential energy.

To perform a MD simulation, the Newton's equations must be solved for each particle in the system. Different algorithms have been

developed to integrate numerically the Newton's equations of motion. All these algorithms have to follow two important requirements: to be able to conserve energy and momentum. In addition, they should also allow a long time step to be used for integration and they should be computationally efficient. The most suitable method is the Stoermer's rule method, generally called *Verlet* method.<sup>304</sup>

The *Verlet* method<sup>304</sup> assume that being  $R(t)$  the positions of atoms in the system at time  $t$ , then the new positions of the atoms at a time  $t + \Delta t$ , and the positions from the previous step  $t - \Delta t$ , can be obtained from a Taylor expansion in terms of the time step  $\Delta t$  :

$$R(t + \Delta t) = R(t) + \Delta \dot{R}(t) + \frac{\Delta^2}{2} \ddot{R}(t) + O(\Delta t^3) \quad (6.3.1.2)$$

$$R(t - \Delta t) = R(t) - \Delta \dot{R}(t) + \frac{\Delta^2}{2} \ddot{R}(t) - O(\Delta t^3) \quad (6.3.1.3)$$

Adding these two equations and subtracting the equation 6.3.1.3 from the equation 6.3.1.2 gives:

$$R(t + \Delta t) \approx 2R(t) - R(t - \Delta t) + \Delta^2 M^{-1} F(t) \quad (6.3.1.4)$$

$$V(t) \approx \frac{1}{2\Delta t} (R(t + \Delta t) - R(t - \Delta t)) \quad (6.3.1.5)$$

Both equations 6.3.1.4 and 6.3.1.5 describe the positions of the particles at  $t + \Delta t$  and the velocities of the particles  $V$  at the current time  $t$ , respectively. The main disadvantage of the *Verlet* method lies in that the velocities at a time  $t$  can only be computed once the positions at a time  $t + \Delta t$  have been obtained, so, this method being subject to relative large errors in the evaluation of velocity, and therefore, at the start of the simulation ( $t=0$ ) a different equation is needed.

## 6. Theoretical Methods

---

To solve this disadvantage, the *Velocity Verlet* method<sup>306</sup> appears which use these equations:

$$R(t + \Delta t) = R(t) + \Delta V(t) + \frac{\Delta^2}{2} M^{-1} F(t) \quad (6.3.1.6)$$

$$V(t + \Delta t) = V(t) + \frac{\Delta}{2} M^{-1} (F(t) + F(t + \Delta t)) \quad (6.3.1.7)$$

Whichever method is employed to carry out the MD, it is important that during the MD some properties of the system are conserved, especially the momentum, angular momentum and the energy. Both total momentum ( $M$ ) and angular momentum ( $L$ ) of a system are defined as:

$$M = \sum_{i=1}^N p_i \quad (6.3.1.8)$$

$$L = \sum_{i=1}^N r_i \wedge p_i \quad (6.3.1.9)$$

There are some important subjects to take into account when conducting MD simulations. The first subject is the value of the time-step of the simulation. The factor that limits the value of the time-step is the nature of the highest frequency motions in the system. In general, the time-step must be small enough so that the fastest vibration described can be followed in detail during the integration of the equations of motion. In practice, the time-step selected is 1 fs.

Another subject is to choose the conditions of the MD simulation; these conditions must be similar to the conditions under which the experiments are carried out. In this sense, the thermodynamic ensembles most used are the *canonical* or *NVT ensemble*, the *isothermal-isobaric* or *NPT ensemble* and the *isobaric-isoenthalpic* or *NPH ensemble*.

In the *canonical ensemble*, the ensemble used in the present Doctoral Thesis, the temperature (T), the particle number (N) and the volume (V) are constants, but the total energy of the system may fluctuate. One method to do the simulation in the *canonical ensemble* was proposed initially by Shuichi Nosé<sup>307,308</sup> and later extended by William G. Hoover.<sup>309</sup> The basis of the method is to introduce an external thermal bath to which the system is coupled. According to the Nosé-Hoover method,<sup>307-309</sup> the modified equations of Newton for the system are:

$$\dot{R} = M^{-1}P \quad 6.3.1.10$$

$$\dot{P} = F(R) - \frac{p_n}{Q}P \quad 6.3.1.11$$

$$\dot{n} = \frac{p_n}{Q} \quad 6.3.1.12$$

$$\dot{p}_n = P^T M^{-1}P - N_{df}k_B T \quad 6.3.1.13$$

where  $N_{df}$  is the number of coordinate degrees of freedom, and the parameter Q represents the *mass* of the thermostat which determines the size of the coupling and it has units of mass times length squared. Large values of Q result in a constant-energy simulation, while small values produce large couplings and lead to MD with poor equilibration. The term  $p_n$  (equations 6.3.1.12 and 6.3.1.13) is the momentum variable and from a physical point of view acts as a friction coefficient.

In some cases the Nosé-Hoover method<sup>307-309</sup> is inadequate to control the temperature. Another approach to control the temperature during the simulation is based on the description of the dynamics of a particle interacting with a thermal bath applying the Langevin equation<sup>310</sup> (Paul Langevin). In the Langevin equation appears two extra force terms



arising from the interaction between the particle and the thermal bath: a random force and a frictional force.

### 6.3.2. Umbrella Sampling

One particular type of free energy is the PMF. The PMF can be obtained as a function of one degree of freedom of the system, the distinguished coordinate,  $\xi$ . It is defined as:

$$U(\xi_0) = c - k_B T \ln \left( \int dR \delta(\xi(R) - \xi_0) e^{-\frac{v(R)}{k_B T}} \right) \quad 6.3.2.1$$

In the equation 6.3.2.1 the term  $\xi_0$  is the value of the degree of freedom  $\xi$  for which the PMF is computed, and the term  $c$  is an arbitrary constant which includes all the terms that are independent of  $\xi$ . If the average of the distribution function of the coordinate is expressed as:

$$\langle \rho(\xi_0) \rangle = \frac{\int dR \delta(\xi(R) - \xi_0) e^{-\frac{v(R)}{k_B T}}}{\int dR e^{-\frac{v(R)}{k_B T}}} \quad 6.3.2.2$$

Then the expression of the PMF can be written in terms of the ensemble average of the  $\rho(\xi_0)$ :

$$U(\xi_0) = c' - k_B T \ln \langle \rho(\xi_0) \rangle \quad 6.3.2.3$$

where  $c'$  is another arbitrary constant.

As it was mentioned in section 6.3, it is important during the simulation explore all the configurations, not only the configurations near to a minimum energy. The method used in this Doctoral Thesis to

overcoming this problem is the *umbrella sampling* method.<sup>303</sup> This method is employed for sampling accurately those configurations with an energy barrier significantly larger than  $k_B T$  (high-energy regions). To perform a correct computation of the whole density along  $\xi_0$ , we should make series of calculations for sampling in different regions of the space, but these regions should overlap in order to cover the whole configurational space and make the integration possible. Since, as stated before, it would be easier for the system to access to low energy regions, so, it is necessary to use an umbrella potential that can focus the sampling for a given value of  $\xi_0$ . Then, the trajectories of each window are added to obtain the full distribution function valid for the whole range of the coordinate  $\xi_0$ . Usually, a harmonic potential is chosen as an umbrella potential:

$$V_{umb}(\xi_0) = \frac{1}{2} k_{umb} (\xi_0 - \xi_{ref})^2 \quad 6.3.2.4$$

where the term  $k_{umb}$  is the force constant for the potential and  $\xi_{ref}$  is the reference value of the reaction coordinate which is changed at each window. The principal advantage of the *umbrella sampling* method is the inclusion of all contributions to the free energy, but this require a very large number of energy evaluations, making it an expensive computational method.

Finally, it is necessary to reconstruct the complete distribution functions computed separately for each window, obtaining the averaged free energy of the system. An efficient method to do this important step in an umbrella sampling calculation was proposed by S. Kumar and co-workers, the *weighted histogram analysis method* (WHAM).<sup>311</sup> An

## 6. Theoretical Methods

---

alternative to WHAM for combining the windows in umbrella sampling simulations with harmonic biases is *umbrella integration* developed by W. Thiel and J. Kästner.<sup>312,313</sup> There are some generalizations of the WHAM method, for example, the *dynamic histogram analysis method* (DHAM) proposed by Hummer and Rosta,<sup>314</sup> this method uses transition probabilities in addition to histogram counts along the selected  $\xi$ . Or more recently, the Bayesian treatment of WHAM developed by Andrew L. Ferguson.<sup>315</sup> This method generates statistically optimal FES estimates in any number of biasing dimensions under arbitrary choices of the Bayes prior. The WHAM method was the one used in the present Doctoral Thesis.

The WHAM method constructs an optimal estimate for the average distribution function in the unbiased ensemble from the biased distribution functions for each window:

$$\langle \rho(\xi_0) \rangle = \frac{\sum_{\alpha=1}^{N_w} n_{\alpha} \langle \rho(\xi_0) \rangle_{biased}^{\alpha}}{\sum_{\beta=1}^{N_w} n_{\beta} e^{-\frac{V_{umb}^{\beta}(\xi_0) - F_{\beta}}{k_{\beta} T}}} \quad 6.3.2.5$$

where the term  $N_w$  is the number of windows and  $n_{\alpha}$  is the number of independent data points used for the generation of the distribution function for a window. And the window free energy ( $F_{\alpha}$ ) is determined according to the following equation:

$$e^{-\frac{F_{\alpha}}{k_{\beta} T}} = \int d \xi_0 \langle \rho(\xi_0) \rangle e^{-\frac{V_{umb}^{\alpha}(\xi_0)}{k_{\beta} T}} \quad 6.3.2.6$$

where  $V_{umb}^{\alpha}$  is the umbrella potential. Both equations 6.3.2.5 and 6.3.2.6 must be solved iteratively and allow to calculate the average distribution function ( $\langle \rho(\xi_0) \rangle$ ), and therefore the PMF from a set of window distribution function.

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 methods. In order to correct the low-level energy function used in the PMF calculations, an interpolated correction scheme developed by Moliner and co-workers has been applied.<sup>316</sup> This scheme, based on a method proposed by Truhlar and co-workers for dynamical calculations of gas phase chemical reactions,<sup>317-319</sup> uses a spline under tension<sup>320,321</sup> to interpolate the correction term at any value of the reaction coordinates selected to generate the FESs. Thus, the new energy function for a PMF is defined as:

$$E = E_{LL/MM} + S \left[ \Delta E_{LL}^{HL}(\zeta) \right] \quad (6.3.2.7)$$

$$E = E_{LL/MM} + S \left[ \Delta E_{LL}^{HL}(\zeta_1, \zeta_2) \right] \quad (6.3.2.8)$$

In the equations 6.3.2.7 and 6.3.2.8,  $S$  denotes the spline function, whose argument  $\Delta E_{LL}^{HL}$  is a correction term taken as the difference between single-point calculations of the QM subsystem using a high-level (HL) method, and a low-level (LL) one (in the present Doctoral Thesis the LL is AM1d). The correction term is expressed as a function of one  $\zeta$  or two coordinates  $\zeta_1, \zeta_2$  to generate the higher level monodimensional-PMF or bidimensional-PMF. The HL calculations were

## 6. Theoretical Methods

---

carry out by means of the M06-2X functional<sup>275</sup> with the 6-31+G(d,p) basis set.<sup>282</sup>

### 6.3.3. Transition State Theory

TST can be used to calculate reaction rates from free energy barriers. Thus, it is the basis for comparing the experiments and the calculations in enzymatic reactions. This theory was initially developed in the 30's of last century,<sup>322-324</sup> and is based on the following postulates:<sup>325</sup>

- the reactant region of phase space is populated according to a Maxwell–Boltzmann distribution. Thermal equilibrium between R and TS;
- the Born–Oppenheimer separation of electronic and nuclear motion is valid, so only one PES is involved. Processes in which electronic motion follows nuclear motion without an electronic state change are called electronically adiabatic, which strictly means that the electronic quantum numbers are conserved;
- the nuclear motion is classical;
- and all trajectories crossing the TS toward P must have originated on the reactant side and must cross the hypersurface only once. This condition is now usually called the *no recrossing* assumption. If the reaction coordinate were indeed globally separable, there would be no recrossing. Thus this condition is sometimes replaced by the assumption that the reaction coordinate is separable.

It is possible to estimate the reaction rate constant ( $k_r$ ) if we are able to determine the free energy between reactants and the transition state,  $\Delta G^\ddagger$ ; the rate constant is defined as:

$$k_r(T) = \frac{k_B T}{h} e^{\left(\frac{-\Delta G(T)^\ddagger}{RT}\right)} \quad 6.3.3.1$$

where the term  $\frac{k_B T}{h}$  gives the rate at which the TS goes to products, the exponential factor is the equilibrium constant for the equilibrium between the reactants and the TS,  $k_B$  is the Boltzmann's constant ( $1.38 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$ ),  $h$  is the Planck's constant ( $6.626 \times 10^{-34} \text{ J} \cdot \text{s}$ ),  $R$  is the universal gas constant ( $8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ) and  $T$  is the absolute temperature in Kelvin.

There are three main corrections to the quasiclassical TST, namely non-equilibrium effects, tunnelling effects, and recrossing. Thus, in order to overcome the limitations and to obtain rate constants which can be compared with the experimental ones, we should modify the TST rate constant multiplying it by a transmission factor ( $\Gamma$ ):

$$k_r(T) = \Gamma(T) \frac{k_B T}{h} e^{\left(\frac{-\Delta G^\ddagger}{RT}\right)} \quad 6.3.3.2$$

The transmission factor, determined at a temperature ( $T$ ) and computed on a value of the reaction coordinate ( $\xi$ ) is defined approximately as a product of these factors:<sup>287</sup>

$$\Gamma(\xi, T) = \gamma(\xi, T) \kappa(T) g(\xi, T) \quad 6.3.3.3$$

## ***6. Theoretical Methods***

---

where  $\gamma(\xi, T)$  corrects the effect of the recrossing trajectories,  $\kappa(T)$  correct the tunnelling effects and  $g(\xi, T)$  correct the non-equilibrium effects. It should be mention that these factors have not been evaluated in the present Doctoral Thesis mainly due to the low effect that can have on the TST rate constants for the kind of reactions subject of study.





## 7. Papers

### 7.1. Paper I

Paper entitled: *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Falcipain-2 Inhibition by the Epoxysuccinate E64.*

Kemel Arafet, Silvia Ferrer, Sergio Martí, and Vicent Moliner.

Biochemistry **2014**, 53(20), 3336-3346. DOI: 10.1021/bi500060h.

Reprinted with permission from (DOI: 10.1021/bi500060h). Copyright (2014) American Chemical Society.

## Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Falcipain-2 Inhibition by the Epoxysuccinate E64

Kemel Arafat, Silvia Ferrer,\* Sergio Martí, and Vicent Moliner\*

Departament de Química Física i Anàlisi, Universitat Jaume I, 12071 Castelló, Spain

Supporting Information

**ABSTRACT:** Because of the increasing resistance of malaria parasites to antimalarial drugs, the lack of highly effective vaccines, and an inadequate control of mosquito vectors, the problem is growing, especially in the developing world. New approaches to drug development are consequently required. One of the proteases involved in the degradation of human hemoglobin is named falcipain-2 (FP2), which has emerged as a promising target for the development of novel antimalarial drugs. However, very little is known about the inhibition of FP2. In this paper, the inhibition of FP2 by the epoxysuccinate E64 has been studied by molecular dynamics (MD) simulations using hybrid AM14/MM and M06-2X/MM potentials to obtain a complete picture of the possible free energy reaction paths. A thorough analysis of the reaction mechanism has been conducted to understand the inhibition of FP2 by E64. According to our results, the irreversible attack of Cys42 on E64 can take place on both carbon atoms of the epoxy ring because both processes present similar barriers. While the attack on the C2 atom presents a slightly smaller barrier (12.3 vs 13.6 kcal mol<sup>-1</sup>), the inhibitor–protein complex derived from the attack on C3 appears to be much more stabilized. In contrast to previous hypotheses, our results suggest that residues such as Gln171, Asp170, Gln36, Trp43, Asn81, and even His174 would be anchoring the inhibitor in a proper orientation for the reaction to take place. These results may be useful for the rational design of new compounds with higher inhibitory activity.



**M**alaria is an infectious disease that, in humans, is caused by five species of parasites of the genus *Plasmodium*: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi*. Malaria caused by *P. falciparum* is the most deadly form and that caused by *P. vivax* the most extended; the other three species are found much less frequently. Malaria parasites are transmitted to humans by the bite of infected female mosquitoes of more than 30 anopheline species. Globally, an estimated 3.4 billion people were at risk of malaria in 2012, with children under five years of age and pregnant women most severely affected.<sup>1</sup> Almost half of the world's population permanently lives under risk of infection where malaria is endemic. Because of the increasing resistance of malaria parasites to antimalarial drugs, the lack of highly effective vaccines, and an inadequate control of mosquito vectors, the problem is growing, especially in the developing world. New approaches to drug development are consequently required.<sup>2–4</sup>

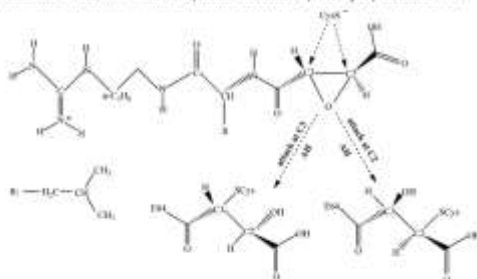
In *P. falciparum*, various proteases catalyze the degradation of human hemoglobin, and the amino acids derived from this process are incorporated into parasite proteins or utilized for energy metabolism.<sup>5</sup> One of the proteases involved in this metabolic process is named falcipain-2 (FP2). FP2, belonging to the family of cysteine proteases (papain-like enzymes known as clan CA), is expressed during the erythrocytic stage of the life cycle of the parasite,<sup>6</sup> and the inhibition of these has proven to be indispensable for the complete blockage of parasitic growth and proliferation.<sup>6</sup> FP2 has emerged as a promising target for the development of novel antimalarial drugs.<sup>7–8</sup>

In the past several years, various types of FP2 inhibitors have been reported to be capable of inactivating the enzyme.<sup>9–15</sup> However, only a few structures reported in the Protein Data Bank correspond to the crystal structure of FP2 with its inhibitors.<sup>16–19</sup> One of them is the crystal structure of FP2 complexed with *N*-[*N*-(1-hydroxycyclohexyl)carboxyl]-leucylaminoethylguanidine, E64 (see Scheme 1), the first epoxysuccinyl discovered and a natural potent and specific irreversible inhibitor of cysteine proteases. One advantage of this type of inhibitors is their stability under physiological conditions toward simple thiols. In addition, although they have limited selectivity toward different cysteine proteases, their reactivity toward cysteine proteases and not other proteases, along with their chemical inactivity, makes this class of inhibitors useful as pharmaceutical agents.<sup>18</sup> E64 was isolated from a culture of *Aspergillus japonicus*,<sup>20</sup> and its structure was determined by Hanada et al. in 1978.<sup>21</sup> E64 contains a *trans* (2*S*,3*S*)-configured epoxide ring, whereas the amino acid residues of the peptidyl part of the inhibitor have the *L*-configuration. The substituents at C2 and C3 on the epoxide ring are in *trans* position to one another,<sup>22,23</sup> which appears to be crucial because the *cis* configuration leads to a total loss of inhibition activity. E64 inhibits cysteine proteases by S-alkylation of the active site cysteine, which results in the

Received: January 14, 2014

Revised: April 25, 2014

Published: May 8, 2014

Scheme 1. General Mechanism of the Inhibition of Cysteine Proteases by the Epoxysuccinate E64<sup>a</sup>

<sup>a</sup>ABE is a source of protons.

opening of the epoxide ring (see Scheme 1) and stabilization of the protein–inhibitor complex.<sup>29,33</sup> Nevertheless, there are some open questions related to the molecular mechanism, because it appears that the attack can take place at either C2 or C3 depending on the orientation of the epoxysuccinate in the active site.<sup>29</sup> On one hand, it has been demonstrated by X-ray investigations that the attack of the alkylation step with epoxide takes place at C2 when an acid is the C2 substituent on this carbon.<sup>17,22–24</sup> The regioselectivity and inhibition potency of epoxide- and aziridine-based inhibitors were studied by theoretical methods by Engeli and co-workers who, on the basis of the exploration of potential energy surfaces (PES) obtained with hybrid quantum mechanics/molecular mechanics (QM/MM) potentials, proposed the attack at C2 as being more favorable.<sup>32</sup> On the other hand, Rihovsky, in an experimental study of the rate and regioselectivity of reactions between 1,3-epoxy carbonyl compounds and methanethiolate in solution, observed that the rate constant for both attacks differs by a factor of only 2, with the C3 attack preferred ( $k_{\text{C3-attack}} = 0.0036 \text{ min}^{-1}$ , and  $k_{\text{C2-attack}} = 0.0064 \text{ min}^{-1}$ ).<sup>35</sup>

An insight into the molecular mechanism of cysteine proteases inhibition by epoxy inhibitors shows that an important feature is the fact that the stereochemistry of the enzyme–inhibitor adduct suffered an inversion of configuration at the reaction site because of a nucleophilic attack by the active site thiolate in an  $S_{\text{N}}2$  reaction.<sup>36</sup> For instance, E64, which has the 2*S*,3*S* configuration before the nucleophilic attack, assumes the 2*R*,3*R* configuration after the covalent bond between the cysteine residue and C2 is formed.<sup>29</sup>

Another matter of debate is the nature of the catalytic acid on the active site. It was initially postulated that when E64 inhibits papain, the oxirane ring would be protonated by His159.<sup>34</sup> Varughese and co-workers<sup>35</sup> rejected this suggestion, on the basis of the crystal structure of papain inhibited by E64. They suggest that the epoxide was more likely to be protonated by a water molecule because of the large distance between the resulting hydroxyl group and His159. Meera and co-workers<sup>37</sup> in mechanistic studies based on the experimental determination of rate constants and the pH dependence of inhibition confirmed that His159 is not necessary for protonation of the epoxide and that water is the predominant source of protons; these results

are likely to be general for all epoxysuccinyl inhibitors with a free carboxylate bound to the C2 atom.<sup>36</sup>

The principal aim of this paper is to gain insights into the inhibition of FP2 by the epoxysuccinate E64 based on molecular dynamics (MD) simulations using hybrid QM/MM methods,<sup>37</sup> which have been shown to be a valuable tool for studying biomolecular systems.<sup>38–40</sup> These calculations allow us to obtain a complete picture of the possible free energy reaction paths, including details of the reaction mechanism, their corresponding free energy barriers and averaged geometries, and interactions between the inhibitor and the protein. Analysis of the results can contribute to our understanding of the inhibition of FP2 by E64, information that can be used to rationally design new inhibitors.

#### ■ COMPUTATIONAL MODEL

The initial coordinates for building our molecular model were taken from the X-ray crystal structure of FP2 inhibited by E64 from *P. falciparum* with PDB entry 3BPF<sup>17</sup> at 2.90 Å resolution. The selected structure contains one polymer of four chains, each chain consisting of 241 amino acids (FP2), the inhibitor E64, 10 water molecules, and one glycerol molecule (see Figure 1). Some missing residues of the structure (Pro189, Leu190, Thr191, Lys192, and Lys193) were added by homology with the crystal structure of FP2 from *P. falciparum* with PDB entry

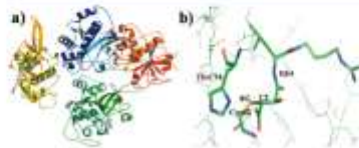
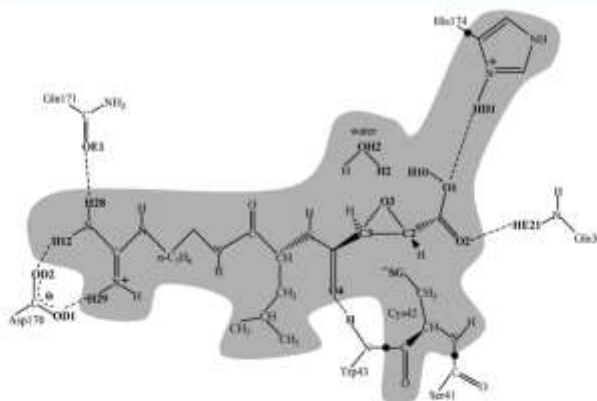


Figure 1. (a) Full tetramer 3BPF structure showing four chains that compose the crystal structure of FP2 inhibited by E64 from *P. falciparum*. (b) Detail of the active site where the inhibitor (C2 atom of E64) binds to the enzyme (SG atom of residue Cys42). E64 and Cys42 are represented as thick sticks.



**Figure 2.** Details of the active site corresponding to the studied model. The gray region corresponds to the QM region, including the inhibitor E64, the side chain of Cys42, a water molecule, and the imidazole ring of His174. The link atoms between the QM and the MM regions are denoted with black dots. The labels of the atoms correspond to the X-ray crystal structure of FP2 inhibited by E64 from *P. falciparum*.

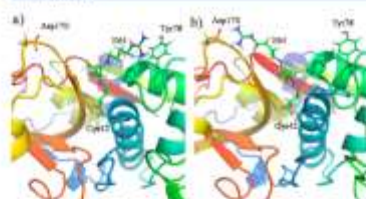
2GHU,<sup>41</sup> in our model, we took only a monomer (chain A) from the tetramer.

Hydrogen atoms were added to the crystal structure of FP2 inhibited by E64 from *P. falciparum* (a productlike structure) using the *ffDYNAMO* library.<sup>42</sup> In this way, the question of the protonation state of Cys42 is initially avoided. The  $pK_a$  values of the structure of PDB entry 3BFF were calculated via the empirical PROPKA 3.1 program of Jensen et al.<sup>43–46</sup> at a physiological pH. According to the results, Glu67, Glu69, and His174 had to be protonated while His19, His27, and His197 were set up as being neutral. A total of six counterions ( $\text{Na}^+$ ) were placed into optimal electrostatic positions around the system, to obtain electroneutrality. Keeping in mind that the digestive vacuole where FP2 is found has a pH of 5.5, we also studied the protonation of the system at this pH to check the influence of pH on the protonation state of the protein. According to PROPKA calculations, a change from pH 7 to 5.5 would affect only four residues: His19, His27, Asp91, and His197. These residues are located 30.7, 28.2, 15.7, and 29.8 Å, respectively, from the C2 atom of the epoxy ring of the inhibitor. These large distances suggest that no significant differences would be expected in the short pH range from 5.5 and 7. Nevertheless, we have performed QM/MM MD simulations on reactant and product states at both protonation states (see Table S1 of the Supporting Information), confirming this assumption and, consequently, indicating that our calculations can be considered as representative of the reaction taking place in the digestive vacuole. Finally, the full system was placed in an orthorhombic box of water molecules (80 Å × 80 Å × 100 Å).

The QM region was described by means of the AM1d semiempirical Hamiltonian,<sup>45</sup> while the rest of the protein and water molecules were described by OPLS-AA<sup>47</sup> and TIP3P<sup>48</sup> force fields, respectively, implemented in the *ffDYNAMO*

library.<sup>42</sup> To saturate the valence of the QM/MM frontier, we used the link atom procedure.<sup>49,50</sup> The number of QM atoms then became 78 and includes the full inhibitor E64, the side chain of Cys42, a water molecule, and the imidazole ring of His174 (see Figure 2). The final system contains a total of 61048 atoms. Because of the size of the system, all residues more than 25 Å from the C1 atom of the inhibitor were kept frozen (54048 atoms of a total of 61048). To treat the nonbonding interactions, a switch function with a cutoff distance in the range of 14.5–18 Å was used. All the QM/MM calculations were conducted using the *ffDYNAMO* library.<sup>42</sup>

The system was relaxed for 700 ps by means of QM/MM MD at 300 K using the NVT ensemble and the Langevin–Verlet integrator. Interestingly, two different conformations, differing in the relative orientation of the E64 in the active site of the protein, were detected: a conformation similar to the initial X-ray structure (PDB entry 3BFF) with the guanidine group of E64 interacting with Tyr78 and a conformation in which the guanidine group of E64 interacts with Asp170 (see Figure 3). Analysis of the time evolution of the potential energy and the total energy obtained during the last 200 ps of the QM/MM MD simulation performed for productlike structures in the two different conformations (see Figure S1 of the Supporting Information) reveals that both structures were stable. Root-mean-square fluctuations of backbone atoms of the protein and atoms of E64 were 0.05 and 0.16 Å for the former and 0.03 and 0.15 Å for the latter, respectively. These results confirm that the system was equilibrated in both conformations. Nevertheless, according to the energetic values, the second structure (Figure 3b) appears to be more stable than the original one. Structures obtained after relaxation were used to compute hybrid AM1d/MM PESs. Stationary structures (including reactants, products, intermediates, and transition state structures) were located and characterized by means of



**Figure 5.** Representative snapshots of the two possible conformations concerning the relative orientation of E64 in the active site. (a) Conformation corresponding to the initial X-ray structure (PDB entry 3BPF) with the guanidine group of E64 interacting with Tyr78. (b) Conformation corresponding to the guanidine group of E64 interacting with Asp170.

the micromacro iterations scheme<sup>41</sup> starting from both productlike conformations. It is important to point out that the reactant state corresponds to the Cys42 as a thiolate anion. According to previous studies based on exploration of QM/MM potential energy surfaces of the mechanism for the addition of the Cys42 thiol to an  $\alpha,\beta$ -unsaturated ether, Gasso et al. showed that activation of Cys42 by His174 is not the rate-limiting step.<sup>42</sup>

Potentials of mean force (PMFs) generated as a function of a distinguished reaction coordinate (RC) were obtained using the weighted histogram analysis method (WHAM) combined with the umbrella sampling approach.<sup>43,44</sup> In particular, the first step of the mechanism involves the attack of sulfur on either C2 or C3 of the epoxide ring, so a one-dimensional PMF (1D-PMF) was obtained as a function of a RC,  $d(CX-O3)-d(SG-CX)$  (with CX being C2 and C3), where 41 and 46 series of simulation windows for the C2 and C3 attacks were required, respectively. After that, the mechanism in which C3 is attacked leads to an intermediate, so two sources for the O3 protonation have been explored: the carboxylic group of the inhibitor and a

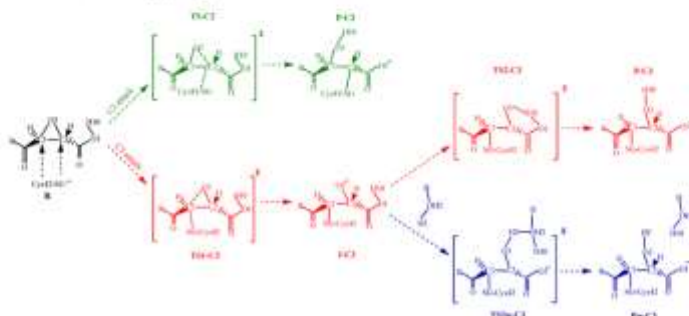
water molecule. In the first case, the 1D-PMF was calculated using  $d(O1-H10)-d(O3-H10)$  as the RC, and in the second case, the 1D-PMF was calculated using  $d(OH2-H2)-d(O3-H2)$  as the RC. This required 87 and 96 series of simulations, respectively. Additionally, to investigate a possible mechanism in which the attack of the proton on O3 of the ring takes place prior to the sulfur attack, we explore three sources of the proton: His174, a water molecule, and an intramolecular transfer from the acid group at C2. The 1D-PMF was calculated using the antisymmetric combination of the  $d(O3-H10)-d(O1-H10)$  distance as the RC for an intramolecular transference, the  $d(O3-H2)-d(OH2-H2)$  distance as the RC for a protonation by a water molecule, and the  $d(O3-HD1)-d(ND1-HD1)$  distance as the RC for a protonation by His174. These 1D-PMFs required 55, 81, and 60 simulation windows, respectively. In all the simulations, the value of the force constant used for the harmonic umbrella sampling was  $2500 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ , and the windows consisted of 20 ps of equilibration and 40 ps of production, with a time step of 1 fs. Afterward, 200 ps of AM1d/MM MD simulations on the stationary points of the free energy surfaces was performed to analyze the average properties of the key states.

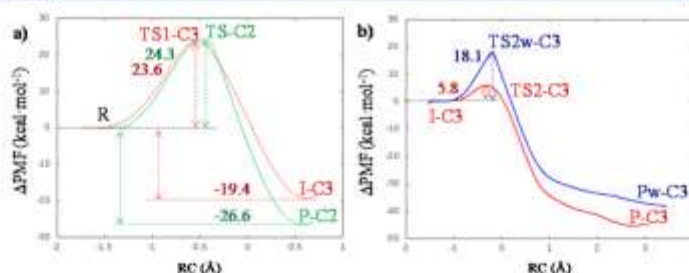
Finally, to correct the low-level AM1d energy function used in the PMF, an interpolated correction scheme developed in our laboratory was applied.<sup>45</sup> In this correction scheme, based on a method proposed by Truhlar and co-workers for dynamical calculations of gas phase chemical reactions,<sup>46</sup> the new energy function employed in the simulation is defined as

$$E = E_{\text{LL/MM}} + S[\Delta E_{\text{LL}}^{\text{HL}}(\zeta)] \quad (1)$$

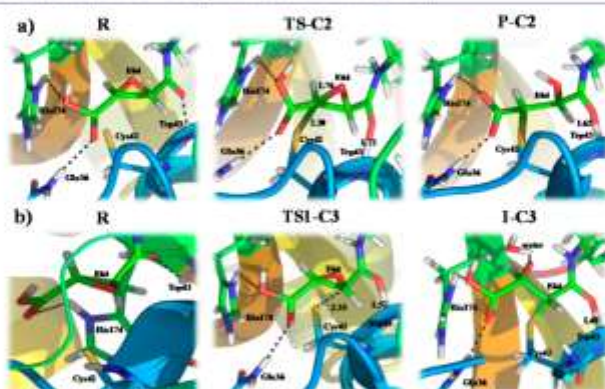
where  $S$  denotes a spline function, whose argument  $\Delta E_{\text{LL}}^{\text{HL}}(\zeta)$  is a correction term taken as the difference between single-point calculations of the QM subsystem using a high-level (HL) method, in this case the hybrid M06-2X<sup>47</sup> with the 6-31+G(d,p)<sup>48</sup> basis set, following the suggestions of Truhlar and co-workers,<sup>47,49</sup> The AM1d Hamiltonian was the low-level (LL) method. The correction term is expressed as a function of the distinguished reaction coordinate  $\zeta$  (RC). These calculations were conducted using Gaussian09.<sup>49</sup>

**Scheme 2.** Proposed Molecular Mechanism As Deduced from the Exploration of the PESs Corresponding to the Attack of Cys42 on C2 or C3 of the Epoxy Ring of the E64 Inhibitor





**Figure 4.** AM1d/MM PMFs of (a) the attack of sulfur on C2 (green line) or C3 (red line) and ring opening. RC corresponds to  $d(CX-O3)-d(SG-CX)$ , with CX being C2 or C3. (b) Transfer of a proton from the carboxylic group to O3 (intramolecular transfer, red line) and from a water molecule (blue line). RC corresponds to  $d(O1-H10)-d(O3-H10)$  for the direct transfer of a proton from the carboxylic group (red line) and  $d(OH2-H2)-d(O3-H2)$  for the transfer of a proton through a water molecule (blue line).



**Figure 5.** Representative snapshots of the key states of the reaction mechanism of inhibition of FP2 by attack of Cys42 on (a) C2 and (b) C3. Average distances are reported in angstroms.

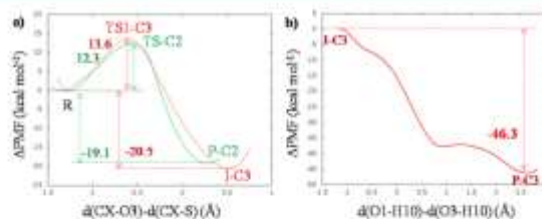
## RESULTS AND DISCUSSION

The mechanism of the inhibition of FP2 by E64 has been studied by exploring the PESs corresponding to the two possible mechanisms, the valine attack of conserved residue Cys42 on C2 or C3 of the epoxide ring (see Scheme 2).

According to the stationary points located on the PESs, the attack of Cys42 on C2 of the epoxy ring of the E64 inhibitor and its protonation would proceed by a single step (see the green path in Scheme 2). In this step, controlled by TS-C2, the SG-C2 bond forms in a concerted fashion with the O3-C2 bond breaking and the transfer of H10 from the carboxylic acid to O3. This was confirmed by tracing the IRC path from the located TS-C2 to reactant and product valleys. On the other hand, the mechanism of the attack on C3 proceeds in a

stepwise manner through a stable intermediate I-C3 where the cysteine is covalently bonded to the inhibitor and O3 is deprotonated (see the red path in Scheme 2). From the I-C3 intermediate, H10 is transferred from the carboxylic acid to O3. Keeping in mind the fact that water molecules are present in the active site of the FP2, we explored this transfer considering two hypotheses: a direct transfer of H10 and a transfer with the participation of a water molecule (colored blue in Scheme 2).

Once the PESs had been fully explored, they were used as the starting point to generate the free energy surfaces that will allow us to gain deeper insight into the mechanisms, as well as kinetic magnitudes that can be directly compared with experimentally measured rate constants. Thus, as explained in the previous section, two 1D-PMFs were obtained for the attack of Cys42



**Figure 6.** Free energy profiles, computed in terms of potential mean force at the M06-2X/6-31+G(d,p) level. (a) Attack of sulfur on C2 (green line) or C3 (red line) and ring opening. (b) Transfer of a proton from the carboxylic group to O3.

on C2 or C3 of the epoxide ring as a function of an antisymmetric combination of the distances defining the breaking and forming bonds,  $d(CX-O3)-d(SG-CX)$  with CX being C2 and C3, depending on the reaction. The PMF of the step corresponding to the transformation from I-C3 to products was calculated using the antisymmetric combination  $d(O1-H10)-d(O3-H10)$  as the reaction coordinate for the direct transfer from the carboxylic group, and using the antisymmetric combination  $d(OH2-H2)-d(O3-H2)$  as the reaction coordinate to explore the possibility of a water molecule being the source of protonation of O3. According to the AM1/d/MM results, presented in Figure 4, the free energy barrier of the attack of Cys42 on C2 (24.3 kcal/mol) is very close to the free energy barrier of the first step of the mechanism of the attack on C3 (23.6 kcal/mol). The corresponding free energy profiles obtained for the conformation of E64 interacting with Tyr78 (Figure 3a) render significantly higher barriers (see Figure S2 of the Supporting Information). This, together with previous analysis of energies of both conformations, suggests that the conformation with E64 interacting with Asp170 appears to be the most stable one. Thus, all simulations and analysis were performed on the basis of this structure after its determination. Snapshots of representative average structures of the states appearing along the different reaction paths are shown in Figure 5 (snapshots of representative structures of the rest of the stationary points are shown in Figures S3 and S4 of the Supporting Information). The free energy barrier of the transformation from I-C3 to P-C3 is much lower and, considering the exothermic character of the full process. As observed in Figure 4b, comparison of free energy barriers of this second step reveals that the transfer of H10 from the carboxylic group to O3 of the epoxy ring will clearly occur in a direct way without participation of a water molecule (5.8 kcal/mol vs 18.1 kcal/mol).

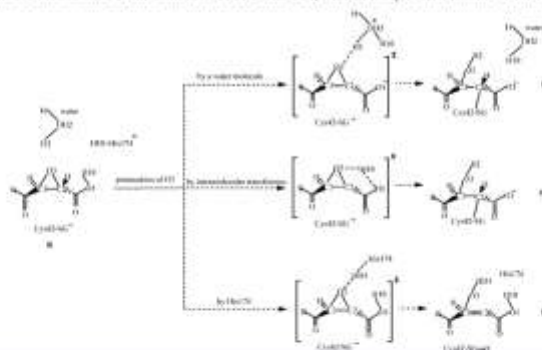
Thus, according to the relative energies of TS-C2 and TS1-C3, our results would be in good agreement with the experimental data of Bihosky.<sup>19</sup> Using the transition state theory at 300 K, his values of the rate constant ( $k_{C2-attack} = 0.0036 \text{ min}^{-1}$ , and  $k_{C3-attack} = 0.0064 \text{ min}^{-1}$ ) can be translated into almost indistinguishable free energy barriers, 23.4 and 23.0 kcal/mol, respectively, which in turn are almost coincident with our predictions for TS-C2 (24.3 kcal/mol) and TS1-C3 (23.6 kcal/mol). Nevertheless, we must keep in mind the fact that the experiments of Bihosky were conducted with nonpeptidic compounds in solution with only one substituent on C2. On

the other hand, because the semiempirical Hamiltonian used in our calculations can render inaccurate barriers, a correction of the profile at the DFT level of theory with the M06-2X functional was conducted (see the previous section for details). The resulting corrected profiles, presented in Figure 6, show a dramatic reduction of the barrier to 12.3 and 13.6 kcal/mol for the attack on C2 and C3, respectively, and a barrier-less process from I-C3 to products. Again, we predict similar DFT/MM free energy barriers for the rate-limiting step of both reactions, in agreement with the very similar rate constants measured by Bihosky in solution.<sup>19</sup> The difference in free energy barriers for attack on C2 and C3 is not as dramatic as the difference in the potential energy barriers deduced from the PESs obtained by Engels and co-workers, who obtained reaction barriers of 1 and 17 kcal/mol for the attack of cysteine on C2 and C3 of the epoxy ring of the E64 inhibitor, respectively.<sup>21</sup>

With regard to the second step of the mechanism of the attack of sulfur on C2, corrections were made at the M06-2X/MM level only for the direct protonation step of O3 from carboxylate group because it was clearly favored with respect to the transfer through a water molecule. In this profile (Figure 6b), a shoulder is observed in the region of products corresponding to a change in the new O3-H10 hydroxyl group. This, initially oriented to the carboxylate group of the inhibitor, points to Asn181 in the most stable product conformation. Keeping in mind that our reactions were conducted in the enzyme, we expected a significant reduction of the barrier by comparison with results obtained in solution. Importantly, our new results predict a very small preference for the attack at C2 over that at C3. Our results differ significantly from those of previous computational studies of Engels and co-workers who, based on QM/MM potential energy profiles, obtained a barrier for the C3 attack of  $\sim 15$  kcal/mol, but an almost barrier-less process for the attack on C2 of the E64c inhibitor. Nevertheless, from a thermodynamic point of view, our calculations predict a more stable enzyme-inhibitor complex when Cys42 attacks C3 than when it attacks C2, which can be deduced from the reaction free energies derived from the profiles shown in Figures 4 and 6.

