Universitat de Girona

COMPUTATIONAL STUDIES OF EPOXIDE HYDROLASE-CATALYZED RING-OPENING REACTIONS

Eila Serrano Hervás

Universitat de Girona

DOCTORAL THESIS:

COMPUTATIONAL STUDIES OF EPOXIDE HYDROLASE-CATALYZED RING-OPENING REACTIONS

Eila Serrano Hervás

2022

Universitat de Girona

DOCTORAL THESIS:

COMPUTATIONAL STUDIES OF EPOXIDE HYDROLASE-CATALYZED RING-OPENING REACTIONS

Eila Serrano Hervás

2022

DOCTORAL PROGRAMME IN CHEMISTRY

Supervised by:
Prof. Dr. Sílvia Osuna Oliveras
Dr. Marc Garcia Borràs

Tutor:
Prof. Dr. Marcel Swart

Presented to obtain the doctoral degree from the University of Girona

List of Publications

This thesis is presented as a compendium of publications.

Published articles included in this thesis:

1. Serrano-Hervás, E.; Garcia-Borràs, M.; Osuna, S. "Exploring the origins of selectivity in soluble epoxide hydrolase from *Bacillus megaterium*", *Org. Biomol. Chem.*, **2017**, 15, 8827-8835. [Chemistry, Organic, 3.876, Q1]

2. Serrano-Hervás, E.; Casadevall, G.; Garcia-Borràs, M.; Feixas, F.; Osuna, S. "Epoxide Hydrolase Conformational Heterogeneity for the Resolution of Bulky Pharmacologically Relevant Epoxide Substrates", *Chem. Eur. J.*, **2018**, 24, 12254. [Chemistry, Multidisciplinary, 5.236, Q1]

List of Abbreviations

Abbreviation	**Description**
AMBER	Assisted model building with energy refinement
anti-HIV	Anti-human immunodeficiency virus
*Bm*EH	*Bacillus megaterium* epoxide hydrolase
CM	Cluster model
CVs	Collective variables
DE	Directed evolution
DFT	Density functional theory
E	Energy
EHs	Epoxide hydrolases
FEL	Free energy landscape
G	Gibbs free energy
H	Hamiltonian
Int1	Alkyl-enzyme intermediate
Int2	Tetrahedral intermediate
kcal	Kilocalorie
LCAO	Linear combination of atomic orbitals
LJ	Lennard-Jones
MD	Molecular dynamics
MM	Molecular mechanics
MO	Molecular orbitals
NACs	Near attack conformations
NCIs	Non-covalent interactions
NGE	Naphthyl glycidyl ether

NMR	Nuclear magnetic resonance
NPT	Isobaric-isothermal ensemble
NVE	Microcanonical ensemble
NVT	Canonical ensemble
ρ	Momentum
PBC	Periodic boundary conditions
PCA	Principal component analysis
PDB	Protein data bank
PGE	Phenyl glycidyl ether
p-NSO	*para*-nitrostyrene oxide
r	Coordinates
QM	Quantum mechanics
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
TICA	Time-structure independent component analysis
TS	Transition state
TST	Transition state theory
U	Potential energy surface
WT	Wild-type enzyme
α	Alpha
β	Beta
Å	Angström
ε	Dielectric constant

Universitat de Girona

Prof. Dr. Sílvia Osuna Oliveras and Dr. Marc Garcia Borràs, of University of Girona,

WE DECLARE:

That the thesis entitled "Computational studies of epoxide hydrolase-catalyzed ring-opening reactions", presented to obtain a doctoral degree, has been carried and completed by Mrs. Eila Serrano Hervás under our supervision and that meets the requirements to opt for an International Doctorate.

For all intents and purposes, we hereby sign this document.

Signature

Prof. Dr. Sílvia Osuna Oliveras Dr. Marc Garcia Borràs

Girona, May 21st, 2022

a "las yayas" y "al abuelo",
a "la mama", "al papa", "al Eloy",
a mis personas favoritas,
a mi media naranja,
y a mi medio limón.

Acknowledgements

The studies included in the present thesis would not have been possible without the following financial support: The *Generalitat* (Government of Catalonia) for a PhD grant for the recruitment of new research staff (2019FI_B2 00144), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001). I also thank the Federation of European Biochemical Societies (FEBS) for a FEBS Short-Term Fellowship, and the Institute for Biochemistry and Technical Biochemistry (IBTB) at the University of Stuttgart, Germany, for hosting me. I would like to also thank the University of Girona (UdG), and in particular the members of the academic committee for all the technical and personal support provided, as well as the Institute of Computational Chemistry and Catalysis (IQCC) for the material and all facilities provided.

No és el primer cop que em poso a escriure els agraïments de la tesi. De fet, no se quantes versions n'hauré fet, però cadascuna em servia per intentar veure més enllà durant uns segons. M'aportaven perspectiva. No crec que funcionés, sincerament, però sigui com sigui, avui som dia 20 de Maig del 2022 i, per impossible que em semblés i encara em sembli, només queda aquesta part de la tesi per finalitzar-la. Pensava que seria la part més fàcil, i típic i tòpic, ara no tinc paraules. Lògicament, arribar fins aquí no hagués estat possible sense el suport (incondicional) de moltes persones. Així que només espero que amb les paraules "que sigui que em surtin", el meu agraïment us arribi de tot cor. Segurament jo tampoc hi seria a aquí, ara mateix, així que gràcies també per això, sobretot per aquesta última part.

El record que tinc dels inicis de la tesi es aquell mes d'agost just acabar el grau, on el Parc era un lloc on hi havia estat cops comptats. No se si hagués acabat a allà sense la insistència d'en **Pedro**, així com tampoc si hagués mai descobert abans pinzellades de la química computacional. Aún no sé el qué, pero gracias por tu tiempo durante "el casal de

verano", por tu tiempo explicando y garabateando conceptos como la escalera de Jacob que aún tengo en alguna libreta. También por todas tus bromas, las que entendí y seguramente las muchas que no entendí...

Tornant al mes d'agost... Començaré amb un gràcies a l'**Adrià** i a en **Ouissam**, per l'energia que transmetien, la seva filosofia, estones al dinar... ja em queda lluny, i el meu cap ha suprimit molts moments, però era d'agrair trobar-te sempre a primera hora del despatx a l'Adrià, amb frases com "relaja la raja" que m'ajudaven a prendre-m'ho tot una mica menys "enserio", o tenir el moment del té o cafè als matins abans de començar la feina. Arribats a aquí, tampoc em puc ni voldria deixar-me a la **Vero**, les converses dels migdies, les estones de pausa, o saber que podies anar al despatx i que et rebria sempre que pogués. En aquell llavors potser tenia molt a dir però el cert es que deia poc, i el record que tinc és que ella em va aportar petits moments de confiança per començar a compartir. Gràcies per aquest espai i també per tot el que vaig anar aprenent de tu, tant de ciència com "de la vida". També vull agrair en **Dani**, la **Carme**, i en **Quim**. De tots tres en vaig aprendre, i malgrat "era cosa de feina", et feien sentir atesa en qualsevol moment, ja fossin problemes tècnics, suport amb les beques, o ajudes en moments més personals. Aquesta etapa tampoc hauria estat el mateix sense ells i ella. Així que malgrat em vaig veure apurada en molts moments durant el Girona Seminar, alhora ben contenta encara de que formar part en aquell llavors de "l'equip d'organització" em donés l'oportunitat de poder passar temps amb vosaltres. Ah! I passi el temps que passi... sempre ens quedaran els dònuts!

A mesura que passava el temps "la gent de sempre" marxava, i persones noves s'incorporaven al grup. Aquí potser arriba el moment d'agrair a en Guillem, en Christian, en Miquel, en Lorenzo, en Javi, la **Clàudia** i la **Cristina**. Gràcies a en **Guillem** per haver estat un pilar durant l'etapa del doctorat, pels malentesos i pels "benentesos". Sempre ric sola al recordar els teus inicis al Parc... sobretot al pensar que jo "et supervisava". Com tots sabem, res més lluny de la realitat! Gràcies per haver format part del "team EHs", per la teva dedicació a la feina, fent codi, fent més preguntes del compte, dedicar temps a fer-me entendre (amb mil tecnicismes que feia impossible entendre't jaja), pels debats i el voler trobar noves formes de fer recerca... Com ja saps, també en l'àmbit personal, per haver-te mostrat perdut i haver-me trobat i parlat ben clar quan era jo la que no trobava el camí. **Curado**, Curado... cuantos ratos de implosión hemos pasado! Fue una alegría cuando llegaste para formar

parte del grupo y creo que lo único que evitó que bajaramos a tomar aire fue el dichoso covid... Gracias por todas las charlas, por todos esos momentos de confianza, de compartir, de hablar de cosas mundanas, de dar vueltas y vueltas a "qué está pasando" y "por qué no me funciona el cálculo", todo tu tiempo también dedicado a hacerme entender algo de código aunque pareciera misión imposible, y sobre todo por darme cobijo. Por compartir conmigo (¡veraaaano!) y por no haberme dado nunca la espalda. Y aquí también tengo que añadir a **Miquel**, porque aquel fin de semana en Andorra fue de los mejores momentos de esta etapa. Me sentí "parte de algo" cuando hacía tiempo que me sentía "fuera de todo" y "sin sentido". A tu **Miquel**, gràcies també per les hores i el temps dedicat amb "la meva feina". Per tenir-te allà al final del despatx, amb un teclat "silenciós" i amb una cadira que ens havia fet servei a les tantes hores de la nit més d'un cop, i per tot el "chiringuito" que tenies muntat per donar-li vida al despatx. Potser ens vam descobrir tard, a vegades "jo no se on era", però ben agraïda de que fos "més tard que mai", de poder-te fer abraçades sabent que no les suportes, tot un privilegi, (sushiii!) Així com també per les classes de natació abans del covid, i també aquells "tornars", tot i que comptats, cap a casa per la vorera del riu. Gràcies a tots tres per haver-me ensenyat tantes coses i pels moments compartits.

Crec que ja arriba el torn d'en **Lorenzo**, "il mio italiano preferito!" que ja sap català "gràcies" a mi i la Vero... Right now I am not sure in which language we speak to each other, Catalan, Spanish, English... but in any case, I will do that in English, although I might not be super good at expressing my feelings or thoughts in that language... Thank you for all our conversations, for having the pleasure to listen to you in more personal situations, for letting me be there and counting on me, and for being there for me as well. I am really happy to have known you, for some of our walks around Girona, for quick conversations during the "espresso-time"... or for all hugs you have given me. Thank you for teaching me the "bad italian words", for all the funny parts in the middle of all the stress that implies doing a Ph.D., and also for all your warm support. You know you can count on me, that I am here and that I will be happy to be there... with you and "your little mozzarella".

Tampoc m'oblido d'en **Javi**, "ay, mi postdoc", qui ha dedicat també molt temps a gestionar i fer-me entendre conceptes que em semblaven impossibles de pair. No només en l'àmbit de feina, també agrair-te tot el suport personal, per ser un exemple per mi o un "model a seguir", per la teva forma de veure la vida i la manera de fer. Per tots els cafès,

els plats dels migdies o els entrepans del matí. I pel temps compartit a Washington D.C., passejant tant entre posters ("poster winner") com pels carrers d'allà, entre botigues i cotxes gegants. Encara em fa riure quan recordo el "Javier Iglesias Fernándes". Gràcies també per haver estat una de les persones amb les que he sentit que em podia expressar, fossin o no pors sense sentit i molt allunyades de la realitat. Per riure, per fer les coses senzilles i per dir també les coses "com son". També un gràcies ben gran a la Laura, i al "teu súper peque", que sempre son motiu d'alegria.

Continuaré amb els supervisors i directors de la tesi, la **Sílvia** i en **Marc**. No només m'heu ensenyat a fer una bona recerca, a qüestionar les dades o a fer-me millors preguntes, sinó d'alguna forma o una altra i en diferents moments, sempre que heu pogut heu estat un suport. Si me'n vaig als primers anys de la tesi, en Marc era a Los Ángeles, i tot i així jo el sentia com un refugi ben proper, algú de qui aprendre, un model i una guia de com anar fent dins el món de la recerca. Per tot això, gràcies Marc! També perquè vas ser algú a qui m'era fàcil escriure mentre "tot aquí s'ensorrava". A part d'això, agrair-te sobretot també aquests últims mesos de suport, d'e-mails, de reunions, i per tenir les ganes de compartir amb mi també les bones notícies. Amb la **Sílvia**, doncs sempre sentiré que em falten paraules, com bé ja saps... Així que gràcies per oferir-me aquesta posició en el seu dia, per embarcar-te amb mi en aquest viatge, per l'oportunitat d'endinsar-me dins el món de la recerca i poder-ho fer per la vessant computacional i al teu costat. Gràcies per haver estat la meva supervisora i directora de tesi, i sobretot pel lligam que s'ha creat durant aquesta etapa més enllà de cap etiqueta. També per la dedicació i la motivació del principi d'aquest viatge, amb cadascun de nosaltres, gràcies per aquella energia que, a mi almenys, m'aportava motius per continuar endavant. Acabo amb un gràcies pel "més enllà de la feina", com les setmana d'esquí, els camins de tornar i anar al Parc amb cotxe, per la complicitat i la confiança, i en definitiva, per haver cregut en mi quan jo no ho feia. Gràcies per aquesta etapa compartida, i per seguir ara també compartint part dels nostres dies. Gràcies també a en Narcís, en Pol i l'Ona. I arribats a aquest punt, tampoc em voldria deixar d'agrair de nou als membres de la Comissió Acadèmica, i sobretot sense dir un gràcies explícit a la **Lidia**, per tota l'assistència i suport, a més pel fet de creure que podia arribar aquest dia.

I will finish that part by writing a huge "thank you" to all people who made me feel at home when I was in Germany during my stay. I did not realize that it was a difficult moment

for me, and although in the beginning I was feeling home-sick, the truth is that when the moment arrived, I did not want really to come back. At the very beginning maybe it was not like that, but for sure in the end, I felt that I belong somewhere and that I was not so useless. I really liked being in the lab (except when we had to make competent cells inside the fridge, bad joke...), especially at the end when I was feeling confident enough to make all the tasks. I still feel good when I remember that experience, how the sound of the German language changed during that period, and how I was happy with taking the "S-Bahn" and walking around the center, although the reality was that the transport was quite a mess. A special thanks to **Lars** and **Bettina**, such experience would not have been possible without both of you. Also a warm thank you to the rest of Master's and Ph.D. Students, technical staff, and postdoctoral researchers that were in the group during that period. I really felt at home.

I finalment arriben els meus suports incondicionals... les meves persones preferides. Gràcies a l'**Iteng**, la **Cris** i la **Nerea**. Gràcies a la **Gemma** i a la **Pat**. Gràcies a l'**Irene**. Gràcies a l'**Òscar**. Gracias "al **Carlos**". Gracias al "Club de Fans de Eduard", **María**, **Pablo** y **Pamela** "la vecina", por haber cuidado tanto de Eduard como de mí durante nuestro tiempo en Holanda, por los momentos compartidos allí, des de los paseos en bicicleta hasta los "brunch", y sobre todo por hacerme sentir cómo parte del "Club" des del primer instante. Gràcies a l'**Edu**, i als seus pares, la **Neus** i en **Carles** "pare", així com també a en **Carles** "fill" i la resta de la **família Masferrer i Rius**. Gracias "a la **mama**", "al **papa**" y "al **Eloy**", a "la **Tam**" i a en "**Martinet**" i a l'**Irina** i les seves abraçades plenes de força. Gracias a "las **yayas** Pepi y Pepa", a "la **tía Tere**" y "la **tía Susana**", "al **abuelo**", i els que ja no hi son o han marxat. En menys o més grau vosaltres sou qui sabeu "tot i res" del que ha passat amb mi durant aquest temps. Qui heu sabut ser al meu costat quan ni tan sols volia acompanyants, els qui heu donat llum i us heu atrevit a mirar la meva foscor, tenint la paciència de fer-me costat mentre no veia el camí. Sense necessitat d'entendre i mostrant-me que també puc ser una bona acompanyant. Definitivament sense la vostra estima i coratge aquesta tesi no hagués estat possible, així com tampoc aquest moment d'ara mateix. **Gràcies!**

Eila

Table of Contents

List of Figures

Summary of the thesis

Billions of years of evolution have made enzymes superb catalysts capable of accelerating reactions by many orders of magnitude (e.g. 10^{11} to 10^{16}), but also to make this task with exquisite selectivities at physiological conditions. This is because keeping life in any cell requires hundreds of chemical reactions to take place at any given moment, in a highly regulated manner, and acting toward particular substances depending on the cells requirements. In a general view, rate accelerations are achieved by lowering the activation energy barriers of reactions, while selectivity mainly arises from the confinement of substrates into the active sites of enzymes. Among all known catalysts, enzymes possess a few important advantages over chemical catalysts that might have promoted their use in synthetic processes. In addition to all their synthetic advantages, such as their high efficiencies, specificities, as well as selectivities, enzymes are also evolvable, biodegradable, non-toxic, and their high selectivities and efficiencies reduce significantly the number of work-up steps, thereby providing compounds in higher yields.

Chiral chemicals are commonly identified as relevant precursors for the synthesis of a huge series of compounds, such as pharmaceuticals, agrochemicals and fine chemicals. In some cases, these optically active compounds come in a form of racemic mixtures, which represent a cheap and easy available starting material for the organic synthesis of bifunctional products. For instance, various β-adrenergic receptor agonists, anti-cancer agents, and anti-human immunodeficiency virus (HIV) drugs are synthesized from epoxide-containing substrates. Enantiopure epoxides and diols can be prepared based on a variety of biotransformation reactions, such as the selective hydrolysis by epoxide hydrolases (EHs). Some EHs might show enantioselectivity, which allows their application in the separation of racemic mixtures of epoxide substrates to provide access to enantiopure epoxide and diols, while some others might display regioselectivity, and therefore, a preference over one particular carbon of the

epoxide ring moiety. In a nutshell, EHs catalyze the opening of epoxide molecules by using only a water molecule as a cosubstrate, thereby producing vicinal diol compounds. In Nature, EHs are ubiquitously found in many tissues and in many different living organisms, carrying out many different functions depending on its origin. In mammals, for instance, EHs are known to metabolize the cleavage of molecules that are potentially toxic and carcinogenic to less harzadous compounds. This might have been one of the reasons these enzymes have been traditionally viewed as *detoxicating* machines, but also because the substrates of EHs are structurally very diverse, representing a broad range of metabolites and xenobiotics. Interestingly, the substrate specificity of individual EHs appears to be diverse as well, being in many cases broad, but occasionally limited to a few available epoxide compounds. In the present thesis, however, the epoxide hydrolase from *Bacillus megaterium* has been the one that caught our attention, but specially due to its potential use as biocatalysts in chiral synthesis. In particular, due to *Bm*EH's ability to preferentially bind and react with (*R*)-epoxide enantiomer of phenyl glycidyl ether derivatives, in contrast to many other related EHs.

In general, it is widely accepted that enzymes commonly catalyze chemical reactions with a high degree of efficiency and selectivity due to their sophisticated structures. On this basis, the present thesis starts with **Chapter 1** by introducing fundamental ideas related to the relationship between enzyme structures and function, how these macromolecules are able to achieve such high levels of acceleration and selectivities from the reduced view of enzymatic active sites to the significant role of their dynamic behaviour. In **Chapter 2**, the strategies and the main concepts of the theory behind the tools applied in the studies included in this thesis are briefly summarized. In the following **Chapter 3** the main objectives of the present thesis are provided, while **Chapters 4** and **5** include our mechanistic and conformational dynamics studies of *Bm*EH enzyme, respectively. Finally, in **Chapter 6** results previously presented on each article are briefly discussed, and in **Chapter 7** the main conclusions drawn from this thesis are provided.

Resum de la tesi

Milers de milions d'anys d'evolució han fet que els enzims siguin excel·lents catalitzadors capaços d'accelerar reaccions en molts ordres de magnitud, com per exemple, ordres des de 10^{11} fins a 10^{16}. Tot i així, aquestes macromolécules han captat també la nostra atenció perquè duen a terme aquesta tasca amb selectivitats altament elevades i en condicions fisiològiques. Això es deu al fet que mantenir la vida a qualsevol cèl·lula requereix que centenars de reaccions químiques tinguin lloc en un moment donat, d'una manera altament regulada, i actuant cap a substàncies particulars en funció de les necessitats d'aquestes mateixes cèl·lules. En una visió general, l'augment de la velocitat de les reaccions s'aconsegueix disminuint les barreres d'activació d'aquestes reaccions, mentre que la selectivitat sorgeix principalment del confinament dels substrats als centres actius dels enzims. Entre tots els catalitzadors coneguts, els enzims tenen alguns avantatges relleants en comparació als catalitzadors químics, propietats inherents que han promogut el seu ús en processos industrials i sintètics. A més de les seves altes eficiències, especificitats i selectivitats, els enzims també són modelables, és a dir, es poden evolucionar, a part de biodegradables, no contaminen, i les seves elevades selectivitats i eficiències redueixen significativament el nombre d'etapes de purificació, proporcionant així compostos purs amb rendiments més elevats.

Els productes químics quirals s'identifiquen habitualment com a precursors rellevants per a la síntesi d'una gran sèrie de compostos, com ara productes farmacèutics, agroquímics i productes de química fina. En alguns casos, aquests compostos òpticament actius es presenten en forma de mescles racèmiques, que representen un material de partida barat i de fàcil disposició per a la síntesi orgànica de productes bifuncionals. Per exemple, a partir de substrats que contenen epòxids es solen sintetitzar diversos agonistes de receptors β-adrenèrgics, agents anticancerígens i fàrmacs contra el virus de inmunodeficiència humana (VIH). Els epòxids i diols enantiopurs es poden preparar a partir d'una varietat de reaccions

de biotransformació, com la hidròlisi selectiva mitjançant enzims epòxid hidrolasa (EHs). Algunes EHs poden mostrar enantioselectivitat, la qual cosa permet la seva aplicació en la separació de mescles racèmiques d'epòxids per a la producció d'epòxids i diols enantiopurs, mentre que algunes altres poden mostrar regioselectivitat i, per tant, una preferència sobre un carboni particular de l'anell de l'epòxid. En poques paraules, les EHs catalitzen l'obertura de molècules d'epòxid utilitzant només una molècula d'aigua com a cosubstrat, produint així compostos diols vicinals. A la natura, les EHs es troben de manera ubiqua en molts teixits i en molts organismes vius diferents, duent a terme moltes funcions diferents segons el seu origen. En els mamífers, per exemple, se sap que les EHs metabolitzen la divisió de molècules que són potencialment tòxiques i cancerígenes a compostos menys perillosos. Aquesta podria haver estat una de les raons per les quals aquests enzims s'han vist tradicionalment com a *màquines desintoxicants*, però també degut a que els substrats de les EHs son estructuralment molt diversos, representant un ampli rang de metabòlits i compostos xenobiotics. De forma similar, l'especificitat de les EHs és també en molts casos àmplia, però alhora limitada a un petit rang de substrats. En la present tesi, però, l'epòxid hidrolasa del *Bacillus megaterium* ha estat l'enzim que ha captivat la nostra atenció, sobretot pel seu potencial ús com a biocatalitzador en la síntesi de compostos quirals, i en particular per la seva capacitat d'unir-se i reaccionar preferentment amb l'enantiòmer R dels derivats de l'epòxid fenil glicidil èter, a diferència de molts altres EHs similars.

En general, és àmpliament acceptat que els enzims catalitzen habitualment reaccions químiques amb un alt grau d'eficiència i selectivitat a causa de les seves estructures sofisticades. En base a això, la present tesi comença amb el **Capítol 1** introduint idees fonamentals relacionades amb la relació entre estructures i funció enzimàtica, com aquestes macromolècules són capaces d'aconseguir nivells tan elevats d'acceleració i selectivitat des de la visió reduïda dels llocs actius enzimàtics fins als significatiu paper del seu comportament dinàmic. En el **Capítol 2**, es resumeixen breument les estratègies i els principals conceptes de la teoria darrere de les eines aplicades en els estudis inclosos en aquesta tesi. En el següent **Capítol 3** es proporcionen els principals objectius de la present tesi, mentre que els **Capítols 4** i **5** inclouen els nostres estudis del mecanisme de reacció i la dinàmica conformacional de l'enzim *Bm*EH, respectivament. Finalment, en el **Capítol 6** es comenten breument els resultats presentats prèviament sobre cada article, i en el **Capítol 7** es presenten les principals conclusions derivades d'aquesta tesi.

Resumen de la tesis

Miles de millones de años de evolución han hecho que las enzimas sean excelentes catalizadores capaces de acelerar reacciones en muchos órdenes de magnitud, como por ejemplo, órdenes desde 10^{11} hasta 10^{16}. Sin embargo, estas macromoléculas han captado también nuestra atención porque llevan a cabo esta tarea con selectividades altamente elevadas y en condiciones fisiológicas. Esto se debe a que mantener la vida en cualquier célula requiere que cientos de reacciones químicas tengan lugar en un momento dado, de forma altamente regulada, y actuando hacia sustancias particulares en función de las necesidades de estas mismas células. En una visión general, el aumento de la velocidad de las reacciones se logra disminuyendo las barreras de activación de estas reacciones, mientras que la selectividad surge principalmente del confinamiento de los sustratos en los centros activos de las enzimas. Entre todos los catalizadores conocidos, las enzimas tienen algunas ventajas relevantes en comparación con los catalizadores químicos, propiedades inherentes que han promovido su uso en procesos industriales y sintéticos. Además de sus altas eficiencias, especificidades y selectividades, las enzimas también son moldeables, es decir, se pueden evolucionar, aparte de ser biodegradables, no contaminan, y sus elevadas selectividades y eficiencias reducen significativamente el número de etapas de purificación, proporcionando así compuestos puros y con rendimientos más elevados.

Los productos químicos quirales se identifican habitualmente como precursores relevantes para la síntesis de una gran serie de compuestos tales como productos farmacéuticos, agroquímicos y productos de química fina. En algunos casos, estos compuestos ópticamente activos se presentan en forma de mezclas racémicas, que representan un material de partida barato y de fácil disposición para la síntesis orgánica de productos bifuncionales. Por ejemplo, a partir de sustratos que contienen epóxidos se suelen sintetizar varios agonistas de receptores β-adrenérgicos, agentes anticancerígenos y fármacos contra el virus de inmunodeficiencia

humana (VIH). Los epóxidos y dioles enantiopuros pueden prepararse a partir de una variedad de reacciones de biotransformación, como la hidrólisis selectiva mediante enzimas epóxido hidrolasa (EHs). Algunas EHs pueden mostrar enantioselectividad, lo que permite su aplicación en la separación de mezclas racémicas de epóxidos para la producción de epóxidos y dioles enantiopuros, mientras que otras pueden mostrar regioselectividad y, por tanto, una preferencia sobre un carbono particular del anillo del epóxido. En pocas palabras, las EHs catalizan la apertura de moléculas de epóxido utilizando sólo una molécula de agua como cosustrato, produciendo así compuestos dioles vicinales. En la naturaleza, las EHs se encuentran de forma ubicua en muchos tejidos y en muchos organismos vivos diferentes, llevando a cabo muchas funciones diferentes según su origen. En los mamíferos, por ejemplo, se sabe que las EHs metabolizan la división de moléculas que son potencialmente tóxicas y cancerígenas en compuestos menos peligrosos. Ésta podría haber sido una de las razones por las que estas enzimas se han visto tradicionalmente como *máquinas desintoxicantes*, pero también debido a que los sustratos de las EHs son estructuralmente muy diversos, representando un amplio rango de metabolitos y compuestos xenobióticos. De forma similar, la especificidad de las EH es también en muchos casos amplia, pero a la vez limitada a un pequeño rango de sustratos. En la presente tesis, sin embargo, la epóxido hidrolasa del *Bacillus megaterium* ha sido la enzima que ha cautivado nuestra atención, sobre todo por su potencial uso como biocatalizador en la síntesis de compuestos quirales, y en particular por su capacidad de unirse y reaccionar preferentemente con el enantiómero R de los derivados del epóxido fenil glicidil éter, a diferencia de muchos otros EHs similares.

Por lo general, es ampliamente aceptado que las enzimas catalizan habitualmente reacciones químicas con un alto grado de eficiencia y selectividad debido a sus sofisticadas estructuras. En base a ello, la presente tesis comienza con el **Capítulo 1** introduciendo ideas fundamentales relacionadas con la relación entre estructura y función enzimática, cómo estas macromoléculas son capaces de alcanzar niveles tan elevados de aceleración y selectividad desde la visión reducida de los sitios activos enzimáticos hasta el significativo papel de su comportamiento dinámico. En el **Capítulo 2**, se resumen brevemente las estrategias y los principales conceptos de la teoría detrás de las herramientas aplicadas en los estudios incluidos en esta tesis. En el siguiente **Capítulo 3** se proporcionan los principales objetivos de la presente tesis, mientras que **Capítulos 4** y **5** incluyen nuestros estudios del mecanismo de reacción y la dinámica conformacional de la enzima *Bm*EH, respectivamente. Por último, en el **Capítulo 6** se

comentan brevemente los resultados presentados previamente sobre cada artículo, y en el **Capítulo 7** se presentan las principales conclusiones derivadas de esta tesis.

Chapter 1

Introduction

There would be no way of keeping the genetic information that makes us without the existence of a biochemical compound known as DNA [1]. Likewise, there would be no way of doing everything that we know as our daily life without taking this stored information to produce the workhorses of our cells. That is, proteins, which are indeed the ones making possible all the daily actions that we might take for granted.

These biomolecules can serve us all sort of different functions, generally, depending upon which shape proteins will take and how they will interact with their surroundings. For instance, proteins like antibodies that form part of our immune system can prevent us from being sick, while we might appreciate that some others let us enjoy our fancy *oatly* cappuccino every morning by digesting its sugars [2]. The truth is that the amount of work that might be actually done by proteins to function as humans, or as any other living organism on Earth, might be inconceivable. And luckily, we do not need to know all of this to exist.

One of the roles of proteins that have specially caught our attention is the ability to speed up (*catalyze*) biochemical reactions [3]. Such type of proteins are known as *enzymes* or biological catalysts, which have naturally evolved to perform efficiently under the mild

conditions required to preserve the integrity of biological systems and support cell function [4]. Each one of the chemical reactions that comprise the metabolism of cells requires to be catalyzed by one specific enzyme [5]. For instance, hemoglobin is an inert oxygen carrier protein essential for respiration, pancreatic enzymes such as proteases, lipases and amylases are key for digesting food, while others are required for muscle and nerve function. As we can see from some of these examples, enzymes typically have *trivial* names that in most cases indicate the reaction they catalyze (generally with the suffix *-ase*, such as oxidase, dehydrogenase, and carboxylase, or with the suffix *-in*, such as trypsin). In some other cases, their names also indicate the substrate on which enzymes act (e.g. *alcohol* dehydrogenase or *glucose* oxidase), while in some other few cases little information about the substrate, the product or the reaction involved can be derived from it (e.g. catalase). Due to their wide range of activities, nowadays there is a more systematic and consistent classification scheme: the Enzyme Commission (EC) numbers [6], whose first digit indicates one of the six major classes of reactions that enzymes catalyze (i.e. oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases).

When we define enzymes as *biocatalysts*, however, this term does not refer to these molecular machines inside our cells, but to the use of enzymes as "process catalysts" under non-native or *in vitro* reaction conditions. In fact, enzymes can be extracted from cells and, as any catalyst, then used to speed up a wide range of industrially important processes without being consumed during the chemical transformation. Some of these applications include from everyday uses (e.g. proteases added to detergents) to those applications at industrial scale in food, agriculture, polymer synthesis, and/or organic synthesis of fine and pharmaceutical compounds [7]. In general, most of these processes initially exploited the chemistry found in Nature. However, for many industrial applications there is no natural counterpart available, as these molecules have naturally evolved for a specific biological purpose. Nevertheless, the field of biocatalysis has greatly advanced from harnessing natural enzymes to evolve them by using protein engineering techniques [8]. And remarkably, many novel biocatalysts with tailor-made functions have been nowadays successfully engineered [9].

1.1 Enzymes as Biocatalysts

The application of enzymes as biocatalysts offers some advantages over chemocatalysts, and during the latter half of the 20th century this ability of enzymes to catalyze chemical transformations started to be viewed as a potential strategy to produce non-natural or synthetic products at the industrial scale. Despite their huge potential, most of the biocatalysts employed are commonly modified to suite the different industrial requirements. Nevertheless, in contrast to the numberless chemical transformations that chemical catalysts make possible, natural enzymes are superb catalysts, but typically restricted to six major types of enzymatic reactions and working at the ambient conditions [10, 11]. In addition, enzymes are inherently specific toward *a small set of substrates*, which means that in practice, they are presumably limited to a particular substrate scope. And therefore, if we are interested on using them toward *non-natural* substrates, such as any starting material of interest, the substrate scope of the enzyme needs to be expanded.

A common way to overcome some of these limitations has been to engineer new enzymes. In other words, enzymes that have been modified by protein evolution techniques. The area that focuses on the generation and optimization of new biocatalysts by altering the properties of wild-type enzymes is indeed widely known as the *enzyme design field* [12]. Up to date, many different classifications can be probably found regarding the enzyme design approaches available. Regardless the information available of the system under study in advance, and either if it is carried out by means of experiments, computations, or a combination of both, all approaches have been typically grouped depending on how the following key points are addressed: (i) first, the selection of amino acid positions subjected to mutagenesis, (ii) second, the generation of the new variants, and finally, (iii) the evaluation of whether the target property has been enhanced [13].

Generally two assumptions need to be made when considering the molecular evolution of functional proteins. That is, the number of possible amino acid sequences, which is astronomically large, and the fact that only a small portion of this protein sequence space is explored during this evolutionary process [14]. As we cannot possibly cover all of it, considering in advance how many changes we want to introduce is quite important, as it will determine how many new enzyme variants from the original wild-type or parent enzyme will

be produced. For instance, if we already have information from X-ray crystallography, we can try to solve this problem on the basis of sequence or structural knowledge by *rationally* targeting active site residues, by choosing the ones responsible for substrate binding, as well as residues that are located at distal positions. In a general view, when the positions and the type of replacement of the amino acids subjected to mutagenesis are guided and selected by prior sequence, structure or functional knowledge, the process is then commonly understood as a rational approach. Interestingly, such amino acid replacements might not only include the 20 most common amino acids found in Nature, but also non-canonical amino acids that might allow us to expand the chemistry found in enzymes even more. Alternatively, if we do not have any information regarding the structure or the amino acid sequence, we can then choose methods that do not require the assistance of that previous knowledge for introducing mutations. In other words, techniques based on directed evolution (DE) strategy, which introduces random mutations into a defined amino acid sequence. In this latter case, you typically end up with a huge library of mutants that are the result of iterating many cycles of random mutagenesis, and suitable high-throughput screening and selection methods are typically required to screen and select the improved variants. Therefore, after exchanging the amino acids, likewise Darwinian evolution, the fittest enzyme variant, such as the one with higher stability or capable of catalyzing the novel target reaction, is identified and artificially selected to be used as the new starting point of the next round of mutagenesis. In other words, replacements are distributed over the whole enzyme sequence. In spite of being a powerful and versatile tool for boosting enzyme activity, worth to mention is that most enzymes that have been dramatically evolved through DE have also included rational approaches during the overall enzyme design process. In a similar way, in spite of the fact that de novo enzymes have been successfully designed by computational means, the application of directed evolution is typically required to enhance the initial low catalytic efficiencies. Overall, thanks to the biotechnological advances, the available toolbox of biocatalysts accessible for organic synthesis is now routinely expanded, both in the laboratory and *in-silico* [10, 15]. Given some of the inherent properties of enzymes, their use in chemical synthesis has become an increasingly popular and attractive approach toward the development of more sustainable and benign processes [16], specially when the specificity of the reaction is a major issue [17–19]. On this direction, enzymes represent an alternative to avoid the formation of by-products, thereby providing compounds in high yields and purity. After all, the reason to produce pure *chiral* pharmaceuticals that might not cause undesired side-effects is biological [20], since the target

receptor of the drug can recognize and bind "only" one particular *enantiomer* of this chiral molecule.

Chiral molecules are compounds in which one or more carbon atoms have four non-identical substituents (frequently termed *stereocenter* or asymmetric centre) [21]. As a chiral molecule, this compound has the ability to exist in two non-superimposable forms that happen to be mirror images of one another, which is how we commonly define the term of *enantiomers* [22]. So, despite sharing identical chemical groups, likewise our right and left hand, these compounds are apparently "the same", but they are indeed different (each enantiomer rotate the plane of *polarized light* in equal but opposite directions) [23]. This is of great relevance, and particularly in the environment of living systems where specific structure-activity relationships may be required for effect, since the physicochemical and biochemical properties of *racemic* mixtures (equal amounts of the two *enantiomer* forms of a given chiral molecule) and individual stereoisomers can differ significantly. In other words, enantiomers have different properties in *chiral* environments, such as in the presence of enzymes, while racemic mixtures are optically inactive, although show chiral property. A world famous example that frame the relevance of producing enantiopure pharmaceutical products is the tragedy of birth defects caused by the drug Thalidomide, sold under the brand name Contergan, a chiral molecule that was used in the late 1950s and early 1960s for the treatment of nausea in pregnant women [24]. While it was found that the (R) enantiomer of thalidomide was effective as a sedative molecule able to stop morning sickness in pregnant women, the (S) enantiomer caused fetal damage and congenital malformations (Fig. 1.1).

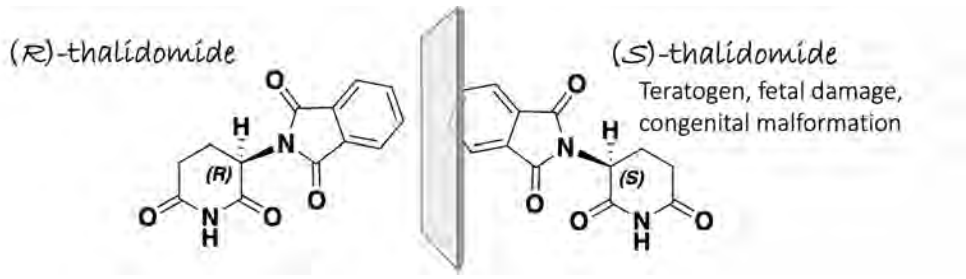


Figure 1.1: Relevance of producing pure chiral molecules in pharmaceuticals.

As opposite enantiomer might lead to harmful side effects, in general, a single enantiomer of a drug is often preferred. Because of this, modern organic chemistry is becoming more and more involved in the discovery, synthesis and testing of new bioactive compounds that

are stereochemically pure [25]. Since the 1990s, several drugs have been the subject of the so-called *chiral switching* [26], that is, the substitution of previously developed racemic mixtures with enantiopure formulations. However, this strategy cannot be applied to the drug Thalidomide, since it has been observed to racemize *in vivo* [27, 28].

In drug development, for instance, much effort has been devoted to exploring efficient approaches for the production of chiral epoxides and diols [29–31]. This is mainly due to the fact that epoxide-containing compounds are ubiquitously found in the environment from both natural and man-made sources [32], such as β-blocker drugs and amino alcohols synthons [33–35]. As shown in Fig. 1.2, an *epoxide* (or oxirane) is a three-membered cyclic ether composed by two carbon atoms and one oxygen atom that has specific reactivity patterns owing to the polarization of the oxygen-carbon bonds and the strain of the ring [36]. Due to their chemical versatility, the synthesis of enantiopure epoxides and their corresponding vicinal diols has been an actively pursued area of research. Indeed, epoxide substrates can be opened by many different type of molecules that we called *nucleophiles*, although in living organisms epoxide hydrolysis seems to be a common route of epoxide transformation [37].

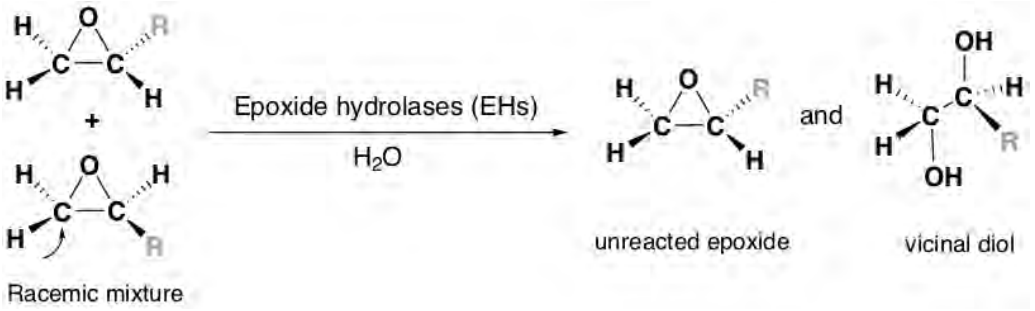


Figure 1.2: General illustration of kinetic resolution of epoxide racemic mixtures in the presence of water catalyzed by epoxide hydrolases (EHs). In this general scheme, the R group refers to any chemical group (e.g. aryl). Ideally, in a kinetic resolution, one of the enantiomers remains unreacted and a maximum theoretical yield of 50 cannot be exceeded.

The catalytic addition of water to oxirane by epoxide hydrolases (EHs) enzymes is only one of several ways that cells employ to transform epoxides [38–40]. This activity was originally reported more than 50 years ago, when only EHs from mammalian sources were known [41, 42]. At that point, their application as biocatalysts was hampered by their poor availability, as well as their insufficient catalytic performance (e.g. low turnover rate or poor enantioselectivity). However, after the discovery of *microbial* epoxide hydrolases, their

potential use as biocatalysts for the kinetic resolution of epoxides at industrial scale was significantly increased (Fig. 1.2) [33, 43–46]. This is because microbial epoxide hydrolases are easier to produce in large quantities, but also because catalytic properties have been modulated by protein engineering techniques such as site-directed or random mutagenesis [47–49]. There might still be a constraint on the account of limited EHs being successfully applied on the preparative scale for industrial biotransformations [17, 50]. However, these enzymes attracted the attention of many organic chemists due to some of their inherent features, such as: i) they are ubiquitous in Nature (found in all types of living organisms, from microorganisms to plants and animals), ii) they are cofactor independent (only requires the presence of water), iii) they can also act in the presence of organic solvents (thereby allowing the use of water-insoluble substrates), and iv) depending upon the substrate, EHs can exhibit different regio- and enantioselectivities (the reaction proceeds at one carbon of the ring and preferentially over one mirror image) [51, 52].

The epoxide hydrolase that has received special attention in the present thesis has been the one from *Bacillus megaterium* (*Bm*EH). As the majority of EHs described to date, *Bm*EH belongs to one of the largest structural enzyme families: the α,β-hydrolase (ABH) superfamily, including a broad range of other types of synthetically useful enzymes in addition to epoxide hydrolase as a result of a divergent evolutionary process (e.g. lipases, esterases, amidases, and dehalogenases) [53–56]. Due to their highly diverse catalytic roles, enzymes belonging to this ABH fold have been widely used in the field of biocatalysis, but also in the protein engineering field, since they are easy to handle in the laboratory and their mechanisms are also well-understood [57]. For the case of the *Bm*EH enzyme, which follows a two-step mechanism, our interest arises mainly due to its *unusual* selectivity toward some arene epoxides identified as relevant building blocks of pharmaceuticals [43, 58]. Some *Bm*EH variants have the ability to accept bulkier epoxide synthons of β-adrenergic drugs such as propranolol [59], which is used to control hypertension, myocardial infarctions, and also to control symptoms related with anxiety disorders [60]. As this thesis has only focused on *Bm*EH enzyme, it is worth to mention that most examples included in this general introduction will refer to this particular enzyme. And so, in this first Chapter, fundamental concepts will be provided throughout all Sections, although in combination with more specific details about *Bm*EH.

1.2 What Are Enzymes?

Enzymes are a type of proteins made up of a long chain of different amino acid residues that are assembled and then folded into the *so-called* three-dimensional arrangement. Even though proteins are usually termed as macromolecules, they are indeed microscopic. Nevertheless, when the scientific community was able to isolate them in *considerable* amounts it was possible to confirm, by using X-ray crystallography, that most enzymes are proteins [3, 61]. Since then, X-ray crystallography has become an important tool for characterizing protein structures but also for providing a mechanistic understanding of protein functions. As it is shown by the huge number of structures that have been deposited up to date to the Protein Data Bank (PDB) [62], there has been an enormous experimental effort to determine protein structure [63–66]. And yet, the current available number of crystal structures only represent a small fraction compared to the billions of known protein sequences [67, 68] (the "protein folding problem" [69]).

In Nature, remarkable only a set of 20 canonical amino acids are required to give rise to the huge functional and structural diversity we found in proteins [70]. Amino acid residues are nothing more than organic molecules containing an acidic carboxylic group, a basic amino group, and a unique side-chain varying in terms of size, polarity, and hydrophobicity (commonly defined as the R group). Therefore, all amino acids are similar but different, each one being characterized by a particular set of physical and chemical properties that will end up having an impact on the final protein structure.

1.2.1 Understanding Enzyme Function from Structure

Protein structures can provide invaluable information, both for reasoning about biological processes and for enabling interventions such as structure-based drug development or targeted mutagenesis [71]. Although at first sight the overall structure of proteins might look lumpy and amorphous, these molecules have a highly precise structure [72]. Proteins can be defined as polymers built from a limited number of amino acid building blocks. The simplest level is the **primary structure**, which is defined as long linear chains of amino acid residues linked to its neighbor through a *covalent peptide bond*. That is, between the carboxylic group of one residue and the amino group of the next one [73] (see Fig. 1.3a). Amino acids molecules are similar, except for the side-chain (or R group), and have a central carbon atom typically

referred as Cα that is an *asymmetric* carbon. As we have already seen, for each asymmetric carbon in a molecule there are two possible configurations. In other words, two completely opposite ways of organizing the same chemical groups around this chiral centre. In the case of amino acid residues, there are two plausible enantiomers, namely L and D configurations, although the L-amino acids are only the ones found in naturally-occurring proteins [74].

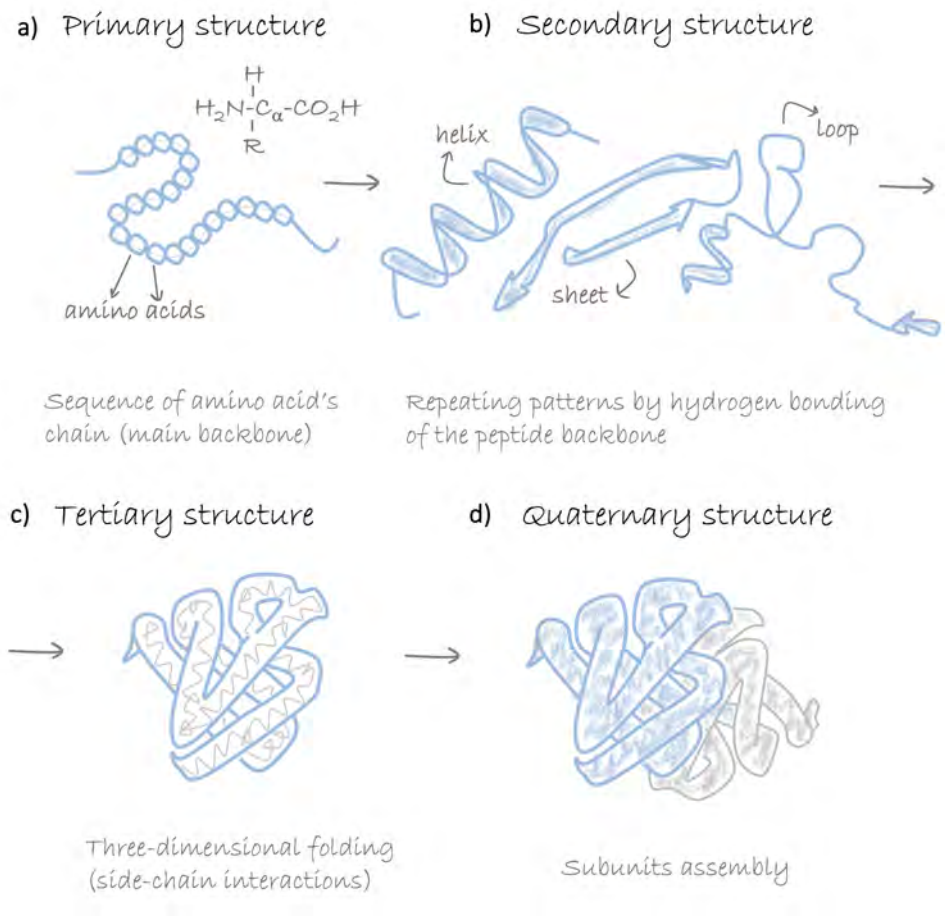


Figure 1.3: The fundamental structure and folding of proteins. a) The protein's amino acid sequence shown schematically as beads on a string. b) Possible folding secondary elements in which some of the total possible amino acid interactions are satisfied. c) The final structure of proteins (native state). d) Assembly of different protein subunits (oligo-state).

The **secondary structure** consists in the *local* spatial arrangement of the linear amino acid sequence due to the interactions between atoms of the backbone, and thus with no regard of the side-chain rotamers. The conformational freedom of the main backbone is *sterically* constrained by the nature of the amino acids and the partial double bond character of peptide bonds [75]. Therefore, depending on the torsional angles and the *hydrogen-bonding patterns*

different secondary structure motifs are found at this level of protein structure (Fig. 1.3b). The two most frequent elements are α-helices and β-sheets, while *bends* or *turns* occur when main chains change sharply their direction. Interestingly, turns are regions where key residues for enzyme function can be located [76]. Overall, the spatial combination of these secondary elements gives rise to different architectures defined as motifs or super secondary structures that might have both functional and structural significance (*domains*) [77]. Additionally, less regular secondary structures called *coils* or *loops* are also characteristic secondary elements. Although more disordered regions, the interest on these irregular and usually more flexible elements have emerged during the past years due to their roles in protein functions. This is one of the reasons why loops are now not being considered as "mere connectors" of frequent secondary structures. Instead, they have become a promising target for protein engineering, in part for their known roles in processes such as substrate recognition or even in catalysis [78–80].

The collective set of attractive and repulsive forces between amino acids drives the protein chain to fold in a three-dimensional space: the "native" structure (Fig. 1.3c). According to the central dogma in structural biology, this level has been traditionally defined as the "vital state", as there is a conviction that biological function is intrinsically encoded in their detailed **tertiary structures**. For instance, proteins that provide structural support might be expected to adopt fibrous shapes, while the majority of enzymes are known to be built from compact and rounded globular protein scaffolds [81]. On this direction, there can be no doubt that protein structure (and dynamics, see Section 1.4) is intimately linked to function [82]. Yet, there are some observations that may call the necessity of a complex tertiary structure questioned. For instance, there are also some proteins that do not need to fold into stable globular structures to carry out their functions (e.g. *intrinsically unstructured proteins (IUPs)* [83]). Some scientists have succeeded in recreating the catalytic function of an enzyme by building a reduced model based on organic chemicals groups resembling the amino acids responsible for catalysis [84, 85]. Nevertheless, the overall structure of a protein does not merely function as a scaffold for the right placing of catalytic (or otherwise active) residues in space, it also includes structural units that couple the principal function of the protein to the biological context. Interestingly, what makes the three-dimensional shape of each protein *specific* is not only the type of amino acids, but also the unique order of them along the chain. In other words, each type of protein has a *unique sequence* of amino acids, which creates a unique pattern of mutual attractive and repulsive *weak* non-covalent forces between amino

acids (including atoms in the main backbone and in the side-chains) [86]. In general, there are three major dominant types of non-covalent bonds that help proteins to fold, that is, van der Waals attractions, electrostatic attractions, and hydrogen bonds [87]. Some of the 20 different amino acid side chains are non-polar and hydrophobic ("water-fearing"), while others are negatively or positively charged, and so, in an aqueous environment (like inside the cells), the major force driving of the global folding seems to be burying and clustering hydrophobic side-chains to minimize their contact with water molecules [88]. In other words, globular proteins are characterized for having hydrophobic cores with their charged groups pointing toward the protein surface (Fig. 1.4).

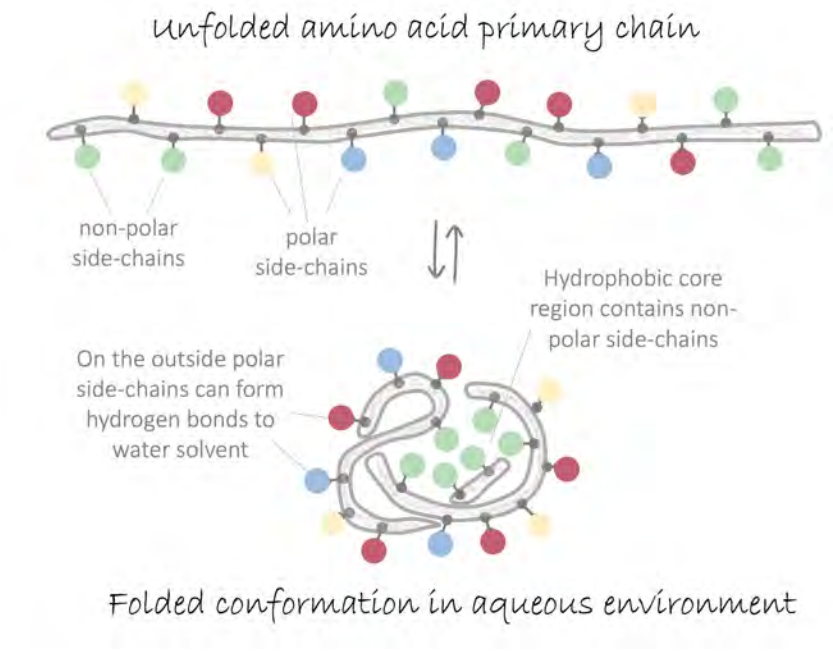


Figure 1.4: Schematic representation of how a protein containing about 35 amino acids folds into a compact conformation. The polar amino acid side chains tend to lie on the outside of the protein (blue and yellow), where they can interact with water. The non-polar amino acid side chains (green) are buried on the inside forming a tightly-packed hydrophobic core of atoms that are hidden from water.

These type of interactions are weaker than covalent bonds, however, they are strong enough to provide tight binding only when many of them are formed simultaneously. For instance, the importance of hydrogen bonds in the stabilization of secondary structures [89], protein folding and stability [87, 90], is well established [91–93]. Nevertheless, these non-covalent forces are not only important on this regard. From a chemist point of view, for instance, a common tool to facilitate a reaction that requires a high energy consumption is the formation and stabilization

of intermediates through different types of non-covalent interactions (NCIs). In fact, there is a large number of fundamental processes in living systems that involve non-covalent interactions, some of them including molecular recognition, binding and catalysis [94, 95].

And finally, although most proteins are active and functional as three-dimensional single subunits, many others need to assemble by means of those weak interactions to finally become functional, thereby giving rise to the fourth and last structural level found in proteins. That is, the **quaternary structure** or the "oligomeric state" [96], whose shape will then depend directly on the primary chain of the linear amino acid sequence (Fig. 1.3d).

