COMPUTATIONAL STUDIES OF ENZYMATIC AND BIOMIMETIC CATALYSTS

Adrian Romero Rivera

Universitat de Girona

DOCTORAL THESIS

Computational studies of enzymatic and biomimetic catalysts

Adrian Romero Rivera

2018

Doctoral programme in Chemistry

Supervised by: Prof. Marcel Swart and Prof. Sílvia Osuna
Tutor: Prof. Marcel Swart

Presented to obtain the degree of PhD at the University of Girona

Universitat de Girona

Prof. Marcel Swart and Prof. Sílvia Osuna from the University of Girona,

WE DECLARE:

That the thesis entitled “Computational studies of enzymatic and biomimetic catalysts”, presented by Adrian Romero Rivera to obtain a doctoral degree, has been completed under our supervision and meets the requirements to opt for an International Doctorate.

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

Signature

Prof. Marcel Swart

Prof. Sílvia Osuna

Girona, November 2th, 2018

Dedicated to all people who contributes directly or indirectly in this thesis

ACKNOWLEDGEMENTS

This work would not have been possible without the following financial support:

Generalitat de Catalunya for the PhD grant FI-2015, Ministeria de Economía y Competividad (MINECO) of Spain for the project CTQ2014-59212-P, the European Commission for COST action CM1305 (ECOSTBio), the European Research Council (ERC) Starting Grant ERC-2015-StG-679001, Career Integration Grant (CIG) 2013-CIG-630978, and the University of Girona.

Special thanks to the Institut de Química Computacional i Catàlisi (IQCC), Barcelona Supercomputing Center (BSC), and Swedish National Infrastructure for Computing (SNIC) for the computational resources.

Ara que aquesta etapa ha acabat vull donar les gràcies als meus supervisors Marcel Swart i Sílvia Osuna. Marcel gràcies per introduïr-me en el món dels complexos i Sílvia gràcies per introduir-me al món dels enzims. Durant tots aquests anys hem fet molts tipus de projectes diferents que alguns han sigut més difícils que altres però tots han anat sortint. Gràcies per l'ajuda i tot el recolzament.

Vull donar també les gràcies als tres post-docs (Javi, Marc, i Ferran) que m'han ajudat en tots el projectes d'enzims, sense ells no s'haguessin acabat. Amb això he aprés la importància essencial de l'ajuda dels post-docs en els projectes. Que tingueu molta sort en la vostra carrera científica.

Els membre del CompBioLab, Miguel Angel pels diferents primers projectes ADH que vam suar. Christian, Eila, Miquel que continueu treballant a tope, que per molt difícils que siguin les coses al final sempre es trobat una solució. Els nous membres, Guillem i Carla, ara que comenceu l'aventura gaudiu-la i apreneu lo màxim que pugueu. Moltes gràcies a tots per resoldrem milers de dubtes que he tingut i m'heu solucionat. També, en Pedro J. i Cristina per deixar-se ajudar amb els TFG ha sigut una bona experiencia, espero que us vagi molt bé.

Lorenzo and Davide “the italian connexion”, many thanks to bring style to the IQCC and for making the IQCC a more international group. I will not forget that “black and blue does not fit!”. Good luck guys.

Per la gent que ja no esta al IQCC, Carme i Quim moltes gràcies per tota la feina tots aquests any, us ho heu currat molt! Espero que estigui anant tot molt bé. També al Ouissam (mai oblidaré la aventura en Uppsala, repetirem al estiu per veure com ens va!), Vero i Abril. Hem coincidit durant molts anys i us he vist la bona feina que heu fet i com heu pogut continuar en el món de la recerca, heu sigut un exemple a seguir, gràcies.

Per descomptat, Dani, gràcies per cuidar les màquines i que funcionin a un gran rendiment. Tothom sap que si no sigués per tu potser les coses no anirien tan bé. Moltes gràcies per tot!

I would also to thank Ulf Ryde for leading me the opportunity to do a research stay in his group to learn new stuff and to discover Sweden.

A la gent del Paleo (Joan, David, Edu, Albert, Mario, Quim, …) gràcies per fer que les tardes siguin diferents a la rutina i manternir-nos en “forma”. Mai oblidaré el special paleo, ens ho vam currar molt, espero que repetiu.

Ahora toca a la família, mi padre por traerme a Girona los domingos cuando los visitaba, mi hermano por las rutas para que se pusiera en forma, al Iván por ver lo friki que se está volviendo con el paso de los años y mi madre por las discusiones que tenemos sobre los temas raro que le pregunto. Gracias por estar allí y vuestro apollo. A la Angela per aguantar-me tot aquest temps al pis i pels viatges que hem fet, que no han sigut pocs, vaja caminata al camino per les muntanyes I mai oblidaré India, va ser dur però al final vam gaudir.

A los amigos que lleváis ahí desde hace años (Sergi, David, Niels, Adriano, Uri, Manel, Sergio, Jordi, Raúl, Santi). Hemos ido a festivales, me habeís venido a ver a Suécia (refugee style), y hemos intentado mantener el contacto y aún se puede contar con vosotros. Sois grandes!!!

Per últim gràcies al IQCC i QTMEM per donar-me la oportunitat de poder fer el doctorat a la UdG, ha sigut una gran oportunitat, perquè gràcies a que és un bon grup vaig poder aconseguir una beca.

Com podeu veure només he donat les gràcies i ja m'ha ocupat molt d'espai, potser m'he oblidat a gent. Han passat moltes coses bones i dolentes durant aquests anys. Hem fet moltes coses com sortir de festa, perdrem per Girona, beure uns zumos que em van deixar el estomac K.O., rogaines, futbol, congressos, calçotades, excursions a muntanyes on mai hagués anat si no fos per vosaltres, esquiar, celebrar que un company serà pare, etc… Explicar tot m'ocuparia massa espai, però quan ens trobem en un futur ja podrem recordar tots aquests moments.

Gràcies!

FULL LIST OF PUBLICATIONS

List of publications results from this thesis

1) Romero-Rivera, A.; Ryde, U.; Feixas, F.; Osuna, S.; Swart, M. Role of a lid domain in the conformational dynamics of 2,4-Quercetinase dioxygenase. *In preparation*.
2) Romero-Rivera, A; Iglesias-Fernández, J.; Osuna, S. Exploring the conversion of a D-sialic acid aldolase into a L-KDO aldolase. *Eur. J. Org. Chem.*, **2018**, 2603-2608.
3) Romero-Rivera, A; Garcia-Borràs, M.; Osuna, S. Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catal.*, **2017**, *7(12)*, 8524-8532.
4) Romero-Rivera, A.; Swart, M. Unraveling the reaction pathway of EUK-8 catalase complex. *In preparation*.
5) Romero-Rivera, A.; Swart, M. New fit parameters for Mössbauer parameters predictions. *In preparation*.
6) Hill, E. A.; Weitz, A. C.; Onderko, E.; Romero-Rivera, A.; Guo, Y.; Swart, M.; Bominaar, E. L.; Green, M. T.; Hendrich, M. P.; Lacy, D. C.; Borovik, A. S. Reactivity of an Fe^{IV}-Oxo complex with protons and oxidants. *J. Am. Chem. Soc.* **2016**, *138*, 13143-13146.

Other publications

7) Maria-Solano, M. A.; Romero-Rivera, A.; Osuna, S. Exploring the reversal of enantioselectivity on a zinc-dependent alcohol dehydrogenase. *Org. Biomol. Chem.*, **2017**, *15*, 4122-4129.
8) Li, G; Maria-Solano, M. A.; Romero-Rivera, A.; Osuna, S., Reetz, M. T. Inducing high activity of a thermophilic enzyme at ambient temperatures by directed evolutions. *Chem. Commun.*, **2017**, *53*, 9454-9457.
9) Garcia-Borràs, M.; Romero-Rivera, A.; Osuna, S.; Luis, J. M.; Swart, M.; Solà, M. The frozen cage model: a computational low-cost tool for predicting the exohedral regioselectvity of cycloaddition reactions involving endohedral metallofullerenes. *J. Chem. Theory Comput.*, **2012**, *8*, 1671-1683.
10) Castro, A. C.; Romero-Rivera, A.; Osuna, S; Houk, K. N.; Swart, M. Computational NMR Spectroscopy for Host-Guest Hemicarcerands. *In preparation.*

11) Call, A.; Casadevall, C.; Romero-Rivera, A.; Sommer, D. J.; Osuna, S.; Ghirlanda, G.; Lloret-Fillol, J. Improved activity and stability of a cobalt reduction catalyst by supramolecular confinement in streptavidin. *Submitted for publication*.