An alternative mechanism of the inhibition of EP2 with E64 consisting of a protonation of the epoxy O3 atom prior to the Cys42 attack and the epoxy ring opening has been also explored. This alternative route was studied just for the attack on C2 because this is the mechanism presenting a lower barrier. Three possible sources of protons have been considered: a water molecule, protonated His174, and a direct intramolecular

Scheme 3. Alternative Mechanisms of the Inhibition of FP2 with E64, with an Early Proton Attack on O3 of the Epoxide Ring

Table 1. Averaged AM1d/MM Distances Obtained for Key States Located along the Reaction Mechanism of Inhibition of FP2 by Attack of Cys42 on C2 (a) and C3 (b)<sup>a</sup>

	(a) Attack of Sulfur on C2 of the E64 Epoxy Ring				
	R	TSS-C2	P-C2		
SU-C2	2.83 ± 0.03	2.30 ± 0.03	1.80 ± 0.03		
O3-C2	1.44 ± 0.03	1.78 ± 0.04	2.37 ± 0.03		
O3-H10	3.82 ± 0.41	2.74 ± 0.43	0.97 ± 0.03		
O3-H10	0.96 ± 0.02	0.98 ± 0.03	4.09 ± 0.39		
O3-HD1 (His174)	2.61 ± 0.26	2.85 ± 0.20	2.55 ± 0.23		
O2-HB21 (Glu36)	3.47 ± 0.47	2.53 ± 0.24	2.25 ± 0.25		
O3-H (His3)	2.55 ± 0.18	2.54 ± 0.13	2.79 ± 0.13		
O3-H (His3)	2.85 ± 0.18	2.81 ± 0.13	2.81 ± 0.12		
O3-H (His3)	3.02 ± 0.30	2.80 ± 0.23	2.63 ± 0.13		
O4-H (Trp4)	3.21 ± 0.78	1.72 ± 0.27	1.62 ± 0.27		
H10-O (Asn8)	4.13 ± 0.37	4.17 ± 0.36	2.38 ± 0.38		
H28-OD1 (Asp70)	1.84 ± 0.10	1.84 ± 0.10	1.65 ± 0.10		
H12-OD3 (Asp70)	1.82 ± 0.09	1.84 ± 0.09	1.66 ± 0.10		
H28-OB1 (Glu71)	3.18 ± 0.19	3.15 ± 0.19	2.18 ± 0.19		
	(b) Attack of Sulfur on C3 of the E64 Epoxy Ring				
	R	TSS-C3	P-C3	TSS-C1	P-C1
SU-C3	3.03 ± 0.04	2.23 ± 0.03	1.81 ± 0.04	1.81 ± 0.04	1.89 ± 0.03
O3-C3	1.44 ± 0.03	1.89 ± 0.03	2.42 ± 0.03	2.28 ± 0.03	2.42 ± 0.03
O3-H10	3.24 ± 0.16	2.87 ± 0.27	2.47 ± 0.31	1.41 ± 0.03	0.97 ± 0.03
O3-H10	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.03	1.12 ± 0.02	3.74 ± 0.21
O3-HD1 (His174)	3.16 ± 0.34	2.89 ± 0.18	2.86 ± 0.31	4.26 ± 0.27	2.57 ± 0.21
O2-HB21 (Glu36)	4.44 ± 0.77	2.60 ± 0.23	2.13 ± 0.18	3.23 ± 0.67	3.23 ± 0.31
O3-HD1 (His174)	5.02 ± 0.37	4.00 ± 0.28	3.81 ± 0.27	2.85 ± 0.38	4.08 ± 0.20
O3-H (His3)	3.50 ± 0.60	2.76 ± 0.13	2.81 ± 0.20	3.18 ± 0.78	2.92 ± 0.24
O3-H (His3)	3.92 ± 0.33	2.73 ± 0.19	2.83 ± 0.14	3.21 ± 0.44	2.89 ± 0.15
O3-H (His3)	4.13 ± 0.34	2.72 ± 0.19	2.64 ± 0.24	2.97 ± 0.28	2.38 ± 0.23
O4-H (Trp4)	4.97 ± 1.01	1.37 ± 0.24	1.49 ± 0.25	3.39 ± 1.71	1.59 ± 0.28
O4-H (Asp73)	2.78 ± 0.78	4.40 ± 0.34	4.48 ± 0.39	3.12 ± 1.05	4.58 ± 0.37
H10-O (Asn8)	3.80 ± 1.89	3.30 ± 1.09	3.89 ± 0.56	3.43 ± 0.52	2.43 ± 0.18
H28-OD1 (Asp70)	1.83 ± 0.10	1.84 ± 0.10	1.86 ± 0.10	1.65 ± 0.10	1.45 ± 0.10
H12-OD3 (Asp70)	1.82 ± 0.09	1.85 ± 0.09	1.88 ± 0.10	1.66 ± 0.10	1.67 ± 0.10
H28-OB1 (Glu71)	2.38 ± 0.19	2.15 ± 0.19	2.17 ± 0.19	2.17 ± 0.19	2.15 ± 0.19

<sup>a</sup>All values are in angstroms and were obtained from 200 ps of AM1d/MM MD simulations on the stationary points of the free energy surfaces.



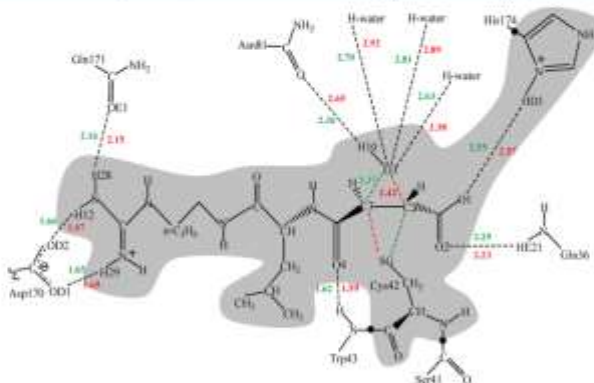
**Table 2.** Averaged AM1d/MM Mulliken Charges Obtained for Key States Located along the Reaction Mechanism of the Inhibition of FP2 by Attack of Sulfur of Cys42 on C2 (a) and C3 (b)<sup>a</sup>

(a) Attack of Sulfur on C2 of the E64 Epoxy Ring			
Mulliken charge	R	TS-C2	P-C2
C2	-0.094 ± 0.028	-0.033 ± 0.024	-0.481 ± 0.033
C3	-0.093 ± 0.021	0.031 ± 0.027	-0.021 ± 0.017
SC1	-0.362 ± 0.383	-0.000 ± 0.061	0.267 ± 0.047
O3	-0.207 ± 0.021	-0.346 ± 0.030	-0.367 ± 0.017
H10	0.238 ± 0.016	0.238 ± 0.018	0.270 ± 0.018
O1	-0.488 ± 0.027	-0.479 ± 0.023	-0.600 ± 0.033
O2	-0.513 ± 0.027	-0.416 ± 0.027	-0.635 ± 0.038
O4	-0.429 ± 0.030	-0.518 ± 0.034	-0.579 ± 0.038

(b) Attack of Sulfur on C3 of the E64 Epoxy Ring					
Mulliken charge	R	TS-C3	I-C3	TS-C3	P-C3
C2	-0.132 ± 0.024	0.069 ± 0.040	0.224 ± 0.029	0.092 ± 0.033	-0.032 ± 0.019
C3	-0.125 ± 0.038	-0.018 ± 0.026	-0.342 ± 0.043	-0.506 ± 0.043	-0.132 ± 0.033
SC1	-0.800 ± 0.061	-0.367 ± 0.066	0.119 ± 0.051	0.259 ± 0.050	0.271 ± 0.044
O3	-0.256 ± 0.023	-0.332 ± 0.023	-0.606 ± 0.033	-0.667 ± 0.016	-0.356 ± 0.017
H10	0.205 ± 0.015	0.293 ± 0.011	0.319 ± 0.010	0.349 ± 0.009	0.283 ± 0.010
O1	-0.436 ± 0.030	-0.493 ± 0.030	-0.494 ± 0.029	-0.521 ± 0.026	-0.697 ± 0.022
O2	-0.241 ± 0.042	-0.427 ± 0.025	-0.378 ± 0.032	-0.620 ± 0.029	-0.620 ± 0.023
O4	-0.423 ± 0.026	-0.316 ± 0.036	-0.423 ± 0.026	-0.563 ± 0.024	-0.336 ± 0.042

<sup>a</sup>All values were obtained from 200 ps of AM1d/MM MD simulations on the stationary points of the free energy surfaces.



**Figure 7.** Schematic representation of the interactions between QM and MM parts obtained in the inhibitor-protein complex: P-C2 (green numbers) and P-C3 (red numbers). All interaction distances are in angstroms. Only distances of <3.00 Å are reported.

transfer of a proton from the carboxylic group of the inhibitor (see Scheme 3). After the required calculations had been performed to determine the free energy profiles in terms of 1D-PMPs, the corresponding free energy barriers are 65.9, 45.4, and 47.4 kcal/mol, respectively (see full PMPs in Figure S5 of the Supporting Information). Consequently, this proposed mechanism does not seem to be feasible.

Once the energetics of possible mechanisms of the inhibition of FP2 with E64 by the attack of Cys42 on C2 or C3 of the inhibitor epoxy ring had been discussed, geometrical and charge population analysis could be conducted to gain deeper insight

into the origins of the formation of the FP2-E64 complex. Thus, average values of key interatomic distances of the different states located along the reaction paths on both reactions are listed in Table 1, and Mulliken charges on key atoms are listed in Table 2. A schematic representation of the inhibitor-protein complex structures obtained in the exploration of both mechanisms (P-C2 and P-C3) is depicted in Figure 7. Values of the second step of the attack of Cys42 on C3, the protonation of O3 by a water molecule, are reported as Supporting Information (Table S2).

The first conclusion that can be drawn from values listed in Table 1 is that the distance between the sulfur atom of Cys42 and the carbon atom of the inhibitor, C2 or C3, is equivalent in the TSs with an uncertain deviation:  $2.30 \pm 0.03$  and  $2.33 \pm 0.03$  Å for TS-C2 and TS1-C3, respectively. Nevertheless, ring opening is in a more advanced stage of the reaction in the latter as deduced from the distance between the epoxy ring oxygen atom and the attacked carbon atom (O3–C3 and O3–C2 distances of 1.89 and 1.76 Å, respectively). This is related to the distance between the O3 atom and the hydrogen atom initially located on the carboxylate group of the inhibitor, H10; a shorter distance is observed in TS-C2 (2.74 Å) than in TS1-C3 (2.87 Å). A longer O3–H10 distance in the latter can explain why intermediate I-C3 is located in the C3 attack mechanism with a remarkable negatively charged O3 atom ( $-0.606$  au). It is also important to point out the slightly higher negative charge developed on O3 in TS-C2 ( $-0.546$  au) by comparison with that in TS1-C3 ( $-0.532$  au), which could help explain why the H10 atom is spontaneously transferred from TS-C2 to products in the C2 attack mechanism. The negatively charged O3 atom is stabilized, both in the TSs and in intermediate II-C3, through hydrogen bond interactions with water molecules of the active site, as shown in Table 1. There are other interesting interactions between the inhibitor and the residues of the active site that are maintained during the full process, such as those with Asp170 and Gln171. These residues would then contribute to anchor the inhibitor to the active site. This is especially important when the covalent bond is not yet formed (reactants). The interactions between O1 and O2 of the inhibitor and His174 and Gln36 are also maintained during the reaction, becoming more intense (shorter distances in Table 1) when the group is deprotonated (products). The large deviations observed in these distances, together with the change in the interaction of O2 in TS2-C3, reveal the flexibility of the active site, thus justifying the statistical simulations performed in this study. Analysis of the time-dependent evolution of structures appearing on the QM/MM MD simulation on TS2-C3 shows that O2 of the carboxylate group can interact, alternatively, with Gln36 or His174 (see Figure S6 of the Supporting Information). Trp43 is also interacting with the inhibitor through the carbonyl O4 atom, although in some of the structures appearing in the mechanism of the attack on C3. Because of the flexibility of the active site (see the standard deviations), the interaction with Trp43 is replaced by an interaction with the polar hydrogen atom of the backbone of Ala175. Finally, as mentioned above, a new interaction is observed in products between Asn81 and the new O3–H10 hydroxyl group, as a result of the rotation of this group once the proton is completely transferred to the oxygen epoxy ring.

## CONCLUSIONS

In this paper, the inhibition of FP2 by the epoxy succinate E64 has been studied by means of MD simulations using hybrid AM1d/MM and M06-2X/MM potentials. The results have allowed us to obtain a complete picture of the possible reaction paths corresponding to the irreversible attack of Cys42 on the carbon atoms of the epoxy ring of the E64 inhibitor. According to our results, the attack of Cys42 on C2 would take place in a single step, while the attack on C3 is described as a two-step mechanism, the first step corresponding to formation of the sulfur–carbon bond, which is the rate-limiting step. In the second step of this mechanism, a hydrogen atom from the

carboxylic group of the inhibitor is finally transferred to the oxygen atom of the opened epoxy ring. This step, when computed at the DFT/MM level, becomes a barrier-less process. Thus, the free energy barriers of the first mechanism and the rate-limiting step of the second are almost equivalent (12.3 and 13.6 kcal/mol for the attack on C2 and C3, respectively), which is in agreement with experimental data. Nevertheless, the inhibitor–protein complex derived from the attack on C3 appears to be much more stabilized. Analysis of averaged geometries and charges of stationary points on the QM/MM free energy surfaces has been used to rationalize the different reaction mechanisms. A thorough analysis of the inhibitor–protein interactions reveals the importance of some of the residues, such as Gln171, Asp170, Gln36, Trp43, Asn81, and His174, in anchoring the inhibitor in a proper orientation for the reaction to take place.

Any attempt to explore an alternative mechanism implying the protonation of the epoxy ring prior to the attack of cysteine on C2 or C3 of the ring, by His174, by a water molecule, or through an intramolecular transfer of a proton from the acid group at C2 as an acid, as suggested in the literature, has been unsuccessful. Finally, it is important to stress that our simulations have allowed detection of two different conformations of a productlike structure, differing in the relative orientation of E64 in the active site of the protein: a conformation similar to the initial X-ray structure (PDB entry 3BPF) with the guanidine group of E64 interacting with Trp74 and a conformation in which the guanidine group of E64 interacts with Asp170. Analysis of our results reveals that, despite both structures being stable, the latter appears to be energetically favored. The flexible character of the system appears to be demonstrated, thus justifying the statistical simulations conducted in this work. Thus, these results may be useful for the rational design of new compounds with higher inhibitory activity against malaria parasites.

## ASSOCIATED CONTENT

### Supporting Information

Averaged AM1d/MM distances obtained from 200 ps of QM/MM MD simulation for the reactant state and product state of the reaction mechanism of inhibition of FP2 by attack of Cys42 on C2 at pH 5.5 and 7 (Table S1), averaged AM1d/MM distances obtained for key states located along the reaction mechanism of the second step of the attack of Cys42 on C3 (Table S2), time evolution of the potential energy and total energy obtained along 200 ps of the QM/MM MD simulation performed for productlike structures in two different conformations (Figure S1), AM1d/MM PMFs of the attack of sulfur on C2 (green line) or C3 (red line) obtained after 200 ps of MD starting from the X-ray conformation (Figure S2), representative snapshots of the key states of the reaction mechanism of the second step of the attack of Cys42 on C3 (protonation of O3 by intramolecular proton transfer) (Figure S3), representative snapshots of the key states of the reaction mechanism of the second step of the attack of Cys42 on C3 (protonation of O3 with participation of a water molecule) (Figure S4), PMFs of the alternative mechanisms of the inhibition of FP2 with E64, with an early transfer of a proton to O3 of the epoxy ring prior to the Cys42 attack (Figure S5), and time-dependent evolution of selected distances on the QM/MM MD simulation constrained on TS2-C3 (Figure S6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

## Corresponding Authors

\*E-mail: sferrer@uji.es.

\*E-mail: moliner@uji.es.

## Author Contributions

K.A. performed the bulk of computational simulations. S.M. wrote the codes and guided some of the calculations. S.M., S.F., and V.M. designed the research. K.A., S.F., and V.M. cowrote the first version of the manuscript. All the authors analyzed data, and V.M. wrote the final version of the manuscript.

## Funding

This work was supported by the Spanish Ministerio de Economía y Competitividad (ref. CTQ2012-36253-C03-01), Generalitat Valenciana (ref. Prometeo/2009/053), Universitat Jaume I (ref P1-IB2011-23), and a predoctoral contract to K.A. from the Spanish Ministerio de Economía y Competitividad.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge the Servei d'Informàtica, Universitat Jaume I, for the generous allotment of computer time.

## REFERENCES

1. [http://www.who.int/malaria/publications/world\\_malaria\\_report\\_2013/report/en/index.html](http://www.who.int/malaria/publications/world_malaria_report_2013/report/en/index.html) (2013).
2. Jara, S., and Palani, J. (2007) Novel molecular targets for antimalarial chemotherapy. *Int. J. Antimicrob. Agents* 30, 4–10.
3. Miel, A. (2007) Recent advances in antimalarial compounds and their patents. *Curr. Med. Chem.* 14, 759–773.
4. Miller, L. H., Baruch, D. I., Marsh, K., and Durrant, O. R. (2002) The pathogenic basis of malaria. *Nature* 415, 673–679.
5. Singh, A., and Rosenthal, P. J. (2004) Selection of cysteine protease inhibitor-resistant malaria parasites is accompanied by amplification of falcipain genes and alteration in inhibitor transport. *J. Biol. Chem.* 279, 35234–35241.
6. Shrestha, B. R., Sirevali, P. S., Singh, A., and Rosenthal, P. J. (2000) Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of *Plasmodium falciparum*. *J. Biol. Chem.* 275, 29000–29010.
7. Sijwal, P. S., Shrestha, B. R., Gai, J., Singh, A., and Rosenthal, P. J. (2001) Expression and characterization of the *Plasmodium falciparum* hemoglobinase falcipain-3. *Biochem. J.* 360, 481–489.
8. Marco, M., and Miguel Cotreras, J. (2012) Falcipain inhibition as a Promising Antimalarial Target. *Curr. Top. Med. Chem.* 12, 408–444.
9. Ehrlich, V., Herold, C., Rottmann, M., Freymond, C., Schweizer, W. B., Fran, B., Stich, A., Schrammer, T., and Diederich, F. (2011) Potent and Selective Inhibition of Cysteine Proteases from *Plasmodium falciparum* and *Trypanosoma brucei*. *ChemMedChem* 6, 273–278.
10. Ehrlich, V., Kichenmüller, F., Herold, C., Cui, K., Huang, J., Schrammer, T., and Diederich, F. (2011) Peptidomimetic nitriles as selective inhibitors for the malarial cysteine protease falcipain-2. *MedChemComm* 2, 800–804.
11. Erturk, R., Bova, F., Zappala, M., Grasso, S., and Micale, N. (2010) Falcipain-2 Inhibitors. *Med. Res. Rev.* 30, 136–167.
12. Erturk, R., Micale, N., Grasso, S., Bova, F., Schrammer, T., Grasso, S., and Zappala, M. (2012) Synthesis and Molecular Modeling Studies of Derivatives of a Highly Potent Peptidomimetic Vinyl Ester as Falcipain-2 Inhibitors. *ChemMedChem* 7, 1594–1600.
13. Erturk, R., Nisi, E., Di Francesco, M. E., Micale, N., Grasso, S., Zappala, M., Vekil, B., and Schrammer, T. (2008) Nonpeptidic vinyl and allyl phosphonates as falcipain-2 inhibitors. *ChemMedChem* 3, 1030–1033.

14. Erturk, R., Zappala, M., Micale, N., Grasso, S., Grefe, S., Schrammer, T., and Grasso, S. (2011) Peptidomimetics containing a vinyl ketone warhead as falcipain-2 inhibitors. *Eur. J. Med. Chem.* 46, 2058–2065.
15. González, F. V., Inquandó, J., Rodríguez, S., McKerron, J. H., and Hansell, E. (2007) Dipeptidyl- $\alpha,\beta$ -epoxysuccinates as potent irreversible inhibitors of the cysteine proteases *craxin* and *thelaxin*. *Bioorg. Med. Chem. Lett.* 17, 6597–6700.
16. Hansens, G., Heilmann, A., Wirt, T., Li, H., Jiang, H., Shen, X., Haasler, V. T., Benzoniang, A., and Hilgerfeld, R. (2011) Structural Basis for the Regulation of Cysteine-Protease Activity by a New Class of Protease Inhibitors in *Plasmodium*. *Structure* 19, 910–929.
17. Kern, I. D., Lee, J. H., Pardee, K. C., Harrison, A., Sgadi, M., Rosenthal, P. J., and Bitens, L. S. (2009) Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: Implications for substrate specificity. *J. Med. Chem.* 52, 852–857.
18. Wang, S. X., Paroley, K. C., Sussosa, J. R., Sijwal, P. S., Kortmann, T., Bitens, L. S., Hemrick, R. J., Rosenthal, P. J., and McKerron, J. H. (2006) Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11503–11508.
19. Suzuki, K., Tsuji, S., and Ishiura, S. (1981) Effect of  $Ca^{2+}$  on the inhibition of calcium-activated neutral protease by isopropyl, anti-pain and spirocycloacetate derivatives. *FEBS Lett.* 136, 119–122.
20. Hanada, K., Tamao, M., Yawaguchi, M., Okamura, S., Sasaki, J., and Tanaka, I. (1978) Isolation and characterization of E-64, a new thiol protease inhibitor. *Agric. Biol. Chem.* 43, 523–528.
21. Hanada, K., Tamao, M., Okamura, S., Sawada, J., Seki, T., and Tanaka, I. (1978) Structure and synthesis of E-64, a new thiol protease inhibitor. *Agric. Biol. Chem.* 42, 529–536.
22. Janes, K. E., Aguiar, J. L., Li, Z. Z., Eicki, O. D., Rubio, J. R., Mikolajczyk, J., Saltsman, G. S., and Powers, J. C. (2004) Design, synthesis, and evaluation of  $\alpha$ -zeta-peptide epoxides as selective and potent inhibitors of caspases-1, -3, -6, and-8. *J. Med. Chem.* 47, 1553–1574.
23. Rosah, W. R., González, F. V., McKerron, J. H., and Hansell, E. (1998) Design and synthesis of diphenyl  $\alpha/\beta'$  epoxy ketones, potent irreversible inhibitors of the cysteine protease *craxin*. *Bioorg. Med. Chem. Lett.* 8, 2809–2812.
24. Matsumoto, K., Yamamoto, D., Ohishi, H., Tomono, K., Ishida, T., Inoue, M., Sadaizumi, T., Kitamura, K., and Mizuno, H. (1989) Mode of binding of E64-C, a potent thiol protease inhibitor, to papain as determined by X-ray crystal analysis of the complex. *FEBS Lett.* 245, 179–180.
25. Yabe, Y., Guillaumet, D., and Rich, D. H. (1988) Irreversible inhibition of papain by spirocycloxy peptides. Carboxyl NMR characterization of the site of alkylation. *J. Am. Chem. Soc.* 110, 4043–4044.
26. Powers, J. C., Aguiar, J. L., Eicki, O. D., and Janes, K. E. (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* 102, 4639–4730.
27. Ghosh, R., Chakraborty, S., Chakrabarti, C., Dattagupta, J. K., and Bhowmik, S. (2008) Structural insights into the substrate specificity and activity of *craxin* and the papain-like cysteine proteases from a tropical plant, *Ervatamia corvorea*. *FEBS J.* 275, 431–434.
28. Kim, M. J., Yamamoto, D., Matsumoto, K., Inoue, M., Ishida, T., Mizuno, H., Saito, S., and Kitamura, K. (1992) Crystal-structure of papain-E64-C complex-binding diversity of E64-C to papain S(2) and S(3) subsites. *Biochem. J.* 287, 797–803.
29. Yamamoto, A., Tomono, K., Matsugi, W., Hara, T., In, Y., Mizuta, M., Kitamura, K., and Ishida, T. (2002) Structural basis for development of cathepsin B-specific noncovalent-type inhibitor: Crystal structure of cathepsin B-E64c complex. *Biochim. Biophys. Acta* 1597, 244–251.
30. Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K., and Mizuno, H. (1999) Refined X-ray structure of papain E-64-C complex at 2.1-Å resolution. *J. Biol. Chem.* 274, 14771–14777.

- (31) Zhao, B. G.; Janson, C. A.; Armpstead, B. Y.; Dalemis, K.; Griffin, C.; Hanning, C. R.; Jonas, C.; Kuzhila, J.; McQuarney, M.; Qiu, X. Y.; Smith, W. W.; and AbdelMegal, S. S. (1997) Crystal structure of human osteocalcin cathepsin K complex with E-64. *Nat. Struct. Biol.* 4, 109–111.
- (32) Mladinovic, M.; Jarsch, K.; Fink, B. F.; Thiel, W.; Schramster, T.; and Ergler, B. (2008) Atomistic insights into the inhibition of cytosine promutase: First QM/MM calculations clarifying the regioselectivity and the inhibition potency of epoxide- and amine-loaded inhibitors. *J. Phys. Chem. B* 112, 5458–5469.
- (33) Ibrivsky, R. (1992) Reactions of  $\alpha,\beta$ -epoxy carbonyl compounds with methanedisulfate-regioselectivity and rate. *J. Org. Chem.* 57, 1029–1031.
- (34) Rich, D. H. (1986) *Proteases Inhibitors*. Elsevier, Amsterdam.
- (35) Varughese, K. I.; Ahmed, F. R.; Carey, P. R.; Houslin, S.; Haber, C. P.; and Steer, A. C. (1999) Crystal structure of a papain-structure of a papain-E64 complex. *Biochemistry* 28, 1330–1332.
- (36) Meaza, J. P., and Rich, D. H. (1996) Mechanistic studies on the inactivation of papain by epoxysuccinyl inhibitors. *J. Mol. Chem.* 39, 3357–3366.
- (37) Warshel, A., and Levitt, M. (1976) Theoretical studies of enzymic reactions. Dielectric, electrostatic and steric stabilization of carbonium-ion reaction transition. *J. Mol. Biol.* 103, 227–249.
- (38) Martí, S.; Rozas, M.; Andrés, J.; Molnar, V.; Silla, E.; Tuñón, I.; and Bertran, J. (2004) Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 33, 98–107.
- (39) Seno, H. M., and Thiel, W. (2009) QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* 48, 1198–1228.
- (40) van der Kamp, M. W., and Mulholland, A. J. (2013) Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* 52, 2708–2728.
- (41) Hogg, T.; Nagarajan, K.; Harberg, S.; Chen, L.; Shen, X.; Jiang, H.; Wicke, M.; Holmberg, C.; Hagemfeld, B.; and Schmidt, C. L. (2008) Structural and functional characterization of Glucop2-1, a homoglobulinase from the radial parasite *Plasmodium falciparum*. *J. Biol. Chem.* 283, 25425–25437.
- (42) Field, M. J., Alde, M.; Bort, C.; Probst-De Martin, F.; and Thomas, A. (2009) The Dynamos library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* 21, 1088–1100.
- (43) Bao, D. C.; Riggs, D. M.; and Jensen, J. H. (2008) Very fast prediction and rationalization of pK<sub>a</sub> values for protein-ligand complexes. *Protein Struct., Funct., Bioinf.* 73, 765–783.
- (44) Li, H.; Robertson, A. D.; and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK<sub>a</sub> values. *Protein Struct., Funct., Bioinf.* 61, 764–771.
- (45) Obozin, M. H. M.; Sondergaard, C. B.; Kothkowski, M.; and Jensen, J. H. (2011) PROPKA: Consistent Treatment of Internal and Surface Residues in Empirical pK<sub>a</sub> Predictions. *J. Chem. Theory Comput.* 7, 525–537.
- (46) Nara, K.; Cai, Q.; Gao, J.; and York, D. M. (2007) Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* 3, 486–504.
- (47) Jorgensen, W. L.; Maxwell, D. S.; and TiradoRives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236.
- (48) Jorgensen, W. L.; Chandrosshan, J.; Madura, J. D.; Impey, R. W.; and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- (49) Field, M. J.; Bash, P. A.; and Karplus, M. (1990) A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.* 11, 700–733.
- (50) Singh, U. C., and Kollman, P. A. (1986) A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems Applications to the CH<sub>3</sub>Cl + Cl<sup>-</sup> Exchange-reaction and gas-phase protonation of polyethers. *J. Comput. Chem.* 7, 718–730.
- (51) Martí, S.; Molner, V.; and Tuñón, I. (2005) 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.* 1, 1008–1016.
- (52) Grassano, G.; Leggiani, L.; Trossi, L.; Ettari, R.; Micale, N.; and De Micheli, C. (2012) Mechanism of Glucop2-1 inhibition by  $\alpha,\beta$ -unsaturated boron 1,4-diazepin-2-one methyl ester. *J. Comput. Aided Mol. Des.* 26, 1035–1043.
- (53) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; and Rosenber, J. M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021.
- (54) Torrie, G. M., and Valleau, J. P. (1977) Non-physical sampling distributions in monte-carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 23, 187–199.
- (55) Ruiz-Pernia, J.; Silla, E.; Tuñón, I.; Martí, S.; and Molnar, V. (2006) Hybrid QM/MM potentials of mean force with interpolated corrections. *J. Phys. Chem. B* 108, 8427–8433.
- (56) Chaoq, Y. Y.; Corchicho, J. C.; and Truhlar, D. G. (1999) 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, 1140–1149.
- (57) Zhao, Y., and Truhlar, D. G. (2008) 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, 215–241.
- (58) Hehre, W. J.; Radom, L.; von R. Schleyer, P.; and Pople, J. A. (1986) *Ab Initio Molecular Orbital Theory*. John Wiley & Sons, New York.
- (59) Lendh, B. J.; Zhao, Y.; and Truhlar, D. G. (2003) Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. *J. Phys. Chem. A* 107, 1384–1388.
- (60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H. C.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kato, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudis, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Cross, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramila, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioabrenski, J.; and Fox, D. J. (2009) *Gaussian 09*, revision A1. Gaussian Inc., Wallingford, CT.

### **7.1.1. Supporting Information**

Supporting Information for the paper entitled: *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Falcipain-2 Inhibition by the Epoxysuccinate E64.*

**Supporting Information**

**QM/MM studies of mechanism of falcipain 2 inhibition by the epoxysuccinate**

**E64.**

*Kemel Arañet, Sílvia Ferrer, Sergio Mari, Vicent Moliner*

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

**Table S1.** Averaged AM1d/MM distances obtained from 200 ps of QM/MM MD simulation for reactants state and products state for the reaction mechanism of FP2 inhibition by the attack of Cys42 on C2 atom. Simulations have been carried out at pH 5.5 and pH 7.

## a) reactants

	pH=7	pH=5.5
SG-C2	2.83 ± 0.03	4.40 ± 0.44
O3-C2	1.44 ± 0.03	1.44 ± 0.03
O3-H10	3.85 ± 0.41	3.31 ± 0.60
O1-H10	0.96 ± 0.02	0.98 ± 0.03
O1-HD1 (His174)	2.61 ± 0.26	4.98 ± 0.53
O2-HI21 (Gln36)	3.47 ± 0.47	3.80 ± 0.84
O4-H (Trp43)	3.21 ± 0.78	4.67 ± 0.82
H10-O (Asn81)	6.13 ± 0.37	4.23 ± 0.68
H29-OD1 (Asp170)	1.64 ± 0.10	2.43 ± 0.55
H12-OD2 (Asp170)	1.62 ± 0.09	2.26 ± 0.30
H28-OE1 (Gln171)	2.18 ± 0.19	4.62 ± 0.84

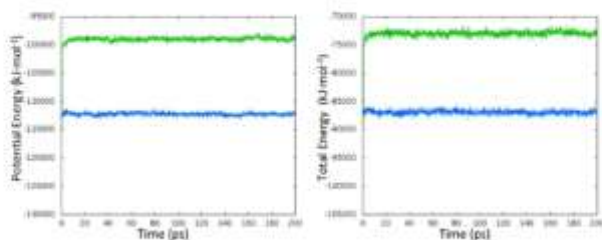
## b) products

	pH=7	pH=5.5
SG-C2	1.80 ± 0.03	1.80 ± 0.03
O3-C2	2.37 ± 0.05	2.45 ± 0.05
O3-H10	0.97 ± 0.03	0.98 ± 0.03
O1-H10	4.09 ± 0.39	2.17 ± 0.23
O1-HD1 (His174)	2.55 ± 0.23	4.68 ± 0.53
O2-HI21 (Gln36)	2.25 ± 0.25	1.96 ± 0.13
O4-H (Trp43)	1.62 ± 0.27	1.65 ± 0.37
H10-O (Asn81)	2.36 ± 0.38	3.18 ± 0.44
H29-OD1 (Asp170)	1.65 ± 0.10	2.08 ± 0.27
H12-OD2 (Asp170)	1.66 ± 0.10	2.02 ± 0.32
H28-OE1 (Gln171)	2.16 ± 0.19	5.01 ± 1.09

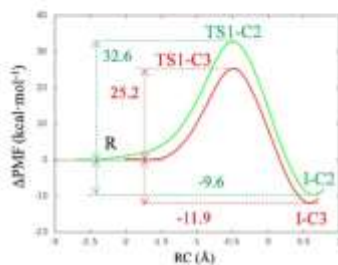
**Table S2.** Averaged AM1d/MM distances obtained for key states located along the reaction mechanism of the second step of the Cys42 attack on C3: the protonation of O3 by a water molecule. All values are in Å and are obtained from 200 ps of AM1d/MM MD simulations on the stationary points of the free energy surfaces.

	I-C3w	TS2-C3w	P-C3w
S <sub>O</sub> -C3	1.82 ± 0.04	1.82 ± 0.04	1.81 ± 0.04
O3-C3	2.44 ± 0.04	2.41 ± 0.05	2.40 ± 0.05
O3-H10	2.60 ± 0.22	2.17 ± 0.15	3.42 ± 0.39
O3-H2w	2.32 ± 0.03	1.23 ± 0.03	0.97 ± 0.03
OH2w-H10	2.13 ± 0.12	0.98 ± 0.03	0.97 ± 0.03
OH2w-H2	0.98 ± 0.03	1.18 ± 0.02	4.34 ± 0.03
O1-H10	0.97 ± 0.03	2.85 ± 0.71	2.61 ± 0.67
O1-HD1 (His174)	2.80 ± 0.20	2.53 ± 0.24	2.47 ± 0.21
O2-HH21 (Gln36)	2.08 ± 0.16	2.04 ± 0.15	2.35 ± 0.23
O4-H (Trp43)	1.51 ± 0.25	1.58 ± 0.27	1.55 ± 0.27
H2w-O (Asn81)	4.32 ± 0.20	3.90 ± 0.27	2.06 ± 0.15
H29-OD1 (Asp170)	1.66 ± 0.10	1.66 ± 0.10	1.65 ± 0.10
H29-OD2 (Asp170)	2.85 ± 0.16	2.85 ± 0.16	2.83 ± 0.16
H12-OD1 (Asp170)	2.73 ± 0.17	2.72 ± 0.17	2.73 ± 0.16
H12-OD2 (Asp170)	1.69 ± 0.10	1.70 ± 0.10	1.67 ± 0.10
H28-OE1 (Gln171)	2.15 ± 0.19	2.14 ± 0.19	2.15 ± 0.19

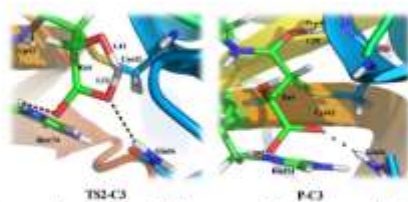




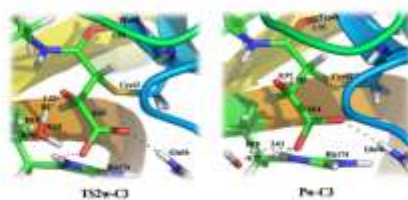
**Figure S1.** Time evolution of potential energy (a) and total energy (b) obtained along 200 ps of QM/MM MD simulation performed for products-like structures at two different conformations; conformation structure corresponding to the initial X-ray structure (PDB code 3BPF) with guanidine group of E64 interacting with Tyr78 (green line), and conformation structure corresponding to guanidine group of E64 interacting with Asp170 (blue line).



**Figure S2.** AM1d/MM PMFs of sulphur attack to C2 (green line) or C3 (red line) obtained after 200 ps of MD starting from X-ray conformation. RC corresponds to  $d(\text{CX-O3})-d(\text{SG-CX})$ , with X being C2 or C3.

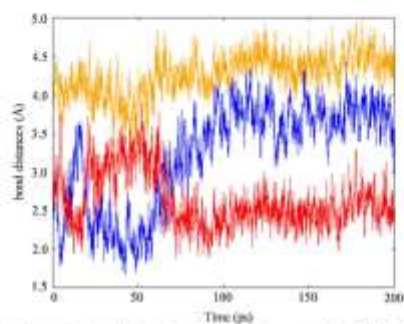


**Figure S3.** Representative snapshots of the key states of the reaction mechanism of the second step of the Cys42 attack on C3: the protonation of O3 by intra-molecular proton transfer. The average distances are reported in Å.



**Figure S4.** Representative snapshots of the key states of the reaction mechanism of the second step of the Cys42 attack on C3: the protonation of O3 with participation of a water molecule. The average distances are reported in Å.





**Figure S6.** Time dependent evolution of selected distances on the QM/MM MD simulation constrained on the TS2-C3. The orange line corresponds with the O1-HD1 (His174) distance, the blue line with the O2-HE21 (Gln36) distance, and the red line with the O2-HD1 (His174) distance.

## **7.2. Paper II**

Paper entitled: *First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones.*

Kemel Arafet, Silvia Ferrer, and Vicent Moliner

Biochemistry **2015**, 54 (21), 3381-3391. DOI: 10.1021/bi501551g.

Reprinted with permission from (DOI: 10.1021/bi501551g). Copyright (2015) American Chemical Society.

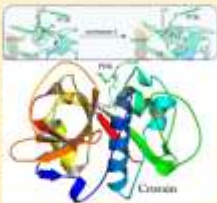
## First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones

Kemal Arafat, Silvia Ferrer,<sup>®</sup> and Vicent Moliner<sup>®</sup>

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

Supporting Information

**ABSTRACT:** Cruzain is a primary cysteine protease expressed by the protozoan parasite *Trypanosoma cruzi* during Chagas disease infection, and thus, the development of inhibitors of this protein is a promising target for designing an effective therapy against the disease. In this paper, the mechanism of inhibition of cruzain by two different irreversible peptidyl halomethyl ketones (PHK) inhibitors has been studied by means of hybrid quantum mechanics/molecular mechanics–molecular dynamics (MD) simulations to obtain a complete representation of the possible free energy reaction paths. These have been traced on free energy surfaces in terms of the potential of mean force computed at AM14/MM and DFT/MM levels of theory. An analysis of the possible reaction mechanisms of the inhibition process has been performed showing that the nucleophilic attack of an active site cysteine, Cys25, on a carbon atom of the inhibitor and the cleavage of the halogen–carbon bond take place in a single step. PCIK appears to be much more favorable than PFK from a kinetic point of view. This result would be in agreement with experimental studies in other papain-like enzymes. A deeper analysis of the results suggests that the origin of the differences between PCIK and PFK can be the different stabilizing interactions established between the inhibitors and the residues of the active site of the protein. Any attempt to explore the viability of the inhibition process through a stepwise mechanism involving the formation of a thiohemiketal intermediate and a three-membered sulfonium intermediate has been unsuccessful. Nevertheless, a mechanism through a protonated thiohemiketal, with participation of His159 as a proton donor, appears to be feasible despite showing higher free energy barriers. Our results suggest that PCIK can be used as a starting point to develop a proper inhibitor of cruzain.