1.2.2 Peering Deeper into Enzymes' Structures

From all the above mentioned sophisticated structure, the specific regions where enzymes are doing life's transformations are typically referred as the **active sites**. As enzymatic reactions take place in this particular location of the protein scaffold, enzymes have the ability to distinguish and bind selectively to their substrate(s). In short, this is due to *geometric* and *electrostatic* complementarity between the substrate binding site or active site of the enzyme and its substrate. That is, due to the structural restrains arising from their complex structures ("size and shape"), and due to the specific non-covalent interactions between the enzyme and substrate(s) chemical groups [97, 98]. Similarly to protein receptors and ligands, which can be hormones, pharmaceutical compounds, or environmental chemicals, enzymes and substrates also "come in closely matched pairs". In the sense that an enzyme has the ability to *recognize* just one (or a few) specific substrate(s) [99–104].

Various theories have been proposed to explain the binding of any protein to a ligand (Fig. 1.5). Initially, specific enzyme-substrate (ES) binding was explained by the **lock-and-key** theory proposed by Emil Fischer in the 19th century [105]. This theory stipulated that enzymatic binding sites ("the lock") are rigid and pre-adjusted geometrically to natural substrates ("the key") (Fig. 1.5a). Accordingly, only those starting materials whose structures have an *exact complementarity* with the active site will then form the ES complex required prior to any chemical transformation [106]. Although this initial idea that substrate(s) fits rigidly into the enzyme's active site was widely accepted, in subsequent years there was increasing evidence indicating that, in many cases, enzyme structures undergo conformational

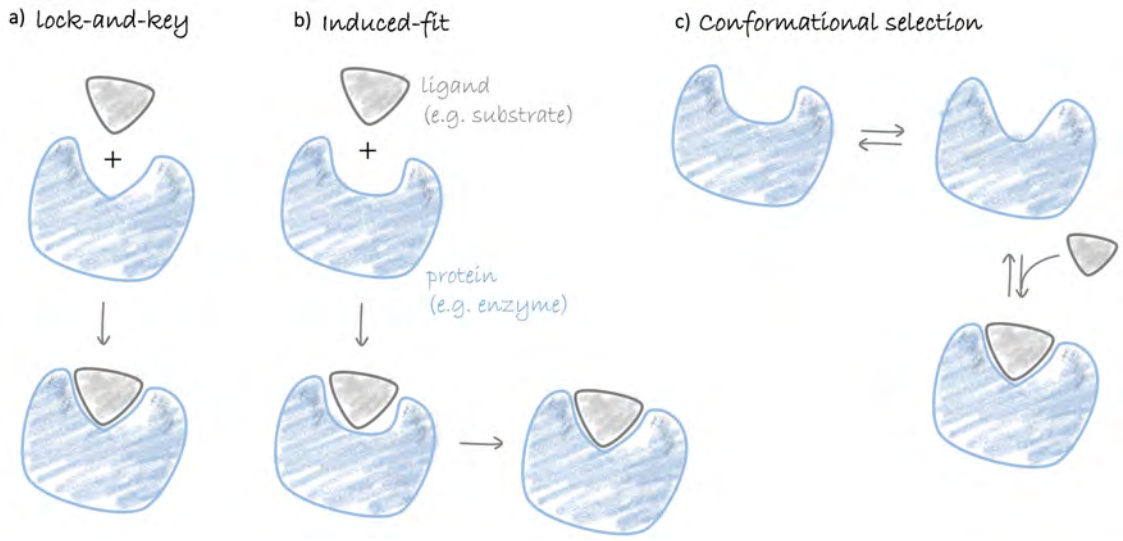


Figure 1.5: Schematic representation of three popular protein-ligand binding theories. The protein is represented as a blue shape and the ligand as a grey shape. a) The "lock-and-key" theory: the ligand-binding site of the protein matches the ligand perfectly. b) The "induced-fit" theory: the binding site generally fits the ligand due to ligand-induced conformational changes in the protein. c) The "conformational selection" theory: proteins in solution are constantly changing their conformations (pre-exist as an ensemble of conformations), with at least one of its matching the ligand.

changes during the binding process [107–109]. In other words, Fischer's model was in a need of a revision, which came in the form of the **induced-fit** theory proposed by Daniel Koshland [110]. This second theory has been validated in numerous proteins [111], and it suggests that enzymes do indeed match their substrates geometrically, although not perfectly. However, once the substrate is present, the binding process is accompanied by conformational adjustments in the binding sites of enzymes that help to reach the perfect match (Fig. 1.5b).

In addition to these two main molecular mechanisms that underlie the binding of any protein to its natural ligand, a few years after the induced-fit theory the *Monod-Wyman-Changeux (MWC)* model of cooperativity appeared [112]. Interestingly, the MWC model withstood that proteins are able to shift spontaneously between (at least) two different conformations even in the absence of a ligand [113]. And therefore, it could explain *allostery*, which is a well-known process in which the binding of the substrate to the catalytic site of the enzyme is affected by ligand binding to a different site [114–116]. Interestingly, the theory underlaying the MWC model has been revised since it was first suggested, completely

changing our very initial perception of proteins, mostly existing as a unique structure with a certain plasticity, to the actual notion of these molecules existing in multiple conformations (or "sub-states") [117–121].

The long-held views on "lock-and-key" *versus* "induced-fit" in binding mechanisms arose from the idea that a protein exists in a single (and most stable) conformation dictated by its amino acid primary sequence. Nevertheless, nowadays it is established that proteins exist in multiple conformations, which finally leads to a third proposal, the **conformational selection** model (Fig. 1.5c) [122, 123]. This last molecular recognition model posits that the ligand binds preferentially to one of the already sampled protein conformations by turning a previously less favorable conformation into the most stable (and probable) one [124]. In other words, by altering the conformational landscape of proteins upon binding (as it will further be explained in Section 1.4). Interestingly, conformational selection does not necessarily refute induced-fit theory, as conformational adjustments further stabilizing the "best-fitting" protein ensemble are indeed usually expected to happen in order "to match the ligand perfectly" [125, 126].

From its very beginning, basic research has focused on many aspects of enzyme action, such as the ways in which enzymes accelerate chemical reactions and maintain specificity and selectivity, their kinetic behavior and how all these aspects can be used for analyzing their efficiency. Before we dive into "how" are enzymes transforming substrates with such remarkable efficiencies upon the enzyme-substrate (ES) complex formation, it might be useful to take first a quick loot at "why" they are so needed to sustain life.

1.3 Unlocking Enzyme Catalysis

Chemical reactions occurring within cells are responsible for sustaining life in all organisms. In fact, according to the molecular needs of any organism, cells are executing thousands of metabolic processes simultaneously and consecutively within a time scale of 10^{-5} to 10^{2} seconds [127]. Every molecule seems to vanish from one point in space and immediately materialize into another, allowing cells to grow or divide, extract energy from food, build complex materials or detoxify harmful substances [128].

In its most simplest case, we can visualize the process: $S \longrightarrow P$, as the reactant substrate molecules needing to negotiate a landscape in which they will need to hike over a hill, the transition state (TS) that stands between them and product molecules along the reaction coordinate. As it is indicated in Fig. 1.6, the transition state acts as an *energy* barrier for the reaction to occur, and this is because the energy of the TS determines "how much energy" the system at least has to gain to reach the top and continue to the formation of products.

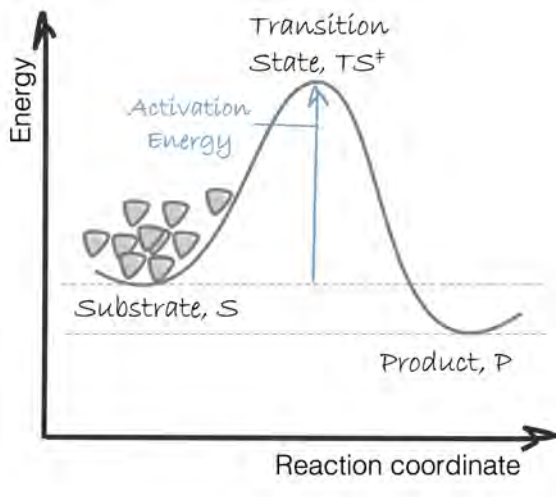


Figure 1.6: Representation of how reactant molecules (grey) need to climb over a single transition state in order to convert into products.

The higher the *activation energy* (E_a) required, the smaller the probability for crossing that energy barrier, and thus a smaller *reaction rate* (k). This dependency between the activation energy value and how fast (or slow) a chemical reaction happens is empirically described by the *Arrhenius equation* [129]:

$$k = A \exp\left(-\frac{E_a}{RT}\right) \tag{1.1}$$

where R is the universal gas constant, T is temperature, RT is the average kinetic energy, and the pre-exponential coefficient A is a "frequency factor". In a chemical reaction in which the reaction rate is linearly dependent on the concentration of only one reactant (a first-order reaction), the A term can be viewed as the total number of random collisions that happen to form products, while the exponential term describes the probability for a successful collision leading to the transition state and then to products. In this context, the reaction rate k is then understood as the "fraction of such successful collisions", being expressed in enzyme units, which are usually given in conversion of substrate per minute under specific conditions.

As it is captured by Eqn. 1.1, the reaction of a chemical reaction depends directly also on the temperature. On this direction, one way to provide reactant molecules sufficient activation energy is by heating them up to a very high value of temperature. In fact, heating the substrate molecules up is one of the common strategies to provide these molecules the necessary energy to reach the transition state.

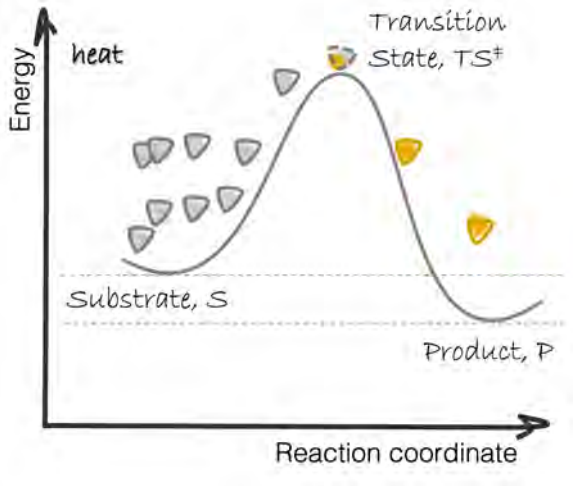


Figure 1.7: Representation of how reactant molecules (grey) need to climb over an energy transition state in order to convert into product molecules (yellow).

As it is represented in Fig. 1.7, this will enable reactants to gain enough energy to move and vibrate faster, and thus overcome the energy barrier imposed by the transition state. The science of "how heat interacts with matter" is called *thermodynamics*. The field of thermodynamics enables to characterize different states in nature at the macroscopic level, mostly by studying the dependency of the energy on magnitudes such as temperature, pressure, volume, and chemical concentration. In fact, almost all of the physical and chemical processes that cause change in our world are driven, at a molecular level, by thermodynamic principles, which in turn are based on these "random molecular collisions". Interestingly, some of the energy available from these random collisions is harvested to construct and maintain living systems as these highly ordered states we can appreciate.

1.3.1 Free Energy Changes and Spontaneous Processes

In terms of thermodynamics, the energy that is available and then transformed into a specific work (e.g. execution of a chemical process or changing molecular configuration) is the so-called

Gibbs free energy (or simply free energy) quantity. This free energy is denoted by G, and it is considered one of the key concepts in this field of physics. As living organisms exist in a state of constant pressure and temperature, the free energy is a more convenient magnitude for determining the direction of a given cellular process. In fact, it is the free energy *variation* (ΔG) what helps us to predict whether a process might take place (or not) effortlessly. In other words, *spontaneously*.

The value of ΔG can be calculated from the *standard* free energy (ΔG^0), assuming that the concentrations of initial and final states (e.g. reactants and products) are known [130]. Interestingly, the free energy change itself can be also obtained from changes in two other thermodynamic quantities:

$$\Delta G = \Delta H \ - \ T\Delta S \tag{1.2}$$

That is, the *enthalpy* (H), which is the general tendency to *minimize the energy*, and the *entropy* (S), which is the general tendency to *maximize the disorder*.

Unlike the free energy, the enthalpy and the entropy quantities can be associated with particular features of the molecular system. For instance, the ΔH component of ΔG in Eqn. 1.2 involves molecular phenomena, such as the formation or breaking of covalent bonds, changes in electrostatic or van der Waals interactions, and changes in thermally-induced atomic motions [131, 132]. For the remaining component $-T\Delta S$, a good example is the process of protein folding, in particular when it involves burial of non-polar atoms inside macromolecular proteins. As we have already seen in Fig. 1.4, this folding process is driven by a hydrophobic effect. "Ordering" events tend to increase the free energy of the system, such as the confinement of substrates inside the active site. Nevertheless, in biological systems there is also "disordering" and energetically events that often compensate this loss of entropy. In the context of the folding process, for instance, it is accompanied by an increase in the entropy of the aqueous solvent, in addition to the attractive non-covalent interactions between amino acids of the protein scaffold [131].

Given a particular set of conditions, any natural state can be indeed characterized by its free energy value (see also Section 1.4). Knowing the free energy of a system is very helpful in predicting the relative stability of states and the direction of processes,

including enzyme-catalyzed reactions, protein-ligand binding, or protein folding events. This is particularly true for biological chemical processes, which involve a change in the molecular properties of the system. We can view a reaction as a system capable of existing in two alternative states with different free energy values. In this context, the reaction will spontaneously shift from the high-energy state to the low-energy one. And therefore, the direction of spontaneous change is always in the direction of decreasing free energy (i.e. $\Delta G < 0$) (as it happens when the final product of a reaction has lower energy than the starting material).

Most of the life-sustaining reactions occur spontaneously, although under physiological conditions they tend to happen very slowly. In fact, they can take hundreds, and even million years to complete [133]. The reason for their slowness has to do with the *energetics*, and with the pathway from reactants to products [134]. For obvious reasons, increasing the temperature to a very high value is an infeasible approach for biological systems, and thus, the other only way left is to accelerate chemical reactions by lowering the activation energy barriers. In the context of living organisms, by using enzymes as catalysts.

1.3.2 Enzyme Catalyzed *versus* Non-enzymatic Catalyzed Reactions

Like any catalyst, enzymes are substances differing from reactants that help to speed up the chemical reactions without being changed and consumed along the process. The transition state might be "the same" as the expected for the uncatalyzed counterpart, although in the presence of a (bio)catalyst its energy is much lower (Fig. 1.8). In other words, the transition state has been stabilized, and thus reactant molecules need less activation energy to overcome the barrier imposed by this transient reaction intermediate.

When we use a non-enzymatic catalyst (Fig. 1.8a), it will bind to reactant A to form a complex denoted here as "Cat-A". After binding, the substrate is transformed into product B, which is finally released to start again the catalytic cycle. Interestingly, it is worth to notice that the energy difference between reactants and products, which is ΔG_R, remains unaltered, and thus the overall energetics of the reaction do not change. In other words, the thermodynamics are not affected by the presence of any catalyst. When the catalyst is

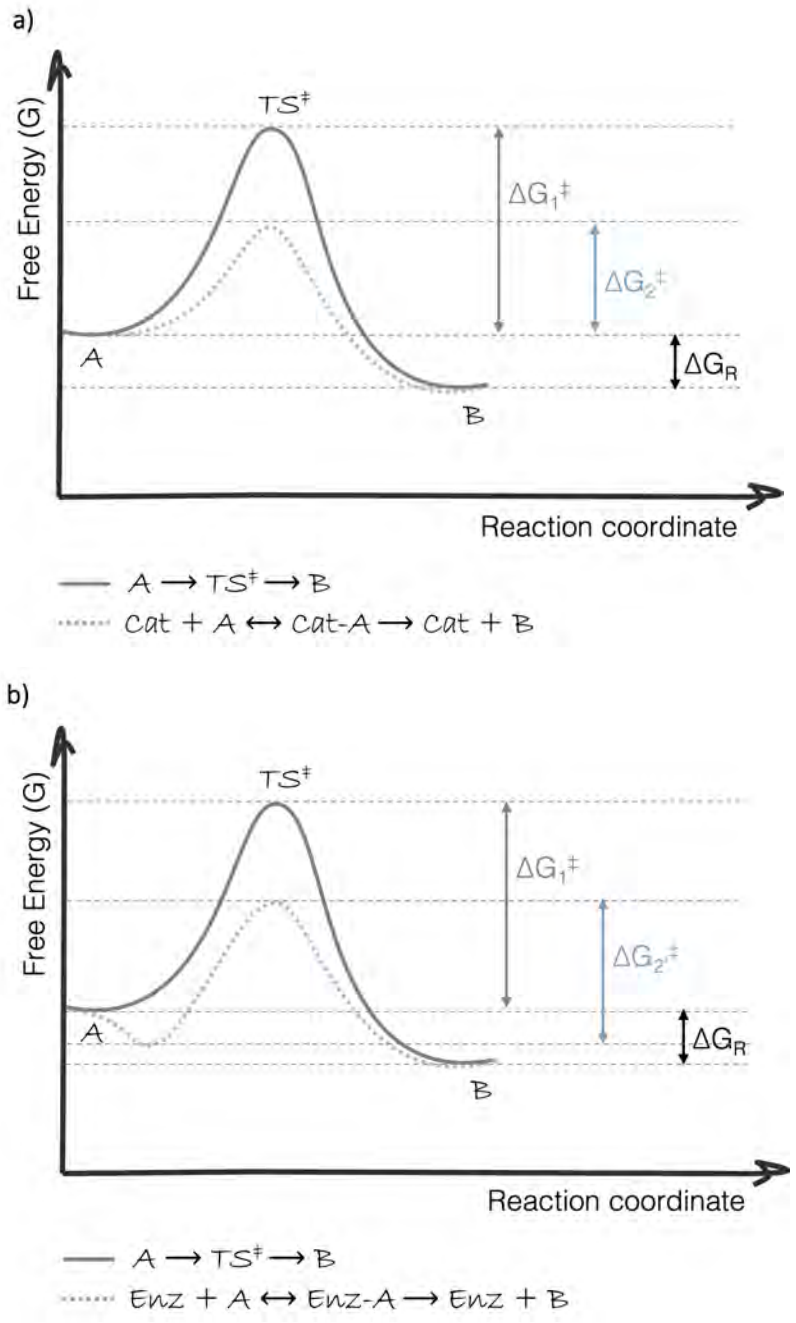


Figure 1.8: Representation of the energetics of a chemical reaction in the absence and presence of a catalyst. The grey plot illustrates the hypothetical reaction: $A \longrightarrow B$. The reaction involves a transition state ($TS^{\ddagger}$), whose free energy exceeds that of the reactant by an activation energy of $\Delta G^{\ddagger}$. Comparison between a) the effect of a chemical catalyst (Cat), and b) the effect of an enzyme (Enz) catalyst.

an enzyme (Fig. 1.8b), the substrate also binds, although to an active site pocket in the enzyme structure that stabilizes it. As we have already explored in Section 1.2, the binding process leads to the formation of a specific enzyme-substrate (ES) complex, which results in a favorable drop in the free energy of the system (i.e. an entropy decrease). As it is represented in Fig. 1.8b, this binding in turn make the activation energy required slightly higher, although the activation energy barrier to overcome is still lower than the one for the uncatalyzed reaction (e.g. $\Delta G^{\ddagger}_{2'} < \Delta G^{\ddagger}_{1}$). In a thermodynamic sense, enzymes seem to follow the same fundamental physical and chemical principles of chemical catalysts, although they are just better designed [135].

1.3.3 A Quick Look at Enzyme Kinetics

In chemical kinetics, reactions are usually characterized in terms of stoichiometry, mechanisms and order. The latter term describes the dependency of the reaction on the concentration of reactants, being able to distinguish between zero, first, and second-order reactions. For instance, when a reaction rate follows a linear dependency on the change in concentration of the reactant A over time, then it is considered a first-order reaction. And therefore, for the hypothetical reaction: $A \longrightarrow B$, the reaction rate could be expressed as it follows:

$$V = r = -\frac{d[A]}{dt} = k[A]^1 \tag{1.3}$$

where k is the *rate constant*, and its value is affected by the environment of the reaction and the corresponding starting materials.

When we think about (enzyme) kinetics, what we want is to simplify the conversion of substrate molecules into products as the elemental change described from reactant A to product B. On this direction, the simplest enzymatic reaction scheme: $S \longrightarrow P$, could be written as it follows:

$$E + S \underset{}{\overset{1}{\rightleftharpoons}} ES \overset{2}{\rightleftharpoons} E + P \tag{1.4}$$

where S is the substrate(s), E is the "free" enzyme, ES is the enzyme-substrate complex formed prior to the chemical transformation, and P is the final product(s). The first step of the reaction corresponds to the binding of substrate(s) to enzymes, while the second step describes the transition from substrate(s) to product(s), each of them having its own *rate*

equations.

When reversible reactions are considered, rate constants for the forward and backward reactions are necessary to describe the kinetics. For the particular case presented in Eqn. 1.4, that would be: $r_1 = k_1[E][S]$ and $r_2 = k_2[ES]$ for the forward direction, and $r_{-1} = k_{-1}[ES]$ and $r_{-2} = k_{-2}[E][P]$ would describe the backward direction. According to these equations, and assuming a constant value of k, the overall reaction rate could be enhanced by increasing either the substrate or the enzyme concentration (i.e. $[S]$ or $[E]$). However, opposite to what should be expected, for enzymatic processes this is not likely to happen. As it is represented in Fig. 1.9, at high values of substrate(s) concentration, all free enzyme has been fully *saturated* into the ES form, and thus enzymatic rates are observed to reach a *plateau*. That is, a situation in which the speed of the reaction reaches a maximum value (V_{max}) and thus it becomes insensitive to further changes in substrate concentrations.

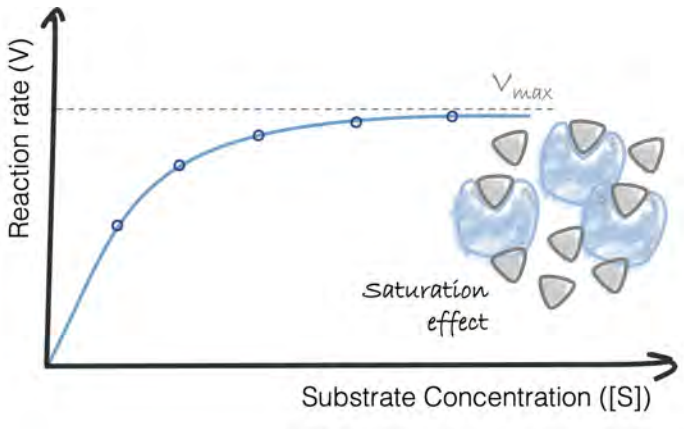


Figure 1.9: Representation of the saturation effect in enzymatic reactions. Enzyme(s) are represented as a blue shape and substrate molecules in grey. Dots in the plot indicate measured initial rate or velocity values at points with different substrate concentrations.

This "saturation kinetics" is indeed a hallmark of enzymatic catalysis, and it is different from what is observed in uncatalyzed reactions, as there are no binding or active sites that can become saturated. Interestingly, the above description explains the molecular origin of the observed saturation kinetics of enzymes, although it does not provide any mechanistic or quantitative insights. Nowadays there are numerous experimental and computational methods available on this regard. Nevertheless, for a long time the only strategy available for understanding enzyme action was to construct mathematical models capable of explaining enzymatic rate acceleration and catalytic steps using "measured parameters". In other words,

by developing and using kinetic models.

Michaelis-Menten Kinetic Model

The most popular model proposed to capture quantitatively enzyme kinetics in a mathematical formalism is so far still nowadays the one provided by Michaelis and Menten, which describes chemical reactions as it follows [136–139]:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \tag{1.5}$$

where E is the enzyme, S is the substrate, ES is the enzyme-substrate complex, P is the product, k_1 is the rate constant of ES formation, k_{-1} is the rate constant of ES breakdown, and k_2 is the rate constant of *the chemical* reaction step. That is, the step in which product(s) are formed irreversibly and then released from the enzyme.

Within Michaelis-Menten formalism, the substrate conversion into product and its subsequent release into solution is considered to be much slower than the enzyme-substrate binding. In other words, k_2 is assumed to be much smaller than k_1 and k_{-1}, being then k_2 the magnitude that determines the overall rate of the reaction. And therefore, it is this step the one determining how fast or slow takes the transformation from substrate(s) to product(s). In this context, k_2 describes the *rate-limiting* step of the entire enzyme-catalyzed reaction, and thus it is often also considered as the reaction's catalytic rate k_{cat}. Accordingly, within Michaelis-Menten model the enzymatic reaction rate could be expressed as it follows:

$$V = \frac{d[P]}{dt} = k_2[ES] \tag{1.6}$$

It should be noted, however, that enzymatic mechanisms are usually a complex series of steps involving different reaction intermediates, and thus k_2 or k_{cat} are often a compilation of several rate constants rather than one. In addition, it should be also noted that Michaelis-Menten model refers to chemical reactions that are catalyzed by an enzyme "of fixed concentration" and that are in a "steady state" [140]. That is, the ES does not change over time, and thus the rates of the ES formation and decomposition (k_1 and k_{-1}, respectively) are considered to be equal:

$$\frac{d[ES]}{dt} = 0; \quad k_1[E][S] = k_{-1}[ES] + k_{cat}[ES] \tag{1.7}$$

And therefore, under saturation conditions we can assume that the amount of total enzyme corresponds to the amount of ES complex, as all substrate molecules are bound to free enzyme ($[E]_f$) (Fig. 1.9). In this context, we can in the end say that it is the total concentration of the enzyme $[E]_t$ the value that does not change either over time.

$$[E]_0 = [E]_f + [ES] = constant; \quad [E]_0 = [ES] \tag{1.8}$$

Considering Eqn. 1.7, as well as the assumptions already mentioned along this Section 1.3.3, we can finally derive the fundamental equation of Michaelis-Menten kinetics, which is:

$$V = \frac{\mathrm{k_{cat}}[S][E]_0}{\mathrm{K_M} + [S]} \tag{1.9}$$

where, K_M is commonly referred as the Michaelis-Menten constant, and its magnitude is indeed a lumped term that includes the three rate constants: $K_M = (k_{-1} + k_{cat})/k_1$. And the remaining $k_{cat}[E]_0$ term results to be equal to the maximum rate V_{max}.

Uses of Michaelis-Menten Parameters: Enzyme Efficiency and Specificity

The qualitative and quantitative relations between the different Michaelis-Menten kinetic parameters enable us to analyze and compare different aspects of enzyme activity among many natural enzymes. Although this model is not necessarily valid for describing the kinetics of enzymatic processes involving complex mechanisms such as those governed by allostery, this model has proven to be useful for understanding and predicting fundamental enzyme behavior. This is because K_M and k_{cat} are "measurable" parameters. In many cases, for instance, the inverse value of Michaelis constant K_M is considered as an approximate measure of the affinity between the enzyme and its substrate(s): "the lower the K_M value, a greater affinity to the substrate". As this constant can easily be measured in the lab, it is often used to approximate the enzyme-substrate affinity. However, it should be noted that this has to be done with extreme caution, since K_M does only reflect *enzyme-substrate affinity* in specific situations. For instance, when the rate-limiting step of a simple reaction is significantly lower than the rates of the ES forming or breaking steps (i.e. $k_{1/-1} >> k_2$). In such case, k_2 can be neglected, and then it is correct to say that $K_M \approx K_S$.

Regarding the catalytic rate constant k_{cat}, this magnitude has been traditionally used as

a measure of enzyme activity and catalytic efficiency, as it can be viewed as the *turnover number*, which is the number of product molecules formed by one enzyme molecule in a unit of time (usually seconds) [74]. As we have already seen in the beginning of Section 1.3, the Arrhenius equation 1.1 describes the dependency between the activation energy required to overcome the transition state barrier and the rate of the reaction. Nevertheless, it does empirically, and thus it does not consider mechanistic aspects such as one stage of the reaction being the rate-limiting step. On this regard, a more adequate description about this relationship is provided by Eyring's *transition state theory* [141, 142], in which the dependency of k_{cat} on the activation free energy ($\Delta G^{\ddagger}$), as well as temperature, is described by what is called the *Eyring-Polanyi equation*:

$$k_{cat} = \frac{k_B T}{h} exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) \tag{1.10}$$

where k_B is the Boltzmann constant, h is Plank's constant, R is the universal gas constant, and T is the temperature. Thanks to Michaelis-Menten formalism, k_{cat} value can also be easily obtained by measuring initial reaction rates (i.e. V_0), since under saturation conditions it is true that: V_0 is equivalent to V_{max}. Nevertheless, the overall efficiency of enzymes is not only affected by the efficiency of the catalytic steps. In other words, it can be also affected by the enzyme's affinity to the substrate. For this reason, the k_{cat}/K_M ratio is a better measure for describing enzyme efficiency. Compared to the isolated terms, the k_{cat}/K_M ratio is a quantity that offers the advantage to include both catalytic and affinity-related contributions to the total enzyme efficiency. And not only that, as in cases in which a single enzyme has the ability of transforming different substrates, a comparison of the efficiency of the enzyme based on this ratio across substrates can be used as a measure of specificity. For this reason, the k_{cat}/K_M magnitude is also typically referred to as the "specificity constant" [139].

1.3.4 Stereochemistry of Enzymatic Reactions

A very common case of specificity is when the enzyme can act on two different enantiomers of its substrate (S and R), but has a clear preference for one of them over the other. This situation is called *enantioselectivity (E)*, and it can be expressed as the ratio between the

specificity constants relating to the enzyme's activity on the two enantiomers [143]. That is,

$$E = \frac{\left(\frac{k_{cat}}{K_M}\right)_S}{\left(\frac{k_{cat}}{K_M}\right)_R} \tag{1.11}$$

Interestingly, the enantioselectivity E can be used to calculate the difference in the activation free energy ($\Delta\Delta G^{\ddagger}$) between the two enantiomers as it follows:

$$\Delta\Delta G^{\ddagger} = -RT \ln E \tag{1.12}$$

and thus, by computing and comparing the activation free energies barriers of the transition state of each enantiomer we could reproduce (or predict) experimentally observed selectivity trends (Fig. 1.10).

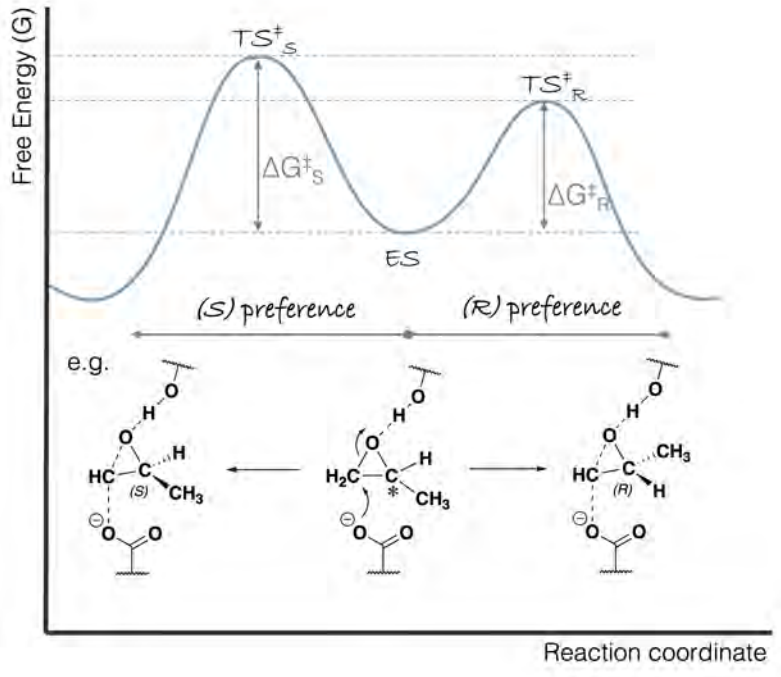


Figure 1.10: Representation of the use of transition state theory (TST) for analyzing enzymatic selectivity. As example, it is represented the opening at the terminal epoxide carbon atom of each enantiomer (S/R) found in a racemic mixture. The chiral carbon atom is indicated with an asterisk in the ES simplified structure.

In fact, this idea is the basis of Eyring's **transition state theory** (TST). According to this proposal, the smallest fraction of the catalytic cycle is spent in the most important step,

that of the transition state. On this direction, collisions between the substrate and enzyme molecules are overlooked, and thus the only physical entities required for describing enzyme efficiency and selectivity are the ground and the transition states, that is, ES and each TS in Fig. 1.10. Worth to notice is the fact that the above equation 1.12 can be used to express *any type of substrate specificity*, and thus it is not only limited to enantiomeric processes. Indeed, computing the activation free energies to elucidate (or predict) enzymatic selectivity is a common strategy while modeling enzymatic mechanisms [144].

Whether or not enantiomers are of biological origin, these molecules do not interact identically with other chiral molecules, such as enzymes. In the absence of any source of chirality, that is, optically inactive reactants and catalysts, the outcome would be a racemic mixture. In fact, many reactions lead to mixtures in which there is exactly equal amounts of the two enantiomers. Yet, enzymes are inherently chiral, thereby being able to recognize one substrate enantiomer over the other. Some enzymes are found to display "absolute" enantioselectivity, while many others show different degrees of preference for a specific substrate enantiomer. Usually, when only one enantiomer of a pair fits properly into the chiral active site of an enzyme, the process is considered to be *enantioselective* (or more generally, stereoselective), which is a concept that should not be confused with the ability of enzymes to form a single configuration in a *pro-chiral* substrate. These type of reactions in which only one pure chiral product is formed should be defined as *stereospecific*, which is indeed a property that relates to the mechanism of the reaction [74]. For example, we can illustrate how this works at a molecular level by considering the ring-opening reaction of epoxide substrates in solution. As it is represented in Fig. 1.11, the choice of the pathway adopted is governed by several factors, such as the reaction conditions and the substitution patterns of the starting materials [58].

The resulting product will vary according to which substitution mechanism dominates during the epoxide ring opening reactions. For example, under basic conditions, this reaction proceeds by an S_N2 reaction mechanism, and although there are two electrophilic carbons in the epoxide moiety, the best target for the nucleophile in this context is the "least hindered" center (Fig. 1.11a). This is because in the S_N2 mechanism, the nucleophile carries out the attack from the backside orientation of the substrate relative to its leaving group, thus resulting in inversion at the electrophilic carbon. On the contrary, when the opening of the

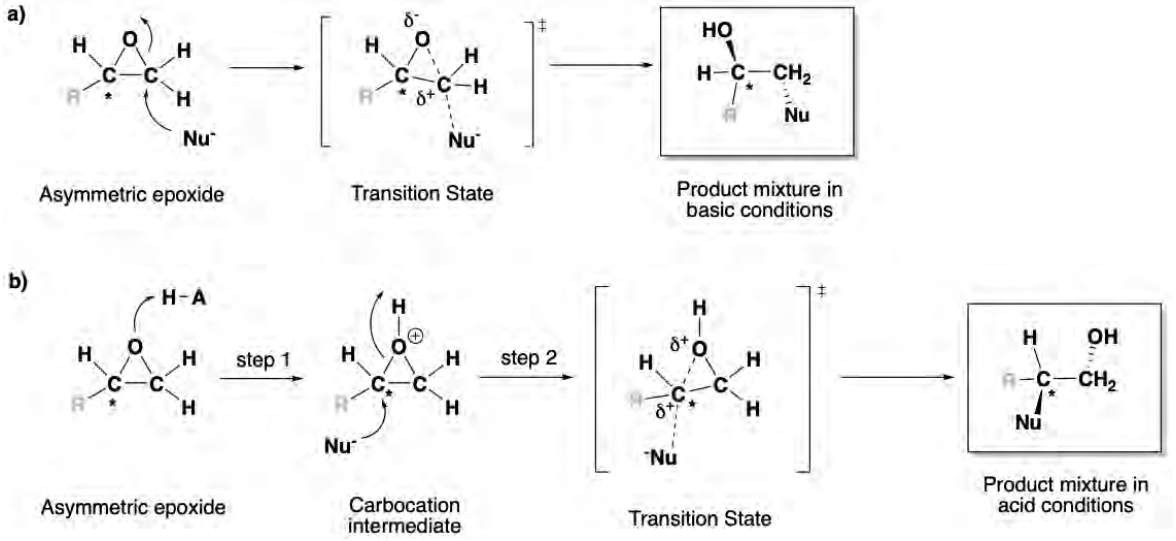


Figure 1.11: Representation of epoxide ring-opening reactions as an example of a stereospecific reaction. Arrows illustrate the electrons movements, R refers to any substituent different from a hydrogen atom, "Nu" stands for any molecule or atom acting as a nucleophile (e.g. in solution, a water molecule). a) Representation of the epoxide ring-opening reaction in basic conditions *via* a S_N2 mechanism. b) Representation of the epoxide ring-opening in acidic conditions *via* a S_N1 mechanism.

epoxide ring proceeds under acidic conditions, the reaction occurs by an S_N1 mechanism, which generally involves two steps. That is, a protonation step followed by the nucleophilic attack. If the nucleophile is neutral, an additional deprotonation step is also required to complete the reaction. Nevertheless, under acid conditions the attack occurs through the formation of carbocation intermediate, which is better stabilized by neighboring carbon atoms. On this direction, in acidic conditions the best target for opening the epoxide is the more substituted carbon, since the positive charge build on that carbon provides the carbocation of greatest stability. In practice, however, a better way to describe the acid-catalyzed epoxide ring-opening is by depicting a hybrid pathway between S_N1 and S_N2 mechanisms (similarly to the transition state depicted in Fig. 1.11b). This is because the nucleophile is indeed reacting with the electrophilic carbon before the carbocation intermediate has the chance to form. And since the carbon-oxygen is still to some extend present, the oxygen of the epoxide partially blocks the attack from the front side.

As represented in Fig. 1.11, if the starting material is an asymmetric epoxide substrate, then the production of a specific compound can be favored by modifying the reaction

conditions. As we have already explored, in basic conditions under steric considerations the attack is being favored at the less substituted carbon. Yet, in acid conditions the process is governed by electrostatic considerations, and so the more substituted carbon is the preferred position for the nucleophiles. In such situations, the chemical process is also referred to be *regioselective*, although we still end up with two different products that are indeed each of them a mixture of two enantiomers. This is important, specially in the biological context, since usually only one enantiomer of a pair causes a desired effect inside our bodies. Many drugs currently on the market are found to be racemic mixtures, causing most of the times undesired effects. And therefore, it is of utmost relevant to provide methods by which racemic mixtures are *resolved*. As it has been shown in the very beginning of this Chapter, the process by which two enantiomers can be separated is called *resolution*, and it is a very different process from the common physical separations. Nevertheless, in the presence of a "chiral probe" such as enzymes, the differences in enantiomers become apparent.

1.3.5 Catalytic Strategies for Enzyme Catalysis

For a long time, the remarkable efficiencies of enzymes have been widely attributed to the precise *pre-organization* of the active site residues for an efficient **transition state stabilization** [106, 145, 146]. To do so, some of the amino acids inside the active sites carry out different catalytic roles. For instance, some amino acids are responsible for the activation of reaction species, others can affect the reaction outcome through factors related with spatial arrangements (i.e. steric clashes), some other residues act as "shuttles" (e.g. donate, accept or relay protons), while many others can become covalently attached to a reaction intermediate in order to greatly accelerate chemical reactions [142, 147].

As we have already explored in Section 1.2.2, the high complementarity between the enzyme and the transition state results from various structural and chemical features of the active site. On this direction, enzymatic specificity arises from the three-dimensional structure of the enzyme's active site, as it is viewed to be much higher *complementary* to the transition state of the reaction than to its substrate(s) [106]. In other words, enzymatic specificity can be understood as most manifest in "the rate that a substrate reacts", although the affinity of substrate binding it is also well-known to play a role [148–150]. A recent example supporting the relevance of binding affinity in determining the substrate specificity is

the case of N-Myristoyltransferases (NMTs), in which the enzyme catalyzes almost exclusively myristoylation when both acetyl-CoA and myristoyl-CoA substrates are present in the reaction [151].

The various catalytic strategies employed by enzymes to promote the formation of the transition state(s) of the reaction involve different aspects of enzyme-substrate (ES) interactions. Some of these aspects include the entropic effects that result from the substrate's confinement in the active site of enzymes (Section 1.2.2), the stabilization of reaction intermediates throughout non-covalent interactions, such as hydrogen-bonding, electrostatic or van der Waal interactions, or the chemical actions carried out by the active site residues on the substrates, such as the formation of covalent bonds or proton/electron transfer processes [152–154].

The **electrostatic pre-organization** and **non-covalent stabilization** of transition state(s) is a very common strategy in enzymatic catalysis, since the conversion from substrate(s) to the corresponding transition state(s) typically implies a "charge delocalization". For instance, if the transition state has an electric charge that is not present on the substrate, the active site is likely to include a chemical group with an opposite charge that will be able to stabilize it. Interestingly, electrostatic interactions are viewed as a major contribution for an efficient transition state stabilization or **ground state destabilization**, although this type of stabilization is not exclusive to enzymatic catalysis [98, 155, 156]. In many cases, the stabilizing charged groups belongs to a side chain of amino acids (e.g. the positively charged guanidinium group of arginine) [157]. In some others, it is the amide or carbonyl group in the protein's backbone that carries a partial positive or negative charge, respectively. Interestingly, the same amide groups can be also said to "strain" the substrate (or any reaction intermediate) into its transition state. Such *induced-strain distortion* of the ground state by enzyme-substrate binding is known as *the Circe effect*. Interestingly, this destabilization was originally assumed to be governed by geometric aspects [158], although later on, electrostatic effects have been found to have a more significant contribution [159, 160]. And finally, in those cases in which the transition state is negatively charged, there is typically also the participation of active site cationic metals for transition state stabilization, such as Mg^{2+}, Cu^{2+}, Zn^{2+}, Mn^{2+}, Co^{2+}, Fe^{2+} and Fe^{3+}, which in enzymes are usually referred to as *cofactors* [161, 162].

Along with non-covalent stabilization of reaction intermediates, the other most common function of catalytic residues in a general view is **acid-base catalysis**. The prevalence of acid-base catalysis as an enzymatic strategy largely accounts for the well-known dependency of enzymatic activity on the pH of the enzyme's environment. For instance, if an enzyme relies on the acid-base behavior of an aspartate, the optimal pH would be expected to be around 3 or 4. In contrast, if an enzyme relies on a certain histidine residue to function as a general base, then the pH for an optimal activity will be around 6 or 7. Interestingly, the protonation or deprotonation of any ionizable group is most affected around its pK_a, as well as its immediate chemical environment inside the active sites. And thus, the amino acids around catalytic residues might be also relevant, as these residues might affect *via* electrostatic effects its pK_a. On this direction, if an enzyme relies on the basicity of a histidine residue, then its local environment will be expected to keep it in a deprotonated state before catalysis starts. However, upon substrate binding, the conversion of the substrate into the transition state will lead to the protonation state of that histidine, which will return to its initial deprotonated state once the catalytic cycle is over.

The simplest way in which acid-base catalysis can exert its effect is to directly extract or transfer a proton. Yet, acid-base catalysis may also contribute indirectly to enzyme action by helping other active site residues to carry out *nucleophilic substitution* on the substrate. That is, promoting **covalent catalysis**, which is a strategy that often leads the formation of an enzyme-substrate covalently bound intermediate after the activation of the nucleophile. In this context, the chemical groups inside the active site are not just merely interacting through non-covalent interactions, but act on it chemically, thereby promoting the change of substrate(s) or intermediate(s) into its transition state(s). Indeed, many enzymes promote the formation of the transition state(s) and the subsequent breaking of such covalent intermediates, although the exact means by which covalent catalysis happens varies across enzymes. Regardless the chemical group acting as a nucleophile, the present strategy usually involves an attack on a substrate's electrophilic center that results in a covalent bonding between the two. In other words, it is also assisted by an electronic polarization. In hydrolases, for instance, the electrophilic center attacked in the substrate is usually a carbonyl group (C=O) that results in an *acyl-enzyme* intermediate. The negatively charged transition state is stabilized electrostatically within the active site, and such stabilization

promotes the subsequent breaking of the covalent bond in the substrate by an active site water molecule [163]. Similarly, in epoxide hydrolase (EHs) there is also the formation of a covalent enzyme-substrate intermediate that is afterwards hydrolyzed. We can view how this works at a molecular level by considering, for instance, the epoxide ring-opening step mediated by *Bacillus megaterium* EH (*Bm*EH).

Bacillus megaterium Epoxide Hydrolase Mechanism

Likewise a large fraction of EHs, *Bm*EH belongs to the α,β-hydrolase (ABH) family, which is one of the largest groups of structurally related enzymes with diverse catalytic functions [163–166]. Enzymes can be grouped into different protein families with related structural and functional similarities. In the particular case of the α,β-hydrolase fold, which is common to several hydrolytic enzymes that appear to be related by divergent evolution, a remarkably conserved *core* domain is shared in spite of the low overall sequence conservation (Fig. 1.12a). In other words, the fold includes enzymes of widely differing phylogenetic origin and catalytic function [53, 54]. And therefore, members of this group display a high sequence diversity and a broad variety of catalytic activities such as protease, lipase, peroxidase, esterase, dehalogenases, and epoxide hydrolase functionality [55]. At the same time, however, it is worth mentioning that interconversion of activities through selective alterations in active site architecture and chemistry within different members of α,β-hydrolase family has been possible, in in spite of sharing the same core catalytic machinery [163]. And therefore, although these different catalytic activities are separated in sequence space due to the divergence process along evolution, the activities are interestingly overlapping rather than diverging [167]. In other words, as protein structure is often more conserved than function in the course of evolution, in practice through few single amino acid mutations it is indeed possible to evolve novel catalytic.

A majority of ABH family members contain three topologically similar residues that form a *catalytic triad*: a nucleophile (serine, aspartate or cysteine), a histidine, and an acidic residue (aspartate or glutamate). Indeed, the ability of the ABH fold enzymes to operate on substrates with different chemical composition or physicochemical properties is based on the evolution of this similar "nucleophile-base-acid" triad located at the core of the ABH fold [168]. In addition to the catalytic triad, the core domain also contains two or three amino acids considered to be essential for the mechanisms of the ABH members, that is, the *oxyanion hole*

residues [169–171]. Generally, these active site residues are located on conserved locations and in loops, and it is the three dimensional ABH fold which brings them together in the catalytic centers. Interestingly, many α,β-hydrolase enzymes also contain further structural motifs in which other catalytic relevant residues have been observed [47, 172], such as the *cap* or *lid* domain located above the catalytic cores (see Fig. 1.12a). However, in contrast to the highly conserved core domain, the structures of these frequent additional motifs vary considerably between the different members within the ABH family [167].

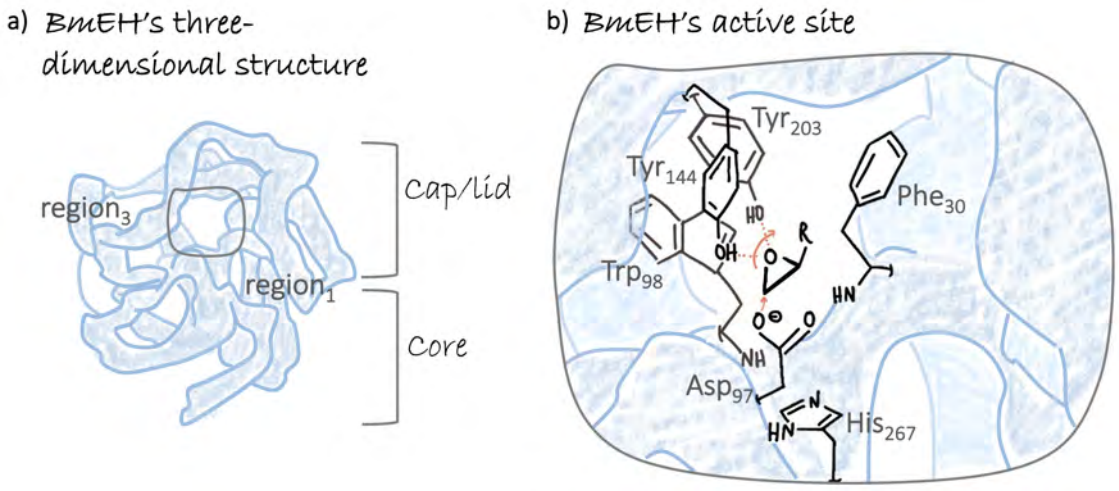


Figure 1.12: Illustration of *Bacillus megaterium* epoxide hydrolase three-dimensional structure and its active site residues. The *Bacillus megaterium* epoxide hydrolase (*Bm*EH) is represented as a blue shape. a) Schematic representation of *Bm*EH core and cap/lid main domains. The active site, substrate entrance (region 1) and product release (region 3) are also indicated in this simple illustration of *Bm*EH three-dimensional structure. b) Zoom in at the active site of the *Bm*EH enzyme: the epoxide substrate (in bold) and relevant catalytic residues are also indicated. Number labelling corresponds to the *Bm*EH wild-type enzyme (PDB ID: 4NZZ). The tethering of the epoxide substrate inside the active site is shown by red dashed lines, whereas red arrows illustrate the movement of electrons as bonds between atoms that will be broken and formed.

EHs are in general characterized for having a deep and hydrophobic active site cavity on the interface of the core and cap domains. In the particular case of the enzyme studied in this thesis, which is the *Bm*EH, its three dimensional structure is featured by a large "open" active site entrance (in this thesis also termed as "region 1"), and an additional cavity not observed in other EHs described as an independent product release site (in this thesis also called "region 3") [59] (Fig. 1.12a). As it is illustrated in Fig. 1.12b, the particular residues that constitute *Bm*EH's catalytic triad are the nucleophile Asp_{97} found in a sharp

turn commonly referred as the "*nucleophile elbow*", the basic His_{267} and the acidic Asp_{239}. In contrast to most of the ABH family members, EHs in general do not employ the common serine-histidine-aspartate catalytic triad. The reason is that this strategy is not suitable, since the nucleophilic attack by a serine residue would facilitate the formation of an ether intermediate that would not be possible to break [173]. For this reason, in EHs we find an aspartate-histidine-aspartate/glutamate triad instead, which facilitates the formation of an *ester* derivative of the aspartate [51, 174]. Far from the situation presented in Fig. 1.11, inside the active site of *Bm*EH enzyme only epoxide substrates that are geometrically and electrostatically complementary will preferentially bound. Similar to many other EH enzyme [175–177], in *Bm*EH there are also two additional tyrosine residues, the Tyr_{144} and Tyr_{203} located on the two adjacent α-helices of *Bm*EH's cap domain (Fig 1.12). [178–182]. These two acidic tyrosine residues are responsible for binding and substrate recognition, but also for assisting the reaction by polarizing the C-O bond through weak hydrogen bonds and facilitating a proton to the oxygen atom once the ring is completely opened (Fig. 1.13). This first step of the reaction leads to the ring-opening of the epoxide, thereby forming an *alkyl-enzyme* intermediate. In addition, there is another residues relevant for *Bm*EH mechanism. That is, Gly_{29} and Phe_{30}, two amino acids that belong to the conserved HGXP motif (X = Phe_{30} in this case). The backbone amide group of these residues is directly involved in the formation of the oxyanion hole, as they are required for stabilizing the negative charged *tetrahedral* intermediate that will arise upon the cleavage of the covalent bond between the enzyme and the substrate (Fig. 1.13). Such stabilization is what promotes at this second stage of the reaction the hydrolysis of the alkyl-intermediate, thereby leading to the final diol product.

It is worth to notice that all the above mentioned mechanisms used by enzymes to accelerate chemical reactions result directly from the pre-organization of enzymatic active sites. This is because the three-dimensional arrangement puts functional groups near each other or near the substrate, thereby enabling many resulting interactions that in turn will be modulated by the physical properties of the active site environment. On this direction, it is not surprising that many scientists aimed to achieve "enzymatic efficiencies" just by mimicking rigid active site models [183, 184]. That is, without accounting for enzymatic conformational dynamics. Far from the notion of proteins as rigid objects with a single structure assigned, however, we now know that enzymes are inherently flexible. Indeed,

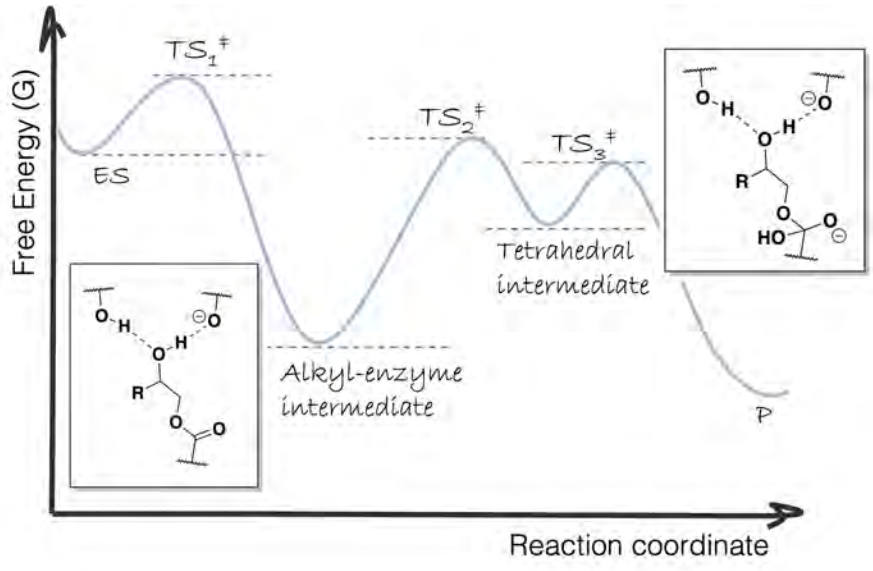


Figure 1.13: Illustration of *Bm*EH two-step energy profile together with the alkyl-enzyme and tetrahedral intermediates schemes. ES stands for the reactant enzyme-substrate complex, $TS_1^{\ddagger}$ stands for the epoxide ring-opening reaction transition state leading to the alkyl-enzyme intermediate, $TS_2^{\ddagger}$ stands for the first-half hydrolysis reaction leading to the formation of the tetrahedral intermediate, and $TS_3^{\ddagger}$ stands for the transient stage leading to the final diol product P. For clarity, only a very simplified representation of the two intermediates are here depicted. More details are also provided in Chapter 4.

enzymatic functions are significantly affected by motions that facilitate substrate diffusion, binding and release, or even by changes that further strengthen the non-covalent interactions during binding. In a general view, it is overall accepted that protein dynamics plays a role on protein functions. However, the association of these motions with catalysis it is also a topic highly debated [185–187], as the exact mechanisms of the involvement of such changes in enzymatic mechanisms is not entirely clear. In particular, there is a large controversy regarding whether conformational changes directly contribute to acceleration, that is, to crossing the activation barrier of the chemical reaction [152, 187, 188].

1.4 The Need of Capturing Dynamics

The static view of proteins started to change when the induced-fit model was postulated. As we have already explored in Section 1.2.2, the enzyme undergoes limited conformational changes to create a better fit between its structure and that of its substrate. Up to date, numerous studies have demonstrated that different structural changes take place in proteins,

supporting the idea of these macromolecules being more than rigid objects with a single structure just assigned [189]. This framework let to a fundamental shift in our understanding of enzymatic behavior, and conformational dynamics became another piece of the puzzle to understand enzymatic mechanisms.