12) Romero-Rivera, A; Garcia-Borràs, M.; Osuna, S. Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.*, **2017**, *53*, 284-297.

13) Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Roles of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.*, **2018**, *54*, 6622-6634.

14) Arqué, X.; Patiño, T.; Romero-Rivera, A.; Feixas, F; Osuna, S; Sánchez, S. Exploring enzymatic reactions to power hollow silica microcapsules. *Submitted for publication*.

LIST OF ABBREVIATIONS

Abbreviations	Description
ADF	Amsterdam density functional
benzimpnOH	N,N,N',N'-tetrakis(2-methylenebenzimidazolyl)-1,3-diaminopropan-2-ol
buea	1,1,1-tris[(N'tert-butylureaylato)-N-ethyl]aminato
ADHs	Alchohol dehydrogenases
AdK	Adenylate kinase
aMD	Accelerated molecular dynamics
ATP	Adenosine triphosphate
CASCO	Catalytic selectivity by computational design
CASTing	Combinatorial active-site test
CDM	Cationic dummy model
CO2	Carbon dioxide
COSMO	Conductor-like screening model
CVs	Collective variables
DE	Directed evolution
DFT	Density functional theory
DHRF	Dihydrofolate reductase
EC	Enzyme comission
EFG	Electric field gradient
EPR	electron paramagnetic resonance
E-S	Enzyme-substrate
FAD+	Flavin adenine dinucleotide
FEL	Free energy landscape
FF	Force fields
FMM	Fast multiple moment
GGA	Generalized gradient approximation
GTOs	Gaussian type orbitals
HF	Hartree-fock
hMGL	Human monoacylglycerol lipase
HMM	Henri-Michaelis-Menten
IGPS	Imidazole-glycerol phosphate synthase
IR	infrared
ISM	Iterative saturation mutagenesis
KS	Kohn-Sham

LBHBs	Low-barrier hydrogen bonds
LDA	Local density approximation
MD	Molecular dynamics
MLs	Mutability landscapes
MM	Molecular mechanics
MSA	Multiple sequence and structure alignments
MDC	Multiple Derived Charge
NACs	Near-attack-conformers
NADP+	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate protonated
N4Py	N,N-bis(2-pyridylmethyl)-N-bis(2-pyridil)methylamine
NMR	Nuclear magnetic resonance
NPT	Isobaric-isothermal ensembles
NVE	Microcanonical ensemble
NVT	Canonical ensembles
PBCs	Periodic boundary conditions
PCA	Principal component analysis
PDB	Protein data bank
PME	Particle mesh ewald
QM	Quantum mechanics
QUILD	Quantum-region interconnected by local description
RA	Retro-aldolases
RMSD	Root mean square deviation
salen	1,2-bis(salicylideneamino)ethane
SES	Solvent excluding surface
SPM	Shortest Path Map
STOs	Slater type orbitals
tICA	Time-structure independent component analysis
TMC	1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane
TS	Transition-state
TPA	tris(5-phenylpyrrol-2-ylmethyl)amine
UV-vis	Ultraviolet-visible
UHF	Unrestricted Hartree-fock
US	Umbrella Sampling
ZORA	Zeroth-order regular approximation

LIST OF FIGURES

LIST OF TABLES

CONTENTS

SUMMARY

In this thesis, different computational strategies are used to provide detailed knowledge of biological and chemical properties of (bio)catalysts. This is the idea behind the computational microscope in which the choice of technique depends on the accuracy needed and the size of the system. Different projects have been selected in which computational techniques are used for understanding conformational dynamics of enzymes, the mechanism of biomimetic complexes, and to characterize iron-oxygen structures through the calculation of Mössbauer parameters. For each project the major feature is the understanding and the characterization of different properties in natural biocatalyst and biomimetic catalysts, which highlight the insights that computational tools can provide a deeper knowledge of the selected topics.

Enzymes are the most efficient biocatalysts in Nature. The primary function of enzymes is the acceleration of chemical reactions that are important to maintain and create life in all of its many forms on Earth. These biocatalysts have evolved for catalyzing biochemical processes under mild conditions; however, enzymes are in general not capable of catalyzing reactions, and/or substrates, which are important for industrial purposes. Hence, biocatalysts need to be engineered by means of directed evolution, which introduces mutations for improving enzymatic properties to do non-biological tasks. The mutations can be introduced randomly in the active site or at distal positions in the enzyme, thereby inducing changes in the conformational dynamics of enzymes, without any clear indication of how these new residues alter the performance of the enzymes. In this thesis, analysis of conformational dynamics have been carried out in a natural copper dependent metalloenzyme (**Chapter 4**), an aldolase with mutations for accepting a non-natural substrate (**Chapters 5**), and *de novo* designed retro-aldolases with mutations for enhancing the catalytic efficiency to reach the levels of natural enzymes (**Chapters 6**). The introduced mutations in the last two enzymes by directed evolution are located all over the protein. These different projects highlight the importance of conformational dynamics for substrate binding and product release, how mutations at distal

positions induce population shifts to conformations that enhance the efficiency to catalyze non-natural substrates, and even to predict *a priori* possible important residues for altering the conformational dynamics of enzymes.

Biomimetic chemistry seeks to design novel efficient metal-based organocatalysts mimicking the structure-function from the enzyme's active site. The problem is that we still do not fully understand how enzymes work, which makes the design and synthesis of non-biological complexes as efficient as biocatalysts a challenging task. In **Chapter 7**, detailed mechanism pathways for EUK-8 have been proposed, a biomimetic complex for the decomposition of hydrogen peroxide. The importance for elucidating a good description of the mechanism will give an idea of which stationary points can be important to enhance the reactivity. Thus, the understanding of the mechanism with different possible pathways and the good description of spin state of the metal in reactivity provide important knowledge to rationally design analogues of EUK-8.

^{57}Fe Mössbauer spectroscopy is a technique that provides information about the chemical nature of Fe systems, regardless of its spin and oxidation states. Because iron is an essential element in all living creatures and is incorporated into many enzymes and organometallic complexes, in **Chapter 8**, a recalibration of a training set to calculate Mössbauer parameters have been presented using DFT functionals (S12g and OPBE) that correctly describe the spin state of iron systems. Since the Mössbauer spectra is not always straightforward to analyze, these new fit parameters support experimental Mössbauer data and also help to elucidate the characterization of Fe-oxygen species and their respective spin state when no X-ray structure is available.

The thesis is divided into eight chapters, which includes: a general introduction, characterization techniques, the principal goals of the thesis, three chapters related with conformational dynamics of enzymes, a chapter on the mechanism of biomimetic complexes, a chapter to characterize Fe-oxygen structures through the calculation of Mössbauer parameters, and a final chapter with the main conclusions obtained from the different projects studied.

RESUM

En aquesta tesi s'han fet servir estratègies computacionals per a proporcionar un coneixement detallat de propietats biològiques i químiques dels (bio)catalitzadors. Aquesta és la idea darrera del microscopi computacional on escollir la tècnica depèn de la precisió necessària i de la mida del sistema. Diferents projectes s'han seleccionat on les tècniques computacionals s'han fet servir per entendre la dinàmica conformacional d'enzims, el mecanisme de complexes biomimètics i la caracterització d'estructures ferro-oxigen a través de càlculs de paràmetres Mössbauer. La característica més important per cada projecte és entendre i caracteritzar diferent propietats en biocatalitzadors naturals i catalitzador biomimètics, destacant la visió de que les eines computacionals poden proporcionar un coneixement profund dels diferents temes seleccionats.