Chagas disease (American trypanosomiasis) is an illness caused by the protozoan parasite *Trypanosoma cruzi* and was discovered in 1909 by the Brazilian doctor Carlos Chagas (1879–1934).<sup>1</sup> It is estimated that the illness affects ~7–8 million people living mainly in endemic Latin American countries.<sup>2</sup> The movement of Chagas disease to areas previously considered nonendemic, resulting from increasing population mobility between Latin America and the rest of the world, represents a serious public health challenge.<sup>3</sup> There is not any available vaccine to prevent Chagas disease, and the two available drugs for treatment, benznidazole and nifurtimox, are not only toxic and have important contra-indications (pregnancy, renal or hepatic failure, and psychiatric and neuronal disorders) but also ineffective for the chronic stage of the disease.<sup>4,5</sup> Clearly, there is an urgent need to develop an effective therapy against Chagas disease. One approach consists of developing inhibitors of cruzain, the primary cysteine protease expressed by *T. cruzi* during infection.<sup>6–9</sup>

Cruzain, which belongs to the family of cysteine proteases, was initially discovered from the parasite cell-free extracts and subsequently heterologously expressed in *Escherichia coli*.<sup>10,11</sup> This protease is expressed in all life cycle stages of the parasite and is involved in nutrition and the fight against host defense mechanisms.<sup>11–14</sup> Addition of a cruzain inhibitor to cultures of mammalian cells exposed to trypomastigotes or to mammalian cells already infected with *T. cruzi* amastigotes blocks

replication and differentiation of the parasite, thus interrupting the parasite life cycle.<sup>15–21</sup>

Several groups have demonstrated that irreversible inhibition of cruzain by small molecules realitates infection of the parasite in cell culture and animal models.<sup>15,22–24</sup> Irreversible inhibitors that contain an electrophilic functional group, such as vinyl sulfones, fluoro methyl ketones, aridines, or nitriles, covalently bind to cruzain via nucleophilic attack of the active site cysteine,<sup>25</sup> showing good inhibition activity.<sup>25,26</sup> In fact, to date, only irreversible inhibitors of cruzain have successfully cured parasitic infection,<sup>25</sup> implying that tight binding to the enzyme may be essential. In this sense, Shoellman and Shaw developed in 1962 the first peptidyl chloromethyl ketones, 1-(1-tosyl-amido-2-phenylethyl) chloromethyl ketone, as specific inhibitors for the serine protease chymotrypsin.<sup>27</sup> The major disadvantage of this inhibitor was its lack of selectivity because of the great chemical reactivity of the chloromethyl functional group. The development of chloromethyl ketone inhibitors led to the investigation of analogous inhibitor structures with different halo leaving groups replacing the chlorine atom. Bromomethyl and iodomethyl ketones have been synthesized

Received: December 22, 2014

Revised: May 7, 2015

Published: May 12, 2015

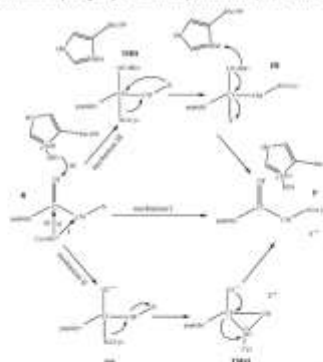
and are typically more reactive but less stable in aqueous solutions. The first peptide fluoromethyl ketone inhibitors were reported in the literature by Rasmick in 1985<sup>28</sup> and by Shaw's group in 1986.<sup>31</sup> These peptide fluoromethyl ketones were shown to be highly reactive and selective irreversible inhibitors for cysteine proteases. They are poor irreversible inactivators for serine proteases and are not reactive toward hemocoagulins. Inhibition of cruzain activity with fluoromethyl ketone-based inhibitors seems to be correlated with the loss of feasibility of parasites in both *ex vivo* tissue culture and *in vivo* mouse models.<sup>19,32,33</sup>

Information derived from the study of the molecular mechanism of hydrolysis catalyzed by cysteine proteases could be used as a starting point to explore the inhibition mechanism at an atomistic level. Early studies revealed a participation of residues Cys25 and His159 (cruzain numbering) from the active site of these proteases.<sup>34–36</sup> The catalytic cysteine mediates protein hydrolysis via a nucleophilic attack on the carbonyl carbon of a susceptible peptide bond. The imidazole group of the histidine polarizes the SH group of the cysteine and allows deprotonation even at neutral to weakly pH, and a highly nucleophilic thiolate/imidazolium ion pair is thereby produced.<sup>37</sup> The ion pair mechanism explains the unusually high reactivity of cysteine proteases toward electrophilic reagents in comparison to the nucleophilicity of the sulfur of cysteine or glutathione, especially in slightly acidic environments.<sup>38</sup>

With regard to the study of the inhibition mechanism of cysteine proteases, only few irreversible inhibitors have been studied theoretically.<sup>39–40</sup> The study of diacetyl ketones based on gas phase MRDO calculations proposed a mechanism in which the Cys25 residue binds to the inner diazo nitrogen thus forming an intermediate that evolves to a stable  $\beta$ -thioketone and molecular nitrogen.<sup>41</sup> More recently, density functional theory (DFT) and semiempirical methods have also been used to study the inhibition process of nitrile-containing compounds in the gas phase.<sup>42</sup> Classical MD simulations and exploration of potential energy surfaces (PESs) obtained with quantum mechanics/molecular mechanics (QM/MM) potentials have been used to study the inhibition of cysteine proteases by epimides and imidazole-based compounds.<sup>43</sup>

Powers and co-workers proposed two possible mechanisms (mechanisms I and II) of irreversible inhibition by peptidyl halomethyl ketones (PHK) based on different crystal structures of cysteine proteases (see Scheme 1).<sup>44</sup> Mechanism I, as shown in Scheme 1, is the direct displacement of the halide group by the thiolate anion. Mechanism II involves formation of a tetrahedral intermediate named thiohemiketal (TH) and a three-membered sulfonium intermediate (TMSI) that rearranges to give the final thioether adduct. The formation of the TH is equivalent to the presence of a tetrahedral intermediate in the serine protease mechanism.<sup>45</sup> Nevertheless, mechanism II has not been supported by theoretical studies. Kollman and co-workers, employing classical molecular mechanics simulations and semiempirical quantum mechanics with model systems to study the cysteine protease catalysis, showed that the attack of S<sup>−</sup> on a carbonyl carbon does not involve a stable anionic tetrahedral structure.<sup>46,47</sup> Later, Suhai and co-workers in a QM/MM study of the active site of free papain and of the NMA–papain complex did not obtain a stable tetrahedral NMA–papain complex.<sup>48</sup> Gao and Byun, in a combined QM/MM study of the nucleophilic addition reaction of methanethiolate and *N*-methylacetamide, concluded that

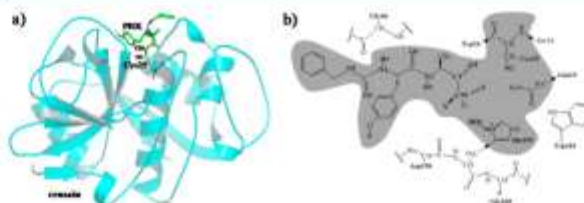
**Scheme 1. Reaction Mechanisms of Inhibition of Cysteine Proteases by Peptidyl Halomethyl Ketones (X = F or Cl)<sup>44</sup>**



<sup>44</sup>Mechanisms I and II were proposed by Powers and co-workers.<sup>44</sup> TH refers to the thiohemiketal intermediate and TMSI to the three-membered sulfonium intermediate. Mechanism III is an alternative mechanism that takes place through a thiohemiketal protonated intermediate, THH.

there is no stable tetrahedral intermediate in going from the reactant to the tetrahedral adduct.<sup>49</sup> Moreover, Warshel and co-workers in an *ab initio* study of general base/acid-catalyzed thiolysis of formamide and the hydrolysis of methyl thioformate showed that the anionic tetrahedral intermediate for the acylation reaction was found to be unstable in aqueous solution and to collapse immediately into the neutral form, which is the only intermediate on the reaction pathway.<sup>50</sup> On the basis of DFT methods with model compounds, Kolundarev and co-workers explored the PESs of the inhibition of halo ketones<sup>48</sup> and diketones<sup>51</sup> in the gas phase. According to their results, the thiohemiketal protonated intermediate would be stable but, in general, less stable than the reactant.<sup>48,49</sup> More recently, Zhan and co-workers in a pseudobond first-principles QM/MM-FE study of the reaction pathway for papain-catalyzed hydrolysis of *N*-acetyl-Phe-Gly 4-nitroline concluded that in the acylation step any path that includes the formation of the TH does not exist.<sup>52</sup> Nevertheless, to date, the mechanism of inhibition of cysteine protease by PHKs has not yet been studied by computational tools, including the protein environment effects.

The principal aim of this paper is to gain insights into the mechanism whereby PHK inhibits bound to a cysteine protease, in particular to cruzain. In a previous theoretical study conducted in our laboratory, the mechanism of inhibition of a protease involved in the degradation of human hemoglobin, namely falcipain-2 (FP2), by the epoxy succinate *N*:[*N*-(1-hydroxycyclohexyl)ethyl-carboxyl]leucylamino-butyl]guanidine, E64, was conducted.<sup>35</sup> The results indicated that the irreversible attack of a conserved cysteine on E64 can take place on both carbon atoms of its epoxy ring because both processes present similar barriers. We herein present a theoretical study of the



**Figure 1.** (a) Monomer 1A1M structure showing the crystal structure of cruzain from *T. cruzi* bound to Bz-Tyr-Ala-CH<sub>2</sub>F (PFK). The inhibitor (CM) is covalently bound to the enzyme through the SG atom from residue Cys25. The inhibitor and Cys25 are represented as thick sticks. (b) Details of the active site corresponding to the studied model. The gray region corresponds to the QM subset of atoms that includes the inhibitor Bz-Tyr-Ala-CH<sub>2</sub>X (X = F or Cl), Cys25, the side chain of Gln19, and the imidazole ring of His159. The link atoms between the QM and MM regions are indicated as black dots.

mechanism of inhibition of cruzain by two PHKs, benzoyl-tyrosine-alanine-fluoromethyl ketone (Bz-Tyr-Ala-CH<sub>2</sub>F, or PFK) and benzoyl-tyrosine-alanine-chloromethyl ketone (Bz-Tyr-Ala-CH<sub>2</sub>Cl, or PCLK). The main difference between these two inhibitors and the previously studied E64 is the presence of the electrophilic functional group, an a halogen atom in PHKs and an epoxide ring and a carboxylic group in E64. The studied proteins belong to the same family of cysteine proteases (clan CA) but are responsible of different human diseases; FP2 is associated with malaria and cruzain with Chagas disease. Hybrid QM/MM<sup>34–36</sup> potentials have been employed to explore the PESs and to run molecular dynamics (MD) simulations that allow the free energy profiles of the reaction to be determined in terms of the potential of mean force (PMF). The reaction mechanisms, with their corresponding free energy barriers, averaged geometries, and interactions between the inhibitor and the protein, have been the subject of a deep analysis that aimed to improve our understanding of the molecular mechanism, paving the way to the design of new inhibitors that allow the effective treatment of Chagas disease.

#### ■ COMPUTATIONAL MODEL

The initial coordinates were taken from the X-ray crystal structure of cruzain from *T. cruzi* bound to Bz-Tyr-Ala-CH<sub>2</sub>F with PDB entry 1A1M<sup>37</sup> at 2.0 Å resolution. The structure consists in one chain of 215 amino acids (cruzain), the inhibitor without the fluor atom, and 46 water molecules (see Figure 1a).

Cruzain is located in the intracellular vesicles during the dormant stage of the parasite at a slightly acidic pH. During the intracellular amastigote stage of the *T. cruzi* life cycle, cruzain is present on the surface of the parasite where the pH of the host cytoplasm is almost neutral (7.4).<sup>32</sup> Thus, hydrogen atoms were added at this physiological pH using the iDYNAMO library,<sup>37</sup> according to the pK<sub>a</sub> values of the residues of the 1A1M structure calculated within the empirical PROPKA 3.1 program of Jensen et al.<sup>38–40</sup> Fluor and chlorine atoms were added using Pyouf<sup>41</sup> to create the two systems being studied here: Bz-Tyr-Ala-CH<sub>2</sub>F and Bz-Tyr-Ala-CH<sub>2</sub>Cl. A total of 12 counterions (Na<sup>+</sup>) were placed into optimal electrostatic positions around both systems, to obtain electro-neutrality. Finally, the systems were placed in a 79.5 Å side cubic box of water molecules.

As depicted in Figure 1b, the inhibitor, Cys25, the side chain of Gln19, and the imidazole ring of His159 (81 atoms) were described by means of the AM1d semiempirical Hamiltonian,<sup>42</sup>

Gln19 was treated quantum mechanically to explore its active role in mechanism II, because formation of a bond or charge transfer from the OT atom of the inhibitor must be considered as a possibility. His159 was treated quantum mechanically to explore mechanism III. The rest of the protein and water molecules were described by OPLS-AA<sup>43</sup> and TIP3P<sup>44</sup> force fields, respectively. To saturate the valence of the QM/MM frontier, we used the link atom procedure.<sup>34,35</sup> Because of the size of the full system, all residues more than 25 Å from the CA) atom of the inhibitor were kept frozen during the simulations (43169 atoms from a total of 50126). A force switch function with a cutoff distance in the range of 14.5–18 Å was applied to treat the nonbonding interactions. All the QM/MM calculations were conducted using the iDYNAMO library.<sup>37</sup>

The two systems were relaxed by means of QM/MM MD for 1 ns at 300 K using the NVT ensemble and the Langevin–Verlet integrator. Analysis of the time evolution of the root-mean-square deviation (rmsd) of backbone atoms of the protein and the inhibitor for the Bz-Tyr-Ala-CH<sub>2</sub>F and the Bz-Tyr-Ala-CH<sub>2</sub>Cl systems confirms that the systems were equilibrated (see Figure S1 of the Supporting Information). Structures obtained after relaxation were used to generate hybrid AM1d/MM PESs. Stationary structures (including reactants, products, intermediates, and transition state structures) were located and characterized by means of a microsearch iterative scheme.<sup>45</sup>

Once the PESs were explored, PMFs generated as a function of a distinguished reaction coordinate (RC) were obtained using the weighted histogram analysis method [WHAM] combined with the umbrella sampling approach.<sup>47,48</sup> The RC is defined depending on the chemical step under study. In general, for the monodimensional PMFs (1D-PMF), the value of the force constant used for the harmonic umbrella sampling was 2500 kJ mol<sup>-1</sup> Å<sup>-2</sup> and the simulation windows consisted of equilibration for 20 ps and production for 40 ps, with a time step of 1 fs. Subsequently, 200 ps AM1d/MM MD simulations of the stationary points were performed to analyze the main geometrical parameters as an average. For the study of mechanism I (see Scheme 1), a 1D-PMF AM1d/MM was obtained using the antisymmetric combination of the bond-breaking and bond-forming distances: RC = [d(X–CM) – d(SG–CM)], where X represents F or Cl. This required series of 92 and 85 simulation windows for Bz-Tyr-Ala-CH<sub>2</sub>F and



Be-Tyr-Ala-CH<sub>2</sub>-Cl systems, respectively. For mechanism II (see Scheme 1), a 1D-PMF AM1d/MM was computed to study the formation of the TH from the reactant state using the antisymmetric combination of the bond-breaking and bond-forming distances [RC =  $d(SG-CT) - d(CT-OT)$ ]. This required series of 61 and 66 simulation windows for Be-Tyr-Ala-CH<sub>2</sub>-F and Be-Tyr-Ala-CH<sub>2</sub>-Cl systems, respectively. A 1D-PMF AM1d/MM was obtained for the formation of the TMSI from the product state using the SG-CT bond-breaking distance as the RC. This required series of 41 and 37 simulation windows for Be-Tyr-Ala-CH<sub>2</sub>-F and Be-Tyr-Ala-CH<sub>2</sub>-Cl systems, respectively. For the study of mechanism III (see Scheme 1), a two-dimensional AM1d/MM PMF (2D-PMF) was computed for the formation of the THH intermediate from the reactant state using the ND1-HD1 bond-breaking distance as RC1 and the SG-CT bond-forming distance as RC2. This required series of 1092 and 1288 simulation windows for Be-Tyr-Ala-CH<sub>2</sub>-F and Be-Tyr-Ala-CH<sub>2</sub>-Cl systems, respectively. For the formation of the IH intermediate from the THH intermediate, an AM1d/MM 2D-PMF was obtained using the SG-CM bond-forming distance as RC1 and the X-CM (X = F or Cl) bond-breaking distance as RC2. This required series of 651 simulation windows for both systems. Finally, a 2D-PMF AM1d/MM was obtained for the formation of the product state from the IH intermediate using the ND1-HD1 bond-forming distance as RC1 and the X-CT (X = F or Cl) bond-breaking distance as RC2. This required series of 1288 and 2310 simulation windows for Be-Tyr-Ala-CH<sub>2</sub>-F and Be-Tyr-Ala-CH<sub>2</sub>-Cl systems, respectively.

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. To correct the low-level AM1d energy function used in the 1D-PMFs, an interpolated correction scheme developed at our laboratory has been applied.<sup>66</sup> In this correction scheme, on the basis of a method proposed by Truhlar et al. for dynamical calculations of gas phase chemical reactions,<sup>78</sup> the new energy function employed in the simulations is defined as

$$E = E_{LL/MM} + S[\Delta E_{LL}^{HL}(\zeta)] \quad (1)$$

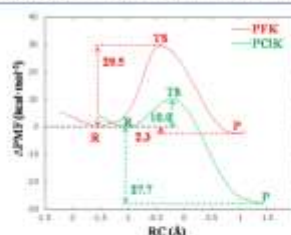
where  $S$  denotes a spline function, whose argument  $\Delta E_{LL}^{HL}(\zeta)$  is a correction term taken as the difference between single-point calculations of the QM subsystem using a high-level (HL) method and the low-level (LL) one. The correction term is expressed as a function of the distinguished reaction coordinate  $\zeta$  (RC). When corrections were made on 2D-PMFs, the correction term was then expressed as a function of the two coordinates,  $\zeta_1$  and  $\zeta_2$ , to generate the higher-level 2D-PMF. The HL calculations were conducted by means of the M06-2X functional<sup>71</sup> with the 6-31+G(d,p) basis set<sup>72</sup> following the suggestions of Truhlar and co-workers.<sup>73,74</sup> The 6-311+G-(2d,2p) basis set<sup>75,76</sup> was also checked for the F systems (mechanism I), because this is the one specifically recommended for this element. Tests of the use of both basis sets rendered almost quantitatively equivalent results (see Figure S2 of the Supporting Information), and consequently, the 6-31+G(d,p) basis set was used for all calculations. These calculations were conducted using Gaussian09.<sup>76</sup>

## RESULTS AND DISCUSSION

The first proposed mechanism of inhibition of cruzain by the two studied PHKs (mechanism I in Scheme 1)<sup>37</sup> has been

initially explored by computing the PESs. Once the stationary points were located on the PESs, the mechanisms of both systems were clearly identified as single-step mechanisms, where the attack of Cys25 on carbon CM takes place concomitantly with the breaking of the bond of the halogen with CM.

In the next step, the free energy profiles have been calculated in terms of 1D-PMFs at the AM1d/MM level (see Figure S3 of the Supporting Information) and corrected at the M06-2X/6-31+G(d,p)/MM level (Figure 2). The first observation is the



**Figure 2.** Free energy profiles of cruzain inhibition with PHKs for mechanism I, computed at the M06-2X/6-31+G(d,p)/MM level. RC corresponds to  $d(X-CM) - d(SG-CM)$ , where X is F (red) or Cl (green).

dramatic influence of increasing the level of the QM region description on the energy barriers and reaction energies. This result justifies the use of high-level Hamiltonians to properly describe the energetics of this chemical reaction when studied by means of QM/MM methods. Consequently, the analysis of energies will be conducted with M06-2X/MM results. As observed in Figure 2, the free energy barrier for the inhibition of cruzain with PCIK ( $\sim 10$  kcal mol<sup>-1</sup>) is much lower than that obtained with PFK ( $\sim 29.5$  kcal mol<sup>-1</sup>). This result would describe a higher chemical reactivity of the peptide inhibitor with clove than with fluce atoms. Despite no experimental kinetic data being available for the inhibition of cruzain with PHKs, experimental studies with cathepsin B (papain-like enzymes known as clan CA)<sup>77</sup> concluded the reactivity of a PCIK is higher than that of a PFK,<sup>27,80,81</sup> in agreement with our results.

Analysis of the reaction energies obtained for both inhibitors shows that the process with PCIK appears to be much more exergonic than the reaction with the PFK. Nevertheless, the reaction is irreversible in any case because the leaving anion will diffuse into the surrounding water shell because of the concentration gradients. Quantitative differences in reaction free energies obtained from the PMFs can be due to the limits of the distinguished reaction coordinate that, obviously, do not cover the full diffusion process of the anion into the bulk solvent. Interestingly, and as can be confirmed by the averaged values of the key distances reported in Table 1, the transition state in the PCIK system appears at a more advanced value of the RC ( $-0.09$  Å) than in the reaction of PFK ( $-0.48$  Å). The fact that a more advanced TS was not associated with a higher activation barrier is due to the differences observed in the reactant complex: the minimum obtained in the PFK is located at a value of the RC ( $-1.60$  Å) more negative than in the PCIK

**Table 1.** Averaged Distances for Key States Located along the Inhibition of Crustin by PHK, through Direct Mechanism I<sup>a</sup>

	<i>d</i> (Å)		
	E	TS	P
(a) PFK			
SG-CM	3.15 ± 0.03	2.32 ± 0.03	1.79 ± 0.03
F-CM	1.55 ± 0.02	2.04 ± 0.03	2.70 ± 0.02
F-HI	4.58 ± 0.11	3.03 ± 0.14	2.55 ± 0.12
F-HI	1.18 ± 0.13	1.41 ± 0.15	6.24 ± 0.01
F-H (H <sub>2</sub> O)	4.67 ± 0.09	3.81 ± 1.30	1.52 ± 0.09
F-H (H <sub>2</sub> O)	4.33 ± 0.00	3.82 ± 0.74	1.75 ± 0.51
SG-H (Trp38)	1.54 ± 0.14	1.55 ± 0.20	3.15 ± 0.22
SG-H (Gly168)	1.39 ± 0.05	3.58 ± 0.81	2.63 ± 0.24
SG-HDI (His159)	3.00 ± 0.38	1.17 ± 0.09	4.12 ± 0.28
OT-H (Cys25)	4.74 ± 0.40	5.46 ± 1.13	3.07 ± 0.33
OT-H (Gln39)	6.12 ± 0.82	7.75 ± 1.03	2.93 ± 0.47
O1-H (Trp24)	1.89 ± 0.20	3.64 ± 0.61	1.58 ± 0.23
O1-H (Gln66)	2.32 ± 0.19	5.55 ± 0.43	3.80 ± 0.40
N2-H (Trp38)	4.44 ± 0.24	4.09 ± 0.44	5.49 ± 0.34
N3-H (Gln66)	3.89 ± 0.31	5.23 ± 0.28	2.68 ± 0.40
H2-O (Gln66)	5.17 ± 0.25	2.18 ± 0.31	2.19 ± 0.87
H1-O (Gln66)	2.04 ± 0.24	3.58 ± 0.26	5.60 ± 0.54
(b) PCIK			
SG-CM	3.80 ± 0.03	2.33 ± 0.03	1.78 ± 0.03
O-CM	1.77 ± 0.03	2.24 ± 0.03	3.10 ± 0.03
O-HDI (His159)	3.50 ± 0.59	4.06 ± 0.36	1.40 ± 0.20
O-HI (Trp181)	3.03 ± 0.78	5.79 ± 0.62	3.78 ± 0.76
O-H (H <sub>2</sub> O)	3.91 ± 0.73	5.58 ± 0.74	2.50 ± 0.26
SG-H (Trp38)	2.18 ± 0.24	2.95 ± 0.29	3.08 ± 0.20
SG-H (His159)	1.82 ± 0.10	3.14 ± 0.29	3.28 ± 0.46
SG-H (Gly168)	2.82 ± 0.66	2.51 ± 0.21	3.03 ± 0.43
OT-H (Cys25)	4.18 ± 0.98	2.55 ± 0.32	3.07 ± 0.84
OT-H (Gln39)	6.74 ± 1.69	2.54 ± 0.33	3.90 ± 0.82
O1-H (Gln66)	3.14 ± 0.73	3.68 ± 0.44	3.92 ± 0.36
O2-H (Trp24)	2.89 ± 1.11	1.62 ± 0.31	1.52 ± 0.20
O2-H (Gln66)	4.42 ± 0.85	2.89 ± 0.45	4.13 ± 0.29
O3-H (Gln66)	3.46 ± 0.68	2.89 ± 0.45	3.01 ± 0.42
N3-H (Gln66)	1.03 ± 0.33	2.74 ± 0.22	2.82 ± 0.24
H2-O (Gln66)	1.81 ± 0.23	2.08 ± 0.20	1.96 ± 0.17
H1-O (Asp138)	3.67 ± 0.47	3.31 ± 0.49	3.43 ± 0.38

<sup>a</sup>Results were obtained from 200 ps AM1d/MM MD simulations on the stationary points taken from the M06-2X/6-31+G(d,p)/MM free energy profiles (in arginons). Interatomic distances associated with the RCs were constrained during the simulations.

(−1.03 Å). Interestingly, the change in the RC from the reactant complex to the transition state in both systems is very similar ( $\Delta RC_{(R,TS)}$  values of 0.94 and 1.12 Å for PCIK and PFK, respectively). Then, the differences in the free energies of activation can not be completely discussed in these terms, and further analysis of average structures of stationary points of the free energy surfaces has to be conducted. Representative snapshots of reactants, transition states, and products of both inhibitory reactions are presented in Figure 3. As observed in the figure, and quantitatively confirmed by the geometrical data listed in Table 1, more hydrogen bond interactions between the inhibitor and the residues of the active site are detected in the TS of PCIK (OT-H<sub>(Cys25)</sub>, O1-H<sub>(Trp24)</sub> and H1-O<sub>(Gln66)</sub>) than in the TS of PFK (H2-O<sub>(Gln66)</sub>), which is in agreement with the lower barrier obtained for the former reaction. The larger amount of favorable inhibitor–protein interactions obtained in

products of the PCIK than in the products of the PFK inhibitor would also support a higher level of stabilization of the PCIK–protein complex (see Table 1). The halogen atoms in products appear to be stabilized mainly by interactions with water molecules, in both reactions. It can be observed that fluorine ion interacts with a hydrogen H2 atom of the inhibitor and two water molecules. In the case of PCIK, chloride ion interacts, mainly, with His159 and two water molecules but also exhibits weak interactions with Gln19 and Trp181. Thus, the thioether adduct (products) presents interactions that are stronger than those of the reactants, supporting a description as an irreversible inhibition.

As mentioned in the introductory section, Powers and co-workers<sup>47</sup> proposed a second mechanism of inhibition of crustin by PFKs (mechanism II in Scheme 1) involving the formation of a thiohemiketal intermediate (TH) and a three-membered sulfonium intermediate (TMSI). Nevertheless, all the attempts conducted to explore the mechanism by means of AM1d/MM PESs were unsuccessful. TH has not been found to be a stable intermediate in the PES. This has been confirmed when exploring the evolution of the PMF from reactants to the TH using the  $d(SG-CT) - d(CT-OT)$  antisymmetric combination as a reaction coordinate. The resulting PMFs obtained for both inhibitors at the M06-2X/6-31+G(d,p)/MM level (see Figure 4) confirm the instability of the proposed TH. The corresponding profiles at the AM1d/MM level, reported in Figure S4 of the Supporting Information, render the same conclusions.

The minimum that appears in Figure 4a at an RC value of 1.17 Å corresponds to an intermediate with SG–CT and SG–CM distances equal to 2.41 ± 0.03 and 3.06 ± 0.10 Å, respectively. Obviously, this structure does not correspond to the TH. Moreover, further attempts to explore mechanism II were based on the exploration of the step from products to the TMSI. The PMFs obtained as a function of the SG–CT distance render profiles that do not allow the localization of stable structures for this intermediate, as shown in Figure 5. The corresponding profiles at the AM1d/MM level (Figure S5 of the Supporting Information) render the same conclusions. Finally, any attempt to equilibrate the PFK–crustin systems at TH or TMSI intermediates by means of constrained QM/MM MD simulations reveals the instability of these structures as soon as all atoms of the system are allowed to move (see Figures S6–S8 of the Supporting Information).

These results suggest that inhibition of crustin by PHKs cannot take place through the proposed mechanism II, a conclusion that is in agreement with previous theoretical studies of different systems showing that the attack of S<sup>−</sup> on a C=O bond does not involve a stable anionic tetrahedral structure.<sup>48–53</sup> Moreover, regardless, experimental studies suggest the possibility of mechanism II taking place in serine proteases through the formation of the TH between the serine–OH and C=O bonds from the inhibitor;<sup>54</sup> no experimental evidence of this mechanism in cysteine proteases exists.

Finally, a third mechanism through a protonated thiohemiketal, THH (mechanism III in Scheme 1), has been explored. From a computational point of view, this mechanism appears to be more complex because of the participation of an acid, His159, that protonates the OT carbonyl oxygen of the inhibitor. Consequently, and as described in Computational Model 2D-PMFs were computed to unequivocally define the mechanism. Thus, the first step corresponds to the attack of

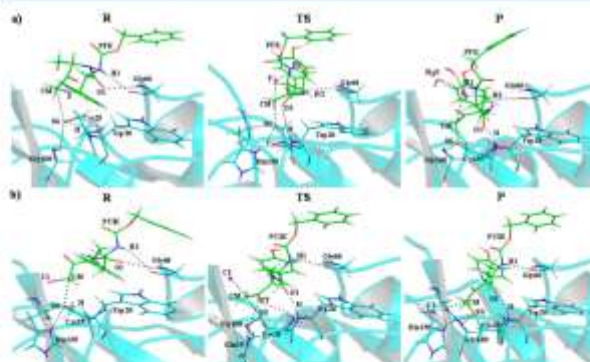


Figure 3. Representative snapshots of the key states of reaction mechanism 1 of inhibition of cruzain by (a) PFK and (b) PCIK.

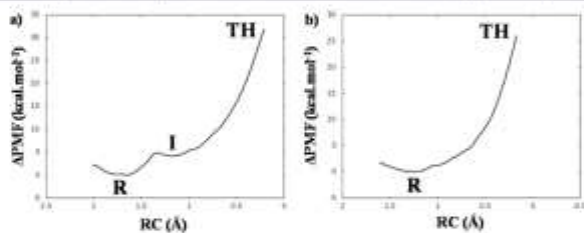


Figure 4. PMFs for the formation of the TH intermediate from reactants computed at the M06-2X/6-31+G(d,p)/MM level: (a) Br-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Br-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. RC corresponds to  $d(SG-CT) - d(OE-CT)$ .

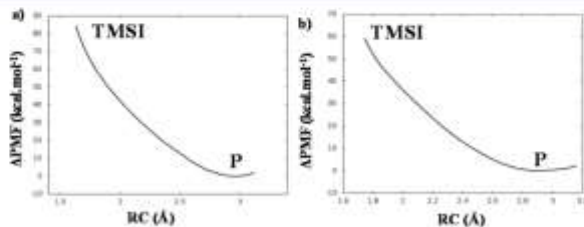
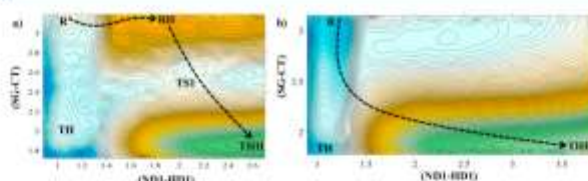


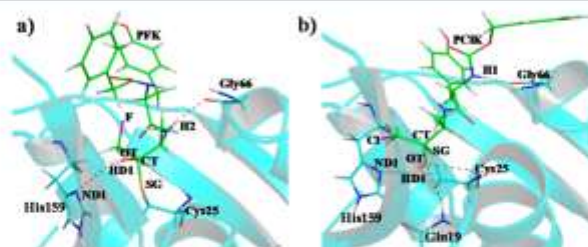
Figure 5. PMFs for the formation of the TMSI intermediate from products, computed at the M06-2X/6-31+G(d,p)/MM level: (a) Br-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Br-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. RC corresponds to the  $d(SG-CT)$  distance.

Cys25 on the CM carbon atom and the transfer of a proton from His159 to the OT oxygen leading to THH. The resulting FESs obtained at the M06-2X/6-31+G(d,p)/MM level are shown in Figure 6. The corresponding surfaces at the AM1d/

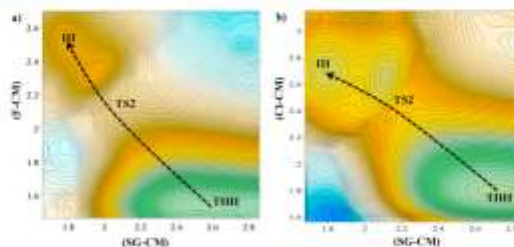
MM level are shown in Figure S9 of the Supporting Information. The first conclusion derived from Figure 6 is that this step takes place without any barrier when inhibition takes place with PCIK. In the case of PFK, the proton from



**Figure 6.** 2D-PMFs for the formation of the protonated thiohemiketal intermediate, THH, from reactants, computed at the M06-2X/6-31+G(d,p)/MM level: (a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in angstroms, and isoenergic lines are displayed every 0.5 kcal mol<sup>-1</sup>.



**Figure 7.** Representative snapshots of the protonated thiohemiketal intermediate, THH, in reaction mechanism III of inhibition of cysteine by (a) PFK and (b) PCIK.



**Figure 8.** 2D-PMFs for the transformation from protonated thiohemiketal, THH, to intermediate IH, computed at the M06-2X/6-31+G(d,p)/MM level: (a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in angstroms, and isoenergic lines are displayed every 1.0 kcal mol<sup>-1</sup>.

His159 is transferred to the OT atom of the inhibitor also without an energy barrier, but rendering a stable intermediate (RH in Figure 6a). Then, the attack of Cys25 on the CT atom proceeds with a free energy barrier defined by TS1. Interestingly, both 2D-PMFs show the protonated thiohemiketal, THH, as a stable species and confirm the previous results (see Figure 4) that indicated the instability of the TH. Representative snapshots of this THH species obtained with both inhibitors are shown in Figure 7.

The second step has also been explored by means of 2D-PMFs, controlling the X-CM and SG-CM distances. The resulting M06-2X/6-31+G(d,p)/MM FESs are presented in Figure 8 (the corresponding surfaces at the AM1d/MM level are shown in Figure S10 of the Supporting Information). In this step, the mechanisms of both inhibitors proceed in an almost equivalent way. Finally, when the last step of the inhibition of cysteine protease by PFK and PCIK inhibitors is explored, differences are found. Via comparison of panels a and b of Figure 9, we observed how the transformation from IH to

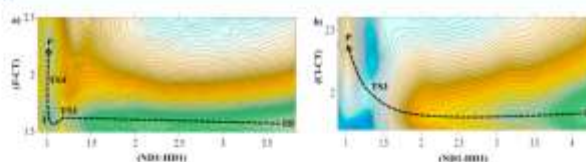


Figure 9. 2D-PMFs for the formation of products from the IH intermediate computed at the M06-2X/6-31+G(d,p)/MM level: (a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in angstroms, and isosurface lines are displayed every 1.0 kcal mol<sup>-1</sup>.

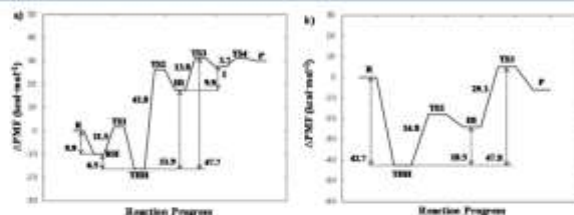


Figure 10. Free energy profiles for the inhibition of cruzain via mechanism III computed at the M06-2X/6-31+G(d,p)/MM level: (a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor and (b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor.

products takes place in a concerted way for PCIK but through a stable intermediate, I, for PFK. The same conclusions are derived from the AM1d/MM free energy surfaces reported in Figure S11 of the Supporting Information.

The complete free energy profiles for the inhibition of cruzain through mechanism III, derived from the M06-2X/MM 2D-PMFs reported in Figures 6, 8, and 9, are summarized in Figure 10. We can conclude that, as was observed for mechanism I, the inhibition of cruzain with PCIK is much more favorable than the inhibition with PFK, once again in agreement with experimental studies of papain-like enzyme studies,<sup>7</sup> leading to the conclusion that the reactivity of a PCIK is higher than that of a PFK.<sup>27,54,51</sup> The rate-limiting step of mechanism III would correspond to TS3 for both inhibitors with effective free energy barriers, computed from the stabilized TH4, of 47.7 and 47.8 kcal mol<sup>-1</sup> for PFK and PCIK, respectively. Nevertheless, as observed on the profiles, if reactants are considered to be the reference state, the inhibition of cruzain with PCIK is clearly more feasible than the inhibition with PFK.

From a thermodynamic point of view, the energetic results derived from the FESs in this mechanism would describe an endergonic process in which inhibition takes place with PFK but an exergonic one for the reaction with the PCIK. Regardless, as mentioned above, the reaction must be irreversible because the leaving anion will diffuse into the surrounding water shell because of the concentration gradients. Quantitative differences obtained from the PMFs can be associated with the limits of the distinguished reaction coordinate in the last step that, obviously, do not cover the full diffusion process of the leaving anion.

Analysis of geometries of stationary point structures (reported in Table S1 of the Supporting Information) reveals

more hydrogen bond interactions between the inhibitor and the residues of the active site in TS3 of PCIK than in TS3 of PFK, which is in agreement with the lower energies obtained for the former reaction. Again, and as observed in the results of mechanism I, the larger amount of favorable inhibitor–protein interactions obtained in products of the PCIK compared to the amount in the products of the PFK inhibitor is in agreement with the higher level of stabilization of the PCIK–protein complex (see Table S1 of the Supporting Information).

## CONCLUSIONS

This paper reports the theoretical study of the inhibition of cruzain by two irreversible peptidyl halomethyl ketones inhibitors, Bz-Tyr-Ala-CH<sub>2</sub>F (PFK) and Bz-Tyr-Ala-CH<sub>2</sub>Cl (PCIK), conducted by means of MD simulations using hybrid AM1d/MM and M06-2X/MM potentials. The reaction mechanisms previously proposed in the literature<sup>27</sup> have been explored. According to our results, the nucleophilic attack of the unprotonated Cys25 on the CM atom of the inhibitor (PFK or PCIK) and the cleavage of the halogen–CM bond take place in a concerted way. Analysis of free energy barriers and reaction free energies, computed at both levels of theory, shows that the inhibition by PCIK would be kinetically and thermodynamically more favorable. In fact, the difference between the barriers of both reactions is ~3-fold (10.0 and 29.5 kcal mol<sup>-1</sup> for PCIK and PFK, respectively), and even larger differences are obtained upon comparison of the reaction energies stabilized (−27.7 and −2.3 kcal mol<sup>-1</sup> for PCIK and PFK, respectively). It is important to point out that the use of a semiempirical Hamiltonian to treat those atoms described quantum mechanically in our QM/MM scheme provokes an important overestimation of the free energy barriers, by comparison with those generated by DFT methods. Our

results confirm the great chemical reactivity of PCIK as an irreversible inhibitor of cruzain. Despite the fact that no experimental kinetic data are available for this particular reaction, our results would be in agreement with experimental studies of other papain-like enzymes (cathepsin B) that concluded that the reactivity of a PCIK is higher than that of a PFK. A deeper analysis of the results suggests that the origin of these differences can be in the different stabilizing interactions established between the inhibitors and the residues of the active site of the protein. Any attempt to explore the viability of the inhibition process through a stepwise mechanism involving the formation of a thiohemiketal intermediate and a three-membered sulfonium intermediate has been unsuccessful. All QM/MM calculations show that these intermediates are unstable, which does not lead to recommending the use of the intermediate as a template to design an efficient inhibitor. Nevertheless, an alternative stepwise mechanism has been determined through a protonated thiohemiketal intermediate. This species, which is generated through the transfer of a proton from a histidine residue located in the active site of cruzain protein, His159, and the attack of Cys25 on the CT atom of the inhibitor, appears to be dramatically stabilized with respect to the initial reactant species. Our results suggest that, in fact, this covalently bound inhibitor-protein complex could be a final state of the inhibition process because significant barriers are obtained in going from this complex to the final product complex.

Altogether, our results suggest that benzoyl-tyrosine-alanine-chloromethyl ketone is a good candidate for the development of a proper inhibitor of cruzain, which in turn could be used in the treatment of Chagas disease.

#### ■ ASSOCIATED CONTENT

##### ● Supporting Information

Time evolution of the root-mean-square deviation along the QM/MM MD simulation for the backbone atoms of the protein (red line) and atoms of the inhibitor (blue line) for both systems (Figure S1), ID-PMFs for inhibition of cruzain by PFK through mechanism I at the M06-2X/MM level with 6-31+G(d,p) and 6-311+G(2d,2p) basis sets (Figure S2), AM1d/MM ID-PMFs of the inhibition of cruzain by PHKs through mechanism I (Figure S3), AM1d/MM ID-PMFs of the formation of the TH intermediate from reactants for both inhibitors (Figure S4), AM1d/MM ID-PMFs of the formation of the TMSI intermediate from products for both inhibitors (Figure S5), time-dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TH of both inhibitors (Figure S6), time-dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TMSI of the Br-Tyr-Ala-CH<sub>2</sub>F inhibitor (Figure S7), time-dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TMSI of the Br-Tyr-Ala-CH<sub>2</sub>Cl inhibitor (Figure S8), AM1d/MM 2D-PMFs for the formation of the THH intermediate from reactants for both inhibitors (Figure S9), AM1d/MM 2D-PMFs for the formation of the IH intermediate from THH for both inhibitors (Figure S10), AM1d/MM 2D-PMFs for the formation of P from the IH intermediate for both inhibitors (Figure S11), and averaged distances for key states located along the inhibition of cruzain by PHK, through a mechanism III (Table S1). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/ht501551g.

#### ■ AUTHOR INFORMATION

##### Corresponding Authors

\*E-mail: sferenc@uji.es. Phone: +34964728084.

\*E-mail: moliner@uji.es.

##### Funding

We thank the FEDER Spanish Ministerio de Economía y Competitividad for Project CTQ2012-36253-C03-01, Generalitat Valenciana for the PrometesII/2014/022 project, and Universitat Jaume I for Project P11R2014-26. K.A. thanks the Spanish Ministerio de Economía y Competitividad for a predoctoral contract.

##### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

We acknowledge the Servei d'Informàtica, Universitat Jaume I, for a generous allotment of computer time.

#### ■ ABBREVIATIONS

1D, monodimensional; 2D, bidimensional; E64, N-[1-(hydroxycarbonyl)ethyl-carboxy]hexylamino-benzyl guanidine; FP2, fálcipain-2; HI, high-level; LL, low-level; MD, molecular dynamics; OPLS-AA, optimized potential for liquid simulations of the all-atom force field; PCIK, benzoyl-tyrosine-alanine-chloromethyl ketone; PDB, Protein Data Bank; PES, potential energy surface; FES, free energy surface; QM/MM, quantum mechanics/molecular mechanics; QM/MM-FE, quantum mechanics/molecular mechanics free energy; RC, reaction coordinate; r.m.d., root-mean-square deviation; TIP3P, transferable intermolecular potential 3P; TH, thiohemiketal; IH, stable intermediate; THH, protonated thiohemiketal; IH, protonated intermediate; I, unprotonated intermediate; TMSI, three-membered sulfonium intermediate; WHAM, weighted histogram analysis method.