Proteins are flexible entities that are better described as an *ensemble* of multiple conformations, sub-states or *microstates* assumed to be in equilibrium, that arise due to their ability to fluctuate within a large range of times (10^{-15} to 10^{4} s), amplitudes (0.01 to 100 Å), and energies (0.1 to 100 $kcal \cdot mol^{-1}$) [190]. As it is illustrated in Fig. 1.14, many of the inherent motions in proteins occur on timescales of central biochemical processes [74]. For instance, vibration and local motions (10^{-15} to 10^{-10} s) correspond to chemical events that occur during enzyme-mediated catalysis, including the formation or breaking of covalent bonds, hydrogen bonds formation, or the transfer of electrons, protons or hydride ions between chemical groups (10^{-12} s or 1 ps). In addition, it also includes larger or slower conformational changes, such as side-chain, secondary elements, and domains motions (10^{-9} to 10^{-3} s), including relevant processes such as water structure re-organization (10^{-8} s), ligand binding (10^{-8} to 10^{1} s), or allostery (10^{-5} to 1 s).

The functional implication of motions occurring around 10^{-8} to 10^{-3} s involving the conformational change of residue's side-chain, secondary structure elements, other segments of the polypeptide chain (e.g. hinges), as well as whole domains motions have been known and documented for some time [78, 191, 192]. This is because such motions account for a wide range of events involved in enzyme efficiencies, such as the creation of gates and tunnels to give access to substrate or promote product release [193], or even those enzymatic motions that make possible the interconversion of short-lived conformational changes that need to occur during catalysis. These type of motions assist enzyme catalysis by facilitating the formation of multiple substrate intermediates [194]. However, suggesting an effect on enzyme catalysis by thermal vibrations (10^{-13} s) and fast motions taking place around 10^{-12} to 10^{-9} s, such as changes in individual non-covalent interactions *via* side-chain rotations, is what have often been more controversial [187]. One of the problems with the proposal that fast motions contribute to enzyme catalysis is the difference between the timescales of these motions and the turnover times of most enzymes (10^{-6} to 10^{-3} s). Yet, this does not necessarily exclude enzyme motions having an effect on catalysis, since in many enzymes

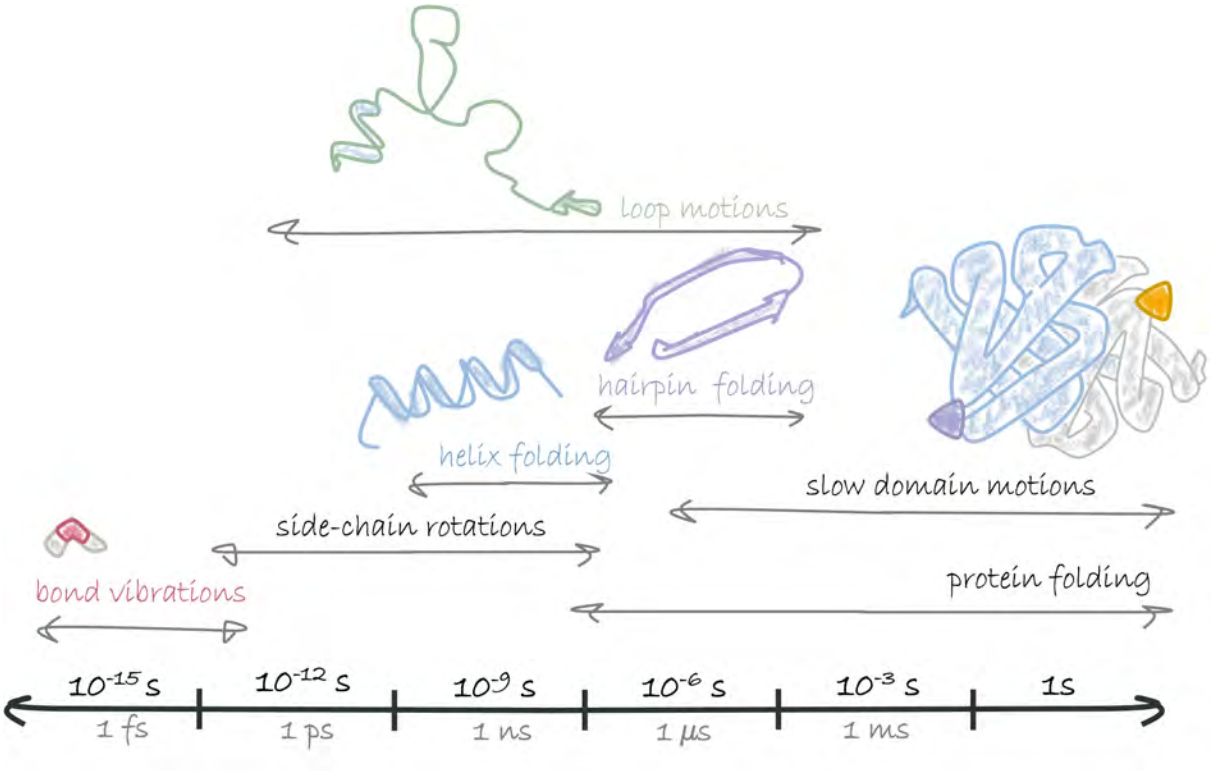


Figure 1.14: Biological time-scales of protein motions.

the rate-limiting steps are substrate binding or product release, rather than the chemical step(s) involved in overcoming the activation energy barrier of the reaction [195]. This is a complex topic and a more detailed discussion is beyond the scope of this thesis. Nevertheless, understanding the structural bases of such dynamic processes is important for realizing the full spectrum of macromolecular function and, eventually, for predicting and engineering protein behavior [196–199].

1.4.1 Proteins are Conformational Ensembles

Each protein exists as an ensemble of conformations that undergo continuous exchange within several spatial and temporal scales (Fig. 1.15). In general, the different sub-states constituting a protein's ensemble share the same fold, as well as the same secondary structure, although they differ in their detailed atomic coordinates [118, 200]. This is particularly true when the energy barriers separating each sub-state are low (within the same ensemble) [121].

As proteins *in vivo* are constantly fluctuating and adopting averaged structures (or a

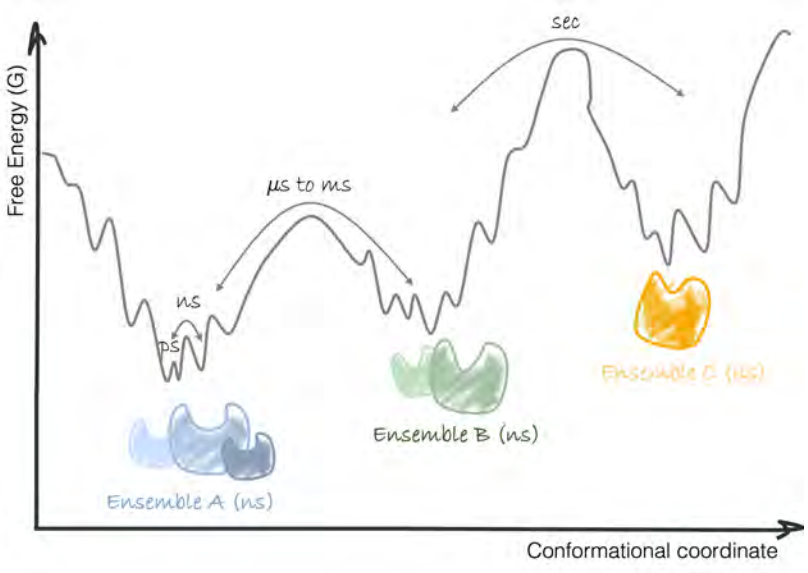


Figure 1.15: As a result of thermal energy, proteins tend to adopt an averaged structure (or set of structures) that can be generally defined as minima in the energy landscapes (A, B, and C). The three different conformational states are shown, as well as the energy barriers separating them [201, 202].

set of structures) as a direct result of thermal energy, there is an inherent uncertainty about the particular state a macromolecule *is* or will evolve to at any given moment. In this context, a detailed description of protein structure and dynamics will demand using *probability distributions* (statistics), as the number of sub-states available to each protein is indeed huge [14].

From *statistical mechanics*, therefore, proteins are best described by the *equilibrium populations* of each state *via* relative free energies (Eqn. 1.13), as well as by the rates and pathways of inter-state conversion.

$$P \propto exp - \frac{\Delta G}{RT} \tag{1.13}$$

This treatment of proteins finally culminated in what is nowadays pretty well-established as the *free energy landscape* framework [203, 204]. In general, the protein is expected to spend most of the time occupying those sub-states that have the slowest free energy (usually referred to as the "native state"). The ensemble that has the least energy is considered to be the most stable one, although the protein is also expected to sample other sub-states that are also of

low energy, but of higher energy than the native one. Worth to notice is that the probability (P) of the protein to acquire a certain sub-state depends inversely and exponentially on its free energy (ΔG). This means that even a small difference in the free energy of the states of the system can be translated as a large change in its probability, thereby inducing a shift from the less stable to the more stable state (e.g. from ensemble A or B to C in Fig. 1.15).

A full account of conformational flexibility would in principle require a reasonable characterization of this available structural space, at equilibrium, and the underlying *free energy surface* that governs accessibility of protein structures and the corresponding transitions between them [205]. On this basis, great advances have been made in the wet and dry laboratories to elucidate macromolecular structure and dynamics. Protein configurations sampled often represent a small fraction of the configurational protein space, and therefore, the *sampling* of the conformational landscape remains a challenging problem. One solution for studying slow (ms-s) to intermediate (ns-μs) motions comes in the form of biochemical methods that allow scientists to "trap" individual protein sub-states and characterize them. Some of these experimental methods include techniques that are also used for structural determination, such as X-ray crystallography, as well as nuclear magnetic resonance (NMR), which has provided significant insights on the structural motions occurring on multiple time frames over the course of a protein life span. Single molecular structures are usually captured from crystallography, specifying the type and location of each atom, in addition to the connectivity. In contrast, NMR provides an ensemble of protein structures that are consistent with the spatial constraints [66].

In addition to experimental methods, computational tools such as molecular dynamics (MD) simulations have provided invaluable information regarding protein motions. Indeed, MD simulations allow to sample numerous configurations of the system, in addition to offer a few important advantages over laboratory techniques. First, it is significantly faster and cheaper compared to experimental methods. Second, it provides atomically detailed descriptions of the simulated systems. Third, all the intermediary states are available provided that protein sub-states are within the timescale covered by the MD simulations (i.e. 10^{-12} to 10^{-9} s). As MD simulations relay on highly explicit models of the system (an *all-atom* description), most MD simulations as a consequence often cover a short period of time that ranges from *ps* to *ns*. Nevertheless, due to recent progress done in supercomputing, such as

the Anton machine [206] and GPU resolution, which has enabled researchers to access to longer timescales, simulations still represent a primary tool for elucidating the conformational dynamics of proteins. Particularly when the data derived from these calculations is combined with data from experimental methods. In fact, MD simulations has commonly served to bridge the gaps between the structural information provided by X-ray crystallography and the kinetic data obtained from NMR techniques. And finally, MD simulations are also capable of describing faster molecular events up to the *ps* timescale. This is because motions occurring on the fs timescale involve processes such as the formation or breaking of covalent bonds, and thus quantum mechanical (QM) calculations are required in order to account for the movement of electrons. Further details will be given in Chapter 2.

1.4.2 Shifts Induced By Environmental Changes

Given a particular protein sequence and environmental conditions, the number of possible configurations leading to a relatively stable structure is huge. However, by using any explicit function of the enzyme coordinates relevant to the process of interest, this high dimensional conformational space can be simplified into one or two-dimensional representation of energy landscapes. If we accept this notion of proteins existing in solution as a broad range of conformational states that are populated based on their relative energies: $G_i \sim -k_B T log(P_i)$, we would not be surprised at all by the fact that any variation on the environmental conditions would induce a *redistribution* on the relative populations or such conformational states. That is, a *population shift* due to the presence of substrate, an amino acid replacement, or for instance, by any variation in temperature or pressure values.

On this direction, there is nowadays a lot of evidence confirming that conformational dynamics leading to different conformational states are indispensable to enzyme function throughout the catalytic cycle, in substrate recruiting, chemical transformation, and product release [207]. In general, conformational transitions may involve induced-fit and conformational selection pathways, although the last one has gained popularity over the years. Within the first model, there is a redistribution of the relative populations in the accessible conformational states after the binding event. In addition, it overlooks the notion of proteins existing in different interconvertible states in the absence of any ligand (and how their ratios are determined by the thermal equilibrium). In contrast, the conformational selection model,

which originated from the earlier allosteric Monod-Wyman-Changeux (MWC) model of cooperativity [112], considers that enzymes can pre-exist in multiple conformations in solution [208]. And a nice example pointing out the role of higher in energy and functionally relevant conformational states is the work provided by Kovermann and co-workers [209]. Although adenylate kinase (AdK) represented a model example of induced-fit theory, through the introduction of a disulfide bond the authors succeeded some years ago in arresting the enzyme in the catalytically competent closed conformation. Until before, in the absence of any ligand, AdK was trapped in its open conformation. Nevertheless, their experimental strategy confirmed that AdK enzyme binding was not governed by induced-fit, as the higher in energy closed conformation required for catalysis was also sampled in the absence of any ligand.

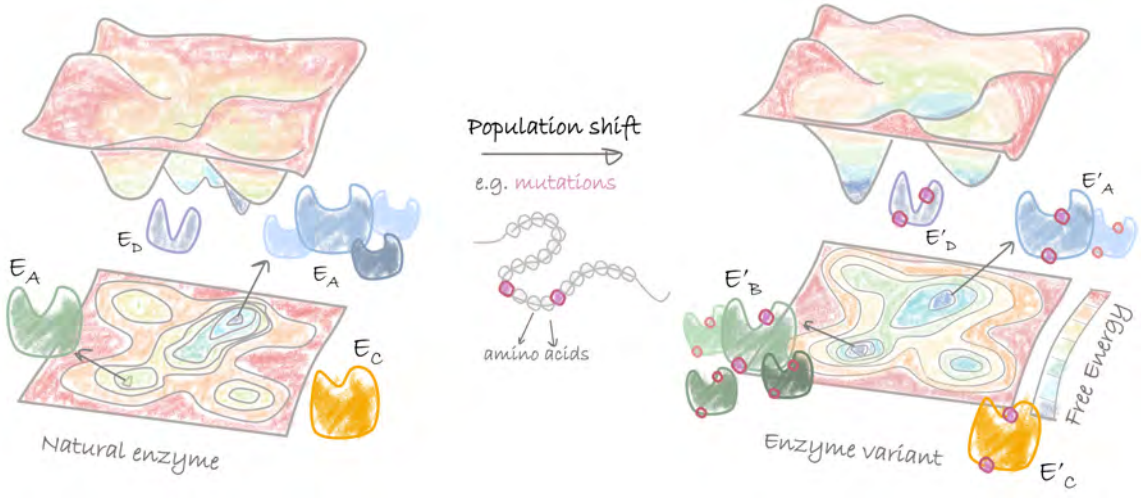


Figure 1.16: Schematic representation of an enzyme three-dimensional conformational free energy landscape and the population shift induced by mutations. Relative populations of each enzyme conformation (i.e. A, B, C, and D) are qualitatively represented by the amount of enzyme conformations. Active site and distal mutations are indicated as pink circles.

As it is illustrated in Fig. 1.16, enzymes have the ability to evolve or change their native functions by gradually stabilizing minor (and higher in energy) conformational states upon the introduction of active site and distal mutations. That is, a population shift happens after the exchange of some amino acid residues. On this direction, there is a remarkable work by Tokuriki and Jackson in which they show a gradual evolutionary change from a phospotriesterase (PTE) enzyme into an arylesterase (AE) functionality [199]. Another recent and insightful study showing how directed evolution (DE) gradually alters the enzyme conformational ensemble to populate a highly active conformational state that dramatically

accelerates the enzymatic reaction is the evolution of Kemp eliminases provided by Hilvert and Kern [210]. The Kemp elimination is a well-studied model for proton transfer that has served as a benchmark for *de novo* design. While distant mutations contributed to this fine-tuning, two active-site residues played an outsized role in sculpting a steric and electrostatic environment conducive to transition state stabilization. Overall, these findings provide a direct and quantitative demonstration of how modulating protein conformational landscapes can speed up a simple chemical reaction, although it is still not something optimized by current design protocols.

Some studies have shown that shifting the population toward the active state, like the PTE into AE example mentioned before, is one factor in increasing the catalytic efficiency during evolution. Indeed, as we have already explored, functional plasticity has emerged critical to explain the ability of enzymes to display novel functions, either through the repurposing of the catalytic residues or through the emergence of novel active sites in promiscuous or ancestral enzyme scaffolds [211]. Yet, whether computationally designed or uncovered in activity screening, enzymes repurposed for biocatalysis do usually present low or unexisting initial activity. In some cases, rate enzymatic enhancements by many orders of magnitude have been reached through directed evolution (DE). Indeed, as pointed out by the award of the Nobel Prize in Chemistry in 2018 to Frances H. Arnold, this strategy has shown to be a highly versatile and powerful tool to meet many challenges within the protein engineering field [212–214]. A key component of experimental techniques based on Darwin's theory of natural selection is the principle that mutations occur *randomly*. That is, species have to wait for the right mutation to come along through random processes, occasionally making the evolved variants *fitter* than their ancestors. And therefore, variations in our genetic code are not generated in response to an environmental change. This is indeed one of the strongest points of DE techniques, the fact that mutations are randomly introduced to the *entire* protein scaffold. However, how these modifications to enzyme structure make such a dramatic difference on catalytic efficiencies has remained something of a mystery. In fact, it is worth mentioning that most of these mutations are deleterious, thereby leading to inactive variants. In other words, in most of the cases, you end up screening for the active enzyme variants from a huge pool of inactive ones. In addition, single mutations are not commonly observed to greatly increase enzyme efficiency, whereas in some cases the combination of two mutants from all the huge library gives rise to rate enhancements due to synergistic effects.

For some of these reasons, approaching the enzyme design as a population shift problem has gained popularity over the years, and a more *rational* design of ligands and mutations targeting the population shift, conformational transitions, fluctuations, and flexibility of proteins shall be core practice for the development of new enzymes and other biomolecular systems [215]. As nicely shown by Kamerlin and Sanchez-Ruiz [216], a tremendously proficient Kemp eliminase enzyme was mainly accomplished due to the conformational variability of the ancestral enzyme. Remarkably, starting the design process from this ancestral scaffold only the insertion of one to two mutations were required for significantly enhancing the Kemp elimination reaction, by 7 orders of magnitude as compared to the uncatalyzed reaction. This rate enhancement is much higher than the previously computationally designed Kemp eliminases [217], but still not above the most proficient Kemp eliminase reported to date by Hilvert (HG3.17). This new variant displays nine orders of magnitude rate enhancement as compared to the uncatalyzed reaction, and was obtained as a result of 17 rounds of directed evolution starting from a designed protein which was already active [218]. In this line, Chica and co-workers studied the changes in the conformational ensemble during evolution of the designed Kemp eliminase HG3. This information based on the changes induced on the protein conformational dynamics due to evolution was then used to engineer a new Kemp eliminase HG4 with a much more pre-organized and rigidified active site for efficient catalysis [219]. These works confirm the relevance of taking advantage of the conformational dynamics of enzymes for designing novel *in-silico* functionalities [220, 221].

Finally, similar to enzyme optimization by random library screening, which remains sluggish, assessing the magnitude of the maximum possible catalytic activity by computational approaches is also still one of the great challenges in the field. Yet, opposite to initial *in-silico* designs focused at the active sites and transition state stabilization, many recent computational approaches are now routinely including the dynamics along the process [188, 220, 222, 223]. Nevertheless, it is worth to note that rationally approaching the enzyme design process as a population shift problem is computationally expensive, as it ideally requires the reconstruction of the free energy landscapes of natural and evolved enzyme variants. So, after all, how feasible it is to computationally construct free energy landscapes and predict which positions can affect and modulate enzymes functions? As shown by many recent examples provided by our group, residue-by-residue correlation and proximity bioinformatic tools such as the

shortest-path-map (SPM) can be used for evaluating and predicting positions that can impact such long-range allosteric networks of interactions [12, 224].

Chapter 2

Methodology

In this Chapter 2, the fundamental concepts behind the quantum and classical methodologies that have been used to carry out the studies included in this thesis will be briefly explained. All energy profiles calculated in the mechanistic studies of *Bacillus megaterium* epoxide hydrolase (*Bm*EH) have been done using the density functional theory (DFT) approach, so Section 2.3 is dedicated to provide the basis of quantum mechanic (QM) calculations. Afterwards, the conformational dynamics of *Bm*EH enzyme and its single-point mutation variants (i.e. F128A and M145A) have been studied using MD simulations based on classical force fields, which are generally described in Section 2.4.

2.1 Quantum and Classical Frameworks in a Nutshell

In a general view, classical mechanics provides a very straightforward picture of the world we see around us. It is tuned to our intuitions, as it involves familiar concepts such as speed, acceleration, momentum and forces. This area of physics governs the behavior of ordinary objects, as it tells us precisely how the *location in space* of any relatively large

object changes with time. And thus, the behavior of macroscopic objects is quite predictable. In quantum mechanics, there is no such a thing because there is a *decoherence filter* that draws an apparent distinction between "what we see" and "what really is" at the microscopic level. In other words, a filter in which the behavior of the atoms and elementary particles is "washed away" in the constantly vibrating thermodynamic interiors of macroscopic objects.

Unlike in classical mechanics, the *state* of the entire quantum system is not *just* given by the position and the velocity of each of its moving parts. Instead, quantum systems are described by something less concrete: *a cloud of probability*. And therefore, the best we can do is to predict the probability of "seeing particles", like electrons, in any particular values of location *or* with any particular values of velocity. As Heisenberg's uncertainty principle says, either position or momentum (mass times velocity) could be certain in an appropriate quantum state, though they cannot be certain simultaneously. Interestingly, that cloud of probability is widely referred to as the *wave function* ψ, since the most probable measurement outcome (e.g. the position of electrons or a fixed wavelength, and thus, momentum) changes, oscillates, like a wave over time (Fig. 2.1).

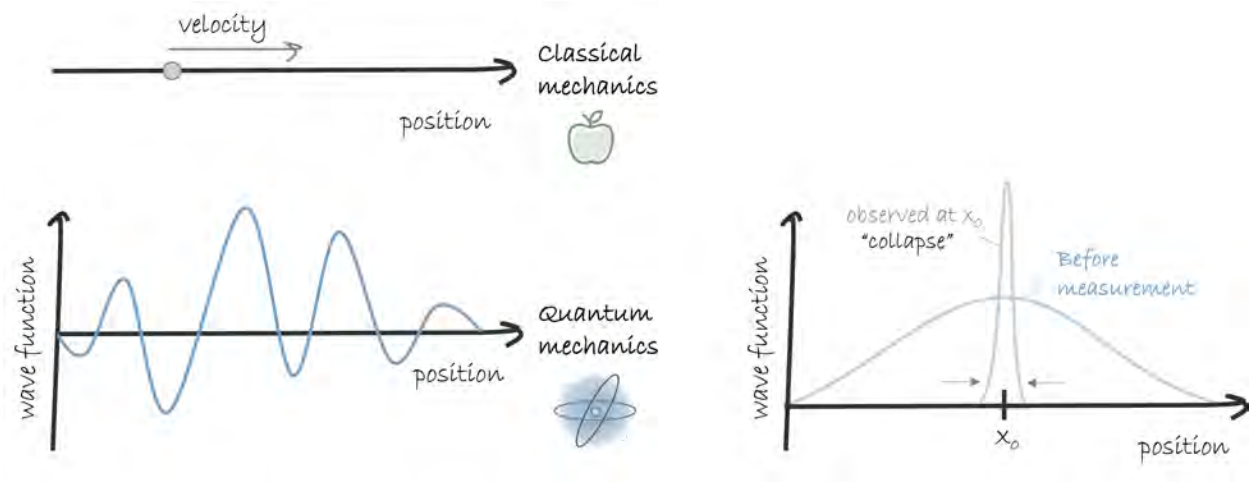


Figure 2.1: Representation of the basic rules of classical and quantum mechanics.

As shown in Fig. 2.1, the wave function describes any variable of the quantum system. In particular, the wave function assigns what is defined as the *amplitude* associated with a particular outcome (e.g. $\psi(x_0)$). And it is the amplitude *squared* which finally gives us the probability of getting a specific outcome when we perform a measurement. Different explanations have been proposed to describe the moment in which it is generally said that

the wave function has "collapsed" (e.g. at a given position) and returned to our classical and familiar perception of the world. Notice that there is more than one interpretation of "why" quantum systems appear to behave so different as they behave. Although this is a complex problem and a more detailed description is beyond the scope of this thesis, the main idea is that this intrinsic uncertainty along with probabilities are core features of quantum physics. In addition there are also other events that set the quantum world apart from the classical world, such as quantized energy levels, wave-particle duality, coherence, entanglement or correlation, as well as tunneling, which all of them are usually perceived to us as impossible situations [225]. Interestingly, a thought experiment often used to illustrate this phenomenon in quantum mechanics is the well-known "paradox" of Schrödinger's cat. In short, it shows how a quantum object's positions (momentum or velocity) can exist as a combination of infinite possibilities *until* you measure it. That is, until you observe its locations or how fast it is going and the wave-function collapses. Once that happens, all combinations vanish and position (or velocity) can assume concrete values. In the context of the Schrödinger's paradox, that is, the cat is either dead or alive. Nevertheless, the wave function can be understood conceptually as the *exact* representation of the quantum system, in a similar way than a set of positions and velocities would represent a classical system. And therefore, on this sense, its mathematical formulation can be viewed as logical and consistent as the more intuitive and familiar laws of motion that were proposed by Newton.

2.2 Computational Chemistry Tools for Enzyme Modeling

Significant advances in the field of computational programs and algorithms have contributed to the evolution of computational (bio)chemistry. In fact, one of the current strengths of computational modeling and simulation relies on its ability to explore, guide, and predict the behavior of many different kinds of chemical and biological systems.

There exists a large number of methods offering the possibility to study enzyme reactivity and their dynamics at different levels of complexity. As it is illustrated in Fig. 2.2, *in-silico* approaches are typically classified according on which approximation is used to solve the Schrödinger equation that describes the energy and the structure of the system. That is,

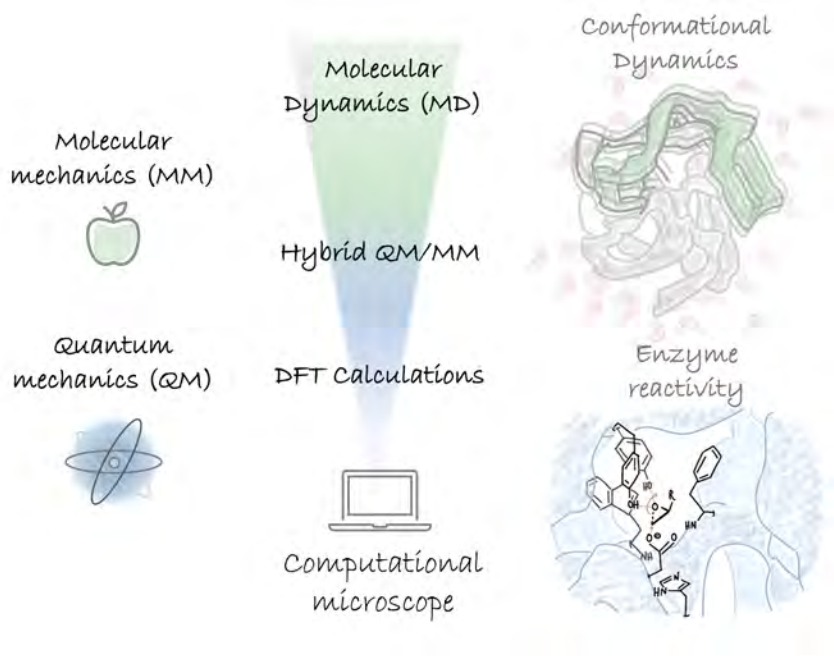


Figure 2.2: Representation of different computational approaches at the atomistic while modeling enzymatic properties.

from the most rigorous to the less complex: quantum mechanics (QM), *hybrid* quantum mechanics/molecular mechanics (QM/MM), and molecular mechanics (MM) methods. In a general view, the QM level of this "virtual computational microscope" refers to the use of quantum mechanics-based descriptions that go right down to the microscopic level, what allow us to study chemical reactivity (i.e. the electrons movement and therefore, bonds being cleaved or formed), whereas the MM level refers to the use of simple potential-energy functions to model molecular systems [13]. As long as we know the analytical expression of the potential energy surface, the dynamics of such system can be also studied analytically. However, except for trivial cases such as two-body systems, the numerical solution of the Schrödinger equation of a *many-body* system is not possible at the QM level, and MM methods offer a computationally much cheaper and also less time-consuming alternative for studying large biomolecules, for example. The term in the middle is referred as QM/MM, which is a combination of the two previous treatments. These type of methods are widely used while studying enzymatic mechanisms, as it allow to reduce the computational cost associated with QM calculations by splitting the system in two distinct parts: a smaller region, usually the substrate and a selection of residues in its surrounding, which is the one treated at the QM

level, whereas a larger region including all remaining parts of the enzyme, usually the ones that are not directly involved in the formation or cleavage of bonds, are treated just classically using MM and what we also define as *force field* methods. In general, therefore, depending on which is the problem under study, as well as the computational resources available [12, 13], we will end up using a classical, a quantum or a combined approach.

2.3 Quantum Mechanics Theory and Calculations

In a general view, quantum mechanics features two things: i) a wave function and, ii) the Schrödinger equation 2.1, which is the mathematical expression that governs how the wave function evolves along time. Throughout this equation, Schrödinger proposed the idea that the *rate* at which the wave function is changing is proportional to how much energy it has, being in its most popular form written as it follows:

$$\frac{\partial \Psi}{\partial t} = \frac{1}{i\hbar} H\Psi \tag{2.1}$$

The left-hand side is the rate at which the wave function is changing over time, whereas on the right-hand side, there is a proportionality constant involving Plank's constant $\hbar$, which is the fundamental unit of quantum mechanics, and i, the square root of minus one. Notice that in Eqn. 2.1, the wave function Ψ is "acted on" by something that is called the *Hamiltonian* H, which is an *operator*. And therefore, this H operator can be viewed as an inquisitor who asks the following question: "How much *energy* do you have?"

$$H\psi = E\psi \tag{2.2}$$

In the above equation the energy E is an *eigenvalue*, whereas the term ψ is defined as an energy *eigenfunction* of the Hamiltonian operator. Conceptually, the message behind Eqn. 2.2 is that "if we know the Hamiltonian operator forms, then we can know everything", as it represents a compact way of capturing all the properties of a physical system. Nevertheless, except for trivial cases, the analytical solution of the Schrödinger equation is not a possible task. This is why there exists different approximations dealing with it, and depending on which approximation is used to solve the Schrödinger's equation, each possibility suggests a particular kind of Hamiltonian, as the form of a universal H is not known.

From the most rigorous to the most approximated approach, that would be, *ab initio*, semi-empirical and MM methods. At the current level of accuracy, there are mainly methods based on wave functions. Broadly speaking, these define the energy of the system, and its molecular properties, in terms of the wave function. An alternative approximation to these conventional wave function-based methods are density-based methods that describe the energy and properties of a physical system as a function of the *electron* density (ρ). Remarkably, such density functional theory (DFT)-based methods reduce the N-body problem into a single-body problem, thereby representing a dramatic simplification to deal with the Schrödinger equation resolution.

2.3.1 Fundamentals of Density Functional Theory

Modern DFT can be viewed as a "recipe" of quantum mechanics substantiated on the Hohenberg-Kohn (HK) theorems [226]. The first theorem of this reformulation relies on the fundamental idea that the energy of the ground state of any *many-electron* system can be completely determined and expressed through the electron density: $E[\rho_{gs}]$. In other words, the electronic energy of the ground state, as well as any other property of a system, can be accurately calculated as a *functional* of the unique electron density of such system. A functional is defined as any mathematical expression that can act on a function. In that case, however, to retrieve a number of the energy as exemplified in Eqn. 2.2.

The electron density only depends on the three-spatial coordinates: $\rho(r)$ where $r = x, y, z$, and so the many-body problem is in turn simplified. This is because, opposite to the wave function, which depends on the coordinates of *all* electrons and nuclei, the electron density does not depend on the system size. In this context, however, we still need to figure out "which is the functional" that connects the electron density of the system with its energy. A question that has been indeed one of the still challenging and unresolved tasks within the DFT framework. Indeed, one of DFT's main limitations arises from the fact that the functional connecting these two quantities is not known.

The second theorem states that the electronic density of a given system is the one that minimizes the electronic energy of the ground state by following the *variational principle*. In other words, it states that the energy calculated from a *guessed* electron density ($\rho_{guess}(r)$),

which might be different from the ground state, is an upper bound to the *real* one [227]. That is:

$$E[\rho_{guess}(r)] \geq E[\rho_{gs}(r)] = E_{min} \tag{2.3}$$

Interestingly, the energy functional E can be divided in three different contributions. That is, the kinetic energy $T[\rho]$, the attraction between the nuclei and electrons $V_{Ne}[\rho]$, and the electron-electron repulsion term $V_{ee}[\rho]$:

$$E_{DFT}[\rho(r)] = T[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] \tag{2.4}$$

In addition, the $V_{ee}[\rho(r)]$ term can then be divided into Coulomb the term, $J[\rho(r)]$ and the exchange-correlation energy, $E_{xc}[\rho]$), which contains the non-classical self-interaction (exchange and correlation effects). The Born-Oppenheimer approximation decouples the movement of the nuclei, which is slower, from the movement of the electrons being much faster. In other words, it assumes that electrons cloud rapidly adapt to a position of the nuclei. Under the Born-Oppenheimer approximation, therefore, the nuclei-nuclei repulsion contribution is not included, and so we end up with an expression for the energy as it follows:

$$E_{DFT}[\rho(r)] = T[\rho(r)] + V_{Ne}[\rho(r)] + J[\rho(r)] + E_{xc}[\rho(r)] \tag{2.5}$$

The foundation for the general use of DFT methods in computational chemistry came when Kohn and Sham introduced orbitals and formulated an expression for the kinetic functional energy under molecular orbital (MO) approximation:

$$\rho(r) = \sum_{i=1}^{N} |\phi_i(r)|^2 \tag{2.6}$$

In this KS-DFT approximation, the $T[\rho(r)]$ term can be divided into two contributions: the kinetic energy of a non-interacting system ("KS system") with the same electron density as the real system, with an additional correction resulting from the difference between the real kinetic energy (T) and the non-interacting system (T_{ni}). As this latter term is included into the $E_{xc}[\rho(r)]$ term, then we get at the following expression:

$$E_{KS-DFT}[\rho(r)] = T_{ni}[\rho(r)] + V_{Ne}[\rho(r)] + J[\rho(r)] + E_{xc}[\rho(r)] \tag{2.7}$$

In the above equation, it is important to mention that the only term that remains elusive is the $E_{xc}[\rho(r)]$ term. And therefore, the main difference between DFT methods arises from how they account for the exchange-correlation part. As it is shown by the huge number of different existing DFT methods, figuring out the adequate description of the exchange-correlation energy has been a common challenge. Yet, two of the most common approximations include the local density approximation (LDA) and the generalized gradient approximation (GGA). The first is considered the most simple description, as the value of $E_{xc}[\rho]$ is estimated at a given position as a functional of an electron density that is at the same local position. In other words, the LDA functionals *solely* depend on the local value of the density ρ at each point r in space. The second one is considered a more sophisticated description, as the energy also depends on derivative on the electron density function. A very popular hybrid GGA functional is B3LYP [228, 229], which expresses the exchange term as a combination of LSDA (local spin density approximation) exchange, pure Hartree-Fock (HF) exchange and an exchange term developed by Becke [229]. Regarding the correlation, this contribution is expressed as a combination of LSDA correlation and a term developed by Lee, Yang, and Parr [230].

DFT *hybrid* functionals such as B3LYP have been highly recommended in the study of reaction mechanisms. Indeed, its accuracy has been extensively assessed against benchmarks, being the method of choice even to treat even relatively large models, such as the quantum mechanical (QM) active site models of enzymes [144, 231–234] (Section 2.3.2). Of course, B3LYP have also its deficiencies, as well as DFT methods in general. One of its critical and well-known limitation being its inadequate description of dispersion forces [235, 236]. However, these are solved a posteriori by adding an empirical correction to the energy [237, 238]. Conventional KS-DFT approaches lack the ability to model accurately non-covalent interactions, which are weak forces that arise from electron correlation and electron density fluctuations. Unlike covalent bonds, non-covalent forces do not involve the sharing of electrons, but rather include different types of interactions arising from nearby atoms (or molecules), including from electrostatic interactions, π-effects, van der Waal or dispersion forces to hydrophobic effects. And this is relevant, because in spite of being defined as "weak", these forces can have an impact on the geometries and energies of modeled systems. Specially, in relatively large systems like enzymatic active sites.

Practical Considerations using DFT

As KS-DFT had become a central methodology in quantum chemistry and weak interactions are ubiquitous and very important in chemical processes, it is not surprising that many efforts have been done in order to enhance the reliability of these methods while describing dispersion energies and therefore, reliable correlation effects. Conventional DFT functionals usually underestimate the dispersion energy values, thereby requiring corrections to obtain more accurate results. There exists a huge variety of dispersion-corrected KS-DFT approaches, and more detailed technical specifications are beyond the scope of this thesis. Nevertheless, it is worth to mention one solution in particular. That is, the incorporation a posteriori of empirical corrections to the total energy, which is an approximation that allows to increase the accuracy without increasing the computational cost [239]. Although trends might remain unchanged, adding dispersion corrections when it is necessary is quite important, as these considerations can directly affect the values of the computed energy barriers, also the optimization or detection of some intermediates, and how they compare to experiments. In the present thesis, the approximation used to account for dispersion forces has been the atom-pairwise DFT-D3(BJ) protocol by Grimme, in which the Becke-Johnson (BJ) damping function is used instead of the standard zero-damping [237, 238]. For reactions where steric repulsions and weak interactions are critical for selectivity, it is not only important to include a certain amount of exact exchange, but also these dispersion additive corrections.

In modeling chemical bonding, specially if we want to adequately describe the electronegative or positive character of every atom of the system, we might also want to account for the uniform distribution of the charge in our system. In such situations, additional functions that describe polarization of the electron density of the atoms in molecules are included to the previously described Kohn Sham orbitals represented in Eqn. 2.6. For example, the minimal basis set for hydrogen atom is one function approximating the 1s atomic orbital, while a simple polarized basis set typically is composed by two s- and one p-function. In practice, this additional description has increased the flexibility of the basis set by allowing molecular orbitals (MO) involving the hydrogen atom to be more asymmetric about the hydrogen nucleus. As basis sets typically come in hierarchies of increasing size, these additions are translated into more accurate outcomes (we become closer to the infinite basis set limit). However, at the expenses of a higher computational cost. And therefore, it is important to

consider in advance the usefulness of each basis function and if it is worth including it for the system under study. In this regard, in the present thesis all geometry optimizations of *Bm*EH active site models were carried out using 6-31G(d) basis set, which is a valence double-zeta polarized basis set that adds to the 6-31G set five d-type Cartesian-Gaussian polarization functions (including polarization functions for all non-hydrogen atoms). As a large amount of valuable information can be derived from the energies computed for each of the chemical species involved, more accurate energy values were computed using the larger 6-311+G(2d,2p) basis set on the previously optimized structure. Notice that in addition to polarized function, it also includes diffuse functions for heavy atoms (indicated with the symbol "+" in Gaussian Software [240]). These type of functions are important for describing dipole moment, and specially become mandatory for describing anionic species along the reaction mechanics (as it is for the case of *Bm*EH).

2.3.2 Density Functional Theory for Computing Reaction Mechanisms

Theoretical mechanistic investigations strictly require to locate and characterize all chemical structures along a reaction coordinate in a *potential energy surface* (PES) (Fig. 2.3) [241]. Within this PES, the atoms are rearranged, in many multiple ways, to finally convert substrate molecules into products thereby creating new bonds or breaking existing ones.

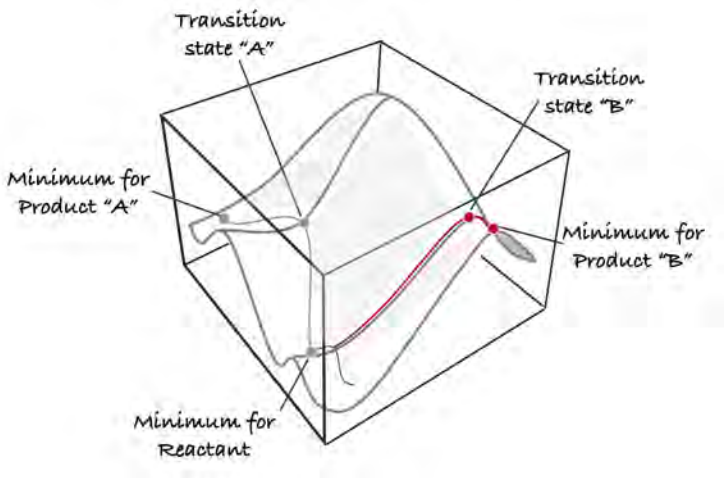


Figure 2.3: Representation of a model potential energy surface (PES) with the most chemically interesting stationary points indicated. This Figure is adapted from [242].

As shown in Fig. 2.3, in general, two different types of *stationary points* are usually found in the PES along a given reaction coordinate: minima and first-order saddle points. Reactant(s), product(s), and intermediate(s) are characterized for being *local* minima along the corresponding reaction coordinate. This means that all eigenvalues of the *Hessian* matrix are positive values. In other words, that the second derivative of the energy with respect to all atomic positions are real values. Interestingly, transition state(s) are characterized for being maxima on the PES but only in the direction of the reaction coordinate, and therefore, for being first-order saddle-points. In that case, all eigenvalues are also positive but one. Interestingly, it is this *negative* (or imaginary) value the one describing the pathway (transition state) connecting substrate(s) and product(s).

Due to the nature of both stationary points, the procedure for locating reactant(s) and product(s) optimized geometries are slightly different. To find a minimum of the PES we need usually only to evaluate the gradient, which is the first derivative of the energy, and follow its slope downhill until it reaches a value equal to zero. To do so, the potential energy and the gradient are commonly computed starting from an initial or "guess" geometry of the system. Then, the geometry of the system changes by moving downhill, giving rise to a novel geometry. At this point, the cycle starts again, and lasts until during the optimization process the change on the geometry and the energy of the system is below a certain defined threshold values. Finally, the Hessian at this point is evaluated to confirm whether it is a real minima on the PES by having all the eigenvalues positive. To find transition state(s) the gradient is also used, although now it is also necessary to include the second derivatives of the energy during the procedure. As building the Hessian matrix at each geometry optimization step is computationally expensive, a common solution to address this limitation is computing the second derivative at the first geometry optimization point, and then updating it following a certain scheme, such as the Bofill updating method [243]. After conversion is reached, the real Hessian is estimated to confirm that one eigenvalue is negative and that the optimized structure is a TS.

2.3.3 Quantum Mechanical Models to Study Enzymatic Reactivity

Most computations for chemical reactivity are done using DFT-based methods, since these have been observed to provide accurate energy values and geometries with medium-sized or considerably large basis sets. The applicability of QM to enzymatic systems has been fueled over the last decades thanks to the development of faster and computational methods offering higher accuracies and thanks to the more efficient computer resources [244]. The cluster approach has been the method used in this thesis to study *Bm*EH mechanism and selectivity trends, as it is considered a valuable technique for elucidating and modeling enzymatic reactivity. In short, it relies on the use of relatively accurate quantum chemical methods that can predict bond strengths, physicochemical properties, as well as spectroscopic features, from a reduced representation of the overall enzyme.

Practical Considerations using Cluster Quantum Approach

Under this cluster approach, a well-chosen part of the enzyme's active site is selected, which is treated explicitly at the QM level. The main focus relies on those active site residues that are expected to capture most of the relevant interactions, whereas the rest of the enzyme is truncated and not considered in the modeling [144, 231, 234].

In a general view, a QM-only or all-QM model, or cluster model approach (all terms are synonyms) the active site is typically designed on the basis of some available crystal structures of the enzyme. Although there is no rigid protocol, it is considered a good practice to start with models that include at least the first shell of residues surrounding the substrate. In practice, however, the selection of the groups to be included in the model will essentially depend on the problem we want to address and the computational power available. Typically, the fixed atoms are the carbons where the truncation has been made, and if it is necessary, the position of one or two of the hydrogen atoms that have been added to saturate that carbon are also (occasionally) maintained. Nevertheless, the "locking-scheme" applied will strongly depend on the system itself, as well as its required flexibility. In general, including enough flexibility (by increasing the system size) can avoid wrong energy profiles affected by artificial strain effects. Indeed, it is well-established that models around 250-300 atoms typically provide good energy profile results [144].

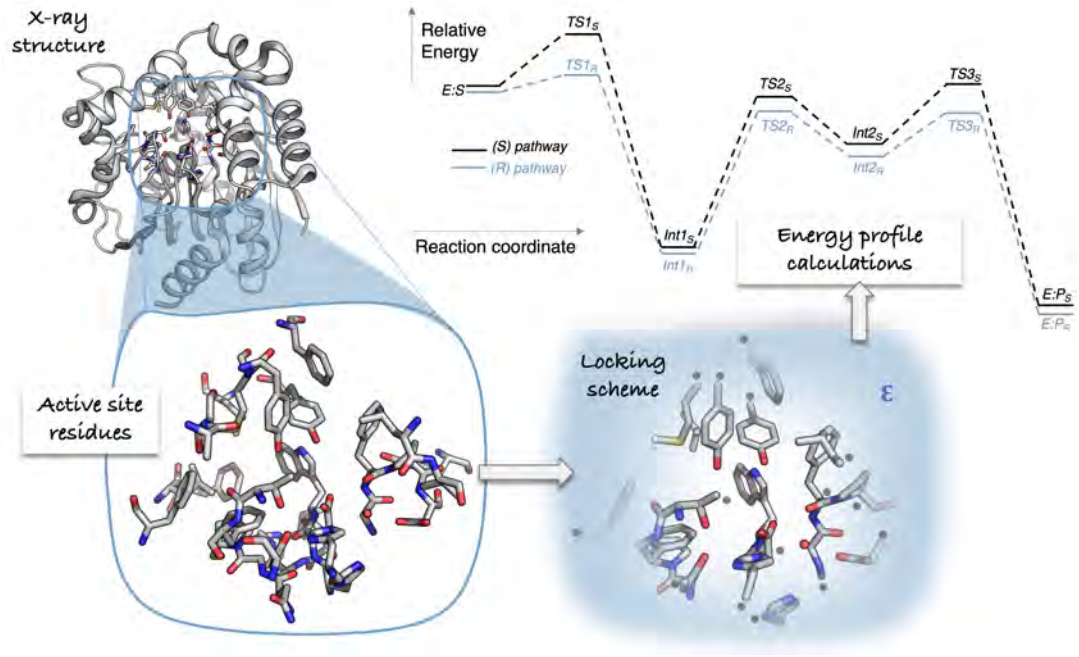


Figure 2.4: Schematic representation of the steps followed to build a quantum chemical cluster model of enzymes active sites. A model of the active site is designed on the basis of a crystal structure (e.g. PDB ID: 4NZZ). At the edge of the model, some positions at the edge of the model are maintained fixed during geometry optimizations, and the surrounding is modeled implicitly as a continuum solvent with a concrete dielectric constant value. Intermediates and transitions states are located, and their energies are used to evaluate the origins of enzymatic preferences derived from calculated mechanisms. This Figure is adapted from [245].

Notice from Fig. 2.4 that the enzyme-substrate complex (ES) is generally the starting point while drawing mechanistic pathways. This is because this approximation considers that the substrate binding as a reversible process, in addition to assume that the release of the product is neither the rate-determining step. In practice, this means that those processes occurring before and after the chemical step(s) are not explicitly considered in the cluster model calculations. Interestingly, overlooking these processes do not affect the study of enzyme mechanism itself. Yet, only because it is also implicitly assumed that entropy effects are quite small in the chemical steps, and therefore, entropy influence can be ignored.

Thanks to the developments on computer hardware, nowadays models routinely reach around the 300 atoms. Yet, it is worth to notice that *increasing* the model size does not always necessary imply more accurate and reliable outcomes. This is generally because, above

300 atoms arises what is often called as the *multiple minima* problem. In such situations, it is even more remarkable the highly recommended practice to test-and-error the size of the QM model before starting any mechanistic study. At the end, what we need to ensure before starting any chemical reactivity study is that from all different substrate rotamers and active site groups combinations, the enzyme-substrate (ES) structure we will use is indeed the one showing the lowest-energy binding mode [144]. Otherwise, artificial movements can take place between various stationary points, which is commonly translated to wrong energy profiles (and thus, wrong mechanistic conclusions) [234]. Another important consideration while building active site QM models is including the crystallographic water molecules adequately. Otherwise, models can "collapse", either because an integral part of the active site structure is missing or due to the presence of artificial voids and cavities that leads to significant errors in the cluster calculations. And finally, another aspect to consider in the cluster model is the *protonation state* of titratable groups, that is, the region of every residue in a protein that can uptake or release protons in an aqueous solution. In principle, every amino acid in a protein can be considered as titratable residue. Nevertheless, we should take special care for those that can be either protonated or deprotonated depending on the local surroundings. Typically, on physiological conditions, glutamate, aspartate, and histidine residues are the ones having a direct impact on the energetics of the reaction usually acting as catalytic residues. For this reason, we need to evaluate both scenarios in case the protonation state of a residue is not clear from the structure, as it can happen with the protonation state of the catalytic triad histidine in some EH enzymes [246].

Although current QM models can contain a considerably large amount of atoms, we have to keep in mind that they still represent a reduced description for the overall enzyme, and thus some additional assumptions have to be made while modeling enzymatic transformations. For instance, as it is represented in Fig. 2.4, the electrostatic influence of the enzyme surrounding is addressed by the use of implicit solvation models, whereas the steric influence that the rest of the enzyme imposes on the active site is modeled by fixing a number of selected atoms, that is, by using a coordination "locking scheme".

Additional Analyses for Studying Reaction Mechanisms

Detailed knowledge about enzymatic reactivity and *selectivity* can provide useful information for the rational design of enzymes as alternative routes to produce highly pure *chiral* molecules.

In general, elucidating the factors that govern enzymatic mechanisms, and in particular of enantioselective enzymes, is often seen as challenging tasks. Due to the inherent chirality of enzymes, their use as biocatalysts can potentially help us to kinetically resolve, for instance, racemic epoxide mixtures. As explored in Chapter 1, by invoking transition state theory (TST), the enantiomeric excess (*ee*) can be extracted from $\Delta\Delta G^{\ddagger}$ values. The cluster model approach is one of the methods that allow us to compute such energies, although there are other already mentioned methods, such as those based on treating the system at QM and MM levels or EVB (e.g. [247]).

In addition, other analyses can be done in order to get further mechanistic insights. For instance, the measurement of geometric parameters in the optimized structures, such as distances, angles, dihedrals between the active site residues and the substrate, is a common practice to evaluate the *pre-organization* of the enzyme's active site. Alternative analyses are based on the measurement of non-covalent interactions, which can be of great value to complement and understand, at the molecular level, the energy profile outcomes. For instance, the analysis of such active site's interactions can provide clues to elucidate the sources of a better transition state stabilization.

In this thesis, the NCIPlot program has been used to study non-covalent interactions (NCI) on the *Bm*EH active site TSs optimized geometries [248]. Due to the size of macromolecules, the most common approach has been to assign van der Waals (vdW) interactions, steric clashes (SC), and hydrogen bonds (HBs) based on pairwise distances between atoms according to their vdW radii. NCI analysis is applicable to large enzymatic systems by using self-consistent fully (SCF) quantum mechanical and promolecular densities. Indeed, NCI index identifies interactions in a chemical system solely on the basis of the electron density and its derivatives. SCF densities are constructed from the wave function information obtained from the wfn file generated from the electronic structure program used (Gaussian09 in our case), whereas promolecular densities are constructed from the atomic positions stored in the xyz coordinate file(s). Once these densities are written as simple sums of exponential functions, the NCI surfaces can be calculated very efficiently for each system, as all the data required can be obtained analytically. Interactions can be selectively displayed, according to different criteria such as the strength of the interaction, its localization in space, and its nature, thereby providing an easy and visual way to detect NCI and compare between optimized geometries.

2.4 Classical Mechanical Modeling and Simulations

The tremendous impact of the conformational dynamics on protein properties and functions has fueled the development of many theoretical methods and a large number of algorithms to assess problems like cellular transport, signal transduction, allostery, cellular recognition, ligand-docking, and enzymatic catalysis [249]. Protein-based systems often include, in addition to the protein itself, atoms of the solvent, as well as any other chemical species, such as ions and cofactors.

At any given moment, each of the atoms occupies a single point in space, giving rise to different system configurations, that is, ensembles of conformations. Due to the large size of protein-based systems, such characterization at the quantum mechanics level is computationally expensive and time-consuming, although it is indeed currently feasible. Consequently, most simulations rely on *molecular mechanics* (MM), as it approximates molecular systems by using empirical force fields potential energy expression, and in combination with Newton's classical mechanics. From this classical perception, in principle we could say that at a given moment the enzyme will have a classical state defined by a position and a velocity. In addition and according to Newton, if no forces act on it, the enzyme will continue to move in a straight line at constant velocity. However, if a force *does* act on it, this will cause some changes, or what we usually call an *acceleration*, in the velocity of the object in direct proportion to how much force is applied. So, in order to figure out the entire trajectory of any large object like enzymes, we then just simply need to specify its position, its initial velocity, and what forces are acting on it. Afterwards, Newton's equations tell us the rest.

These are basically the fundamental ideas behind classical mechanics, and thus, behind molecular dynamics (MD) simulations based on classical *force fields*. At this level of accuracy, quantum effects are ignored, and so, we cannot study the formation or breakage of bonds. However, the conformational dynamics of enzymes can be efficiently studied through the analysis of the movement of every single atom making up their three dimensional structures. As the method of choice in the present thesis to study the conformational dynamics of *Bm*EH enzyme at the atomistic level, a more detailed explanation of the MD steps followed will be provided in the present Section.