Els enzims són els biocatalitzadors més eficients a la Natura. La funció primària dels enzims és l'acceleració de reaccions químiques importants per a mantenir i crear vida en totes les formes de vida a la Terra. Aquests biocatalitzadors han evolucionat per catalitzar processos bioquímics sota condicions suaus, no obstant, en general els enzims no són capaços de catalitzar reaccions, i/o substrats que són importants per a propòsits industrials. Per tant, els biocatalitzadors necessiten ser modificats a través de l'evolució dirigida, que introdueix mutacions per a millorar les propietats enzimàtiques per fer tasques no biològiques. Les mutacions poden ser introduïdes aleatòriament en el centre actiu o en posicions llunyanes, així s'altera la dinàmica conformacional dels enzims, sense una clara indicació de com aquests nous residus modifiquen el rendiment dels enzims. En aquesta tesi, hem realitzat un anàlisi en de la dinàmica conformacional en metal·loenzims naturals que depenen del coure (**Capítol 4**), una aldolasa amb mutacions per acceptar un substrat no natural (**Capítol 5**) i una "*de novo"* retro-aldolasa dissenyada computacionalment amb mutacions que milloren la eficiència catalítica fins arribar a nivells d'enzims naturals (**Capítol 6**). La introducció de mutacions en els dos últims enzims a través d'evolució dirigida estan localitzats a tota la proteïna. Aquests projectes diferents ressalten la importància de la dinàmica conformacional de l'enzim per acceptar substrats i alliberar el producte, així com les mutacions llunyanes del centre actiu in-

dueixen un canvi a les poblacions de les conformacions que milloren l'eficiència a l'hora d'acceptar substrats no naturals, i fins i tot es poden predir *a priori* possibles residus importants per alterar la dinàmica conformacional dels enzims.

La química biomimètica cerca dissenyar nous organocatalitzadors eficients imitant la funció estructural del centre actiu de l'enzim. El problema és que encara no entenem totalment com funcionen els enzims, això fa que el disseny i síntesi de complexos no biològics amb eficiències similars als enzims sigui un repte. En el **Capítol 7**, mostrem un mecanisme detallat per EUK-8, un complex biomimètic per la descomposició d'aigua oxigenada. La importància per eludir una bona descripció del mecanisme donarà una idea de quins punts de la reacció són importants per a millorar la reactivitat. Per tant, entendre el mecanisme amb les diferent vies possibles i una bona descripció dels estats d'espín del metall en la reactivitat proporciona un coneixement important per a dissenyar racionalment anàlegs de EUK-8.

L'espectroscòpia Mössbauer de ^{57}Fe és una tècnica que proporciona informació sobra la naturalesa química de sistemes de Fe, respecte els estat d'espín i oxidació. Ja que el ferro és un element essencial per totes les criatures vives i és incorporat en molts enzims i complexos organometàl·lics, en el **Capítol 8**, hem presentat una recalibració d'un conjunt de càlculs per a predir paràmetres de Mössbauer fent servir funcionals DFT (S12g i OPBE) que descriuen correctament l'estat d'espín del sistema de ferro. Com que els espectres de Mössbauer no sempre són fàcils d'analitzar, aquests nous paràmetres d'ajust recolzen les dades experimentals de Mössbauer i també ajuden a caracteritzar espècies de Fe-oxigen i el respectiu estat d'espín quan no hi han estructures de raig-X disponibles.

La tesi esta dividida en vuit capítols, que inclouen: una introducció general, tècniques de caracterització, els objectius principals de la tesi, tres capítols relacionats amb la dinàmica conformacional d'enzims, un capítol del mecanisme de un complex biomimètic, un capítol de la caracterització d'estructures Fe-oxigen a través de càlculs de paràmetres de Mössbauer, i un capítol final amb les conclusions principals obtingudes en els diferent projectes estudiats.

RESUMEN

En esta tesis se han utilizado estrategias computacionales para proporcionar un conocimiento detallado de las propiedades biológicas y químicas de los (bio)catalizadores. Esta es la idea detrás del microscopio computacional donde escoger la técnica depende de la precisión necesaria y del tamaño del sistema. Se han seleccionado diferentes proyectos donde varias técnicas computacionales se han utilizado para entender la dinámica conformacional de enzimas, el mecanismo de complejos biomiméticos y la caracterización de la estructura hierro-oxígeno a través de cálculos de parámetros Mössbauer. La característica más importante para cada proyecto es entender y caracterizar diferentes propiedades de los biocatalizadores naturales y catalizadores biomiméticos, destacando la visión de que las herramientas computacionales pueden proporcionar un conocimiento profundo de los diferentes proyectos seleccionados.

Las enzimas son los biocatalizadores más eficientes de la Naturaleza. La función primaria de las enzimas es la aceleración de reacciones químicas importantes para mantener y crear la vida en todas las formas de vida en la Tierra. Estos biocatalizadores han evolucionado para catalizar procesos bioquímicos bajo condiciones suaves, no obstante, en general las enzimas no son capaces de catalizar reacciones, y/o aceptar sustratos que son importantes para propósitos industriales. Por tanto, los biocatalizadores necesitan ser modificados a través de la evolución dirigida, que introduce mutaciones para mejorar las propiedades enzimáticas para realizar funciones no biológicas. Las mutaciones pueden ser introducidas aleatoriamente en el centro activo o en posiciones lejanas de la enzima, así se introducen cambios en la dinámica conformacional de las enzimas, sin una clara indicación de cómo estos nuevos residuos alteran el rendimiento de las enzimas. En esta tesis se ha realizado un análisis de la dinámica conformacional en metaloenzimas naturales que dependen de cobre (**Capítulo 4**), una aldolasa con mutaciones para aceptar un sustrato no natural (**Capítulo 5**) y una "*de novo"* retro-aldolasa diseñada computacionalmente con mutaciones que mejoran la eficiencia catalítica hasta llegar a niveles de enzimas naturales (**Capítulo 6**). La introducción de mutaciones en las dos últimas enzimas a través de evolución dirigida están localizadas en toda la proteína. Es-

tos diferentes proyectos resaltan la importancia de la dinámica conformacional para aceptar sustratos y liberar productos, así como las mutaciones lejanas del centro activo inducen un cambio en las poblaciones de las conformaciones que mejoran la eficiencia para catalizar sustratos no naturales, y hasta predecir *a priori* posibles residuos importantes para alterar la dinámica conformacional de las enzimas.

La química biomimética busca diseñar nuevos organocatalizadores imitando la función estructural del centro activo de las enzimas. El problema es que no entendemos totalmente cómo funcionan las enzimas, esto hace que el diseño y síntesis de complejos no biológicos con eficiencias comparables a las enzimas sea un reto. En el **Capítulo 7**, se muestra un mecanismo detallado para EUK-8, un complejo biomimético para la descomposición de agua oxigenada. La importancia de porporcionar una buena descripción del mecanismo dará una idea de que puntos de reacción son importantes para mejorar la reactividad. Por tanto, entender el mecanismo con las diferentes vías posibles y una buena descripción de los estados espín del metal en la reactividad proporcionará un conocimiento importante para diseñar racionalmente análogos de EUK-8.

La espectroscopia Mössbauer de ^{57}Fe es una técnica que proporciona información sobre la naturaleza química de sistemas de Fe, respecto a los estados de espín y oxidación. Dado que el hierro es un elemento esencial para todas las criaturas vivas y es incorporado en muchas enzimas y complejos organometálicos, en el **Capítulo 8**, se presenta una recalibración de un conjunto de cálculos para predecir parámetros de Mössbauer utilizando funcionales DFT (S12g y OPBE) que describen correctamente los estados de espín en sistemas de hierro. Como los espectros de Mössbauer no siempre son fáciles de analizar, estos nuevos parámetros de ajustes apoyan a los datos experimentales de Mössbauer y también ayudan a caracterizar especies de Fe-oxígeno y el respectivo estado de espín cuándo no hay estructuras de rayos-X disponibles.

La tesis está dividida en ocho capítulos, que incluyen: una introducción general, técnicas de caracterización, los objetivos principales de la tesis, tres capítulos relacionados con la dinámica conformacional de enzimas, un capítulo del mecanismo de un complejo biomimético, un capítulo de la caracterización de estructuras de Fe-oxígeno a través de

cálculos de parámetros de Mössbauer, y un capítulo final con las conclusiones principales obtenidas en los diferentes proyectos estudiados.