#### ■ REFERENCES

- (1) Clayton, J. (2010) Chagas disease. *Nature* 465, 54–55.
- (2) <http://www.who.int/mediacentre/factsheets/fs340/en> (2014).
- (3) Jackson, Y., Argüelles, A., Carrero Fernández, B., Jansa I Lopez del Valado, J. M., Jarrín, J. G., and Albarrán-Vinaz, P. (2009) Mesogester of Chagas disease in Europe: Experiences and challenges in Spain, Switzerland and Italy. *Bull. Soc. Pathol. Exot. Sér. Fil.* 102, 326–334.
- (4) Castro, J. A., de Mico, M. M., and Bartel, L. C. (2006) Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Helm. Exp. Tassiol.* 28, 471–479.
- (5) Filardi, L. S., and Brenner, Z. (1987) Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas disease. *Trans. R. Soc. Trop. Med. Hyg.* 81, 755–759.
- (6) McKerrow, J. H., Engel, J. C., and Caffrey, C. E. (1999) Cysteine protease inhibitors as chemotherapy for parasitic infections. *Drug Met. Chem.* 7, 639–644.
- (7) McKerrow, J. H., McGrath, M. H., and Engel, J. C. (1995) The cysteine protease of *Trypanosoma cruzi* as a model for antiparasitic drug design. *Parasitol. Today* 11, 279–282.
- (8) Rendó, A. B., and McKerrow, J. H. (2006) Drug discovery and development for neglected parasitic diseases. *Nat. Chem. Biol.* 2, 701–710.
- (9) Eskin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., and Czulk, C. S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *J. Biol. Chem.* 267, 7411–7420.

- (10) Jiro, S., and Camargo, E. P. (1977) Proteolytic activities in cell extracts of *Trypanosoma cruzi*. *J. Protozool.* 24, 591–595.
- (11) Campetelli, O., Henttonen, J., Ahlund, L., Franck, A. C. C., Peterson, U., and Cazzulo, J. J. (1992) The major cysteine proteinase (Crupain) from *Trypanosoma cruzi* is encoded by multiple polymorphic tandemly repeated genes located on different chromosomes. *Mol. Biochem. Parasitol.* 50, 225–234.
- (12) Fujis, M., Mallin, J. P., Hansell, E. J., Mackay, Z., Doyle, P., Zhou, Y. M., Goh, J., Rosenthal, P. J., McKerrow, J. H., and Gay, R. K. (2005) Discovery of potent thioester-carbamate inhibitors of chondroitinase and crupain. *Bioorg. Med. Chem. Lett.* 16, 121–125.
- (13) Scharfstein, J., Schachter, M., Sereca, M., Peralta, J. M., Mendonça-Pereira, L., and Miles, M. A. (1988) *Trypanosoma cruzi* characterization and isolation of a 57/51000 MW surface glycoprotein (GP57/51) expressed by epimastigotes and blood stream trypomastigotes. *J. Immunol.* 137, 1536–1541.
- (14) Schenay, A. R., Bickhoff, C. S., Stevenson, D., Curtis, R., and Hoff, D. F. (2002) Crupain induces both mucosal and systemic protection against *Trypanosoma cruzi* in mice. *Infect. Immun.* 70, 5065–5074.
- (15) Cazzulo, J. J. (2002) Proteinases of *Trypanosoma cruzi*: Potential targets for the chemotherapy of Chagas disease. *Curr. Top. Med. Chem.* 2, 1261–1271.
- (16) Cazzulo, J. J., Steka, V., and Turk, V. (1997) Crupain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*. *Biol. Chem.* 378, 1–10.
- (17) Engel, J. C., Doyle, P. S., Hsieh, L., and McKerrow, J. H. (1998) Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *J. Exp. Med.* 186, 725–734.
- (18) Engel, J. C., Doyle, P. S., Palmer, J., Hsieh, L., Buzarov, D. F., and McKerrow, J. H. (1999) Cysteine protease inhibitors alter Golgi complex ultrastructure and function in *Trypanosoma cruzi*. *J. Cell Sci.* 112, 597–606.
- (19) Harth, G., Andrews, N., Mills, A. A., Engel, J. C., Smith, R., and McKerrow, J. H. (1993) Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 58, 17–24.
- (20) McKerrow, J. H. (1999) Development of cysteine protease inhibitors as chemotherapy for parasitic diseases: Insights on safety, target validation, and mechanism of action. *Int. J. Parasitol.* 29, 833–857.
- (21) Pukicelli, F., Zaini, G., Belli, A., Antonini, G., Gradoni, L., and Accanti, P. (2005) Probing the crupain 5-2 recognition substrate: A kinetic and binding energy calculation study. *Biochemistry* 44, 2781–2789.
- (22) Barr, S. C., Warner, E. L., Korneic, R. G., Pictalk, J., Wolfe, A., Beset, L., and McKerrow, J. H. (2005) A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi*. *Antimicrob. Agents Chemother.* 49, 5160–5161.
- (23) Brub, K., Doyle, P. S., McKerrow, J. H., and Hlman, J. A. (2008) Identification of a new class of nonpeptidic inhibitors of crupain. *J. Am. Chem. Soc.* 130, 6404–6410.
- (24) Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007) A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrob. Agents Chemother.* 51, 3932–3939.
- (25) Jacobsen, W., Christian, U., and Beset, L. Z. (2000) In vitro evaluation of the disposition of a novel cysteine protease inhibitor. *Drug Metab. Dispos.* 28, 1343–1351.
- (26) Roush, W. B., Cheng, J. M., Knapp-Reed, B., Abarrate-Hernandez, A., McKerrow, J. H., Hansell, E., and Engel, J. C. (2001) Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease crupain. *Bioorg. Med. Chem. Lett.* 12, 2759–2762.
- (27) Otto, H. H., and Schmeissner, T. (1997) Cysteine proteinases and their inhibitors. *Chem. Rev.* 97, 133–171.
- (28) Lecaillon, F., Kala, J., and Bommi, D. (2002) Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* 102, 4459–4488.
- (29) Schoellman, G., and Shaw, E. (1963) Direct evidence for presence of histidine in active center of chymotrypsin. *Biochemistry* 2, 232–235.
- (30) Rantick, D. (1985) Synthesis of peptide fluoromethyl ketones and the inhibition of human cathepsin-B. *Aust. Biochem.* 149, 461–465.
- (31) Rader, P., Anglifer, H., Wilkes, B., and Shaw, E. (1986) The synthesis of peptidylfluoromethylamides and their properties as inhibitors of serine proteinases and cysteine proteinases. *Biochem. J.* 239, 633–640.
- (32) Gilmore, S. A., Crak, C. S., and Fletcher, R. J. (1997) Structural determinants of specificity in the cysteine protease crupain. *Protein Sci.* 6, 1605–1611.
- (33) McGrath, M. E., Eakin, A. E., Engel, J. C., McKerrow, J. H., Crak, C. S., and Fletcher, R. J. (1995) The crystal structure of crupain: A therapeutic target for Chagas disease. *J. Mol. Biol.* 247, 251–259.
- (34) Gilles, A. M., and Kail, B. (1984) Evidence for an active-center cysteine in the 5H-proteinase  $\alpha$ -clostripin through use of N-tuoyl-L-lysine fluoromethyl ketones. *FEBS Lett.* 173, 58–62.
- (35) Felgar, L. (1973) Mode of activation of catalytically essential sulphydryl group of papain. *Biochem. J.* 135, 104–109.
- (36) Felgar, L., and Hagan, F. (1982) Current problems in mechanistic studies of serine and cysteine proteinases. *Biochem. J.* 207, 1–10.
- (37) Kellier, J. W., and Brown, R. S. (1992) Attack of cysteine protease inhibitors on a disordered amide as a model for the acylation of papain by amides: A simple determination of a bell-shaped pH rate profile. *J. Am. Chem. Soc.* 114, 7983–7989.
- (38) Brockhurst, K. (1979) Specific covalent modification of thiols. Applications in the study of enzymes and other biomolecules. *Int. J. Biochem.* 10, 259–274.
- (39) Azari, K., Ferrer, S., Martí, S., and Moliner, V. (2014) Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Falcipain-2 Inhibition by the Trypanosuccinate 264. *Biochemistry* 53, 3336–3346.
- (40) Barreiro, G., deAlencastro, R. B., and Neto, J. D. D. (1997) A semiempirical study on laprazep: An inhibitor of cysteine proteases. *Int. J. Quantum Chem.* 65, 1125–1134.
- (41) Mendes-Lacín, O., Roura-Marcillo, A., Medina-Franco, J. L., and Castiella, R. (2012) Computational study on the inhibitive mechanism of crupain by nitrile-containing molecules. *J. Mol. Graphics Modelling* 35, 28–35.
- (42) Mladjenovic, M., Amberg, K., Fiedl, R. F., Thiel, W., Schmeissner, T., and Engel, B. (2008) Aromatic insights into the inhibition of cysteine proteases: Firm QM/MM calculations clarifying the stereoselectivity of epoxide-based inhibitors. *J. Phys. Chem. B* 112, 11798–11808.
- (43) Shankar, R., Kolandavel, P., and Senthil Kumar, K. (2010) Reaction Mechanism of Cysteine Proteinase Model Compound HSH with Diketone Inhibitor PhCOC(O)CH<sub>2</sub>nX<sub>2</sub> (X = F, Cl, n = 0, 1, 2). *Int. J. Quantum Chem.* 110, 1660–1674.
- (44) Tarnowska, M., Chelini, S., Liao, A., Kania, P., Kasprzykowski, F., and Gronowicz, Z. (1992) MNDO study of the mechanism of the inhibition of cysteine proteinases by fluoromethyl ketones. *Eur. Biophys. J.* 21, 217–222.
- (45) Vučk, B., Helmut, H., Schmeissner, T., and Engel, B. (2006) Rational design of amide-containing cysteine protease inhibitors with improved potency: Studies on inhibition mechanism. *ChemMedChem* 1, 1021–1028.
- (46) Vijayarajam, S., and Kolandavel, P. (2008) Reaction mechanism of HSH and CH<sub>2</sub>SH with NH<sub>2</sub>CH<sub>2</sub>COC(O)CH<sub>2</sub>X (X = F and Cl) molecules. *Int. J. Quantum Chem.* 108, 927–936.
- (47) Powers, J. C., Anglin, J. L., Bacci, O. D., and Jansa, K. E. (2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* 102, 4639–4750.

- (48) Atad, D., Langridge, R., and Kollman, P. A. (1990) A simulation of the sulfur attack in the catalytic pathway of papain using molecular mechanics and semiempirical quantum-mechanics. *J. Am. Chem. Soc.* 112, 491–502.
- (49) Howard, A. E., and Kollman, P. A. (1984) OH versus SH nucleophilic attack on amides: Dramatically different gas-phase and solvation energies. *J. Am. Chem. Soc.* 106, 7195–7200.
- (50) Han, W. G., Tajkhanlou, R., and Subai, S. (1999) QM/MM study of the active site of free papain and of the NMA-papain complex. *J. Biomol. Struct. Dyn.* 16, 1019–1032.
- (51) Byan, K., and Gao, J. L. (2000) A combined QM/MM study of the nucleophilic addition reaction of methanethiolate and N-methylacetamide. *J. Mol. Graph. Model.* 18, 50–55.
- (52) Strahl, M., Florian, J., and Warshel, A. (2001) Ab initio evaluation of the free energy surfaces for the general base/acid catalyzed thiolysis of formamide and the hydrolysis of methyl thioformate: A reference reaction for studies of cysteine proteases. *J. Phys. Chem. B* 105, 4471–4484.
- (53) Wei, D. H., Huang, X. Q., Liu, J., Tang, M. S., and Zhao, C. G. (2011) Reaction Pathway and Free Energy Profile for Papain-Catalyzed Hydrolysis of N-Acetyl-Phe-Gly + Nitroamide. *Biochemistry* 52, 5145–5154.
- (54) Warshel, A., and Levitt, M. (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of carbocationic reaction intermediates. *J. Mol. Biol.* 103, 227–249.
- (55) Field, M. J., Bash, P. A., and Karplus, M. (1990) A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.* 11, 700–733.
- (56) Gao, J. L., Amara, P., Abambra, C., and Field, M. J. (1998) A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* 102, 4714–4721.
- (57) Field, M. J., Ajtey, M., Bretz, C., Proust-De Martin, F., and Thomas, A. (2008) The Dyanmo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* 29, 1084–1100.
- (58) Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pK<sub>a</sub> values for protein-ligand complexes. *Protein Struct. Funct. Bioinf.* 73, 765–783.
- (59) Li, H., Rohmeyer, A. D., and Jernag, J. H. (2005) Very fast empirical prediction and rationalization of protein pK<sub>a</sub> values. *Protein Struct. Funct. Bioinf.* 81, 704–721.
- (60) Olesen, M. H. M., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK<sub>a</sub> Predictions. *J. Chem. Theory Comput.* 7, 525–535.
- (61) www.pyrolog.org (2009).
- (62) Nara, K., Cai, Q., Gao, J., and York, D. M. (2007) Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* 7, 486–504.
- (63) Jorgensen, W. L., Maxwell, D. S., and TiradoRives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236.
- (64) Jorgensen, W. L., Chandrosskhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- (65) Singh, U. C., and Kollman, P. A. (1996) A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH<sub>3</sub>Cl + Cl<sup>-</sup> exchange reaction and gas-phase protonation of polyethers. *J. Comput. Chem.* 17, 718–730.
- (66) Marti, S., Molnar, V., and Turán, I. (2005) 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.* 5, 1008–1016.
- (67) Kumar, S., Boniolo, D., Swendsen, R. H., Kollman, P. A., and Rosenber, J. M. (1992) The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021.
- (68) Torrie, G. M., and Vallejos, J. P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 23, 187–199.
- (69) Riza-Ponja, J. J., Siba, E., Turán, I., Marti, S., and Molnar, V. (2004) Hybrid QM/MM potentials of mean force with integrated corrections. *J. Phys. Chem. B* 108, 8427–8433.
- (70) Chuang, Y. Y., Conchado, J. C., and Truhlar, D. G. (1999) 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, 1140–1149.
- (71) Zhao, Y., and Truhlar, D. G. (2008) 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, 215–241.
- (72) Helzer, W. J., Baden, L., von R. Schleyer, P., and Pople, J. (1986) *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York.
- (73) Lynch, B. J., Zhao, Y., and Truhlar, D. G. (2005) Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* 107, 1394–1398.
- (74) Frisch, M. J., Pople, J. A., and Brédler, J. S. (1984) Self-Consistent Molecular-Orbital Methods. 25. Supplementary functions for Gaussian-Basis Sets. *J. Chem. Phys.* 80, 3265–3269.
- (75) Krishnan, R., Binkley, J. S., Seeger, R., and Pople, J. A. (1980) Self-Consistent Molecular-Orbital Methods. 20. Basis set for correlated wave functions. *J. Chem. Phys.* 72, 650–654.
- (76) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalapino, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., C. Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukui, R., Hasegawa, J., Ishida, M., Nakajima, T., Horiuchi, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Ciolekowski, J., and Fox, D. J. (2009) Gaussian 09: version 9.1. Gaussian, Inc., Wallingford, CT.
- (77) Rasthans, N. D., Barrett, A. J., and Bateman, A. (2012) MERCPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 40, D343–D350.
- (78) Malthasso, J. P. G., Mackenzie, N. B., Boyd, A. S. P., and Scott, A. I. (1983) Detection of a tetrahedral adduct in a triphenylchloromethyl ketone specific inhibitor complex by C-13 NMR. *J. Am. Chem. Soc.* 105, 1685–1686.



### **7.2.1. Supporting Information**

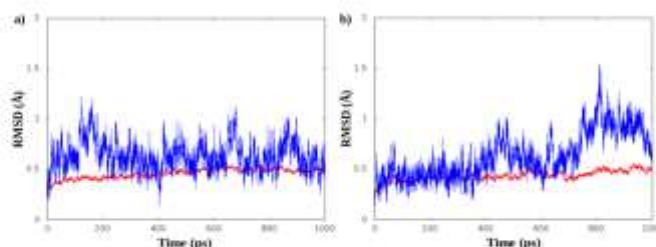
Supporting Information for the paper entitled: *First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones.*

**Supporting Information**

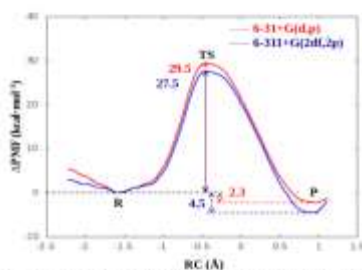
**First QM/MM Studies of Inhibition Mechanism of Cruzain by Peptidyl  
Halomethyl Ketones.**

*Kemel Arafet, Silvia Ferrer, Vicent Moliner*

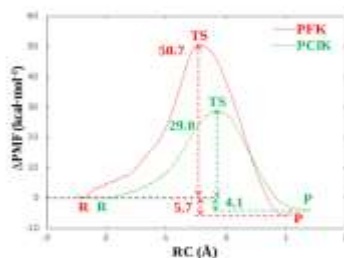
Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain



**Figure S1.** Time evolution of the root-mean-square-deviation (RMSD) along the QM/MM MD simulation for the backbone atoms of the protein (red line) and atoms of inhibitor (blue line) for (a) the Bz-Tyr-Ala-CH<sub>3</sub>F and (b) the Bz-Tyr-Ala-CH<sub>2</sub>Cl systems. Simulations performed on the protein-inhibitor bonded state (products).

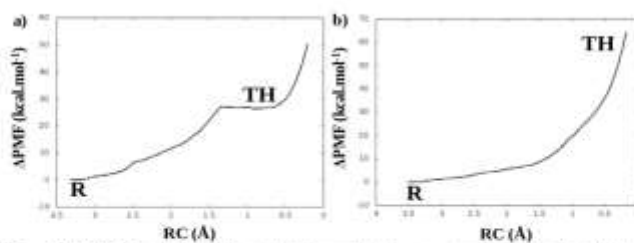


**Figure S2.** 1D-PMFs for cruzain inhibition by PFK through mechanism 1 computed at M062X/MM level with 6-31+G(d,p) and 6-311+G(2df,2p) basis set. RC corresponds to  $d(\text{F-CM})-d(\text{SG-CM})$ .

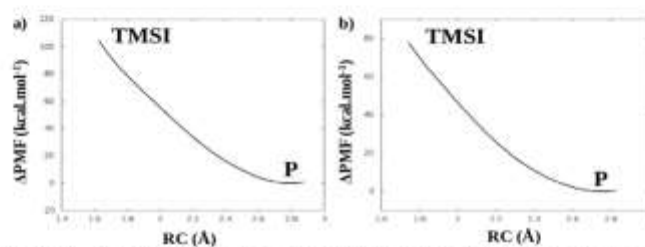


**Figure S3.** Free energy profiles of the cruzain inhibition with PHKs through mechanism 1 computed at AM1d/MM level. RC corresponds to  $d(\text{X-CM})-d(\text{SG-CM})$ , with X: F, Cl.

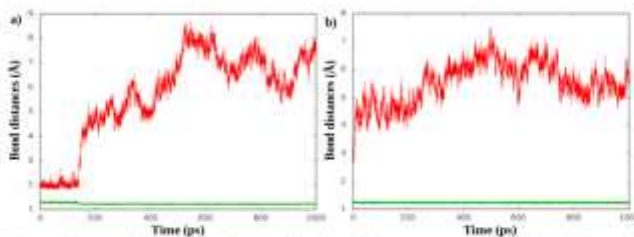
52



**Figure S4.** PMFs for the formation of TH intermediate from reactants computed at AM1d/MM level: a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor; and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. RC corresponds to  $d(\text{SG-CT})-d(\text{OT-CT})$ .

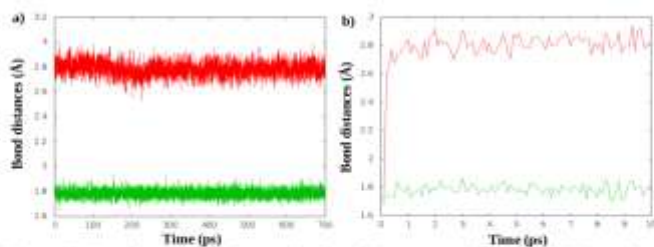


**Figure S5.** PMFs for the formation of TMSI intermediate from products, computed at AM1d/MM level: a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor; and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. RC corresponds to the distance  $d(\text{SG-CT})$ .

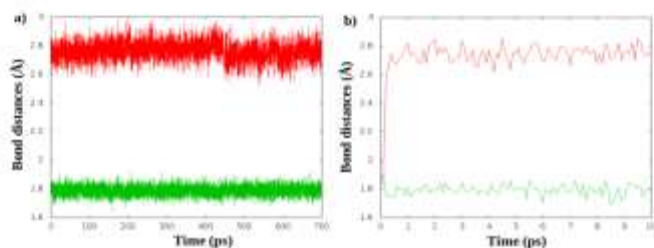


**Figure S6.** Time dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TH of the a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. The green line corresponds to the OT-CT distances, and the red line with the SG-CT distance.

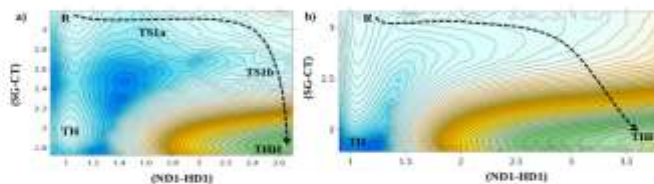
53



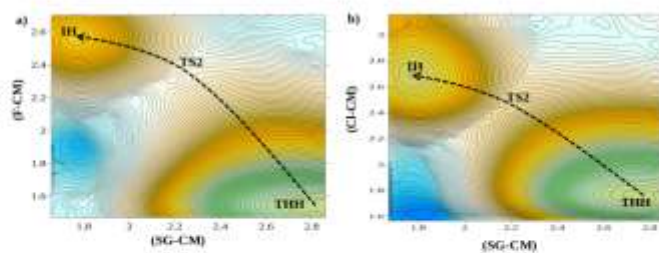
**Figure S7.** Time dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TMSI of the Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor. a) 700 ps of MD, b) first 10 ps of the 700 ps MD. The green line corresponds with the SG-CM distances, and the red line with the SG-CT distance.



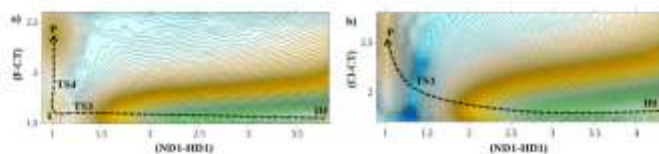
**Figure S8.** Time dependent evolution of selected distances on the QM/MM MD simulation corresponding with the TMSI of the Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. a) 700 ps of MD, b) first 10 ps of the 700 ps MD. The green line corresponds with the SG-CM distances, and the red line with the SG-CT distance.



**Figure S9.** 2D-PMFs for the formation of protonated thiohemiketal intermediate, THH, from reactants, computed at AM1d/MM level: a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor, and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in Å and iso-energetic lines are displayed every 1.0 kcal·mol<sup>-1</sup>.



**Figure S10.** 2D-PMFs for the transformation from protonated thiohemiketal, THH to intermediate IH, computed at AM1d/MM level: a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor, and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in Å and iso-energetic lines are displayed every 1.0 kcal·mol<sup>-1</sup>.



**Figure S11.** 2D-PMFs for the formation of products from IH intermediate computed at AM1d/MM level: a) Bz-Tyr-Ala-CH<sub>2</sub>F inhibitor, and b) Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitor. Distances are in Å and iso-energetic lines are displayed every 1.0 kcal·mol<sup>-1</sup>.

**Table S1.** Averaged distances for key states located along the inhibition of cruzain by PHK, through a mechanism III. Results were obtained from 200 ps of AM1d/MM MD simulations on the stationary points taking from the M06-2X/6-31+G(d,p)/MM free energy profiles (in Å). Intramolec distances associated to the RCs were constrained during the simulations.

a) PHK.

CO	4	40	104	168	232	300	364	428	492	556
SC-CT	4.23±0.00	4.38±0.04	5.17±0.10	2.69±0.06	2.27±0.03	1.76±0.02	1.77±0.03	1.76±0.03	1.80±0.04	1.80±0.03
SC-CT	3.07±0.01	3.07±0.01	2.96±0.01	1.83±0.02	1.86±0.04	2.03±0.06	2.40±0.08	2.63±0.04	2.77±0.05	2.81±0.08
P-CT	1.55±0.02	1.66±0.03	1.64±0.02	1.64±0.02	2.24±0.03	2.17±0.02	2.16±0.04	2.16±0.04	2.70±0.05	2.81±0.08
P-CT	2.53±0.04	2.51±0.03	2.32±0.03	2.43±0.03	1.38±0.04	1.16±0.02	1.16±0.02	1.44±0.03	1.84±0.03	2.40±0.03
CT-CT	1.23±0.01	1.35±0.02	1.31±0.02	1.42±0.02	1.41±0.01	1.44±0.03	1.42±0.01	1.33±0.01	1.28±0.01	1.24±0.01
CT-101(101/10)	4.09±0.25	3.93±0.03	3.91±0.03	3.98±0.03	3.98±0.03	3.97±0.03	3.99±0.04	4.63±0.22	3.71±0.48	4.47±0.54
WD-101(101/10)	1.03±0.02	1.73±0.03	1.93±0.03	3.03±0.03	7.13±0.03	5.26±0.55	3.35±0.03	3.80±0.03	3.80±0.03	3.80±0.03
R-101(101/10)	6.27±0.17	4.44±0.07	3.45±0.17	5.31±0.25	1.91±0.51	2.09±0.13	2.52±0.06	3.03±0.13	6.48±0.35	8.33±0.38
F-101	2.67±0.14	3.03±0.13	4.76±0.08	5.26±0.19	5.03±0.14	4.62±0.08	4.63±0.07	6.77±0.15	3.62±0.12	3.03±0.15
F-101	2.07±0.00	2.97±0.13	3.00±0.18	3.04±0.18	3.07±0.11	3.33±0.20	3.35±0.13	2.71±0.10	3.20±0.11	2.76±0.12
1C-101(10)	2.26±0.30	1.23±0.11	1.50±0.11	3.31±0.34	4.03±0.20	5.03±0.30	3.61±0.27	3.62±0.28	3.34±0.71	2.11±0.34
2C-101(101/10)	1.97±0.10	2.08±0.14	3.13±0.15	2.05±0.13	3.16±0.21	5.55±0.41	3.71±0.14	5.86±0.14	4.12±0.44	3.07±0.25
3C-101(101/10)	3.38±0.17	2.68±0.22	3.27±0.31	2.07±0.20	2.43±0.40	3.03±0.22	3.64±0.20	4.43±0.41	3.84±0.73	3.10±0.48
11-101(10)	6.08±0.10	5.66±0.17	4.01±0.34	5.61±0.45	6.38±0.32	4.16±0.17	4.18±0.36	3.83±0.36	8.01±0.27	5.47±0.62
11-101(10)	5.62±0.25	3.37±0.12	3.60±0.17	4.34±0.03	3.33±0.41	3.09±0.60	3.72±0.42	3.83±0.40	4.72±1.09	3.98±1.14
12-101(10)	2.37±0.25	4.02±0.32	3.49±0.36	5.25±0.17	3.33±0.46	3.99±0.17	3.91±0.24	4.34±0.34	3.70±0.43	3.42±0.61
13-101(10)	4.18±0.33	2.77±0.32	3.54±0.42	3.61±0.56	3.58±0.42	3.24±0.42	3.89±0.45	3.36±0.34	3.76±0.44	4.73±0.44
11-101(10)	4.26±0.27	3.43±0.27	3.42±0.48	3.77±0.25	4.18±0.34	3.86±0.30	3.71±0.24	3.86±0.27	4.49±0.40	3.81±0.38
10-101(10)	2.04±0.25	2.74±0.10	4.11±0.29	3.02±0.03	2.08±0.01	3.31±0.17	3.34±0.25	4.57±0.26	3.33±0.06	2.91±0.30
11-101(10)	6.03±0.20	3.78±0.24	4.30±0.79	3.73±0.22	3.99±0.25	3.38±0.41	3.43±0.25	3.12±0.24	4.31±0.40	3.38±1.04
11-101(10)	4.89±0.41	1.92±0.14	2.75±0.38	2.18±0.33	2.28±0.28	2.22±0.22	2.14±0.20	3.83±0.20	2.44±0.42	1.76±0.68
10-101(10)	6.27±0.25	5.98±0.28	6.31±0.31	5.75±0.18	5.69±0.21	5.66±0.57	4.66±0.30	3.91±0.16	6.31±0.23	5.90±0.18
12-101(10)	3.33±0.26	3.03±0.20	5.86±0.43	4.06±0.31	3.32±0.41	3.09±0.41	3.09±0.20	3.77±0.23	3.96±0.48	4.52±0.38

b) PCIK.

d (Å)	R	THH	TS2	HI	TS1	P
SG-CM	3.57 ± 0.21	2.78 ± 0.06	1.98±0.02	1.77±0.02	1.79±0.04	1.79±0.03
SG-CT	3.05 ± 0.03	1.88 ± 0.02	2.90±0.07	2.77±0.05	2.81±0.05	2.77±0.06
CI-CM	1.77 ± 0.03	1.77 ± 0.03	2.55±0.02	2.69±0.02	2.77±0.06	2.96±0.13
CI-CT	2.71 ± 0.05	2.75 ± 0.05	1.80±0.04	1.81±0.04	1.94±0.02	2.57±0.03
OT-CY	1.24 ± 0.02	1.42 ± 0.02	1.41±0.03	1.41±0.02	1.37±0.03	1.24±0.02
OT-HD1 (His159)	3.74 ± 0.36	0.97 ± 0.03	0.98±0.03	0.97±0.03	1.10±0.04	4.86±0.34
ND1-HD1 (His159)	1.03 ± 0.02	3.70 ± 0.03	6.14±0.42	5.86±0.60	1.43±0.03	1.00±0.02
CI-HD1 (His159)	3.93 ± 0.35	2.58 ± 0.18	3.05±0.22	2.87±0.20	2.81±0.11	4.66±0.51
SG-H (Tyr26)	2.99 ± 0.23	3.16 ± 0.19	2.66±0.25	2.93±0.27	3.51±0.44	3.80±0.24
SG-H (His159)	1.93 ± 0.09	2.93 ± 0.11	3.91±0.23	3.81±0.23	3.15±0.20	3.93±0.52
SG-H (Gly160)	2.60 ± 0.21	3.00 ± 0.26	3.29±0.33	3.40±0.35	3.74±0.59	2.80±0.32
OT-H (Cys25)	3.48 ± 0.36	2.51 ± 0.23	4.10±0.91	2.89±0.34	4.53±0.37	3.07±0.56
OT-H (Gln19)	3.90 ± 0.63	2.65 ± 0.33	4.72±0.50	4.00±0.62	5.48±0.56	3.51±0.91
O1-H (Gly66)	3.50 ± 0.39	3.29 ± 0.35	3.75±0.65	3.27±0.58	3.06±0.40	3.72±0.42
O1-H (Tyr26)	1.71 ± 0.24	1.66 ± 0.24	1.76±0.36	2.03±0.62	1.90±0.39	1.53±0.24
O3-H (Gly66)	2.95 ± 0.47	3.21 ± 0.38	2.84±0.52	3.21±0.52	3.54±0.40	3.26±0.64
N1-H (Gly66)	2.71 ± 0.22	2.47 ± 0.21	2.96±0.29	2.80±0.30	2.48±0.24	2.95±0.26
H1-O (Gly66)	1.94 ± 0.15	1.92 ± 0.15	1.93±0.19	1.96±0.18	2.00±0.18	1.97±0.19
H1-N (Gly66)	3.23 ± 0.24	3.09 ± 0.22	3.47±0.28	3.40±0.30	3.13±0.24	3.51±0.28
H2-O (Asp158)	2.92 ± 0.33	3.10 ± 0.29	3.44±0.57	3.21±0.35	4.05±0.37	3.65±0.46



### 7.3. Paper III

Paper entitled: *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Cysteine Protease Inhibition by Peptidyl-2,3-epoxyketones.*

Kemel Arafet, Silvia Ferrer, Florenci V. González, and Vicent Moliner

Physical Chemistry Chemical Physics **2017**, 19, 12740-12748.

DOI: 10.1039/C7CP01726J.

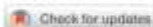
Reproduced by permission of The Royal Society of Chemistry.

<http://pubs.rsc.org/en/content/articlelanding/2017/cp/c7cp01726j#!divAbstract>



PCCP

PAPER

View Article Online  
View Journal | View IssueCite this: *Phys. Chem. Chem. Phys.*,  
2017, **19**, 12740

## Quantum mechanics/molecular mechanics studies of the mechanism of cysteine protease inhibition by peptidyl-2,3-epoxyketones†

Keriel Arafet,<sup>a</sup> Silvia Ferrer,<sup>a</sup> Florenci V. González<sup>b</sup> and Vicent Moliner<sup>a\*</sup>

Cysteine proteases are the most abundant proteases in parasitic protozoa and they are essential enzymes to the life cycle of several of them, thus becoming attractive therapeutic targets for the development of new inhibitors. In this paper, a computational study of the inhibition mechanism of cysteine protease by dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph(S), a recently proposed inhibitor, has been carried out by means of molecular dynamics (MD) simulations with hybrid QM/MM potentials. The computed free energy surfaces of the inhibition mechanism of cysteine proteases by peptidyl epoxyketones showing how the activation of the epoxide ring and the attack of Cys25 on either C2 or C3 atoms take place in a concerted manner. According to our results, the acid species responsible for the protonation of the oxygen atom of the ring would be able to conserve His159. In contrast to previous studies that proposed a water molecule as the activating species, the low activation free energies for the reaction where Cys25 attacks the C2 atom of the epoxide ring (2.3 kcal mol<sup>-1</sup>) or to the C3 atom (3.4 kcal mol<sup>-1</sup>), together with the high negative reaction energies suggest that the derivatives of peptidyl-2,3-epoxyketones can be used to develop new potent inhibitors for the treatment of Chagas disease.

Received 17th March 2017,  
Accepted 19th April 2017

DOI: 10.1039/c7cp01729g

rsc.li/pccp

### Introduction

Cysteine proteases are the most abundant proteases in many parasitic protozoa and they are essential enzymes to the life cycle of several of them. For example, cruzain is crucial for the development and survival of the protozoan *Trypanosoma cruzi*, the etiologic agent of Chagas disease,<sup>1–3</sup> affecting six to seven million people worldwide mostly in Latin America.<sup>4</sup> Hence, cysteine proteases have become attractive therapeutic targets for the development of new inhibitors. A diverse number of peptide molecules with a group susceptible to suffer from a nucleophilic attack can inhibit cysteine proteases. Inhibitors such as vinyl sulfones, halo methyl ketones, aziridines, nitriles,

enones, nitroalkenes or epoxides have shown good inhibition activity.<sup>5–10</sup> In 1998, Roush *et al.* designed and synthesized a new class of cysteine protease inhibitors,<sup>7</sup> dipeptidyl-2,3-epoxyketone as represented in Scheme 1 (Cbz-Phe-Hph(S) epoxyketone, where Hph is homophenylalanine). This potent and irreversible inhibitor was designed by combining a portion of the epoxide moiety of the known epoxy-succinyl inhibitor E64c<sup>11</sup> and the peptide sequence of the dipeptidyl fluoromethyl ketone, Cbz-Phe-Ala-CH<sub>2</sub>F.<sup>12</sup> In fact, the second-order rate constant (*k*<sub>app</sub>) for the inhibition of cruzain by Cbz-Phe-Hph(S) epoxyketone proved to be 4.5-fold greater than that by E64c (333 000 M<sup>-1</sup> s<sup>-1</sup> and 70 600 M<sup>-1</sup> s<sup>-1</sup> for Cbz-Phe-Hph(S)<sup>8</sup> and E64c,<sup>13</sup> respectively).

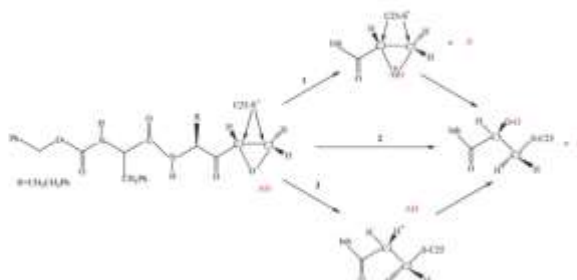
Several studies have been performed to establish an order of reactivity of different substituents in the C2 atom of the epoxy ring.<sup>14–17</sup> In 1992, Ron Bihovsky, in an experimental study of the reactions of 2,3-epoxy carbonyl compounds in solution with methanethiolate, observed for this non-enzymatic thiolate that the reactivity decreases in the order COCH<sub>3</sub> > CO<sub>2</sub>CH<sub>3</sub> > CONH<sub>2</sub> > H > CO<sub>2</sub>H.<sup>16</sup> However, the following year, Bihovsky *et al.* synthesized and tested several epoxy-succinates as inhibitors of papain, and observed the inverse trend for the inhibition rate constants: CO<sub>2</sub>H > CONH<sub>2</sub> > CO<sub>2</sub>R > COCH<sub>3</sub> > H.<sup>14</sup> They concluded that the electrostatic attraction between the protonated His159 (cruain numbering) and the carboxylate of the inhibitor facilitates the docking of the inhibitor at the active site of papain,

<sup>a</sup>Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain. E-mail: moliner@qf.uji.es; Tel.: +3496723994

<sup>b</sup>Departament de Química Inorgànica i Orgànica, Universitat Jaume I, 12071 Castelló, Spain

<sup>†</sup>Department of Chemistry, University of Bath, Bath BA2 7AY, UK

† Electronic supplementary information (ESI) available: Free energies of the BMSD of backbone atoms of the protein and the inhibitor, free energy surfaces compared as ID-PMF and ID-PMF at the AM16MM level, representative snapshots of key states of the reaction derived from the QM/MMs, selected averaged distances obtained for the key states located along the ID-PMFs, representative snapshots of the states involved in the reaction with the water molecule acting as acid. See DOI: 10.1039/c7cp01729g



**Scheme 1** Proposed general mechanisms of inhibition of cysteine protease by dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph(S). The mechanism only refers to the Cys25 attack on the C2 atom, but as indicated with the arrows, the Cys25 attack can also occur on C3 atoms.

thus the rate-limiting step in the enzyme inhibition is not the reactivity of the epoxide, but rather appears to be the rate at which the inhibitor docks in the active site prior to the formation of the covalent complex. Later, Engels *et al.*, in a theoretical study about the influence of different ring substituents on the nucleophilic ring opening of three-membered heterocyclic, concluded that the acidic substituent increases the reaction barrier of the allylation step for epoxides.<sup>14</sup>

The inhibition mechanism of cysteine proteases by peptidyl epoxyketones does not yet seem to be clear. Albeck *et al.* proposed that it is based on an enzyme-catalyzed allylation of the active-site cysteine from a Michaelis complex between the enzyme and the inhibitor.<sup>18–20</sup> They suggested that the protonation of the oxirane ring by His159 occurs first or in a concerted manner with the nucleophilic attack on the amide scissile bond (mechanisms 1 and 2 in Scheme 1). Albeck and Kliper proposed that the inhibition mechanism of cysteine proteases by epoxysuccinyl E64 is different from the inhibition by peptidyl epoxides,<sup>18</sup> whereas the peptidyl epoxides are simple epoxides that interact directly with the thiolate, E64 is an activated epoxide that interacts with the enzyme through an initial attack on the carbonyl carbon, in analogy with the inhibition mechanism by peptidyl diazomethanes.<sup>19</sup> However, several authors are agreeing that epoxide is more likely protonated by a water molecule than by the His159 residue.<sup>17,21</sup> Kim *et al.*, in a mechanistic study of the inhibition of papain with 2-benzyl-3,4-epoxybutanoic acid esters,<sup>22</sup> concluded that the sulfur attack of Cys25 on the oxirane ring can occur without a prior protonation of oxirane (mechanism 3 in Scheme 1). Engels *et al.*, in several theoretical studies,<sup>15,22–24</sup> proposed that the protonation of the oxirane ring takes place far behind the transition state (TS). In our laboratory, theoretical studies, based on multiscale quantum mechanical/molecular mechanical (QM/MM) potentials,<sup>25–28</sup> were carried out to study the inhibition mechanism of cysteine protease falcipain-2 (FP2) by E64.<sup>29</sup> Our results indicated that after the attack of cysteine, an intramolecular hydrogen transfer from the carboxylic group of E64 to the oxygen atom of the opened epoxy

ring takes place. Clearly, this protonation is not possible with the Cbz-Phe-Hph(S) epoxyketone inhibitor due to the absence of the carboxylic group. An alternative mechanism, in which the protonation of the epoxy ring occurs prior to the attack of cysteine on C2 or C3 atoms, was unsuccessful. In no case the cysteine attacked the carbonyl carbon.

Another interesting question related to the inhibition molecular mechanism is that the cysteine attack can take place on either the C2 or C3 atom, depending on the orientation of the epoxy ring at the active site.<sup>7</sup> Bibovsky observed that the C3 attack is preferred for 2,3-epoxy amides, while for 2,3-epoxy esters and carboxylic acids the cysteine can attack both on C2 or C3 atoms.<sup>20</sup> However, Engels *et al.*, based on the exploration of potential energy surfaces (PES) obtained with hybrid QM/MM potentials, concluded that for epoxysuccinyl peptides with a free acid at C2, the attack preferably takes place on the C2 atom.<sup>18</sup> Our results on the inhibition mechanism of FP2 by E64 indicated that the irreversible attack of a cysteine on E64 can take place on both carbon atoms (C2 and C3) of its epoxy ring since both processes present almost similar barriers.<sup>25</sup> To date, the inhibition mechanism of cysteine protease by peptidyl epoxyketones has not been studied by computational tools explicitly including the protein environment effects.