2.4.1 Molecular Mechanics Force Fields in a Nutshell

Within the current all-atom MM treatment, all atoms of the system are explicitly considered, and the energy of the system is described by using a mathematical equation, a *force field*, which treats atoms and covalent bonds as balls and springs, respectively. Within this approach, quantum effects are neglected, and the movement of large systems on potential energy surfaces can be simulated. In general, force fields that have been designed to treat relatively large protein-based systems include several separate terms, each of them describing a *potential energy* resulting from all covalent bonds and non-covalent interactions in a single configuration of the system:

$$U_{tot} = U_{bond} + U_{non-bonded} \tag{2.8}$$

where U_{tot} is the total potential energy, U_{bond} is the potential energy resulting from properties of covalent bonds, $U_{non-bonded}$ is the potential energy resulting from electrostatic interactions, non-polar interactions, and van der Waals interactions. Within this framework, therefore, the total potential energy of the system (U_{tot}) can be captured by classical interaction functions, each referring to a different type of atom-atom interaction. For instance, Hook's law is commonly used to describe bonded interactions, while the non-bonded atoms might be as inelastic hard spheres or interact according to a Lennard-Jones potential. CHARMM, GROMACS and AMBER refer to computer simulations packages, but also to three common molecular dynamic force fields designed for running simulations of biomolecules. AMBER force field has been the force field used in this thesis, and its expression forms are particularly given as it follows:

$$\begin{aligned} U(\vec{r}) &= \overbrace{U_{bond} + U_{angle} + U_{torsion}}^{bonded} + \overbrace{U_{vdw} + U_{electrostatic}}^{non-bonded} \\ &= \sum_{i}^{bonds} k_{r,i}(r_i - r_{eq,i})^2 + \sum_{i}^{angles} (\theta_i - \theta_{eq,i})^2 + \sum_{i}^{dihedrals} \sum_{n} \frac{\upsilon_{n,i}}{2}[1 + \cos(n\phi_i - \gamma_i)] \\ &+ \sum_{i<j}^{atoms} (\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}) + \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} \end{aligned} \tag{2.9}$$

- The **first term** (U_{bond}) contains a sum over all the bonds and gives the potential energy increment as a consequence of a distortion of the covalent bonds distances from the equilibrium bond distances. The $r_{eq,i}$ term is the equilibrium bond distance of the bond

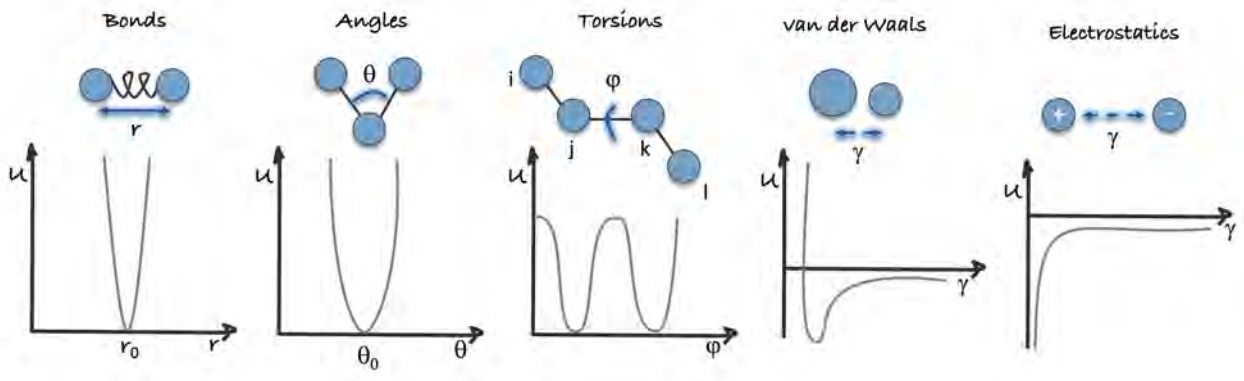


Figure 2.5: Representation of molecular mechanics models consisting of spherical atoms connected by springs that represent bonds. Simple potential energy functions, such as harmonic oscillator or Coulombic potentials, to model molecular systems are also illustrated. Forces experienced in the model structure are described using these simple mathematical expressions.

i and $k_{r,i}$ is the associated harmonic force constant to that particular bond. In essence, this means that force constants and equilibrium distances will vary significantly from "one type of atom-atom bond" to another. And therefore, different forces and distances values are expected for single C-C bonds, C-N bonds, as well as for double or π C=C bonds. After all, force constants and equilibrium distances will not be the same, but they will depend on what is know as the "atom type". In other words, these parameters will be only shared between molecules with the same type of bond (e.g. Csp^3-Csp^3 bond in ethane and butane, but not for Csp^2=Csp^2 in ethylene). In practice, this means that experimental and computational data obtained from "small" molecules can be employed to define large systems such as proteins In other words, there is a transferability of the type of atoms. Finally, it is worth to mention that stretching potentials for a bond between atoms i and j are indeed given by Taylor series, although as shown in Eqn. 2.9, it takes usually the expression only of the first quadratic term. This is the reason behind their poor performance at large distances, which can be corrected by using what is known as Morse Potential [227].

- The **second term** (U_{angle}) contains a sum over all angles formed by two consecutive bonds and gives the potential energy increment caused by a distortion of the angles from the equilibrium angle. The $\theta_{eq,i}$ term is the equilibrium value of the angle i, and k_{θ_i} is the harmonic force related to that angle. As it is illustrated in Fig. 2.5, angle bending energy potential are often treated in a similar way than the previous bond or

stretching potentials. And therefore, the energy is assumed to increase quadratically with the displacement of the bond angle from equilibrium (i.e. the first term of the Taylor series).

- The **third term** ($U_{torsion}$) contains a sum over all the dihedrals formed by three consecutive bonds, thereby giving the potential energy as a consequence of bond torsions. The $\upsilon n, i$ term is the harmonic force associated to the dihedral angle i, and γ is the phase angle. In that case, the total torsional energy is expressed as a Fourier series, which is a way of representing a periodic function as a (possibly infinite) sum of sine and cosine functions. As it is depicted in Fig. 2.5, the torsional angle θ is the angle between two planes defined by atoms i, j, and k and by j, k, and l. After all, the torsional energy what attempts is to capture some of the steric and electrostatic non-bonded contributions between atoms i and l, which are connected through an intermediate $j-k$ bond. It is worth to notice that the adequate description of torsional angles is particularly important in macromolecular modeling, as torsions are the main responsible of the different protein configurations.

- The **fourth term** (U_{vdw}) contains a double summation over all the atoms and gives the potential energy due to the van der Waals (vdW) between the atoms pairs. The R_{ij} term are the distances between atoms i and j, and A_{ij} and B_{ij} are the van der Waals constants. The energy associated to vdW interactions often arises from interactions between "electron clouds" around two non-bonded atoms (Fig. 2.5). This attraction is due to electron correlation that results in "dispersion" or "London" forces, that is, instantaneous multipole/induced multipole. At short distances vdW interactions are strongly repulsive, being close to an exponential increase of the potential energy, while at intermediate to long distances the attraction is proportional to $1/r^6$. Typically, these long range vdW interactions are the dominant cost of a force field. This is because, opposite to bonded terms, which are all proportional to the number of atoms (N), non-bonded interactions increase as the square of the number of the atoms (N^2). One natural solution to reduce the computational cost associated is by using a less expensive expression for this term such as the Lennard-Jones (LJ) potential. In addition, a popular way to speed up the calculation of non-bonded interactions is to define a cutoff radius. Typically a cutoff value equal to 8-12 $\mathring{A}$ from the atoms we are interested in is recommended, which means that all atom pairs whose distances are greater than the

established cutoff have an interaction energy equal to zero.

- The fifth and **last term** ($U_{electrostatic}$) also contains a double summation over all the atoms. In this case, however, the electrostatic interaction between the atoms pairs is provided, in which the q_i term is the electric charge of the atom i. Alongside with the torsional energy term, electrostatic interactions are the second term of special relevance. Nevertheless, under MM treatment it is not possible to describe how electrons are delocalized, as atoms do not have charges. On this regard, we need to use force fields that accounts for polarization by assigning electrons to atoms.

In a general view, a good parameter set should not only reproduce the experimental data for the molecules included in the training set, but also for those molecules outside it. There exists many different set of parameters that have been fitted to study different type of molecules. As stated before, AMBER force field allows the description of relatively large objects such as proteins that contain conventional amino acids, while the General Amber force field (GAFF), for instance, is the one commonly applied (in an automatic fashion) to describe a wide range of small organic molecules that are composed of H, C, N, O, S, P, and halogens.

One of the main shortcomings of MD simulations based on MM force fields comes from the quality of such force fields, which are often compromised between accuracy and computational efficiency. This is because, although the potential energy is expressed as a sum of classical potentials, still in energy minimization calculations and MD simulations, the computation of first and second derivatives of the energy respect to the atom coordinates is required. This can be done analytically for the energy expressions previously described. On this direction, choosing an appropriate energy function for describing the interactions of our systems becomes critical to a successful MD simulation. Nevertheless, as force fields are empirical, different parameters that will ultimately depend on the atom types that are required. This is why when using force fields, it becomes strictly necessary and very important to assign an atom type to each atom of the system under study. In general, this process is often quite straightforward when using common general force fields, except for those cases in which force fields lack the parameters of some of the atoms present in our system, such as in enzymes containing a metal atom or other type of cofactors. In those cases, the corresponding parameters need to be generated and provided.

2.4.2 Classical Trajectories using Molecular Dynamics

By using these simple models, a molecular dynamics simulation numerically solves in an iterative way Newton's equations of motion that allow structural fluctuations to be observed with respect to time. Solving the equations of motion usually requires the integration of Hamilton's equations of motion for each coordinate r_i and component of linear momentum p_i, which can be obtained from the definition of the Hamiltonian and the second law of motion.

In summary:

(i) The derivation of the first Hamilton's equation is given by:

$$\frac{dr_i}{dt} = \frac{\partial H(\mathbf{r},\mathbf{p})}{\partial p_i} \tag{2.10}$$

(ii) whereas the derivation of the second Hamilton's equation is provided by the following expression:

$$\frac{dp_i}{dt} = -\frac{\partial H(\mathbf{r},\mathbf{p})}{\partial r_i} = -\frac{\partial U(\mathbf{r})}{\partial r_i} \tag{2.11}$$

where $H(r,p)$ is the expression of the Hamiltonian, and $U(r_i)$ is the potential energy surface (PES) of the system.

According to Newtonian mechanics, an atom described as "a ball" would keep moving at the same speed and direction unless it experiences a force (F) acting on it that will perturb, accelerate or decelerate its current "classical state". In molecular assemblies, however, the force on each individual atom i is affected by its position relative to the rest of atoms:

$$F_i(t) = m_i a_i(t) \quad (i = 1, 2, 3, ..., N) \tag{2.12}$$

which is equivalent to:

$$F_i(t) = m_i \frac{d^2 r_i(t)}{dt^2} \tag{2.13}$$

Yet, if the total potential energy can be expressed as the sum of purely classical potential energies, then force field equations can be in turn differentiated to obtain forces on *each* of the atoms of the system since there is the following equality:

$$F_i(t) = -\frac{dU}{dr_i} \tag{2.14}$$

and therefore:

$$-\frac{dU}{dr_i} = m_i \frac{d^2 r_i(t)}{dt^2} = m_i a_i(t) \tag{2.15}$$

Based on classical physics, then forces on the atoms in their new positions can be propagated by using the second Newton's law of motion, since the rate of change in the *momentum* of a body is directly proportional to the applied forced F. This calculation leads to new values of positions and velocities at a time $t + \delta t$, which finally generates a specific biological system's *trajectory*, through an iterative procedure.

As the motions of all particles are coupled together, the calculation of the forces applied on a particle at a particular configuration has to be done at a *given* and *finite* time interval (e.g. Δt). This is because the *many-body* problem cannot be solved analytically, but numerically using methods equivalent to Euler method. In this context, the Eqn. 2.15 that describes the motion of a particle of mass m_i along the coordinate r_i can be expressed as the sum of its interactions with other particles. And therefore, once forces are derived from force field, then the corresponding accelerations can be numerically integrated, typically in small stages around the *fs* (10^{-15} s) range, to get new velocities and positions for the atoms of the molecular assembly at a new given and finite time (e.g. $t + \Delta t$). In other words, its numerical solution allow to determine the position and momentum of the system at step $n + 1$ from the previous position of the system at step n.

The essential idea is that by using finite difference methods, the integration is divided into many small stages. In fact, the calculation of a trajectory involves a huge number of such small steps, and therefore choosing an adequate time-step is another critical point to keep under consideration. In short, this is because the numerical error of the equations of motions is proportional to its value, and therefore, the decrease of Δt reduces the numerical error, although at the expenses of an increment of the computational cost associated. In general, a time-step equal to 1-2 fs is often recommended to cover in an efficient way what is defined as the *phase space*. That is, for mechanical systems, all possible values of position and momentum variables. In statistical mechanics, in addition to be considered as coordinates of this multidimensional space, position and momenta magnitudes also define what is known as a mechanical or *microscopic state* of a system. Interestingly, molecular simulations describe conformational changes at this microscopic level, and therefore, statistical mechanics become

necessary if we want to translate such information to macroscopic (experimental) observables.

States, Ensembles and Ensemble Averages

The dynamic properties of complex and biological systems are typically described with *statistical mechanics*. As the force field-based calculations used in classical MD simulations provide only the potential energy of the system, adequate sampling of each state is required in order to account for the entropic component of the free energy (G). The advances achieved in recent years in computational resources, as well as in the sampling methods and algorithms, have enabled all-atom MD simulations to become longer and more efficient on this regard [250]. Although classical MD simulations of such large systems aim to generate enough representative conformations to connect MD data to experiments, this technique is often unable to produce statistically meaningful configurational ensembles. In other words, there is often a sampling problem.

An ensemble is a collection of points of the *phase space* satisfying the conditions of a particular thermodynamic or macroscopic state of a system. In terms of thermodynamics, such macroscopic state is fully identified by values of a suitable set of parameters, such as the temperature (T), pressure (P), and the type and number of atoms (N_i). On this basis, there exist different type of ensembles depending on which variables are used to define the thermodynamic state:

- In a microcanonical ensemble (NVE), the thermodynamic state is characterized by a fixed number of the different type of atoms (N_i), a fixed volume (V), and a fixed energy (E). This macroscopic state is the one describing an isolated system.

- In a canonical ensemble (NVT), the thermodynamic state is determined by a fixed number of the different type of atoms (N_i), a fixed volume (V), and a fixed value of temperature (T).

- In an isobaric-isothermal ensemble (NPT) the thermodynamic state is determined by a fixed number of the different type of atoms (N_i), a fixed pressure (P), and a fixed temperature (T).

- In a grand canonical ensemble (mVT), the thermodynamic state is characterized by a fixed chemical potential (μ), a fixed volume (V), and a fixed temperature (T).

A molecular dynamics simulation generates a sequence of nuclear configurations in terms of the time that belong to the *same* ensemble. In statistical mechanics, an experimental observable of a particular property is given by its ensemble averages, which is defined as the mean of a property for all microscopic states of an ensemble, according to the distribution of the system on its microscopic states. Such ensemble average is typically taken over a large number of replicas of the system considered simultaneously. This leads to one of the widely accepted and established statement in statistical mechanics, the Ergodic hypothesis, which assumes that the *time average* of a property A obtained from MD trajectories is equal to the *ensemble average* of A:

$$< A >_{ensemble} = < A >_{time} \tag{2.16}$$

$$ensemble\, average = time\, average$$

which is true for a trajectory computed for an infinite time, as it will able to run (eventually) *all* over microscopic states of the ensemble. This is because, strictly speaking, the ensemble average of a property A is given by an integration over all possible microscopic states. Alternatively, and according to the Ergodic hypothesis, experimental observables are extracted from this time integral of all configurations described in molecular dynamics trajectories. And this is why one common strategy involves the computation of many replicas of the system under study rather than one long trajectory, with the final goal to generate enough representative conformations such that equality in Eqn. 2.16 is satisfied.

2.4.3 Setting Up, Running and Analyzing Molecular Dynamics Simulations

When we perform MD simulations we go through different steps with our system. Before starting the calculation, it is often necessary to establish an **initial configuration of the system** and build a realistic atomistic model. In case the crystal structure of the system under study has been resolved by X-ray (and NMR) technique, then initial coordinates are often taken from PDB files. Due to the inherent nature of hydrogen atoms, these are not located accurately and precisely by X-ray crystallography, and therefore, we have to **add** all these **hydrogens** to the system. On this regard, there are different methods available that predict and assign the protonation state of the residues in our initial coordinates by default,

such as H++ and PropKa servers [251, 252]. Yet, as protonation states of residues are key for processes such as substrate binding or catalysis, it is a good practice to ensure that the resulting protonated configuration of the system is adequate for addressing the problem under study.

In order to carry out an MD simulation the steps hereunder described are followed:

Step 1: Building an Atomistic Model of the System in Solution

In reality, biological systems are rarely isolated. Instead, enzyme-based systems are typically immersed in aqueous solvent, and **explicit solvent** water molecules should be included if we want to compare experiments and the data derived from simulations. One of the problems of considering explicitly solvent molecules is the drastic increase in the number of atoms (e.g. from around 9000 atoms without the solvent to 52000 atoms with solvent), and consequently, in the computational cost. Another critical problem related to this explicit solvent inclusion is how to perform a realistic simulation of the boundaries, as a correct treatment of such boundaries, as well as its effects, is what enables us to extract "macroscopic" properties from molecular dynamics trajectories.

An efficient approach to remove the effect of the box's surface are **periodic boundary conditions** (PBC), which are a set of conditions that are often used to simulate an infinite system by modeling a small system instead. In such a way, a simulation box (or unit cell) is surrounded by other boxes, which are exact copies of the simulation box. Under this treatment, particles experience forces as if they were in bulk fluids, since when an atom leaves the unit cell it automatically reappears on the opposite face of the simulation box with the same velocity. As it is replaced by an image particle of the central box, the number of atoms in the simulation box is overall conserved. And therefore, from a practical point of view, this operation is translated in no surface box effects anymore, being the entire system properly solvated. As far as the cell unit fills all of space by translational operations in the central box, the shape it is in principle not relevant. Nevertheless, a cubic shell is normally the description employed. In the present thesis, water molecules and counter ions have been added using the AMBER leap module. Explicit solvent is treated as rigid water molecules that only interacts through non-bonded interactions, and from the most popular models, the one adopted has been TIP3P, which is consistent with the SHAKE algorithm that constraints the hydrogen

bonds to fixed lengths.

Step 2: Protein and Solvent Minimization

Before running the molecular dynamic simulation the starting (experimental) configuration of the system has to be optimized using the chosen force field. This step is relevant, as it is essential to find a minimum on the potential energy surface (PES) in order to begin the MD simulation from such optimized geometry. During this step, it removes any bad contact between the atoms, allowing the system to reach a temperature near absolute zero Kelvin and the plausible minimum energy configuration.

In order to do so, it is necessary to **assign initial velocities** to the atoms. This can be done by using a random number generator from a Maxwell-Boltzmann distribution at the temperature of interest. Such temperature is defined by the average kinetic of the system, and depending at which temperature the simulation is running the system under study is able to cross over different energy barriers and visit multiple configurations. Having set up the system and assigned the initial velocities the molecular simulation to minimize the system can start, since at each step (typically in 2000 steps) the force on each atom can be calculated by differentiating the potential energy function. Once finished, as we are typically interested in simulating protein-based systems at room temperature, the whole system must be brought up to the temperature of interest, which leads us to the following step.

Step 3: The Heating Process

As we cannot directly start the heating from 0K to 300K, at this stage the system is heated progressively. Otherwise, the system can collapse. A common way to proceed is to heat the system every "N" steps to a higher value of temperature (e.g. every 1000 molecular dynamic simulation steps). In practice, this is done by assigning velocities at some low temperature and then running a short molecular dynamics. After a number of steps in the molecular dynamic simulations, the atom velocity scaled upwards, which in turn increases the temperature of the system. In general this process is done systematically, until the desired temperature has been reached. After heating the system at 300K, we now need to check for the variation of the ambient pressure and temperature, which according to our "real conditions", have to be stable with respect to time.

Step 4: Simulations at Constant Temperature and Pressure

To do so, a small **equilibration** calculation is done using an NVT ensemble and/or NPT ensemble. The first is the statistical ensemble that represents the possible states of a mechanical system in thermal equilibrium with a heat bath at a fixed temperature, the NPT ensemble is used in the equilibration step to check that we have the correct density of the system before generating the trajectory. Many of the methods used for pressure control are analogous to those used for temperature control, and therefore, the system can be also coupled to a pressure bath to maintain the pressure constant during the simulation (typically of a length of 1 ns).

Step 5: Production Run

Traditionally, the ensemble naturally used in molecular dynamics simulations has been the NVE (or microcanonical) ensemble. Nevertheless, experiments are usually done at constant pressure and/temperature, and this is why the two already mentioned NVT and NPT ensembles are the other most common alternative choices. In the present thesis, the NPT ensemble is the one adopted during the previous equilibration step, whereas the trajectory of the system is produced in the NVT ensemble. At this stage, the data obtained is collected to extract thermodynamic averages.

Step 6: Analyze Molecular Dynamics Data

The type of analyses adopted at this stage often strongly depends on the system under study and what information is intended to be derived from it. In other words, what problem needs to be addressed.

- **Visual Inspection.** In general, before doing any analysis, it is considered a good practice to look at our trajectory files using a molecular visualization program, such as VMD [253] or PyMOL softwares [254].

- **RMSD and RMSF Calculations of Atomic Coordinates.** The root-mean-square deviation (RMSD) is usually calculated, as it represents a simple way to ensure that all chemical and physical properties of the simulated system has reached an equilibrium. RMSD is commonly fitted on the backbone atoms, and using the starting structure

as a reference. Nevertheless, any structure of interest can be employed as a reference, and such deviation it is not only limited to the backbone atoms. In this line, the root-mean-square fluctuation (RMSF) can be also calculated, as it represents a simple tool to measure the rigidity of different parts of the protein, as different trends of flexibility can be easily identified. The main difference between RMSD and RMSF is that the latter is averaged over time, giving a value for each particle i, whereas for the RMSD the average is taken over the particles (thereby giving time-specific values).

- **Distances and Cavities in the Structure.** Although using force fields and MM formation and cleavage of bonds cannot be modeled, the measurement of distances between catalytic residues can also provide insights on the catalytic proficiency of enzymes [255]. In a more general view, computing the distances between two atoms or the minimum distance between two groups of atoms is a common and simple way to get information about contacts in the system structure. Usually, the corresponding plots of distances *versus* time are drawn, thereby making easier the task of confirming a relevant conformational change arising from formed/broken non-covalent interaction in the protein structure. In a similar way, the measurement of bend angles, torsions, as well as hydrogen-bonding patterns can allow us to identify conformational behavior at the secondary level of the protein structure.

The analysis of the pocket size and shape in the active site area is also of great relevance while studying receptor dynamics and changes in substrate preferences. In the present thesis, for example, the shape and volumes of the active site and binding pockets of selected representative structures of *Bm*EH wild-type and its variants have been measured by using an algorithm implemented in POcket Volume MEasurement (POVME) software (Chapter 5). This software accepts a multiframe PDB file as input, including the resulting MD simulation files, multiple crystal structure or NMR conformations. As POVME algorithm assumes that the pocket being measured does not translate or rotate in space, before any volume computation is strictly necessary to align the trajectory. In such a way, it is possible to successfully define a pocket-encompassing region that allow the subsequent identification of the pocket over all trajectory frames. In a qualitative way, the volumes can be graphically represented, thereby facilitating the analysis and comparisons of the active site shape between different enzymatic variants.

- **Mapping Conformational Free Energy Landscapes.** The resulting MD dynamic trajectory consists in a high dimensional data. And therefore, working with such big sized and complex systems demands the application of tools that allow to reduce the dimensionality of the data extracted from MD simulations. In such a way, the analysis of the conformational changes along the trajectory is in principle simplified.

  As the data from MD trajectories consists in a huge number of dimensions, that is, the dynamics of protein-based systems is typically defined by lots of motions and atomic coordinates. The calculation of the probabilities and the reconstruction of the free energy landscape (FEL) using the whole data set is not feasible at a reduced computational cost. One natural solution to identify relevant conformational configurations of the system under study is to construct free energy landscapes that only depends on a reduced set of global properties (i.e. collective variables (CVs)), while less relevant motions are equally distributed over the chosen CVs (e.g. backbone atoms distances, backbone dihedral angles, distances between catalytic residues). In such a way, all motions explored along MD trajectories can be then projected onto these *selected* properties, thereby obtaining the corresponding probability distributions and associated free energies for those particular features.

  It is worth to mention that making the proper *feature selection*, that is, the choice of a subset of the original set of CVs, to accurately describe and discriminate between the enzyme conformations without overlooking other relevant enzyme motions is a challenging task. A detailed knowledge of the biological system is required. And therefore, when little or no knowledge is available, one might be interested in using approaches that automatically reduce the dimensionality of the MD data set while accounting as much as possible for the rest of motions. On this regard, principal component analysis (PCA) and time-lagged independent component analysis (TICA) are popular linear methods used for that task [256–259].

  The main idea behind these approaches is that new representative variables (e.g. components) are constructed as linear combinations of the initial variables. In the case

of PCA, the dimensions of the data are reduced to maximize the variance. However, the motions that are usually associated to kinetically relevant processes (i.e. the slowest motions), do not have necessarily to correspond to the large amplitude conformational changes. Likewise PCA, the majority of the dimensionality reduction techniques do not include the evolution of the data along time. In contrast, TICA provides a reduced space by maximizing the auto-correlation of the data instead of the variance. Consequently, with this method one can minimize the loss of kinetic data. In the present thesis, TICA has been used as a dimensionality reduction technique, which seeks to represent the MD data in a two-dimensional free energy landscape according to the first two first TIC components. And this is why one of the critical problems related to these techniques is the selection of the feats employed as collective variables (CVs) in order to reduce the dimensionality. In case of not choosing an unbiased and adequate magnitude, the analysis will not lead to meaningful conclusions.

Chapter 3

Objectives

In Nature, enzymes are the most efficient catalyst known, however, many limitations arise once they are used as biocatalysts for manufacturing targets of great interest in industries. Due to their inherent properties, enzyme-based strategies have become an attractive alternative to conventional chemical catalysis, although it might not be still currently well implemented. Since many of the factors responsible for their proficient activity are still not well established, many efforts have been put forward to rationalize their catalytic performances.

For the particular case of epoxide hydrolases (EHs), their potential application as biocatalyst for asymmetric epoxide hydrolysis of epoxide racemic mixtures has encouraged their study and implementation as an alternative for the synthesis of target compounds like precursors of β adrenergic drugs. Up to date, many examples provide evidence in which the use of EHs can represent a suitable path in which chiral important precursors could be obtained from low-cost and also easy-to-access epoxide racemic mixtures.

From all EH enzymes, some features of the *Bacillus megaterium* EH (*Bm*EH) enzyme have particularly encouraged us to further explore and understand its selectivity towards aryl-based epoxide substrates. These include its unusual (*R*)-selectivity towards phenyl

glycidyl ethers (**PGE**) and its ability to accept bulkier substrates like naphthyl glycidyl ether (**NGE**), both compounds identified as relevant β adrenergic drugs precursors. Accordingly, the specific objectives of the present thesis have been:

In Chapter 4:

- First, to define the catalytic residues to be essential to investigate and understand *Bm*EH site- and stereoselectivities at QM level by using the cluster model framework.
- Second, to build a representative truncated model of *Bm*EH active site.
- Third, to get insights and rationalize the molecular basis of the *Bm*EH mechanism and selectivity using *para*-nitrostyrene oxide (*p*-**NSO**) as model substrate.

In Chapter 5:

- Fourth, to explore whether the *Bm*EH conformational flexibility plays a decisive role in determining the *Bm*EH stereoselectivity preference during the catalytic cycle.
- Fifth, to characterize the conformational free energy landscape of *Bm*EH and identify the major and minor conformational states.
- Sixth, to investigate and rationalize the effect of mutations F128 and M145 on the conformational landscape of *Bm*EH towards the acceptance of bulky substrates like naphthyl glycidyl ether (**NGE**).

As a final goal, we aim to provide the basis of the substrate control of the selective epoxide ring-opening catalyzed by *Bm*EH, with a keen eye on the future goal of predict and computationally design new enzyme variants with enhanced activities towards relevant bulky chiral compounds of pharmacological interest.

Chapter 4

Exploring the origins of selectivity in soluble epoxide hydrolase from *Bacillus megaterium*

This chapter corresponds to the following publication:
Serrano-Hervás, E.; Garcia-Borràs, M.; Osuna, S. "Exploring the origins of selectivity in soluble epoxide hydrolase from *Bacillus megaterium*", *Org. Biomol. Chem.*, **2017**, 15, 8827-8835.

Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2017, **15**, 8827

Exploring the origins of selectivity in soluble epoxide hydrolase from *Bacillus megaterium*†

Eila Serrano-Hervás, [a] Marc Garcia-Borràs *[b] and Sílvia Osuna *[a]

Received 26th July 2017,
Accepted 5th October 2017

DOI: 10.1039/c7ob01847a

rsc.li/obc

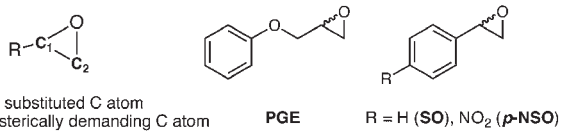
Epoxide hydrolase (EH) enzymes catalyze the hydration of racemic epoxides to yield their corresponding vicinal diols. These enzymes present different enantio- and regioselectivity depending upon either the substrate structure or the substitution pattern of the epoxide ring. In this study, we computationally investigate the *Bacillus megaterium* epoxide hydrolase (BmEH)-mediated hydrolysis of racemic styrene oxide (*rac*-SO) and its *para*-nitro styrene oxide (*rac*-*p*-NSO) derivative using density functional theory (DFT) and an active site cluster model consisting of 195 and 197 atoms, respectively. Full reaction mechanisms for epoxide ring opening were evaluated considering the attack at both oxirane carbons and considering two possible orientations of the substrate at the BmEH active site. Our results indicate that for both SO and *p*-NSO substrates the BmEH enantio- and regioselectivity is opposite to the inherent (*R*)-BmEH selectivity, the attack at the benzylic position (C1) of the (*S*)-enantiomer being the most favoured chemical outcome.

1. Introduction

Enantiomerically pure compounds are commonly identified as key synthons for the manufacturing of bioactive products in the pharmaceutical and agrochemical industries. Both epoxide-containing substrates and their corresponding 1,2-diol products are essential chiral building blocks of pharmaceuticals.[1–5] For instance, aryl glycidyl and naphthyl ethers are potentially useful compounds for the production of chiral amino alcohols such as (*S*)-alprenolol and (*S*)-propranolol β-blocker drugs.[6]

The resolution of epoxide racemic mixtures is an attractive synthetic strategy for obtaining optically pure bioactive compounds of pharmacological interest. To this end, many different strategies are available that include the use of metal or organo-based catalysts,[7] but also biocatalysts such as monooxygenases.[8] However, these (bio)chemical approaches are not very efficient, and usually offer moderate yields of pure enantio-enriched epoxides, thereby making them less useful for organic synthesis.[9] In pursuit of more efficient methodologies, the use of epoxide hydrolases (EHs) has emerged as a potential synthetic route due to some of the EH key properties such as: (i) it is a cofactor independent enzyme; (ii) it is found in a huge number of organisms; (iii) it is capable of operating in organic solvents; and (iv) it can exhibit high enantio- and regioselectivity.[10,11] In particular, *Bacillus megaterium* ECU1001 epoxide hydrolase (BmEH) has an inherently high (*R*) enantioselectivity towards phenyl glycidyl ethers (PGE, see Scheme 1),[12–14] for which some other EHs have only shown modest enantioselectivity or are (*S*)-specific.[15–17] The BmEH enzyme also exhibits an excellent enantiomeric ratio ($E > 200$) towards a *para*-nitro styrene oxide (*p*-NSO) compound.[12] For the latter case, it was found that the nature of the *para*-substituent switches both BmEH enantio- and regioselectivity, with the hydrolysis of the (*S*)-enantiomer *via* the attack at the benzylic position being favoured (see Scheme 1).[12,18]

The BmEH enzyme belongs to the α,β-hydrolase fold, consisting of a three-dimensional (3D) structure that possesses an α,β sheet core domain and a lid domain that caps the active site.[19–21] The EH members of this α,β-hydrolase superfamily



Scheme 1 Left, general representation of epoxide (carbon atoms labelled). Right, phenyl glycidyl ether (PGE) and the substrates used in this study: styrene oxide (SO), and *p*-nitro styrene oxide (*p*-NSO).

[a]*Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, 17003 Girona, Spain. E-mail: silvia.osuna@udg.edu*

[b]*Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095, USA. E-mail: marcgbq@gmail.com*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c7ob01847a

are characterised for having a highly conserved catalytic triad (Asp–His–Asp/Glu), two tyrosine residues positioned on the lid domain responsible for substrate recognition and for assisting the epoxide ring opening, and a conserved oxyanion motif consisting of HGXP residues (X = F in BmEH, see Fig. 1).[20,22,23]

EHs catalyse the addition of a water molecule to racemic epoxide-containing substrates to yield optically active 1,2-diols *via* biocatalytic hydrolytic kinetic resolution (HKR) and enantioconvergent hydrolysis of epoxides.[24–28] The mechanism by which these enzymes operate has long been debated.[29–33] However, it is generally accepted that EHs from the α,β-hydrolase fold share a common mechanism that takes place *via* a two-step mechanism through the formation of alkyl-enzyme intermediates (see Scheme 2, with BmEH labelling).[34–36] Upon initial binding of the epoxide substrate with Tyr144 and Tyr203 (see Fig. 2), the first step involves the nucleophilic attack by the Asp97 side chain at one of the epoxide carbons to generate a covalently bound alkyl-enzyme intermediate (**Int1** in Scheme 2). In the second step of the proposed mechanism, His267 acts as a general base to facilitate the nucleophilic activation of a water molecule to attack the Asp97 carbonyl, generating a new tetrahedral intermediate hereafter called **Int2**. The negatively charged tetrahedral intermediate is stabilized through hydrogen bond interactions between the backbone amide groups of the oxyanion hole residues Gly29–Phe30 and Asp97–Trp98. In the final step, the tetrahedral intermediate rearranges and dissociates to yield the corresponding 1,2-diol product (**Prod** in Scheme 2).

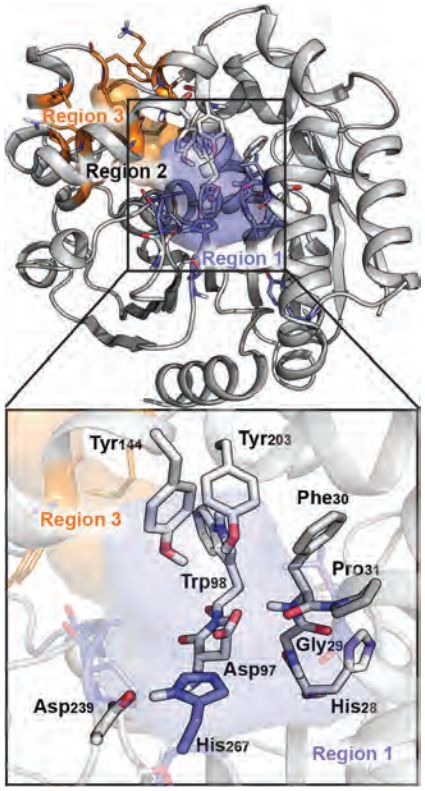
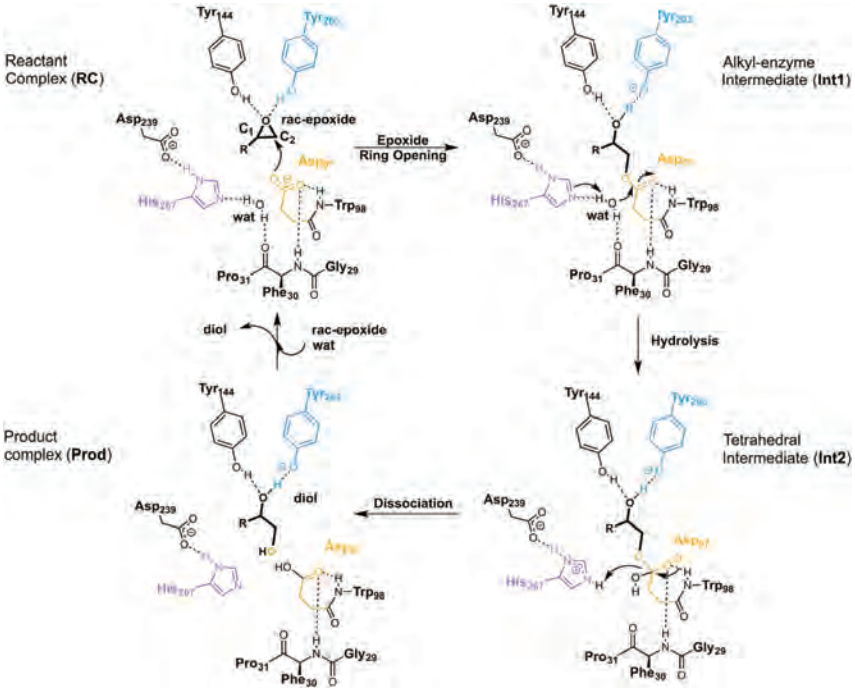


Fig. 1 Representation of the active site (region 2) of the BmEH enzyme (PDB: 4NZZ). The most important residues for the reaction are represented in sticks, and non-polar hydrogens are omitted for clarity. The BmEH region 1 (substrate entrance) and region 3 (product release) surfaces are highlighted in blue and orange, respectively. The two possible binding modes are also indicated as "region 1", *i.e.* substrate substituent pointing towards His267, and "region 3", substituent pointing towards Trp98.

Different computational approaches[37] have been used to explore the overall mechanism and selectivity of soluble EHs (sEHs). Truncated *theozyme* and cluster model calculations have been successfully applied to study the role and impact of enzyme active site residues in catalysed epoxide ring opening reaction mechanisms and selectivities.[38,39] Hopmann and Himo applied the cluster model (CM) approach on the X-ray structure of human soluble EH (sEH) to explore the role of the two conserved catalytic tyrosine residues.[23] Their results showed that one Tyr is enough for the alkylation reaction to occur, although the barrier increases by *ca.* 6.7 kcal mol^{-1} as compared to the wild-type. A drastic change in the activation barriers was observed for the double tyrosine mutant (*i.e.* larger barriers of *ca.* 24.8 kcal mol^{-1} were found after mutating both Tyr to Phe) indicating that this enzyme variant must be inactive. These results demonstrate that the presence of H-bond donors to activate the epoxide ring and stabilize the negative charge developed during the course of the reaction is essential.[40]

Amrein *et al.* carried out empirical valence bond (EVB) simulations of the enantio- and regioselective hydrolysis of *trans*-stilbene oxide (TSO) catalysed by *Solanum tuberosum* epoxide hydrolase I (StEH1).[30] EVB calculations suggested that the alkylation step for (*S*,*S*)-TSO at the benzylic C1 position is preferred by 1.7 kcal mol^{-1}, whereas the attack at the terminal C2 is favoured by 3.6 kcal mol^{-1} for (*R*,*R*)-TSO. However, high-energy barriers were found for both enantiomers for the hydrolysis step of the alkyl-enzyme intermediate that is formed after the Asp nucleophilic attack at C1 (**Int1** in Scheme 2). Therefore, the regioselectivity of the asymmetric epoxide ring opening of TSO by StEH1 is determined in the hydrolysis step, only the trajectories following the Asp attack at C2 for both enantiomers being productive. They also highlighted the relevance of considering a second active site histidine residue doubly protonated to properly describe the system. This histidine, which interacts with the nucleophilic aspartate, was thought to balance the negative charge developed during the catalysis.

In a very recent study, Lind and Himo applied the CM approach to explore the enantioconvergent resolution of racemic styrene oxide (SO) by using StEH1.[41] To investigate the origins of StEH1 selectivity, a very large CM system (of 279 atoms) was designed based on the StEH1 X-ray crystal in complex with the competitive inhibitor valpromide (PDB: 2CJP). The authors investigated the reaction mechanism considering both enantiomers, and examined the two possible binding modes of each enantiomer considering both the shape of the active site pocket and the substitution pattern of the substrate. Their results showed that (*S*)-SO is preferably attacked at C1 by the catalytic Asp, whereas for the (*R*)-SO



Scheme 2 General reaction mechanism of soluble EH enzymes. BmEH wild-type (PDB: 4NZZ labelling).

enantiomer the attack at C2 is favoured. Both cases lead to the formation of the (*R*)-diol, thus highlighting the enantioconvergent behaviour of StEH1 for the SO substrate. This is in contrast to previous results by Amrein *et al.* obtained for the TSO substrate, for which the reaction was favoured at the C2 position for both enantiomers. For TSO and (*S*)-SO substrates, the hydrolysis step was found to be the selectivity determining,[30,41] whereas for the (*R*)-SO substrate the alkylation step is found to determine the regioselectivity of the process.

It should also be pointed out that Lind and Himo investigated the role of the protonation state of the histidine placed close to the catalytic Asp in the catalytic reaction. Their computed energy profiles suggested that including the doubly protonated His104 has a minor effect on their computed energy barriers. These studies show how despite all mechanistic studies carried out to date, the fine details of the sEHs mechanism still remain incompletely understood.

As most EHs preferentially accept the (*S*)-epoxide enantiomer, the comprehension of how (*R*)-selective BmEH operates, and the main factors that contribute to its selectivity and efficiency is of great interest. In this study, we computationally investigate the origins of the enantio- and regioselectivity in BmEH towards *rac*-SO and its derivative *rac-p*-NSO using density functional theory (DFT) calculations within the cluster model framework. We have computed the full reaction profile using a consciously designed quantum CM from the BmEH active site (PDB: 4NZZ) consisting of 195 and 197 atoms for *rac*-SO and *rac-p*-NSO, respectively (see Fig. 2 and Computational methods). Our results show that styrene oxide and its *para*-nitro derivative switch the inherent BmEH enantio- and regioselectivity, and demonstrate the minor effect of the nitro group on the enzyme selectivity.

2. Results and discussion

The BmEH enzyme presents different features in comparison with other EHs. The BmEH crystal structure, shown in Fig. 1, revealed that this enzyme has an independent product-release site (called region 3) that is not found in other EHs.[6] The active site tunnel (region 2) connecting the substrate-entrance site (region 1) and the product-release site (region 3) was previously identified by an 80 ns Molecular Dynamics (MD) simulation.[6] BmEH exhibits opposite (*R*)-selectivity in comparison with many other (*S*)-specific EHs, thereby providing an alternative approach to obtain those optically pure (*S*) aromatic epoxides, commonly identified as building blocks of active β-adrenergic agents.[12] Many experimental studies reported on how the structure of the substrate or the presence of substituents on the aromatic ring moiety can switch the inherent selectivity of EHs.[42] The inherent (*R*)-selectivity that BmEH exhibits for phenyl glycidyl ethers (PGE) is switched towards (*S*)-selectivity when the styrene oxide (SO) substrate and its *p*-NO_2 derivative are considered.[12] In order to shed some light on the role of the substrate structure and its substitution pattern, as well as to explore the origins of BmEH selectivity, we performed DFT-D3BJ calculations of both racemic styrene oxide (*rac*-SO) and racemic *para*-nitro styrene oxide (*rac-p*-NSO). Based on the X-ray BmEH wild-type structure and our MD simulations on the alkyl intermediate **Int1**, we have devised a truncated cluster model (CM) as done by Himo in

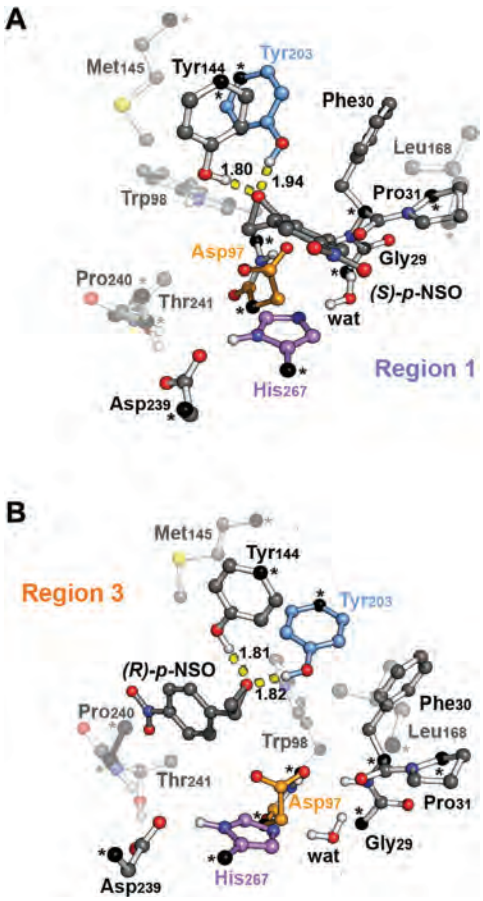


Fig. 2 B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) optimized geometries of the CM enzyme-structure complex (**RC**) in the active site model with (A) the phenyl substituent of (*S*)-*p*-NSO epoxide pointing towards the catalytic His267 residue (region 1, **RC-S-r1**) and (B) the phenyl substituent of (*R*)-*p*-NSO pointing towards Trp98 (region 3, **RC-R-r3**) (residue labelling as in BmEH). Active site pocket residues are shown as transparent sticks and spheres. Non-polar hydrogen atoms are omitted for clarity. The Asp97 nucleophile, the His267 base and the acid Tyr203 are highlighted in orange, violet, and blue, respectively. Atoms in black spheres and asterisks are kept fixed.

previous studies (see Fig. 2 and Computational methods in the ESI†).[29,43]

To investigate the origins of BmEH selectivity we have computed the EH mechanism described in Scheme 2, considering the first nucleophilic attack either at the epoxide-ring benzylic C1 or the terminal C2 positions for each enantiomer, but also accounting for the two possible substrate orientations regarding the BmEH active site tunnel regions (see Fig. 1 and 2).

Mechanism and selectivity of the BmEH enzyme with the *rac-p*-NSO substrate

The computed reaction energy profiles at the B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory for all possible outcomes of *rac-p*-NSO BmEH hydrolysis are represented in Fig. 3. The DFT optimized structures of intermediates (**Int**), products (**Prod**), and transition states (**TS**) for the attack at the benzylic position (C1) of (*S*)-*p*-NSO oriented towards region 1 are shown in Fig. 4.

From our computed energy profiles, small energy differences exist between both region 1 and 3 orientations of the substrate (*ca.* 0.6 kcal mol^{-1}) in the enzyme–substrate complexes (**RC**) for the (*S*)-enantiomer, whereas a difference of *ca.* 3.0 kcal mol^{-1} is found for (*R*)-*p*-NSO **RC**. As shown in Fig. 2, in the DFT optimized **RC** geometries, the epoxide ring forms hydrogen bonds with both Tyr144 and Tyr203 residues, and C1 and C2 carbons are well pre-organized for the subsequent Asp97 nucleophilic attack. The carboxylate side-chain of Asp97 is well positioned for the catalysis thanks to hydrogen bond interactions with the amide backbone groups of Asp97–Trp98 and Gly29–Phe30. The ion-pair charge relay system consisting of Asp239–His267 residues properly interacts for assisting the water molecule activation. In addition, the position of the nucleophilic water is maintained fixed at the active site by hydrogen bonds to the general base His267 and to the backbone carbonyl group of Phe30–Pro31 residues.

The first step of the reaction mechanism after the enzyme–substrate complex (**RC-S-r1**) consists of the nucleophilic attack of Asp97 either at the most substituted carbon C1 (inverting its configuration) or at the less sterically hindered C2 of the epoxide ring to form an ester intermediate (retaining configuration, **Int1-S-C1-r1**, see Fig. 4B). The optimized TS for the Asp97 nucleophilic attack at the C1 position in (*S*)-*p*-NSO (**TS1-S-C1-r1**) exhibits a 2.25 Å distance between the O-Asp97 side-chain and the epoxide C1 atom. The distance between the epoxide oxygen and C1 atom is elongated up to 1.88 Å (see Fig. 4A), corresponding to an S_N2-like concerted TS. The O–H distance of Tyr203 is slightly elongated (from 0.98 Å to 1.00 Å) in the alkylation TS, and the proton is completely transferred from Tyr203 to the epoxide O atom in the optimized covalently-bound enzyme intermediate **Int1-S-C1-r1** (see Fig. 4B). The associated reaction barrier for **TS1-S-C1-r1** is 1.8 kcal mol^{-1} relative to the reactant complex (**RC-S-r1**), and is 0.3 kcal mol^{-1} lower in energy than **TS1-S-C2-r1** corresponding to the attack at the less hindered C2 position (see Fig. 3A). These results show that the intrinsic preference of the enzyme is to perform the nucleophilic attack at the C1 position of (*S*)-*p*-NSO. The resulting **Int1-S-C1-r1** is stabilized by 18.4 kcal mol^{-1}. Regarding the (*R*)-*p*-NSO energy profile diagram, the lowest alkylation TS (**TS1-R-C1-r3**) has an energy of 5.5 kcal mol^{-1} relative to **RC-R-r1** (see Fig. 3B). Comparing both enantiomers, the lowest alkylation transition state (**TS1**) for each enantiomer differs by *ca.* 3.7 kcal mol^{-1}. Note that the orientation of the phenyl ring of the epoxide is different for each case (see Fig. 4A and Fig. S2†). For the (*S*)-enantiomer, the phenyl substituent points towards His267, *i.e.* here called region 1, whereas the lowest alkylation energy barrier for the (*R*)-enantiomer is found when in the **TS1** the phenyl substituent points towards Trp98 *i.e.* region 3 (see Fig. 2B). The analysis of both TS geometries indicates that the additional stabilization of **TS1-S-C1-r1** might be due to the π-stacking interaction between (*S*)-*p*-NSO and the His267 residue, which is not possible in the **R-C1-r3** case (see Fig. 4A and 5). Thus, at this first

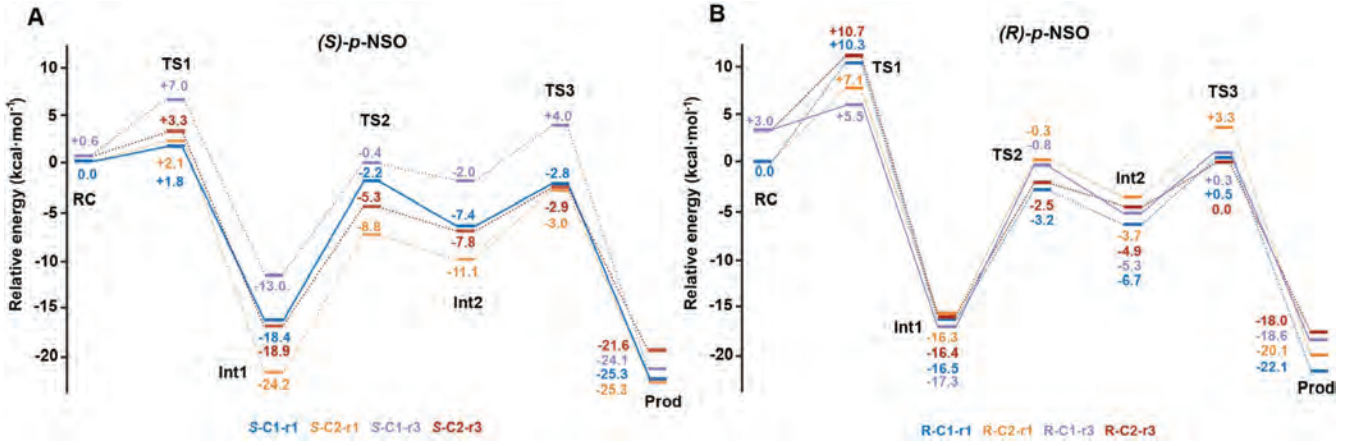


Fig. 3 Energy profiles at the B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory for all possible BmEH-mediated epoxide *rac*-*p*-NSO hydrolysis outcomes. All energies are in kcal mol^{-1} and referenced to the lowest energy **RC** for each enantiomer, **RC-S-r1** and **RC-R-r1**, respectively. The kinetically favoured attack for each enantiomer is shown as solid lines (**S-C1-r1** and **R-C1-r3**), the other possible pathways are shown as dashed lines.

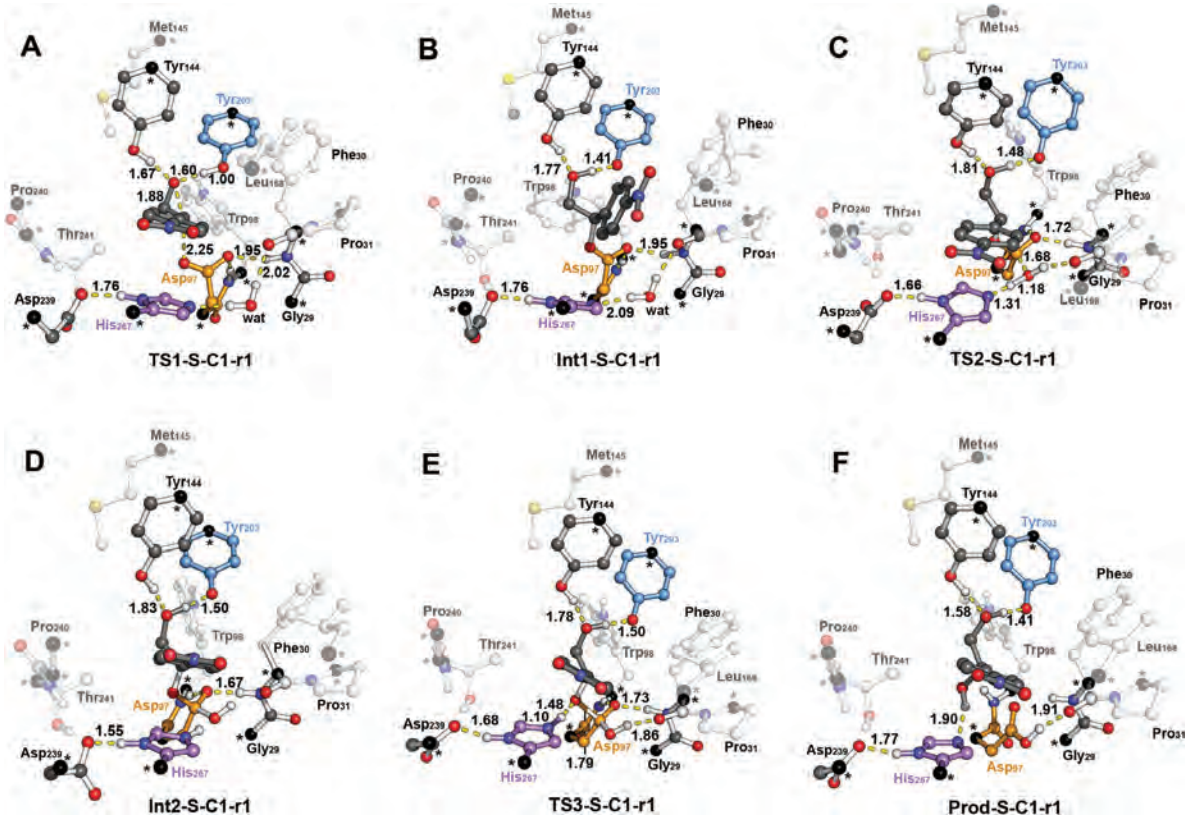


Fig. 4 B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) optimized geometries corresponding to the different steps of the BmEH mechanism of action for the (*S*)-*p*-NSO substrate oriented towards region 1 for the attack at the benzylic position (C1): (A) epoxide ring opening (**TS1-S-C1-r1**), (B) alkyl-enzyme intermediate (**Int1-S-C1-r1**), (C) hydrolysis of the alkyl-enzyme intermediate (**TS2-S-C1-r1**), (D) tetrahedral intermediate (**Int2-S-C1-r1**), (E) dissociation of tetrahedral (**TS3-S-C1-r1**) and (F) product complex (**Prod-S-C1-r1**). All distances are in Å. Active site pocket residues are shown as transparent sticks and spheres and non-polar hydrogen atoms are omitted for clarity. Atoms in black spheres and asterisks are kept fixed.

stage of the BmEH mechanism, our computed energy profiles indicate that the epoxide ring opening of (*S*)-*p*-NSO is more favoured than the (*R*)-enantiomer.