Chapter 1. Introduction

Biocatalysts

Since ancient times, living organisms have been used as catalysts for making wine, vinegar, bread, cheese, among others. Most of the biocatalyst utilization was related to food and drink production, but these biocatalysts had been used without any knowledge about them. The earliest applications came through empirical observations generation by generation. However, during the last four centuries scientists have been studying these living organisms shedding light that the catalysts responsible of accelerating the chemical reactions is what we call enzymes (from the Greek *énzyma* meaning "in yeast"). Enzymes are proteins speeding up the chemical reactions that are important to maintain and create life in all its many and varied forms on Earth.

1.1 Enzymatic catalysis

Enzymes are made of 20 different building blocks, called amino acids. Amino acids are small structures composed of a carboxylic group, an amino group, and a side-chain all attached to a central carbon atom. The side-chain varies in all the different building blocks and thus it determines the amino acid properties. There are different groups; hydrophobic amino acids with carbon rich side-chains; polar or hydrophilic amino acids; and charged amino acids interacting with oppositely charged building blocks or water molecules.

The linear sequence of amino acids encoded *via* the DNA and bound by a peptide bond is the primary structure (**Fig. 1.1**). The regions of amino acids grouping together confer different angles forming secondary structures organized into hydrogen bond structures such as right-handed α-helix and β-pleated sheets (**Fig. 1.1**). The α-helix is stabilized by a network of hydrogen bonds between the amine and carboxyl group of nearby amino acids. On the other hand, β-pleated sheets are made when hydrogen bonds stabilize two or more adjacent strands of the sheet. Finally, the specific arrangement of the secondary structure defines the folded tertiary structure (**Fig. 1.1**); its shape is determined by the characteristics of the amino acid, which in turn define their function and properties. The most common enzyme fold in the Protein Data Bank[1] (PDB, data base containing the protein X-ray structures) is the $(\beta\alpha)_8$ barrel enzymes.[2] Nevertheless, not all enzymes are active with a single folded tertiary structure. There are cases where two

or more folded tertiary structures are required for the functionality of the system,[3] here the individual protein of these multi-structures is called a subunit. This complex system with more than one protein is called the quaternary structure (**Fig. 1.1**).

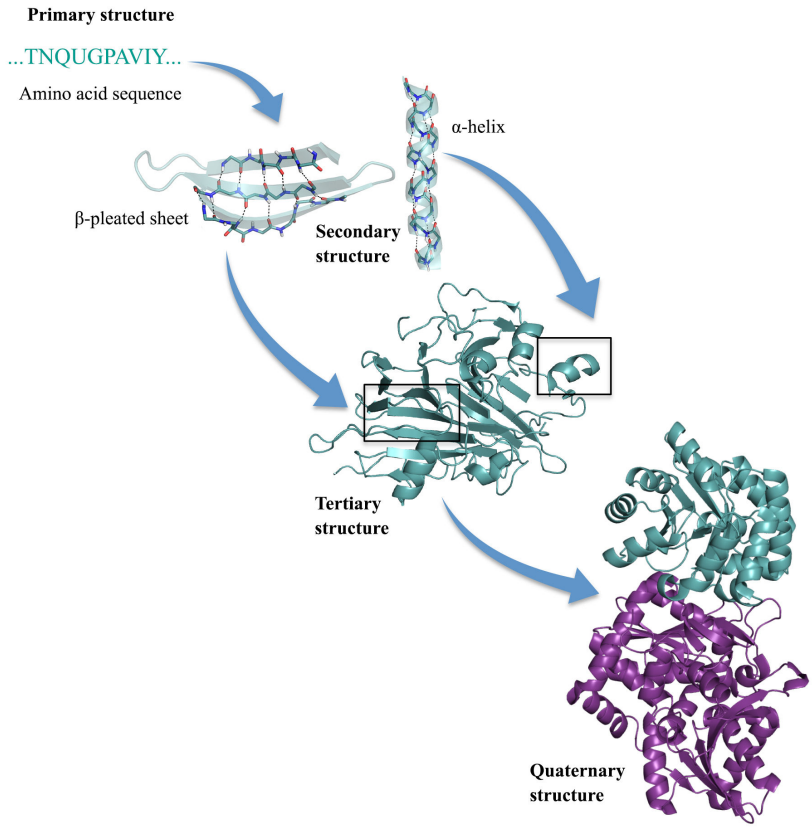


Figure 1.1 Representation of the different possible structures found in proteins. PDB 1JUH was used for the tertiary structure and PDB 1QOP was used for the quaternary structure.

The fold of the tertiary structure of enzymes induces a specific binding pocket for the substrate ligand. Inside the binding pocket the catalytic amino acids are present setting up the active site, which is the location where the chemical reaction takes place. The shape, size, and feature of the active site confer the enzyme the specificity towards some substrates and chemical reactions, thus it is commonly said that enzymes accommodate a small set of natural substrates, forming an enzyme-substrate (E-S) complex. This theory corresponds to the famous "*lock-and-key*" model proposed by Fischer in the 19th century.[4] Nevertheless, this definition is not accurate because many natural en-

zymes are also capable of accepting different substrates and even catalyze different reactions (i.e. promiscuous reactions).[5]

Enzymes have evolved to catalyze efficiently its natural reaction, however additional promiscuous side-reactions are also present in many proteins.[6] The ability of enzymes to visit different thermally accessible conformations plays a key role in enzyme promiscuity, regulation and inhibition (i.e. compounds blocking enzyme activity), but also in essential steps in enzyme catalysis such as substrate binding and product release.[7,8] The existence of a link between active site dynamics and catalysis of the chemical steps of the reaction has also long been debated. However, this observation is irrespective of the fact that enzymes by nature can adopt multiple conformations along the catalytic cycle.

As mentioned before, the primary function of enzymes is the acceleration of chemical reactions by many orders of magnitude to make them compatible with life. The rate acceleration is achieved by decreasing the activation barrier from substrate to product under mild conditions (**Fig. 1.2**). Thanks to billions of years of evolution, enzymes have acquired a high efficiency to operate usually under mild temperature and pressure, and in aqueous solution to produce high yield of products.

The extraordinary catalytic power arises from their precisely pre-organized active sites that properly position the catalytic residues for efficient transition-state (TS) stabilization, followed by breaking and formation of covalent bonds. However, many enzymes use non-protein cofactors to catalyze reactions, which can directly participate in the formation of the product but also can act as provider of the thermodynamic driving force

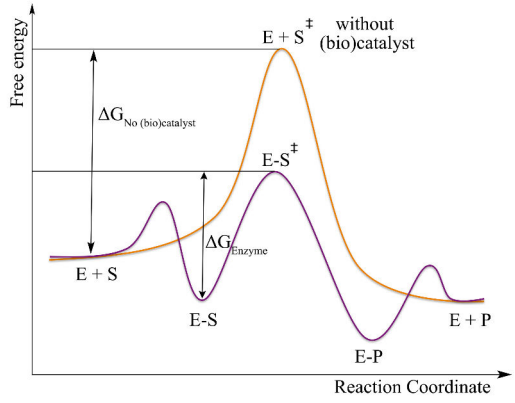


Figure 1.2 Representative energetic profile of simple reaction without and with (bio)catalyst. The enzyme (E), substrate (S), and product (P) are represented forming the corresponding enzyme-substrate (E-S) and the enzyme-product complex (E-P).

for catalysis. The different cofactors can be classified as: inorganic cofactors such as metal ions (*e.g.* Fe^{+2}, Cu^{+2}, Zn^{+2}), in so-called metalloenzymes; and organic cofactors such as nicotinamide adenine dinucleotide phosphate ($NADP^{+}$), flavin adenine dinucleotide (FAD^{+}), and adenosine triphosphate (ATP).[9]

1.1.1 Transition-state stabilization

During the last twenty years some key factors have been identified regarding enzyme catalysis.[10] Different proposals for transition-state stabilization have arisen, supporting this hypothesis by Pauli in 1946.[11] The general idea is that each enzyme becomes structurally complementary to the TS making it easier for the substrate to reach the right configuration to enhance the chemical reaction. Enzymes can follow different mechanisms to accelerate the chemical reaction:

A. Electrostatic catalysis is performed through effects in protein charges, making the active site a local and specific environment thus allowing ionic TS stabilization by nearby permanent dipoles. In contrast to bulk waters, the ionic groups are pre-organized by the 3D protein structure to further facilitate small reorganization energy for catalysis. It is worth to mention that the presence of metal/organic cofactors can also stabilize the TS with electrostatic interactions.[12] Warshel and coworkers showed that the main contribution for enzyme catalysis came through electrostatic stabilization of the TS.[13]

B. Acid/base catalysis is done by the stabilization of opposite charge interactions occurring between acid and base groups in enzymes, which depend on the pK_a (*i.e.* strength of the acidity in solution) values of the side-chain of the amino acids. This usually makes histidines good acid/base catalysts due to its pK_a around seven, allowing the correct ionizable state for catalysis. Acid/base catalysis includes reactions that are pH-dependent and perform abstraction and donation of protons from the substrate to groups in the enzyme, but also could be done by the presence of water molecules.[14]

C. Covalent catalysis uses a mechanism by which the reaction occurs through the formation of a covalent bond between the enzyme and the substrate(s). This covalent catalysis uses either electrophile or nucleophile groups to form the covalently bound

adduct. Hence, the reaction occurs thanks to the formation of intermediates and its subsequent breakdown making it easier to overcome the TS barrier to product formation.

D. Distortion catalysis consists of a slight destabilization of the substrate while still being at the ground state. The distortion could be induced by bond strain and polarization, E-S conformational change and/or electrostatic effect. The best version of this hypothesis is reported by the work of Jencks[15], called Circe effect. The hypothesis is that the substrate binding energy helps the catalytic residues to approach and reach the TS. Moreover, Bruice proposed that reacting atoms of substrate and enzymes must first come together inside a suitable reaction distance and angle to form the covalent bond, this term is called Near-Attack-Conformers (NACs), allowing the reaction to occur.[16]

E. Bringing two or more molecules together has an associated to entropic penalty. Enzymes perform the catalysis at the E-S complex with the catalytic groups being part of the same molecule as the substrate, avoiding the loss of entropy in the TS stabilization and thus it enhances the rate of the reaction. This hypothesis was reported by Jencks and coworkers,[17] however, Villà and coworkers suggested that those entropy contributions are much smaller than previously thought.[18]

F. Schowen, Cleland and Kreevoy suggested another stabilization of the TS through low-barrier hydrogen bonds (LBHBs).[9,19] This hypothesis suggests that when nitrogens or oxygens, with closer pKa values, are sharing the hydrogen, it can enhance the hydrogen bond for stabilizing the TS and for promoting the chemical reaction.

G. Despite the different possible stabilizations of the TS, it is clear that enzymes are dynamic structures. The important motions of enzymes could also affect the stabilization of the TS through a better pre-organization of the active site.[20] The effect of motions in catalysis is related to a revealed network of hydrogen bonds that decreases the activation free energy barrier of the hydride transfer.[21]

Overall it is important to know that the whole enzymatic mechanism from binding the substrate to form the E-S complex to the catalysis of the chemical reaction can occur through the modes mentioned above, with different energetic contributions of each

mode depending on the enzyme. All hypotheses described in this subsection have been postulated thanks to experimental studies and computational calculations, unraveling important aspects in enzymatic reactions.[13,22-24]

1.1.2 Enzymatic reaction mechanism

The diversity of enzymes on Earth gives us a broad range of possible chemical reactions. The active site topology with respect to stereochemistry, hydrophobicity, and electrostatic character determines which kind of substrates bind and undergo catalysis. The Enzyme Commission (EC) and ENZYME database classifies enzymatic reactions in different levels. The first level considers six general classes of reactions shown in **Table 1.1**. The other levels describe different characteristics and properties of the classes shown in **Table 1.1**. The last level is referred to substrate specificity of enzymes.[25,26] The mechanism followed by enzyme depends also on whether it uses a cofactor or not, and thus a different mechanism pathway for similar reactions can occur. I selected just two examples of interest for this thesis, highlighting important single and multi-step enzymatic reaction mechanisms that are possible in natural enzymes.

Table 1.1 Classification of enzyme reactions in the first level.

EC Class Number	Enzyme Class	Reaction
1	Oxidoreductases	Oxidation/reduction
2	Transferases	Group-transfer
3	Hydrolases	Hydrolytic cleavage
4	Lyases	Addition and elimination of functional groups to unsaturated or saturated carbon atoms
5	Isomerases	Rearrangement of atoms or group of atoms
6	Ligases	Joining of molecule of functional groups

A common reaction is the reversible catalysis of alcohol/aldehyde to the corresponding ketone by alcohol dehydrogenases (ADHs).[27] The oxidation-reduction reaction of these enzymes is as simple as the transfer or abstraction of a hydride. Many ADHs are zinc dependent enzymes requiring the $NADP^+$ and its protonated state (NADPH) cofactor as a source to abstract or deliver the hydride to the corresponding ketone (**Fig. 1.3**). ADHs are a good example to explain which is the function of the cofactor in the reac-

tion and also to rationalize the shape and topology of the binding pockets originating from the 3D structures. The reaction takes place when the substrate is bound to the zinc metal atom, while the NADPH is committed to deliver its *pro*-(R) hydride to a specific face of the ketone substrate. Hence, the reaction tends to form the stereospecific *(S)*-alcohol product. Nevertheless, Lamed and coworkers reported an interesting study showing the asymmetric reduction of variety of ketones catalyzing different stereospecific alcohols using ADHs. They suggested the existence of two binding pockets (big and small) in the active site of the enzyme. These two different active site pockets allow the substrate to accommodate the large alkyl chain in the big pocket and the small alkyl chain in the small one, thus it allows the proper orientation of the substrate showing the same face close to the cofactor for catalysis.[28] Enzymes evolved to adopt a 3D structure with a specific binding pocket topology for the acceptance of a particular substrate, being also highly efficient. However, the broad substrate scope of some ADHs also indicates the possibility that many enzymes are neither totally specific nor rigid, and thus enzymes are able to accept a range of different substrates.[6]

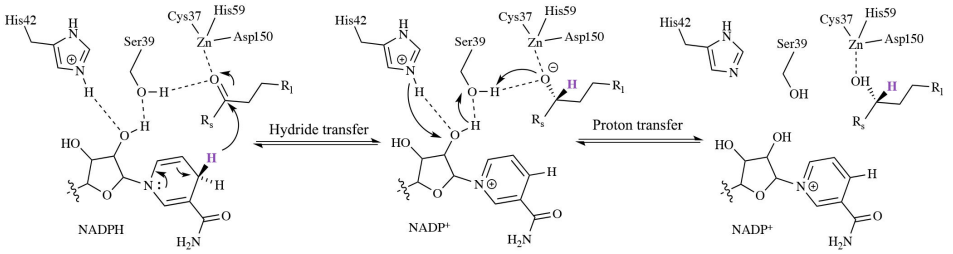


Figure 1.3 Mechanism of ketones/aldehyde reduction to alcohols in ADHs. The active site is composed of zinc, coordinated to Cys37, His59, and Asp150. His42 and Ser39 are proposed to support the deprotonation/protonation steps. The *pro*-(R) hydride atom is highlighted in purple. R_s is the region for the small alkyl chains of the ketone and the hydrogen of the aldehyde, while R_l region is for the larger alkyl chains.

In many enzymatic reactions multiple catalytic steps are needed. Those enzymatic reactions require the formation of a covalent intermediate between the substrate and the enzyme to fulfill the reaction. Generally, enzymes that utilize covalent catalysis evolved to perform this type of reaction in two main steps: the formation and breakdown of the covalent intermediate. The covalent adduct helps the TS to overcome the activation barrier of the reaction. As mentioned before, one of the mechanisms to promote the covalent intermediate formation is electrophilic catalysis. The covalent adduct is formed *via* a

cationic electrophile and an electron-rich region of the substrate compound. One example of using a covalent intermediate is the retro-aldolases (RA) computational designed enzymes. These enzymes follow a multi-step pathway involving enzyme-bound Schiff base intermediates similar to those of natural aldolases, converting the activated substrate by the lysine residue into the product(s). RA catalyzes the [4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone] (methodol) to 6-methoxy-2-naphthaldehyde and acetone (**Fig. 1.4**). Summarizing the complex multi-step mechanism: firstly, the methodol and a lysine catalytic residue form a Schiff base intermediate (i.e. imine compound). Thereafter, deprotonation occurs in the Schiff base releasing 6-methoxy-2-naphthaldehyde, followed by the formation again of another imine covalent intermediate for the final decomposition of the Schiff base expelling acetone, the final product (**Fig. 1.4**).