The principal aim of this paper is to gain insights into the inhibition mechanism of cysteine protease by dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph(S). In previous theoretical studies in our laboratory, the inhibition mechanisms of a cysteine protease FP2 by epoxysuccinate E64,<sup>28</sup> and eruzain by peptidyl halomethyl ketones (PHKs)<sup>30</sup> were carried out. We proposed that benzoyl-tyrosine-alanine-chloromethyl ketone (PCCK) is a good candidate to develop a proper inhibitor of cysteine protease.<sup>30</sup> The main difference between these inhibitors and Cbz-Phe-Hph(S) is the electrophilic functional group, a chloro atom in PCCK and an epoxide ring together with a carboxylic group in E64. Regarding the enzymes, both proteins belong to the same family of cysteine proteases (clan CA) but they are responsible for different human diseases; FP2 is associated with malaria and eruzain with Chagas disease. Hybrid QM/MM potentials have been employed in the

PCCP

View Article Online

Paper

present study to explore the PESs and to run molecular dynamics (MD) simulations that allow obtaining the free energy surfaces of the reaction in terms of the Potential of Mean Force (PMF). The activation free energies can be directly compared with experimentally measured rate constants, while the predicted averaged geometries and the interactions between the inhibitor and the protein can be used to improve the design of cruzain inhibitors.

### Computational model

The molecular model was constructed from the X-ray crystal structure of cruzain from *Trypanosoma cruzi* with PDB entry 1AM1<sup>31</sup> and a resolution of 2.0 Å, by replacing the bound inhibitor Bz-Tyr-Ala-CH<sub>2</sub>F with the Cbz-Phe-Hph-[5] molecule (see Fig. 1). The final system consists of one chain of 215 amino acids of cysteine protease cruzain and the inhibitor.

Cruzain is located in the intracellular vesicles during the dormant stage of the parasite at a slightly acidic pH. During the intracellular amastigote stage of the *T. Cruzi* life cycle, cruzain is present on the surface of the parasite where the pH of the host cytoplasm is almost neutral (pH 7.4).<sup>32</sup> Thus, and because the crystallographic structure does not contain the position of the hydrogen atoms, these were added at this physiological pH using the fDYNAMO library,<sup>33</sup> according to the pK<sub>a</sub> values of the titratable residues calculated within the empirical PROPKA 3.1 program of Jensen *et al.*<sup>34–36</sup> A total of 11 counter ions (Na<sup>+</sup>) were placed into optimal electrostatic positions around the protein, in order to obtain electro neutrality of the full system. Finally, the system was placed in a 79.5 Å side cubic box of water molecules.

As depicted in Fig. 2, the inhibitor, Cys25, the imidazole ring of His159, and a water molecule (W atoms) were described by means of the AM10 semiempirical Hamiltonian.<sup>36</sup> The rest of the protein and water molecules were described by OPLS-AA<sup>37</sup> and TIP3P<sup>38</sup> force fields, respectively. To saturate the valence of the QM/MM frontier atoms, three link atoms<sup>38,39</sup> were introduced in the positions indicated in Fig. 2. Because of the size of the full system, all residues more than 25 Å from the CA1 atom of the inhibitor were kept frozen during the simulations (43 671 atoms

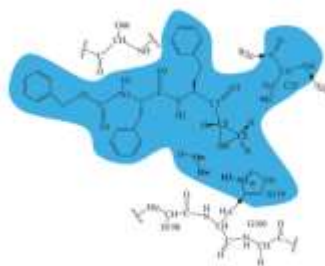


Fig. 2 Details of the active site corresponding to the studied model. The blue region corresponds to the QM subset of atoms that includes the inhibitor Cys25, the imidazole ring of His159, and a water molecule. The link atoms between the QM and MM regions are indicated as black dots.

from a total of 50 028). A force switch function with a cutoff distance in the range of 14.5–18 Å was applied to treat the non-bonding interactions. All the QM/MM calculations were carried out using the fDYNAMO library.<sup>33</sup>

The system was relaxed by means of 1 ns of QM/MM MD at 300 K using the NVT ensemble and the Langevin–Verlet integrator. The analysis of the time evolution of the root-mean-square-deviation of backbone atoms of the protein and the inhibitor confirms that the system was equilibrated (see Fig. S1 of the ESI<sup>†</sup>). Structures obtained after relaxation were used to generate hybrid AM10/MM PESs. Once the PESs were obtained, PMFs were traced by means of a series of molecular dynamics simulations in which a distinguished reaction coordinate is constrained around particular values by the umbrella sampling procedure,<sup>40</sup> while the remaining degrees of freedom (including those of the protein environment) are conveniently sampled. The values of the variables sampled during the simulations are then pieced together to construct a distribution function using the weighted histogram analysis method (WHAM).<sup>41</sup>

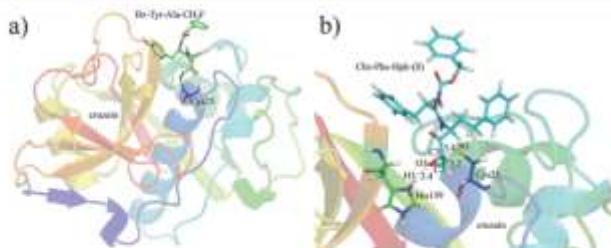


Fig. 1 (a) Crystal structure of cruzain from *T. cruzi* bound to Bz-Tyr-Ala-CH<sub>2</sub>F (PDB code 1AM1). (b) Detail of the active site of the studied model cruzain and the dipyrrolyl-2,5-epoxyketone Cbz-Phe-Hph-[5] inhibitor, after docking and equilibration (see the text for details). The inhibitor and the key residues Cys25 and His159 are represented by thick sticks, while key interactions are represented by dashed lines. Distances are in Å.

For the study of the possible stepwise mechanism where Cys25 attacks the epoxy ring before the activation of the ring by a proton transfer (mechanism 3 in Scheme 1), monodimensional PMFs (1D-PMFs) at the AM1d/MM level were obtained using the antisymmetric combination of the bond-forming and bond-breaking distances:  $\zeta = [d(\text{SG}-\text{CX}) - d(\text{O3}-\text{CX})]$ , where CX represents C2 or C3. The labels of the epoxide ring (C2, C3 and O3) have been assigned for facilitating direct comparisons with previous inhibitors such as E64. These required series of 33 and 30 simulation windows for the C2 and C3 attacks. For the study of the concerted mechanism or the stepwise mechanism where the activation of the epoxy ring precedes the Cys25 attack (mechanism 1 and 2 in Scheme 1), bidimensional PMFs (2D-PMFs) at the AM1d/MM level were computed using the SG-CY bond-forming distance as  $\zeta_1$  (where CY represents C2 or C3 atoms) and the O3-H3 bond-forming distance as  $\zeta_2$ . This required a series of 600 and 360 simulation windows for the attack of Cys25 on C2 and C3 respectively. The value of the force constant used for the harmonic umbrella sampling to generate PMFs was  $2500 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$  and the simulation windows consisted of 20 ps of equilibration and 40 ps of production, with a time step of 1 fs, in all PMFs. Subsequently, 200 ps of AM1d/MM MD simulations constrained at the stationary points located on the free energy surfaces (FESs) were performed to analyze the main geometrical parameters as averages.

Additionally, to investigate a possible mechanism in which a water molecule is the acid species responsible for the protonation of the oxirane ring, 2D-PMFs at the AM1d/MM level were computed using the SG-CX bond-forming distance as  $\zeta_1$  (where CX represents C2 or C3 atoms) and the antisymmetric combination of the  $\text{O}_w-\text{H}_w$  bond-breaking and the  $\text{O3}-\text{H}_w$  bond-forming distances as RC2, with  $\text{H}_w$  and  $\text{O}_w$  being the proton and the oxygen atom of the water molecule. This required series of 585 and 720 simulation windows for the attack of Cys25 on C2 and C3, respectively.

Because of the large number of structures that must be evaluated to generate the free energy calculations, QM/MM calculations are usually restricted to the use of semiempirical Hamiltonians. In order to correct the low-level AM1d energy function used in the 1D- and 2D-PMFs, an interpolated correction scheme developed in our laboratory has been applied.<sup>42</sup> This scheme, based on a method proposed by Truhlar *et al.* for dynamical calculations of gas phase chemical reactions,<sup>43–45</sup> uses a spline under tension<sup>46,47</sup> to interpolate the correction term at any value of the RCs selected to generate the FESs. Thus, the new energy function for a 1D-PMF is defined as

$$\tilde{E} = E_{\text{CL}/\text{MM}} + S[\Delta E_{\text{CL}}^{\text{HL}}(\zeta)] \quad (1)$$

where  $S$  denotes a spline function, whose argument  $\Delta E_{\text{CL}}^{\text{HL}}(\zeta)$  is a correction term taken as the difference between single-point calculations of the QM subsystem using a high-level (HL) method, and the low-level (LL) one. The correction term is expressed as a function of the distinguished reaction coordinate  $\zeta$ . When corrections were done on 2D-PMF, the correction term in eqn (1) is then expressed as a function of the two coordinates,  $\zeta_1$  and  $\zeta_2$ , to generate the higher level 2D-PMF. The AM1d

Hamiltonian was the LL method, and the HL calculations were carried out by means of the M06-2X functional<sup>48</sup> with the 6-31+G(d,p) basis set<sup>49</sup> following Truhlar *et al.*<sup>46,47</sup> These calculations were carried out using the Gaussian09 program.<sup>51</sup>

## Results and discussion

As mentioned in the previous section, the inhibition mechanism of the cysteine protease craxin by Chx-Phe-Hph-[8] has been studied by exploring FESs corresponding to the stepwise and concerted mechanisms depicted in Scheme 1. The resulting profiles assuming that the protonation of the oxirane ring takes place far behind the TS, as suggested by Engels *et al.*,<sup>15,31–33</sup> were obtained in terms of 1D-PMFs at the M06-2X/6-31+G(d,p):AM1d/MM level and are shown in Fig. 3. The corresponding 1D-PMFs at the AM1d/MM level are shown in Fig. S2 of the ESI,<sup>†</sup> representative snapshots of the three states are shown in Fig. S3, and Table S1 (ESI<sup>†</sup>) lists the average values of some key distances. According to the TS structures located on the FESs presented in Fig. 3, the attack of Cys25 on either C2 or C3 atoms of the epoxy ring and the protonation of O3 occur in a concerted manner. The principal conclusion that can be drawn from Table S1 and Fig. S3 (ESI<sup>†</sup>) is that the protonation of O3 spontaneously occurs when the reaction is controlled by just the antisymmetric combination of the SG-CX and O3-CX distances (where CX represents C2 or C3), as it can be confirmed by the structures of the products in both possible paths, with the source of the proton being the residue His159. Nevertheless, the average values of the distance O3-H<sub>159</sub> in the TS-C2 ( $2.13 \pm 0.03 \text{ \AA}$ ) and in the TS-C3 ( $0.98 \pm 0.03 \text{ \AA}$ ) reveal a different degree of synchronicity: while the proton is almost completely transferred in the TS of the reaction path corresponding to the attack of Cys25 on C3, the proton transfer from His159 to O3 takes place clearly after the TS is reached when the attack occurs on C2. It is important to point out that the distance O3-H<sub>159</sub> in both reactant complexes ( $2.14 \pm 0.03 \text{ \AA}$  in R-C2 and

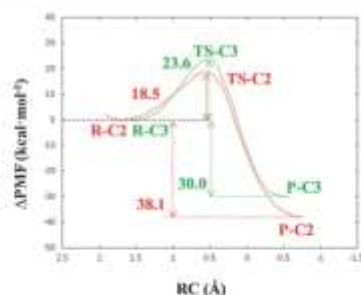


Fig. 3 2D-PMFs of the mechanism II computed at the M06-2X/6-31+G(d,p):AM1d/MM level. Cys25 attack on the C2 atom (red line) or C3 atom (green line). RC corresponds to  $d(\text{SG}-\text{C}2) - d(\text{O}3-\text{C}2)$ , with CX, C2, C3.

$2.31 \pm 0.03$  Å in R-C3) is shorter than the distance with any other possible proton donor at the active site, as the water molecule (the O3-H<sub>w</sub> distance is  $4.24 \pm 0.70$  and  $1.27 \pm 0.67$  Å in R-C2 and R-C3, respectively). This result initially supports the role of the His159 as the acid species. Interestingly, His159 suffers a dramatic displacement from the epoxide once the transfer takes place, as revealed in the interatomic distance between O3 and the N3 atom of His159 listed in Table S1 (ESI<sup>†</sup>).

The free energy profiles of Fig. 3 show both reactions as very exergonic processes ( $18.1$  and  $30.0$  kcal mol<sup>-1</sup> for the attack on

C2 and C3, respectively) thus supporting the inhibitory character of the epoxide. Nevertheless, significantly different free energies of activation are obtained;  $18.5$  and  $23.6$  kcal mol<sup>-1</sup> for the attack on C2 and C3, respectively. Within these results, it could be concluded that the reaction preferentially takes place through the attack on the C2 carbon atom with the protonation of the O3 atom of the epoxide taking place in a concerted but very asynchronous way. Nevertheless, taking into account the fact that the proton transfer from His159 to O3 has not been included in the distinguished reaction coordinate used to generate the

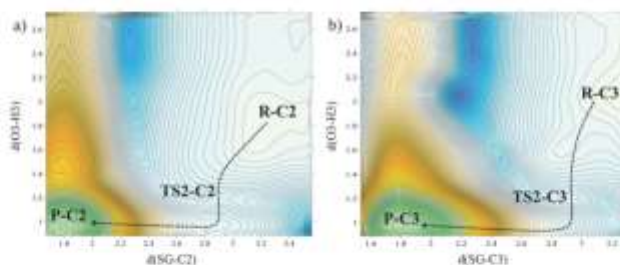


Fig. 4 FEPs computed at the M06-2X/6-31+G(d,p)/AM03/MBM level for the Cys25 attack on the (a) C2 atom and (b) C3 atom, of the inhibitor. Distances are in Å and iso-energetic lines are displayed every 1.0 kcal mol<sup>-1</sup>.

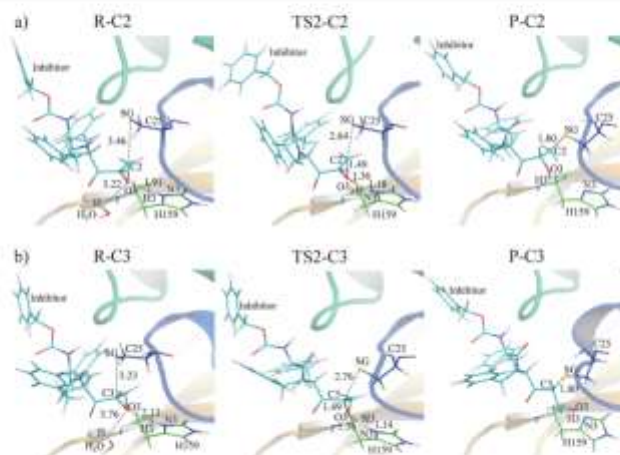


Fig. 5 Representative snapshots taken from the 2D-PMFs corresponding to the three states of the inhibition mechanism of cruzain by the attack of Cys25 on (a) C2; and (b) C3 of the inhibitor. Key distances are reported in Å.

1D-PMF, these values must be considered with caution. A description of the mechanisms where the protonation of the oxirane ring occurs after the attack of Cys25, as deduced from these 1D-PMFs, could be the result of a bad selection of the distinguished reaction coordinate. Consequently, in order to explore a possible concerted mechanism or the stepwise mechanism where the protonation of the ring precedes the attack of Cys25, depicted as mechanisms 2 and 1 in Scheme 1, 2D-PMFs at the M06-2X/6-31+G(d,p)/AM1d/MM level were generated using the SG-CX bond-forming distance as RC1 (where CX represents C2 or C3) and the O3-H3 bond-forming distance as RC2 (being H3 the proton from His159). The resulting FESs, shown in Fig. 4, describe the attack of Cys25 on either C2 or C3 atoms of the epoxy ring and the protonation of O3 as a concerted but asynchronous mechanism. The corresponding 2D-PMFs at the AM1d/MM level, shown in Fig. S4 of the ESI† render the same information of the process, at the qualitative level. The description of the reaction from a deeper analysis of Fig. 4 reveals important differences with respect the previous analysis based on Fig. 3. First, despite the trend in activation free energies of the attack of Cys25 on C2 and C3 atoms is maintained, the values deduced from the 2D-PMFs are significantly lower: 12.1 and 13.4 kcal mol<sup>-1</sup> for the attack on C2 and C3, respectively. These values clearly indicate that the irreversible attack of a cysteine on the inhibitor takes place preferably on the C2 atom of the epoxy ring. As commented above, the proton transfer from His159 is coupled to the nucleophilic attack of Cys25 and, consequently, coordinates associated with this proton transfer must be included in the distinguished reaction coordinate to not generate computationally biased results. From a thermodynamic point of view, the 2D-PMFs confirm the reaction to be a very exergonic process, but with much closer reaction free energies for the attack on C2 and C3 atoms than before: -56.4 and -54.9 kcal mol<sup>-1</sup> for the Cys25 attack on C2 and C3 atoms, respectively.

Representative snapshots of the three states involved in the reaction are presented in Fig. 5 and Fig. S5 (ESI†), while the average values of interatomic distances of the key states located along the reaction of both attacks of Cys25 on the C2 and C3 atoms are listed in Table 1. The data reported in Table 1, apart from confirming the concerted character of both processes, show important differences with respect to the conclusions derived from the analysis of 1D-PMFs. Thus, the position of TS-C2 is significantly shifted to a less advanced nucleophilic attack of Cys25 on C2 and a more advanced proton transfer from His159 to O3. Regarding TS-C3, the opposite trend is observed: it is significantly shifted to a more advanced attack of Cys25 on C3 and a less advanced proton transfer from His159 to O3. The consequence of these shifts is that TS-C2 and TS-C3 are more alike than those located on the FESs of Fig. 3. Regarding the interactions between the peptide and the residues of the active site measured in the structures generated along both reaction paths, it can be noticed how the interaction between the O2 atoms of the epoxide and Gly66 becomes more intense (closer interatomic O2-H<sub>66</sub> distance) as the reaction progresses in both reactions. It is also noticeable that the displacement of His159 occurs from the peptide once the proton has been transferred.

Table 1 Averaged AM1d/MM distances obtained for key states located along the inhibition mechanism of cruzain by the attack of Cys25 on the C2 atom (a), and on the C3 atom (b). All values are in Å and are obtained from 200 ps of AM1d/MM MD simulations on the stationary points extracted from the M06-2X/6-31+G(d,p) free energy profiles shown in Fig. 5. Some interatomic distances associated with the RCs were constrained during the simulations

(a) Cys25 attack on the C2 atom of the epoxy ring.			
d [Å]	RC2	TS2-C2	PC2
O2-C2	1.44 ± 0.03	1.49 ± 0.02	2.43 ± 0.05
O3-C3	1.44 ± 0.03	1.46 ± 0.03	1.42 ± 0.02
SG[Cys25]-C2	1.46 ± 0.03	2.04 ± 0.01	2.79 ± 0.03
SG[Cys25]-C3	1.41 ± 0.20	2.56 ± 0.11	2.79 ± 0.06
O3-H3(His159)	1.91 ± 0.02	1.36 ± 0.03	0.97 ± 0.02
N3-H3(His159)	1.01 ± 0.01	1.18 ± 0.01	5.10 ± 0.55
O3-N3(His159)	2.83 ± 0.07	2.70 ± 0.04	5.19 ± 0.52
O3-H <sub>66</sub> (water)	3.22 ± 0.44	4.16 ± 0.56	3.72 ± 0.64
O-H(Gly66)	1.63 ± 0.49	3.69 ± 0.86	3.62 ± 0.47
O2-H(Gly66)	3.40 ± 0.47	3.20 ± 0.43	2.10 ± 0.22
N1-H(Gly66)	3.22 ± 0.27	3.51 ± 0.29	3.35 ± 0.30
H1-O(Gly66)	2.16 ± 0.28	2.19 ± 0.25	2.17 ± 0.12
H2-O(Asp154)	4.26 ± 0.51	3.03 ± 0.32	3.28 ± 0.49
SG[Cys25]-H(Trp26)	1.49 ± 0.09	1.53 ± 0.10	2.92 ± 0.35
SG[Cys25]-H(Gly66)	2.76 ± 0.30	3.02 ± 0.23	5.29 ± 0.41
H4(His159)-O3(Asn175)	3.85 ± 0.69	4.63 ± 0.99	2.04 ± 0.23

(b) Cys25 attack on the C3 atom of the epoxy ring.			
d [Å]	RC3	TS2-C3	PC3
O2-C2	1.44 ± 0.03	1.45 ± 0.01	1.42 ± 0.02
O3-C3	1.45 ± 0.03	1.49 ± 0.02	2.38 ± 0.05
SG[Cys25]-C2	2.07 ± 0.12	2.84 ± 0.09	2.77 ± 0.06
SG[Cys25]-C3	3.23 ± 0.03	2.76 ± 0.03	1.80 ± 0.03
O3-H3(His159)	2.13 ± 0.01	1.36 ± 0.03	0.98 ± 0.03
N3-H3(His159)	1.03 ± 0.03	1.14 ± 0.02	6.81 ± 0.75
O3-N3(His159)	3.00 ± 0.21	2.47 ± 0.04	5.42 ± 0.45
O3-H <sub>66</sub> (water)	3.76 ± 0.46	4.10 ± 0.41	3.99 ± 1.21
O-H(Gly66)	3.18 ± 0.43	3.53 ± 0.41	3.56 ± 0.34
O2-H(Gly66)	3.57 ± 0.72	2.23 ± 0.20	3.12 ± 0.16
N1-H(Gly66)	3.13 ± 0.32	3.48 ± 0.27	3.06 ± 0.26
H1-O(Gly66)	2.38 ± 0.31	2.18 ± 0.23	1.98 ± 0.21
H2-O(Asp154)	4.46 ± 0.50	2.94 ± 0.50	3.57 ± 0.47
SG[Cys25]-H(Trp26)	1.48 ± 0.09	1.60 ± 0.11	3.31 ± 0.10
SG[Cys25]-H(Gly66)	3.48 ± 0.35	4.57 ± 0.28	5.87 ± 0.32
H4(His159)-O3(Asn175)	3.27 ± 0.56	2.27 ± 0.61	2.30 ± 0.70

The rest of the interaction between the epoxide and the protein remains almost invariant along the mechanism, or with no significant changes. Interestingly, the distance between Cys25 and Trp26 is dramatically elongated once the nucleophilic attack on C2 or C3 takes place. His159, once the H3 atom is transferred to the O3 atom of the epoxy, interacts with Asn175.

Finally, in order to confirm whether a water molecule can be responsible for the activation of the epoxide ring, instead of His159, 2D-PMFs at the M06-2X/6-31+G(d,p)/AM1d/MM level were generated using the SG-CX bond-forming distance as RC1 (CX represents C2 or C3) and the antisymmetric combination of the O<sub>2</sub>-H<sub>66</sub> bond-breaking and the O3-H<sub>66</sub> bond-forming distances as RC2, with H<sub>66</sub> and O<sub>66</sub> being the proton atom and the oxygen atom of the water molecule. The resulting M06-2X/6-31+G(d,p)/AM1d/MM FESs are shown in Fig. 6. The corresponding 2D-PMFs at the AM1d/MM level are deposited in the ESI† (Fig. S6). The FESs presented in Fig. 6 show how, despite the process is feasible, the reaction would take place through

PCCP

View Article Online

Paper

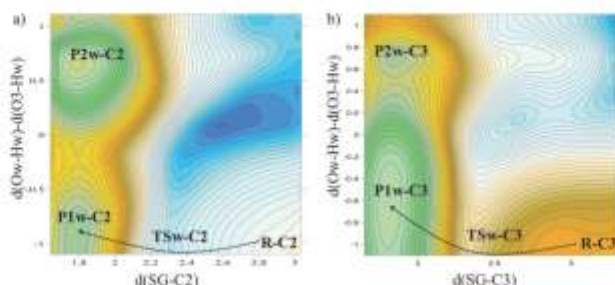


Fig. 6 FESs computed at the M06-2X/6-31+G(d,p)/AM12/MM level for the Cys25 attack on the (a) C2 atom and (b) C3 atom of the inhibitor, together with the transfer of a proton from a water molecule to the O3 atom of the ring. Distances are in Å and iso-energetic lines are displayed every 1.5 kcal mol<sup>-1</sup>.

higher activation energies than when His159 is the acid species: 42.1 and 29.9 kcal mol<sup>-1</sup> for the reaction where Cys25 attacks on C2 and C3, respectively. This result confirms our previous predictions based on the structural analysis of the reactant states. In fact, the FESs suggest that the water molecule would not activate the inhibitors but the bond between the Cys25 and the inhibitor would be formed concertedly with a spontaneous proton transfer from His159. P1w-C2 and P1w-C3 in Fig. 6a and b, respectively, correspond to the structures similar to P-C2 and P-C3 displayed in Fig. 5, where His159 has spontaneously transferred the proton to the O3 atom. As observed on both surfaces, although more dramatically in Fig. 6a, another local minimum is located when forcing the proton transfer from the water molecule to the O3 atom: P2w-C2 and P2w-C3 in Fig. 6a and b, respectively. These minima correspond to the structures where the proton from His159 that was spontaneously transferred to O3 is transferred back to the water. Thus, the FESs describe the reaction as the concerted mechanism labeled as 2 in Scheme 1. It is important to point out that the fact that different activation energies are obtained (by comparison with those in Fig. 4) is the consequence of not using the proper reaction coordinates to describe the proton transfer from His159, as occurred when comparing Fig. 3 and 4. Representative snapshots of the states involved in the exploration of these mechanisms are depicted in Fig. S7 and S8 of the ESI†

## Conclusions

A computational study of the inhibition mechanism of cysteine protease by dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph-[S] has been carried out by means of MD simulations with hybrid QM/MM potentials. The first conclusion derived from our study is that the activation of the epoxide ring and the attack of Cys25 on either C2 or C3 atoms take place in a concerted manner, in agreement with the predictions of Albeck *et al.*<sup>18,19</sup> and in contrast to the suggestions of Kim *et al.*<sup>12</sup> A mechanism where

the protonation of the oxirane ring takes place far behind the transition state, as suggested by Engels *et al.*,<sup>23,25–27</sup> is only observed when the transfer of the proton is not explicitly controlled during the simulations. According to our results, the acid species responsible for the protonation of the oxygen atom of the ring would be able to conserve His159, in contrast to previous studies that proposed a water molecule as the activating species.<sup>17,41</sup> Any attempt to use a water molecule of the active site as the source of the proton transfer to the oxirane ring renders significantly higher activation energies than when using His159. Thus, the activation free energy for the reaction where Cys25 attacks the C2 atom of the epoxide ring, activated by His159, is noticeably lower (12.1 kcal mol<sup>-1</sup>) than the one obtained by the attack on the C3 atom (15.4 kcal mol<sup>-1</sup>). These barriers are in turn lower than these that can be deduced (within the Transition State Theory at 298 K) from the  $k_{on}$  values measured for the inhibition of Papain and Cathepsin B with different peptidyl epoxides, which range between 20.0 and 22.2 kcal mol<sup>-1</sup>.<sup>14,20</sup>

This result is different from the one derived from our previous theoretical studies of the inhibition mechanism of cysteine protease Falcipain-2 (FP2) by a different inhibitor (the epoxysuccinate E64),<sup>20</sup> which gave similar barriers for the irreversible attack of Cys42 on both carbon atoms of the epoxy ring. This suggests that the mechanism of inhibition can depend not only on the inhibitor but on the particular cysteine protease.

The comparison of the free energy values for the inhibition mechanism of the cysteine protease cruzain studied in the present study, 12.1 kcal mol<sup>-1</sup>, with the barriers obtained in our previous study of the inhibition of the same protein by PHRs (29.5 and 10.0 kcal mol<sup>-1</sup> for PFK and PCIK, respectively)<sup>20</sup> allows a direct comparison of the inhibition reactivity toward cruzain of both families of inhibitors. In addition, it is important to highlight that the dipeptidyl-2,3-epoxyketone Cbz-Phe-Hph-[S] studied in the present paper presents the advantage of a large stabilization of the protein-inhibitor covalent complex (-56.4 kcal mol<sup>-1</sup>) by comparison with the PFK (-2.3 kcal mol<sup>-1</sup>) and the PCIK (-27.7 kcal mol<sup>-1</sup>).



In all, the results of the present study suggest that the tested epoxylone can be a good candidate to design new cruzain inhibitors with higher potency. The analysis of average structures of the key states involved in the inhibitory reaction obtained in the present study, together with those derived from our previous studies of cysteine protease inhibition,<sup>24–30</sup> can be used to design new inhibitors that kept the favorable interactions established in the covalent complex and stabilize the TS of the protein–inhibitor covalent complex.

## Acknowledgements

This work was supported by the Spanish Ministerio de Economía y Competitividad for project CTQ2013-66223-C2, Universitat Jaume I (project P1 1B2014-26) and Generalitat Valenciana (projects PROMETEOII/2014/022 and AICO/2016/932). V. M. is grateful to the University of Bath for the award of a David Parkin Visiting Professorship. K. A. thanks the Spanish Ministerio de Economía y Competitividad for a predoctoral contract. Authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I.

## References

- M. Marco and J. Miguel Coterón, *Curr. Top. Med. Chem.*, 2012, **12**, 408.
- J. H. McKerrow, J. C. Engel and C. R. Caffrey, *Bioorg. Med. Chem.*, 1999, **7**, 639.
- J. H. McKerrow, M. E. McGrath and J. C. Engel, *Parasitol. Today*, 1995, **11**, 279.
- A. R. Remiso and J. H. McKerrow, *Nat. Chem. Biol.*, 2006, **2**, 703.
- P. S. Sijwali, B. H. Shenai, J. Gut, A. Singh and P. J. Rosenthal, *Biochem. J.*, 2001, **360**, 481.
- World Health Organization, <http://www.who.int/mediacentre/factsheets/fs340/en/>, (accessed March 2017).
- J. C. Powers, J. L. Asgiao, O. D. Elici and K. E. James, *Chem. Rev.*, 2002, **102**, 4639.
- W. R. Roush, F. V. Gonzalez, J. H. McKerrow and E. Hansell, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 2809.
- S. Royo, S. Rodríguez, T. Schirmeister, J. Kesselring, M. Kaiser and F. V. Gonzalez, *ChemMedChem*, 2015, **10**, 1484.
- A. Latorre, T. Schirmeister, J. Kesselring, S. Jung, P. Johé, U. A. Hellmich, A. Heilos, B. Engels, R. L. Kramth-Siegel, N. Dirdjaja, L. Bou-Isaie, S. Rodríguez and F. V. Gonzalez, *ACS Med. Chem. Lett.*, 2016, **7**, 1073.
- K. Hanada, M. Tamai, S. Ohmura, J. Sawada, T. Seki and I. Tanaka, *Agric. Biol. Chem.*, 1978, **42**, 529.
- P. Rauder, H. Anglikler, B. Walker and E. Shaw, *Biochem. J.*, 1986, **239**, 633.
- W. R. Roush, A. A. Hernandez, J. H. McKerrow, P. M. Selzer, E. Hansell and J. C. Engel, *Tetrahedron*, 2000, **56**, 9747.
- R. Bibrowsky, J. C. Powers, C. M. Kam, R. Walton and R. C. Loewi, *Enzyme Inhib.*, 1993, **7**, 15.
- H. Helten, T. Schirmeister and B. Engels, *J. Org. Chem.*, 2005, **70**, 231.
- R. Bibrowsky, *J. Org. Chem.*, 1992, **57**, 1029.
- J. P. Meera and D. H. Rich, *J. Med. Chem.*, 1996, **39**, 3357.
- A. Albeck, S. Fluss and R. Pinsky, *J. Am. Chem. Soc.*, 1996, **118**, 3391.
- A. Albeck and S. Kliper, *Biochem. J.*, 1997, **322**, 879.
- A. Albeck, R. Pinsky and S. Kliper, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 1767.
- K. L. Varughese, F. R. Ahmed, P. E. Carry, S. Haszain, C. P. Huber and A. C. Storer, *Biochemistry*, 1989, **28**, 1330.
- D. H. Kim, Y. H. Jin and C. H. Ryu, *Bioorg. Med. Chem.*, 1997, **5**, 2103.
- H. Helten, T. Schirmeister and B. Engels, *J. Phys. Chem. A*, 2004, **108**, 7691.
- M. Mladenovic, K. Junold, R. F. Fink, W. Thiel, T. Schirmeister and B. Engels, *J. Phys. Chem. B*, 2008, **112**, 5458.
- R. Vicić, H. Helten, T. Schirmeister and B. Engels, *ChemMedChem*, 2006, **1**, 1021.
- M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700.
- J. L. Gan, P. Amara, C. Alhambra and M. J. Field, *J. Phys. Chem. A*, 1998, **102**, 4714.
- A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
- K. Arafe, S. Ferrer, S. Martí and V. Moliner, *Biochemistry*, 2014, **53**, 3336.
- K. Arafe, S. Ferrer and V. Moliner, *Biochemistry*, 2015, **54**, 3381.
- S. A. Gillmor, C. S. Craik and R. J. Fletcher, *Protein Sci.*, 1997, **6**, 1603.
- M. J. Field, M. Albe, C. Bret, F. Proust-De Martin and A. Thomas, *J. Comput. Chem.*, 2009, **21**, 1088.
- D. C. Bas, D. M. Rogers and J. H. Jensen, *Protein: Struct., Funct., Bioinf.*, 2008, **73**, 765.
- H. Li, A. D. Robertson and J. H. Jensen, *Protein: Struct., Funct., Bioinf.*, 2005, **61**, 704.
- M. H. M. Olsson, C. R. Sondergaard, M. Roszkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525.
- K. Nam, Q. Cui, J. Gao and D. M. York, *J. Chem. Theory Comput.*, 2007, **3**, 486.
- W. L. Jorgensen, D. S. Maxwell and J. Tiradollites, *J. Am. Chem. Soc.*, 1996, **118**, 11225.
- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718.
- G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187.
- S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011.
- J. J. Ruiz-Pernia, E. Silla, I. Turbón, S. Martí and V. Moliner, *J. Phys. Chem. B*, 2004, **108**, 8427.
- Y. Y. Chuang, J. C. Corchado and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 1140.
- J. C. Corchado, E. L. Coitinho, Y.-Y. Chuang, P. L. Fast and D. G. Truhlar, *J. Phys. Chem. A*, 1998, **102**, 2424.
- K. A. Nguyen, I. Rossi and D. G. Truhlar, *J. Chem. Phys.*, 1995, **103**, 5522.
- R. J. Renka, *SIAM Journal on Scientific and Statistical Computing*, 1987, **8**, 393.
- R. J. Renka, *ACM Trans. Math. Softw.*, 1993, **19**, 81.

[View Article Online](#)

PCCP

Paper

- 48 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 213.
- 49 W. J. Hehre, L. Radom, P. v. R. Schleyer and J. A. Pople, *Ab initio Molecular Orbital Theory*, New York, 1986.
- 50 B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384.
- 51 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. Montgomery, Jr., 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision A.1*, 2009.

### **7.3.1 Supporting Information**

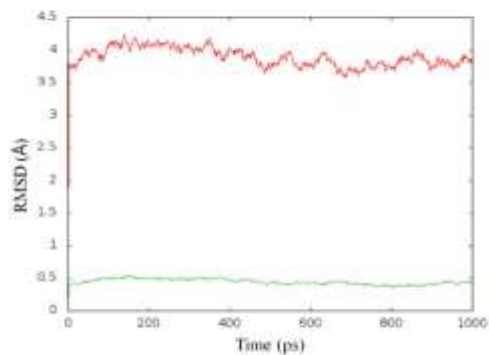
Supporting Information for the paper entitled: *Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Cysteine Protease Inhibition by Peptidyl-2,3-epoxyketones.*

Supporting Information

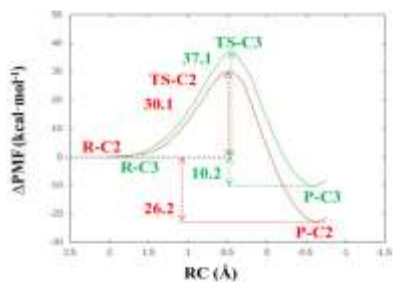
Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Cysteine  
Protease Inhibition by Peptidyl-2,3-epoxyketones.

*Kemel Arufet,<sup>1</sup> Silvia Ferrer,<sup>1</sup> Florenci V. González,<sup>2</sup> Vicent Moliner<sup>1,3\*</sup>*

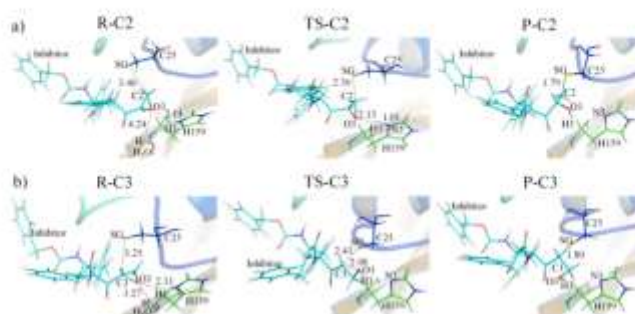
1. Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain
2. Departament de Química Inorgànica i Orgànica, Universitat Jaume I, 12071 Castelló, Spain
3. Department of Chemistry, University of Bath, Bath BA2 7AY, (United Kingdom)



**Figure S1.** Time evolution of the root-mean-square-deviation (RMSD) along the QM/MM MD simulation for the backbone atoms of the protein (green line) and atoms of inhibitor (red line). Simulations performed on the reactants state.



**Figure S2.** 1D-PMFs computed at AM1d/MM level. Cys25 attack to the C2 atom (red line) or C3 atom (green line). RC correspond to  $d(\text{SG-CX}) - d(\text{O3-CX})$ , with CX: C2, C3.



**Figure S3.** Representative snapshots taken from the 1D-PMFs corresponds to the three states of the inhibition mechanism of cruzain by the attack of Cys25 on (a) C2; and (b) C3 atoms of the inhibitor. The reported values of key interatomic distances are in Å.

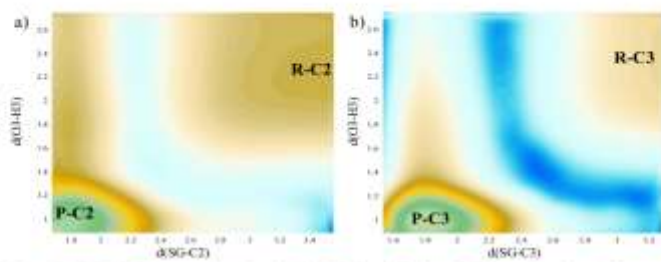
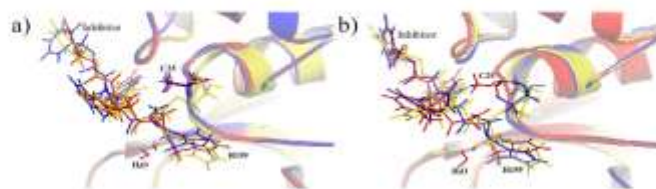
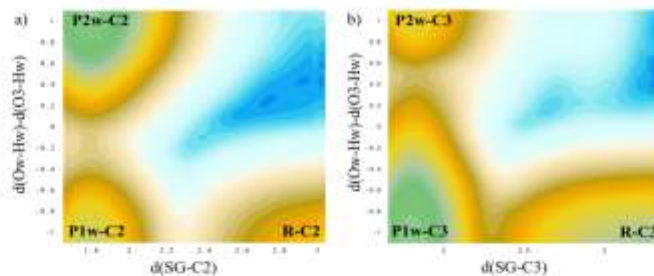


Figure S4. 2D-PMFs computed at the AM1d/MM level for the Cys25 attack to the a) C2 atom, and b) C3 atom, of the inhibitor. Distances are in Å and iso-energetic lines are displayed every 1.0 kcal·mol<sup>-1</sup>.

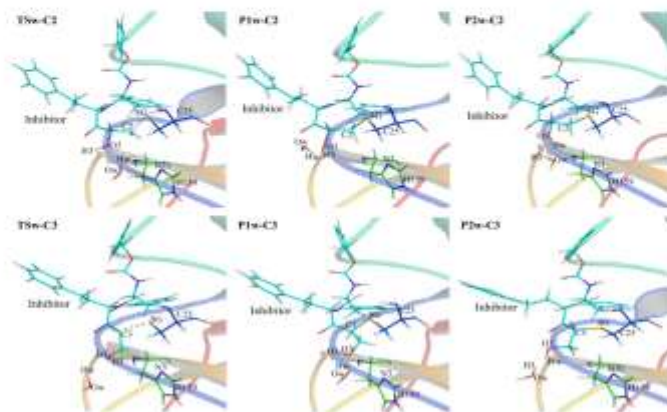


**Figure S5.** Superposition of representative snapshots presented in Figure 5 of the text corresponding to the three states of the inhibition mechanism of cruzain by the attack of Cys25 on (a) C2; and (b) C3 atoms of the inhibitor. Reactants are in red, TS in blue and Products in yellow.

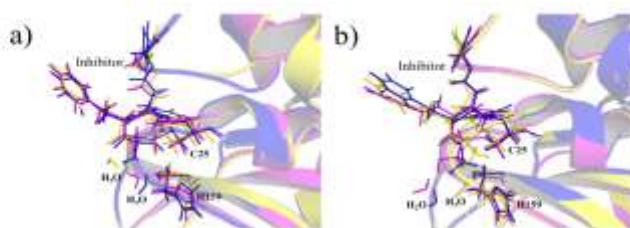




**Figure S6.** 2D-PMFs computed at the AM1d/MM level for the Cys25 attack to the a) C2 atom, and b) C3 atom, of the inhibitor, together with the transfer of a proton from a water molecule to the O3 atom of the ring. Distances are in Å and iso-energetic lines are displayed every 1.5 kcal·mol<sup>-1</sup>.