The next step in the reaction mechanism includes the nucleophilic attack by a water molecule (**wat**) to hydrolyse the previously formed ester bond, and the dissociation of the tetra-

hedral intermediate (**Int2**) generated at this half-hydrolytic reaction to yield the vicinal diol (**Prod**, see Scheme 2). In the DFT optimized **TS2-S-C1-r1**, the neutral His267 residue acts as a general base to activate the water molecule that attacks Asp97 generating the **Int2-S-C1-r1** intermediate, as shown in Fig. 4C. The activation of the water molecule is possible because at **TS2-S-C1-r1** the Asp239–His267 distance has been substantially shortened, allowing the stabilization of the protonated state of His267. Likewise, the observed distances in **Int2-S-C1-r1** between the carbonyl oxygen atom of Asp97 and the amide bonds of Asp97–Trp98 and Gly29–Phe30 in the oxyanion hole indicate that the negative charge that arises at this stage is well stabilized (shorter distances of *ca*. 1.67 Å, see Fig. 4D). All these observed networks of interactions allow **TS2** stabilization, thereby making it possible to overcome the computed **Int1** → **TS2** barriers of 16.2 kcal mol^{-1}. Similar energy barriers were found for the most favoured (*R*)-enantiomer attack (**TS2-R-C1-r3**, $\Delta E^{\ddagger}$ = 16.5 kcal mol^{-1}), in which the covalent enzyme intermediate **Int1-R-C1-r3** was found to be stabilized by 17.3 kcal mol^{-1}.

To generate the final diol product, a final step is needed in which the C–O bond between the epoxide substrate and Asp97 is cleaved, followed by subsequent protonation (see Scheme 2). The corresponding transition state **TS3-S-C1-r1** ($\Delta E^{\ddagger}$ = 4.6 kcal mol^{-1} with respect to **Int2**) presents elongated C–O distances (*ca.* 1.79 Å), and the protonated His267 interacts with the (*S*)-*p*-NSO substrate delivering a proton for 1,2-diol formation (*ca.* 1.48 Å) as shown in Fig. 4E. Therefore, the catalytic His267 found to be important in the previous step for catalysing the hydrolysis acting as a general base has also a critical role at this final step for 1,2-diol generation. Note that this C–O bond breaking generates a diol product in which one of the oxygen atoms comes from the nucleophilic Asp97 residue (see Scheme 2). The activation barrier for the dissociative transition state **TS3** for the (*R*)-enantiomer (**TS3-R-C1-r3**) is 5.6 kcal mol^{-1} with respect to **Int2**. As previously mentioned for the alkylation transition state (**TS1**), since for (*R*)-*p*-NSO the phenyl ring points towards region 3, this might lead to a less stabilized **TS3** due to the lack of π-stacking interaction with the catalytic His267 residue (see Fig. S3†).

As shown in Fig. 4F, the new hydroxyl group in the product **Prod-S-C1-r1** interacts with Tyr144-O–H (*ca.* 1.58 Å) and the deprotonated Tyr203-O^{-} (*ca.* 1.41 Å). **Prod-S-C1-r1** is −25.3 kcal mol^{-1} more stable than **RC-S-r1**, whereas for the (*R*)-enantiomer **Prod-R-C1-r3** is −18.6 kcal mol^{-1} lower in energy than **RC-R-r1**. In all optimized **Prod** structures, Asp97 is in its protonated state whereas Tyr203 remains deprotonated. Thus, an additional acid–base step is needed, and the binding of a new epoxide and water molecule is required to regenerate the catalytic cycle (see Scheme 2). These steps are difficult to be accurately described with the current methodology.[41] As proposed by Himo in previous studies,[41] the overall energetics of the process can be roughly approximated by computing the free energy of the (*S*)-epoxide + water → (*R*)-1,2-diol reaction, which is exergonic by 8 kcal mol^{-1}. This large energetic span between **Prod-S-C1-r1** and **RC** (−8 + 25.3 = 17.3 kcal mol^{-1}) should be added to the alkylation activation barrier of the next cycle, thus yielding an approximated barrier of *ca.* 19 kcal mol^{-1} for (*S*)-*p*-NSO. This estimation indicates that both alkylation and hydrolysis steps present similar barriers. Due to the large inaccuracies associated with the estimation of the **Prod** → **React** process, we focus our analysis on the computed reaction pathway displayed in Fig. 3, which shows that alkylation is selectivity-determining, whereas hydrolysis is rate-determining. Thus, to directly compare the reactivities of (*S*) and (*R*) substrates, we evaluate the activation barriers for the rate-determining hydrolysis step in the pathways that have the lowest alkylation barriers (**TS1**) for each enantiomer, *i.e.* **S-C1-r1** and **R-C1-r3**, respectively.

Overall our computed reaction pathways indicate that the lowest energy intermediate corresponds to **Int1**, and the rate-determining transition state is **TS2** for **S-C1-r1** and **TS3** for **R-C1-r3** both corresponding to the dissociation of the covalently bound enzyme intermediate (**Int1** in Scheme 2). The computed barriers are 16.2 kcal mol^{-1} for (*S*)-*p*-NSO (**S-C1-r1** in Fig. 3A), whereas 17.6 kcal mol^{-1} for (*R*)-*p*-NSO (**R-C1-r3** in Fig. 3B). Therefore, the hydrolysis of the (*S*)-enantiomer is kinetically favoured by *ca.* 1.4 kcal mol^{-1}, which is in agreement with experimental observations for BmEH.[12] It should also be noted that our computed activation barriers are in line with the experimental rate constants of 3–10 s^{-1} at 30 °C (*ca.* 16–17 kcal mol^{-1}) for the related StEH1 enzyme.[44,45]

Substituent effect on the BmEH selectivity

Using DFT-D3BJ calculations we also examined the hydrolysis of the racemic styrene oxide (*rac*-SO) substrate to further explore how the substrate structure and its substitution pattern affect BmEH selectivity. To this aim, we also considered all possible outcomes for *rac*-SO, *i.e.* the nucleophilic attack either at benzylic C1 or terminal carbon C2 for each enantiomer oriented towards region 1 or 3 as described above (see Fig. 1 and Fig. S4†).

The computed energy profiles for the *rac*-SO substrate follow a similar trend to those discussed in the previous section for its *para*-nitro derivative substrate *rac-p*-NSO (see Fig. 3 and Fig. S4†). As observed for *rac-p*-NSO energy profiles, the most favoured outcome for (*S*)-SO epoxide comes from the nucleophilic attack at the benzylic position (C1) when the phenyl moiety of the substrate is oriented towards His267 ($\Delta E^{\ddagger}$ = 4.0 kcal mol^{-1} with respect to **RC-S-SO-r1**, and $\Delta E^{\ddagger}$ = 15.3 kcal mol^{-1} computed from the lowest energy **Int1**, for **TS3-S-SO-C1-r1**).

For the (*R*)-enantiomer, the kinetically favored attack is found when the phenyl ring is oriented towards region 3. The alkylation barrier of **TS1-R-SO-C1-r3** is 6.6 kcal mol^{-1} regarding the lowest energy **RC** for (*R*)-SO (**RC-R-SO-r1**, see Fig. S4†). The alkylation at position C2 is *ca.* 0.5 kcal mol^{-1} less favorable than the attack at C1 (**TS1-R-SO-C2-r3**). The hydrolysis step of **Int1** is kinetically more favoured at the terminal carbon C2 as the computed activation barriers are *ca.* 19.8 and 17.4 kcal mol^{-1} for **TS3-R-SO-C1-r3** and **TS3-R-SO-C2-r3**,

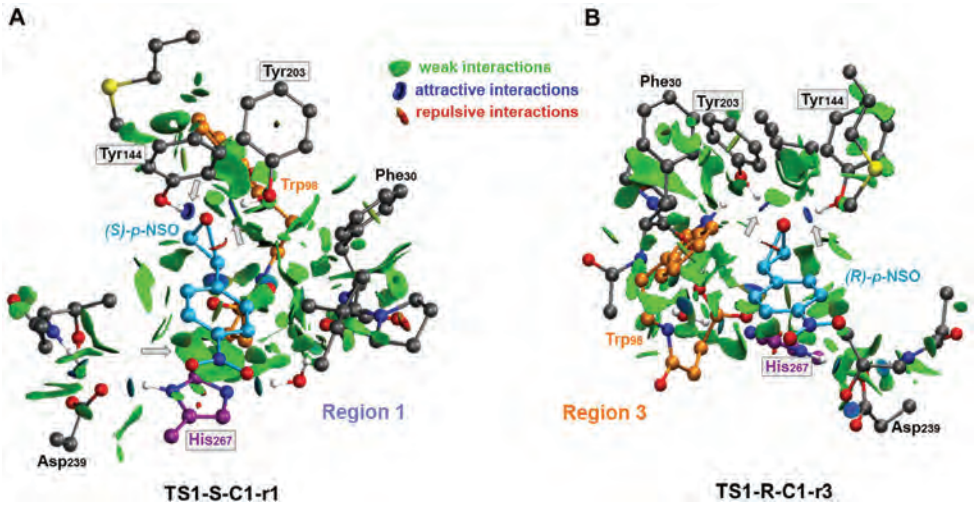


Fig. 5 Non-covalent interaction representation of the lowest energy alkylation transition states (A) **TS1-S-C1-r1** and (B) **TS1-R-C1-r3** for the *rac*-*p*-NSO epoxide ring opening reaction using the NCIplot computational tool.[44,45] NCI surfaces show intermolecular interactions between the epoxide substrate and the active site residues included in the CM model. All residues are represented as balls and sticks and the two possible orientations for each substrate are indicated for each case (region 1, His267 and region 3, Trp98). Only the most important catalytic residues are labelled regarding BmEH (PDB: 4NZZ numbering) and non-polar hydrogen are not shown for clarity. Grey arrows are used to highlight the most important noncovalent interactions. The most important residues are framed.

respectively (computed from the lowest energy intermediate **Int1**).

As observed for *p*-NSO, the activation barriers obtained (shown in Fig. S4†) for (*S*)-SO at C1 are *ca.* 2.6 kcal mol^{-1} lower than those obtained for (*R*)-SO, thus suggesting that the nitro group has a very minor effect on the enantio- and regioselectivity of the process.

Overall our mechanistic studies also indicate that both *rac*-*p*-NSO and *rac*-SO switch the inherent (*R*)-BmEH selectivity observed towards PGE substrates, as described in previous experimental studies. A higher enantioselectivity of BmEH towards *rac*-*p*-NSO ($ee_s \geq 99\%$) than for *rac*-SO ($ee_s = 53\%$) was observed experimentally.[12] The energy differences between our computed activation barriers overestimate the ee for *rac*-SO. However, the relative stabilities of the **RCs** indicate that the binding of (*R*)-epoxides for both *p*-NSO and SO is more favourable than for (*S*), especially in region 1 due to the higher stabilization by non-covalent interactions with the active site residues, including π···π with His267 (see Fig. 5). Our computed energy differences between substrate bound **RCs** ((*R*) *vs.* (*S*)) for *p*-NSO are *ca.* 3.0 kcal mol^{-1}, while it is only 1.4 kcal mol^{-1} for *rac*-SO. Thus, the higher enantioselectivity observed experimentally for *rac*-*p*-NSO can be attributed to the more favourable unproductive binding of (*R*)-*p*-NSO in region 1, as in this region the activation barriers for the alkylation and hydrolysis step are *ca.* 2 kcal mol^{-1} higher than in region 3.

Our calculations show that in the BmEH catalysed hydration of *rac*-*p*-NSO and *rac*-SO, **TS3** energies are in general lower than **TS1**. This is in contrast to *rac*-SO hydrolysis by StEH1,[41] where Himo and coworkers found that the hydrolysis step (**TS3**) is generally higher in energy than **TS1**. Although the energy differences are rather small (*ca.* ±3 kcal mol^{-1}), our results show that the shape and conformation of the active site pocket, as well as the substrate orientation will modulate the selectivity-determining step, *i.e.* the alkylation (**TS1**) or the hydrolysis (**TS3**).

3. Conclusions

Our DFT-D3BJ calculations within the CM framework indicate that the substrate structure switches the usual BmEH (*R*)-selectivity towards (*S*) when racemic styrene oxide (*rac*-SO) and racemic *para*-nitro styrene oxide (*rac*-*p*-NSO) are considered. The computed energy profiles for the different substrates and enantiomers suggest that the first alkylation step (**TS1**) is responsible for the regioselectivity of the process, while the hydrolysis half reaction is the rate-limiting step. In both *rac*-SO and *rac*-*p*-NSO the (*S*)-enantiomer is preferentially attacked at the benzylic position (C1) when the phenyl ring moiety points towards region 1 (*i.e.* the substrate entrance). This orientation maximizes the π···π interactions between the substrate (both SO, and *p*-NSO) and the active site residues, in particular His267. Therefore, the catalytically competent poses of the (*S*)-enantiomer of the aromatic SO and *p*-NSO substrates are found in region 1. The least reactive (*R*)-epoxide has a preferential binding for region 1, however in this region the activation barriers are substantially higher as compared to those in region 3. This work demonstrates that the enantiopreference of BmEH towards aromatic epoxides can be directly assessed by the combination of DFT-D3BJ and the analysis of non-covalent interactions, and provides an alternative strategy for the engineering of EHs towards the resolution of synthetically useful epoxides.

Computational methods

Full geometry optimizations were performed with the hybrid DFT B3LYP functional[46,47] and the 6-31G(d) basis set[48,49] using Gaussian 09,[50] a widely accepted level of theory for studying mechanisms of organic reactions.[51,52] The effects of the active site environment of the protein were implicitly included in all calculations using the Conductor-like Polarizable Continuum Model (CPCM)[53,54] with diethyl ether as the solvent ($\varepsilon = 4$). The transition states (TS) were located by using a Bofill update TS search on a reduced potential energy surface.[55] Analytical frequency calculations were performed at the same level of theory as the geometry optimizations to obtain the zero point energies (ZPE). The nature of the stationary points was determined in each case according to the Hessian matrix eigenvalues. We have also checked that imaginary frequencies exhibit the expected motion, while transition states have also been verified by intrinsic reaction coordinate (IRC) calculations. More accurate energies were obtained by single-point calculations including the DFT hybrid B3LYP-D3BJ dispersion correction with the larger 6-311+G(2d,2p) basis set. All energies are ZPE-corrected.

Active site model. The quantum chemical model used has been devised based on the most populated cluster from a 1000 ns MD simulation on the alkyl-enzyme intermediate (**Int1**, see full details in the ESI†). The starting structure was obtained from the X-ray crystal structure of wild-type BmEH from the PDB structure: 4NZZ. The cluster model consists of: an Asp239–His267–Asp97 catalytic triad, two active Tyr144 and Tyr203 residues, a nucleophilic water molecule and two hydrogen bonded amide bonds (Gly29–Phe30 and Phe30–Pro31 from the oxyanion motif HGFP). In addition, Trp98, Leu168, Met145, Thr241, and Pro240 residues were also included to appropriately describe the shape of the active site cavity of the $BmEH_{4NZZ}$ enzyme. All amino acids were truncated at the α- or β-carbon, except for the tyrosines, which are modelled as phenols and for Pro240, in which only the α- and its side-chain carbon attached are included to properly simulate the conformational rigidity of proline. Atoms in black (and highlighted with asterisks) shown in Fig. 2 are kept fixed to their initial Cartesian coordinates during the optimizations. Hydrogen atoms were included manually based on either the literature or the $BmEH_{4NZZ}$ active site rearrangement, resulting in two aspartate Asp97 and Asp239 active site residues modelled as in the negatively charged state, whereas His267 was modelled in the neutral form, obtaining a quantum model with an overall charge of −2 (see Fig. 2).

The epoxide substrates used to model the mechanisms were the racemic styrene oxide (*rac*-SO) and its racemic *para* nitro-substituted derivative (*rac-p*-NSO, cluster models of 195 and 197 atoms, respectively). To explore the origins of the enantio- and regioselectivity of BmEH and the effect of the presence of the NO_2 group, all plausible reaction mechanisms should be considered. Thus, a total of 16 reaction pathways need to be computed to take into account that the EH hydrolysis of *rac*-epoxides can take place *via* an attack on either carbon atom of each enantiomer, and also two possible binding orientations of the substrate on the basis of the proposed BmEH active site tunnel.[6] The two possible binding poses arise from the orientation of the phenyl moiety of the substrate towards either region 1 (**r1**) placing the phenyl moiety close to His267, or region 3 (**r3**) when it points towards Trp98 (see Fig. 1 and 2).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

E. S.-H. thanks the Generalitat de Catalunya for a PhD fellowship (2017-FI-B-00118), M. G.-B. is grateful to the Ramón Areces Foundation for a Postdoctoral Fellowship. S. O. thanks the Spanish MINECO CTQ2014-59212-P, Ramón y Cajal contract (RYC-2014-16846), the European Community for CIG project (PCIG14-GA-2013-630978), and the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2015-StG-679001). We are grateful for the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center – Centro Nacional de Supercomputacion.

Notes and references

1 A. Archelas and R. Furstoss, *Curr. Opin. Chem. Biol.*, 2001, **5**, 112–119.
2 E. J. de Vries and D. B. Janssen, *Curr. Opin. Biotechnol.*, 2003, **14**, 414–420.
3 A. Archelas and R. Furstoss, *Trends Biotechnol.*, 1998, **16**, 108–116.
4 P. Besse and H. Veschambre, *Tetrahedron*, 1994, **50**, 8885–8927.
5 K. Michael, A. Alain and W. Roland, *Curr. Org. Chem.*, 2012, **16**, 451–482.
6 X.-D. Kong, S. Yuan, L. Li, S. Chen, J.-H. Xu and J. Zhou, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 15717–15722.
7 F. E. Held, S. Wei, K. Eder and S. B. Tsogoeva, *RSC Adv.*, 2014, **4**, 32796–32801.
8 A. Schmid, K. Hofstetter, H. J. Feiten, F. Hollmann and B. Witholt, *Adv. Synth. Catal.*, 2001, **343**, 732–737.
9 P. Berglund, *ChemBioChem*, 2006, **7**, 1280–1280.
10 M. Smit and M. Labuschagné, *Curr. Org. Chem.*, 2006, **10**, 1145–1161.
11 B. Van Loo, J. Kingma, M. Arand, M. G. Wubbolts and D. B. Janssen, *Appl. Environ. Microbiol.*, 2006, **72**, 2905–2917.
12 J. Zhao, Y.-Y. Chu, A.-T. Li, X. Ju, X.-D. Kong, J. Pan, Y. Tang and J.-H. Xu, *Adv. Synth. Catal.*, 2011, **353**, 1510–1518.

13 Y.-F. Tang, J.-H. Xu, Q. Ye and B. Schulze, *J. Mol. Catal. B: Enzym.*, 2001, **13**, 61–68.
14 P.-F. Gong and J.-H. Xu, *Enzyme Microb. Technol.*, 2005, **36**, 252–257.
15 Y. Xu, J.-H. Xu, J. Pan and Y.-F. Tang, *Biotechnol. Lett.*, 2004, **26**, 1217–1221.
16 M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand and A. Cronin, *Org. Lett.*, 2004, **6**, 177–180.
17 M. T. Reetz, L. W. Wang and M. Bocola, *Angew. Chem., Int. Ed.*, 2006, **118**, 1258–1263.
18 P. Moussou, A. Archelas, J. Baratti and R. Furstoss, *J. Org. Chem.*, 1998, **63**, 3532–3537.
19 S. Barth, M. Fischer, R. D. Schmid and J. Pleiss, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 846–855.
20 M. Nardini and B. W. Dijkstra, *Curr. Opin. Struct. Biol.*, 1999, **9**, 732–737.
21 D. Lindberg, S. Ahmad and M. Widersten, *Arch. Biochem. Biophys.*, 2010, **495**, 165–173.
22 R. N. Armstrong and C. S. Cassidy, *Drug Metab. Rev.*, 2000, **32**, 327–338.
23 K. H. Hopmann and F. Himo, *J. Phys. Chem. B*, 2006, **110**, 21299–21310.
24 R. N. Patel, *Green Biocatalysis*, John Wiley & Sons, 2016.
25 W. J. Choi, *Appl. Microbiol. Biotechnol.*, 2009, **84**, 239–247.
26 P. Bauer, A. J. Carlsson, B. A. Amrein, D. Dobritzsch, M. Widersten and S. C. L. Kamerlin, *Org. Biomol. Chem.*, 2016, **14**, 5639–5651.
27 D. Lindberg, A. Gogoll and M. Widersten, *FEBS J.*, 2008, **275**, 6309–6320.
28 E. Y. Lee and M. L. Shuler, *Biotechnol. Bioeng.*, 2007, **98**, 318–327.
29 K. H. Hopmann and F. Himo, *Chem. – Eur. J.*, 2006, **12**, 6898–6909.
30 B. A. Amrein, P. Bauer, F. Duarte, A. J. Carlsson, A. Naworyta, S. L. Mowbray, M. Widersten and S. C. L. Kamerlin, *ACS Catal.*, 2015, **5**, 5702–5713.
31 K. H. Hopmann and F. Himo, *J. Phys. Chem. B*, 2006, **110**, 21299–21310.
32 B. Schiott and T. C. Bruice, *J. Am. Chem. Soc.*, 2002, **124**, 14558–14570.
33 M. T. Reetz, M. Bocola, L.-W. Wang, J. Sanchis, A. Cronin, M. Arand, J. Zou, A. Archelas, A.-L. Bottalla, A. Naworyta and S. L. Mowbray, *J. Am. Chem. Soc.*, 2009, **131**, 7334–7343.
34 R. N. Armstrong, *Drug Metab. Rev.*, 1999, **31**, 71–86.
35 G. M. Lacourciere and R. N. Armstrong, *J. Am. Chem. Soc.*, 1993, **115**, 10466–10467.
36 C. Morisseau and B. D. Hammock, *Annu. Rev. Pharmacol. Toxicol.*, 2005, **45**, 311–333.
37 A. Romero-Rivera, M. Garcia-Borràs and S. Osuna, *Chem. Commun.*, 2017, **53**, 284–297.
38 K. Hotta, X. Chen, R. S. Paton, A. Minami, H. Li, K. Swaminathan, I. I. Mathews, K. Watanabe, H. Oikawa, K. N. Houk and C.-Y. Kim, *Nature*, 2012, **483**, 355–358.
39 Y. Zou, M. Garcia-Borràs, M. C. Tang, Y. Hirayama, D. H. Li, L. Li, K. Watanabe, K. N. Houk and Y. Tang, *Nat. Chem. Biol.*, 2017, **13**, 325–332.
40 L. T. Elfström and M. Widersten, *Biochem. J.*, 2005, **390**, 633–640.
41 M. E. S. Lind and F. Himo, *ACS Catal.*, 2016, 8145–8155.
42 P. Moussou, A. Archelas, J. Baratti and R. Furstoss, *Tetrahedron: Asymmetry*, 1998, **9**, 1539–1547.
43 K. H. Hopmann, B. M. Hallberg and F. Himo, *J. Am. Chem. Soc.*, 2005, **127**, 14339–14347.
44 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.
45 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.
46 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter*, 1988, **37**, 785–789.
47 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
48 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.
49 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
50 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, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT, 2013.
51 G. Kiss, N. Celebi-Oelcuem, R. Moretti, D. Baker and K. N. Houk, *Angew. Chem., Int. Ed.*, 2013, **52**, 5700–5725.
52 F. Himo, *Theor. Chem. Acc.*, 2006, **116**, 232–240.
53 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
54 V. Barone, M. Cossi and J. Tomasi, *J. Comput. Chem.*, 1998, **19**, 404–417.
55 S. K. Burger and P. W. Ayers, *J. Chem. Phys.*, 2010, **132**, 234110.

Organic & Biomolecular Chemistry

SUPPORTING INFORMATION

Exploring the origins of selectivity in soluble Epoxide Hydrolase from *Bacillus megaterium*

Eila Serrano-Hervás,[a] Marc Garcia-Borràs*[b] and Sílvia Osuna*[a]

SUPPORTING INFORMATION

Computational methods

Full geometry optimizations were performed with the hybrid DFT B3LYP functional[1, 2] and the 6-31G(d) basis set[3, 4] basis set using Gaussian 09,[5] a widely accepted level of theory for studying mechanisms of organic reactions.[6, 7] The effects of the active site environment of the protein were implicitly included in all calculations using the Conductor-like Polarizable Continuum Model (CPCM)[8, 9] with diethyl ether as the solvent (ε=4). The transition states (TS) were located by using a Bofill update TS search on a reduced potential energy surface.[10] Analytical frequency calculations were performed at the same level of theory as the geometry optimizations to obtain the zero point energies (ZPE). The nature of the stationary points was determined in each case according to the Hessian matrix eigenvalues. We have also checked that imaginary frequencies exhibit the expected motion, while transition states have also been verified by intrinsic reaction coordinate (IRC) calculations. More accurate energies were obtained by single-point calculations including the DFT hybrid B3LYP-D3BJ dispersion correction with the larger 6-311+G(2d,2p) basis set. All energies are ZPE-corrected.

Active site model. The quantum chemical model used has been devised based on the most populated cluster from a 1000-ns MD simulation on the alkyl-enzyme intermediate (**Int1**, see full details of MD simulation below). The starting structure was obtained from the X-ray crystal structure of wild-type BmEH from the PDB structure: 4NZZ. The cluster model consists of: Asp239-His267-Asp97 catalytic triad, the two active Tyr144 and Tyr203, the nucleophilic water molecule and two hydrogen bonded amide bonds (Gly29-Phe30 and Phe30-Pro31 from the oxyanion motif HGFP). In addition, Trp98, Leu168, Met145, Thr241, Pro240 residues were also included to appropriately describe the shape of the active site cavity of $BmEH_{4NZZ}$ enzyme. All amino acids were truncated at the α– or β–carbon, except for the tyrosines, which are modelled as phenols and for the Pro240, in which only the α– and its side-chain carbon attached are included to properly simulate the conformational rigidity of proline. Atoms in black in Figure 2 are those maintained fixed to their initial Cartesian coordinates during the optimizations. Hydrogen atoms were included manually based either on literature and $BmEH_{4NZZ}$ active site rearrangement, resulting the two aspartate Asp97 and Asp239 active site residues modelled as in the negatively charged state, whereas His267 was modelled in the neutral form, obtaining a quantum model with an overall charge of -2 (see Figure 2).

The epoxide substrates used to model the mechanisms were the racemic styrene oxide (*rac*-SO) and its racemic *para* nitro substituted derivative (*rac*-*p*-NSO, cluster models of 195 and 197 atoms, respectively). To explore the origins of the enantio- and regioselectivity of BmEH and the effect of the presence of the NO_2 group, all plausible reaction mechanism should be considered. Thus, a total of 16 reaction pathways need to be computed to take into account that EHs hydrolysis of *rac*-epoxides can take place via attack on either carbon atom of each enantiomer, but also two possible binding orientations of the substrate on the basis of the proposed BmEH active site tunnel.[11] The two possible binding poses arise from the orientation of the phenyl moiety of the substrate towards either region 1 (**r1**) placing the phenyl moiety close to His267, or region 3 (**r3**) when it is pointing towards Trp98 (see Figure 1 and 2).

Molecular Dynamics of the alkyl intermediate. 1000 ns MD simulation in explicit water was performed using AMBER 16 package[12] in our in-house GPU cluster *Galatea*.

Parameters for the non-standard alkyl-intermediate were generated within the *antechamber* module of AMBER 16 using the general AMBER force field (GAFF),[13] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the restrained electrostatic potential (RESP) model.[14] The charges were calculated according to the Merz-Singh-Kollman scheme[15, 16] using Gaussian 09.[5] Amino acid protonation states were predicted using the H++ server (http://biophysics.cs.vt.edu/H++).[17] Then, the covalently modified enzyme (PD: 4NZZ) was solvated in a pre-equilibrated truncated cuboid box with a 10-Å buffer of TIP3P[18] water molecules using the AMBER16 *leap* module, resulting in the addition of ~9,000 solvent molecules. The systems were neutralized by addition of explicit counterions (Na^+). All subsequent calculations were done using the widely tested Stony Brook modification of the Amber 99 force field (ff99SB).[19]

A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on solute by a harmonic potential with a force constant of 500 kcal·mol^{-1}·Å^{-2}, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems are gently heated using six 50-ps steps, incrementing the temperature 50 K each step (0–300 K) under constant volume and periodic boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.[20]

An 8-Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 10 kcal·mol^{-1} were applied to the solute, and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated without restrains for 2 ns with a 2-fs time-step at a constant pressure of 1 atm and temperature of 300 K. After the systems were equilibrated in the NPT ensemble, 1 µs MD simulation was performed under the NVT ensemble and periodic-boundary conditions.

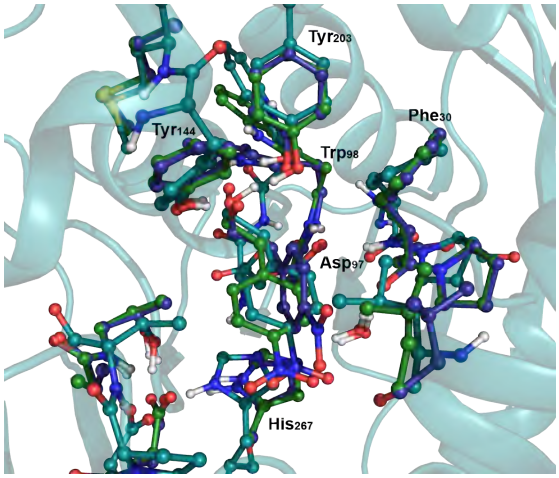


Figure S1. Representation of the overlay of the most populated 1 µs MD BmEH simulation (in teal), the cluster model structures **TS1-S-C1-r1** (in green) and **Int1-S-C1-r1** (in purple).

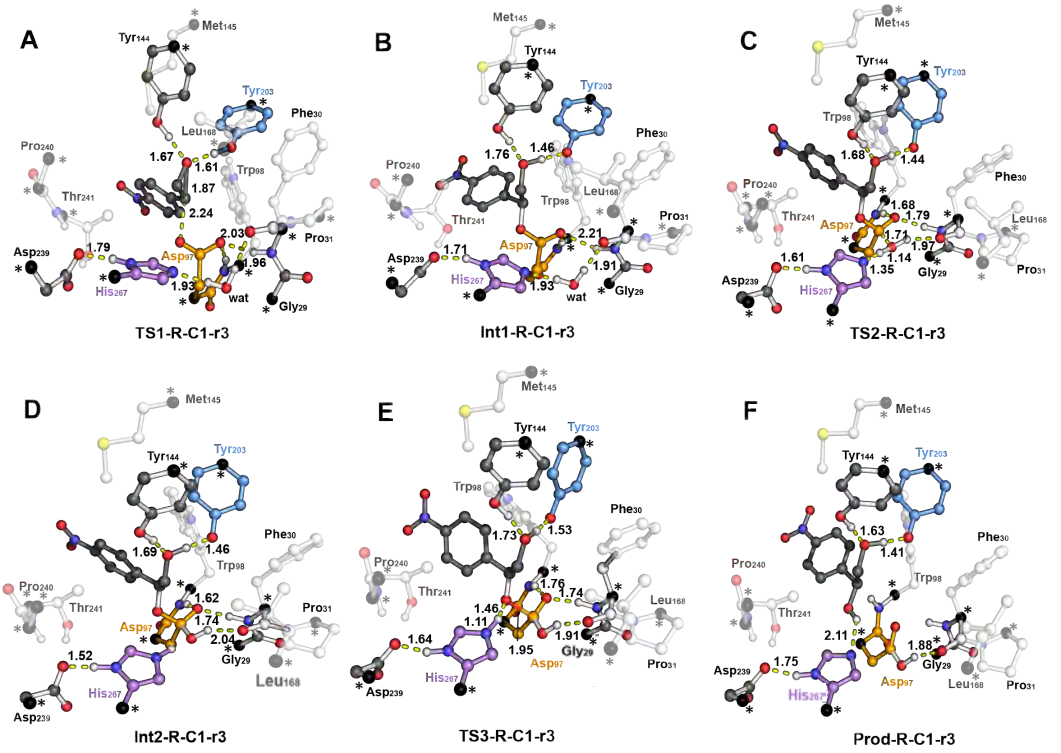


Figure S2. B3LYP-D3BJ/6-311+g(2d,2p)//6-31g(d) level corresponding to the different steps of the BmEH mechanism of action for *(R)-p*-NSO substrate oriented towards region 3 for the attack at the benzylic positon (C1): **A)** epoxide ring opening (**TS1-R-C1-r3**), **B)** alkyl-enzyme intermediate (**Int1-R-C1-r3**), **C)** hydrolysis of alkyl-enzyme intermediate (**TS2-R-C1-r3**), **D)** tetrahedral intermediate (**Int2-R-C1-r3**), **E)** dissociation of tetrahedral (**TS3-R-C1-r3**) and **F)** product complex (**Prod-R-C1-r3**). All distances are in Å. Active site pocket residues are shown as transparent sticks and spheres and non-polar hydrogen atoms are omitted for clarity. Atoms in black spheres and asterisks are those maintained fixed.

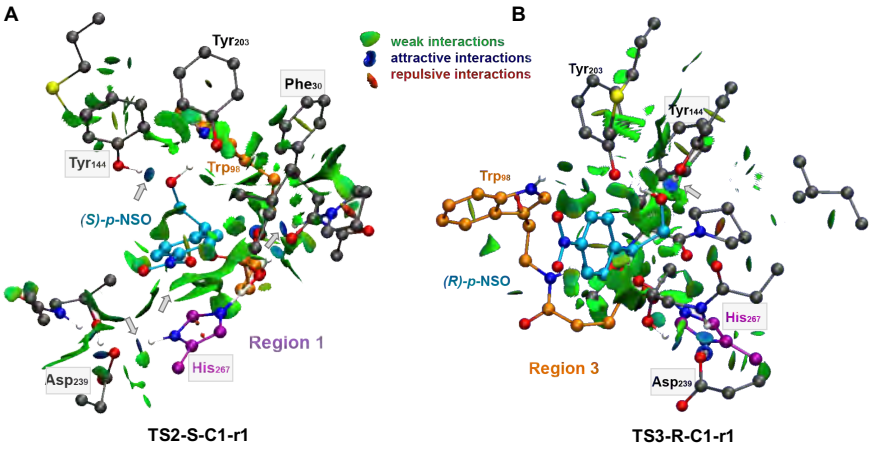


Figure S3. Non-covalent interactions representation of transition states **A) TS2-S-C1-r1** and **B) TS3-R-C1-r1** for the most favoured attack of *(S)-p*-NSO hydrolysis using the NCIplot computational tool.[21, 22] NCI surface show intermolecular interactions between the epoxide substrate and the active site residues included in the CM model. All residues are represented as balls and sticks and the two possible orientations of each substrate is indicated for each case (region 1, His267 and region 3, Trp98). Only the most important catalytically residues are labelled regarding BmEH (PDB: 4NZZ) numbering and non-polar hydrogen are not shown for clarity. Grey arrows are used to highlight the most important noncovalent interactions. The most important residues are framed.

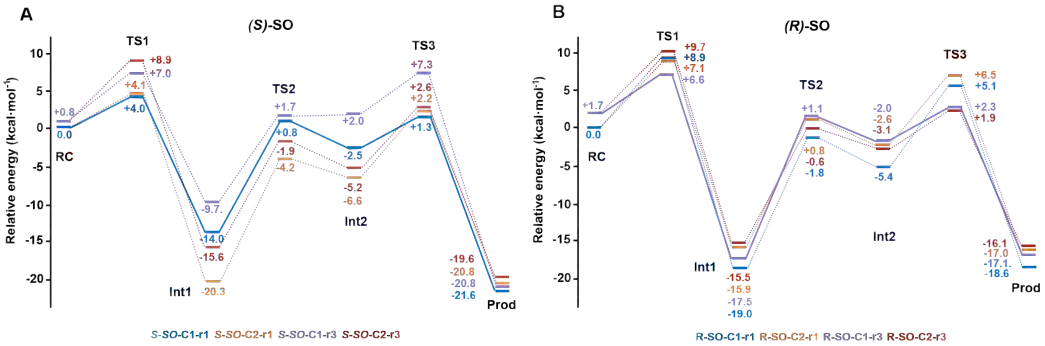


Figure S4. Energy profiles at B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory for all possible BmEH-mediated epoxide *rac*-SO hydrolysis outcomes. All energies are in kcal·mol^{-1} and referenced to the lowest energy **RC** for each enantiomer, **RC-S-SO-r1** and **RC-R-SO-r1**, respectively.

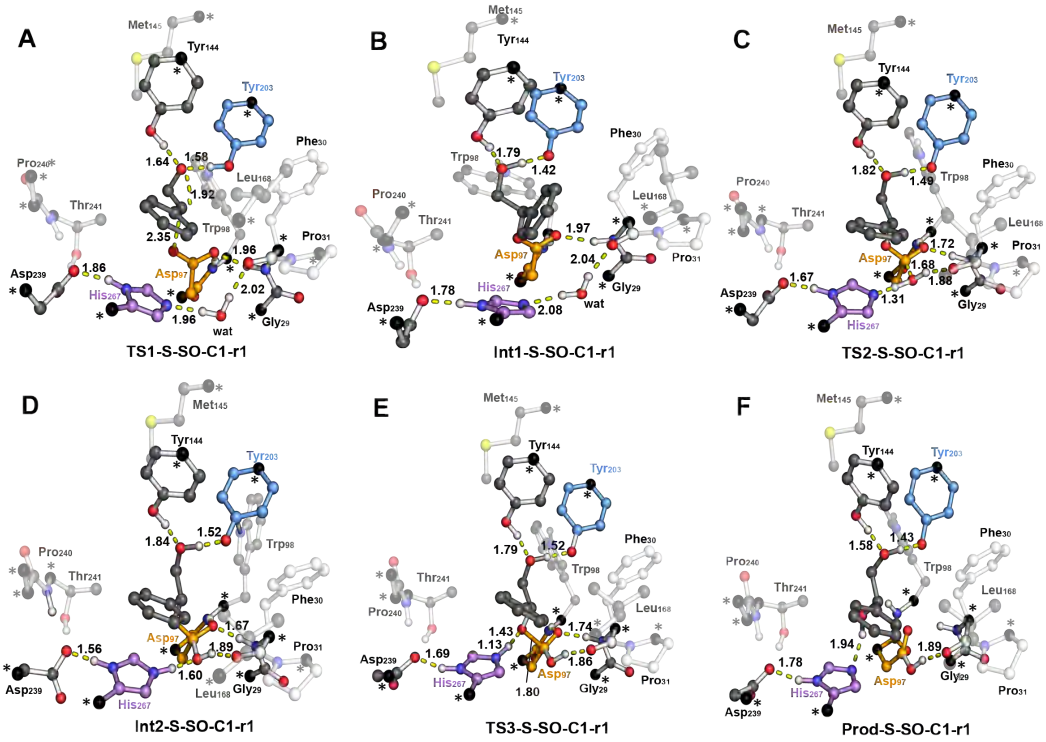


Figure S5. B3LYP-D3BJ/6-311+g(2d,2p)//6-31g(d) level corresponding to the different steps of the BmEH mechanism of action for *(S)*-SO substrate oriented towards region 1 for the attack at the benzylic positon (C1): **A)** epoxide ring opening (**TS1-S-SO-C1-r1**), **B)** alkyl-enzyme intermediate (**Int1-S-SO-C1-r1**), **C)** hydrolysis of alkyl-enzyme intermediate (**TS2-S-SO-C1-r1**), **D)** tetrahedral intermediate (**Int2-S-SO-C1-r1**), **E)** dissociation of tetrahedral (**TS3-S-SO-C1-r1**) and **F)** product complex (**Prod-S-SO-C1-r1**). All distances are in Å. Active site pocket residues are shown as transparent sticks and spheres and non-polar hydrogen atoms are omitted for clarity. Atoms in black spheres and asterisks are those maintained fixed.

References

1. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
2. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
3. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213-222.
4. W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257-2261.
5. G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria,, J. R. C. M. A. Robb, G. Scalmani, V. Barone, B. Mennucci,, H. N. G. A. Petersson, M. Caricato, X. Li, H. P. Hratchian,, J. B. A. F. Izmaylov, G. Zheng, J. L. Sonnenberg, M. Hada,, K. T. M. Ehara, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima,, O. K. Y. Honda, H. Nakai, T. Vreven, J. A. Montgomery, Jr.,, F. O. J. E. Peralta, M. Bearpark, J. J. Heyd, E. Brothers,, V. N. S. K. N. Kudin, T. Keith, R. Kobayashi, J. Normand,, A. R. K. Raghavachari, J. C. Burant, S. S. Iyengar, J. Tomasi,, N. R. M. Cossi, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross,, C. A. V. Bakken, J. Jaramillo, R. Gomperts, R. E. Stratmann,, A. J. A. O. Yazyev, R. Cammi, C. Pomelli, J. W. Ochterski,, K. M. R. L. Martin, V. G. Zakrzewski, G. A. Voth,, J. J. D. P. Salvador, S. Dapprich, A. D. Daniels,, J. B. F. O. Farkas, J. V. Ortiz, J. Cioslowski, and a. D. J. Fox, *Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2013.*
6. G. Kiss, N. Celebi-Oelcuem, R. Moretti, D. Baker and K. N. Houk, *Angew. Chem. Int. Ed.*, 2013, **52**, 5700-5725.
7. F. Himo, *Theor. Chem. Acc.*, 2006, **116**, 232-240.
8. V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995-2001.
9. V. Barone, M. Cossi and J. Tomasi, *J. Comp. Chem.*, 1998, **19**, 404-417.
10. S. K. Burger and P. W. Ayers, *J. Chem. Phys.*, 2010, **132**, 234110.
11. X.-D. Kong, S. Yuan, L. Li, S. Chen, J.-H. Xu and J. Zhou, *Proc. Nat. Acad. Sci. USA*, 2014, **111**, 15717-15722.
12. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J.

Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 16, University of California, San Francisco, 2016*.

13. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comp. Chem.*, 2004, **25**, 1157-1174.
14. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
15. B. H. Besler, K. M. Merz and P. A. Kollman, *J. Comp. Chem.*, 1990, **11**, 431-439.
16. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, 1984, **5**, 129-145.
17. R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537-W541.
18. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
19. J. Wang, P. Cieplak and P. A. Kollman, *J. Comp. Chem.*, 2000, **21**, 1049-1074.
20. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
21. J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625-632.
22. E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498-6506.

Chapter 5

Epoxide hydrolase conformational heterogeneity for the resolution of bulky pharmacologically-relevant epoxide substrates

This chapter corresponds to the following publication:
Serrano-Hervás, E.; Casadevall, G.; Garcia-Borràs, M.; Feixas, F.; Osuna, S. "Epoxide Hydrolase Conformational Heterogeneity for the Resolution of Bulky Pharmacologically Relevant Epoxide Substrates", *Chem. Eur. J.*, **2018**, 24, 12254.

Epoxide hydrolase conformational heterogeneity for the resolution of bulky pharmacologically-relevant epoxide substrates

Eila Serrano-Hervás,[†[a]] Guillem Casadevall,[†[a]] Marc Garcia-Borràs,[[c]] Ferran Feixas,[*[a]] and Sílvia Osuna[*[a,b]]

Abstract: The conformational landscape of *Bacillus megaterium* epoxide hydrolase (*Bm*EH) and how it is altered by mutations that confer the enzyme the ability to accept bulky epoxide substrates has been investigated. Extensive molecular dynamics (MD) simulations coupled to active site volume calculations have unveiled relevant features of the enzyme conformational dynamics and function. Our long-timescale MD simulations identify a key conformational state not previously observed by means of X-ray crystallography and short MD simulations that presents the loop containing one of the catalytic residues Asp239 in a wide-open conformation, which is likely involved in the binding of the epoxide substrate. Introduction of mutations M145S and F128A dramatically alters the conformational landscape of the enzyme. These singly mutated variants are able to accept bulkier epoxide substrates due to the disorder in the α-helix containing the catalytic Tyr144 and some parts of the lid domain induced by mutation. These changes impact the enzyme active site, which is substantially wider and more complementary to the bulky pharmacologically-relevant epoxide substrates.

Enzymes are inherently dynamic and sample alternative conformations to the native state.[1] Many studies have shown that conformational heterogeneity is a key feature of enzymes, which is important for their function and for their evolution towards novel activities.[2] The enzyme free energy landscape originates from the different conformations that exist in dynamic equilibrium, which if separated by small energy barriers can be significantly populated in solution. Mutations introduced in the enzyme, but also the binding of substrate, inhibitor or product can induce a redistribution of the populations of the conformational states that pre-exist in solution.[3] The ability of some enzymes to accelerate additional promiscuous reactions has indeed been connected to this conformational plasticity and flexibility.[4] Experimental evidences have been provided supporting the importance of enzyme conformational dynamics that include B-factor analysis, careful inspection of crystallographic structures along evolutionary pathways, changes in NMR spectra, among others.[2a, 4a, 5] Computational methods have been shown to be particularly useful in this regard.[5c, 6] We have recently found that molecular dynamics (MD) simulations coupled to correlation-based analysis can provide important insights on the alternative conformations that the enzyme can adopt in solution, but most importantly can predict which amino acid positions are key for favouring a desired population shift to enhance a promiscuous activity.[2c]

Most of the epoxide hydrolases (EH) structurally characterized up to date possess an α,β-sheet core and a flexible lid domain that covers the enzyme active site (see **Figure 1A**).[7] In the lid, two catalytically relevant tyrosine residues are contained (Tyr144 and Tyr203), which are responsible for the recognition and activation of the epoxide substrate to yield the vicinal diol in high enantio- and regioselectivity (see Figure 1B).[7c, 8] In the core domain, EHs present a conserved catalytic triad (Asp97-His267-Asp239), as well as the HGXP oxyanion motif responsible for stabilizing the negative charge developed after the epoxide ring opening by the nucleophilic Asp 97.

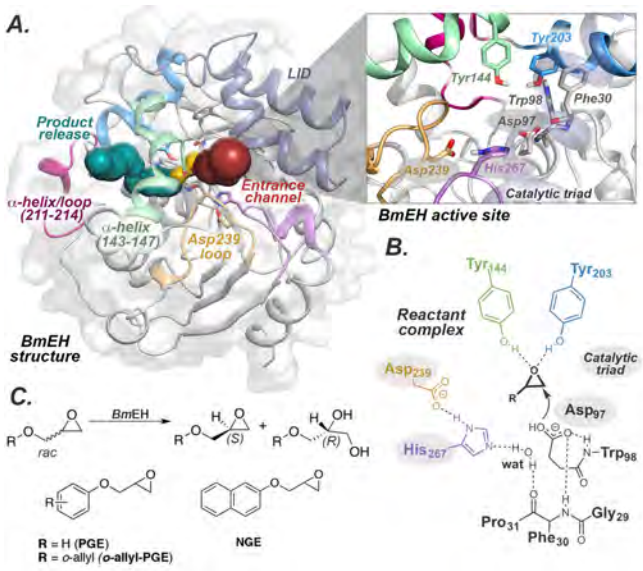


Figure 1. A) Representation of *Bm*EH enzyme wild-type structure (PDB: 4NZZ), and active site access tunnels: substrate entrance (zone 1, in red), active site (zone 2, yellow) and product release (zone 3, teal). Relevant regions for the enzyme catalytic activity are highlighted: the loops containing the catalytic triad residues Asp239 (in gold), His267 (violet), and the nucleophilic Asp97 (grey); the α-helices of Tyr144 and Tyr203 in green and blue, respectively. **B**) Schematic representation of *Bm*EH active site with an epoxide substrate bound (reactant complex). **C**) General reaction scheme of epoxide hydrolysis to yield the corresponding vicinal diol, together with the representation of bulky pharmacologically-relevant epoxide substrates phenyl glycidyl ether (**PGE**) and naphthyl glycidyl ether (**NGE**) structures.

The mechanism followed by EHs for catalysing the hydration of racemic epoxides has been long debated.[9] Many computational studies have been published in the literature providing a detailed atomistic description of the accepted two-step mechanism, and how active site mutations affect the activity and selectivity of the enzyme.[9] In contrast to other reported EHs,[9d, 10] *Bacillus megaterium* EH (*Bm*EH) has an unusual *(R)*-selectivity, being also active towards *ortho*-substituted phenyl glycidyl ethers (**PGE,** see **Figure 1C**) providing excellent enantiomeric ratios.[11] This makes *Bm*EH a potential industrial biocatalyst for the production of β-blocker drugs such as *(S)*-alprenolol and *(S)*-propranolol.[11b] In a recent study by our group, the origin of selectivity of *Bm*EH towards racemic styrene oxide and its *para*-nitro derivative was computationally explored.[12] Interestingly, substantially lower activation barriers were found for *(S)*-epoxides, indicating that the inherent *(R)*-selectivity of the enzyme can be modulated by the nature of the epoxide substrate as well.

*Bm*EH was crystallized in the presence of phenoxyacetamide, a substrate analogue (PDB: 4O08).[11b] Interestingly, three different potential binding sites were identified corresponding to either the active site region, named zone 2, that contains the abovementioned catalytic triad as well as the oxyanion motif, a deep cleft identified as zone 1 or entrance channel, and a bowl-like area corresponding to the product release or zone 3 (see active site tunnels highlighted in **Figure 1A**). Alanine scanning experiments at two positions located in the substrate entrance and product release zones, i.e. residues F128 and M145, were used to support the hypothesis of the different active site tunnels.[11b] Interestingly, mutation of both positions to smaller residues (serine and alanine) was found to be crucial for enhancing the catalytic activity of the enzyme towards the resolution of bulky epoxide substrates such as α-naphthyl glycidyl ether (**NGE**, the activity is from 25 to 434 times higher than for the wild type enzyme).[11a, b] This increase in activity was

attributed to the wider product release zone 3 of the variants that is supposed to facilitate the product dissociation step.[11b]

The hypothesis that different access tunnels for the substrate association and product dissociation exist in *Bm*EH, combined with its unique ability to accept bulkier substrates of industrial interest make the study of the conformational variability of *Bm*EH highly appealing. Of particular relevance is the evaluation of the role of positions F128 and M145. In this communication, we evaluate the conformational heterogeneity of BmEH by means of long-timescale molecular dynamics (MD) simulations, and explore how the included mutations affect the conformational landscape and the ability of the enzyme to accept bulkier pharmacologically-relevant epoxide substrates.

BmEH conformational heterogeneity:

The free energy landscape (FEL) constructed from an accumulated simulation time of 19 microseconds of MD simulations for the wild-type enzyme in the *apo* state is shown in **Figure 2A**. Dimensionality reduction of the highly complex MD dimensional data was achieved by applying the kinetically-relevant time-lagged independent components analysis (TICA) (see computational details and **Scheme S1** for free energy construction). Wild-type EH can adopt four major conformations, among which the most populated one corresponds to the X-ray-like conformation $\mathbf{1^{WT}}$ (see **Figure 2A**). In this conformational state, the catalytic machinery of the enzyme is well-positioned for catalysis: both Tyr144 and Tyr203 are properly positioned, and the Asp293-His267-Asp97 triad is in a catalytically competent conformation for performing the nucleophilic attack after substrate binding (see **Figure 1C**). However, this conformational state presents quite small volumes in both region 1 and 3 (an approximate total volume of 133 $Å^3$). The other three conformational states sampled are not as catalytically competent as $\mathbf{1^{WT}}$ as their catalytic triad arrangements present deviations from the ideal quantum mechanics geometry.[12] However, in the absence of ligand they are also substantially populated in solution.

Conformational state $\mathbf{2^{WT}}$ exhibits a different conformation of the sidechain of the oxyanion hole residue Phe30. This change of Phe30 affects the nearby residue Trp98, also involved in the stabilization of the developed negative charges during the course of the epoxide ring opening reaction. In $\mathbf{3^{WT}}$, the lid that covers the active site of the enzyme is slightly deviated from the X-ray position. This displacement affects the loop that contains His267, which is still situated at a short distance from the catalytic Asp97 (at *ca.* 3.6 Å). The loop containing Asp239 that completes the catalytic triad is also affected, and adopts a partially open conformation (see black arrows in **Figure 2A**, $\mathbf{3^{WT}}$). Conformations $\mathbf{2^{WT}}$ and $\mathbf{3^{WT}}$ present substantially larger volumes as compared to the X-ray conformation $\mathbf{1^{WT}}$ (*ca.* 170 and 230 $Å^3$ for $\mathbf{2^{WT}}$ and $\mathbf{3^{WT}}$, respectively, see **Table S1**). Both conformational states are relatively stable and separated from $\mathbf{1^{WT}}$ by small energy barriers in this reaction coordinate. These results are indicating that in the absence of epoxide substrate, these three states are thermally accessible and coexist in solution.

The observed displacement of the lid and Asp239-containing loop is even more pronounced in **4**^{WT} (see black arrows in **Figure 2A**, **4**^{WT}). Conformational state **4**^{WT} is quite stable and is separated from conformations **1**^{WT}-**3**^{WT} by an energy barrier of *ca*. 3 kcal/mol in this reaction coordinate. The lid adopts a closed conformation substantially shifting the α-helix that contains the catalytically relevant Tyr144, which is displaced from the active site. Tyr203 is also not properly positioned for catalysis as it locates its sidechain in region 3. The lid conformation has a dramatic impact into the loop containing Asp239 that adopts a catalytically unproductive open conformation. The calculation of the volume of zones 1 and 3 evidences that this conformational state **4**^{WT} has a wider access to the active site zone (a total volume of *ca.* 617 Å^3, see **Table S1**). These results are suggesting a key role of both the Asp239 loop and the coupled Tyr144 conformational change for binding the **PGE** epoxide substrate.

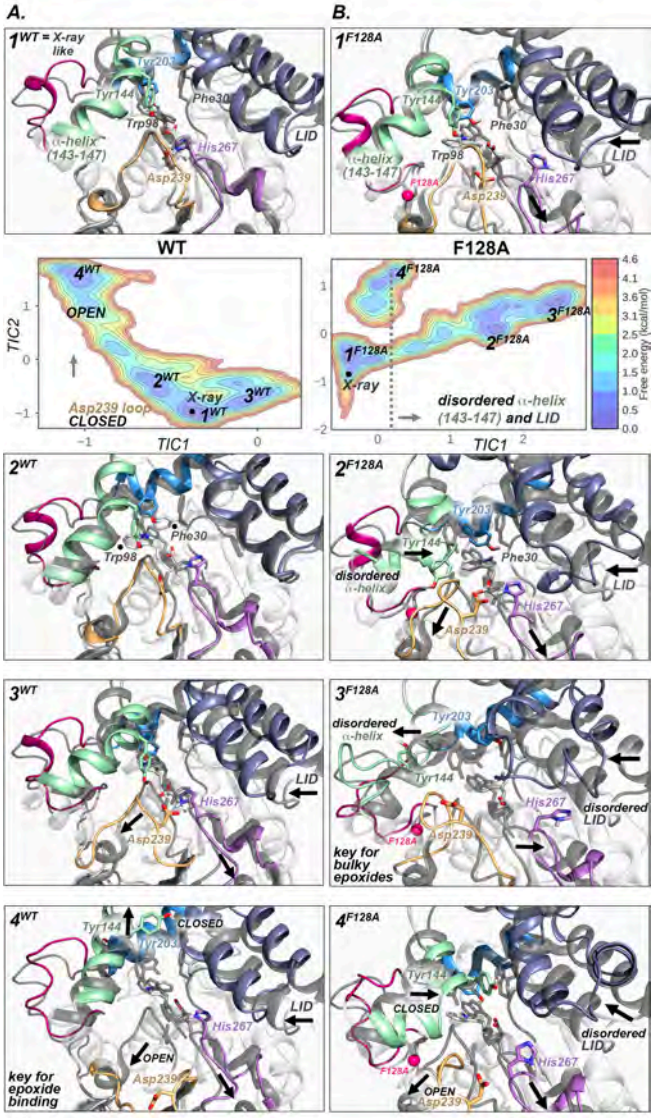


Figure 2. Representation of the free energy landscape obtained after applying the dimensionality reduction TICA to the accumulated MD data set for **A**) *Bm*EH wild-type, and **B**) F128A variant. TIC1 describes the degree of α-helix and lid disorder, while TIC2 differentiates open and closed conformational states of Asp239-containing loop. The four most stable conformational states for: **A**) *Bm*EH wild-type (i.e. **1**^{WT}- **4**^{WT}), and **B**) F128A variant (i.e. **1**^{F128A}-**4**^{F128A}) are overlaid to the crystallographic wild-type structure of BmEH (shown in gray). The most important changes occurring in each conformational state (as compared to the X-ray conformation) are highlighted with black arrows and bold text.