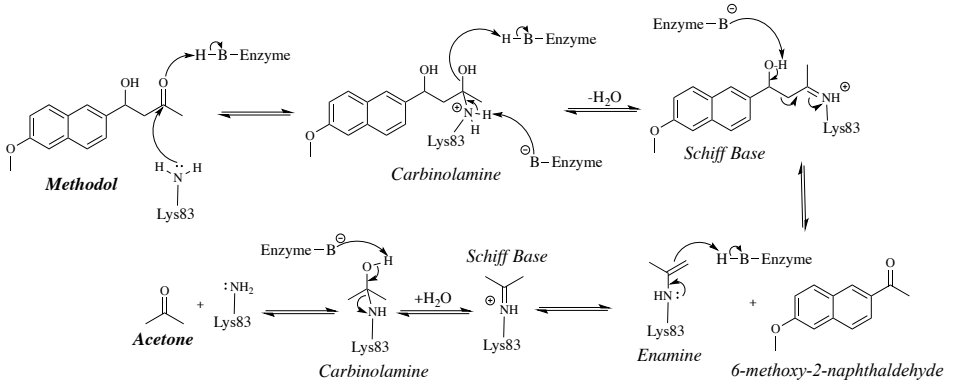


Figure 1.4 Representation of the multi-step mechanism reaction of decarboxylation of acetoacetate to acetone by acetoacetate decarboxylase. The lysine is supposed to form the covalent intermediate with the substrate, while other amino acids act as a base.

Although RA and many aldolases catalyze the reaction *via* a lysine residue, there are enzymes that follow the same mechanism in presence of cofactors. For example, aspartate aminotransferase and carbonic anhydrase use pyridoxal phosphate and Zn^{+2} as cofactors, respectively.[29,30] These cofactors help to overcome the energy barrier needed to complete the chemical reaction. This example shows the importance of the formation of covalent intermediates in some enzyme for breaking/forming of carbon-carbon bond reactions, which are needed to reach the activation barrier to catalyze the final product(s).

1.1.3 Kinetic parameters

The experimental characterization of kinetic parameters of enzyme-catalyzed reactions is the most common technique to elucidate enzyme mechanisms. Wurtz was the first to introduce the E-S complex idea in 1880,[31] later this idea was used by Brown to describe the general qualitative representation of enzymatic-catalyzed reactions in 1902, shown in equation 1.1.[32] Henri was the first to make a mathematical description just one year later in his doctoral thesis,[33] but Michaelis and Menten are more recognized for their contribution in 1913.[34] The reader should realize that in the majority of literature the equation is commonly referred as a Michaelis-Menten equation, but I will refer to it as the Henri-Michaelis-Menten (HMM) equation during this thesis as it is done nowadays.[35] Expression 1.1 represents an enzyme (E) in which a substrate (S) binds to form an E-S complex, followed by the release of the corresponding product (P) and recovering the free enzyme (E).

$$E + S \xleftrightarrow{K_S} ES \xrightarrow{k_2} P + E \tag{1.1}$$

The associated constant, K_S, is defined as the ratio of the forward and reverse reaction rate constant, while k_2 is the first-order rate constant of the reaction. Thus, in general equation 1.1 expresses a rapid equilibrium between enzyme and substrate and the E-S complex, followed by slower conversion of the E-S complex to the final product and enzyme. At high concentrations of substrate, the enzyme will be mostly present as E-S complex that can be rearranged as the expression 1.2.

$$[ES] = \frac{[E][S]}{K_S + [S]} \tag{1.2}$$

Such conditions make the velocity dependent on the rate of the chemical transformations to convert the E-S complex to product and recovering the free enzyme. The simplest case with just one chemical step is defined as the first-order rate constant. The expression can be described as equation 1.3 for the product formation in the single first-order rate constant.

$$v = k_2[S] \tag{1.3}$$

By combining equation 1.2 and 1.3, we obtain the famous HMM equation:

$$v = \frac{k_2[E][S]}{K_S + [S]} \tag{1.4}$$

All these derivations are dependent on the rapid equilibrium approach to enzyme reactions. Briggs and Haldane found that the equilibrium-binding of the HMM equation could be depicted by the steady-state approach.[36] The hypothesis suggests that the rate of E-S formation is equal to the rate of E-S decomposition into product and recovering the free enzyme. The condition is accomplished when the concentration of substrate is much greater than the enzyme. These conditions allow the simplification of the mathematical treatment of the kinetics as shown in equation 1.5.

$$v = \frac{V_{\max}[S]}{K_M + [S]} = \frac{V_{\max}}{1 + \frac{K_M}{[S]}} \tag{1.5}$$

This simplified equation for the steady state enzyme kinetics differs from the equilibrium expression derived by HMM equation, however, it is universally referred also as the HMM equation. The equation usually describes the velocity as a hyperbolic function with the maximum velocity, V_{max}, representing the maximum rate constant by the enzyme, k_{cat} *[E]*, at infinite concentration of substrate, *[S]*. The k_{cat} is a general rate constant involving more complex chemical steps in enzymatic reactions. The K_M term is used as an abbreviation for the kinetic constants (*Ks* and k_2, eq. 1.1) to make the equation less complicated. The latter two parameters (k_{cat} and K_M) are quite important to understand the enzymatic reactions used by experimental studies since its discovery.

K_M is the substrate concentration that provides the half-maximal velocity for an enzymatic reaction, and thus it can be also seen as the substrate concentration when half of the enzyme active sites are occupied by substrate molecules in the steady state. This pa-

rameter allows the relative measure of the substrate binding, but sometimes depending on the solution conditions these parameters can be altered by destabilization of the E-S complex making the ratio k_{cat}/K_M a better measurement for the substrate binding. On the other hand, the k_{cat} is normally referred as the turnover number (i.e. number of reaction processes per unit of time) for the enzyme. The meaning of k_{cat} is the maximal velocity that the enzymatic reaction can proceed with a fixed concentration of enzyme and infinite amount of substrate. In other words, it also relates to the chemical steps after substrate binding, thus normally it will provide information about the rate constant of the slowest step. However, for complicated mechanisms the k_{cat} may be a function of several rate constants.[37]

The catalytic efficiency of enzymes could be defined as the ratio k_{cat}/K_M, which is generally used to compare the efficiencies of different enzymes, for instance, to catalyze different substrates. It is also possible to be used as better binding affinity because the ratio k_{cat}/K_M captures the effects of different substrates on both kinetic constant parameters showed in the expression 1.1, regarding to E-S complex formation and the product release.

1.2 Designing new enzymes

Many chemical transformations of industrial interest do not have natural enzymes capable of catalyzing the reaction nor the proper binding of the desired industrial substrate for catalysis. For these reasons the ability to modify enzymes is highly important, and to do so efficiently is essential for understanding how an enzyme works. All this knowledge will give us the capacity to design enzymes with the ability to catalyze a scope of substrates, the enhancement towards a specific stereoselectivity, and/or design enzymes for new chemical reactions that are interesting for industrial purposes.

The substitution of amino acids in the enzyme sequence (i.e. mutations) is crucial to successfully modifying natural enzymes. The discovery of molecular biology methods revolutionized the engineering of enzyme fields thanks to the possibility to modify enzymes using an *in vitro* version of Darwinian evolution, mimicking the natural selection. The strategy is called directed evolution (DE),[38,39] which consists of enhancing or

altering the substrate specificity, catalytic activity, thermostability (i.e. stability towards high temperatures), or stereoselectivity for a selected enzyme. Initial iterative cycles of random amino acid changes are introduced into the enzyme and followed by selecting and screening the best variants with improved properties, and also further optimization can be done enhancing another specific property. This technique was recently laureated by the Nobel Prize in Chemistry 2018 to Francis H. Arnold for the directed evolution of enzymes.