**Figure S7.** Representative snapshots of the states involved in the inhibition reaction when using a water molecule as the acid species activating the epoxide ring.



**Figure S8.** Superposition of representative snapshots presented in Figure S7 corresponding to the states of the inhibition mechanism of cruzain by the attack of Cys25 on (a) C2; and (b) C3 atoms of the inhibitor, assuming a water molecule as the acid species activating the epoxirane ring. TSw are in blue, P1w in yellow and P2w in magenta.

## 7. Papers

**Table S1.** Selected averaged distances obtained for the key states located along the inhibition mechanism of cruzain by attack of Cys25 on C2 and C3 atoms by 1D-PMFs. All values are in Å and are obtained from 200 ps of AM1d/MM MD simulations on the windows extracted from the M06-2X/MM free energy profiles.

<i>d</i> (Å)	Cys25 attack on C2			Cys25 attack on C3		
	R-C2	TS-C2	P-C2	R-C3	TS-C3	P-C3
O3-C2	1.44 ± 0.03	1.83 ± 0.06	2.42 ± 0.06	1.44 ± 0.03	1.41 ± 0.02	1.42 ± 0.03
O3-C3	1.44 ± 0.03	1.39 ± 0.03	1.42 ± 0.02	1.45 ± 0.03	2.08 ± 0.05	2.38 ± 0.05
SG(Cys25)-C2	3.40 ± 0.03	2.56 ± 0.03	1.79 ± 0.03	3.26 ± 0.31	3.05 ± 0.08	2.78 ± 0.06
SG(Cys25)-C3	4.36 ± 0.19	3.29 ± 0.11	2.79 ± 0.06	3.25 ± 0.03	2.43 ± 0.05	1.80 ± 0.03
O3-H3(His159)	2.14 ± 0.03	2.13 ± 0.03	0.97 ± 0.03	2.31 ± 0.03	0.98 ± 0.03	0.98 ± 0.03
N3-H3(His159)	1.00 ± 0.03	1.01 ± 0.03	4.81 ± 0.69	1.01 ± 0.03	5.77 ± 0.46	5.24 ± 0.44
O3-N3(His159)	2.99 ± 0.11	3.02 ± 0.09	5.05 ± 0.68	3.04 ± 0.20	5.69 ± 0.45	5.00 ± 0.53
O3-H <sub>w</sub> (water)	4.24 ± 0.70	2.33 ± 0.33	5.61 ± 0.81	3.27 ± 0.67	3.08 ± 0.79	4.08 ± 1.15

#### 7.4. Paper IV

Paper entitled: *Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease.*

Kemel Arafet, Silvia Ferrer, and Vicent Moliner

ACS Catalysis **2017**, 7, 1207-1215. DOI: 10.1021/acscatal.6b03096.

Reprinted with permission from (DOI: 10.1021/acscatal.6b03096).

Copyright (2017) American Chemical Society.

## Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease

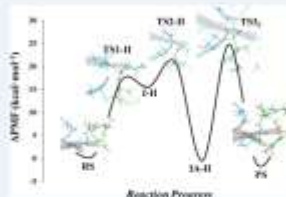
Kemel Arafet, Silvia Ferrer,<sup>†</sup> and Vicent Moliner<sup>\*,‡</sup>

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

Supporting Information

**ABSTRACT:** Cysteine proteases of the papain family are involved in many diseases, making them attractive targets for developing drugs. In this paper, different catalytic mechanisms of cruzain cysteine protease have been studied on the basis of molecular dynamics simulations within hybrid quantum mechanics/molecular mechanics potentials. The obtained free energy surfaces have allowed characterizing every single step along the mechanism. The results confirm that the full process can be divided into an acylation and a deacylation stage, but important differences with respect to previous studies can be deduced from our calculations. Thus, our calculations suggest that the acylation stage takes place in a stepwise mechanism where a proton from a conserved His159 is transferred first to the N1 atom of the peptide and, after a transient intermediate is located, the Cys25 attacks the carbonyl carbon atom. The stabilization of the activated Cys25 is achieved by an effect of the local environment through interactions with residues Trp26, Gly166, and His159, rather than by a less complex Cys25<sup>−</sup>/His159H<sup>+</sup> ion pair. In contrast, the deacylation stage, which was proposed to occur via a general base-catalyzed reaction whereby His159 activates a water molecule that attacks the peptide, would take place through a concerted mechanism. In this stage, the role of some residues of the active site, such as Gln19, Asn175, and Trp181, appears to also be crucial. Interestingly, the local environment of His159 would be modulating its pK<sub>a</sub> value to act as an acid in the acylation stage and as a base in the following deacylation stage.

**KEYWORDS:** cysteine proteases; catalytic mechanism; QM/MM; molecular dynamics; PMF



## INTRODUCTION

Cysteine proteases are an important class of enzymes involved in the hydrolysis of peptides and proteins. These enzymes are present in a variety of living organisms, from fungi to humans.<sup>1</sup> The cysteine proteases of the papain family are involved in many diseases, making them attractive targets for developing new drugs. For example, crizanant and falcipain-2 are drug targets against Chagas disease and malaria.<sup>2,3</sup>

The catalytic mechanism of the cysteine proteases depends on two residues that are located in the active site: Cys25 and His159 (cruzain numbering). It has been proposed that the imidazole group of His159 polarizes the SH group of the Cys25 and enables deprotonation of the SH group, thus forming a highly nucleophilic Cys25<sup>−</sup>/His159H<sup>+</sup> ion pair.<sup>4</sup> This ion pair explains the unusually high reactivity of these proteases toward electrophilic reagents in comparison with the nucleophilic power of the sulfur of cysteine or glutathione, especially in slightly acidic environments.<sup>5</sup> The existence of the ion pair was experimentally proven by different studies<sup>6–8</sup> and confirmed by computational tools.<sup>9–11</sup>

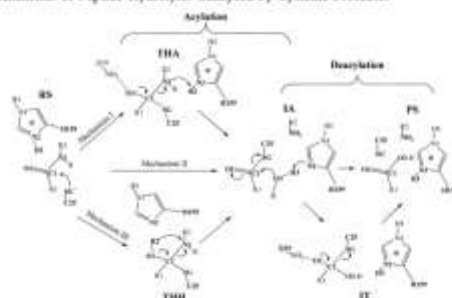
In 1976, Shafer and co-workers, in a potentiometric determination of ionization states at the active site of papain, concluded that the active site thiol group exists mainly as a thiol anion in the physiological pH range.<sup>7</sup> Hiller and co-workers, in a quantum mechanics/molecular mechanics (QM/MM) study

of the catalytic mechanism of the enzyme papain, showed that in the absence of the enzymatic environment the pair of neutral residues is strongly favored; however, with the presence of the enzyme, the potential energy surface (PES) for proton transfer was very flat, with the ion pair being more stabilized than the neutral species.<sup>12</sup> This was further supported by Suhai and co-workers that, in a QM/MM study of the active site of free papain and the NMA-papain complex, concluded that the ion pair becomes more stable than its neutral form in the protein environment.<sup>13</sup>

In general, it is accepted that the catalytic mechanism of cysteine protease consists of two stages: acylation and deacylation (see Scheme 1). The acylation stage basically consists of the nucleophilic attack on the carbonyl carbon of the peptide bond (C1) by the thiolate anion of Cys25 producing an intermediate (IA in Scheme 1). Then, the deacylation occurs via a general base-catalyzed reaction, whereby the imidazole nitrogen of His159 (N2) activates a water molecule that attacks the C1 atom of the peptide to produce the final products (PS in Scheme 1).

Received: October 30, 2016  
Revised: December 21, 2016  
Published: December 27, 2016

Scheme 1. Reaction Mechanisms of Peptide Hydrolysis Catalyzed by Cysteine Proteases



Several experimental<sup>17–19</sup> and computational<sup>20–23,25–28</sup> studies have been carried out in order to explain the manner in which the cysteine proteases catalyze the breaking of peptide bonds. Different mechanisms have been proposed in the acylation stage. Thus, Drenth and co-workers, on the basis of X-ray diffraction studies, proposed that this stage occurs via a serine protease like mechanism (mechanism I in Scheme 1),<sup>17,18</sup> the thiolate anion attacks the C1 atom to produce a thiohemiketal intermediate, denoted THA. This intermediate would be stabilized by hydrogen bond interactions with the backbone NH group of Cys25 and to the NH<sub>2</sub> group of the Gln19 side chain that create an oxyanion hole similar to that proposed in serine proteases. Then a proton is transferred from His159 to the nitrogen N1 of the peptide, concomitant with the C1–N1 peptide bond cleavage, to form IA, as proposed by Lowe in 1970.<sup>15</sup> However, the stability of the THA intermediate has been questioned by several authors.<sup>8,11,29–31,33,39</sup> We recently carried out QM/MM studies of the inhibition mechanism of cruzain by pepstatin halomethyl ketones, and we concluded that the THA is not a stable intermediate.<sup>30</sup> On the other hand, studies based on molecular mechanics and semiempirical QM methods with reduced models of the papain-catalyzed reaction carried out by Kollman and co-workers suggested that the protonation of O3 (or N1 atoms) must occur prior to (see mechanism III in Scheme 1) or concurrently with (see mechanism II in Scheme 1) the nucleophilic attack of the thiolate anion.<sup>20,21</sup> Hiller and co-workers found that this stage is concerted (mechanism II), rather than stepwise, by exploring QM/MM potential energy pathways with *N*-methylacetamide as substrate.<sup>21</sup> Kolandaivel and co-workers explored the PES of the inhibition of haloketones<sup>30</sup> and diketones<sup>30</sup> in the gas phase at the DFT level, concluding that the THH intermediate would be stable but, in general, less stable than the reactant.

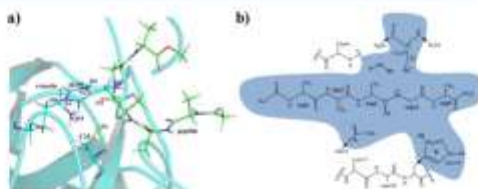
After acylation, the deacylation occurs when a water molecule attacks the acyl–enzyme complex (IA intermediate) at the carbonyl carbon C1 to produce PS. The formation of the ion pair also makes the thiol a good leaving group during the deacylation due to its low *pK<sub>a</sub>* value.<sup>1</sup> Welch and co-workers, in a discussion of the catalytic pathway of cysteine proteases on the basis of AM1 calculations, showed that the protonation of His159 by a water molecule and the formation of the C1–O3

bond (formation of PS, as depicted in Scheme 1) occur simultaneously with an activation barrier of 39.0 kcal mol<sup>-1</sup>.<sup>24</sup> Years later, Gao and co-workers, in a study of the catalytic pathway of a human cathepsin K, on the basis of QM/MM MD simulations, reported that the proton transfer from the water molecule to the imidazole ring of the catalytic histidine (His162 in the human cathepsin K sequence) was fully concerted with the nucleophilic attack from O3 atom to the carbonyl carbon C1 and presented a free energy of activation of 16.7 kcal mol<sup>-1</sup>.<sup>22</sup> However, Warshel and co-workers, in an ab initio study of several reference reactions in solution (the histidine-assisted thiomethanolysis of formamide and the hydrolysis of the resulting thioester), had concluded that the deacylation has stepwise character through an intermediate denoted IT (see Scheme 1) that would be formed prior to the cleavage of the SG–C1 bond, IT formation being the rate-limiting step.<sup>23</sup> More recently, Zhan and co-workers<sup>31</sup> in a pseudobond first-principles QM/MM FE study of the reaction pathway for papain-catalyzed hydrolysis of *N*-acetyl-Phe-Gly-4-nitroimidate arrived at the same conclusions as Warshel and co-workers. Thus, to date, many aspects around the catalytic mechanism of cysteine proteases are unclear and are under discussion. How does the acylation stage really occur? Is the deacylation concerted, or is it a stepwise process?

The principal aim of this paper is to gain insights into the catalytic mechanism of cysteine proteases on the basis of MD simulations within hybrid QM/MM potentials. We study the catalytic mechanism of the crainin cysteine protease by exploring the free energy surfaces (FES) of the different proposed mechanisms. Analysis of energetics, averaged geometries, and interactions between the peptide and the protein will allow improving our understanding of this important enzyme, which could be used for a rational design of new inhibitors with potential application in the treatment of severe human diseases.

## COMPUTATIONAL METHODS

The starting structure for the construction of our enzyme-peptide model was the X-ray crystal structure of cruzain from *Trypanosoma cruzi* bound to Bi-Tyr-Ala-CH<sub>2</sub>F with PDB entry 1AEM<sup>32</sup> and 2.0 Å resolution. This particular X-ray structure was already used in our laboratory to study the inhibition



**Figure 1.** (a) Representative snapshot of the active site in the reactants state (RS) of the studied model. (b) Details of the active site corresponding to the standard model. The blue region corresponds to the QM atoms. The link atoms between the QM and MM regions are indicated as black dots.

mechanism of cruzain by peptidyl halomethyl ketones.<sup>30</sup> The initial geometry of the cruzain peptide was constructed by replacing the covalent inhibitor Bz-Tyr-Ala-CH<sub>2</sub>F by a small peptide: Ac-Ala-Ala-Ala-Gly-Ala-OCH<sub>3</sub>. The structure of the peptide was manually modified until a favorable (reactive) conformation was reached using the PyMol program.<sup>43</sup> Hence, our model consists of one chain of 215 amino acids (cruzain) and the peptide Ac-Ala-Ala-Ala-Gly-Ala-OCH<sub>3</sub> (see Figure 1a).

Cruzain is located in the intracellular vesicles during the dormant stage of the parasite at a slightly acidic pH. During the intracellular amastigote stage of the *T. cruzi* life cycle, cruzain is present on the surface of the parasite, where the pH of the host cytoplasm is almost neutral (pH 7.4).<sup>30</sup> Thus, hydrogen atoms were added to the X-ray structure at this physiological pH using the IDYNAMO library,<sup>44</sup> according to the pK<sub>a</sub> values of the titratable residues calculated within the empirical PROPKA 3.1 program of Jensen et al.<sup>45</sup> A total of 11 counterions (Na<sup>+</sup>) were placed into optimal electrostatic positions around the system, in order to obtain electro-neutrality. Finally, the systems were placed in a 79.5 Å side cubic box of water molecules.

As depicted in Figure 1b, the full peptide, Cys26, the imidazole ring of His159, the side chain of Gln19, and one water molecule (88 atoms) were described by means of the AM1d semiempirical Hamiltonian.<sup>37</sup> The rest of the protein and the water molecules were described by OPLS-AA<sup>46</sup> and TIP3P<sup>47</sup> force fields, respectively. To saturate the valence of the QM/MM frontier, we used the link atoms procedure.<sup>36</sup> Because of the size of the full system, all residues further than 25 Å from the CA1 atom of the peptide were kept frozen during the simulations (43115 atoms from a total of 50029). A force switch function with a cutoff distance in the range of 14.5–18 Å was applied to treat the nonbonding interactions, and periodic boundary conditions were employed. All the QM/MM calculations were carried out using the IDYNAMO library.<sup>44</sup>

The system was relaxed by means of 10 ns of QM/MM MD at 300 K using the NVT ensemble and the Langevin–Verlet integrator. Analysis of the time evolution of the root-mean-square deviation of the backbone atoms of the protein and the peptide confirms that the system was equilibrated. The structure obtained after relaxation was used to generate the initial hybrid AM1d/MM PESs. Stationary point structures (including reactants, products, intermediates, and transition state structures) were located and characterized, guided by means of a micromacro iterations scheme.<sup>37</sup>

Once the PESs corresponding to every chemical step were explored, potentials of mean force (PMFs) were generated as a function of different distinguished reaction coordinates (RCs) using the weighted histogram analysis method combined with

the umbrella sampling approach.<sup>38,39</sup> In general, the value of the force constant used for the harmonic umbrella sampling to generate the PMFs was 2300 kJ mol<sup>-1</sup> Å<sup>-2</sup> and the simulation windows consisted of 20 ps of equilibration and 40 ps of production, with a time step of 1 fs. Subsequently, 200 ps of AM1d/MM MD simulations of the stationary points were performed to analyze the main geometrical parameters as an average.

For the study of mechanism I of the acylation stage (see Scheme 1), a monodimensional PMF (1D-PMF) AM1d/MM was obtained using the bond-forming distance,  $d(\text{SG}-\text{C}1)$ , as RC. This required a series of 41 simulation windows. For the acylation stage through mechanism II, a two-dimensional PMF (2D-PMF) AM1d/MM was computed using as RC1 the antisymmetric combination of the bond-forming and bond-breaking distances,  $d(\text{SG}-\text{C}1)-d(\text{N}1-\text{C}1)$ , and as RC2 the antisymmetric combination of the bond-forming and bond-breaking distances,  $d(\text{N}1-\text{H}2)-d(\text{N}2-\text{H}2)$ . This required a series of 4941 simulation windows. For the acylation stage through mechanism III, a 2D-PMF AM1d/MM was computed for the formation of the THH intermediate from the reactant state using the SG–C1 bond-forming distance as RC1 and the O1–H2 bond-forming distance as RC2. This required a series of 980 simulation windows. For the formation of the IA intermediate from the THH intermediate, a 1D-PMF AM1d/MM was obtained using the antisymmetric combination of the bond-forming and bond-breaking distances,  $d(\text{N}1-\text{H}2)-d(\text{O}1-\text{H}2)$ . This required a series of 65 simulation windows.

For the study of the deacylation stage (see Scheme 1), a 2D-PMF AM1d/MM was computed for the formation of the PS state from the IA intermediate state using the SG–C1 bond-forming distance as RC1 and the O3–H3 bond-breaking distance as RC2. This required a series of 1330 simulation windows. Within these coordinates, both possible mechanisms were explored.

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 correct the low-level AM1d energy function used in the 1D- and 2D-PMFs, an interpolated correction scheme developed in our laboratory has been applied.<sup>37</sup> This scheme, based on a method proposed by Tumbler and co-workers for dynamic calculations of gas-phase chemical reactions,<sup>48–51</sup> uses a spline under tension<sup>52–55</sup> to interpolate the correction term at any value of the RCs selected to generate the PESs. Thus, the new energy function for a 1D-PMF is defined as

$$E = E_{\text{AM1d}} + \Delta E_{\text{IT}}^{\text{IT}}(\zeta) \quad (1)$$

where  $S$  denotes the spline function, whose argument  $\Delta R_{ij}^{\text{HL}}(\zeta)$  is a correction term taken as the difference between single point calculations of the QM subsystem using a high-level (HL) method, and a low-level (LL) method (in our case AM1d). When corrections were done on 2D-PMF, the correction term was then expressed as a function of the two coordinates  $\zeta_1$  and  $\zeta_2$  to generate the higher level 2D-PMF. The HL calculations were carried out by means of the M06-2X functional<sup>58</sup> with the 6-31+G(d,p) basis set<sup>57</sup> following the suggestions of Truhlar and co-workers.<sup>54,56</sup> These calculations were carried out using the Gaussian09 program.<sup>59</sup>

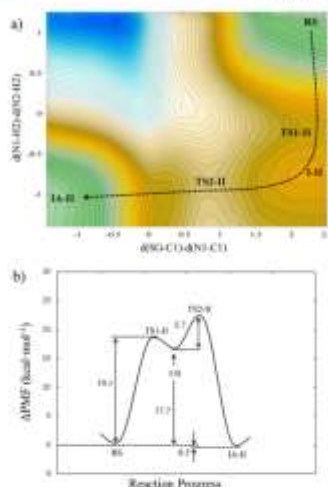
## RESULTS AND DISCUSSION

The catalytic mechanism of the cysteine protease cruzain has been initially studied by exploring the PESs of all the possible chemical steps corresponding in the two stages of the mechanism: acylation and deacylation. As explained in Computational Methods, the PESs were then used as the starting point to generate FESs.

The first avenue for the acylation stage is through the formation of the thiohemiacetal intermediates denoted THA (see mechanism I in Scheme 1); the exploration of this step by AM1d/MM PESs shows that THA is not a stable intermediate. The free energy profiles in terms of 1D-PMFs, computed at the AM1d/MM level as well as after correction at the M06-2X/6-31+G(d,p):AM1d/MM level (see Figure S1a,b in the Supporting Information), confirm the instability of the THA. After the DFT corrections, the THA is 48 kcal mol<sup>-1</sup> higher in energy than the reactant state and does not appear as a stable intermediate. Although this value can be slightly overestimated and can be reduced if the interactions of the anionic THA are improved by increasing the size of the QM region, dramatic changes are not expected to occur, considering that QM/MM MD simulations are carried out and, consequently, proper distribution of molecules around the charges developed in the QM subsystem during the reaction step are sampled. These results suggest that the acylation stage cannot take place through the formation of THA, which is in correspondence with several computational studies.<sup>60,61,62,63,64</sup>

According to the stationary points located on the PESs generated to explore the acylation stage through mechanism II, it appears that it does not take place in a single concerted step but in a stepwise manner. First the proton from His159 is transferred to the N1 atom of the peptide leading to an intermediate (I-II), and then the Cys25 attacks the C1 atom of the peptide, forming the products of the acylation stage: the IA-II intermediate. The FES, in terms of 2D-PMF at the M06-2X/6-31+G(d,p):AM1d/MM level, is presented in Figure 2a. The corresponding 2D-PMF at AM1d/MM level is shown in Figure S2 in the Supporting Information. According to the results, the protonation of the N1 atom of the peptide, through TS1-II, proceeds with a free energy barrier of 19.1 kcal mol<sup>-1</sup> (see Figure 2b). Then, the attack of Cys25 at the C1 atom of the peptide takes place with a low free energy barrier (5.7 kcal mol<sup>-1</sup>) defined by TS2-II. Representative snapshots of some key states are presented in Figure S3 in the Supporting Information.

Finally, the FESs generated to explore the third possible mechanism for the acylation stage, mechanism III in Scheme 1, are shown in Figure 3a,b (the corresponding FESs at the AM1d/MM level are shown in Figure S4 in the Supporting Information). The complete free energy profiles for mechanism III are summarized in Figure 3c, and representative snapshots

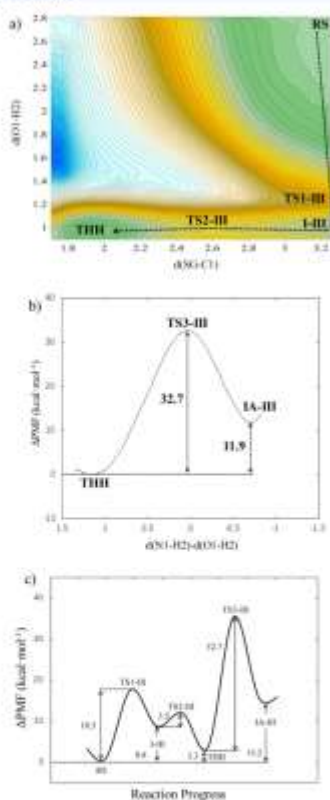


**Figure 2.** (a) FESs of the acylation stage through mechanism II computed at the M06-2X/6-31+G(d,p):AM1d/MM level. Distances are in Å, and isomeric lines are displayed every 0.8 kcal mol<sup>-1</sup>. (b) Corresponding free energy profile for the acylation stage through mechanism II.

of the key states are presented in Figure S5 in the Supporting Information. It is important to point out that names assigned to each state reflect just the order of intermediates or TSs appearing along the explored mechanism. Thus, for instance, I-III is much more stable than I-II but just because they correspond to significant different intermediates (see Figures S3 and S5 in the Supporting Information). According to the results, this mechanism, initially proposed to take place in two steps, is in fact characterized by three steps. First the proton from His159 is transferred to the O1 atom of the peptide with a free energy barrier of 18.5 kcal mol<sup>-1</sup> defined by TS1-III, leading to the intermediate I-III (see the corresponding snapshots in Figure S5). Afterward, the attack of Cys25 on the C1 atom of the peptide proceeds with a low free energy barrier defined by TS2-III, as occurred in mechanism II, leading to the stable intermediate I-IIIH. The last step, the cleavage of the C1–N1 peptide bond, leads to the formation of IA-III with a high free energy barrier of 32.7 kcal mol<sup>-1</sup>. The rate-limiting step of the acylation stage through mechanism III would consequently correspond to the decomposition of I-IIIH, with an effective free energy barrier, computed from RS, of 36.0 kcal mol<sup>-1</sup>.

A comparison of the free energy profiles of the acylation stage through mechanisms II and III (see Figures 2b and 3c) provides an important conclusion. Considering the effective free energy barriers, computed from RS, the difference between the barriers of both mechanisms is significant: 23.4 and 36.0





**Figure 3.** (a) FESs of the acylation stage through mechanism III computed at the M06-2X/6-31+G(d,p)/AM1d/MM level; formation of the THH intermediate from RS (asymmetric lines are displayed every 0.5 kcal mol<sup>-1</sup>). (b) Free energy profile of the formation of IA intermediate from THH. (c) Overall free energy profile for the acylation stage. All distances are in Å.

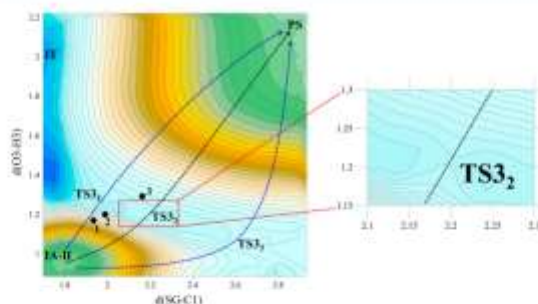
kcal mol<sup>-1</sup> for mechanisms II and III, respectively. Clearly, the acylation stage through mechanism II is kinetically more favorable than that through mechanism III. Moreover, mechanism II has a slightly exergonic character (−0.5 kcal mol<sup>-1</sup>), unlike mechanism III, which has an endergonic character (15.2 kcal mol<sup>-1</sup>); the intermediate IA-II is 15.7 kcal mol<sup>-1</sup> more stable than the intermediate IA-III. Obviously,

both intermediates correspond to the same state but are reached from different paths and, consequently, the obtained difference in energy must be attributed to different conformational spaces explored in the generation of the PMFs. An analysis of the geometries reveals that the main difference between the two averaged structures is the orientation of the imidazole ring of His159 (see Figure S6 in the Supporting Information); the imidazole ring of His159 in IA-II is oriented to interact with the peptide ( $d[\text{N2}-\text{H2}] = 2.13$  Å), while in IA-III it is oriented toward residues Ser176 and Asp175.

Once the intermediate IA is formed, a water molecule attacks the carbonyl carbon C1 of the peptide taking place at the deacylation stage. The FES computed at the M06-2X/6-31+G(d,p)/AM1d/MM level is presented in Figure 4, while the corresponding surface at the AM1d/MM level is shown in Figure S7 in the Supporting Information. According to our results, the deacylation stage is identified as a concerted mechanism defined by TS3, at which the water molecule deprotonates the proton to His159 and attacks the carbonyl carbon C1 simultaneously. As mentioned in the Introduction, several authors<sup>17,23</sup> proposed that the deacylation has stepwise character through the formation of an intermediate denoted IT (see Scheme 1). Nevertheless, this IT intermediate, which would be located in the upper left corner of the FES of Figure 4, has not been found to be a stable intermediate, being in fact 32.2 kcal mol<sup>-1</sup> less stable than the IA-II intermediate.

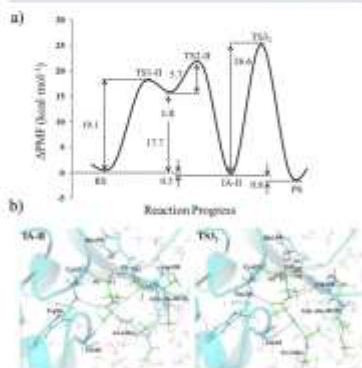
As observed in Figure 4, different reaction paths could be traced from IA-II to P5 since the quadratic region of the TS dividing surface is associated with saddle points of order 1 with a wide range of values for the two selected reaction coordinates, all of them with small energetic differences. Thus, while the black dashed line reaction path traced on the figure would describe a quite synchronous bond formation and breaking process, other paths (see blue dashed lines) could be traced describing a more asynchronous-like process with similar free energy barriers. In any event, none of these paths are close to a possible stepwise mechanism through an IT intermediate-like structure. In fact, the mechanism with the most asynchronous character, which could be closer to a stepwise process, takes place through the opposite corner of the surface (with the SG–C1 bond breaking in a very advanced stage of the process and the forming O3–H3 bond in a early stage of the process). In order to check the reliability of the description of the mechanism derived from the corrected FES presented in Figure 4, structures of saddle points have been reoptimized at the M06-2X/MM level to test whether the geometries obtained at the LL/MM level were representative of the HL/MM geometries. The position of the resulting TS structures optimized at the M06-2X/MM level from TS3, TS2, and TS1, structures selected from the FES, displayed as black circles 1–3 in Figure 4, confirm the step as a concerted reaction. As observed, all three reoptimized TS structures lie in the quadratic region of the corrected FES, thus supporting the computational correction technique. IRC paths were traced down from the three M06-2X/MM TSs and further confirm that the three of them connect IA-II with P5. Cartesian coordinates and key interatomic distances of the three TSs are reported in Tables S1 and S2 in the Supporting Information. According to these results, it appears that the path through TS3, was the most favorable one. Thus, our results are far from supporting a stepwise deacylation reaction.

The complete free energy profile for the most favorable mechanism of the peptide hydrolysis catalyzed by cruzain



**Figure 4.** FES of the deacylation stage computed at the M06-2X/6-31+G(d,p):AM1-d/MM level. These competitive reaction energy paths are traced as dashed lines. Black dots labeled as 1–3 correspond to the position of TSs optimized at the M06-2X/6-31+G(d,p)/MM level of structures selected from TS<sub>1</sub>, TS<sub>2</sub>, and TS<sub>3</sub>, respectively. Distances are given in Å, and isosurface lines are displayed every 1.0 kcal mol<sup>-1</sup>.

cysteine protease, derived from our M06-2X/6-31+G(d,p):AM1-d/MM results, are summarized in Figure 5a, while the mechanism is schematically summarized in Scheme 2.



**Figure 5.** (a) Free energy profile for the reaction mechanism of peptide hydrolysis catalyzed by cysteine proteases computed at the M06-2X/6-31+G(d,p):AM1-d/MM level. (b) Representative snapshots of 1A-II and TS<sub>3</sub>. Distances are given in Å.

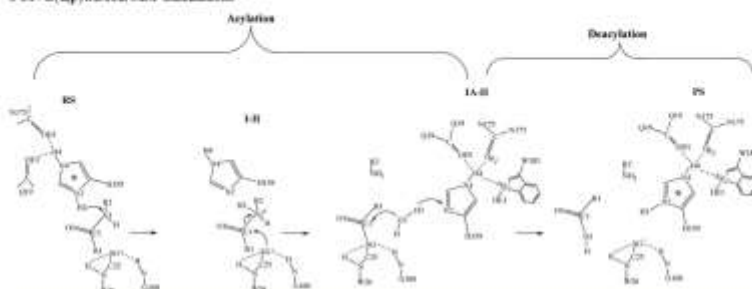
The rate-limiting step would correspond to TS<sub>3</sub>, with an activation free energy barrier of 26.6 kcal mol<sup>-1</sup>. It is important to point out that the predicted activation free energy for the rate-limiting step appears to be slightly overestimated with respect to other cysteine proteases such as papain. For instance, free energy barriers ranging from 16.9 to 17.9 kcal mol<sup>-1</sup> can be deduced within the transition state theory (TST) from the rate constants of papain-catalyzed hydrolysis of some *p*-nitrophenyl

esters and *p*-nitroindoles at pH 6.0 and 35 °C, reported by Low et al. for a series of substrates.<sup>19</sup> Nevertheless, the comparison must be made with caution because, apart from not being the same cysteine protease, the activity of cysteine proteases depends on the specific substrate. In any event, further improvements can be done in the computational methodology in the future that could slightly modify the results at a quantitative level.

Average values of interatomic distances of the key states located along the proposed catalytic mechanism are given in Table 1. The first conclusion that can be drawn from Figure 5a and Table 1 is that the high free energy barrier corresponding to the first step of the acylation is associated with the fact that TS1-II is closer to the intermediate 1-II than to 8S, as can be confirmed by the average values of the key distances N1–H2 (1.23 Å) and N2–H2 (1.48 Å). In the intermediate 1-II, both N1–H2 and N1–C1 bonds are weak, according to the interatomic distances of 1.09 and 1.49 Å, respectively. These values are just slightly larger than their corresponding standard bond distances. From 1-II to 1A-II, the N1–H2 bond becomes stronger and the N1–C1 bond is breaking at the same time as Cys25 attack at the C1 atom of the peptide. The values given in Table 1 confirm that the deacylation stage is concerted, the O3–C1 and N2–H3 bonds are forming concomitantly with the SG–C1 and O3–H3 breaking bonds, in TS<sub>3</sub>. Regarding the interactions between the peptide and the protein, the analysis of the interactions along the reaction profile reveals, as mentioned above, the flexibility present in the full system. There are some interactions between the active site and the residues of the peptide that are strengthened while others are weakened along the reaction mechanism (see Table S3 in the Supporting Information). It appears that only the interaction O(Asp158)–H(Ala3,peptide) remains invariant in all key states of the mechanism. In fact, an analysis of interaction energies by amino acid reveals that this is a favorable interaction to stabilize all states but, in particular, the TSs of the first and last steps of the reaction (see Figures S8 and S9 in the Supporting Information).

Regarding the interactions that are established with the two key residues of the active site, Cys25 and His159, analysis of the

**Scheme 2.** Reaction Mechanism of Peptide Hydrolysis Catalyzed by Cruzin Cysteine Protease As Deduced from our M06-2X/6-31+G(d,p):AM1d/MM Calculations



**Table 1.** Average Distances for Key States Located along the Mechanism of Peptide Hydrolysis Catalyzed by Cysteine Proteases, Obtained from 200 ps of AM1d/MM MD Simulations on the Stationary Points Extracted from the M06-2X/MM Free Energy Profiles\*

	<i>d</i> (Å)							
	RS	TS1-II	II-II	TS2-II	IA-II	TS3-II	PS	
N2(H14)–H2(H19)	1.00 ± 0.03	1.49 ± 0.03	1.88 ± 0.03	1.05 ± 0.03	7.90 ± 1.08	6.18 ± 1.24	5.41 ± 0.72	
N1(G4 <sub>pp</sub> )–H2(H19)	2.13 ± 0.03	1.13 ± 0.03	1.09 ± 0.03	1.06 ± 0.03	1.00 ± 0.03	1.00 ± 0.03	1.00 ± 0.03	
SG(C25)–C1(A <sub>pp</sub> )	3.17 ± 0.03	3.56 ± 0.03	3.14 ± 0.03	1.41 ± 0.03	1.82 ± 0.02	2.20 ± 0.01	1.77 ± 0.03	
N1(G4 <sub>pp</sub> )–C1(A <sub>pp</sub> )	1.41 ± 0.03	1.45 ± 0.02	1.49 ± 0.02	1.54 ± 0.02	1.77 ± 0.06	5.35 ± 1.20	4.34 ± 0.65	
O1(H21)–H2(H19)	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03	1.22 ± 0.03	2.09 ± 0.03	
O1(H21)–C1(A <sub>pp</sub> )	4.76 ± 0.03	4.36 ± 0.50	5.72 ± 1.11	5.58 ± 0.64	6.07 ± 0.77	1.43 ± 0.02	1.27 ± 0.02	
N2(H18)–H2(H19)	4.04 ± 0.44	4.57 ± 0.44	4.72 ± 0.37	4.62 ± 0.58	4.42 ± 0.90	1.30 ± 0.03	1.01 ± 0.03	
SG(C25)–H(G60)	1.53 ± 0.11	1.40 ± 0.09	1.31 ± 0.10	2.24 ± 0.39	2.72 ± 0.30	3.03 ± 0.23	2.74 ± 0.29	
SG(C25)–H(G160)	2.80 ± 0.28	3.16 ± 0.33	3.69 ± 0.84	2.78 ± 0.30	3.65 ± 0.57	2.53 ± 0.22	3.32 ± 0.17	
SG(C25)–H2(H19)	1.07 ± 0.39	3.16 ± 0.36	3.61 ± 0.21	3.59 ± 0.11	3.48 ± 0.33	7.05 ± 1.16	5.38 ± 1.02	
SG(C25)–H2(H19)	4.12 ± 0.97	5.54 ± 0.60	6.55 ± 0.89	6.67 ± 0.73	3.46 ± 0.73	2.80 ± 0.14	1.67 ± 0.34	
H4(H18)–O1(H21)	2.53 ± 0.30	2.58 ± 0.32	4.90 ± 1.97	3.80 ± 0.30	3.41 ± 0.33	3.94 ± 0.41	3.59 ± 0.36	
H4(H18)–O1(H17)	2.22 ± 0.33	3.46 ± 0.39	5.07 ± 1.44	3.34 ± 0.44	4.11 ± 0.44	2.94 ± 0.42	3.51 ± 0.31	
H4(H18)–N1(W181)	4.34 ± 0.53	4.76 ± 0.64	6.25 ± 0.99	3.84 ± 0.66	2.88 ± 0.38	3.73 ± 0.51	2.69 ± 0.38	

\*Some steric distances associated to the RCs were constrained during the simulations.

key interatomic distances reported in Table 1 renders interesting conclusions. The activated (negatively charged) Cys25 in RS is stabilized by hydrogen bond interactions with Trp26 ( $1.53 \pm 0.11$  Å) and, to a lesser extent, with Gly160 ( $2.80 \pm 0.28$  Å) and His159 ( $3.07 \pm 0.39$  Å) through the H2 atom. In fact, the strong interaction with Trp26 is maintained until the TS, where Cys25 attacks the C1 atom of the peptide (TS2-II). Once the Cys25 bond with the C1 atom is broken in the deacylation stage, the negatively charged sulfur atom of Cys25 is stabilized by interactions with Trp26 ( $2.74 \pm 0.29$  Å), with Gly160 ( $2.32 \pm 0.17$  Å), and with the hydrogen atom now bonded back to His159, the H3 atom ( $2.67 \pm 0.34$  Å). This analysis of distances would suggest that the originally suggested ion pair formed between His159 and Cys25 appears to be more complex, since the activation of Cys25 is carried out by additional contributions of Trp26 and Gly160.

The geometrical analysis can be supplemented by the calculation of the interaction energies between the QM and the MM regions decomposed by amino acid (see Figures S8 and S9 in the Supporting Information). The results confirm the role

of some residues such as Asp158, Gly160, and Trp181 that stabilize the TSs with respect to their corresponding preceding stable stationary state.

Finally, another interesting conclusion can be derived from the analysis of the interactions established with the N4–H4 group of His159, deduced from the distances reported in Table 1. Thus, by inspection of distances between the H4 atom of His159 and residues such as Gln19 (which is treated quantum mechanically in all of the simulations), Asn175, and Trp181, it seems that the local environment is capable of modulating the  $pK_a$  value of this titratable residue. The shortest values of the average distances to the H4 atom in the reactants are detected with Gln19 ( $2.55 \pm 0.30$  Å) and Asn179 ( $2.22 \pm 0.33$  Å), while in the IA intermediate H4 interacts with Gln19 ( $2.61 \pm 0.35$  Å) and Trp181 ( $2.68 \pm 0.38$  Å). We must keep in mind that His159 must act as an acid in the acylation stage but as a base in the deacylation stage. In this regard, the  $pK_a$  value of His159 has been computed in RS, IA, and PS by means of the PROPKA program of Jensen and co-workers<sup>42</sup> and the results show how this residue presents the lowest  $pK_a$  value in the IA

intermediate (4.36), in comparison with the values in RS (3.95) and PS (3.90).

### CONCLUSIONS

Our results confirm that the full process can be divided into an acylation and a deacylation stage, as previously described in the literature. Nevertheless, important new conclusions can be deduced from our QM/MM study.

First of all, the acylation stage, consisting of the nucleophilic attack on a carbonyl carbon of the peptide bond by the formation of an activated Cys25, takes place by a stepwise mechanism where a proton from a conserved His159 is transferred first to the N1 atom of the peptide and, after the transient intermediate I-II is located, the Cys25 attacks the carbonyl carbon atom. In this stage, the stabilization of the activated Cys25 is achieved by an effect of the local environment through interactions with residues Trp26, Gly160, and His159, rather than by a less complex Cys25S<sup>-</sup>/His159H<sup>+</sup> ion pair, as previously suggested in the literature.

The second main conclusion is that the deacylation stage, which was proposed to occur via a general base-catalyzed reaction whereby His159 activates a water molecule that attacks the peptide to produce the final products, would take place through a concerted mechanism. In this stage, the role of some residues of the active site, such as Gln19, Asn175, and Trp181, also appear to be crucial. Interestingly, it seems that residues Gln19, Asn175, and Trp181 would be modulating the pK<sub>a</sub> value of the catalytic His159, which would act as an acid in the acylation stage and as a base in the following deacylation stage. This has been confirmed by computing the pK<sub>a</sub> value of His159 along the reaction, rendering the highest pK<sub>a</sub> value of His159 in the intermediate located between both stages, IA.

According to our results, the most favorable free energy profile, shown in Figure 5a, indicates that, despite the fact that the activation free energies can be slightly overestimated by comparison with experimental rate constants of related reactions in other cysteine proteases, the process would be kinetically controlled by the deacylation stage that would take place in a single step. The quadratic TS region is defined by a wide variety of values of the selected reaction coordinate, ranging from an almost perfectly synchronous process to a quite asynchronous situation where the covalent bond between the Cys25 and the peptide would be in a very advanced stage of the process. This result could be interesting for the future design of inhibitors based on TS analogues (TSA<sub>ts</sub>). An understanding of this important enzyme at an atomistic level could be used for the rational design of new inhibitors with potential applications in the treatment of severe human diseases.