Partial unfolding of Tyr144 alpha-helix is responsible for BmEH tolerance towards bulky epoxide substrates:

Mutations F128A and M145S were found to enhance the enzyme ability for accepting the bulkier *o*-allyl phenyl glycidyl ether (*o*-allyl-**PGE**) and naphthyl glycidyl ether (**NGE**) substrates, respectively. To elucidate the effect of the mutations in altering the conformational landscape of the wild-type enzyme, we ran additional MD simulations of both singly-mutated variants (see computational details).

The crystal structure available for the F128A variant (PDB: 4IO0) presents the diol product of **NGE** bound to zone 3 of the enzyme. As shown in **Figure 2B**, the X-ray conformation that presents all catalytic residues in a catalytically competent pose is not the most stable conformation of the enzyme in the *apo* state (see X-ray minima in **Figure 2B**). Instead, conformational state $\mathbf{1}^{\mathbf{F128A}}$ is visited that presents the lid domain slightly shifted with respect to the crystal structure. This lid displacement affects the conformation of the loop containing His267 that positions its sidechain far away from the active site (the distance between Nε and Asp OD1 is *ca.* 7.7 Å). F128A mutation generates more space in zone 3, which is now occupied by the indole ring of Trp98 (see **Figure 2B**, $\mathbf{1}^{\mathbf{F128A}}$). Apart from these deviations, the rest of the catalytic and oxyanion hole residues are properly positioned. This conformation has a similar volume to $\mathbf{1}^{\mathbf{WT}}$.

Conformational state $\mathbf{2}^{\mathbf{F128A}}$ is equally populated as $\mathbf{1}^{\mathbf{F128A}}$, and is separated from $\mathbf{1}^{\mathbf{F128A}}$ by an energy barrier of *ca.* 2.5 kcal/mol through different minor conformational states that present the α-helix of Tyr144 slightly disordered (see black arrows in **Figure 2B,** $\mathbf{2}^{\mathbf{F128A}}$). Thus, converting $\mathbf{1}^{\mathbf{F128A}}$ into $\mathbf{2}^{\mathbf{F128A}}$ induces a partial disorder to the α-helix of the catalytic Tyr144, which is displaced from the active site and occupies the space left by F128A mutation. $\mathbf{2}^{\mathbf{F128A}}$ presents His267 closer to the nucleophilic Asp97 (at a distance between Nε and OD1 of *ca.* 5.8 Å), however as the loop that contains Asp239 is in a partially closed conformation the His267··Asp267 distance is elongated up to *ca.* 7 Å. This conformation is separated from $\mathbf{3}^{\mathbf{F128A}}$ by a small energy barrier of less than 1 kcal/mol.

$\mathbf{3}^{\mathbf{F128A}}$ presents the α-helices of the lid and the helix that contains residues 143-147 highly disordered, as well as the Asp239 loop in a partially open conformation. These changes increase the volume of both zone 1 and 3 quite substantially leading to a total volume of *ca.* 631 $Å^3$ (see **Table S1**). As compared to the wild-type enzyme $\mathbf{4}^{\mathbf{WT}}$, the wider volume found in region 1 suggest that conformational state $\mathbf{3}^{\mathbf{F128A}}$ is likely to play a key role for binding the bulky **NGE** epoxide substrate and facilitate the diol product release. Conformational state $\mathbf{4}^{\mathbf{F128A}}$ presents the lid in a closed conformation and the Asp239 loop open. Similarly to $\mathbf{4}^{\mathbf{WT}}$, Tyr144 is displaced from the active site (see **Figure 2**). However, the computed volume for $\mathbf{4}^{\mathbf{F128A}}$ is substantially smaller than in $\mathbf{4}^{\mathbf{WT}}$ (the total volumes are *ca.* 188 and 615 $Å^3$, respectively, see **Table S1**).

Our long-timescale MD simulations indicate that partial unfolding of the α-helix that contains the catalytically relevant Tyr144 is required for *Bm*EH to accept the bulkier **NGE** epoxide substrate. Of special relevance is conformational state $\mathbf{3}^{\mathbf{F128A}}$, not observed in the wild-type enzyme, that presents a wide active site pocket ready to bind **NGE** and facilitate the product release. The partial unfolding of Tyr144 α-helix observed in our MD simulations is in line with the experimental observation that doubly mutated

F128S/M145S variant presents severe protein-folding problems.[11a] To further evaluate our hypothesis, we decided to computationally evaluate the M145S variant.

The free energy landscape of variant M145S shares some common features of the wild-type enzyme and F128A variant (see **Figure S2**, SI). As observed in the wild-type enzyme, conformational state $\mathbf{1}^{\mathbf{M145S}}$ presents all catalytic residues in the proper conformation for catalysis, and $\mathbf{2}^{\mathbf{M145S}}$ has the oxyanion residues Phe30 and Trp98 with a different side-chain conformation. In this particular case, Trp98 can favorably position its indole ring in the space left by the M145S mutation. Conformational state $\mathbf{4}^{\mathbf{M145S}}$ is also similar to $\mathbf{4}^{\mathbf{WT}}$, as it presents the Asp239 loop in a wide-open conformation and Tyr144 displaced from the active site. This conformation presents a large volume of the active site (*ca.* 644 $Å^3$), which is bigger than for wild-type enzyme in line with the ability of M145S to accept *o*-allyl-**PGE**. An additional conformation is observed for M145S, i.e. $\mathbf{5}^{\mathbf{M145S}}$, that resembles $\mathbf{4}^{\mathbf{M145S}}$ although in this case a different conformation for the Trp98 sidechain is observed.

Interestingly M145S shares with F128A some conformational states that present the α-helix of Tyr144 and the lid domain highly disordered. Although with the simulation time accumulated for M145S, the stable conformational states $\mathbf{2}^{\mathbf{F128A}}$ and $\mathbf{3}^{\mathbf{F128A}}$ observed in F128A could not be visited. Still, our findings for M145S are again demonstrating that both mutations induce partial unfolding of α-helix 143-147, which generates a wider active site pocket and thus confers the enzyme the ability to accept bulky epoxide substrates.

Conclusions

Our extensive MD simulations combined with active site volume measurements reveal interesting features of *Bm*EH conformational dynamics. A new conformational state is characterized for the first time that presents the loop containing the catalytic Asp239 in a wide-open conformation. This conformational change positions Tyr144 out of the active site, and thus leads to a wide active site and a broad substrate access channel. Based on these observations, we propose that this conformation is key for binding the epoxide substrate. Introduction of mutations M145S and F128A alters the conformational landscape of the enzyme. Of special relevance is the appearance of some major conformational states that present the α-helix containing the catalytic Tyr144 highly disordered as well as some parts of the lid domain. These conformational states have a substantially wider active site pockets and access channels and thus correspond to the major reason for the ability of the M145S and F128A variants to accept bulkier pharmacologically-relevant epoxide substrates. Our study on the *Bm*EH conformational dynamics provides key relevant information for the pursue of new engineered variants with higher lid and α-helix 143-147 plasticity for the resolution of bulky industrially-relevant epoxide substrates.

Computational Section

Molecular Dynamics Simulations. Long-timescale MD simulations in explicit water were performed using AMBER 16 package[13] in our in-house GPU cluster Galatea. The structures used were taken from the

protein data bank (PDB); wild-type (WT, 4NZZ) and F128A variant complexed with the **NGE**-diol product (4IO0). M145S variant was generated by introducing the single mutation on different wild-type *apo* conformations explored through the long MD simulations using PyMOL (http://www.pymol.org). Each system was solvated in a pre-equilibrated truncated cuboid box with a 10-Å buffer of TIP3P water molecules and was neutralized by addition of explicit sodium counterions. All subsequent calculations were done using the widely tested Stony Brook modifications of the Amber 99 force field (ff99SB). The data used for the following analyses contained MD simulations of 2 trajectories of 5.2 μs and 3 of 0.2 μs for WT, 2 trajectories of 3.8 and 3.2 μs and 0.2 μs for F128A and 1 trajectory of 2 μs for M145S starting from the X-ray conformation. Additional MD simulations were performed in multiple rounds starting from the FEL less populated clusters generated with *k*-means algorithm of the previous long MD simulations. The total time accumulated is 19 μs for WT, 20 μs F128A, and 18 μs for M145S (see **Scheme S1** in SI).

FEL construction. The Free Energy Landscape (FEL) was constructed from the MD simulation data using *pyEMMA* [14] (http://pyemma.org). We applied the dimensionality reduction technique time-lagged independent component analysis (TICA),[15] by considering the minimum group distance between the Cα of every three residues as feature. TICA performs a dimensionality reduction by considering the time correlation of the data, thus minimizing the loss of kinetic information. The two slowest TICA components identified correspond to: TIC1 describing the disorder of α-helix and LID domain, TIC2 accounting for Asp239-loop movement. TIC1 and TIC2 were used to construct the free energy landscape of the enzyme. Uniform distance clustering (k-means algorithm as implemented in pyEMMA[14]) was applied to obtain 30 clusters for WT and M145S, and 50 for F128A that clearly distinguished the different protein conformations explored along the MD simulations. The obtained clusters were further clusterized according to the Cα coordinates using cpptraj.[13] Volume calculations were performed for the most populated and significant clusters with the computational tool POcket Volume MEasurer (POVME) 2.0.[16] Full description of computational methods, and additional figures and tables as described in the text are provided in the Supporting Information (SI).

Acknowledgements

E. S.-H. thanks the Generalitat de Catalunya for a PhD fellowship (2017-FI-B-00118), M. G.-B is grateful to the Ramón Areces Foundation for a Postdoctoral Fellowship. F.F. thanks the European Community for MSCA-IF-2014-EF-661160-MetAccembly grant. S. O. thanks the Spanish MINECO CTQ2014-59212-P, Ramón y Cajal contract (RYC-2014-16846), the European Community for CIG project (PCIG14-GA-2013-630978), and the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2015-StG-679001). We thank the Generalitat de Catalunya for grup emergent 2017 SGR-1707. We are grateful for the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center - Centro Nacional de Supercomputación.

Keywords: keyword 1 • keyword 2 • keyword 3 • keyword 4 • keyword 5

[1] N. Tokuriki and D. S. Tawfik, *Science* **2009**, *324*, 203-207.
[2] a) E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, N. Tokuriki and C. J. Jackson, *Nat. Chem. Biol.* **2016**, *12*, 944-950; b) B. Ma and R. Nussinov, *Nat. Chem. Biol.* **2016**, *12*, 890-891; c) A. Romero-Rivera, M. Garcia-Borràs and S. Osuna, *ACS Catal.* **2017**, *7*, 8524-8532.
[3] D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.* **2009**, *5*, 789-796.
[4] a) E. C. Campbell, G. J. Correy, P. D. Mabbitt, A. M. Buckle, N. Tokuriki and C. J. Jackson, *Curr. Opin. Struct. Biol.* **2018**, *50*, 49-57; b) G. G. Hammes, S. J. Benkovic and S. Hammes-Schiffer, *Biochem.* **2011**, *50*, 10422-10430; c) A. Pabis, V. A. Risso, J. M. Sanchez-Ruiz and S. C. L. Kamerlin, *Curr. Opin. Struct. Biol.* **2018**, *48*, 83-92; d) A. Ramanathan, A. Savol, V. Burger, C. S. Chennubhotla and P. K. Agarwal, *Acc. Chem. Res.* **2014**, *47*, 149-156.
[5] a) J. M. Axe, E. M. Yezdimer, K. F. O'Rourke, N. E. Kerstetter, W. You, C. A. Chang and D. D. Boehr, *J. Am. Chem. Soc.* **2014**, *136*, 6818-6821; b) A. Neu, U. Neu, A.-L. Fuchs, B. Schlager and R. Sprangers, *Nat. Chem. Biol.* **2015**, *11*, 697-704; c) M. Orozco, *Chem. Soc. Rev.* **2014**, *43*, 5051-5066.
[6] a) A. Romero-Rivera, M. Garcia-Borràs and S. Osuna, *Chem. Commun.* **2017**, *53*, 284-297; b) J.-P. Colletier, A. Aleksandrov, N. Coquelle, S. Mraihi, E. Mendoza-Barberá, M. Field and D. Madern, *Mol. Biol. Evol.* **2012**, *29*, 1683-1694; c) G. Jiménez-Osés, S. Osuna, X. Gao, M. R. Sawaya, L. Gilson, S. J. Collier, G. W. Huisman, T. O. Yeates, Y. Tang and K. N. Houk, *Nat. Chem. Biol.* **2014**, *10*, 431-436; d) S. Osuna, G. Jiménez-Osés, E. L. Noey and K. N. Houk, *Acc. Chem. Res.* **2015**, *48*, 1080-1089.
[7] a) S. Barth, M. Fischer, R. D. Schmid and J. Pleiss, *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 846-855; b) D. Lindberg, A. Gogoll and M. Widersten, *FEBS J.* **2008**, *275*, 6309-6320; c) M. Nardini and B. W. Dijkstra, *Curr. Opin. Struct. Biol.* **1999**, *9*, 732-737.
[8] a) R. N. Armstrong and C. S. Cassidy, *Drug Metab. Rev.* **2000**, *32*, 327-338; b) K. H. Hopmann and F. Himo, *J. Phys. Chem. B.* **2006**, *110*, 21299-21310.
[9] a) B. A. Amrein, P. Bauer, F. Duarte, A. J. Carlsson, A. Naworyta, S. L. Mowbray, M. Widersten and S. C. L. Kamerlin, *ACS Catal.* **2015**, *5*, 5702-5713; b) K. H. Hopmann, B. M. Hallberg and F. Himo, *J. Am. Chem. Soc.* **2005**, *127*, 14339-14347; c) K. H. Hopmann and F. Himo, *Chem. Eur. J.* **2006**, *12*, 6898-6909; d) M. T. Reetz, M. Bocola, L.-W. Wang, J. Sanchis, A. Cronin, M. Arand, J. Zou, A. Archelas, A.-L. Bottalla, A. Naworyta and S. L. Mowbray, *J. Am. Chem. Soc.* **2009**, *131*, 7334-7343; e) B. Schiott and T. C. Bruice, *J. Am. Chem. Soc.* **2002**, *124*, 14558-14570.
[10] a) Y. Xu, J.-H. Xu, J. Pan and Y.-F. Tang, *Biotechnol Lett.* **2004**, *26*, 1217-1221; b) M. T. Reetz, L. W. Wang and M. Bocola, *Angew. Chem.* **2006**, *118*, 1258-1263.
[11] a) X.-D. Kong, Q. Ma, J. Zhou, B.-B. Zeng and J.-H. Xu, *Angew. Chem. Int. Ed.* **2014**, *53*, 6641-6644; b) X.-D. Kong, S. Yuan, L. Li, S. Chen, J.-H. Xu and J. Zhou, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 15717-15722; c) J. Zhao, Y.-Y. Chu, A.-T. Li, X. Ju, X.-D. Kong, J. Pan, Y. Tang and J.-H. Xu, *Adv. Synth. Catal.* **2011**, *353*, 1510-1518.
[12] E. Serrano-Hervás, M. Garcia-Borràs and S. Osuna, *Org. Biomol. Chem.* **2017**, *15*, 8827-8835.
[13] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 16, University of California, San Francisco, 2016*.
[14] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.* **2015**, *11*, 5525-5542.
[15] a) G. Pérez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis and F. Noé, *J. Chem. Phys.* **2013**, *139*, 015102; b) C. R. Schwantes and V. S. Pande, *J. Chem. Theory Comput.* **2013**, *9*, 2000-2009.
[16] J. D. Durrant, L. Votapka, J. Sørensen and R. E. Amaro, *J. Chem. Theory Comput.* **2014**, *10*, 5047-5056.

CHEMISTRY
A European Journal

Supporting Information

Epoxide Hydrolase Conformational Heterogeneity for the Resolution of Bulky Pharmacologically Relevant Epoxide Substrates

Eila Serrano-Hervás[+],[a] Guillem Casadevall[+],[a] Marc Garcia-Borràs,[c] Ferran Feixas,*[a] and Sílvia Osuna*[a, b]

chem_201801068_sm_miscellaneous_information.pdf

Computational Section with full references

Molecular Dynamics Simulations. Long-timescale MD simulations in explicit water were performed using AMBER 16 package[1] in our in-house GPU cluster *Galatea*. Amino acid protonation states were predicted using the H++ serve (http://biophysics.cs.vt.edu/H++).[2] Then, the enzyme was solvated in a pre-equilibrated truncated cuboid box with a 10-Å buffer of TIP3P[3] water molecules. The systems were neutralized by addition of explicit counterions (Na+). All subsequent calculations were done using the widely tested Stony Brook modifications of the Amber 99 force field (ff99SB).[4] The structures used were taken from the protein data bank (PDB); wild-type (WT) (4NZZ), WT complexed with a substrate analogue (40O8) and F128A variant complexed with the NGE product (4IO0). M145S variant was generated by introducing the single mutation on different wild-type *apo* conformations explored along the MD simulations using the PyMOL (http://www.pymol.org).

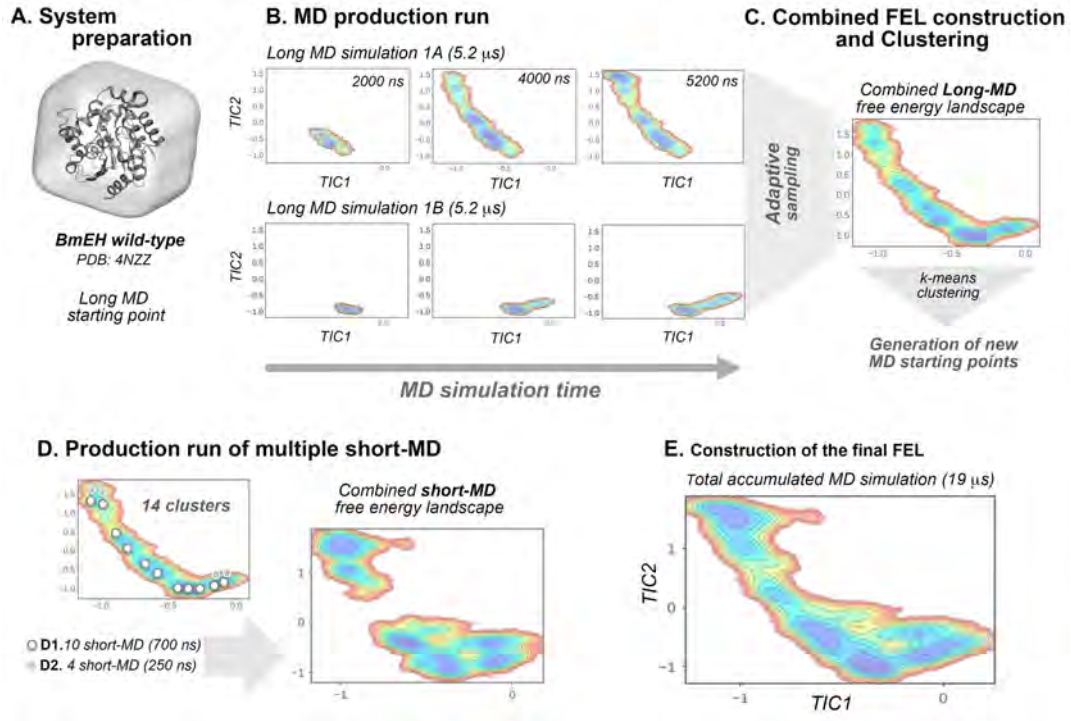
A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on solute by a harmonic potential with a force constant of 500 kcal·mol^{-1}· Å^{-2}, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems are gently heated using six 50-ps steps, incrementing the temperature 50 K each step (0-300 K) under constant volume and periodic boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[5]

An 8-Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 10 kcal·mol^{-1} were applied to the solute, and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated without restraints for 2 ns with a 2 fs time-step at a constant pressure of 1 atm and temperature of 300 K. After the systems were equilibrated in the NPT ensemble, 2 trajectories of 5.2 μs and 3 of 0.2 μs for WT, 2 trajectories of 3.8 and 3.2 μs 0.2 μs for F128A and 1 trajectory of 2 μs for M145S of MD simulations starting from the X-ray conformation were performed under the NVT ensemble and periodic-boundary conditions using our Galatea cluster (composed by 178 GTX1080 GPUs).

Additional MD simulations were performed in multiple rounds starting from the FES less populated clusters generated with *k*-means algorithm of the previous long MD runs: 10 trajectories of 0.7 μs and 4 of 0.25 μs for the WT, 10 of 0.7 μs and 10 of 0.6 μs for F128A and 20 of 0.8 μs starting from the FES WT clusters for M145S. The total simulation time sums up to *ca.* 57 μs.

FEL construction. The Free Energy Landscape (FEL) was constructed from the MD simulation data by combining functionalities of *pyEMMA* software package[6] (http://pyemma.org). After some tests, using the alpha carbon minimum group distance feature every 3 residues, resulted numerically feasible and therefore these *feats* were used as input coordinates. Then, the slow linear subspace of these particular definition of the system was estimated by computing the time-lagged independent component analysis (TICA).[7] A dimension reduction was achieved by projecting on the slowest TICA components: TIC1 describing the LID motion and TIC2 accounting for Asp239-loop movement. Then, uniform distance clustering[8] (*k*-means algorithm) was applied to obtain 30 clusters for the WT and M145S and 50 clusters for the F128A variant of the microstates that clearly distinguished the different protein conformations explored in the long MD simulations. Furthermore, volume calculations were performed for all the most populated and significant clusters with the computational tool POcket Volume MEasurer (POVME) 2.0[9] and Caver (v. 3.0; 0.7 Å radius probe).[10]

Scheme S1 | Protocol for FEL construction.



Input file for cpptraj module for computing the most significant conformations for each cluster/trajectory.

reference WT_Xray_MD.pdb
trajin cluster_conf_X.dcd

center origin :1-287
rms reference mass
cluster C1 rms :1-287 clusters 10 epsilon 3.0 summary summary.dat info info.dat repout cluster_conf_X repfmt pdb

Table S1 | Volumes in Å^3 of the different BmEH variants studied calculated with POVME 2.0.

Conformational State	$V_{ZONE\ 1}$ (Å^3)	$V_{ZONE\ 3}$ (Å^3)	V_{TOTAL} (Å^3)
1^{WT}	111	22	133
2^{WT}	163	4	167
3^{WT}	120	107	227
4^{WT}	128	489	**617**
1^{F128A}	131	0	131
2^{F128A}	24	96	120
3^{F128A}	211	420	**631**
4^{F128A}	116	72	188
1^{M145S}	177	80	257
2^{M145S}	150	0	150
3^{M145S}	78	111	189
4^{M145S}	220	424	**644**
5^{M145S}	120	88	208

Figure S1 | Free energy landscapes of **A**) WT, **B**) F128A and **C**) M145S with the most stable explored conformational states (**1-4**) indicated.

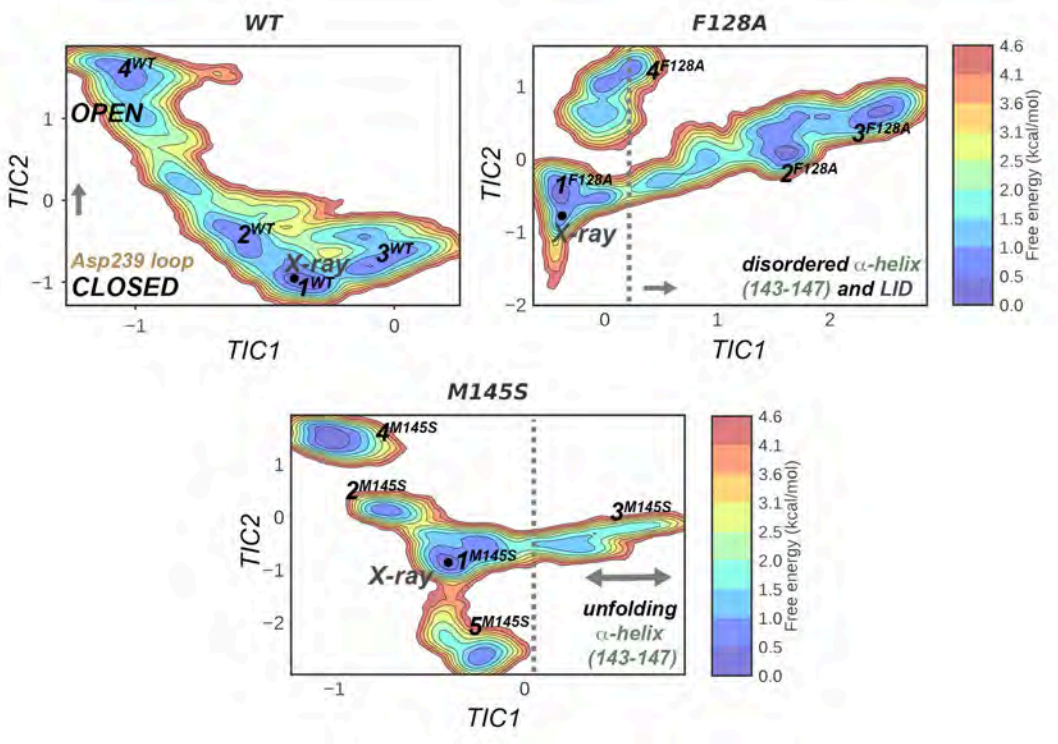


Figure S2 | Most stable conformations of BmEH M145S variant compared with X-ray wild-type structure. Conformations found in BmEH M154S (i.e. $\mathbf{1}^{M145S}$-$\mathbf{4}^{M145S}$) (colored as in **Figure 1.B.**) are matched by crystallographic wild-type structure of BmEH (gray).

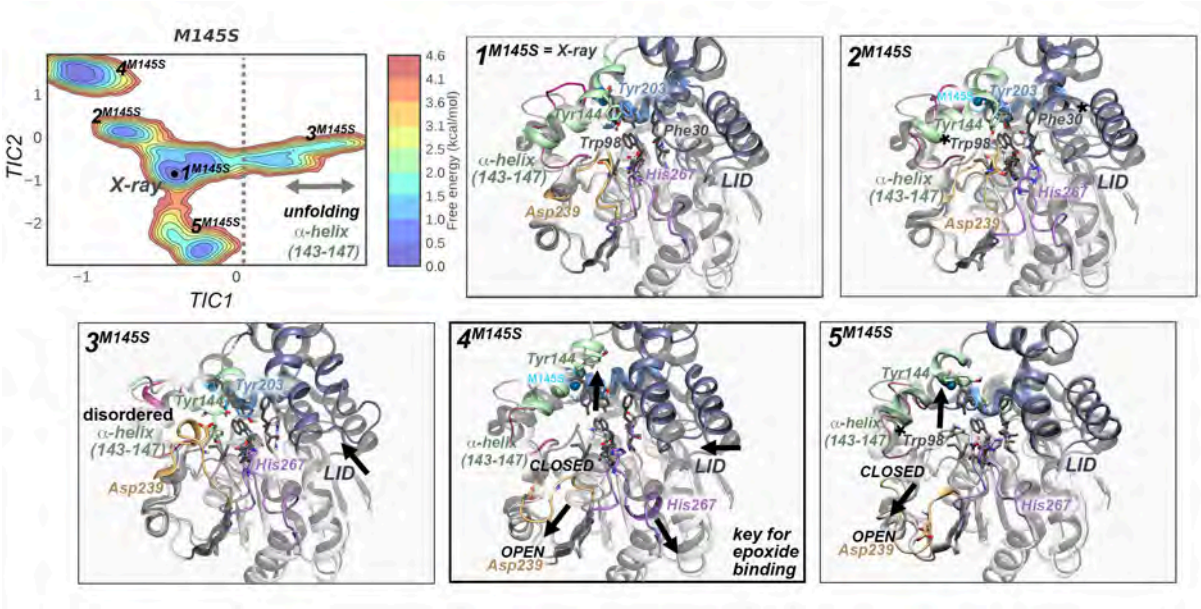
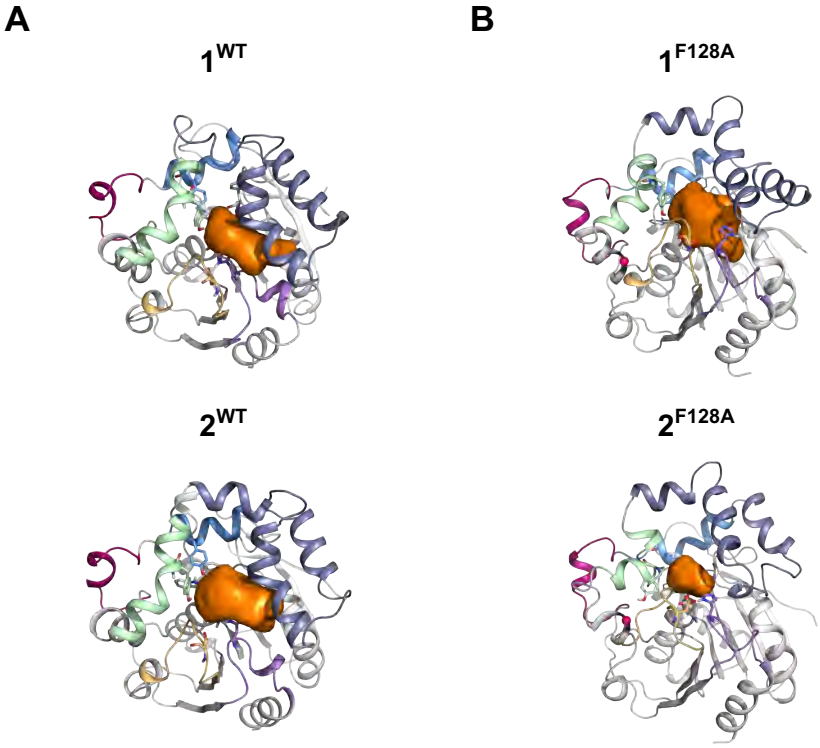
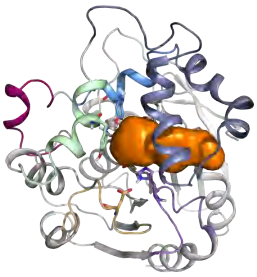


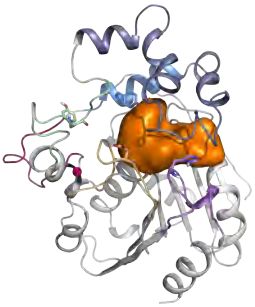
Figure S3 | Volume representation of substrate entrance (zone 1) and product release (zone 3) for **A**) WT (zone 1), **B**) F128A (zone 1), **C**) WT (zone 3), and **D**) F128A (zone 3) most stable explored conformational states (**1-4**). Volume calculations were carried out with POVME 2.0.



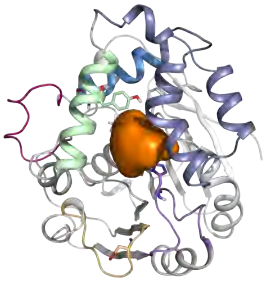
3^{WT}



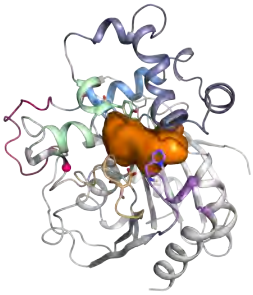
3^{F128A}



4^{WT}

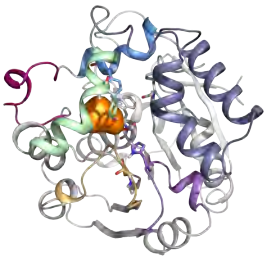


4^{F128A}

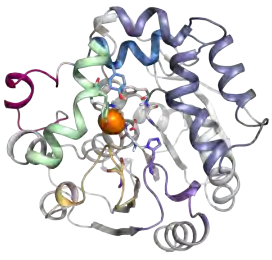


C

1^{WT}



2^{WT}

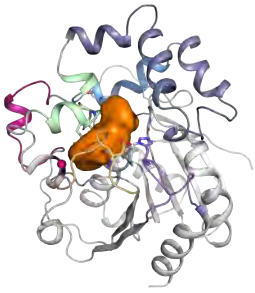


3^{WT}

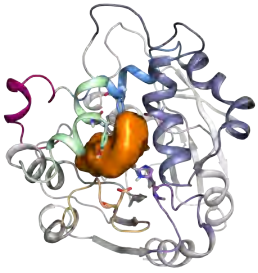
D

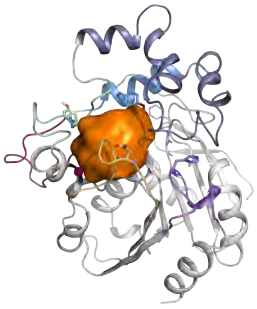
1^{F128A}

2^{F128A}



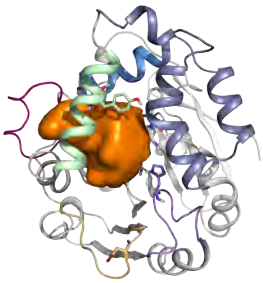
3^{F128A}





4^{WT}

4^{F128A}



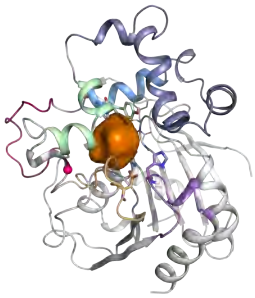


Figure S4 | Plot of root mean square fluctuation (RMSF) in Å of long MD trajectories for wild-type (WT), variant F128A and M145S.

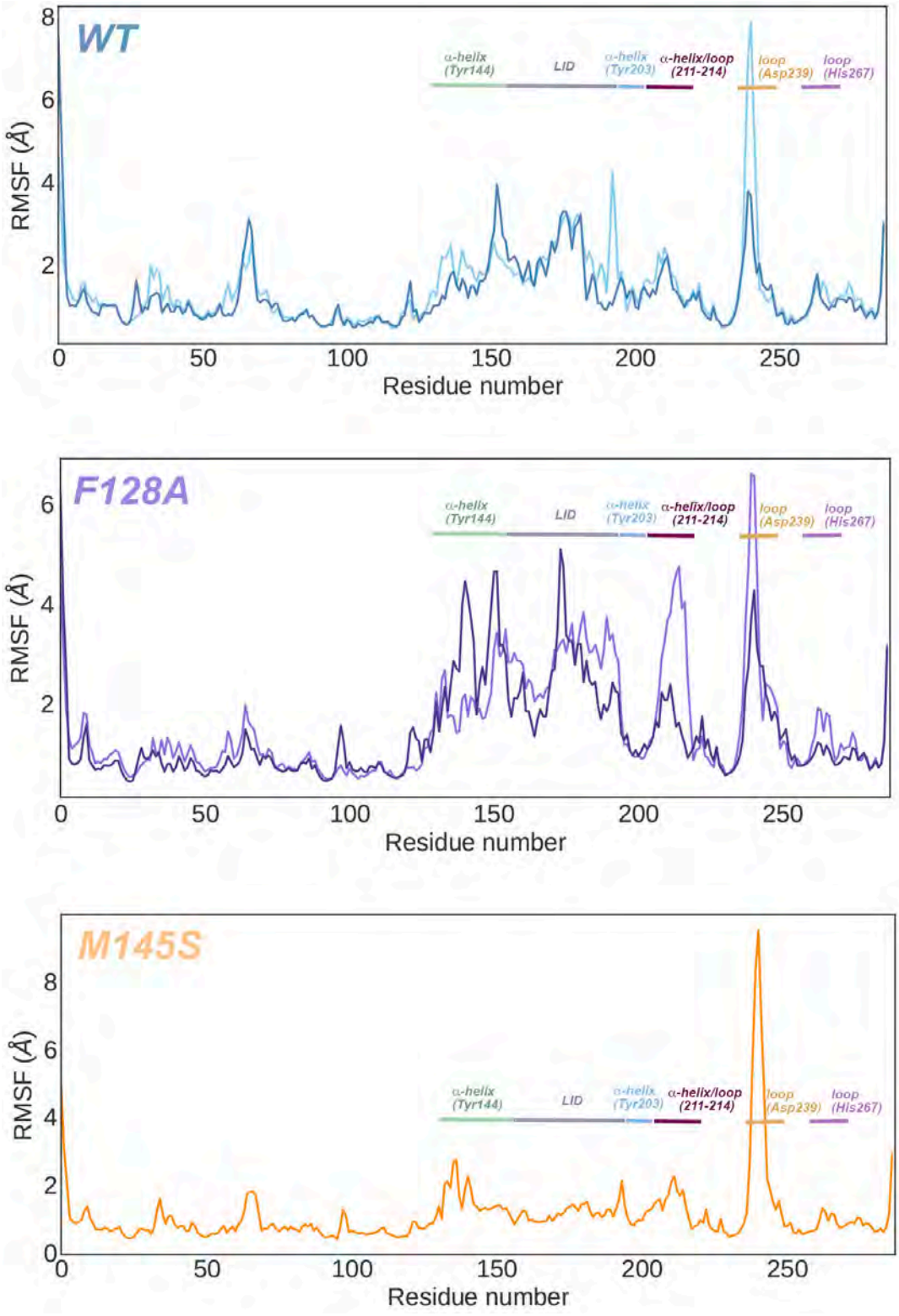
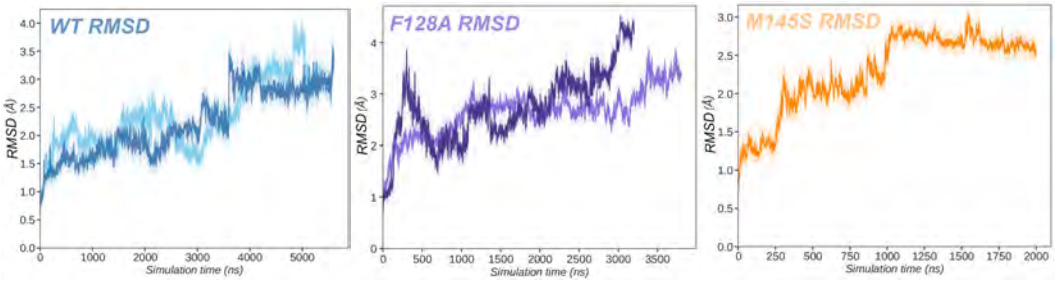


Figure S5 | Plot of root mean square deviation (RMSD) in Å of long MD trajectories for wild-type (WT), variant F128A and M145S.



References

[1] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 16, University of California, San Francisco, 2016.*

[2] R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537-W541.

[3] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926-935.

[4] J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049-1074.

[5] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.

[6] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.* **2015**, *11*, 5525-5542.

[7] a) G. Pérez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis and F. Noé, *J. Chem. Phys.* **2013**, *139*, 015102; b) C. R. Schwantes and V. S. Pande, *J. Chem. Theory Comput.* **2013**, *9*, 2000-2009.

[8] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.* **2011**, *134*, 174105.

[9] J. D. Durrant, L. Votapka, J. Sørensen and R. E. Amaro, *J. Chem. Theory Comput.* **2014**, *10*, 5047-5056.

[10] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLOS Comput. Biol.* **2012**, *8*, e1002708.

Chapter 6

Results and Discussion

In this Chapter 6 the main goals achieved in this thesis will be summarized and discussed in two main sections. According to both publications, this Chapter will start with the insights derived from the mechanistic studies of *Bm*EH stereo- and regioselectivity shown in Chapter 4. Afterwards, the discussion will continue with the impact of point mutations and the conformational dynamics of *Bm*EH, as well as how these amino acid exchanges might have an impact on its activity. Particularly, the effect on its activity towards bulky epoxide substrates, as described in Chapter 5. At the end of Chapter 6, a brief explanation about the link between both studies will be also provided.

6.1 Mechanistic Aspects on *Bm*EH's Stereoselectivity

The production of enantiopure epoxides, as well as 1,2-diols compounds, is of special interest, mostly because these molecules can be used as building blocks in synthetic chemistry, such as for the production of β-blocker pharmaceutical compounds. In general, achieving these compounds from cheap racemic mixture in high purities, that is, in high enantio- and regioselectivity, is challenging by means of both traditional synthetic chemistry and biocatalytic approaches. For instance, a common non-enzymatic strategy involves the

asymmetric epoxidation of olefins and enantioselective resolution of racemic epoxides using chemocatalysts [260]. Nevertheless, due to the inherent chirality of enzymes, in addition to some other advantages over metal or organic-based catalysts, the trend has been moving towards the development and the application of more sustainable and "greener" biocatalytic approaches. For instance, some enzymatic-based processes that allow to access enantiopure epoxides include the direct epoxidation of olefin substrate(s) by monooxygenases or peroxidases, the dehalogenation of halohydrin by halohydrin dehalogenases, and/or the epoxide ring-opening by epoxide hydrolases (EHs), which can catalyze the reaction selectively by distinguishing between racemic mixtures of epoxide substrate(s). That is, through the kinetic resolution of epoxide racemic mixtures (Fig. 6.1a). Opposite to most EHs enzymes, *Bm*EH shows an unusual (*R*)-selectivity towards phenyl glycidyl ether (**PGE**) and the bulky naphthyl glycidyl ether (**NGE**) substrates, both β-blockers precursors, while (*S*)-preference is observed towards the widely studied styrene oxide (**SO**) substrate (Fig. 6.1b). Worth to notice is that depending on the EH itself and the substrate's molecular structure (e.g. the substitution pattern of the epoxide substrate), different selectivities towards a specific substrate's enantiomer can be indeed observed in the resolution of epoxide racemic mixtures.

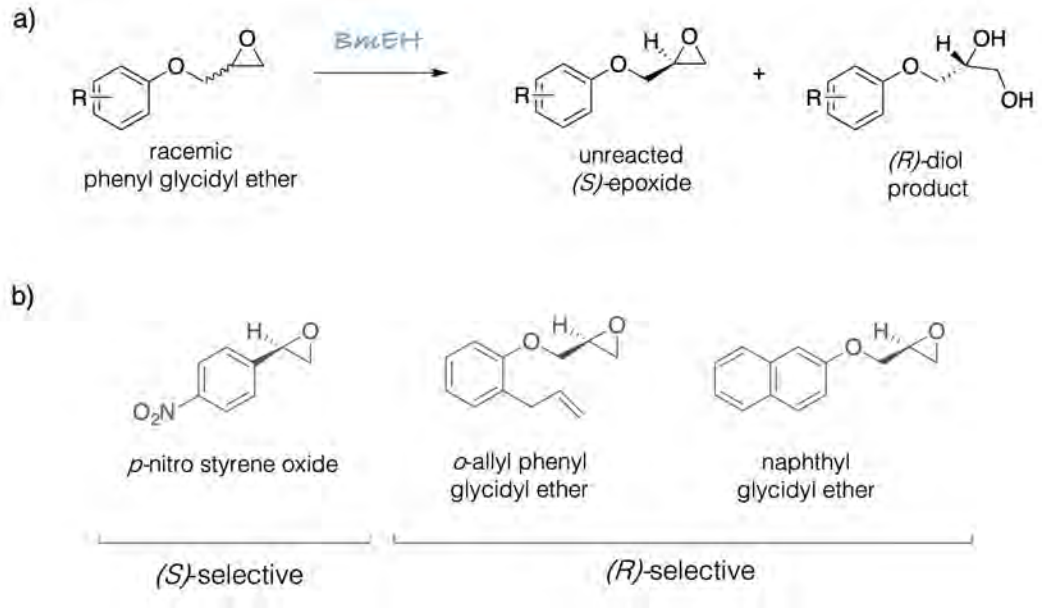


Figure 6.1: a) Schematic representation of the kinetic resolution of the terminal epoxide racemic mixture catalyzed by *Bm*EH and the molecular structure of *Bm*EH epoxide substrate(s). **b)** The substrates used in this study: styrene oxide (**SO**) and *p*-nitrostyrene oxide (***p*-NSO**), and two relevant building blocks of β-blocker drugs are also represented.

With the aim to address the origins of such preferences, in the present work we have particularly studied the mechanism for the epoxide ring-opening catalyzed by the *Bm*EH enzyme. We considered the nucleophilic attack at both oxirane carbon atoms of styrene oxide (**SO**) and its *para*-nitrostyrene oxide (***p*-NSO**) derivative. As Kong and co-workers identified a *Bm*EH active site tunnel in which three regions are defined (i.e. the substrate entrance, the active site, and the product release), our mechanistic studies have also been performed considering two different orientations of the substrates at the *Bm*EH pocket, that is, the aromatic moiety pointing towards the substrate entrance (i.e. *region 1*), and pointing towards the product release (i.e. *region 3*) (Fig. 6.2).

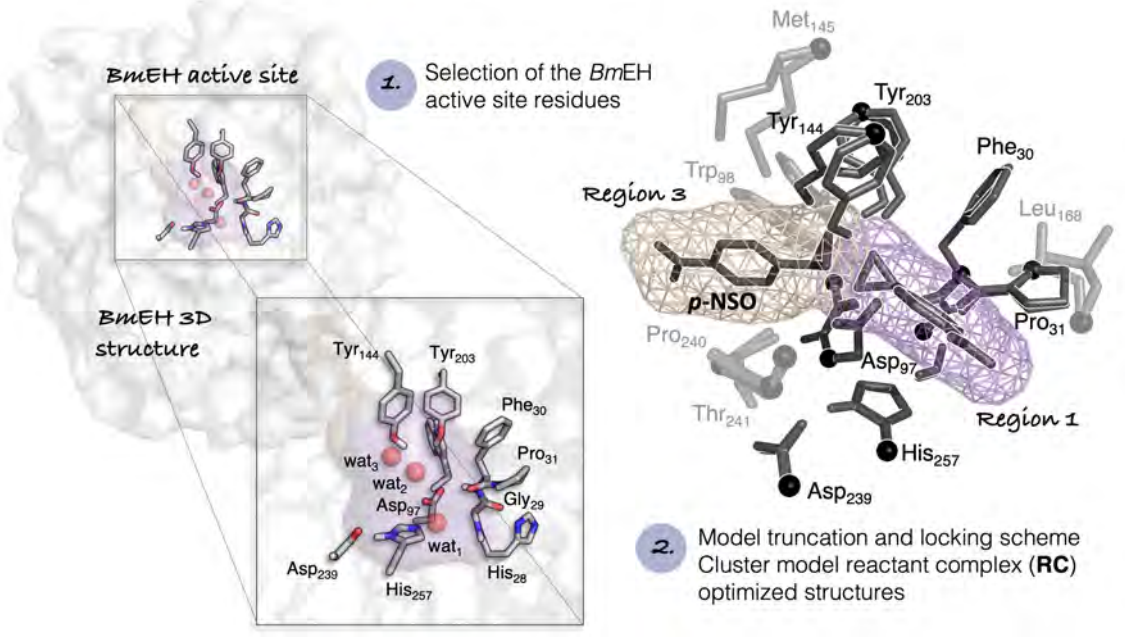


Figure 6.2: Schematic representation of the construction of *Bm*EH active site model. Representation of the X-ray *Bm*EH wild-type structure (PDB ID: 4NZZ) with the active site catalytic residues highlighted. The truncated and the locking scheme (in black spheres and asterisks) of the *Bm*EH are also represented, as well as the two plausible orientations of the substrate ***p*-NSO** inside the *Bm*EH active site (*region 1* or substrate entrance is colored in purple, while *region 3* or product release is colored in yellowish).

In more detail, in this study our mechanistic investigations have been carried out using the quantum mechanical cluster approach (CM) [231, 232]. Our cluster model has been constructed following the same protocol as done in previous studies by Himo and co-workers. However, our truncated model has been particularly devised based on the X-ray *Bm*EH wild-type structure (PDB ID: 4NZZ), and also from the information gathered on our 1000 *ns*

MD simulations on the alkyl-enzyme intermediate (**Int1**) complex (Fig. 6.3). In addition to the analysis of the geometric parameters on each of the optimized reaction intermediates, in order to rationalize the selectivities observed on the energy profile outcomes non-covalent interactions in selected transition states (TSs) were also analyzed using the NCIplot tool [248].

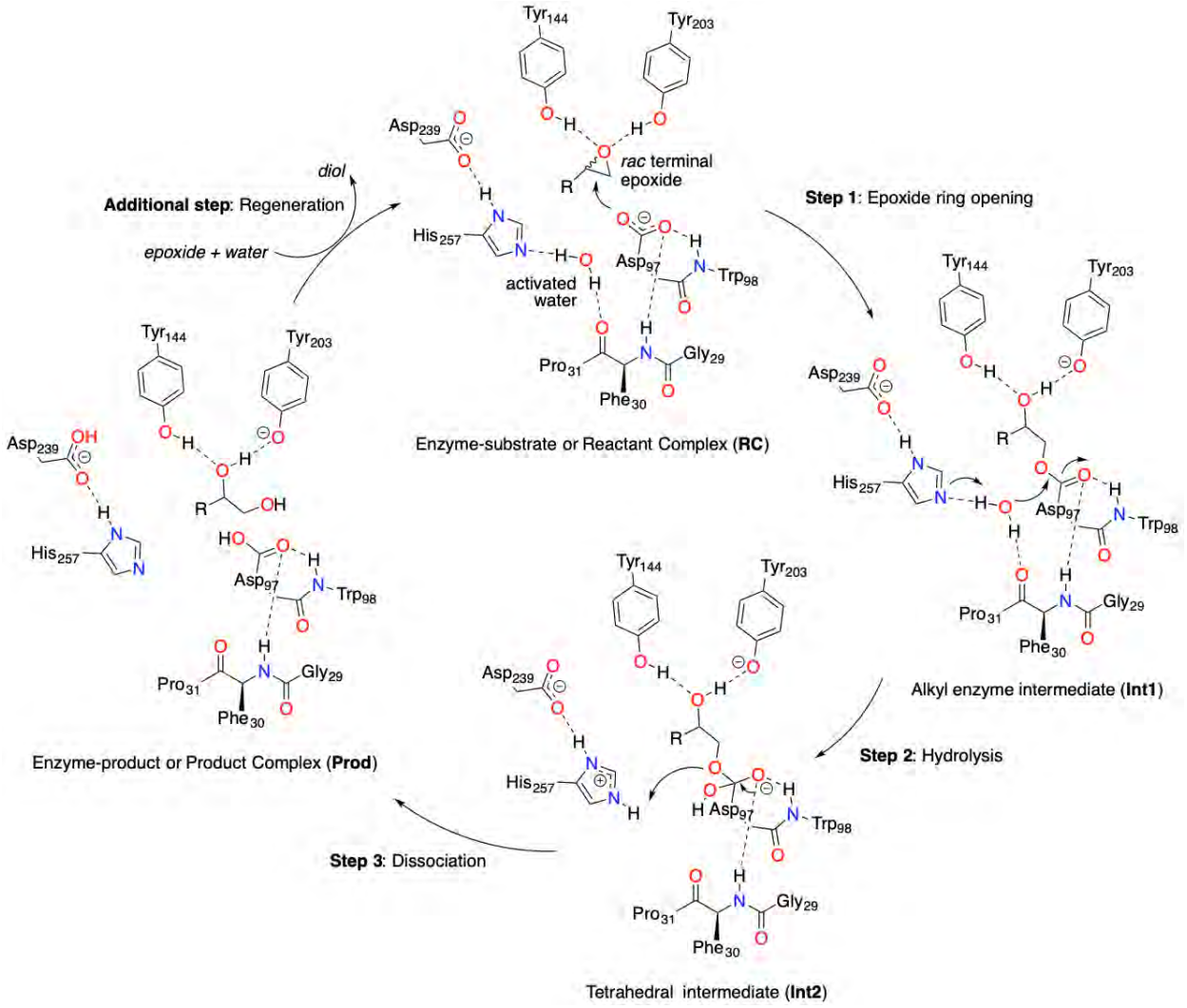


Figure 6.3: General reaction mechanism scheme for α,β-hydrolase EHs. Catalytic residues are labelled according to the X-ray structure of *Bm*EH wild-type (PDB ID: 4NZZ).

As shown in Fig. 6.2, the position of two crystallographic waters in the active site (i.e. wat_2 and wat_3) indicates two plausible locations for binding the oxygen of the epoxide substrate. According to the proposed EHs mechanism (Fig. 6.3), the first step is assisted by two tyrosine residues. As shown in the overlay of the optimized enzyme-substrate (i.e. reactant complex,

RC) truncated model structures for both (*R*) and (*S*) ***p*-NSO** enantiomers, the epoxide establishes hydrogen bond interactions with Tyr_{144} and Tyr_{203}. Regardless the region where the phenyl moiety is pointing, these interactions facilitate the epoxide ring-opening during the nucleophilic attack from Asp_{97}. Nucleophilic Asp_{97} is in turn properly oriented thanks to hydrogen bond interactions between the amide backbone groups of Asp_{97}-Trp_{98} and the Gly_{29}-Phe_{30} oxyanion hole residues. Furthermore, regarding the second step of the reaction mechanism, the ion-pair charge relay system formed by Asp_{239} and His_{267} establishes a hydrogen bond to activate the water molecule responsible for the hydrolysis of the covalently bound alkyl-enzyme intermediate (**Int1**) (Fig. 6.3). As it is also shown in Fig. 6.2, this active site water is maintained fixed by hydrogen bonds between the nitrogen atom of the base His_{267} and the carbonyl group of the oxyanion hole residues Phe_{30}-Pro_{31}. The formation of the second covalently bound tetrahedral intermediate (**Int2**) is generally accepted to be the rate-determining step of the α,β-hydrolase EH mechanism. And thus, once it is formed, it is widely accepted that the tetrahedral intermediate **Int2** quickly rearranges and dissociates on the last step to afford the final diol product (**Prod**).