Many improvements have been obtained related to protein engineering,[40] gene synthesis,[41] sequence analysis,[42] bioinformatic tools,[43] and high-throughput screening techniques[44] since DE appeared in molecular biology studies. For example, a dehalogenase enzyme was evolved for the synthesis of *lipitor*, in which by combining statistical analysis of the protein sequence and random mutagenesis the productivity was increased 4000-fold.[45] This hybrid method allowed the identification of beneficial mutations even in variants with low function ensuring the desired activity in a stepwise manner.

DE has become a powerful technique to (re)design enzymes with enhanced activity and has the ability to introduce mutations in the enzyme active site, nearby, and/or at distal positions from the active site. The latter mutations have been found to be important for increasing the enzyme catalytic activity (k_{cat}), as highlighted recently by Kell and coworkers.[46] However, the main drawback of DE is that little information can be extracted regarding how these mutations induced enhanced enzyme proficiency. Many studies are being done for the rationalization of how DE introduces the new mutations altering the enzyme catalytic activities.[47]

1.2.1 Strategies for (re)designing enzymes

The enhancement of the enzyme activity towards promiscuous reactions, a specific stereoselectivity, and even to catalyze non-natural chemical reactions, can be achieved with protocols based on: (1) choice of the residues to be mutated, (2) introduction of mutations in the enzyme, and (3) screening of the new variants for enhanced properties. Three main strategies are commonly used to (re)design enzymes: non-rational, rational, and semi-rational.[48]

The non-rational strategy mostly used is DE, capable of generating engineered variants of enzymes without the knowledge of either structure or the function of enzymes. The main drawback of DE is the necessity of a high-throughput screening methodology, which allows to identify the desired property under specific conditions for a large number of random mutations, forming large quantity of libraries.[49] Not all enzymes activities are manageable to implement this methodology. Despite that, huge advances have been made in non-natural substrate specificity, stability, product selectivity, among others.[50-52] For example, Arnold and coworkers reported for the first time a laboratory evolution of a cytochrome $P450_{BM3}$ monooxygenase enzyme, which naturally catalyzes the hydroxylation of fatty acids, into an efficient biocatalyst for the conversion of alkanes to alcohols. The best variant generated contained 11 mutations with a total library size screening of *ca.* 10^4, being highly active, soluble, and without the requirement of any additional electron transfer protein for catalysis.[53]

The rational strategy requires a detailed structural knowledge about the catalytic mechanism of the desired enzymes. Many computational strategies have been used, for instance, the multiple sequence and structure alignments (MSA).[54] This computational tool permits the analysis of sequences to identify functional residues of unknown proteins. Hence, it provides information about evolutionary relationships, functional sites, and correlated mutations.[54] Specifically, the HotSpot Wizard server combines information from sequences and 3D structure database to create a mutability map for the desired protein.[55] Similarly, 3DM database uses enzyme sequence and structure information from GenBank and PDB to further create alignments able to guide the design of mutations altering the protein properties and functionalities.[56] The combination of the huge number of crystal structures in PDB and the close link of 3D structure with functionality have prompted advances in homology modeling, offering to seek key residues near active sites and at domain interfaces to design smarter libraries.[57]

One step forward has been made through the *de novo* computational enzyme design. This strategy has the potential to create new biocatalysts for any chemical transformation, even non-natural reactions. Thus, allowing to unravel important knowledge between the 3D structure and its functionality.[57] One of the most successful strategies in

de novo design is the Houk and Baker computational *inside-out* methodology that combines structure prediction utilities in the Rosetta software (RosettaMatch[58] and RosettaDesign[47,59]) with Quantum Mechanics (QM) calculations (i.e. *theozyme,* ideal active site calculation with protein functional groups maximizing TS stabilization).[60] This protocol was successfully used to generate biocatalysts for Kemp elimination,[61] retro-aldol,[43] and Diels–Alder[62] reactions. This methodology for designing a new enzyme will be explained in more detail in the next subsection. RossettaMatch has the capability to search scaffolds in which the *theozyme* could be incorporated, while RosettaDesign protocol is used to optimize residues for TS binding affinity and catalysis.[63,64] Other strategies for matching the *theozyme* into a protein active site have also been developed in the enzyme design field.[65,66] More computational methods, for instance, Molecular Dynamics (MD) simulations have also been found to be key for ranking and identifying the best enzyme mutants.[67,68] Combined MD-Rosetta protocols were found to overcome some of the Rosetta sampling limitations, and MD is highly complementary to the Rosetta refinement.[68,69] Janssen and coworkers developed the catalytic selectivity by computational (CASCO) design framework that involves high-throughput MD to engineer enzyme stereoselectivity and replace most of the experimental screening assays.[70] Many additional protocols for (re)designing proteins with a large range of possibilities exist depending on the properties one wants to change.[48,57,71]

A recent example reported by Janssen and coworkers showed the computational redesign of an aspartase enzyme to a set of hydroamination biocatalysts.[72] Aspartase catalyzes the deamination of aspartate to fumarate and ammonia.[73] This enzyme is a highly selective enzyme without promiscuous reactions difficult to evolve by DE. However, they have been able to redesign the enzyme using *in silico* sampling of sequence and conformational space. Specifically, it was used for the synthesis of four important chiral β–amino acids that are both bioactive and building blocks for pharmaceutical compounds and natural products. The libraries generated were enclosing a restricted number of variants with new substituents to modify the binding pockets, which are not involved in the original enzyme to stabilize the β–carboxylate of the new substrates.[72]

Despite all the important advances in the field of (re)designing proteins, DE and rational strategies have their own limitations. DE has the limitation of the requirement of

some initial activity to use the high-throughput screening methodology for all enzymes, as mentioned before, and usually a huge number of libraries needs to be screened for obtaining better variants than the predecessor,[44] thus it implies highly economic cost to evolve an enzyme. The rational design has also some limitations. Notwithstanding the initial successes, for instance, computationally designed enzymes perform much less than natural and laboratory engineered enzymes. This observation reflects the extremely challenging task of enzyme design itself and indicates that rational computational enzyme design is still far from being a robust and systematic strategy for designing new biocatalysts useful for manufacturing and industrial purposes. The reasons for the low activity of computationally designed enzymes are highly debated, where one of reasons could be the lack of information in the topology of the active site and the non-inclusion of the conformational dynamics for selecting the proper scaffold to incorporate the *theozyme*.[74]

To further optimize the latter strategies, semi-rational approaches have been advantageous against these limitations. This strategy combines the benefits of DE and rational design to increase the efficiency of the (re)design process. The use of *in silico* methods allows rational mutations, thus reducing the complexity to smaller and smarter libraries. In turn, DE represents clearly the most effective strategy to enhance activity and identify the best variants in those focused libraries.[57,71] For instance, Munro, Reetz and coworkers reported a cytochrome $P450_{BM3}$ monooxygenase enzyme, able to control the regio- and diastereoselectivity of steroid hydroxylation introducing mutations nearby the binding pocket.[75] They used exploratory mutational experiments based on mutability landscapes (MLs) for choosing optimal amino acids for an iterative saturation mutagenesis (ISM, i.e. mutation of a specific residues for the other 19 possible amino acids) at a reduced set of residues lining the binding pocket (combinatorial active-site test, CASTing),[76-78] MD simulations were also used as extra guides to unravel the effects of mutations introduced in the enzyme. The combination of rational design and DE methodologies constitute a viable protocol to keep the screening effort as low as possible.[75] Another important example was recently done for improving the thermostability and catalytic efficiency of nitrile hydratase.[79] The semi-rational approached used was based on the pmut scan application of Rosetta (i.e. preliminary prediction of single point mutations) and MD simulations to select a small set of libraries. Subsequently, kinetic char-

acterization was performed for the best hits that showed better thermostability. Interestingly, three single mutations with enhancement in thermostability have shown also better catalytic efficiency than the wild-type enzyme. Thus, the semi-rational protocol is also an efficient strategy to select single point mutations with the capability to the enhancement of thermostability and catalytic efficiency, which are important features for industrial purposes. Moreover, these semi-rational approaches have also been used for designing new enzymes with non-natural chemical reactions, as mentioned before.[43,61,62]

Table 1.2 Summary of representative examples of (re)designed enzymes using non-rational, rational, and semi-rational strategies.