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.catal.8b03096.

Free energy surfaces corresponding to every single step of the reaction, computed as 1D PMF and 2D PMF, at the AM1d/MM level, 1D-PMFs for the acylation stage through mechanism I computed at the M06-2X/6-31+G(d,p):AM1d/MM level, representative snapshots of key states of the reaction, Cartesian coordinates of the QM atoms of the three TSs of the deacylation step computed at the M06-2X/6-31+G(d,p)/MM level,

selected interatomic distances for the key states located along the reaction mechanism, and average interaction energies by residues for key states for the reaction mechanism (PDF)

### AUTHOR INFORMATION

#### Corresponding Authors

\*S.F.: e-mail, sferrer@uji.es

\*V.M.: e-mail, molinerv@uji.es; tel. +34964728084.

#### ORCID

Vicent Moliner: 0000-0002-3665-1310

#### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

This work was supported by the Spanish Ministerio de Economía y Competitividad (project CTQ2015-66223-C2-1-P), Universitat Jaume I (project P1-1B2014-26), and Generalitat Valenciana (PROMETEOII/2014/022). K.A. thanks the Spanish Ministerio de Economía y Competitividad for a predoctoral contract. The authors acknowledge computational resources from the Servei d'Informàtica de Universitat Jaume I.

### REFERENCES

- (1) Otsu, H. H.; Schrammer, T. *Chem. Rev.* **1997**, *97*, 333.
- (2) McKerrow, J. H.; Engel, J. C.; Caffrey, C. R. *Bioorg. Med. Chem.* **1999**, *7*, 639.
- (3) Rendo, A. B.; McKerrow, J. H. *Nat. Chem. Biol.* **2006**, *2*, 701.
- (4) Kelller, J. W.; Brown, K. S. *J. Am. Chem. Soc.* **1992**, *114*, 7983.
- (5) Bruckelhart, K. *Int. J. Biochem.* **1978**, *10*, 259.
- (6) Compton, D. J.; Genotouras, M. S.; Heslop, J. M. *FEBS Lett.* **1980**, *110*, 319.
- (7) Lewis, S. D.; Johnson, F. A.; Shafir, J. A. *Biochemistry* **1976**, *15*, 3009.
- (8) Lewis, S. D.; Johnson, F. A.; Shafir, J. A. *Biochemistry* **1981**, *20*, 48.
- (9) Han, W. G.; Tajbakhsh, E.; Sahai, S. *J. Biomed. Struct. Dyn.* **1999**, *16*, 1019.
- (10) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12385.
- (11) Wei, D. H.; Huang, X. Q.; Liu, J. J.; Tang, M. S.; Zhan, C. G. *Biochemistry* **2013**, *52*, 5145.
- (12) Angelides, K. J.; Park, A. I. *Biochemistry* **1979**, *18*, 2355.
- (13) Angelides, K. J.; Park, A. I. *Biochemistry* **1979**, *18*, 2365.
- (14) Dentith, J.; Kalk, K. H.; Swen, H. M. *Biochemistry* **1976**, *15*, 3731.
- (15) Dentith, J.; Swen, H. M.; Hogerstraten, W.; Sjöström, L. A. A. *Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen Serie C-Biological and Medical Sciences* **1975**, *78*, 104.
- (16) Herthield, B.; Schmitz, G. L. *J. Am. Chem. Soc.* **1972**, *94*, 1263.
- (17) Kelller, J. W.; Nesome, A. A.; Brown, K. S. *J. Am. Chem. Soc.* **1994**, *116*, 4669.
- (18) Lowe, G.; Yuthavong, Y. *Biochim. J.* **1971**, *126*, 107.
- (19) Anal, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491.
- (20) Bryan, K.; Gao, J. L. *Mol. Graphics Modell.* **2000**, *18*, 50.
- (21) Howard, A. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 7105.
- (22) Ma, S.; Devi-Kosovan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2007**, *129*, 13033.
- (23) Stražič, M.; Florin, J.; Wastel, A. *J. Phys. Chem. B* **2001**, *105*, 4471.
- (24) Welsh, W. J.; Liu, Y. *J. Mol. Struct.: THEOCHEM* **1997**, *401*, 315.
- (25) Lowe, G. *Philo. Trans. R. Soc. B* **1970**, *237*, 237.

- (26) Araki, K.; Ferrer, S.; Moliner, V. *Biochemistry* **2015**, *54*, 3381.
- (27) Vijayakumar, S.; Kolandaivel, P. *Int. J. Quantum Chem.* **2008**, *108*, 927.
- (28) Shankar, B.; Kolandaivel, P.; Senthilkumar, K. *Int. J. Quantum Chem.* **2009**, *109*, 1660.
- (29) Gillens, S. A.; Crak, C. S.; Flatterick, R. *J. Protein Sci.* **1997**, *6*, 1623.
- (30) [www.pymol.org](http://www.pymol.org), 2009.
- (31) Field, M. J.; Albe, M.; Bret, C.; Proust-de Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1098.
- (32) Olesen, M. H. M.; Sandergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525.
- (33) Nam, K.; Choi, Q.; Gan, J.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486.
- (34) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (35) Jorgensen, W. L.; Chandrosskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (36) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 708.
- (37) Martí, S.; Moliner, V.; Tuñón, I. *J. Chem. Theory Comput.* **2005**, *1*, 1008.
- (38) Kumar, S.; Bezuid, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (39) Torro, G. M.; Valhan, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
- (40) Ruiz-Perrós, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2004**, *108*, 8427.
- (41) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140.
- (42) Corchado, J. C.; Gattuso, E. L.; Chuang, Y. Y.; Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 2424.
- (43) Nguyen, S. A.; Bassi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.
- (44) Benka, R. *J. SIAM Journal on Scientific and Statistical Computing* **1987**, *8*, 393.
- (45) Benka, R. *J. ACM Trans. Math. Softw.* **1993**, *19*, 81.
- (46) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, *120*, 235.
- (47) Helm, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.
- (48) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.
- (49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalapino, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H. C.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09; Gaussian, Inc.: Wallingford, CT, 2009.

### **7.4.1. Supporting Information**

Supporting Information for the paper entitled: *Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease.*

**Supporting Information**

**A Computational Study of the Catalytic Mechanism of the Cruzaïn Cysteine  
Protease**

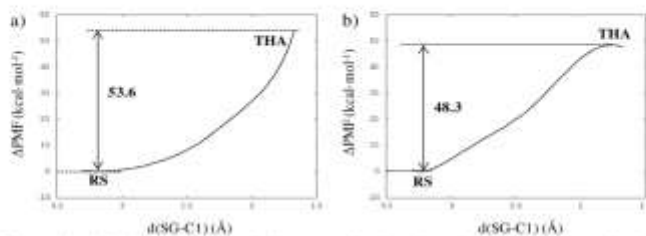
*Kemel Arafet, Sílvia Ferrer,\* Vicent Moliner\**

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

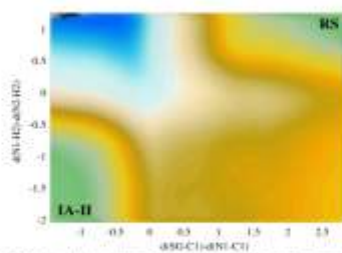
Corresponding authors

S. Ferrer: [sferrer@uji.es](mailto:sferrer@uji.es)

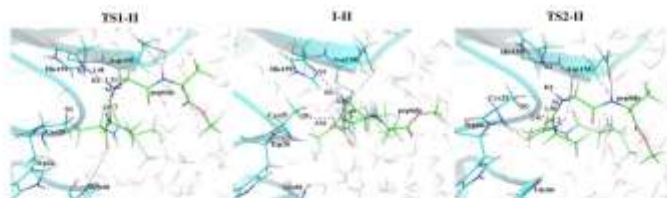
V. Moliner: [moliner@uji.es](mailto:moliner@uji.es); tel: +34964728084



**Figure S1.** 1D-PMFs for the acylation stage through the mechanism I computed at the AM1d/MM level (a) and computed at the M06-2X/6-31+G(d,p):AM1d/MM level (b).

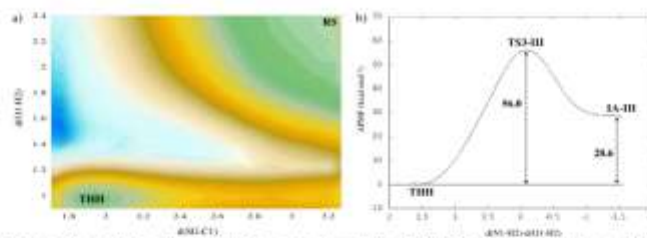


**Figure S2.** 2D-PMF for the acylation stage through a mechanism II computed at the AM1d/MM level. Distances are in Å and iso-energetic lines are displayed every 0.6 kcal mol<sup>-1</sup>.

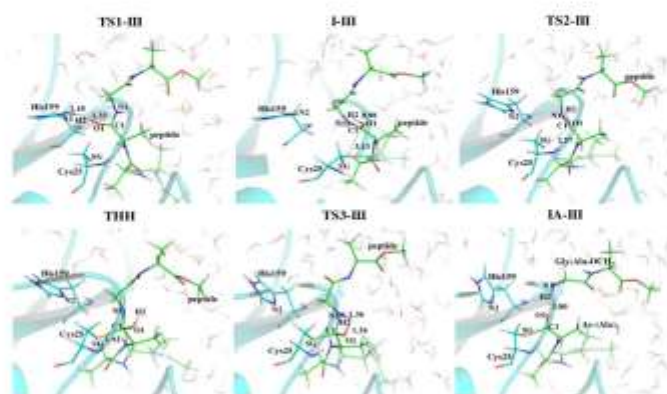


**Figure S3.** Representative snapshots of some key states for the acylation stage through a mechanism II. Distances are in Å.





**Figure S4.** Acylation stage through a mechanism III. a) 2D-PMFs for the formation of THH intermediate from R5 computed at the AM1d/MM level. Iso-energetic lines are displayed every 1.2 kcal·mol<sup>-1</sup>. b) 1D-PMFs for the formation of IA intermediate from THH computed at the AM1d/MM level. All distances are in Å.



**Figure S5.** Representative snapshots of some key states for the acylation stage through a mechanism III. Distances are in Å.



H	-8.562973	16.394413	-0.079181	-6.472830	16.213971	-6.055126	-0.575003	11.792080	0.041961
S	8.801897	9.171888	-0.344883	1.174724	8.819327	-8.207381	1.561264	8.202007	0.031093
H	8.809097	8.878489	1.181168	2.016868	8.487749	1.277716	1.223763	8.380301	4.878973
C	8.942904	8.603556	0.085194	2.121584	8.375965	0.688984	2.543625	8.944836	10.877988
H	1.898774	7.961790	0.149307	2.681913	7.470111	0.189918	2.691516	8.687994	0.688993
C	-8.795484	9.300247	13.173934	-1.258640	8.894836	16.114596	0.048068	16.158421	14.191255
H	-8.261448	9.307373	14.369982	-1.648650	8.382757	13.864646	-0.044909	11.286779	23.897847
H	-1.047533	8.529066	13.862993	-8.784432	8.255587	18.923248	-0.284361	8.997368	15.231611
H	-8.808834	16.247822	0.010713	-2.806048	8.137616	18.557913	-0.010103	8.714257	13.901761
C	8.903921	8.809254	22.819823	-0.266949	7.805320	13.588784	1.107714	8.308250	13.336649
G	8.801153	8.821819	23.778676	0.418398	8.024889	16.403648	2.038907	8.471401	18.893198
S	8.909186	8.161859	21.800679	-0.081766	7.848882	12.716973	1.074292	8.601768	12.018669
H	8.108208	8.968159	0.877427	0.489178	8.568915	11.978674	1.774115	8.157864	11.677101
C	-0.094948	8.578919	0.384270	-0.817350	8.232015	11.204383	0.067748	10.221815	11.216596
H	-1.018352	9.463280	10.899852	-1.867055	8.410112	11.471742	-0.011371	10.230198	11.381056
C	8.161467	18.982443	0.862564	-0.870232	8.258496	0.082236	0.444412	11.774148	19.910314
H	-8.601217	13.247823	9.135660	0.951769	9.353957	10.663425	-0.129188	12.254679	10.200807
H	1.171922	11.016138	9.362583	-0.063071	10.189784	11.288669	1.822467	11.789784	10.428614
H	8.117226	11.839288	10.803705	-0.801503	10.119911	10.173558	0.108974	12.243411	11.847488
C	-8.801179	8.341984	0.118184	-8.771722	7.416716	8.930003	0.018048	8.738051	9.918593
G	8.579999	7.861325	8.946838	0.184261	8.825200	0.731000	0.028726	8.786484	9.571572
M	-1.503845	8.530867	8.664542	-1.699780	7.752812	8.585181	-1.187196	8.707143	9.159846
H	-1.888560	9.139403	8.779307	-2.848888	8.198297	8.273973	-1.801003	10.258617	9.535647
C	8.621212	7.710223	2.286419	11.380188	7.268811	5.851186	1.212173	8.843761	7.914888
C	-1.744986	7.850383	8.786389	-0.443872	7.312350	7.335141	-0.126643	8.934436	7.538483
C	-8.410111	8.801903	7.688179	-2.023135	8.583690	1.425534	-1.898604	7.614127	8.193117
H	-2.818398	8.197386	7.854813	-1.087380	8.806399	7.871743	3.993148	7.776403	8.489741
H	8.801280	8.858585	8.131154	-1.406881	8.728586	1.878486	1.472863	7.904189	9.973736
H	-1.550746	-5.772352	8.501924	-1.890146	8.486757	8.954484	-1.861344	-8.066825	7.258458
C	-2.383124	8.479114	6.548887	-2.158149	8.218996	6.782720	-2.024388	8.918685	6.918077
G	-1.429456	8.070510	6.212028	-1.551648	8.206653	6.784812	-3.217973	10.022281	7.085780
S	8.017122	9.507181	7.148411	-1.451221	8.011788	8.021249	1.186668	16.388384	5.556761
H	-8.625324	9.851401	6.019417	-8.666446	8.284311	8.638158	-0.381453	10.371431	10.009889
C	-1.866398	18.198680	4.873807	-1.810342	8.968679	3.028589	1.888663	11.888081	4.888674
H	-1.501925	9.327631	8.789160	-1.418680	8.895431	4.959035	-2.822875	10.973031	4.612289
C	-2.710413	11.844886	8.298389	-2.448180	11.346192	5.751118	-2.028182	12.668817	2.710771
H	-7.715666	11.811148	8.434549	-2.955211	12.082164	11.739883	-2.573789	12.454842	6.036689
H	-1.266881	12.445489	8.569333	-2.867140	11.236150	6.738985	-1.665330	13.106487	5.972542
H	1.251216	11.884321	6.280144	-1.429879	11.688847	1.892889	2.605190	13.129777	5.104388
C	-2.087711	16.173189	3.986629	-1.899302	16.118827	1.621488	1.018166	13.973031	1.684494
G	-2.678941	10.848337	3.451278	-1.717877	10.868687	8.912513	1.488108	11.584789	2.851487
S	-8.889929	13.448789	1.918184	-8.617841	17.173778	2.582286	-0.103111	19.381934	2.073611
H	-1.472369	14.189188	1.509152	-8.581439	17.788185	2.480182	0.475531	14.851570	3.337889
C	-2.940628	14.313897	2.916405	-8.281777	17.957515	4.471289	-1.479736	18.431225	2.135865
H	-2.111622	13.442442	7.588524	-8.892117	12.284213	3.324271	-2.188448	19.892591	2.362574
H	-3.188912	15.121181	2.821289	0.067542	14.488388	3.298686	1.823558	14.481136	5.718193
C	-1.739428	15.662759	0.018803	-1.861288	14.953183	4.866979	-1.705870	16.258603	4.388688
G	-2.999116	14.817611	8.181793	-1.382846	13.088271	4.074015	2.882988	16.781889	4.387411
S	-8.801888	15.799188	8.918888	-1.354889	14.868888	8.679884	16.858887	16.118887	3.718818
H	8.025188	15.888874	4.588788	-8.250723	15.847888	8.477887	0.188823	19.881847	5.088888
C	-1.474581	16.821644	8.998614	-2.018264	16.758218	8.388818	-0.011472	17.888187	6.361372
H	-2.561744	16.879279	5.764257	-2.488314	17.149884	3.488834	-1.428889	18.149881	6.059775
C	-8.972619	18.167887	3.827388	-1.292118	17.888658	8.891181	8.959144	17.763681	6.218381
H	-2.267156	18.881812	6.881348	-1.812366	18.721286	1.717137	0.018818	18.821789	7.685986
H	1.088786	18.478883	8.888888	8.888888	18.127888	8.148881	1.888811	16.788886	6.018288
H	8.211388	13.878815	3.978821	-8.991147	13.891872	7.848579	1.188971	16.988889	7.148811
C	-1.212177	14.988889	3.288887	-1.212181	16.281844	2.782021	2.818877	16.888838	7.487181
G	-1.028890	14.748781	7.717887	-1.078888	15.573888	8.888888	-1.183880	15.918181	7.948882
G	-1.432845	16.743829	8.248117	-1.594384	17.068888	1.978111	-2.598818	17.284811	7.032818
C	-1.311082	16.217188	9.574446	-2.272541	16.888286	8.888884	5.364443	16.788127	0.008511
H	-1.488816	11.078188	8.212888	-1.888888	16.068888	7.121188	-8.888125	15.888125	8.668818
H	8.968887	15.348174	9.731881	-1.888888	17.781288	8.881888	-8.118888	17.981288	9.111188
H	-1.278989	15.888888	9.731811	-8.871273	16.788989	8.687843	2.757187	16.868687	9.888888
G	-8.907174	11.882039	2.488478	-8.414718	10.821279	3.788888	0.681223	11.788448	9.907288
H	8.578849	16.688272	2.568462	11.285779	10.231899	2.678891	1.848888	13.771888	2.558467
H	-0.972614	18.858889	8.528202	-8.364511	11.822181	1.611882	0.799912	11.488181	4.788147
H	1.380586	9.261817	-2.541258	3.337388	8.221188	2.738882	2.888112	8.802987	2.888014
H	-1.258885	9.518889	1.002882	-8.381812	8.931885	2.999184	2.735577	8.268884	2.168541
H	-1.762115	8.268818	1.974813	-8.382843	8.884826	1.938812	-8.988653	9.732887	1.367812
H	-8.071133	18.388413	-1.388888	-1.788888	8.878497	4.532115	-8.934118	16.798881	1.488888

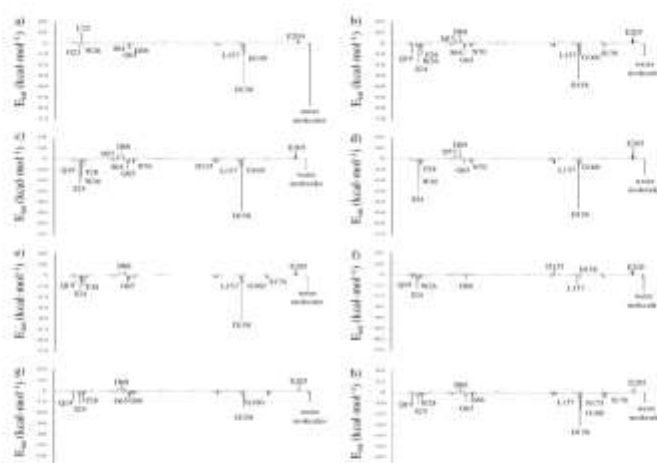
**Table S2.** Interatomic distances for the TS3s of the deacylation step of the mechanism of peptide hydrolysis catalyzed by cysteine proteases computed at M06-2X/6-31+G(d,p)/MM level from each TS computed at M06-2X/6-1+G(d,p):AM1d/MM level.

d (Å)	TS3 <sub>1</sub>	TS3 <sub>2</sub>	TS3 <sub>3</sub>
N2(H159)-H2(H159)	3.86	4.20	5.61
N1(G4, pep)-H2(H159)	1.02	1.03	1.03
SG(C25)-C1(A3, pep)	1.93	2.00	2.23
N1(G4, pep)-C1(A3, pep)	3.67	3.31	3.98
O3(H <sub>2</sub> O)-H3(H <sub>2</sub> O)	1.18	1.20	1.27
O3(H <sub>2</sub> O)-C1(A3, pep)	1.69	1.53	1.54
N2(H159)-H3(H <sub>2</sub> O)	1.30	1.26	1.27
SG(C25)-H(W26)	3.32	3.09	1.34
SG(C25)-H(G160)	2.23	2.44	2.75
SG(C25)-H2(H159)	4.42	4.65	6.65
SG(C25)-H3(H <sub>2</sub> O)	2.94	2.75	2.60
H4(H159)-OE1(Q19)	2.95	3.91	3.17
H4(H159)-OG(S176)	1.75	1.72	2.30
H4(H159)-OD1(N175)	3.37	3.70	3.46
H4(H159)-NE1(W177)	3.43	3.13	3.23
O(A2, pep)-H(G66)	2.46	2.47	2.54
O(A2, pep)-H(W26)	3.44	4.42	4.86
O(L157)-H(A1, pep)	2.18	4.75	2.47
OD158)-H(A3, pep)	1.88	1.79	1.66
OD2(D158)-H(A5, pep)	4.47	3.75	4.46
O(G66)-H(A2, pep)	3.59	2.53	3.90

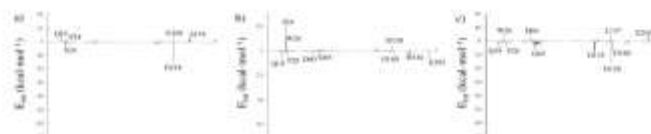
**Table S3.** Averaged distances between the active site and the residues for key states located along the mechanism of peptide hydrolysis catalyzed by cysteine proteases, obtained from 200 ps of AM1d/MM MD simulations on the stationary points extracted from the M06-2X/MM free energy profiles.

d (Å)	RS	TS1-R	I-R	TS2-R	IA-R	TS3	PI
O(A2, pep)-H(G66)	2.34±0.17	2.27±0.22	3.76±0.76	3.71±0.48	3.04±0.62	3.08±0.91	2.77±0.44
O(A2, pep)-H(W26)	3.47±0.46	3.75±0.75	4.42±0.76	2.39±0.44	2.14±0.51	1.77±0.27	1.91±0.28
O(L157)-H(A1, pep)	2.74±0.48	3.25±0.77	3.05±0.47	2.53±0.38	3.39±0.90	3.61±0.63	2.63±0.44
OD158)-H(A3, pep)	1.89±0.19	1.99±0.23	1.97±0.20	2.08±0.22	2.16±0.27	2.08±0.17	2.08±0.19
OD158)-H(G4, pep)	2.00±0.17	1.85±0.14	2.29±0.29	1.89±0.15	0.35±1.13	2.25±0.91	6.78±0.62
OD1(D158)-H(G4, pep)	3.71±0.30	3.91±0.41	2.43±0.49	3.96±0.28	4.35±1.07	3.22±0.71	3.33±0.49
OD1(D158)-H(A5, pep)	3.96±0.48	2.07±0.35	2.80±1.00	2.85±0.34	4.90±0.90	5.93±1.05	2.44±0.47
OD2(D158)-H(A5, pep)	2.83±0.55	2.26±0.35	3.01±0.88	2.37±0.38	5.34±0.51	3.02±1.02	3.06±0.36

S6



**Figure S8.** Averaged interaction energies by residues for key states for the reaction mechanism of peptide hydrolysis catalyzed by cysteine proteases, obtained from 200 ps of AM1d/MM MD simulations on the stationary points extracted from the M06-2X/MM free energy profiles. a) RS-A, the interaction energies were calculated taking into account the interaction between the cruzain and the peptide. For the rest of the key state was calculated taking into account the interaction between the cruzain and the QM region that participating in the reaction (Cys25, His159, H<sub>2</sub>O, and the peptide). b) RS. c) TS1-II. d) I-II. e) TS2-II. f) IA-II. g) TS3<sub>2</sub>. h) PS.



**Figure S9.** The difference in the averaged interaction energies by residue between a) TS1-II and RS, b) TS2-II and I-II, and c) TS3<sub>2</sub> and IA-II. The average interaction energies were obtained from 1 ns of AM1d/MM MD simulations on the stationary points extracted from the M06-2X:AM1d/MM free energy profiles.



## 8. Conclusions and Outlook

### 8.1 Conclusions

In correspondence with the proposed objectives, three different inhibition mechanisms have been studied in the present Doctoral Thesis corresponding to the inhibition mechanism of cruzain by PEK and PHK, and to the inhibition mechanism of FP2 by E64. In addition, the reaction mechanism of the peptide scission catalyzed by cruzain has been also studied. The main conclusions drawn from this research can be summarized as follows:

- **Study of the Mechanism of FP2 Inhibition by the Epoxysuccinate E64:** the results obtained during this study have allowed us to obtain a complete picture of the possible reaction paths corresponding to the irreversible attack of Cys42 on the carbon atoms of the epoxy ring of the E64 inhibitor. According to our results, the attack of Cys42 on C2 would take place in a single step, while the attack on C3 is described as a two-step mechanism, the first step corresponding to formation of the sulfur-carbon bond, which is the rate-limiting step. In the second step of this mechanism, a hydrogen atom from the carboxylic group of the inhibitor is finally transferred to the oxygen atom of the opened epoxy ring. From a kinetic point of view, the irreversible attack of Cys42 on E64 can take place on both carbon atoms of the epoxy ring because both processes present similar barriers. Nevertheless, the inhibitor-protein complex derived from the attack on C3 appears to be much more stabilized. A thorough analysis of the E64-FP2 interactions reveals the importance of some of the residues, such as Gln171, Asp170, Gln36,

Trp43, Asn81, and His174, in anchoring the inhibitor in a proper orientation for the reaction to take place.

- **Study of the Mechanism of Cruzain Inhibition by PHK:** according to this study, the nucleophilic attack of the unprotonated Cys25 on the C atom from the halogen-C bond of the inhibitor (Bz-Tyr-Ala-CH<sub>2</sub>F or Bz-Tyr-Ala-CH<sub>2</sub>Cl) and the cleavage of the halogen-C bond take place in a concerted way. Analysis of free energy barriers and reaction free energies, computed at both levels of theory, shows that the inhibition by Bz-Tyr-Ala-CH<sub>2</sub>Cl would be kinetically and thermodynamically more favorable. Then, our results suggest that inhibitor Bz-Tyr-Ala-CH<sub>2</sub>Cl can be used as a starting point to develop a proper inhibitor of cruzain.

- **Study of the Inhibition Mechanism of Cruzain by the epoxyketone Cbz-Phe-Hph-(S):** The first conclusion derived from our study is that the activation of the epoxide ring and the attack of Cys25 to either C2 or C3 atoms take place in a concerted manner. According to our results, the acid species responsible of the protonation of the oxygen atom of the ring would be the conserved His159; any attempt to use a water molecule of the active site as the source of the proton transfer to the oxyrane ring render significantly higher activation energies than when using the His159. From a kinetic point of view, the irreversible attack of a cysteine to the inhibitor takes place preferably on the C2 atom of the epoxy ring.

The comparison of the free energy values for the inhibition mechanism of cruzain by the epoxyketone Cbz-Phe-Hph-(S), 12.1 kcal·mol<sup>-1</sup>, with the barrier obtained in the inhibition of the same protein by Bz-Tyr-Ala-CH<sub>2</sub>Cl, 10.0 kcal·mol<sup>-1</sup>, allows a direct comparison of the inhibition reactivity toward cruzain of both families of inhibitors. In addition, it is important to highlight that the inhibitor



## *8. Conclusions and Outlook*

---

Cbz-Phe-Hph-(S) studied presents the advantage of a large stabilization of the cruzain-inhibitor covalent complex ( $-56.4 \text{ kcal}\cdot\text{mol}^{-1}$ ) by comparison with the inhibitor Bz-Tyr-Ala-CH<sub>2</sub>Cl ( $-27.7 \text{ kcal}\cdot\text{mol}^{-1}$ ). These results suggest that the tested epoxyketone Cbz-Phe-Hph-(S) can be a good candidate to design new cruzain inhibitors with higher potency. However given the low free energy barrier of the inhibitor Bz-Tyr-Ala-CH<sub>2</sub>Cl, the information derived from the reaction with both inhibitors could be combined to design an improved drug.

- Regarding with the inhibitor proposed for cruzain, the epoxyketone Cbz-Phe-Hph-(S), the analysis of average structures of the key states involved in the inhibitory reaction shows the role of the residues Gly66 and Asp158 of the active site, the interactions between these residues and the inhibitor stabilize the TS of the cruzain-inhibitor covalent complex. Then, it will be important to keep these favorable interactions, and probably substitute the phenyl group from the Hph part of the inhibitor for another one that can favor the interactions with the enzyme, to design new inhibitors.

- **Study of the Catalytic Mechanism of the Cruzain Cysteine Protease:** this study confirms that the catalytic process can be divided into an acylation and a deacylation stage, as previously described in the literature. Nevertheless, important new conclusions can be deduced from our study. First of all, the acylation stage, consisting of the nucleophilic attack on a carbonyl carbon of the peptide bond by the thiolate anion of an activated Cys25, takes place by a stepwise mechanism where a proton from a conserved His159 is transferred first to the N1 atom of the peptide and, after the transient intermediate I-II is located, the Cys25 attacks the carbonyl carbon atom. The second main conclusion is that the

deacylation stage, which was proposed to occur via a general base-catalyzed reaction whereby His159 activates a water molecule that attacks the peptide to produce the final products, would take place through a concerted mechanism. Importantly, this is the rate-limiting step of the full catalysis process. An understanding of this important enzyme at molecular level could be used for the rational design of new inhibitors based on not covalent bond interactions with the protein.

### **8.2 Outlook**

The main results obtained from the inhibition mechanism studies conducted during the present Doctoral Thesis allow us obtaining valuable information to design potent covalent inhibitors of cruzain and FP2. In the case of the cruzain, both Cbz-Phe-Hph-(S) and Bz-Tyr-Ala-CH<sub>2</sub>Cl inhibitors have been shown promising kinetic and thermodynamic parameters for the inhibition of this enzyme. Nevertheless, it should be noted that the cruzain-Cbz-Phe-Hph-(S) covalent complex is much more stable than the cruzain-Bz-Tyr-Ala-CH<sub>2</sub>Cl covalent complex. In the same way occurred with the natural inhibitor E64 of FP2; the inhibitor E64 could be used as a starting point to develop a proper inhibitor of FP2.

In the computational study of the catalytic mechanism of the cruzain, three TS are involved in both acylation and deacylation stage. Of these, in both TS1-II and TS2-II (see Paper IV in Chapter 7) the interactions between the peptide and cruzain are stronger than in the TS of the deacylation stage. Consequently, both TS1-II and TS2-II can be used to design TSA inhibitors of cruzain. These types of inhibitors have

## *8. Conclusions and Outlook*

---

become a promising target for the design of new inhibitors with potential applications in the treatment of severe human diseases.

Molecular dynamics simulations and binding free energy calculations have demonstrated to be appropriate tools to analyze the interactions between either cruzain or FP2 CPs and the inhibitors studied in the present Doctoral Thesis. In addition to the exploration of the reaction of the peptide bond breaking catalysed by cruzain provides clues for the design of TSA like inhibitors. Other families of CP inhibitors could be added to this study, as for example nitriles, vinyl sulfones, and other peptidic inhibitors with potent electrophilic groups. Regarding binding free energy calculations, rigorous but computationally expensive pathway methods such as as thermodynamic integration (TI), the Bennett acceptance ratio (BAR), or multistate extension of BAR called multistate Bennett acceptance ratio (MBAR), could be used for this purpose.

In the particular case of Malaria, despite the implementation of the RTS,S/AS01 malaria vaccine in the beginning of 2018 (recall that the vaccine only provides partial protection in young children), a new hope is opening with the discovery of the AP2-G protein. This is a protein involved in an especial phase, that allows the organism to pass from the asexual phase in which it proliferates in humans, to the sexual phase, the gametocytes, which is the one that survives if it is absorbed by anopheles mosquitoes. Blocking this protein could would break the passage of the parasite to the mosquitoes.<sup>326</sup> In this sense, the computational study of the AP2-G protein would allow improving our understanding of this important enzyme, which could be used for a rational design of new inhibitors with potential application in the treatment of Malaria.



## References

- (1) D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 5th ed.; W. H. Freeman and Company, New York, **2008**.
- (2) M. Sajid and J. H. McKerrow, *Molecular and Biochemical Parasitology*, **2002**, 121, 159-159.
- (3) N. C. S. Price, L., *Fundamentals of enzymology*; Oxford University Press, New York, **1999**.
- (4) J. B. Sumner, *Journal of Biological Chemistry*, **1926**, 69, 435-441.
- (5) C. H. W. Hirs, S. Moore and W. H. Stein, *Journal of Biological Chemistry*, **1960**, 235, 633-647.
- (6) [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1972](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1972).
- (7) C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Nature*, **1965**, 206, 757-761.
- (8) B. Gutte and R. B. Merrifield, *Journal of the American Chemical Society*, **1969**, 91, 501-502.
- (9) PDB, *Nature: New Biology*, **1971**, 233, 223.
- (10) T. T. Lah, J. L. Clifford, K. M. Helmer, N. A. Day, K. Moin, K. V. Honn, J. D. Crissman and B. F. Sloane, *Biochimica Et Biophysica Acta*, **1989**, 993, 63-73.
- (11) F. Lecaille, J. Kaleta and D. Bromme, *Chemical Reviews*, **2002**, 102, 4459-4488.
- (12) K. Komatsu, K. Tsukuda, J. Hosoya and S. Satoh, *Experimental Neurology*, **1986**, 93, 642-646.
- (13) H. C. Castro, P. A. Abreu, R. B. Geraldo, R. C. A. Martins, R. dos Santos, N. I. V. Loureiro, L. M. Cabral and C. R. Rodrigues, *Journal of Molecular Recognition*, **2011**, 24, 165-181.
- (14) M. Drag and G. S. Salvesen, *Nature reviews. Drug discovery*, **2010**, 9, 690-701.
- (15) O. L. Tavano, *Journal of Molecular Catalysis B: Enzymatic*, **2013**, 90, 1-11.
- (16) K. Saeki, K. Ozaki, T. Kobayashi and S. Ito, *Journal of Bioscience and Bioengineering*, **2007**, 103, 501-508.

- (17) J. H. McKerrow, J. C. Engel and C. R. Caffrey, *Bioorganic & Medicinal Chemistry*, **1999**, 7, 639-644.
- (18) A. R. Renslo and J. H. McKerrow, *Nature Chemical Biology*, **2006**, 2, 701-710.
- (19) R. E. Bernstein, *Journal of the Royal Society of Medicine*, **1984**, 77, 608-609.
- (20) L. S. Filardi and Z. Brener, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **1987**, 81, 755-759.
- (21) J. A. Castro, M. M. de Mecca and L. C. Bartel, *Human & Experimental Toxicology*, **2006**, 25, 471-479.
- (22) B. R. Shenai, P. S. Sijwali, A. Singh and P. J. Rosenthal, *Journal of Biological Chemistry*, **2000**, 275, 29000-29010.
- (23) S. Glushakova, J. Mazar, M. F. Hohmann-Marriott, E. Hama and J. Zimmerberg, *Cellular Microbiology*, **2009**, 11, 95-105.
- (24) W. R. Roush, J. M. Cheng, B. Knapp-Reed, A. Alvarez-Hernandez, J. H. McKerrow, E. Hansell and J. C. Engel, *Bioorganic & Medicinal Chemistry Letters*, **2001**, 11, 2759-2762.
- (25) R. Lonsdale, K. E. Ranaghan and A. J. Mulholland, *Chemical Communications*, **2010**, 46, 2354-2372.
- (26) M. W. van der Kamp and A. J. Mulholland, *Natural Product Reports*, **2008**, 25, 1001-1014.
- (27) A. Warshel and M. Karplus, *Journal of the American Chemical Society*, **1972**, 94, 5612-5625.
- (28) A. Warshel and M. Levitt, *Journal of Molecular Biology*, **1976**, 103, 227-249.
- (29) M. J. Field, P. A. Bash and M. Karplus, *Journal of Computational Chemistry*, **1990**, 11, 700-733.
- (30) [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2013/](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/).
- (31) W. L. Jorgensen, D. S. Maxwell and J. TiradoRives, *Journal of the American Chemical Society*, **1996**, 118, 11225-11236.
- (32) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, **1983**, 79, 926-935.
- (33) M. W. Roberts, *Catalysis Letters*, **2000**, 67, 1-4.

## **References**

---

- (34) [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1907](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1907).
- (35) E. Fischer, *Berichte der deutschen chemischen Gesellschaft*, **1894**, 27, 2985-2993.
- (36) D. E. Koshland, *Proceedings of the National Academy of Sciences of the United States of America*, **1958**, 44, 98-104.
- (37) L. Michaelis and M. L. Menten, *Biochemische Zeitschrift*, **1913**, 49, 333-369.
- (38) G. E. Briggs and J. B. S. Haldane, *Biochemical Journal*, **1925**, 19, 338-339.
- (39) D. Voet, J. Voet and C. W. Pratt, *Fundamentos de Bioquímica: La vida a nivel molecular*, Second ed.; Médica Panamericana, Buenos Aires, **2007**.
- (40) L. Pauling, *Chem. Eng. News*, **1946**, 24, 1375-1377.
- (41) S. Martí, M. Roca, J. Andres, V. Moliner, E. Silla, I. Tuñón and J. Bertran, *Chemical Society Reviews*, **2004**, 33, 98-107.
- (42) A. Warshel, *Proceedings of the National Academy of Sciences of the United States of America*, **1978**, 75, 5250-5254.
- (43) A. Warshel, *Journal of Biological Chemistry*, **1998**, 273, 27035-27038.
- (44) J. Villà and A. Warshel, *The Journal of Physical Chemistry B*, **2001**, 105, 7887-7907.
- (45) T. C. Bruice, *Annual Review of Biochemistry*, **1976**, 45, 331-373.
- (46) F. C. Lightstone and T. C. Bruice, *Journal of the American Chemical Society*, **1996**, 118, 2595-2605.
- (47) T. C. Bruice and S. J. Benkovic, *Biochemistry*, **2000**, 39, 6267-6274.
- (48) J. L. Griffin, M. W. Bowler, N. J. Baxter, K. N. Leigh, H. R. W. Dannatt, A. M. Hounslow, G. M. Blackburn, C. E. Webster, M. J. Cliff and J. P. Waltho, *Proceedings of the National Academy of Sciences of the United States of America*, **2012**, 109, 6910-6915.
- (49) M. I. Page and W. P. Jencks, *Proceedings of the National Academy of Sciences of the United States of America*, **1971**, 68, 1678-1683.

- (50) J. Hermans and L. Wang, *Journal of the American Chemical Society*, **1997**, 119, 2707-2714.
- (51) R. V. Stanton, M. Perakyla, D. Bakowies and P. A. Kollman, *Journal of the American Chemical Society*, **1998**, 120, 3448-3457.
- (52) B. Kuhn and P. A. Kollman, *Journal of the American Chemical Society*, **2000**, 122, 2586-2596.
- (53) P. A. Kollman, B. Kuhn, O. Donini, M. Perakyla, R. Stanton and D. Bakowies, *Accounts of Chemical Research*, **2001**, 34, 72-79.
- (54) A. Kohen, R. Cannio, S. Bartolucci and J. P. Klinman, *Nature*, **1999**, 399, 496-499.
- (55) M. J. Knapp and J. P. Klinman, *European Journal of Biochemistry*, **2002**, 269, 3113-3121.
- (56) M. J. Knapp, K. Rickert and J. P. Klinman, *Journal of the American Chemical Society*, **2002**, 124, 3865-3874.
- (57) J. P. Klinman, *Philosophical Transactions of the Royal Society B-Biological Sciences*, **2006**, 361, 1323-1331.
- (58) J. N. Bandaria, C. M. Cheatum and A. Kohen, *Journal of the American Chemical Society*, **2009**, 131, 10151-10155.
- (59) J. P. Klinman and A. Kohen In *Annual Review of Biochemistry*, Vol 82; Kornberg, R. D., Ed. **2013**; Vol. 82, p 471-496.
- (60) N. Kanaan, M. Roca, I. Tuñón, S. Martí and V. Moliner, *Physical Chemistry Chemical Physics*, **2010**, 12, 11657-11664.
- (61) M. Roca, M. Oliva, R. Castillo, V. Moliner and I. Tuñón, *Chemistry-a European Journal*, **2010**, 16, 11399-11411.
- (62) A. J. Adamczyk, J. Cao, S. C. L. Kamerlin and A. Warshel, *Proceedings of the National Academy of Sciences of the United States of America*, **2011**, 108, 14115-14120.
- (63) G. Bhabha, J. Lee, D. C. Ekiert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic and P. E. Wright, *Science*, **2011**, 332, 234-238.
- (64) N. Boekelheide, R. Salomon-Ferrer and T. F. Miller, III, *Proceedings of the National Academy of Sciences of the United States of America*, **2011**, 108, 16159-16163.
- (65) E. J. Loveridge, E. M. Behiry, J. Guo and R. K. Allemann, *Nature Chemistry*, **2012**, 4, 292-297.