As shown in Figure 6.3, to regenerate the catalytic cycle (i.e. the binding of a new epoxide and water molecule, and recover the initial state of the catalytic residues) an additional acid-base step is required (i.e. **Prod** $\longrightarrow$ **RC**). Although this process cannot be properly described within the cluster model framework, as proposed by Himo in previous studies [233], by computing the free energy of the corresponding $epoxide + water \longrightarrow diol$ reaction scheme, the overall energetics can be roughly estimated. For clarity, in this present discussion we have selected the two lowest energy profiles outcomes for each enantiomer, which have been computed at the B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory (Fig. 6.4). In the first step of the reaction, the Asp_{97} can attack either the C_1 or the C_2 of the epoxide ring to form an ester intermediate. According to our computed energy profiles, the most hindered C_1 position is kinetically favored regardless the enantiomer, forming the **Int1** intermediate inducing to an inversion of the configuration at C_1 (**Int1-S-C1**). The lowest alkylation transition state for (*S*)-***p*-NSO** (i.e. **TS1-S-C1**) is 1.8 $kcal \cdot mol^{-1}$, while for the opposite enantiomer the equivalent transition state (i.e. **TS1-R-C1**) is 3.7 $kcal \cdot mol^{-1}$ higher. In both **TS1** optimized structures similar catalytic distances are observed. Therefore, in order to investigate the additional stabilization of the **TS1-S-C1**, we analyzed non-covalent interactions occurring with surrounding residues with the NCIplot tool (Fig. 6.5). Although

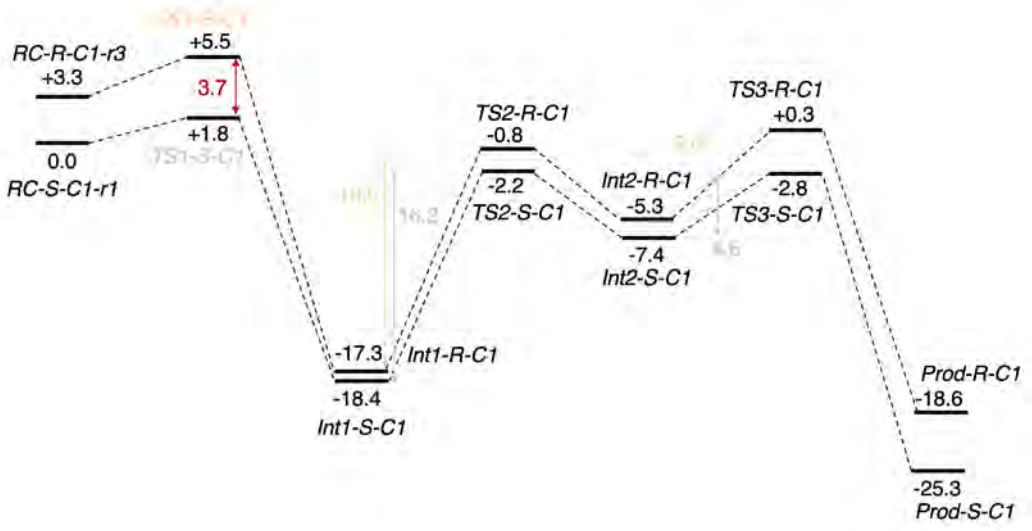


Figure 6.4: Representation of the two kinetically favored computed energy profiles for the epoxide hydrolysis mediated by *Bm*EH enzyme for each ***p*-NSO** enantiomer at the B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory. All relatives energies are in $kcal \cdot mol^{-1}$ and referenced for each case to its lowest RC structure (**RC-S-r1** and **RC-R-r1**, respectively).

the same regioselectivity towards C_1 attack is observed for both enantiomers of ***p*-NSO**,it is worth to point out that each enantiomer is positioned differently inside the *Bm*EH active site, being the (*S*)-*p*-**NSO** oriented towards *region 1* (or nearby His_{267}) and (*R*)-*p*-**NSO** placed in *region 3* (or nearby Trp_{98}). According to these findings, the former can be further stabilized through π-stacking interactions between the substrate and the catalytic His_{267} residue (Fig. 6.5), such non-covalent interactions cannot indeed be established for the lowest path obtained for the *R* of the substrate.

For the case of the (*S*)-***p*-NSO** enantiomer, the lowest in energy **TS1-S-C1** displays a 2.25 Å distance between the side-chain oxygen of Asp_{97} and the C_1 atom of the epoxide substrate, in addition to an elongation of the single bond between the oxygen and the C_1 of the epoxide (up to 1.88 Å), as well as the O-H distance of Tyr_{203} (up to 1.00 Å). The optimized structure of the protonated alkyl-enzyme intermediate (**Int1-S-C1**) suggests that the opening of the epoxide Asp_{97} and its protonation by Tyr_{203} is taking place in a concerted manner, in line with an S_N2 reaction mechanism.

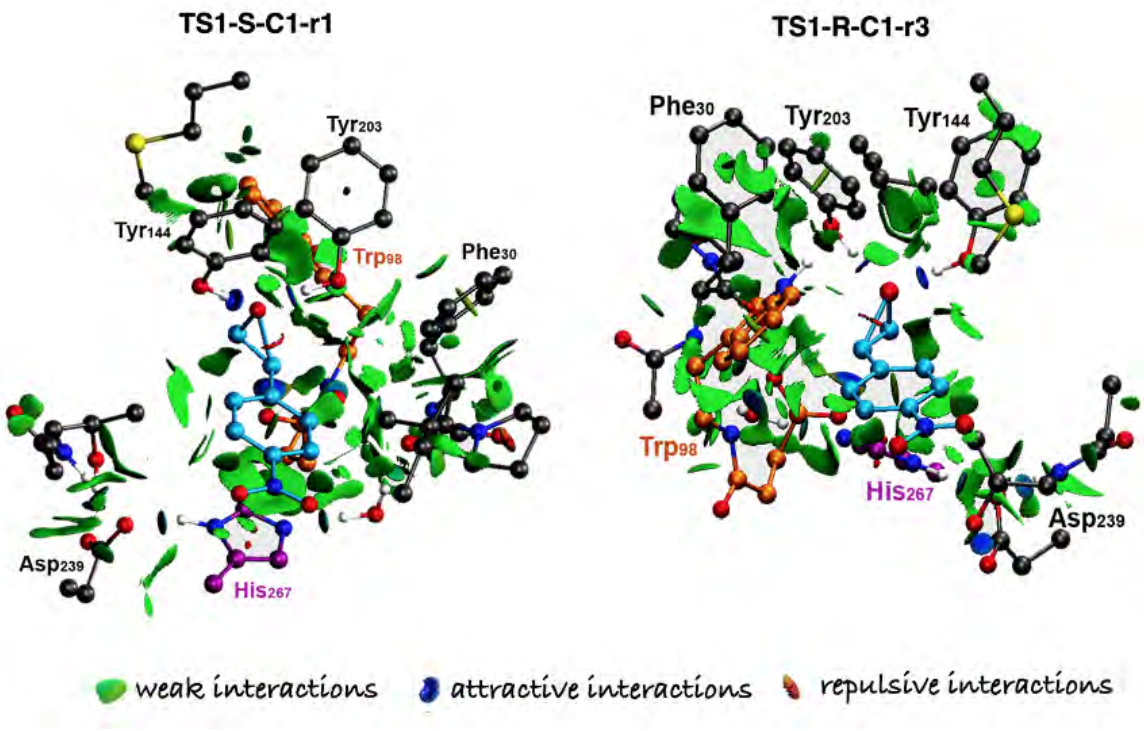


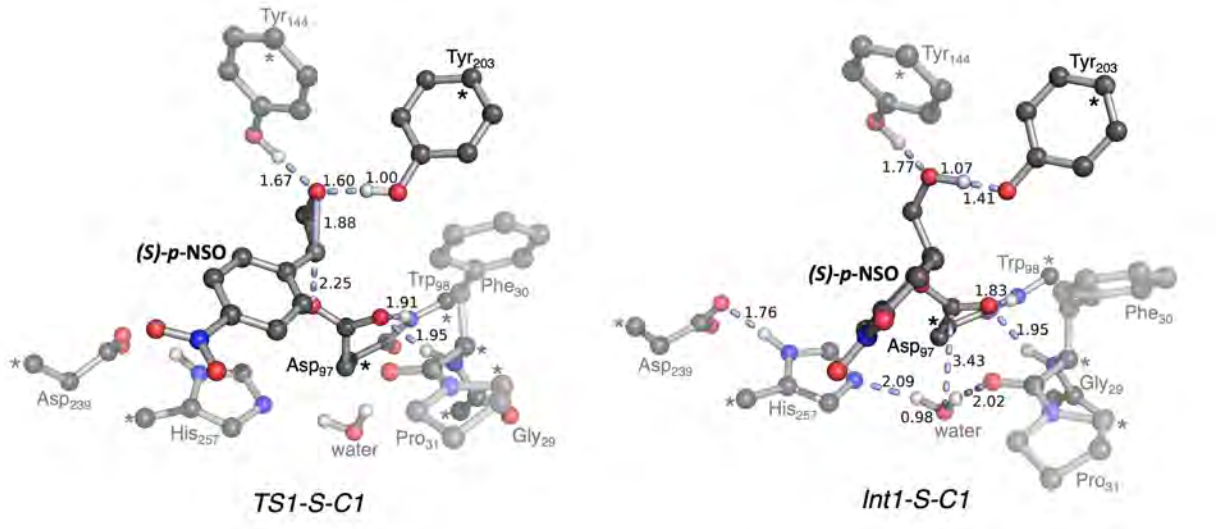
Figure 6.5: Representation of non-covalent interactions (NCIs) within the lowest in energy alkylation transition states. **TS1-S-C1-r1** and **TS1-R-C1-r3** stands for the *rac-p*-**NSO** epoxide ring-opening reaction using the NCIplot. NCI surface shows intermolecular interactions between the epoxide substrate and the active site residues included in the cluster model. All residues are represented as balls and sticks and the two possible orientations of each substrate are indicated for each case (His_{267} stands for *region 1*, and Trp_{98} stands for *region 3*). Only relevant catalytic residues are labelled regarding *Bm*EH wild-type numbering (PDB ID: 4NZZ). Non-polar hydrogens are omitted for clarity.

Regarding the second step of the reaction mechanism, similar energy barriers were found for both enantiomers (i.e. 16.2 $kcal \cdot mol^{-1}$ and 16.5 $kcal \cdot mol^{-1}$ for (*S*)- and (*R*)-***p*-NSO**, respectively). Although both alkyl-enzyme intermediates are highly stabilized (by 18.4 for the **Int1-S-C1** and 17.3 $kcal \cdot mol^{-1}$ for the **Int1-R-C1**), the interactions established by these structures indicate a good pre-organization to overcome the following reaction barrier. As indicated on Fig. 6.6, all distances have been shortened at **TS2** to hydrolyze the alkyl-enzyme intermediate and generate the tetrahedral intermediate **Int2**.For instance, in this step the active site water molecule has to be activated by the catalytic His_{267}, which is at a distance of 2.09 Å on **Int1-S-C1**. In turn, the basic character of His_{267} is enhanced by the hydrogen bond established with Asp_{239} at a distance of 1.76 Å. In addition, the oxygen atom of the Asp_{97} is properly oriented and interacting at a **Int1** distance of 1.83 Å and 1.95 Å with the

amide backbone of the oxyanion hole residues (Trp_{98} and Phe_{30}, respectively), relevant to stabilize the negative charge that is being formed at this position during this step.

In the last step of the mechanism, similar activation energy barriers are observed for both enantiomers, being 4.6 $kcal \cdot mol^{-1}$ for **TS3-S-C1** with respect to **Int2**, whereas **TS3-R-C1** is 5.6 $kcal \cdot mol^{-1}$. The analysis of the non-covalent interactions on the optimized model **TS3** suggests that **TS3-R-C1** is higher in energy due to the lack of π-stacking interactions with His_{267}, because the phenyl group of the substrate is pointing towards *region 3*. As shown in Fig. 6.3, in this final step of the reaction, the diol product is generated after the oxygen atom (originally from Asp_{97}) takes a proton from His_{267} at the same time that the C-O bond between the substrate and the nucleophile is broken. This is in line with the observed distances in our **TS3** optimized structures. For instance, in **TS3-S-C1** an elongated C-O distance up to 1.79 Å is observed, while the distance between the now protonated His_{267} and the substrate is at 1.48 Å (Fig. 6.6).

Therefore, by comparing the lowest energy profiles obtained for both enantiomers, our findings suggest that overall the selectivity of *Bm*EH enzyme towards the hydrolysis of ***p*-NSO** is determined during the first alkylation step of the reaction, being the *R* enantiomer less stabilized due to the lack of π-stacking interactions with the surrounding active site residues.



(continue in the next page)

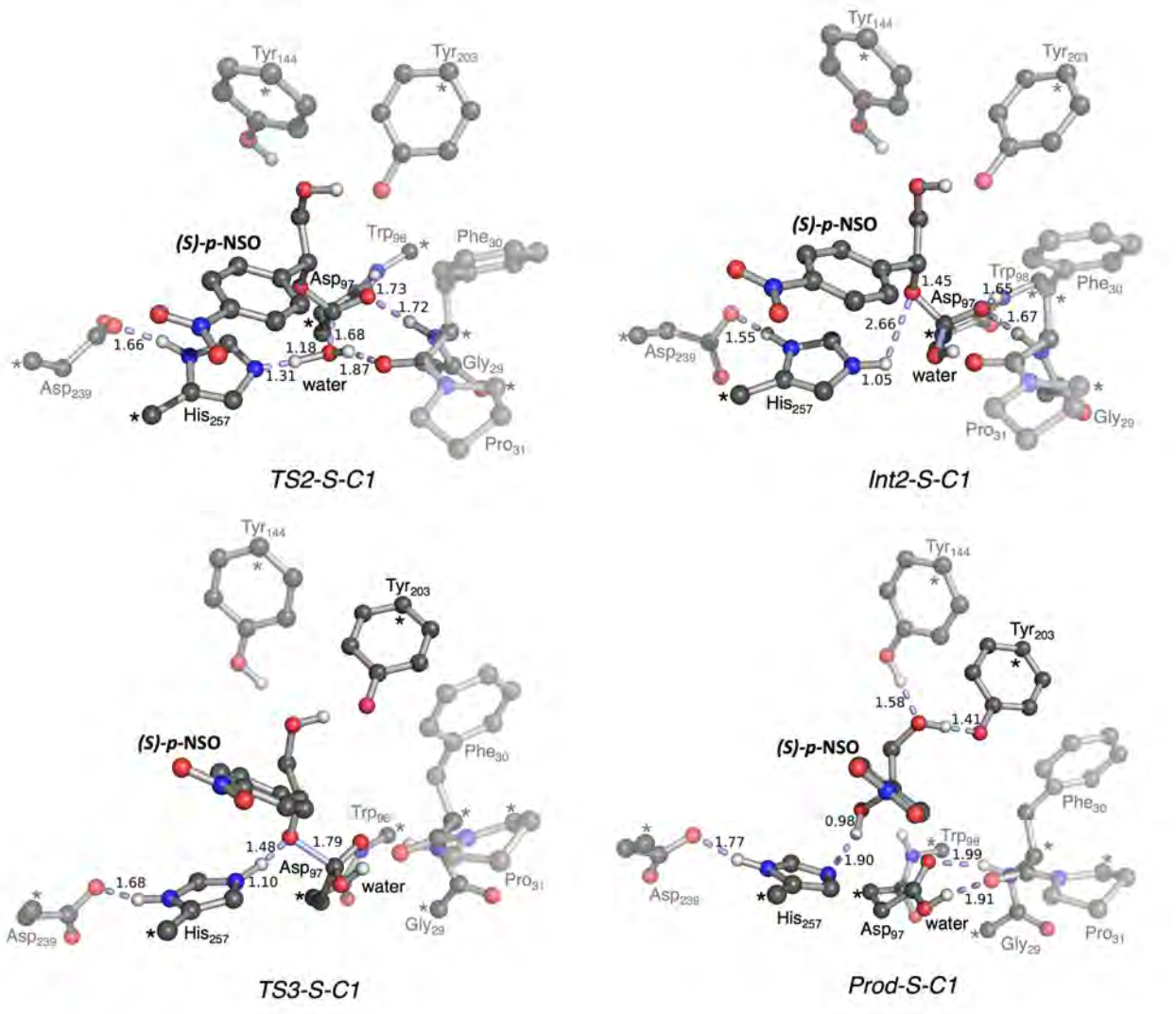


Figure 6.6: Selected optimized cluster model structures for the hydrolysis of (*S*)-***p*-NSO** substrate mediated by *Bm*EH throughout the attack at the benzylic position (C_1) (**S-C1-r1**) at B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP-D3BJ/6-31g(d) level. For clarity, a simplified version of the truncated models are shown, non-polar hydrogens are omitted, and only a selected set of distances and residues are shown regarding the chemical process occurring on each step of the reaction mechanism. Active site pocket residues are shown as sticks and spheres. Depending on the reaction step different transparencies are displayed, and bonds breaking or forming are pointed out in blue color. All distances are in Å. Asterisks indicate the applied locking scheme.

With the aim to investigate the effect of the *nitro* (NO_2) substituent on the *Bm*EH selectivity, the energy profiles for both styrene oxide (**SO**) enantiomers were also computed at the same level of theory. Our calculated energy profiles describe a similar trend as for ***p*-NSO** derivative, being the attack at the benzylic position C_1 the preferred path in both *S* and

R enantiomers, although their orientations in the active site are not the same. As in the case of ***p*-NSO**, the lowest activation energy barrier imposed by **TS1** is found for the *S* enantiomer when the phenyl group is pointing towards *region 1* (**S-SO-C1-r1**), whereas the most favored path for the opposite enantiomer is when the phenyl group is located on *region 3* (**R-SO-C1-r3**).

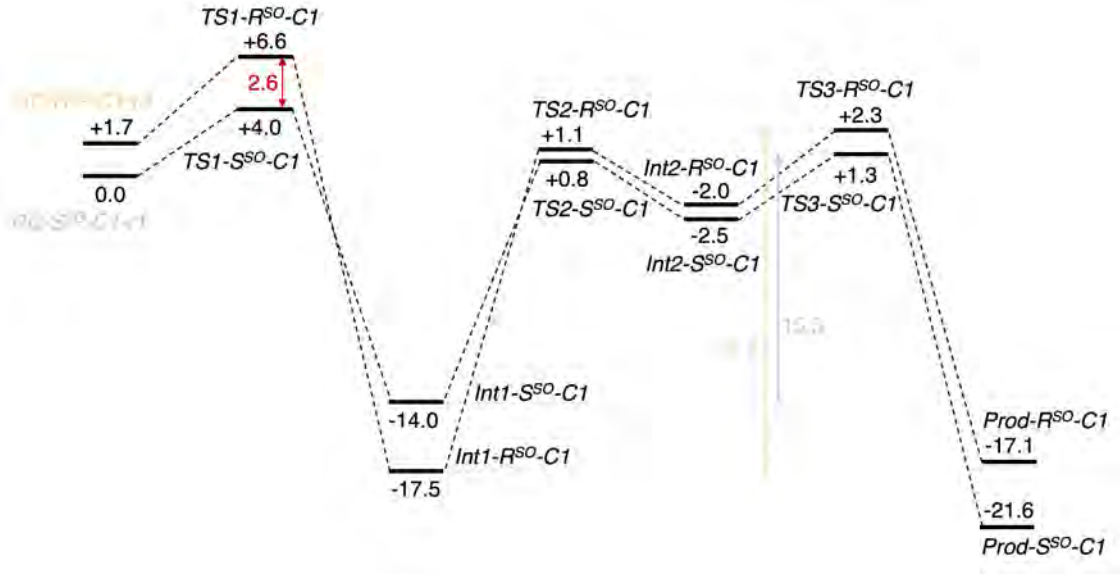


Figure 6.7: Representation of the two lowest computed energy profiles for the styrene oxide (**SO**) hydrolysis catalyzed by *Bm*EH enzyme considering the two enantiomers of **SO** at the B3LYP-D3BJ/6-311+g(2d,2p)//B3LYP/6-31g(d,p) level of theory. All relatives energies are in $kcal \cdot mol^{-1}$ and referenced for each case to its lowest RC structure (**RC-S-SO-r1** and **RC-R-SO-r1**, respectively).

As shown in Fig. 6.7, the activation barrier for **TS1-S-SO-C1** in *region 1* is 4.0 $kcal \cdot mol^{-1}$, while this value is 2.6 $kcal \cdot mol^{-1}$ higher for **TS1-R-SO-C1** which is oriented towards *region 3*. Although similar energies are calculated for each **TS3** structures, the alkyl-enzyme intermediate (**Int1**) for the case of the *R* enantiomer is stabilized by 3.5 $kcal \cdot mol^{-1}$. In other words, a total barrier of 19.8 $kcal \cdot mol^{-1}$ has to be crossed to generate the diol from the *R* enantiomer, while a total barrier of 15.3 $kcal \cdot mol^{-1}$ is found for the rate-determining step of *S*-**SO**. According to these findings, a substituent on **SO** at the *para* position, like a nitro group, is not playing a major effect on *Bm*EH selectivity.

6.2 Effect of Mutations in *Bm*EH Conformational Landscape

Most of the structurally characterized EHs up to date belong to the α,β-hydrolase superfamily, sharing an α,β-sheet core domain and a flexible lid domain covering the enzyme active site. The catalytic machinery of α,β-EHs consists on a highly conserved catalytic triad (i.e. $Asp_{97} - His_{267} - Asp_{239}$), the ion-pair charge relay (i.e. $Asp_{239} - His_{267}$), the oxyanion hole residues (i.e. $Phe_{30} - Trp_{98}$), and two acidic tyrosine residues (i.e. Tyr_{144} and Tyr_{203}) located in the lid that are responsible for the substrate activation and recognition (Fig. 6.8).

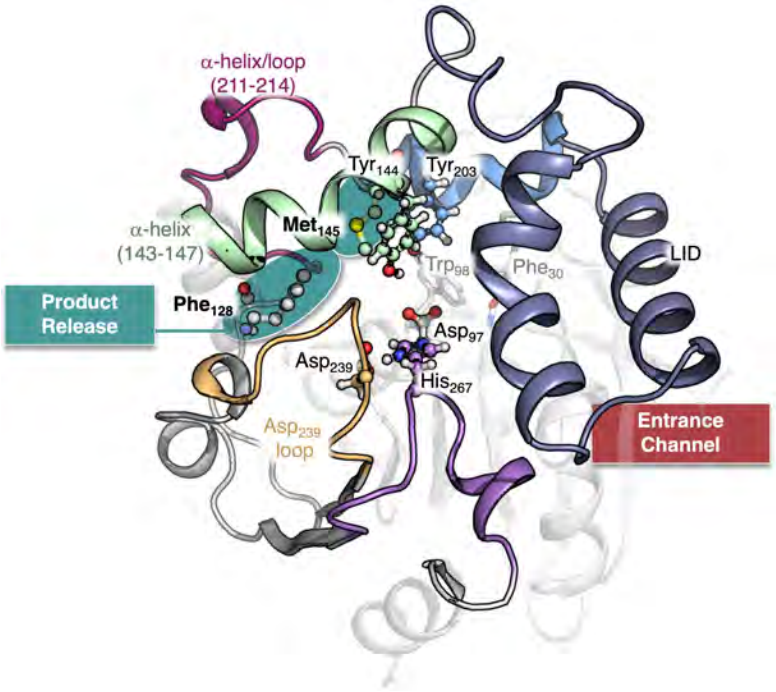


Figure 6.8: Schematic representation of *Bm*EH enzyme wild-type structure with relevant active site residues and secondary structures highlighted: the loop containing the catalytic triad residues Asp_{239} in gold, His_{267} in violet, and the nucleophilic Asp_{97} in grey; the α-helix of Tyr_{144} in green, and the α-helix of Tyr_{203} in blue. Substrate entrance is indicated in red, product release in teal. Residues at positions Phe_{128} and Met_{145} that were mutated are also represented.

As enzymes are not static, different conformational states existing in dynamic equilibrium are sampled in solution, as well as along the whole catalytic cycle. The ability to adopt alternative conformations other than the native state, which is a concept known as conformational heterogeneity, might be relevant towards enhanced activities or novel functionalities. Such conformational states can be mapped and identified by constructing what is known as

conformational free energy landscapes (FEL), a simplified picture that shows the population of each state and the height of the barriers separating them. As described in Section 1.4, each conformational free energy landscape is linked to a particular set of conditions (i.e. the enzyme sequence, temperature, pressure, etc.). Therefore, the introduction of mutations can modify the conformational free energy landscapes, thereby leading to what is known as the population shift concept. In the particular case of *Bm*EH, the insertion of mutations on the product release region (i.e. *region 3*) at position Phe_{128} and Met_{145} have been experimentally found to alter dramatically the ability of *Bm*EH to accept bulky *o*-allyl-phenyl glycidyl ether and naphthyl glycidyl ether (**NGE**) substrates compared to the wild-type *Bm*EH. Therefore, with the aim to identify the most relevant conformational states of the wild-type *Bm*EH enzyme, as well as its variants, in the present work we have partially reconstructed the conformational free energy landscapes for each *Bm*EH variant (i.e. wild-type, singly-mutated **F128A**, and **M145S**) by computing short and long MD simulations. In addition, to rationalize the experimentally enhanced activities of *Bm*EH variants, apart from the analysis and comparison of the most relevant conformational changes compared to the wild-type *Bm*EH X-ray structure, active site volumes of representative structures have been estimated using the POVME tool [261].

To investigate the impact of mutations on *Bm*EH ability to accept bulkier substrates, a FEL for wild-type *Bm*EH (**WT**) was build from an accumulated MD simulation time of 19 microseconds in the *apo* state (i.e. in the absence of any ligand), while singly-mutated variants **F128A** and **M145S** FELs were constructed from 20 and 18 microseconds (μs), respectively (Fig. 6.9). In a nutshell, all these accumulated high dimensional MD data was reduced by applying the time-lagged independent component analysis (TICA) technique (for further details, see Chapters 2 and 5). The FEL is represented regarding the two slowest TICA components. Considering the minimum group distance of the Cα applied every 3 residues. The physical meaning of the two slowest TICA components identified (i.e. TIC_1 and TIC_2) describe the disorder of the alpha-helix and lid domain of the *Bm*EH enzyme (i.e. x), and the conformational variation of the loop containing the catalytic residue Asp_{239} (i.e. y) (Fig. 6.9).

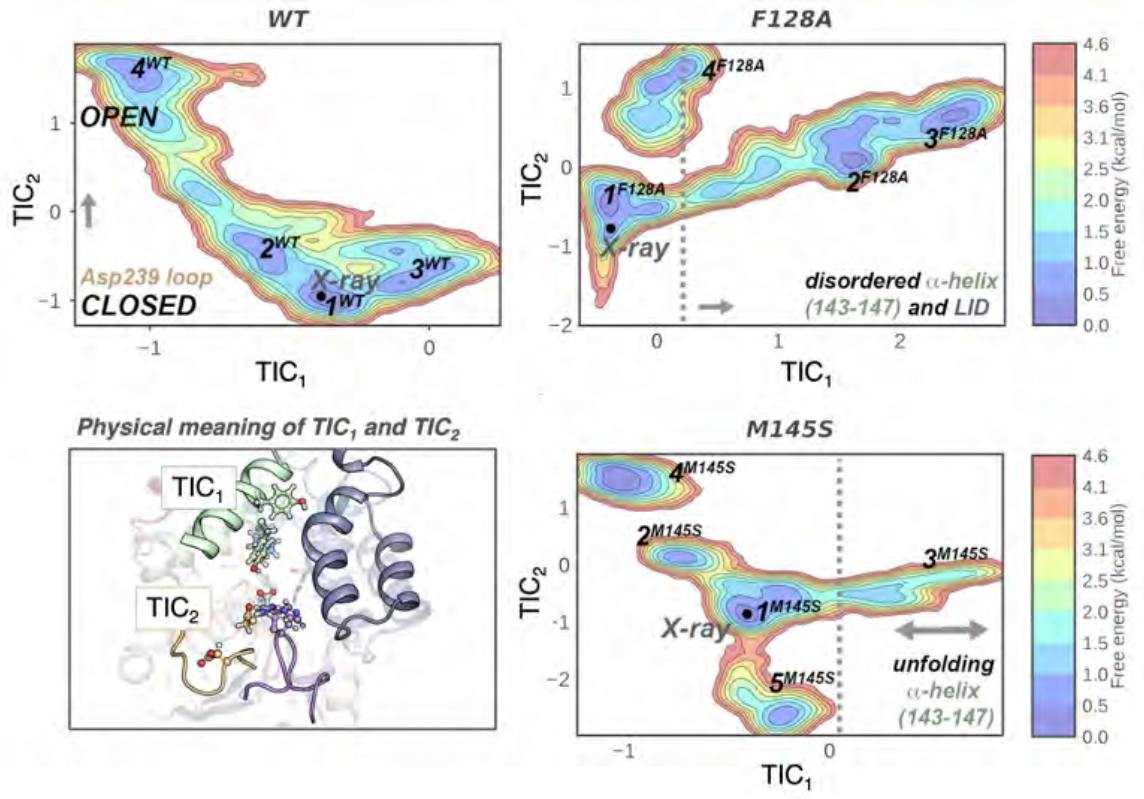


Figure 6.9: Representation of partially reconstructed *Bm*EH wild-type, and singly-mutated **F128A** and **M145S** variants free energy landscapes. The physical meaning of each TICA component is also indicated.

As shown in Figure 6.9, four major conformational states for the wild-type enzyme and its **F128A** variant were identified, while for the **M145S** variant an additional conformational state was observed. Note, however, that the conformational space explored by the wild-type enzyme in our simulated time is indeed smaller than for the singly-mutated variants, thereby suggesting that these additional states visited after the insertion of mutations **F128A** and **M145S** might be relevant towards the acceptance of bulkier substrates.

In more detail, regarding the conformations explored by the wild-type *Bm*EH (Fig. 6.10), its most populated and representative conformation (i.e. 1^{WT}) is structurally similar to the crystallographic structure, as all residues involved in the first step of the mechanism are properly oriented for catalysis. In particular, the two tyrosine residues responsible for the substrate recognition and binding are well oriented, as well as the catalytic triad $Asp_{239} - His_{267} - Asp_{97}$ to perform the first nucleophilic attack. Regarding the second

conformational state (i.e. 2^{WT}), the most relevant deviations from the X-ray structure correspond to conformational changes of the oxyanion hole residues Phe_{30} and Trp_{98} side-chains. Furthermore, in the third conformational state (i.e. 3^{WT}), subtle changes regarding the loop containing the Asp_{239} of the ion-pair charge relay and the lid domain are observed. Compared to the crystallographic structure (PDB ID: 4NZZ), 3^{WT} is characterized for having a partially open conformation of the lid. Although the 4^{WT} conformational state is separated by a 3 $kcal \cdot mol^{-1}$ free energy barrier from the three previously discussed states, relevant conformational changes that lead to a wider active site *Bm*EH pocket are observed. On one hand, a deviation of the lid affects the orientation of the Tyr_{144}, which assists the binding of the epoxide and the alkylation step of the reaction (Fig. 6.3). On the other hand, a conformational change of the loop containing Asp_{239} leads to a lid open conformation, both changes leading to wider active site pocket (Fig. 6.10).

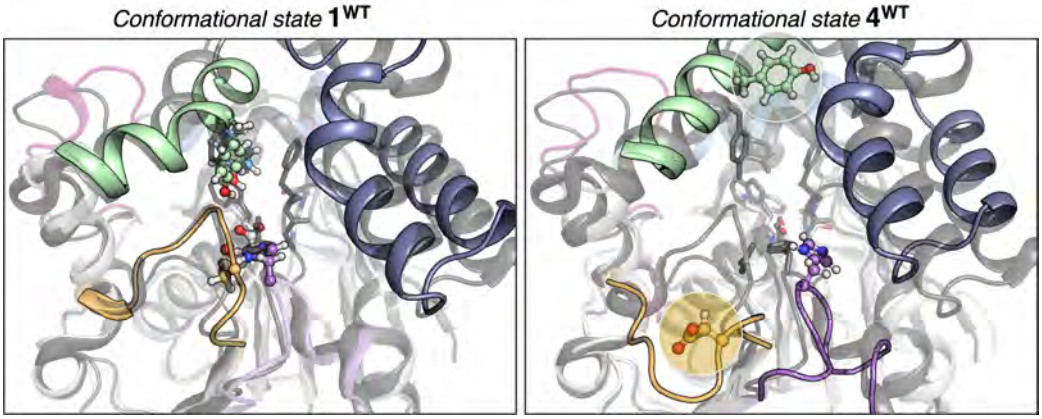


Figure 6.10: Representation of *Bm*EH wild-type conformational states 1^{WT} and 4^{WT}. *Bm*EH wild-type conformations are overlaid to the X-ray of *Bm*EH (PDB ID: 4NZZ, shown in gray). The most relevant conformational side-chain changes are highlighted with residues represented in bold spheres and sticks. Bold cartoon is also used to emphasized the other relevant secondary elements of *Bm*EH conformational states that are deviating from the crystallographic structure.

Although these changes might be relevant for the acceptance and binding of epoxide substrates, noteworthy to mention is that the catalytically relevant residues are not found in this conformational state in a productive orientation for catalysis. For instance, regarding the first step of the mechanism, Tyr_{203} side-chain is not properly oriented in 4^{WT} to assist the binding and to act as a proton donor to the epoxide during the nucleophilic attack by Asp_{97}. In a similar way, in the third conformational state 3^{WT}, the Asp_{239} is not properly positioned to interact with His_{267}, an interaction required to activate the catalytic water molecule

on the second step of the reaction. In this line, opposite to the 1^{WT} conformational state, 2^{WT} is not displaying a good pre-organization of the oxyanion hole residues responsible to stabilize the negative charge that arises along the catalytic cycle. Nevertheless, in spite of the deviations from the ideal productive arrangement, the active site volume calculations suggest that these conformational changes might be indeed necessary for effectively binding bulky epoxide substrates like **NGE**. On the other hand, when the catalytic machinery is well oriented, only significantly small volumes are observed. The estimated total volumes (i.e. the sum of the independently calculated volumes in *region 1* and *3*) for 1^{WT}, 2^{WT}, and 3^{WT} are 133, 170, and 230 Å^3, respectively, while remarkably, an estimated total volume of 617 Å^3 is found for the 4^{WT} conformation. Nevertheless, the estimated volumes in *region 1* are quite similar in all wild-type conformational states, while the enlargement of the active site volumes in 4^{WT} mainly occurs due to the wider volume found in *region 3*, as shown in Figure 6.11.

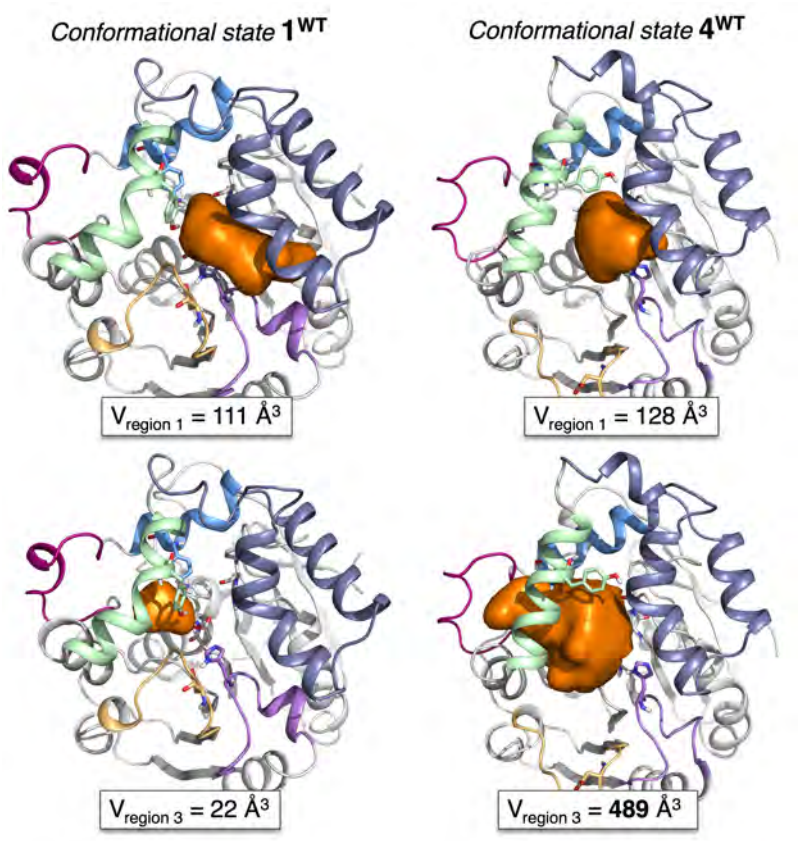


Figure 6.11: Representation of the 1^{WT} and 4^{WT} estimated active site pocket volumes in *region 1* and *3*.

The insertion of the smaller residue alanine at position Phe_{128} (i.e. the singly-mutated *Bm*EH variant **F128A**) generates more space in *region 3*, the product release region. As drawn from the analysis of the crystallographic structure (PDB ID: 4IO0), the diol product of **NGE** epoxide is found in *region 3*, thereby confirming its ability to hydrolyze bulky substrates. Nevertheless, in the first most populated conformational state of this variant (i.e. 1^{F128A}), the lid slightly deviates from this crystallographic structure, which in turn affects the conformation of the loop that contains His_{267} (found at 7.7 Å respect to Asp_{239}). In the equally populated conformational state 2^{F128A}, the alpha helix that contains Tyr_{144} is slightly disordered and Tyr_{144} occupies the space left by the **F128A** mutation. In addition, in this conformational state the residues that belong to the ion-charge pair relay (i.e. Asp_{239} and His_{267}) are not forming a hydrogen bond, instead these residues are at a distance of c.a. 7 Å due to the conformations adopted by their respective loops. The third major state identified for this variant (i.e. 3^{F128A}) is characterized for displaying more disordered regions in the lid domain and the alpha helix containing Tyr_{144}. 3^{F128A} is also characterized for displaying an open conformation of the loop containing Asp_{239}. Finally, in the fourth and the last state (i.e. 4^{F128A}), similar conformational changes compared to the previously discussed wild-type enzyme (4^{WT}) are observed. Although the alpha helix containing Tyr_{144} is more deviated respect to the crystallographic structure, Tyr_{144} side-chain is pointing outwards the active site, while the loops containing Asp_{239} and His_{267} are in an open conformation.

Although the conformations adopted by 4^{F128A} are not significantly differing from the ones previously observed on 4^{WT}, it is worth to mention that according to the active site volume calculations, smaller total volumes are observed (c.a. 188 and 615 Å^3, respectively). Nevertheless, as shown in Fig. 6.12, the conformational changes observed in 3^{F128A} lead to a significant increase on the total volumes (c.a. 631 Å^3). Noteworthy, the wider volume found in *region 1* of state 3^{F128A} indicates that, compared to the 4^{WT}, this conformational state might play a key role for accepting bulky substrates, while facilitating also the diol release after the reaction.

With the simulation time accumulated for the other *Bm*EH variant (i.e. **M145S**), five different conformational states are identified by partially reconstructing its corresponding FEL. In a general view, **M145S** conformations share similar structural features with **F128A** (e.g. partial unfolding of alpha helix containing Tyr_{144}), but also with the **WT** enzyme (e.g.

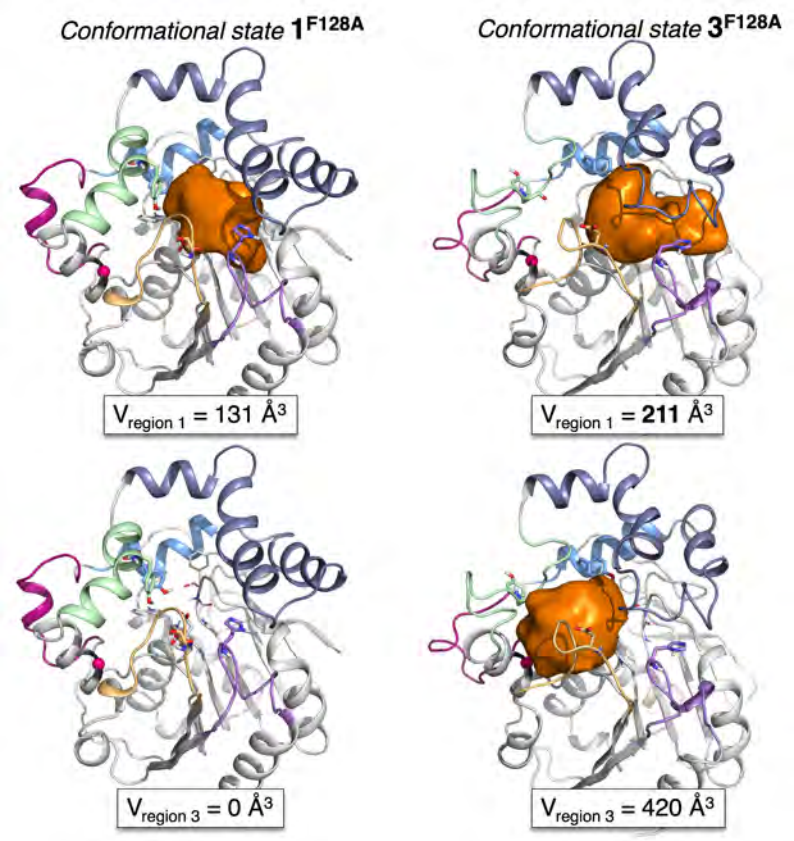


Figure 6.12: Representation of the 1^{F128A} and 4^{F128A} estimated active site pocket volumes in *region 1* and *3*.

the catalytic residues are properly oriented for catalysis). For instance, 1^{M145S} and 2^{M145S} conformational states resemble to the previously discussed states for the *Bm*EH wild-type. Likewise in **WT**, in 1^{M145S} the active site arrangement leads to a productive orientation of the catalytic residues, while the only deviations observed on 2^{M145S} are the Phe_{30} and Trp_{98} side-chains. In the third conformational state (i.e. 3^{M145S}), the insertion of **M145S** mutation induces a high disorder in the alpha helix containing Tyr_{144}, in addition to the deviations observed for the lid. Opposite to **F128A**, the conformational state 3^{M145S} has a smaller total volume. However, similar to **WT** and **F128A**, in the forth conformational state 4^{M145S} wider active site volumes are also observed. This is because the loops containing Asp_{239} and His_{267} are found in an open conformation, while the lid is slightly deviated, and the Tyr_{144} side-chain is pointing towards the solvent.

Chapter 7

Conclusions

In the present thesis, we have investigated the mechanism and selectivity of the *Bm*EH enzyme and variants toward different substrates and how the introduction of mutations affects the active site and its conformational dynamics toward the acceptance of bulkier substrates.

The main conclusions drawn from this thesis are hereunder presented:

First:
In Chapter 4, the truncated cluster model framework was applied to computationally explore the mechanism of the hydrolysis of *p*-**NSO** and **SO** substrates by *Bacillus megaterium* epoxide hydrolase at the DFT level. The information derived from the computed energy profiles for all plausible outcomes (i.e. attack at C_1 and C_2 carbons for each enantiomer and considering different substrate orientations in the active site), showed that the inherent *Bm*EH preference to bind and react with (*R*)-epoxide enantiomer is switched for *p*-**NSO** and **SO** substrates. Similar reactivity trends where found for both substrates. The two kinetically most favored paths correspond the attack at the benzylic position for both enantiomers but pointing towards different active site regions (i.e. region 1 for (*S*)-enantiomer and region 3 for the (*R*)). The differences on the activation energy barriers indicate that *Bm*EH selectivity

is determined during the first step of the reaction (i.e. the epoxide ring opening, alkylation step), while the rate-determining step corresponds to the second step (i.e. the hydrolysis of the covalently bound intermediate). Additional analysis of non-covalent interactions suggests that the presence of weak interactions like π-stacking interactions between the substrate and the residues of the active site pocket are key to dictate selectivity. Overall, the data derived in the present work shows how mechanistic insights derived from QM modeling serves as a strategy to rationalize the enantioselectivity and experimental trends with a keen eye on the future goal of prediction for further enzyme active site design aiming to expand the substrate scope.

Second:
In Chapter 5, extensive MD simulations combined with structural active site analysis were employed to explore the conformational changes induced by single-point *Bm*EH's mutations **F128A** and **M145S**. The comparison between the partially reconstructed free energy landscapes for each *Bm*EH variant revealed the conformational changes that lead to variants with wider active site pockets, thereby being able to accept bulkier substrates such as **NGE**. From our accumulated MD data, the alpha-helix and lid domain disorder of *Bm*EH was captured by the first slowest TICA component TIC_1, while the second slowest component TIC_2 was described by the conformational variation of the loop containing the catalytic residue Asp_{239}. Accordingly, for *Bm*EH wild-type four different conformational states were identified, four different major states for **F128A** *Bm*EH variant, and five for **M145S**. Based on these observations, we concluded that the ability of *Bm*EH variants to accept bulkier substrates arises not only for the space generated in the product release region, but specially from their ability to visit additional conformational states. Overall, this study on *Bm*EH conformational dynamics based on the partially reconstruction of free energy landscapes serves to identify key conformational states induced by mutations, which allow to rationalize their ability to accept and act upon bulkier substrates, and therefore to further guide the design of novel variants able to produce pure building blocks of pharmacologically-relevant compounds.

Bibliography

[1] M. Eigen, W. Gardiner, P. Schuster, and R. Winkler-Oswatitsch, "The origin of genetic information," *Scientific American*, vol. 244, no. 4, pp. 88–119, 1981.

[2] P. J. Butterworth, F. J. Warren, and P. R. Ellis, "Human alfa-amylase and starch digestion: An interesting marriage," *Starch - Stärke*, vol. 63, no. 7, pp. 395–405, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/star.201000150

[3] A. Fersht, *Enzyme Structure and Mechanism*. W.H. Freeman, 1985. [Online]. Available: https://books.google.es/books?id=C1OgQgAACAAJ

[4] A. Illanes, *Enzyme Biocatalysis*. Springer Netherlands, 2008. [Online]. Available: http://dx.doi.org/10.1007/978-1-4020-8361-7

[5] G. Cooper, *The Cell: A Molecular Approach*, 2nd ed. Sinauer Associates Inc, 2000.

[6] S. Boyce and K. F. Tipton, *Enzyme Classification and Nomenclature*. American Cancer Society, 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0000710

[7] S. Jemli, D. Ayadi-Zouari, H. B. Hlima, and S. Bejar, "Biocatalysts: application and engineering for industrial purposes," *Critical Reviews in Biotechnology*, vol. 36, no. 2, pp. 246–258, Nov 2014. [Online]. Available: http://dx.doi.org/10.3109/07388551.2014.950550

[8] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins, "Engineering the third wave of biocatalysis," *Nature*, vol. 485, no. 7397, pp. 185–194, May 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11117

[9] S. C. Hammer, A. M. Knight, and F. H. Arnold, "Design and evolution of enzymes for non-natural chemistry," *Current Opinion in Green and Sustainable Chemistry*, vol. 7, pp. 23–30, 2017, new Synthetic Methods 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452223617300354

[10] A. Barrozo, R. Borstnar, G. Marloie, and S. C. L. Kamerlin, "Computational protein engineering: Bridging the gap between rational design and laboratory evolution," *International Journal of Molecular Sciences*, vol. 13, no. 12, pp. 12 428–12 460, Sep 2012. [Online]. Available: http://dx.doi.org/10.3390/ijms131012428

[11] J. M. W. Stephanie G Burton, Don A Cowan, "The search for the ideal biocatalyst," *Nat Biotechnol*, vol. 20, pp. 37–45, 2002.

[12] S. Osuna, "The challenge of predicting distal active site mutations in computational enzyme design," *WIREs Computational Molecular Science*, vol. 11, no. 3, p. e1502, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1502

[13] A. Romero-Rivera, M. Garcia-Borràs, and S. Osuna, "Computational tools for the evaluation of laboratory-engineered biocatalysts," *Chem. Commun.*, vol. 53, pp. 284–297, 2017.

[14] D. T. Dryden, A. R. Thomson, and J. H. White, "How much of protein sequence space has been explored by life on earth?" *Journal of The Royal Society Interface*, vol. 5, no. 25, pp. 953–956, Apr 2008. [Online]. Available: http://dx.doi.org/10.1098/rsif.2008.0085

[15] U. T. Bornscheuer, "The fourth wave of biocatalysis is approaching," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2110, p. 20170063, Nov 2017. [Online]. Available: http://dx.doi.org/10.1098/rsta.2017.0063

[16] K. R. Jegannathan and P. H. Nielsen, "Environmental assessment of enzyme use in industrial production: a literature review," *Journal of Cleaner Production*, vol. 42, pp. 228–240, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095965261200594X

[17] N. Gurung, S. Ray, S. Bose, and V. Rai, "A broader view: Microbial enzymes and their relevance in industries, medicine, and beyond," *BioMed Research International*, vol. 2013, pp. 1–18, 2013. [Online]. Available: http://dx.doi.org/10.1155/2013/329121

[18] N. J. Turner and L. Humphreys, *Biocatalysis in Organic Synthesis*. The Royal Society of Chemistry, 2018.

[19] A. Liese, K. Seelbach, and C. Wandrey, *Industrial biotransformations*. John Wiley & Sons, 2006.

[20] S. W. Smith, "Chiral toxicology: It's the same thing... only different," *Toxicological Sciences*, vol. 110, no. 1, pp. 4–30, 05 2009. [Online]. Available: https://doi.org/10.1093/toxsci/kfp097

[21] J. Clayden, N. Greeves, S. Warren, and P. Wothers, *Organic Chemistry*. Oxford university press Oxford, 20001.

[22] J. McMurry, *Organic Chemistry*. Cengage Learning, 2015. [Online]. Available: https://books.google.es/books?id=3BKdBQAAQBAJ

[23] J. Eames and P. J. M., *Stereochemistry at a Glance*. Wiley, 2003.

[24] J. H. Kim and A. R. Scialli, "Thalidomide: The tragedy of birth defects and the effective treatment of disease," *Toxicological Sciences*, vol. 122, no. 1, pp. 1–6, 04 2011. [Online]. Available: https://doi.org/10.1093/toxsci/kfr088

[25] N. R. Council, *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals*. Washington, DC: The National Academies Press, Jun 2015. [Online]. Available: http://dx.doi.org/10.17226/19001

[26] I. Agranat, H. Caner, and J. Caldwell, "Putting chirality to work: the strategy of chiral switches," *Nature Reviews Drug Discovery*, vol. 1, no. 10, pp. 753–768, Oct 2002. [Online]. Available: http://dx.doi.org/10.1038/nrd915

[27] E. Tokunaga, T. Yamamoto, E. Ito, and N. Shibata, "Understanding the thalidomide chirality in biological processes by the self-disproportionation of enantiomers," *Scientific Reports*, vol. 8, no. 1, Nov 2018. [Online]. Available: http://dx.doi.org/10.1038/s41598-018-35457-6

[28] T. Eriksson, S. Björkman, and P. Höglund, "Clinical pharmacology of thalidomide," *European Journal of Clinical Pharmacology*, vol. 57, no. 5, pp. 365–376, Jul 2001. [Online]. Available: http://dx.doi.org/10.1007/s002280100320

[29] A. Li, J. Liu, S. Q. Pham, and Z. Li, "Engineered p450pyr monooxygenase for asymmetric epoxidation of alkenes with unique and high enantioselectivity," *Chem. Commun.*, vol. 49, pp. 11 572–11 574, 2013. [Online]. Available: http://dx.doi.org/10.1039/C3CC46675B

[30] R. Kadyrov, R. Koenigs, C. Brinkmann, D. Voigtlaender, and M. Rueping, "Efficient enantioselective synthesis of optically active diols by asymmetric hydrogenation with modular chiral metal catalysts," *Angewandte Chemie International Edition*, vol. 48, no. 41, pp. 7556–7559, Sep 2009. [Online]. Available: http://dx.doi.org/10.1002/anie.200902835

[31] Y. Nie, Y. Xu, M. Yang, and X.-Q. Mu, "A novel nadh-dependent carbonyl reductase with unusual stereoselectivity for (r)-specific reduction from an (s)-1-phenyl-1,2-ethanediol-producing micro-organism: purification and characterization," *Letters in Applied Microbiology*, vol. 44, no. 5, pp. 555–562, May 2007. [Online]. Available: http://dx.doi.org/10.1111/j.1472-765X.2006.02100.x

[32] A. R. Gomes, C. L. Varela, E. J. T. da Silva, and F. M. Roleira, "Epoxide containing molecules: A good or a bad drug design approach," *European Journal of Medicinal Chemistry*, vol. 201, p. 112327, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0223523420302968

[33] A. Archelas and R. Furstoss, "Synthetic applications of epoxide hydrolases," *Current Opinion in Chemical Biology*, vol. 5, no. 2, pp. 112–119, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367593100001794

[34] N. Bala and S. S. Chimni, "Recent developments in the asymmetric hydrolytic ring opening of epoxides catalysed by microbial epoxide hydrolase," *Tetrahedron: Asymmetry*, vol. 21, no. 24, pp. 2879–2898, 2010.

[35] M. Kotik, A. Archelas, and R. Wohlgemuth, "Epoxide hydrolases and their application in organic synthesis," *Current Organic Chemistry*, vol. 16, no. 4, pp. 451–482, Feb 2012. [Online]. Available: http://dx.doi.org/10.2174/138527212799499840

[36] R. E. Parker and N. S. Isaac, "Mechanisms of epoxide reactions," *Chem. Rev.*, no. 264, pp. 9310–13, 1959.

[37] C. A. Weijers and J. A. de Bont, "Epoxide hydrolases from yeasts and other sources: versatile tools in biocatalysis," *Journal of Molecular Catalysis B: Enzymatic*, vol. 6, no. 3, pp. 199–214, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1381117798001234

[38] F. Oesch, "Mammalian epoxide hydrases: Inducible enzymes catalysing the inactivation of carcinogenic and cytotoxic metabolites derived from aromatic and olefinic compounds," *Xenobiotica*, vol. 3, no. 5, pp. 305–340, Jan 1973. [Online]. Available: http://dx.doi.org/10.3109/00498257309151525

[39] A. Marowsky, A. Cronin, F. Frère, M. Adamska, and M. Arand, "Mammalian epoxide hydrolases," in *Comprehensive Toxicology (Second Edition)*, second edition ed., C. A. McQueen, Ed. Oxford: Elsevier, 2010, pp. 275–294. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080468846004152

[40] J. Gautheron and I. Jéru, "The multifaceted role of epoxide hydrolases in human health and disease," *International Journal of Molecular Sciences*, vol. 22, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1422-0067/22/1/13

[41] G. Brooks, A. Harrison, and S. Lewis, "Cyclodiene epoxide ring hydration by microsomes from mammalian liver and houseflies," *Biochemical Pharmacology*, vol. 19, no. 1, pp. 255–273, 1970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0006295270903461

[42] C. Morisseau and B. D. Hammock, "Gerry brooks and epoxide hydrolases: four decades to a pharmaceutical," *Pest Management Science*, vol. 64, no. 6, pp. 594–609, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ps.1583

[43] Y.-F. Tang, J.-H. Xu, Q. Ye, and B. Schulze, "Biocatalytic preparation of (s)-phenyl glycidyl ether using newly isolated bacillus megaterium ecu1001," *Journal of Molecular Catalysis B: Enzymatic*, vol. 13, no. 4, pp. 61 – 68, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1381117700002307

[44] B. van Loo, J. Kingma, M. Arand, M. G. Wubbolts, and D. B. Janssen, "Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis," *Applied and Environmental Microbiology*, vol. 72, no. 4, pp. 2905–2917, 2006.