## *References*

---

- (66) K. L. Grant and J. P. Klinman, *Biochemistry*, **1989**, 28, 6597-6605.
- (67) T. Jonsson, M. H. Glickman, S. J. Sun and J. P. Klinman, *Journal of the American Chemical Society*, **1996**, 118, 10319-10320.
- (68) B. J. Bahnson, T. D. Colby, J. K. Chin, B. M. Goldstein and J. P. Klinman, *Proceedings of the National Academy of Sciences of the United States of America*, **1997**, 94, 12797-12802.
- (69) A. Kohen, T. Jonsson and J. P. Klinman, *Biochemistry*, **1997**, 36, 2603-2611.
- (70) S. C. Tsai and J. P. Klinman, *Biochemistry*, **2001**, 40, 2303-2311.
- (71) M. H. M. Olsson, W. W. Parson and A. Warshel, *Chemical Reviews*, **2006**, 106, 1737-1756.
- (72) A. V. Pislakov, J. Cao, S. C. L. Kamerlin and A. Warshel, *Proceedings of the National Academy of Sciences of the United States of America*, **2009**, 106, 17359-17364.
- (73) S. C. L. Kamerlin and A. Warshel, *Proteins-Structure Function and Bioinformatics*, **2010**, 78, 1339-1375.
- (74) A. Kohen, *Accounts of Chemical Research*, **2015**, 48, 466-473.
- (75) L. Y. P. Luk, E. J. Loveridge and R. K. Allemann, *Physical Chemistry Chemical Physics*, **2015**, 17, 30817-30827.
- (76) A. Warshel and R. P. Bora, *Journal of Chemical Physics*, **2016**, 144, 1-17.
- (77) G. Careri, P. Fasella and E. Gratton, *Annual Review of Biophysics and Bioengineering*, **1979**, 8, 69-97.
- (78) B. Gavish and M. M. Werber, *Biochemistry*, **1979**, 18, 1269-1275.
- (79) J. A. McCammon, P. G. Wolynes and M. Karplus, *Biochemistry*, **1979**, 18, 927-942.
- (80) M. Karplus and J. A. McCammon, *Annual Review of Biochemistry*, **1983**, 52, 263-300.
- (81) K. Henzler-Wildman and D. Kern, *Nature*, **2007**, 450, 964-972.
- (82) K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Huebner and D. Kern, *Nature*, **2007**, 450, 838-844.

- (83) K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus and D. Kern, *Nature*, **2007**, 450, 913-916.
- (84) P. Singh, K. Francis and A. Kohen, *ACS Catalysis*, **2015**, 5, 3067-3073.
- (85) J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry*, Fifth ed.; W. H. Freeman and Company, New York, **2002**.
- (86) R. B. Silverman, *The Organic Chemistry of Enzyme-Catalyzed Reactions*; Academic Press, California, **2000**.
- (87) J. F. Morrison and C. T. Walsh, *Advances in Enzymology and Related Areas of Molecular Biology*, **1988**, 61, 201-301.
- (88) J. V. Schloss, *Accounts of Chemical Research*, **1988**, 21, 348-353.
- (89) J. C. Powers, J. L. Asgian, O. D. Ekici and K. E. James, *Chemical Reviews*, **2002**, 102, 4639-4750.
- (90) S. A. Bernhard and L. E. Orgel, *Science*, **1959**, 130, 625-626.
- (91) R. Wolfenden, *Nature*, **1969**, 223, 704-705.
- (92) R. Wolfenden, *Annual Review of Biophysics and Bioengineering*, **1976**, 5, 271-306.
- (93) D. W. Christianson and W. N. Lipscomb, *Journal of the American Chemical Society*, **1988**, 110, 5560-5565.
- (94) V. L. Schramm, *Accounts of Chemical Research*, **2015**, 48, 1032-1039.
- (95) V. L. Schramm In *Annual Review of Biochemistry, Vol 80*; Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., Thorner, J. W., Eds. **2011**; Vol. 80, p 703-732.
- (96) V. L. Schramm, *ACS chemical biology*, **2013**, 8, 71-81.
- (97) R. E. Babine and S. L. Bender, *Chemical Reviews*, **1997**, 97, 1359-1472.
- (98) J. A. Gutierrez, T. Crowder, A. Rinaldo-Matthis, M.-C. Ho, S. C. Almo and V. L. Schramm, *Nature Chemical Biology*, **2009**, 5, 251-257.
- (99) M. B. Sturm, P. C. Tyler, G. B. Evans and V. L. Schramm, *Biochemistry*, **2009**, 48, 9941-9948.
- (100) Y. Zhang and V. L. Schramm, *Journal of the American Chemical Society*, **2010**, 132, 8787-8794.

## References

---

- (101) Y. Zhang, G. B. Evans, K. Clinch, D. R. Crump, L. D. Harris, R. F. G. Froehlich, P. C. Tyler, K. Z. Hazleton, M. B. Cassera and V. L. Schramm, *Journal of Biological Chemistry*, **2013**, 288, 34746-34754.
- (102) G. L. Patrick, *An Introduction to Medicinal Chemistry*, Third ed.; Oxford University Press, New York, **2005**.
- (103) J. K. Gierse, C. M. Koboldt, M. C. Walker, K. Seibert and P. C. Isakson, *Biochemical Journal*, **1999**, 339, 607-614.
- (104) D. S. Johnson, C. Stiff, S. E. Lazerwith, S. R. Kesten, L. K. Fay, M. Morris, D. Beidler, M. B. Liimatta, S. E. Smith, D. T. Dudley, N. Sadagopan, S. N. Bhattachar, S. J. Kesten, T. K. Nomanbhoy, B. F. Cravatt and K. Ahn, *Acs Medicinal Chemistry Letters*, **2011**, 2, 91-96.
- (105) F. E. Kwarcinski, C. C. Fox, M. E. Steffey and M. B. Soellner, *ACS chemical biology*, **2012**, 7, 1910-1917.
- (106) C. W. Zapf, B. S. Gerstenberger, L. Xing, D. C. Limburg, D. R. Anderson, N. Caspers, S. Han, A. Aulabaugh, R. Kurumbail, S. Shakya, X. Li, V. Spaulding, R. M. Czerwinski, N. Seth and Q. G. Medley, *Journal of medicinal chemistry*, **2012**, 55, 10047-10063.
- (107) A. B. Dounay, M. Anderson, B. M. Bechle, B. M. Campbell, M. M. Claffey, A. Evdokimov, E. Evrard, K. R. Fonseca, X. Gan, S. Ghosh, M. M. Hayward, W. Horner, J.-Y. Kim, L. A. McAllister, J. Pandit, V. Paradis, V. D. Parikh, M. R. Reese, S. Rong, M. A. Salafia, K. Schuyten, C. A. Strick, J. B. Tuttle, J. Valentine, H. Wang, L. E. Zawadzke and P. R. Verhoest, *Acs Medicinal Chemistry Letters*, **2012**, 3, 187-192.
- (108) A. Warshel, *Angewandte Chemie-International Edition*, **2014**, 53, 10020-10031.
- (109) S. V. Gulnik, L. I. Suvorov, B. S. Liu, B. Yu, B. Anderson, H. Mitsuya and J. W. Erickson, *Biochemistry*, **1995**, 34, 9282-9287.
- (110) N. Singh, M. P. Frushicheva and A. Warshel, *Proteins-Structure Function and Bioinformatics*, **2012**, 80, 1110-1122.
- (111) H. H. Otto and T. Schirmeister, *Chemical Reviews*, **1997**, 97, 133-171.
- (112) P. J. Rosenthal, *International Journal for Parasitology*, **2004**, 34, 1489-1499.
- (113) N. D. Rawlings, A. J. Barrett and R. Finn, *Nucleic Acids Research*, **2016**, 44, D343-D350.

- (114) J. W. Keillor and R. S. Brown, *Journal of the American Chemical Society*, **1992**, 114, 7983-7989.
- (115) K. Brocklehurst, *International Journal of Biochemistry*, **1979**, 10, 259-274.
- (116) S. D. Lewis, F. A. Johnson and J. A. Shafer, *Biochemistry*, **1976**, 15, 5009-5017.
- (117) D. J. Creighton, M. S. Gessouroun and J. M. Heapes, *Febs Letters*, **1980**, 110, 319-322.
- (118) S. D. Lewis, F. A. Johnson and J. A. Shafer, *Biochemistry*, **1981**, 20, 48-51.
- (119) M. J. Harrison, N. A. Burton and I. H. Hillier, *Journal of the American Chemical Society*, **1997**, 119, 12285-12291.
- (120) W. G. Han, E. Tajkhorshid and S. Suhai, *Journal of Biomolecular Structure & Dynamics*, **1999**, 16, 1019-1032.
- (121) M. Mladenovic, K. Junold, R. F. Fink, W. Thiel, T. Schirmeister and B. Engels, *Journal of Physical Chemistry B*, **2008**, 112, 5458-5469.
- (122) G. Grazioso, L. Legnani, L. Toma, R. Ettari, N. Micale and C. De Micheli, *Journal of Computer-Aided Molecular Design*, **2012**, 26, 1035-1043.
- (123) D. H. Wei, X. Q. Huang, J. J. Liu, M. S. Tang and C. G. Zhan, *Biochemistry*, **2013**, 52, 5145-5154.
- (124) G. Lowe, *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **1970**, 257, 237-248.
- (125) G. Lowe and Yuthavon.Y, *Biochemical Journal*, **1971**, 124, 107-115.
- (126) R. Hershfield and G. L. Schmir, *Journal of the American Chemical Society*, **1972**, 94, 1263-1270.
- (127) J. Drenth, H. M. Swen, W. Hoogenstraaten and L. A. A. Sluyterman, *Proceedings of the Koninklijke Nederlandse Akademie Van Wetenschappen Series C-Biological and Medical Sciences*, **1975**, 78, 104-110.
- (128) J. Drenth, K. H. Kalk and H. M. Swen, *Biochemistry*, **1976**, 15, 3731-3738.

## *References*

---

- (129) K. J. Angelides and A. L. Fink, *Biochemistry*, **1979**, 18, 2355-2363.
- (130) K. J. Angelides and A. L. Fink, *Biochemistry*, **1979**, 18, 2363-2369.
- (131) J. W. Keillor, A. A. Neverov and R. S. Brown, *Journal of American Chemical Society*, **1994**, 116, 4669-4673.
- (132) A. E. Howard and P. A. Kollman, *Journal of the American Chemical Society*, **1988**, 110, 7195-7200.
- (133) D. Arad, R. Langridge and P. A. Kollman, *Journal of the American Chemical Society*, **1990**, 112, 491-502.
- (134) W. J. Welsh and Y. Lin, *Journal of Molecular Structure-Theochem*, **1997**, 401, 315-326.
- (135) K. Byun and J. L. Gao, *Journal of Molecular Graphics & Modelling*, **2000**, 18, 50-55.
- (136) M. Strajbl, J. Florian and A. Warshel, *Journal of Physical Chemistry B*, **2001**, 105, 4471-4484.
- (137) S. Ma, L. S. Devi-Kesavan and J. Gao, *Journal of the American Chemical Society*, **2007**, 129, 13633-13645.
- (138) S. Vijayakumar and P. Kolandaivel, *International Journal of Quantum Chemistry*, **2008**, 108, 927-936.
- (139) R. Shankar, P. Kolandaivel and K. Senthilkumar, *International Journal of Quantum Chemistry*, **2010**, 110, 1660-1674.
- (140) M. Shokhen, N. Khazanov and A. Albeck, *Proteins-Structure Function and Bioinformatics*, **2011**, 79, 975-985.
- (141) A. Fekete and I. Komaromi, *Physical Chemistry Chemical Physics*, **2016**, 18, 32847-32861.
- (142) S. Itow and E. P. Camargo, *Journal of Protozoology*, **1977**, 24, 591-595.
- (143) A. E. Eakin, A. A. Mills, G. Harth, J. H. McKerrow and C. S. Craik, *Journal of Biological Chemistry*, **1992**, 267, 7411-7420.
- (144) J. Scharfstein, M. Schechter, M. Senna, J. M. Peralta, L. Mendoncapreviato and M. A. Miles, *Journal of Immunology*, **1986**, 137, 1336-1341.

- (145) O. Campetella, J. Henriksson, L. Aslund, A. C. C. Frasc, U. Pettersson and J. J. Cazzulo, *Molecular and Biochemical Parasitology*, **1992**, 50, 225-234.
- (146) A. R. Schnapp, C. S. Eickhoff, D. Sizemore, R. Curtiss and D. F. Hoft, *Infection and Immunity*, **2002**, 70, 5065-5074.
- (147) N. Fujii, J. P. Mallari, E. J. Hansell, Z. Mackey, P. Doyle, Y. M. Zhou, J. Gut, P. J. Rosenthal, J. H. McKerrow and R. K. Guy, *Bioorganic & Medicinal Chemistry Letters*, **2005**, 15, 121-123.
- (148) J. Clayton, *Nature*, **2010**, 465, S4-S5.
- (149) A. C. Aufderheide, W. Salo, M. Madden, J. Streitz, J. Buikstra, F. Guhl, B. Arriaza, C. Renier, L. E. Wittmers, G. Fornaciari and M. Allison, *Proceedings of the National Academy of Sciences of the United States of America*, **2004**, 101, 2034-2039.
- (150) Y. Jackson, A. Angheben, B. Carrilero Fernandez, J. M. Jansa i Lopez del Vallado, J. G. Jannin and P. Albajar-Vinas, *Bulletin de la Societe de pathologie exotique* (1990), **2009**, 102, 326-329.
- (151) J. Gascon, C. Bern and M.-J. Pinazo, *Acta Tropica*, **2010**, 115, 22-27.
- (152) <http://www.who.int/mediacentre/factsheets/fs340/en/>, accessed **March 2017**.
- (153) J. H. McKerrow, M. E. McGrath and J. C. Engel, *Parasitology Today*, **1995**, 11, 279-282.
- (154) G. Harth, N. Andrews, A. A. Mills, J. C. Engel, R. Smith and J. H. McKerrow, *Molecular and Biochemical Parasitology*, **1993**, 58, 17-24.
- (155) J. J. Cazzulo, V. Stoka and V. Turk, *Biological Chemistry*, **1997**, 378, 1-10.
- (156) J. C. Engel, P. S. Doyle, I. Hsieh and J. H. McKerrow, *Journal of Experimental Medicine*, **1998**, 188, 725-734.
- (157) J. C. Engel, P. S. Doyle, J. Palmer, I. Hsieh, D. F. Bainton and J. H. McKerrow, *Journal of Cell Science*, **1998**, 111, 597-606.
- (158) J. H. McKerrow, *International Journal for Parasitology*, **1999**, 29, 833-837.
- (159) J. J. Cazzulo, *Current topics in medicinal chemistry*, **2002**, 2, 1261-1271.

## *References*

---

- (160) F. Polticelli, G. Zaini, A. Bolli, G. Antonini, L. Gradoni and P. Ascenzi, *Biochemistry*, **2005**, 44, 2781-2789.
- (161) W. Jacobsen, U. Christians and L. Z. Benet, *Drug Metabolism and Disposition*, **2000**, 28, 1343-1351.
- (162) S. C. Barr, K. L. Warner, B. G. Kornreic, J. Piscitelli, A. Wolfe, L. Benet and J. H. McKerrow, *Antimicrobial Agents and Chemotherapy*, **2005**, 49, 5160-5161.
- (163) P. S. Doyle, Y. M. Zhou, J. C. Engel and J. H. McKerrow, *Antimicrobial Agents and Chemotherapy*, **2007**, 51, 3932-3939.
- (164) K. Brak, P. S. Doyle, J. H. McKerrow and J. A. Ellman, *Journal of the American Chemical Society*, **2008**, 130, 6404-6410.
- (165) J. Lavrado, Z. Mackey, E. Hansell, J. H. McKerrow, A. Paulo and R. Moreira, *Bioorganic & Medicinal Chemistry Letters*, **2012**, 22, 6256-6260.
- (166) J. M. dos Santos Filho, D. R. M. Moreira, C. A. de Simone, R. S. Ferreira, J. H. McKerrow, C. S. Meira, E. T. Guimaraes and M. B. Pereira Soares, *Bioorganic & Medicinal Chemistry*, **2012**, 20, 6423-6433.
- (167) J. W. Choy, C. Bryant, C. M. Calvet, P. S. Doyle, S. S. Gunatilleke, S. S. F. Leung, K. K. H. Ang, S. Chen, J. Gut, J. A. Oses-Prieto, J. B. Johnston, M. R. Arkin, A. L. Burlingame, J. Taunton, M. P. Jacobson, J. M. McKerrow, L. M. Podust and A. R. Renslo, *Beilstein Journal of Organic Chemistry*, **2013**, 9, 15-25.
- (168) B. D. Fennell, J. M. Warren, K. K. Chung, H. L. Main, A. B. Arend, A. Tochowicz and M. G. Goetz, *Journal of Enzyme Inhibition and Medicinal Chemistry*, **2013**, 28, 468-478.
- (169) R. S. Ferreira, M. A. Desso, I. Pauli, M. L. Souza, R. Krogh, A. I. L. Sales, G. Oliva, L. C. Dias and A. D. Andricopulo, *Journal of medicinal chemistry*, **2014**, 57, 2380-2392.
- (170) A. Laveran, *Bull Acad Nat Med*, **1888**, 9, 1234-1235.
- (171) [http://www.nobelprize.org/nobel\\_prizes/medicine/laureates/1907/](http://www.nobelprize.org/nobel_prizes/medicine/laureates/1907/).
- (172) S. Y. Tan and H. Sung, *Singapore Medical Journal*, **2008**, 49, 370-371.
- (173) [http://www.nobelprize.org/nobel\\_prizes/medicine/laureates/1902/](http://www.nobelprize.org/nobel_prizes/medicine/laureates/1902/).

- (174) <http://www.who.int/malaria/publications/world-malaria-report-2016/report/en/>, accessed April 2017.
- (175) A. J. Birkett, *Vaccine*, **2016**, 34, 2915-2920.
- (176) <http://www.malariavaccine.org>, accessed April 2017.
- (177) J. A. Regules, S. B. Cicatelli, J. W. Bennett, K. M. Paolino, P. S. Twomey, J. E. Moon, A. K. Kathcart, K. D. Hauns, J. L. Komisar, A. N. Qabar, S. A. Davidson, S. Dutta, M. E. Griffith, C. D. Magee, M. Wojnarski, J. R. Livezey, A. T. Kress, P. E. Waterman, E. Jongert, U. Wille-Reece, W. Volkmuth, D. Emerling, W. H. Robinson, M. Lievens, D. Morelle, C. K. Lee, B. Yassin-Rajkumar, R. Weltzin, J. Cohen, R. M. Paris, N. C. Waters, A. J. Birkett, D. C. Kaslow, W. R. Ballou, C. F. Ockenhouse and J. Vekemans, *Journal of Infectious Diseases*, **2016**, 214, 762-771.
- (178) D. Kazmin, H. I. Nakaya, E. K. Lee, M. J. Johnson, R. van der Most, R. A. van den Berg, W. R. Ballou, E. Jongert, U. Wille-Reece, C. Ockenhouse, A. Aderem, D. E. Zak, J. Sadoff, J. Hendriks, J. Wrammert, R. Ahmed and B. Pulendran, *Proceedings of the National Academy of Sciences of the United States of America*, **2017**, 114, 2425-2430.
- (179) S. Mahmoudi and H. Keshavarz, *Human vaccines & immunotherapeutics*, **2017**, 1-4.
- (180) S. Mahmoudi and H. Keshavarz, *Human vaccines & immunotherapeutics*, **2017**, 0-0.
- (181) R. Ettari, F. Bova, M. Zappala, S. Grasso and N. Micale, *Medicinal Research Reviews*, **2010**, 30, 136-167.
- (182) J. H. McKerrow, E. Sun, P. J. Rosenthal and J. Bouvier, *Annual Review of Microbiology*, **1993**, 47, 821-853.
- (183) S. E. Francis, R. Banerjee and D. E. Goldberg, *Journal of Biological Chemistry*, **1997**, 272, 14961-14968.
- (184) V. L. Lew, T. Tiffert and H. Ginsburg, *Blood*, **2003**, 101, 4189-4194.
- (185) P. J. Rosenthal, J. H. McKerrow, M. Aikawa, H. Nagasawa and J. H. Leech, *Journal of Clinical Investigation*, **1988**, 82, 1560-1566.
- (186) S. E. Francis, I. Y. Gluzman, A. Oksman, A. Knickerbocker, R. Mueller, M. L. Bryant, D. R. Sherman, D. G. Russell and D. E. Goldberg, *Embo Journal*, **1994**, 13, 306-317.



## *References*

---

- (187) J. Hill, L. Tyas, L. H. Phylip, J. Kay, B. M. Dunn and C. Berry, *Febs Letters*, **1994**, 352, 155-158.
- (188) K. K. Eggleston, K. L. Duffin and D. E. Goldberg, *Journal of Biological Chemistry*, **1999**, 274, 32411-32417.
- (189) R. Banerjee, J. Liu, W. Beatty, L. Pelosof, M. Klemba and D. E. Goldberg, *Proceedings of the National Academy of Sciences of the United States of America*, **2002**, 99, 990-995.
- (190) P. J. Rosenthal, P. S. Sijwali, A. Singh and B. R. Shenai, *Current Pharmaceutical Design*, **2002**, 8, 1659-1672.
- (191) I. Y. Gluzman, S. E. Francis, A. Oksman, C. E. Smith, K. L. Duffin and D. E. Goldberg, *The Journal of Clinical Investigation*, **1994**, 93, 1602-1608.
- (192) A. Singh and P. J. Rosenthal, *Journal of Biological Chemistry*, **2004**, 279, 35236-35241.
- (193) P. S. Sijwali, B. R. Shenai, J. Gut, A. Singh and P. J. Rosenthal, *Biochemical Journal*, **2001**, 360, 481-489.
- (194) M. Marco and J. Miguel Coteron, *Current topics in medicinal chemistry*, **2012**, 12, 408-444.
- (195) I. D. Kerr, J. H. Lee, K. C. Pandey, A. Harrison, M. Sajid, P. J. Rosenthal and L. S. Brinen, *Journal of medicinal chemistry*, **2009**, 52, 852-857.
- (196) V. Ehmke, C. Heindl, M. Rottmann, C. Freymond, W. B. Schweizer, R. Brun, A. Stich, T. Schirmeister and F. Diederich, *ChemMedChem*, **2011**, 6, 273-278.
- (197) R. Ettari, N. Micale, G. Grazioso, F. Bova, T. Schirmeister, S. Grasso and M. Zappala, *ChemMedChem*, **2012**, 7, 1594-1600.
- (198) A. F. Marques, D. Esser, P. J. Rosenthal, M. U. Kassack and L. M. T. R. Lima, *Bioorganic & Medicinal Chemistry*, **2013**, 21, 3667-3673.
- (199) S. Chatterjee, K. Tanabe and E. A. Nodiff, *Bioorganic & Medicinal Chemistry Letters*, **2014**, 24, 4106-4109.
- (200) D. J. Weldon, F. Shah, A. G. Chittiboyina, A. Sheri, R. R. Chada, J. Gut, P. J. Rosenthal, D. Shivakumar, W. Sherman, P. Desai, J.-C. Jung and M. A. Avery, *Bioorganic & Medicinal Chemistry Letters*, **2014**, 24, 1274-1279.

- (201) S. X. Wang, K. C. Pandey, J. R. Somoza, P. S. Sijwali, T. Kortemme, L. S. Brinen, R. J. Fletterick, P. J. Rosenthal and J. H. McKerrow, *Proceedings of the National Academy of Sciences of the United States of America*, **2006**, 103, 11503-11508.
- (202) G. Hansen, A. Heitmann, T. Witt, H. Li, H. Jiang, X. Shen, V. T. Heussler, A. Rennenberg and R. Hilgenfeld, *Structure*, **2011**, 19, 919-929.
- (203) K. Suzuki, S. Tsuji and S. Ishiura, *Febs Letters*, **1981**, 136, 119-122.
- (204) K. Hanada, M. Tamai, M. Yamagishi, S. Ohmura, J. Sawada and I. Tanaka, *Agricultural and Biological Chemistry*, **1978**, 42, 523-528.
- (205) K. Hanada, M. Tamai, S. Ohmura, J. Sawada, T. Seki and I. Tanaka, *Agricultural and Biological Chemistry*, **1978**, 42, 529-536.
- (206) W. R. Roush, F. V. Gonzalez, J. H. McKerrow and E. Hansell, *Bioorganic & Medicinal Chemistry Letters*, **1998**, 8, 2809-2812.
- (207) K. E. James, J. L. Asgian, Z. Z. Li, O. D. Ekici, J. R. Rubin, J. Mikolajczyk, G. S. Salvesen and J. C. Powers, *Journal of medicinal chemistry*, **2004**, 47, 1553-1574.
- (208) Y. Yabe, D. Guillaume and D. H. Rich, *Journal of the American Chemical Society*, **1988**, 110, 4043-4044.
- (209) K. Matsumoto, D. Yamamoto, H. Ohishi, K. Tomoo, T. Ishida, M. Inoue, T. Sadatome, K. Kitamura and H. Mizuno, *Febs Letters*, **1989**, 245, 177-180.
- (210) K. I. Varughese, F. R. Ahmed, P. R. Carey, S. Hasnain, C. P. Huber and A. C. Storer, *Biochemistry*, **1989**, 28, 1330-1332.
- (211) D. Yamamoto, K. Matsumoto, H. Ohishi, T. Ishida, M. Inoue, K. Kitamura and H. Mizuno, *Journal of Biological Chemistry*, **1991**, 266, 14771-14777.
- (212) M. J. Kim, D. Yamamoto, K. Matsumoto, M. Inoue, T. Ishida, H. Mizuno, S. Sumiya and K. Kitamura, *Biochemical Journal*, **1992**, 287, 797-803.
- (213) K. I. Varughese, Y. Su, D. Cromwell, S. Hasnain and N. H. Xuong, *Biochemistry*, **1992**, 31, 5172-5176.
- (214) B. G. Zhao, C. A. Janson, B. Y. Amegadzie, K. Dalessio, C. Griffin, C. R. Hanning, C. Jones, J. Kurdyla, M. McQueney, X. Y. Qiu,

## References

---

- W. W. Smith and S. S. AbdelMeguid, *Nature Structural Biology*, **1997**, 4, 109-111.
- (215) A. Yamamoto, K. Tomoo, K. Matsugi, T. Hara, Y. In, M. Murata, K. Kitamura and T. Ishida, *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, **2002**, 1597, 244-251.
- (216) R. Ghosh, S. Chakraborty, C. Chakrabarti, J. K. Dattagupta and S. Biswas, *Febs Journal*, **2008**, 275, 421-434.
- (217) R. Bihovsky, *Journal of Organic Chemistry*, **1992**, 57, 1029-1031.
- (218) D. H. Rich, *Proteinase Inhibitors*; Elsevier, The Netherlands, **1986**.
- (219) J. P. Meara and D. H. Rich, *Journal of medicinal chemistry*, **1996**, 39, 3357-3366.
- (220) D. H. Kim, Y. H. Jin and C. H. Ryu, *Bioorganic & Medicinal Chemistry*, **1997**, 5, 2103-2108.
- (221) H. Helten, T. Schirmeister and B. Engels, *Journal of Physical Chemistry A*, **2004**, 108, 7691-7701.
- (222) H. Helten, T. Schirmeister and B. Engels, *Journal of Organic Chemistry*, **2005**, 70, 233-237.
- (223) R. Vicik, H. Helten, T. Schirmeister and B. Engels, *ChemMedChem*, **2006**, 1, 1021-1028.
- (224) F. V. Gonzalez, J. Izquierdo, S. Rodriguez, J. H. McKerrow and E. Hansell, *Bioorganic & Medicinal Chemistry Letters*, **2007**, 17, 6697-6700.
- (225) S. Royo, S. Rodriguez, T. Schirmeister, J. Kesselring, M. Kaiser and F. V. Gonzalez, *ChemMedChem*, **2015**, 10, 1484-1487.
- (226) A. Latorre, T. Schirmeister, J. Kesselring, S. Jung, P. Johe, U. A. Hellmich, A. Heilos, B. Engels, R. L. Krauth-Siegel, N. Dirdjaja, L. Bou-Iserte, S. Rodriguez and F. V. Gonzalez, *Acs Medicinal Chemistry Letters*, **2016**, 7, 1073-1076.
- (227) G. Schoellmann and E. Shaw, *Biochemistry*, **1963**, 2, 252-255.
- (228) D. Rasnick, *Analytical Biochemistry*, **1985**, 149, 461-465.
- (229) P. Rauber, H. Angliker, B. Walker and E. Shaw, *Biochemical Journal*, **1986**, 239, 633-640.

- (230) F. Ashall, H. Angliker and E. Shaw, *Biochemical and Biophysical Research Communications*, **1990**, 170, 923-929.
- (231) M. E. McGrath, A. E. Eakin, J. C. Engel, J. H. McKerrow, C. S. Craik and R. J. Fletterick, *Journal of Molecular Biology*, **1995**, 247, 251-259.
- (232) S. A. Gillmor, C. S. Craik and R. J. Fletterick, *Protein Science*, **1997**, 6, 1603-1611.
- (233) R. Bihovsky, J. C. Powers, C. M. Kam, R. Walton and R. C. Loewi, *Journal of Enzyme Inhibition*, **1993**, 7, 15-25.
- (234) W. R. Roush, A. A. Hernandez, J. H. McKerrow, P. M. Selzer, E. Hansell and J. C. Engel, *Tetrahedron*, **2000**, 56, 9747-9762.
- (235) A. Albeck, R. Persky and S. Kliper, *Bioorganic & Medicinal Chemistry Letters*, **1995**, 5, 1767-1772.
- (236) A. Albeck, S. Fluss and R. Persky, *Journal of the American Chemical Society*, **1996**, 118, 3591-3596.
- (237) A. Albeck and S. Kliper, *Biochemical Journal*, **1997**, 322, 879-884.
- (238) A. Warshel and R. M. Weiss, *Journal of the American Chemical Society*, **1980**, 102, 6218-6226.
- (239) A. Warshel, *Annual Review of Biophysics and Biomolecular Structure*, **2003**, 32, 425-443.
- (240) P. E. M. Siegbahn and F. Himo, *Wiley Interdisciplinary Reviews-Computational Molecular Science*, **2011**, 1, 323-336.
- (241) F. Himo, *Journal of the American Chemical Society*, **2017**, 139, 6780-6786.
- (242) M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *Journal of Physical Chemistry*, **1996**, 100, 19357-19363.
- (243) S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma and M. J. Frisch, *Journal of Molecular Structure-Theochem*, **1999**, 461, 1-21.
- (244) *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*; Royal Society of Chemistry, Cambridge, **2017**.
- (245) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *Journal of the American Chemical Society*, **1984**, 106, 765-784.

## References

---

- (246) S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *Journal of Computational Chemistry*, **1986**, 7, 230-252.
- (247) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society*, **1995**, 117, 5179-5197.
- (248) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, **1983**, 4, 187-217.
- (249) A. D. MacKerell, J. Wiorkiewicz-Kuczera and M. Karplus, *Journal of the American Chemical Society*, **1995**, 117, 11946-11975.
- (250) A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B*, **1998**, 102, 3586-3616.
- (251) A. D. MacKerell, N. Banavali and N. Foloppe, *Biopolymers*, **2000**, 56, 257-265.
- (252) A. D. Mackerell, M. Feig and C. L. Brooks, *Journal of Computational Chemistry*, **2004**, 25, 1400-1415.
- (253) W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society*, **1988**, 110, 1657-1666.
- (254) J. Pranata, S. G. Wierschke and W. L. Jorgensen, *Journal of the American Chemical Society*, **1991**, 113, 2810-2819.
- (255) M. J. Field, *A Practical Introduction to the Simulation of Molecular Systems*, Second Edition ed.; Cambridge University Press, Cambridge, UK, **2007**.
- (256) U. C. Singh and P. A. Kollman, *Journal of Computational Chemistry*, **1986**, 7, 718-730.
- (257) G. G. Ferenczy, J. L. Rivail, P. R. Surjan and G. Narayszabo, *Journal of Computational Chemistry*, **1992**, 13, 830-837.
- (258) V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret and G. G. Ferenczy, *Journal of Computational Chemistry*, **1994**, 15, 269-282.

- (259) G. Monard, M. Loos, V. Thery, K. Baka and J. L. Rivail, *International Journal of Quantum Chemistry*, **1996**, 58, 153-159.
- (260) N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *Journal of Physical Chemistry A*, **2000**, 104, 1720-1735.
- (261) J. L. Gao and D. G. Truhlar, *Annual Review of Physical Chemistry*, **2002**, 53, 467-505.
- (262) J. L. Gao, P. Amara, C. Alhambra and M. J. Field, *Journal of Physical Chemistry A*, **1998**, 102, 4714-4721.
- (263) H. M. Senn and W. Thiel, *Angewandte Chemie-International Edition*, **2009**, 48, 1198-1229.
- (264) M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *Journal of the American Chemical Society*, **1985**, 107, 3902-3909.
- (265) J. J. P. Stewart, *Journal of Computational Chemistry*, **1989**, 10, 209-220.
- (266) M. J. S. Dewar and W. Thiel, *Journal of the American Chemical Society*, **1977**, 99, 4899-4907.
- (267) M. J. S. Dewar and W. Thiel, *Journal of the American Chemical Society*, **1977**, 99, 4907-4917.
- (268) K. Nam, Q. Cui, J. Gao and D. M. York, *Journal of Chemical Theory and Computation*, **2007**, 3, 486-504.
- (269) W. Thiel and A. A. Voityuk, *Theoretica Chimica Acta*, **1996**, 93, 315-315.
- (270) P. Hohenberg and W. Kohn, *Physical Review B*, **1964**, 136, B864-B871.
- (271) W. Kohn and L. J. Sham, *Physical Review*, **1965**, 140, 1133-1138.
- (272) E. Runge and E. K. U. Gross, *Physical Review Letters*, **1984**, 52, 997-1000.
- (273) A. D. Becke, *Physical Review A*, **1988**, 38, 3098-3100.
- (274) C. T. Lee, W. T. Yang and R. G. Parr, *Physical Review B*, **1988**, 37, 785-789.
- (275) Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, **2008**, 120, 215-241.

## ***References***

---

- (276) Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, **2008**, 4, 1849-1868.
- (277) Y. Zhao and D. G. Truhlar, *Accounts of Chemical Research*, **2008**, 41, 157-167.
- (278) J. Zheng, Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, **2009**, 5, 808-821.
- (279) Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, **2010**, 6, 1104-1108.
- (280) J. C. Slater, *Physical Review*, **1930**, 36, 0057-0064.
- (281) S. F. Boys, *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences*, **1950**, 200, 542-554.
- (282) W. J. Hehre, L. Radom, P. v. R. Schleyer and J. A. Pople, *Ab Initio Molecular Orbital Theory*, New York, **1986**.
- (283) B. J. Lynch, Zhao, Y., Truhlar, D. G., *Journal of Physical Chemistry A*, **2003**, 107, 1384-1388.
- (284) M. Born and R. Oppenheimer, *Annalen Der Physik*, **1927**, 84, 0457-0484.
- (285) D. Frenkel and B. Smit; Academic Press: **1996**.
- (286) G. Vanderzwan and J. T. Hynes, *Chemical Physics*, **1984**, 90, 21-35.
- (287) M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science*, **2004**, 303, 186-195.
- (288) R. Fletcher, *Computer Journal*, **1970**, 13, 317-322.
- (289) D. Goldfarb, *Mathematics of Computation*, **1970**, 24, 23-26.
- (290) D. F. Shanno, *Mathematics of Computation*, **1970**, 24, 647-656.
- (291) R. H. Byrd, P. H. Lu, J. Nocedal and C. Y. Zhu, *Siam Journal on Scientific Computing*, **1995**, 16, 1190-1208.
- (292) K. E. Ranaghan and A. J. Mulholland, *International Reviews in Physical Chemistry*, **2010**, 29, 65-133.
- (293) S. Fischer and M. Karplus, *Chemical Physics Letters*, **1992**, 194, 252-261.

- (294) V. Moliner, A. J. Turner and I. H. Williams, *Chemical Communications*, **1997**, 1271-1272.
- (295) A. J. Turner, V. Moliner and I. H. Williams, *Physical Chemistry Chemical Physics*, **1999**, 1, 1323-1331.
- (296) G. Monard, X. Prat-Resina, A. Gonzalez-Lafont and J. M. Lluch, *International Journal of Quantum Chemistry*, **2003**, 93, 229-244.
- (297) X. Prat-Resina, A. Gonzalez-Lafont and J. M. Lluch, *Journal of Molecular Structure-Theochem*, **2003**, 632, 297-307.
- (298) T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel and M. J. Frisch, *Journal of Computational Chemistry*, **2003**, 24, 760-769.
- (299) X. Prat-Resina, J. M. Bofill, A. Gonzalez-Lafont and J. M. Lluch, *International Journal of Quantum Chemistry*, **2004**, 98, 367-377.
- (300) S. Martí, V. Moliner and I. Tuñón, *Journal of Chemical Theory and Computation*, **2005**, 1, 1008-1016.
- (301) S. Marti, V. Moliner, M. Tunon and I. H. Williams, *Journal of Physical Chemistry B*, **2005**, 109, 3707-3710.
- (302) J. Kästner, H. M. Senn, S. Thiel, N. Otte and W. Thiel, *Journal of Chemical Theory and Computation*, **2006**, 2, 452-461.
- (303) G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, **1977**, 23, 187-199.
- (304) L. Verlet, *Phys. Rev.*, **1967**, 159, 98-103.
- (305) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Journal of Chemical Physics*, **1953**, 21, 1087-1092.
- (306) W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *Journal of Chemical Physics*, **1982**, 76, 637-649.
- (307) S. Nosé and M. L. Klein, *Molecular Physics*, **1983**, 50, 1055-1076.
- (308) S. Nosé, *The Journal of Chemical Physics*, **1984**, 81, 511-519.
- (309) W. G. Hoover, *Physical Review A*, **1985**, 31, 1695-1697.
- (310) G. S. Grest and K. Kremer, *Physical Review A*, **1986**, 33, 3628-3631.
- (311) S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *Journal of Computational Chemistry*, **1992**, 13, 1011-1021.



## *References*

---

- (312) J. Kästner and W. Thiel, *The Journal of Chemical Physics*, **2005**, 123, 144104.
- (313) J. Kästner and W. Thiel, *Journal of Chemical Physics*, **2006**, 124, 234106.
- (314) E. Rosta and G. Hummer, *Journal of Chemical Theory and Computation*, **2015**, 11, 276-285.
- (315) A. L. Ferguson, *Journal of Computational Chemistry*, **2017**, 38, 1583-1605.
- (316) J. J. Ruiz-Pernia, E. Silla, I. Tuñón, S. Martí and V. Moliner, *Journal of Physical Chemistry B*, **2004**, 108, 8427-8433.
- (317) K. A. Nguyen, I. Rossi and D. G. Truhlar, *Journal of Chemical Physics*, **1995**, 103, 5522-5530.
- (318) J. C. Corchado, E. L. Coitiño, Y.-Y. Chuang, P. L. Fast and D. G. Truhlar, *The Journal of Physical Chemistry A*, **1998**, 102, 2424-2438.
- (319) Y. Y. Chuang, J. C. Corchado and D. G. Truhlar, *Journal of Physical Chemistry A*, **1999**, 103, 1140-1149.
- (320) R. J. Renka, *SIAM Journal on Scientific and Statistical Computing*, **1987**, 8, 393-415.
- (321) R. J. Renka, *ACM Trans. Math. Softw.*, **1993**, 19, 81-94.
- (322) M. G. Evans and M. Polanyi, *Transactions of the Faraday Society*, **1935**, 31, 0875-0893.
- (323) H. Eyring, *Journal of Chemical Physics*, **1935**, 3, 107-115.
- (324) H. Eyring and A. E. Stearn, *Chemical Reviews*, **1939**, 24, 253-270.
- (325) D. G. Truhlar, *Journal of Physical Organic Chemistry*, **2010**, 23, 660-676.
- (326) B. F. C. Kafsack, N. Rovira-Graells, T. G. Clark, C. Bancells, V. M. Crowley, S. G. Campino, A. E. Williams, L. G. Drought, D. P. Kwiatkowski, D. A. Baker, A. Cortes and M. Llinas, *Nature*, **2014**, 507, 248-252.



## **Appendices**

### **Appendix A: Others Scientific Contributions Related with the Doctoral Thesis**

#### **Poster Communications**

1. *QM/MM studies of mechanism of inhibition of cruzain by peptidyl halomethyl ketones.* Kemel Arafet, Silvia Ferrer, and Vicent Moliner. 8<sup>th</sup> Congress on Electronic Structure: Principles and Applications (ESPA 2012). 26-29/06/2012. Barcelona (Spain).
2. *Inhibition of Falcipain-2 by the Epoxisuccinate E64: QM/MM Studies of the Mechanism.* Kemel Arafet, Silvia Ferrer, and Vicent Moliner. XXXIX International Conference Theoretical Chemists of Latin Expression (QUITEL 2013). 30/06/2013-05/07/2013. Granada (Spain).
3. *QM/MM studies of mechanism of inhibition of two cysteine proteases: cruzain and falcipain-2.* Kemel Arafet, Silvia Ferrer, and Vicent Moliner. 9<sup>th</sup> Congress on Electronic Structure: Principles and Applications (ESPA 2014). 02-04/07/2014. Badajoz (Spain).
4. *QM/MM studies of the catalytic mechanism of the cysteine proteases: acylation step.* Kemel Arafet, Silvia Ferrer, and Vicent Moliner. 7<sup>th</sup> International Theoretical Biophysics Symposium (TheoBio2015). 08-12/06/2015. Cagliari (Italy).
5. *QM/MM studies of the catalytic mechanism of the cysteine proteases.* Kemel Arafet, Silvia Ferrer, and Vicent Moliner. 10<sup>th</sup> Congress on Electronic Structure: Principles and Applications (ESPA 2016). 28/06/2016-01/07/2016. Castellón (Spain).

**Appendix B: Scientific Contributions not Directly  
Related with the Doctoral Thesis**

**Paper V**

1. Katarzyna Świderek, Kemel Arafet, Amnon Kohen, and Vicent Moliner. *Benchmarking Quantum Mechanics/Molecular Mechanics (QM/MM) Methods on the Thymidylate Synthase-Catalyzed Hydride Transfer. Journal of Chemical Theory and Computation.* **2017.** 13, 1375-1388. DOI: 10.1021/acs.jctc.6b01032.