[45] M. Widersten, A. Gurell, and D. Lindberg, "Structure-function relationships of epoxide hydrolases and their potential use in biocatalysis," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1800, no. 3, pp. 316–326, 2010, includes Special Section: Mitochondrial Research and Medicine. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030441650900333X

[46] H. Lin, J.-Y. Liu, H.-B. Wang, A. A. Q. Ahmed, and Z.-L. Wu, "Biocatalysis as an alternative for the production of chiral epoxides: A comparative review," *Journal of Molecular Catalysis B: Enzymatic*, vol. 72, no. 3, pp. 77–89, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1381117711001925

[47] R. Rink, J. H. Lutje Spelberg, R. J. Pieters, J. Kingma, M. Nardini, R. M. Kellogg, B. W. Dijkstra, and D. B. Janssen, "Mutation of tyrosine residues involved in the alkylation half reaction of epoxide hydrolase from agrobacterium radiobacter ad1 results in improved enantioselectivity," *Journal of the American Chemical Society*, vol. 121, no. 32, 1999.

[48] M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand, A. Cronin, Y. Genzel, A. Archelas, and R. Furstoss, "Enhancing the enantioselectivity of an epoxide hydrolase by directed evolution," *Organic Letters*, vol. 6, no. 2, pp. 177–180, Dec 2003. [Online]. Available: http://dx.doi.org/10.1021/ol035898m

[49] B. van Loo, J. H. Spelberg, J. Kingma, T. Sonke, M. G. Wubbolts, and D. B. Janssen, "Directed evolution of epoxide hydrolase from a. radiobacter toward higher enantioselectivity by error-prone pcr and dna shuffling," *Chemistry and Biology*, vol. 11, no. 7, pp. 981–990, Jul 2004. [Online]. Available: http://dx.doi.org/10.1016/j.chembiol.2004.04.019

[50] P. Saini and D. Sareen, "An overview on the enhancement of enantioselectivity and stability of microbial epoxide hydrolases," *Molecular Biotechnology*, vol. 59, no. 2-3, pp. 98–116, Mar 2017. [Online]. Available: http://dx.doi.org/10.1007/s12033-017-9996-8

[51] C. Morisseau and B. D. Hammock, "Epoxide hydrolases: Mechanisms, inhibitor designs, and biological roles," *Annual Review of Pharmacology and Toxicology*, vol. 45, no. 1, pp. 311–333, 2005, pMID: 15822179. [Online]. Available: https://doi.org/10.1146/annurev.pharmtox.45.120403.095920

[52] A. Archelas, G. Iacazio, and M. Kotik, *Epoxide Hydrolases and their Application in Organic Synthesis*. John Wiley Sons, Ltd, 2016, ch. 8, pp. 179–229. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118828083.ch8

[53] P. D. Carr and D. L. Ollis, "Alpha/beta hydrolase fold: an update," *Protein and peptide letters*, vol. 16, no. 10, pp. 1137–1148, 2009.

[54] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. Verschueren, and A. Goldman, "The alpha/beta hydrolase fold," *Protein Engineering, Design and Selection*, vol. 5, no. 3, pp. 197–211, 04 1992. [Online]. Available: https://doi.org/10.1093/protein/5.3.197

[55] M. Nardini and B. W. Dijkstra, "alpha/beta-hydrolase fold enzymes: the family keeps growing," *Current Opinion in Structural Biology*, vol. 9, no. 6, pp. 732–737, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959440X99000378

[56] D. M. Nedrud, H. Lin, G. Lopez, S. K. Padhi, G. A. Legatt, and R. J. Kazlauskas, "Uncovering divergent evolution of alpha/beta-hydrolases: a surprising residue substitution needed to convert hevea brasiliensis hydroxynitrile lyase into an esterase," *Chem. Sci.*, vol. 5, no. 11, pp. 4265–4277, 2014. [Online]. Available: http://dx.doi.org/10.1039/C4SC01544D

[57] H. Jochens, M. Hesseler, K. Stiba, S. K. Padhi, R. J. Kazlauskas, and U. T. Bornscheuer, "Protein engineering of alpha/beta-hydrolase fold enzymes," *ChemBioChem*, vol. 12, no. 10, pp. 1508–1517, 2011. [Online]. Available: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cbic.201000771

[58] J. Zhao, Y.-Y. Chu, A.-T. Li, X. Ju, X.-D. Kong, J. Pan, Y. Tang, and J.-H. Xu, "An unusual (r)-selective epoxide hydrolase with high activity for facile preparation of enantiopure glycidyl ethers," *Advanced Synthesis &*

Catalysis, vol. 353, no. 9, pp. 1510–1518, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adsc.201100031

[59] X.-D. Kong, Q. Ma, J. Zhou, B.-B. Zeng, and J.-H. Xu, "A smart library of epoxide hydrolase variants and the top hits for synthesis of (s)-beta-blocker precursors," *Angewandte Chemie International Edition*, vol. 53, no. 26, pp. 6641–6644, May 2014. [Online]. Available: http://dx.doi.org/10.1002/anie.201402653

[60] A. A. Al-Majed, A. H. Bakheit, H. A. Abdel Aziz, F. M. Alajmi, and H. AlRabiah, "Propranolol," in *Profiles of Drug Substances, Excipients and Related Methodology*, H. G. Brittain, Ed. Academic Press, 2017, vol. 42, pp. 287–338. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1871512517300067

[61] T. L. Blundell, "The first resolution revolution in protein structure analysis: X-ray diffraction of polypeptide conformations and globular protein folds in 1950s and 1960s," *Progress in Biophysics and Molecular Biology*, vol. 167, pp. 32–40, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S007961072100105X

[62] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 01 2000. [Online]. Available: https://doi.org/10.1093/nar/28.1.235

[63] M. C. Thompson, T. O. Yeates, and J. A. Rodriguez, "Advances in methods for atomic resolution macromolecular structure determination." *F1000Res*, vol. 9, 2020.

[64] X.-c. Bai, G. McMullan, and S. H. Scheres, "How cryo-em is revolutionizing structural biology," *Trends in Biochemical Sciences*, vol. 40, no. 1, pp. 49–57, Jan 2015. [Online]. Available: http://dx.doi.org/10.1016/j.tibs.2014.10.005

[65] M. Jaskolski, Z. Dauter, and A. Wlodawer, "A brief history of macromolecular crystallography, illustrated by a family tree and its nobel fruits," *The FEBS Journal*, vol. 281, no. 18, pp. 3985–4009, 2014. [Online]. Available: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.12796

[66] K. Wüthrich, "The way to nmr structures of proteins," *Nature Structural*

Biology, vol. 8, no. 11, pp. 923–925, Nov 2001. [Online]. Available: http://dx.doi.org/10.1038/nsb1101-923

[67] A. L. Mitchell, A. Almeida, M. Beracochea, M. Boland, J. Burgin, G. Cochrane, M. R. Crusoe, V. Kale, S. C. Potter, L. J. Richardson, E. Sakharova, M. Scheremetjew, A. Korobeynikov, A. Shlemov, O. Kunyavskaya, A. Lapidus, and R. D. Finn, "MGnify: the microbiome analysis resource in 2020," *Nucleic Acids Research*, vol. 48, no. D1, pp. D570–D578, 11 2019. [Online]. Available: https://doi.org/10.1093/nar/gkz1035

[68] M. Steinegger, M. Mirdita, and J. Söding, "Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold," *Nature Methods*, vol. 16, no. 7, pp. 603–606, Jun 2019. [Online]. Available: http://dx.doi.org/10.1038/s41592-019-0437-4

[69] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, "The protein folding problem," *Annual Review of Biophysics*, vol. 37, no. 1, pp. 289–316, Jun 2008. [Online]. Available: http://dx.doi.org/10.1146/annurev.biophys.37.092707.153558

[70] R. P. Bywater, "Why twenty amino acid residue types suffice(d) to support all living systems," *PLOS ONE*, vol. 13, no. 10, p. e0204883, Oct 2018. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0204883

[71] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, and D. Hassabis, "Highly accurate protein structure prediction for the human proteome," *Nature*, vol. 596, no. 7873, pp. 590–596, Jul 2021. [Online]. Available: http://dx.doi.org/10.1038/s41586-021-03828-1

[72] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. New York: Garland Science, 2002.

[73] F. Sanger, E. O. P. Thompson, and R. Kitai, "The amide groups of insulin," *Biochemical Journal*, vol. 59, no. 3, pp. 509–518, 03 1955. [Online]. Available: https://doi.org/10.1042/bj0590509

[74] A. Kessel and N. Ben-Tal, *Introduction to Proteins: Structure, Function, and Motion.* CRC PRess, 2010.

[75] J. S. Richardson, *The Anatomy and Taxonomy of Protein Structure*, ser. Advances in Protein Chemistry, C. Anfinsen, J. T. Edsall, and F. M. Richards, Eds. Academic Press, 1981, vol. 34. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0065323308605203

[76] D. G. Udatha, K. M. Madsen, G. Panagiotou, and L. Olsson, "Multiple nucleophilic elbows leading to multiple active sites in a single module esterase from sorangium cellulosum," *Journal of Structural Biology*, vol. 190, no. 3, pp. 314–327, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047847715000817

[77] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211

[78] E. Papaleo, G. Saladino, M. Lambrughi, K. Lindorff-Larsen, F. L. Gervasio, and R. Nussinov, "The role of protein loops and linkers in conformational dynamics and allostery," *Chemical Reviews*, vol. 116, no. 11, pp. 6391–6423, Feb 2016. [Online]. Available: http://dx.doi.org/10.1021/acs.chemrev.5b00623

[79] B. M. Nestl and B. Hauer, "Engineering of flexible loops in enzymes," *ACS Catalysis*, vol. 4, no. 9, pp. 3201–3211, Aug 2014. [Online]. Available: http://dx.doi.org/10.1021/cs500325p

[80] M. M. Malabanan, T. L. Amyes, and J. P. Richard, "A role for flexible loops in enzyme catalysis," *Current Opinion in Structural Biology*, vol. 20, no. 6, pp. 702 – 710, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959440X10001405

[81] P. Privalov, "Stability of proteins small globular proteins," ser. Advances in Protein Chemistry, C. Anfinsen, J. T. Edsall, and F. M. Richards, Eds. Academic Press, 1979, vol. 33, pp. 167–241. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S006532330860460X

[82] R. Reis and I. Moraes, "Structural biology and structure - function relationships of membrane proteins," *Biochemical Society Transactions*, vol. 47, no. 1, pp. 47–61, 12 2018. [Online]. Available: https://doi.org/10.1042/BST20180269

[83] P. E. Wright and H. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal of Molecular Biology*, vol. 293, no. 2, pp. 321–331, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283699931108

[84] J. P. Collman, N. K. Devaraj, R. A. Decréau, Y. Yang, Y.-L. Yan, W. Ebina, T. A. Eberspacher, and C. E. D. Chidsey, "A cytochrome c oxidase model catalyzes oxygen to water reduction under rate-limiting electron flux," *Science*, vol. 315, no. 5818, pp. 1565–1568, 2007. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1135844

[85] K. U. Walter, K. Vamvaca, and D. Hilvert, "An active enzyme constructed from a 9-amino acid alphabet," *Journal of Biological Chemistry*, vol. 280, no. 45, pp. 37 742–37 746, Nov 2005. [Online]. Available: http://dx.doi.org/10.1074/jbc.M507210200

[86] D. A. C. Beck, D. O. V. Alonso, D. Inoyama, and V. Daggett, "The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins," *PNAS; Proceedings of the National Academy of Sciences*, vol. 105, no. 34, pp. 12 259–12 264, 2008.

[87] K. A. Dill, "Dominant forces in protein folding," *Biochemistry*, vol. 29, no. 31, pp. 7133–7155, Aug 1990. [Online]. Available: http://dx.doi.org/10.1021/bi00483a001

[88] G. D. Rose, "Prediction of chain turns in globular proteins on a hydrophobic basis," *Nature*, vol. 272, no. 5654, pp. 586–590, Apr 1978. [Online]. Available: http://dx.doi.org/10.1038/272586a0

[89] D. Bordo and P. Argos, "The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins," *Journal of Molecular Biology*, vol. 243, no. 3, pp. 504–519, 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283684716767

[90] A. R. Fersht and L. Serrano, "Principles of protein stability derived from protein engineering experiments," *Current Opinion in Structural Biology*, vol. 3, no. 1, pp. 75–83, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0959440X9390205Y

[91] E. Baker and R. Hubbard, "Hydrogen bonding in globular proteins," *Progress in Biophysics and Molecular Biology*, vol. 44, no. 2, pp. 97–179, 1984. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0079610784900075

[92] W. Saenger and G. Jeffrey, *Hydrogen bonding in biological structures.* Springer-Verlag, 1994.

[93] I. K. McDonald and J. M. Thornton, "Satisfying hydrogen bonding potential in proteins," *Journal of Molecular Biology*, vol. 238, no. 5, pp. 777–793, 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283684713349

[94] H.-X. Zhou and X. Pang, "Electrostatic interactions in protein structure, folding, binding, and condensation," *Chemical Reviews*, vol. 118, no. 4, pp. 1691–1741, 2018.

[95] P. Chien and L. M. Gierasch, "Challenges and dreams: physics of weak interactions essential to life," *Molecular Biology of the Cell*, vol. 25, no. 22, pp. 3474–3477, Nov 2014. [Online]. Available: http://dx.doi.org/10.1091/mbc.e14-06-1035

[96] M. D. W. Griffin and J. A. Gerrard, "The relationship between oligomeric state and protein function," *Protein Dimerization and Oligomerization in Biology*, pp. 74–90, 2012. [Online]. Available: http://dx.doi.org/10.1007/978-1-4614-3229-6_5

[97] A. Shurki, M. Štrajbl, J. Villà, and A. Warshel, "How much do enzymes really gain by restraining their reacting fragments?" *Journal of the American Chemical Society*, vol. 124, no. 15, pp. 4097–4107, 2002.

[98] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. M. Olsson, "Electrostatic basis for enzyme catalysis," *Chemical Reviews*, vol. 106, no. 8, pp. 3210–3235, Aug 2006. [Online]. Available: http://dx.doi.org/10.1021/cr0503106

[99] W. R. Loewenstein, "On the specificity of a sensory receptor," *Journal of Neurophysiology*, vol. 24, no. 2, pp. 150–158, Mar 1961. [Online]. Available: http://dx.doi.org/10.1152/jn.1961.24.2.150

[100] A. Ullrich, A. Gray, A. Tam, T. Yang-Feng, M. Tsubokawa, C. Collins, W. Henzel, T. Le Bon, S. Kathuria, and E. Chen, "Insulin-like growth factor i receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity." *The EMBO Journal*, vol. 5, no. 10, pp. 2503–2512, 1986. [Online]. Available: https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1986.tb04528.x

[101] M. Rask-Andersen, S. Masuram, and H. B. Schiöth, "The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication," *Annual Review of Pharmacology and Toxicology*, vol. 54, no. 1, pp. 9–26, Jan 2014. [Online]. Available: http://dx.doi.org/10.1146/annurev-pharmtox-011613-135943

[102] W. I. Weis and B. K. Kobilka, "The molecular basis of g protein–coupled receptor activation," *Annual Review of Biochemistry*, vol. 87, no. 1, pp. 897–919, 2018.

[103] S. Trowitzsch and R. Tampé, "Abc transporters in dynamic macromolecular assemblies," *Journal of molecular biology*, vol. 430, no. 22, pp. 4481–4495, 2018.

[104] A. S. F. Oliveira, C. J. Edsall, C. J. Woods, P. Bates, G. V. Nunez, S. Wonnacott, I. Bermudez, G. Ciccotti, T. Gallagher, R. B. Sessions, and A. J. Mulholland, "A general mechanism for signal propagation in the nicotinic acetylcholine receptor family," *Journal of the American Chemical Society*, vol. 141, no. 51, pp. 19 953–19 958, 2019.

[105] E. Fischer, "Einfluss der configuration auf die wirkung der enzyme," *Berichte der deutschen chemischen Gesellschaft*, vol. 27, no. 3, pp. 2985–2993, 1894. [Online]. Available: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cber.18940270364

[106] L. PAULING, "Molecular architecture and biological reactions," *Chemical and Engineering News Archive*, vol. 24, no. 10, pp. 1375–1377, May 1946. [Online]. Available: http://dx.doi.org/10.1021/cen-v024n010.p1375

[107] J. R. Schnell, H. J. Dyson, and P. E. Wright, "Structure, dynamics, and catalytic function of dihydrofolate reductase," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, no. 1, pp. 119–140, 2004, pMID: 15139807. [Online]. Available: https://doi.org/10.1146/annurev.biophys.33.110502.133613

[108] Y. Savir and T. Tlusty, "Conformational proofreading: The impact of conformational changes on the specificity of molecular recognition," *PLoS ONE*, vol. 2, no. 5, p. e468, May 2007. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0000468

[109] C. A. Sotriffer, O. Krämer, and G. Klebe, "Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations," *Proteins: Structure, Function, and Bioinformatics*, vol. 56, no. 1, pp. 52–66, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.20021

[110] D. E. Koshland JR., "Enzyme flexibility and enzyme action," *Journal of Cellular and Comparative Physiology*, vol. 54, no. S1, pp. 245–258, 1959. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcp.1030540420

[111] M. Gerstein and N. Echols, "Exploring the range of protein flexibility, from a structural proteomics perspective," *Current Opinion in Chemical Biology*, vol. 8, no. 1, pp. 14–19, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367593103001753

[112] J. Monod, J. Wyman, and J.-P. Changeux, "On the nature of allosteric transitions: A plausible model," *Journal of Molecular Biology*, vol. 12, no. 1, pp. 88–118, 1965. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283665802856

[113] L. Fetler, E. R. Kantrowitz, and P. Vachette, "Direct observation in solution of a preexisting structural equilibrium for a mutant of the allosteric aspartate transcarbamoylase," *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 495–500, 2007. [Online]. Available: https://www.pnas.org/content/104/2/495

[114] C.-J. Tsai and R. Nussinov, "A unified view of "how allostery works"," *PLoS Computational Biology*, vol. 10, no. 2, p. e1003394, Feb 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1003394

[115] K. Gunasekaran, B. Ma, and R. Nussinov, "Is allostery an intrinsic property of all dynamic proteins?" *Proteins: Structure, Function, and Bioinformatics*, vol. 57, no. 3, pp. 433–443, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.20232

[116] B. Ramakrishnan, P. Balaji, and P. K. Qasba, "Crystal structure of 1,4 galactosyltransferase complex with udp gal reveals an oligosaccharide acceptor binding site," *Journal of Molecular Biology*, vol. 318, no. 2, pp. 491–502, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283602000207

[117] A. Kitao, S. Hayward, and N. Go, "Energy landscape of a native protein: Jumping-among-minima model," *Proteins: Structure, Function, and Bioinformatics*, vol. 33, no. 4, pp. 496–517, 1998.

[118] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, pp. 585–590, 1977.

[119] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, "Dynamics of ligand binding to myoglobin," *Biochemistry*, vol. 14, no. 24, pp. 5355–5375, 1975.

[120] G. A. Petsko and D. Ringe, "Fluctuations in protein structure from x-ray diffraction," *Annual Review of Biophysics and Bioengineering*, vol. 13, no. 1, pp. 331–371, 1984, pMID: 6331286. [Online]. Available: https://doi.org/10.1146/annurev.bb.13.060184.001555

[121] H. Frauenfelder, F. Parak, and R. D. Young, "Conformational substates in proteins," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 17, no. 1, pp. 451–479, 1988, pMID: 3293595. [Online]. Available: https://doi.org/10.1146/annurev.bb.17.060188.002315

[122] G. Weber, "Ligand binding and internal equilibiums in proteins," *Biochemistry*, vol. 11, no. 5, pp. 864–878, Feb 1972. [Online]. Available: http://dx.doi.org/10.1021/bi00755a028

[123] H. R. Bosshard, "Molecular recognition by induced fit: How fit is the concept?" *Physiology*, vol. 16, no. 4, pp. 171–173, 2001. [Online]. Available: https://doi.org/10.1152/physiologyonline.2001.16.4.171

[124] S. Kumar, B. Ma, C.-J. Tsai, N. Sinha, and R. Nussinov, "Folding and binding cascades: Dynamic landscapes and population shifts," *Protein Science*, vol. 9, no. 1, pp. 10–19, 2000. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.9.1.10

[125] D. D. Boehr and P. E. Wright, "How do proteins interact?" *Science*, vol. 320, no. 5882, pp. 1429–1430, 2008. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1158818

[126] L. C. James, P. Roversi, and D. S. Tawfik, "Antibody multispecificity mediated by conformational diversity," *Science*, vol. 299, no. 5611, pp. 1362–1367, 2003. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1079731

[127] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo, "The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters," *Biochemistry*, vol. 50, no. 21, pp. 4402–4410, May 2011. [Online]. Available: http://dx.doi.org/10.1021/bi2002289

[128] M. Yang and J. M. Quayle, "Chapter 11 - assessing the real-time metabolism and bioenergetics of single cells using fluorescence biosensors," in *Clinical Bioenergetics*, S. Ostojic, Ed. Academic Press, 2021, pp. 269–284. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128196212000115

[129] S. Arrhenius, "Über die dissociationswärme und den einfluss der temperatur auf den dissociationsgrad der elektrolyte," *Zeitschrift für Physikalische Chemie*, vol. 4U, no. 1, pp. 96–116, Jul 1889. [Online]. Available: http://dx.doi.org/10.1515/zpch-1889-0408

[130] J. B. Chaires, "Calorimetry and thermodynamics in drug design," *Annual Review of Biophysics*, vol. 37, no. 1, pp. 135–151, Jun 2008. [Online]. Available: http://dx.doi.org/10.1146/annurev.biophys.36.040306.132812

[131] K. A. Dill, S. Bromberg, and D. Stigter, "Molecular driving forces," Oct 2010. [Online]. Available: http://dx.doi.org/10.4324/9780203809075

[132] A. Cooper, "Thermodynamics of protein folding and stability," *Protein: A comprehensive treatise*, vol. 2, pp. 217–270, 1999.

[133] R. Wolfenden and M. J. Snider, "The depth of chemical time and the power of enzymes as catalysts," *Accounts of Chemical Research*, vol. 34, no. 12, pp. 938–945, Oct 2001. [Online]. Available: http://dx.doi.org/10.1021/ar000058i

[134] R. Wolfenden, "Thermodynamic and extrathermodynamic requirements of enzyme catalysis," *Biophysical Chemistry*, vol. 105, no. 2, pp. 559–572, 2003, walter

Kauzmann's 85th Birthday. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301462203000668

[135] J. R. Knowles, "Enzyme catalysis: not different, just better," *Nature*, vol. 350, no. 6314, pp. 121–124, Mar 1991. [Online]. Available: http://dx.doi.org/10.1038/350121a0

[136] H. W. Wiley, "Lois générales de l'action des diastases." *Journal of the American Chemical Society*, vol. 25, no. 7, pp. 780–782, Jul 1903. [Online]. Available: http://dx.doi.org/10.1021/ja02009a024

[137] A. Cornish-Bowden, J.-P. Mazat, and S. Nicolas, "Victor henri: 111 years of his equation," *Biochimie*, vol. 107, pp. 161–166, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0300908414002673

[138] M. L. Michaelis, Leonor; Menten, "The kinetics of the inversion effect," *Biochemistry*, 1913.

[139] K. A. Johnson and R. S. Goody, "The original michaelis constant: Translation of the 1913 michaelis–menten paper," *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, Oct 2011. [Online]. Available: http://dx.doi.org/10.1021/bi201284u

[140] G. E. Briggs and J. B. S. Haldane, "A note on the kinetics of enzyme action," *Biochemical Journal*, vol. 19, no. 2, pp. 338–339, Jan 1925. [Online]. Available: http://dx.doi.org/10.1042/bj0190338

[141] H. Eyring, "The activated complex in chemical reactions," *The Journal of Chemical Physics*, vol. 3, no. 2, pp. 107–115, 1935. [Online]. Available: https://doi.org/10.1063/1.1749604

[142] G. E. Lienhard, "Enzymatic catalysis and transition-state theory," *Science*, vol. 180, no. 4082, pp. 149–154, 1973. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.180.4082.149

[143] C. S. Chen, Y. Fujimoto, G. Girdaukas, and C. J. Sih, "Quantitative analyses of biochemical kinetic resolutions of enantiomers," *Journal of the American Chemical Society*, vol. 104, no. 25, pp. 7294–7299, Dec 1982. [Online]. Available: http://dx.doi.org/10.1021/ja00389a064

[144] X. Sheng, M. Kazemi, F. Planas, and F. Himo, "Modeling enzymatic enantioselectivity using quantum chemical methodology," *ACS Catalysis*, vol. 10, no. 11, pp. 6430–6449, May 2020. [Online]. Available: http://dx.doi.org/10.1021/acscatal.0c00983

[145] R. L. Schowen, "Catalytic power and transition-state stabilization," *Transition States of Biochemical Processes*, pp. 77–114, 1978. [Online]. Available: http://dx.doi.org/10.1007/978-1-4684-9978-0_2

[146] T. C. Bruice and F. C. Lightstone, "Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions," *Accounts of Chemical Research*, vol. 32, no. 2, pp. 127–136, Nov 1998. [Online]. Available: http://dx.doi.org/10.1021/ar960131y

[147] R. M. Bell and D. E. Koshland, "Covalent enzyme-substrate intermediates," *Science*, vol. 172, no. 3989, pp. 1253–1256, 1971. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.172.3989.1253

[148] J. J. Perona and C. S. Craik, "Structural basis of substrate specificity in the serine proteases," *Protein Science*, vol. 4, no. 3, pp. 337–360, 1995. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5560040301

[149] H.-M. Chu, T.-P. Ko, and A. H.-J. Wang, "Crystal structure and substrate specificity of plant adenylate isopentenyltransferase from Humulus lupulus: distinctive binding affinity for purine and pyrimidine nucleotides," *Nucleic Acids Research*, vol. 38, no. 5, pp. 1738–1748, 12 2009. [Online]. Available: https://doi.org/10.1093/nar/gkp1093

[150] G. Huyer, J. Kelly, J. Moffat, R. Zamboni, Z. Jia, M. J. Gresser, and C. Ramachandran, "Affinity selection from peptide libraries to determine substrate specificity of protein tyrosine phosphatases," *Analytical Biochemistry*, vol. 258, no. 1, pp. 19–30, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003269797925410

[151] D. Su, T. Kosciuk, M. Yang, I. R. Price, and H. Lin, "Binding affinity determines substrate specificity and enables discovery of substrates for n-myristoyltransferases," *ACS Catalysis*, vol. 11, no. 24, pp. 14 877–14 883, Nov 2021. [Online]. Available: http://dx.doi.org/10.1021/acscatal.1c03330

[152] S. J. Benkovic and S. Hammes-Schiffer, "A perspective on enzyme catalysis," *Science*, vol. 301, no. 5637, pp. 1196–1202, 2003. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1085515

[153] G. J. Bartlett, C. T. Porter, N. Borkakoti, and J. M. Thornton, "Analysis of catalytic residues in enzyme active sites," *Journal of Molecular Biology*, vol. 324, no. 1, pp. 105–121, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022283602010367

[154] M. Garcia-Viloca, J. Gao, M. Karplus, and D. G. Truhlar, "How enzymes work: Analysis by modern rate theory and computer simulations," *Science*, vol. 303, no. 5655, pp. 186–195, 2004. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1088172

[155] A. Warshel, "Energetics of enzyme catalysis," *PNAS; Proceedings of the National Academy of Sciences*, vol. 75, no. 11, pp. 5250–5254, 1978.

[156] ——, "Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites," *Journal of Biological Chemistry*, vol. 273, no. 42, pp. 27 035–27 038, Oct 1998. [Online]. Available: http://dx.doi.org/10.1074/jbc.273.42.27035

[157] A. Gutteridge and J. M. Thornton, "Understanding nature's catalytic toolkit," *Trends in Biochemical Sciences*, vol. 30, no. 11, pp. 622–629, Nov 2005. [Online]. Available: http://dx.doi.org/10.1016/j.tibs.2005.09.006

[158] E. F. Armstrong, "Enzymes. by j. b. s. haldane, m. a. monographs on biochemistry," *Journal of the Society of Chemical Industry*, vol. 49, no. 44, pp. 919–920, 1930.

[159] A. Warshel and M. Karplus, "Calculation of ground and excited state potential surfaces of conjugated molecules. i. formulation and parametrization," *Journal of the American Chemical Society*, vol. 94, no. 16, pp. 5612–5625, Aug 1972. [Online]. Available: http://dx.doi.org/10.1021/ja00771a014

[160] A. R. Fersht, "Profile of martin karplus, michael levitt, and arieh warshel, 2013 nobel laureates in chemistry," *PNAS; Proceedings of the National Academy of Sciences*, vol. 110, no. 49, pp. 19 656–19 657, 2013.

[161] L. Que and W. B. Tolman, "Biologically inspired oxidation catalysis," *Nature*, vol. 455, no. 7211, pp. 333–340, Sep 2008. [Online]. Available: http://dx.doi.org/10.1038/nature07371

[162] J. C. Lewis, P. S. Coelho, and F. H. Arnold, "Enzymatic functionalization of carbon–hydrogen bonds," *Chem. Soc. Rev.*, vol. 40, no. 4, pp. 2003–2021, 2011. [Online]. Available: http://dx.doi.org/10.1039/C0CS00067A

[163] A. Rauwerdink and R. J. Kazlauskas, "How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of alpha/beta-hydrolase fold enzymes," *ACS Catalysis*, vol. 5, no. 10, pp. 6153–6176, 2015, pMID: 28580193. [Online]. Available: https://doi.org/10.1021/acscatal.5b01539

[164] H. M., "Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms," *Current Protein and Peptide Science*, vol. 1, no. 2, pp. 209–235, 2000.

[165] N. Lenfant, T. Hotelier, E. Velluet, Y. Bourne, P. Marchot, and A. Chatonnet, "ESTHER, the database of the Î±/Î²-hydrolase fold superfamily of proteins: tools to explore diversity of functions," *Nucleic Acids Research*, vol. 41, no. D1, pp. D423–D429, 11 2012. [Online]. Available: https://doi.org/10.1093/nar/gks1154

[166] J. T. Mindrebo, C. M. Nartey, Y. Seto, M. D. Burkart, and J. P. Noel, "Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom," *Current Opinion in Structural Biology*, vol. 41, pp. 233–246, 2016, multi-protein assemblies in signaling * Catalysis and regulation. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959440X16301269

[167] T. L. Bauer, P. C. F. Buchholz, and J. Pleiss, "The modular structure of alpha/beta-hydrolases," *The FEBS Journal*, vol. 287, no. 5, pp. 1035–1053, 2020. [Online]. Available: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.15071

[168] A. Denesyuk, P. S. Dimitriou, M. S. Johnson, T. Nakayama, and K. Denessiouk, "The acid-base-nucleophile catalytic triad in abh-fold enzymes is coordinated by a set of structural elements," *PLOS ONE*, vol. 15, no. 2, p. e0229376, Feb 2020. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0229376

[169] A. Ordentlich, D. Barak, C. Kronman, N. Ariel, Y. Segall, B. Velan, and A. Shafferman, "Functional characteristics of the oxyanion hole in human acetylcholinesterase," *Journal of Biological Chemistry*, vol. 273, no. 31, pp. 19 509–19 517, Jul 1998. [Online]. Available: http://dx.doi.org/10.1074/jbc.273.31.19509

[170] R. MÉNARD and A. C. STORER, "Oxyanion hole interactions in serine and cysteine proteases," *Biological Chemistry Hoppe-Seyler*, vol. 373, no. 2, pp. 393–400, Jan 1992. [Online]. Available: http://dx.doi.org/10.1515/bchm3.1992.373.2.393

[171] P. S. Dimitriou, A. I. Denesyuk, T. Nakayama, M. S. Johnson, and K. Denessiouk, "Distinctive structural motifs co-ordinate the catalytic nucleophile and the residues of the oxyanion hole in the alpha/beta-hydrolase fold enzymes," *Protein Science*, vol. 28, no. 2, pp. 344–364, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3527

[172] R. Rink, J. Kingma, J. H. Lutje Spelberg, and D. B. Janssen, "Tyrosine residues serve as proton donor in the catalytic mechanism of epoxide hydrolase from agrobacterium radiobacter," *Biochemistry*, vol. 39, no. 18, pp. 5600–5613, Apr 2000. [Online]. Available: http://dx.doi.org/10.1021/bi9922392

[173] E. D. Schuiten, C. P. S. Badenhorst, G. J. Palm, L. Berndt, M. Lammers, J. Mican, D. Bednar, J. Damborsky, and U. T. Bornscheuer, "Promiscuous dehalogenase activity of the epoxide hydrolase coreh from corynebacterium sp. c12," *ACS Catalysis*, vol. 11, no. 10, pp. 6113–6120, May 2021. [Online]. Available: http://dx.doi.org/10.1021/acscatal.1c00851

[174] K. H. Hopmann and F. Himo, "Theoretical study of the full reaction mechanism of human soluble epoxide hydrolase," *Chemistry - A European Journal*, vol. 12, no. 26, pp. 6898–6909, Sep 2006. [Online]. Available: http://dx.doi.org/10.1002/chem.200501519

[175] J. Zou, B. M. Hallberg, T. Bergfors, F. Oesch, M. Arand, S. L. Mowbray, and T. A. Jones, "Structure of aspergillus niger epoxide hydrolase at 1.8 å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases," *Structure*, vol. 8, no. 2, pp. 111–122, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969212600000873

[176] M. Nardini, I. S. Ridder, H. J. Rozeboom, K. H. Kalk, R. Rink, D. B. Janssen, and B. W. Dijkstra, "The x-ray structure of epoxide hydrolase from agrobacterium radiobacter ad1," *Journal of Biological Chemistry*, vol. 274, no. 21, pp. 14 579–14 586, May 1999. [Online]. Available: http://dx.doi.org/10.1074/jbc.274.21.14579

[177] G. M. Lacourciere and R. N. Armstrong, "Microsomal and soluble epoxide hydrolases are members of the same family of cx bond hydrolase enzymes," *Chemical research in toxicology*, vol. 7, no. 2, pp. 121–124, 1994.

[178] R. N. Armstrong and C. S. Cassidy, "New structural and chemical insight into the catalytic mechanism of epoxide hydrolases," *Drug Metabolism Reviews*, vol. 32, no. 3-4, pp. 327–338, Jan 2000. [Online]. Available: http://dx.doi.org/10.1081/DMR-100102337

[179] T. Yamada, C. Morisseau, J. E. Maxwell, M. A. Argiriadi, D. W. Christianson, and B. D. Hammock, "Biochemical evidence for the involvement of tyrosine in epoxide activation during the catalytic cycle of epoxide hydrolase," *Journal of Biological Chemistry*, vol. 275, no. 30, pp. 23 082–23 088, Jul 2000. [Online]. Available: http://dx.doi.org/10.1074/jbc.M001464200

[180] M. Argiriadi, C. Morisseau, B. Hammock, and D. Christianson, "Detoxification of environmental mutagens and carcinogens: Structure, mechanism, and evolution of liver epoxide hydrolase," *PNAS; Proceedings of the National Academy of Sciences*, vol. 96, no. 19, pp. 10 637–10 642, 1999.

[181] M. A. Argiriadi, C. Morisseau, M. H. Goodrow, D. L. Dowdy, B. D. Hammock, and D. W. Christianson, "Binding of alkylurea inhibitors to epoxide hydrolase implicates active site tyrosines in substrate activation," *Journal of Biological Chemistry*, vol. 275, no. 20, pp. 15 265–15 270, May 2000. [Online]. Available: http://dx.doi.org/10.1074/jbc.M000278200

[182] G. A. Gomez, C. Morisseau, B. D. Hammock, and D. W. Christianson, "Structure of human epoxide hydrolase reveals mechanistic inferences on bifunctional catalysis in epoxide and phosphate ester hydrolysis,," *Biochemistry*, vol. 43, no. 16, pp. 4716–4723, Apr 2004. [Online]. Available: http://dx.doi.org/10.1021/bi036189j

[183] D. J. Tantillo, C. Jiangang, and K. N. Houk, "Theozymes and compuzymes: theoretical models for biological catalysis," *Current Opinion in Chemical Biology*, vol. 2, no. 6, pp.

743–750, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367593198801129

[184] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, and D. Baker, "De novo enzyme design using rosetta3," *PLoS ONE*, vol. 6, no. 5, p. e19230, May 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0019230

[185] A. Kohen, "Role of dynamics in enzyme catalysis: Substantial versus semantic controversies," *Accounts of Chemical Research*, vol. 48, no. 2, pp. 466–473, Dec 2014. [Online]. Available: http://dx.doi.org/10.1021/ar500322s

[186] V. C. Nashine, S. Hammes-Schiffer, and S. J. Benkovic, "Coupled motions in enzyme catalysis," *Current Opinion in Chemical Biology*, vol. 14, no. 5, pp. 644 – 651, 2010, nanotechnology and Miniaturization/Mechanisms. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367593110001031

[187] S. C. L. Kamerlin and A. Warshel, "At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?" *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 6, pp. 1339–1375, 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22654

[188] G. Bhabha, J. Lee, D. C. Ekiert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic, and P. E. Wright, "A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis," *Science*, vol. 332, no. 6026, pp. 234–238, 2011. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1198542

[189] C. K. Winkler, J. H. Schrittwieser, and W. Kroutil, "Power of biocatalysis for organic synthesis," *ACS Central Science*, vol. 7, no. 1, pp. 55–71, Jan 2021. [Online]. Available: http://dx.doi.org/10.1021/acscentsci.0c01496

[190] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," *Nature*, vol. 450, no. 7172, pp. 964–972, Dec 2007. [Online]. Available: http://dx.doi.org/10.1038/nature06522

[191] H. X. Kondo, N. Okimoto, G. Morimoto, and M. Taiji, "Free-energy landscapes of protein domain movements upon ligand binding," *The Journal of Physical*

Chemistry B, vol. 115, no. 23, pp. 7629–7636, May 2011. [Online]. Available: http://dx.doi.org/10.1021/jp111902t

[192] N. Kreß, J. M. Halder, L. R. Rapp, and B. Hauer, "Unlocked potential of dynamic elements in protein structures: channels and loops," *Current Opinion in Chemical Biology*, vol. 47, pp. 109 – 116, 2018, energy / Mechanistic Biology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367593118300395

[193] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *Nature Chemical Biology*, vol. 5, no. 11, pp. 789–796, Oct 2009. [Online]. Available: http://dx.doi.org/10.1038/nchembio.232

[194] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern, "A hierarchy of timescales in protein dynamics is linked to enzyme catalysis," *Nature*, vol. 450, no. 7171, pp. 913–916, Nov 2007. [Online]. Available: http://dx.doi.org/10.1038/nature06407

[195] P. K. Robinson, "Enzymes: principles and biotechnological applications," *Essays in Biochemistry*, vol. 59, pp. 1–41, 10 2015. [Online]. Available: https://doi.org/10.1042/bse0590001

[196] M. C. Zwier and L. T. Chong, "Reaching biological timescales with all-atom molecular dynamics simulations," *Current Opinion in Pharmacology*, vol. 10, no. 6, pp. 745–752, 2010, endocrine and metabolic diseases/New technologies - the importance of protein dynamics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1471489210001463

[197] I. R. Kleckner and M. P. Foster, "An introduction to nmr-based approaches for measuring protein dynamics," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1814, no. 8, pp. 942–968, 2011, protein Dynamics: Experimental and Computational Approaches. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570963910002864

[198] L. C. James and D. S. Tawfik, "Conformational diversity and protein evolution – a 60-year-old hypothesis revisited," *Trends in Biochemical Sciences*, vol. 28, no. 7, pp. 361–368, Jul 2003. [Online]. Available: http://dx.doi.org/10.1016/S0968-0004(03)00135-X

[199] E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, and et al., "The role of protein dynamics in the evolution of new enzyme function," *Nature Chemical Biology*, vol. 12, no. 11, pp. 944–950, Sep 2016. [Online]. Available: http://dx.doi.org/10.1038/nchembio.2175

[200] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The energy landscapes and motions of proteins," *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1749933

[201] G. Kar, O. Keskin, A. Gursoy, and R. Nussinov, "Allostery and population shift in drug discovery," *Current Opinion in Pharmacology*, vol. 10, no. 6, pp. 715–722, 2010, endocrine and metabolic diseases/New technologies - the importance of protein dynamics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1471489210001384

[202] M. A. Maria-Solano, E. Serrano-Hervás, A. Romero-Rivera, J. Iglesias-Fernández, and S. Osuna, "Role of conformational dynamics in the evolution of novel enzyme function," *Chemical Communications*, vol. 54, no. 50, pp. 6622–6634, 2018.

[203] P. G. Wolynes, "Recent successes of the energy landscape theory of protein folding and function," *Quarterly Reviews of Biophysics*, vol. 38, no. 4, pp. 405–410, 2005.

[204] R. Nussinov and P. G. Wolynes, "A second molecular biology revolution? the energy landscapes of biomolecular function," *Physical Chemistry Chemical Physics*, vol. 16, no. 14, p. 6321, 2014. [Online]. Available: http://dx.doi.org/10.1039/C4CP90027H

[205] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu, "Principles and overview of sampling methods for modeling macromolecular structure and dynamics," *PLOS Computational Biology*, vol. 12, no. 4, p. e1004619, Apr 2016. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1004619

[206] R. O. Dror, C. Young, and D. E. Shaw, *Anton, A Special-Purpose Molecular Simulation Machine*. Boston, MA: Springer US, 2011, pp. 60–71. [Online]. Available: https://doi.org/10.1007/978-0-387-09766-4_199

[207] B. Ma and R. Nussinov, "Enzyme dynamics point to stepwise conformational selection in catalysis," *Current Opinion in Chemical Biology*, vol. 14, no. 5, pp. 652–659, 2010, nanotechnology and Miniaturization/Mechanisms. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367593110001146

[208] A. D. Vogt and E. Di Cera, "Conformational selection or induced fit? a critical appraisal of the kinetic mechanism," *Biochemistry*, vol. 51, no. 30, pp. 5894–5902, 2012, pMID: 22775458. [Online]. Available: https://doi.org/10.1021/bi3006913

[209] M. Kovermann, C. Grundström, A. Sauer-Eriksson, U. Sauer, and M. Wolf-Watz, "Structural basis for ligand binding to an enzyme by a conformational selection pathway," *PNAS; Proceedings of the National Academy of Sciences*, vol. 114, no. 24, pp. 6298–6303, 2017.

[210] R. Otten, R. A. P. Pádua, H. A. Bunzel, V. Nguyen, W. Pitsawong, M. Patterson, S. Sui, S. L. Perry, A. E. Cohen, D. Hilvert, and D. Kern, "How directed evolution reshapes the energy landscape in an enzyme to boost catalysis," *Science*, vol. 370, no. 6523, pp. 1442–1446, 2020. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abd3623

[211] D. Petrović, V. A. Risso, S. C. L. Kamerlin, and J. M. Sanchez-Ruiz, "Conformational dynamics and enzyme evolution," *Journal of The Royal Society Interface*, vol. 15, no. 144, p. 20180330, Jul 2018. [Online]. Available: http://dx.doi.org/10.1098/rsif.2018.0330

[212] P. A. Dalby, "Optimising enzyme function by directed evolution," *Current Opinion in Structural Biology*, vol. 13, no. 4, pp. 500–505, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959440X03001015

[213] H. Renata, Z. J. Wang, and F. H. Arnold, "Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution," *Angewandte Chemie International Edition*, vol. 54, no. 11, pp. 3351–3367, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201409470

[214] C. Zeymer and D. Hilvert, "Directed evolution of protein catalysts," *Annual Review of Biochemistry*, vol. 87, no. 1, pp. 131–157, 2018, pMID: 29494241. [Online]. Available: https://doi.org/10.1146/annurev-biochem-062917-012034

[215] D. Li, M. S. Liu, and B. Ji, "Mapping the dynamics landscape of conformational transitions in enzyme: The adenylate kinase case," *Biophysical Journal*, vol. 109, no. 3, pp. 647–660, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006349515006700

[216] V. A. Risso, S. Martinez-Rodriguez, A. M. Candel, D. M. Krüger, D. Pantoja-Uceda, M. Ortega-Muñoz, F. Santoyo-Gonzalez, E. A. Gaucher, S. C. L. Kamerlin, M. Bruix, and et al., "De novo active sites for resurrected precambrian enzymes," *Nature Communications*, vol. 8, no. 1, Jul 2017. [Online]. Available: http://dx.doi.org/10.1038/ncomms16113

[217] G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, and K. N. Houk, "Computational enzyme design," *Angewandte Chemie International Edition*, vol. 52, no. 22, pp. 5700–5725, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201204077

[218] R. Blomberg, H. Kries, D. M. Pinkas, P. R. E. Mittl, M. G. Grütter, H. K. Privett, S. L. Mayo, and D. Hilvert, "Precision is essential for efficient catalysis in an evolved kemp eliminase," *Nature*, vol. 503, no. 7476, pp. 418–421, Oct 2013. [Online]. Available: http://dx.doi.org/10.1038/nature12623

[219] A. Broom, R. V. Rakotoharisoa, M. C. Thompson, N. Zarifi, E. Nguyen, N. Mukhametzhanov, L. Liu, J. S. Fraser, and R. A. Chica, "Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico," *Nature Communications*, vol. 11, no. 1, Sep 2020. [Online]. Available: http://dx.doi.org/10.1038/s41467-020-18619-x

[220] A. Pabis, V. A. Risso, J. M. Sanchez-Ruiz, and S. C. Kamerlin, "Cooperativity and flexibility in enzyme evolution," *Current Opinion in Structural Biology*, vol. 48, pp. 83–92, 2018, folding and binding in silico, in vitro and in cellula * Proteins: An Evolutionary Perspective. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959440X17300933

[221] R. Otten, R. A. P. Pádua, H. A. Bunzel, V. Nguyen, W. Pitsawong, M. Patterson, S. Sui, S. L. Perry, A. E. Cohen, D. Hilvert, and D. Kern, "How directed evolution reshapes the energy landscape in an enzyme to boost

catalysis," *Science*, vol. 370, no. 6523, pp. 1442–1446, 2020. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abd3623

[222] D. J. Mandell and T. Kortemme, "Backbone flexibility in computational protein design," *Current Opinion in Biotechnology*, vol. 20, no. 4, pp. 420 – 428, 2009, protein technologies / Systems and synthetic biology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0958166909000913

[223] J. K. Lassila, "Conformational diversity and computational enzyme design," *Current Opinion in Chemical Biology*, vol. 14, no. 5, pp. 676 – 682, 2010, nanotechnology and Miniaturization/Mechanisms. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367593110001122

[224] M. A. Maria-Solano, T. Kinateder, J. Iglesias-Fernández, R. Sterner, and S. Osuna, "In silico identification and experimental validation of distal activity-enhancing mutations in tryptophan synthase," *ACS Catalysis*, vol. 11, no. 21, pp. 13 733–13 743, Oct 2021. [Online]. Available: http://dx.doi.org/10.1021/acscatal.1c03950

[225] J. McFadden and J. Al-Khalili, "The origins of quantum biology," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 474, no. 2220, p. 20180674, Dec 2018. [Online]. Available: http://dx.doi.org/10.1098/rspa.2018.0674

[226] A. V. Arbuznikov, "Hybrid exchange correlation functionals and potentials: Concept elaboration," *Journal of Structural Chemistry*, vol. 48, no. S1, pp. S1–S31, Jan 2007. [Online]. Available: http://dx.doi.org/10.1007/s10947-007-0147-0

[227] F. Jensen, *Introduction to computational chemistry*, second edition ed. John Wiley & Sons, 2007.

[228] C. Lee, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785–789, 1988.

[229] A. D. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, Apr 1993. [Online]. Available: http://dx.doi.org/10.1063/1.464913

[230] P. M. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, "The performance of the becke—lee—yang—parr (b—lyp) density functional theory with various basis

sets," *Chemical Physics Letters*, vol. 197, no. 4, pp. 499–505, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/000926149285807M

[231] P. E. Siegbahn and F. Himo, "The quantum chemical cluster approach for modeling enzyme reactions," *WIREs Computational Molecular Science*, vol. 1, no. 3, pp. 323–336, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.13

[232] P. E. M. Siegbahn and F. Himo, "Recent developments of the quantum chemical cluster approach for modeling enzyme reactions," *JBIC Journal of Biological Inorganic Chemistry*, vol. 14, no. 5, pp. 643–651, May 2009. [Online]. Available: http://dx.doi.org/10.1007/s00775-009-0511-y

[233] M. E. S. Lind and F. Himo, "Quantum chemical modeling of enantioconvergency in soluble epoxide hydrolase," *ACS Catalysis*, vol. 6, no. 12, pp. 8145–8155, Nov 2016. [Online]. Available: http://dx.doi.org/10.1021/acscatal.6b01562

[234] F. Himo, "Recent trends in quantum chemical modeling of enzymatic reactions," *Journal of the American Chemical Society*, vol. 139, no. 20, pp. 6780–6786, May 2017. [Online]. Available: http://dx.doi.org/10.1021/jacs.7b02671

[235] G. A. DiLabio, M. Koleini, and E. Torres, "Extension of the b3lyp–dispersion-correcting potential approach to the accurate treatment of both inter- and intra-molecular interactions," *Theoretical Chemistry Accounts*, vol. 132, no. 10, Aug 2013. [Online]. Available: http://dx.doi.org/10.1007/s00214-013-1389-x

[236] E. Torres and G. A. DiLabio, "A (nearly) universally applicable method for modeling noncovalent interactions using b3lyp," *The Journal of Physical Chemistry Letters*, vol. 3, no. 13, pp. 1738–1744, Jun 2012. [Online]. Available: http://dx.doi.org/10.1021/jz300554y

[237] S. Grimme, "Density functional theory with london dispersion corrections," *WIREs Computational Molecular Science*, vol. 1, no. 2, pp. 211–228, 2011. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.30

[238] M. Bursch, E. Caldeweyher, A. Hansen, H. Neugebauer, S. Ehlert, and S. Grimme, "Understanding and quantifying london dispersion effects in organometallic complexes,"

Accounts of Chemical Research, vol. 52, no. 1, pp. 258–266, Dec 2018. [Online]. Available: http://dx.doi.org/10.1021/acs.accounts.8b00505

[239] L. Goerigk, "Chapter 6 - a comprehensive overview of the dft-d3 london-dispersion correction," in *Non-Covalent Interactions in Quantum Chemistry and Physics*, A. Otero de la Roza and G. A. DiLabio, Eds. Elsevier, 2017, pp. 195–219. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128098356000074

[240] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian˜16 Revision D.01," 2016.

[241] C. J. Cramer, *Essentials of computational chemistry: theories and models*. John Wiley & Sons, 2013.

[242] F. Weinhold and P. Schleyer, "Encyclopedia of computational chemistry," *New York: John Wiley & Sons*, vol. 3, p. 1792, 1998.

[243] J. M. Bofill, "Updated hessian matrix and the restricted step method for locating transition structures," *Journal of Computational Chemistry*, vol. 15, no. 1, pp. 1–11, 1994. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540150102

[244] K. N. Houk and P. H.-Y. Cheong, "Computational prediction of small-molecule catalysts," *Nature*, vol. 455, no. 7211, pp. 309–313, Sep 2008. [Online]. Available: http://dx.doi.org/10.1038/nature07368

[245] F. Himo and S. P. de Visser, "Status report on the quantum chemical cluster approach for modeling enzyme reactions," *Communications Chemistry*, vol. 5, no. 1, pp. 1–4, 2020.

[246] B. A. Amrein, P. Bauer, F. Duarte, Å. Janfalk Carlsson, A. Naworyta, S. L. Mowbray, M. Widersten, and S. C. L. Kamerlin, "Expanding the catalytic triad in epoxide hydrolases and related enzymes," *ACS Catalysis*, vol. 5, no. 10, pp. 5702–5713, Aug 2015. [Online]. Available: http://dx.doi.org/10.1021/acscatal.5b01639

[247] P. Bauer, Å. J. Carlsson, B. A. Amrein, D. Dobritzsch, M. Widersten, and S. C. L. Kamerlin, "Conformational diversity and enantioconvergence in potato epoxide hydrolase 1," *Organic and Biomolecular Chemistry*, vol. 14, no. 24, pp. 5639–5651, 2016. [Online]. Available: http://dx.doi.org/10.1039/C6OB00060F

[248] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, and W. Yang, "Nciplot: A program for plotting noncovalent interaction regions," *Journal of Chemical Theory and Computation*, vol. 7, no. 3, pp. 625–632, Jan 2011. [Online]. Available: http://dx.doi.org/10.1021/ct100641a

[249] D. J. Rigden, *From Protein Structure to Function with Bioinformatics*, 2009. [Online]. Available: https://doi.org/10.1007/978-1-4020-9058-5

[250] A. Perez, J. A. Morrone, C. Simmerling, and K. A. Dill, "Advances in free-energy-based simulations of protein folding and ligand binding," *Current Opinion in Structural Biology*, vol. 36, pp. 25–31, 2016, folding and binding * Nucleic acids-protein complexes. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959440X15001888

[251] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, " H++: a server for estimating p Ka s and adding missing hydrogens to macromolecules ," *Nucleic Acids Research*, vol. 33, no. suppl˙2, pp. W368–W371, 07 2005. [Online]. Available: https://doi.org/10.1093/nar/gki464

[252] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "Propka3: Consistent treatment of internal and surface residues in empirical pka predictions," *Journal of Chemical Theory and Computation*, vol. 7, no. 2, pp. 525–537, Jan 2011. [Online]. Available: http://dx.doi.org/10.1021/ct100578z

[253] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0263785596000185

[254] W. L. DeLano *et al.*, "Pymol: An open-source molecular graphics tool," *CCP4 Newsl. Protein Crystallogr*, vol. 40, no. 1, pp. 82–92, 2002.

[255] X. Zhang and K. Houk, "Why enzymes are proficient catalysts: beyond the pauling paradigm," *Accounts of chemical research*, vol. 38, no. 5, pp. 379–385, 2005.

[256] G. A. Tribello and P. Gasparotto, "Using dimensionality reduction to analyze protein trajectories," *Frontiers in Molecular Biosciences*, vol. 6, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmolb.2019.00046

[257] G. Pérez-Hernández and F. Noé, "Hierarchical time-lagged independent component analysis: Computing slow modes and reaction coordinates for large molecular systems," *Journal of Chemical Theory and Computation*, vol. 12, no. 12, pp. 6118–6129, Nov 2016. [Online]. Available: http://dx.doi.org/10.1021/acs.jctc.6b00738

[258] M. M. Sultan and V. S. Pande, "tica-metadynamics: Accelerating metadynamics by using kinetically selected collective variables," *Journal of Chemical Theory and Computation*, vol. 13, no. 6, pp. 2440–2447, May 2017. [Online]. Available: http://dx.doi.org/10.1021/acs.jctc.7b00182

[259] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for markov model construction," *The Journal of Chemical Physics*, vol. 139, no. 1, p. 015102, Jul 2013. [Online]. Available: http://dx.doi.org/10.1063/1.4811489

[260] E. N. Jacobsen, "Asymmetric catalysis of epoxide ring-opening reactions," *Accounts of Chemical Research*, vol. 33, no. 6, pp. 421–431, May 2000. [Online]. Available: http://dx.doi.org/10.1021/ar960061v

[261] J. R. Wagner, J. Sørensen, N. Hensley, C. Wong, C. Zhu, T. Perison, and R. E. Amaro, "Povme 3.0: Software for mapping binding pocket flexibility," *Journal of Chemical Theory and Computation*, vol. 13, no. 9, pp. 4584–4592, 2017, pMID: 28800393. [Online]. Available: https://doi.org/10.1021/acs.jctc.7b00500