Enzyme	Enhance property	Strategy	Library size
Dehalogenase[45]	activity	non-rational	10^6
Cytochrome $P450_{BM3}$[53]	catalysis of new substrates	non-rational	10^4
Aspartase[72]	regio- and enantioselectivity hydroamination	rational	10^2
Cytochrome $P450_{BM3}$[75]	regio- and stereoselectivity for 16C steroids	semi-rational	10^3
Nitrile Hydratase[79]	thermostability and efficiency	semi-rational	10^3-10^2

1.2.2 Successfully designed enzymes for new reactions

One of the most representative cases of *de novo* computationally designed enzymes was the creation of Kemp eliminases, which catalyze a proton abstraction from a carbon by a base, a simple acid-base catalysis that is normally restricted by high energy barriers. The first *de novo* designs were generated using the *inside-out* protocol.[61,80] The first step of the protocol consists of selecting a set of catalytic residues, and subsequent QM calculations to create the *theozyme*.[60] Once the desired functional groups are chosen, the RossettaMatch algorithm is used to look for protein scaffolds. Afterwards, residues on the binding pocket are redesigned (RossettaDesign) to further increase the stabilization of the TS. The new designs were experimentally and computationally characterized to evaluate the activity towards the Kemp elimination.[61,68,81] Nevertheless, the *inside-out* Kemp eliminases exhibited quite low activities due to the lack of precision to generate the perfect arrangement of the active site for catalysis.[61] For this reason, designed en-

zymes that show good efficiency for the Kemp elimination reaction have been selected to be further optimized by DE, making use of iterative design protocols that yielded new variants containing 10–15 mutations and exhibiting higher activities.[61,81] It is worth to mention that other Kemp eliminases were designed using other methodologies such as cavitands and antibodies.[82,83]

Despite all the efforts to develop Kemp eliminases enzymes using the *inside-out* protocol, it has been shown that the most proficient Kemp eliminase reported so far was recently created by Kamerlin, Sanchez-Ruiz, and coworkers using an alternative approach. They showed that through a single hydrophobic-to-ionizable mutation an ancestral β-lactamase could be efficiently converted into a Kemp eliminase.[84] Some enzymes are able to have ionizable side-chains inside the hydrophobic cavity of enzymes, promoting stability, modulating the dynamics, and/or stabilizing residues near the active site or even the TS.[85,86] It was remarkable that with only 1–2 mutations the new variant was more efficient in accelerating the Kemp elimination than any of the previously designed Kemp eliminases. This efficiency was accomplished mainly due to the conformational flexibility of the ancestral enzyme. This study further confirms the importance of taking into account conformational dynamics for computationally designing novel enzyme functions.

A similar protocol was used for the computationally designed retro-aldolases (RA, **Fig. 1.5**).[43] Many artificial aldolases have been created by including a reactive lysine residue in a hydrophobic pocket.[43,87,88] The RA follows a multi-step pathway involving enzyme-bound Schiff base intermediates, converting the activated substrate by the lysine residue into the product. The most active enzyme designed, RA95, which catalyzes the methodol to 6-methoxy-2-naphthaldehyde and acetone (mechanism explained in **Fig. 1.4**), revealed only a modest *ca.* 10^4-fold rate acceleration.[87] Further DE engineering of the computationally designed RA95 boosted its specific activity by more than $>10^9$-fold in the final RA variant obtained, RA95.5-8F.[89,90] The improvement was achieved manly by increasing the k_{cat}. The evolution path towards the highly active RA95.5-8F showed improvements involving rearrangement of the active site[89,90] and a shift in the rate-limiting step.[91] The high efficiency of RA95.5-8F compared to class I aldolases was attributed to the catalytic tetrad that emerged along the DE pathway,

highlighting the sophisticated H-bond network of the four active site residues arising one by one over the course of laboratory evolution. Nevertheless, the capacity of the catalytic lysine to form Schiff base intermediate has increased the scope of many carbon-carbon bond formation reactions accelerated by the RA95 variants, giving promiscuous behavior towards Michael additions,[92,93] Knoevenagel,[94] and Henry reactions.[95]

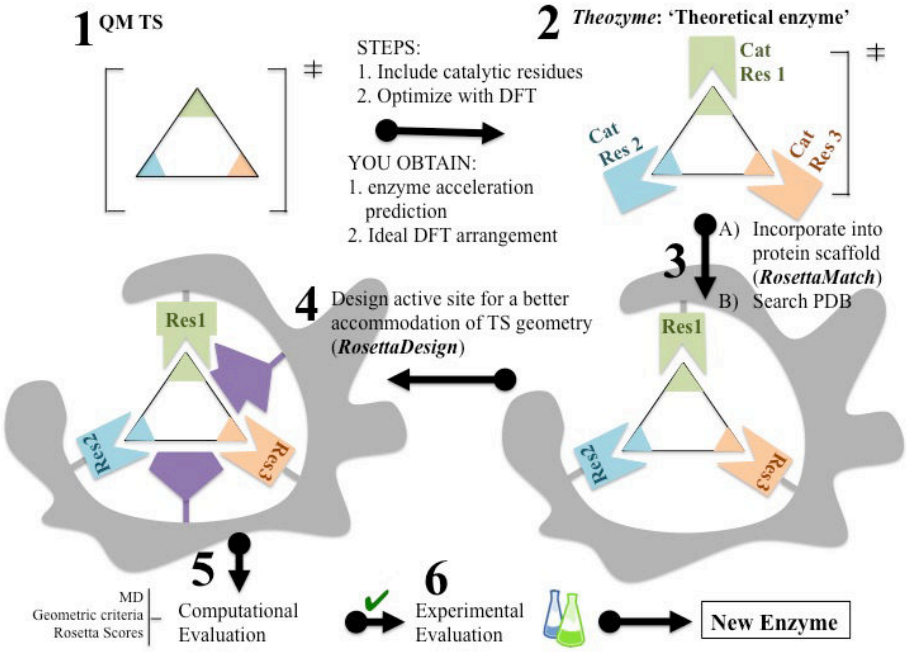


Figure 1.5 General summary of the important points for *de novo* design using *inside-out* protocol.

1.3 Conformational Dynamics

Enzymes are dynamic systems by nature, and can undergo conformational changes in a variety of timescales (**Fig. 1.6**).[96] On the shortest timescale are bond vibrations (10–100 femtoseconds) and side-chain conformational changes (picoseconds to millisecond), while the loop (i.e. disordered secondary structures) motions – which are often key for catalysis, substrate binding and product release – occur on larger timescales (nanosecond to millisecond). Slow domain motions and allosteric transitions (i.e. regulation of enzyme function by distal positions) can take place on the longest timescales (milliseconds to seconds).[97] All this variety of motions can be present before or after the chemical steps, and most importantly those conformational changes can be the rate-limiting

(i.e. the slowest step of the reaction) in some natural and/or laboratory engineered enzymes.[98]

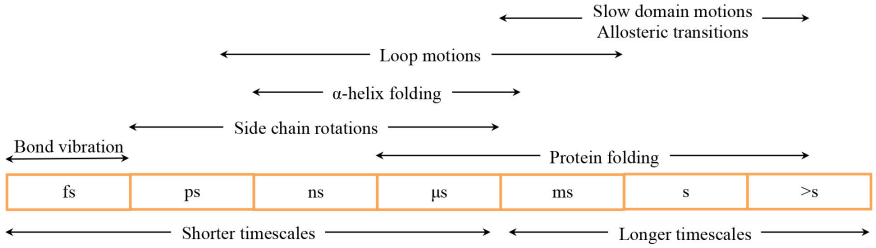


Figure 1.6 Timescales of different types of proteins motions.

The different conformational states that the enzyme can adopt in solution can be indirectly characterized through experimental methods such as X-ray crystallography and nuclear magnetic resonance (NMR) techniques. Both spectroscopic techniques have revolutionized structural biology by obtaining the 3D structure of these complex systems ever since the first structure was obtained through X-ray analysis in the late 1950s.[99] The amount of protein structures crystallized has increased exponentially over the years as shown in **Fig. 1.7**.

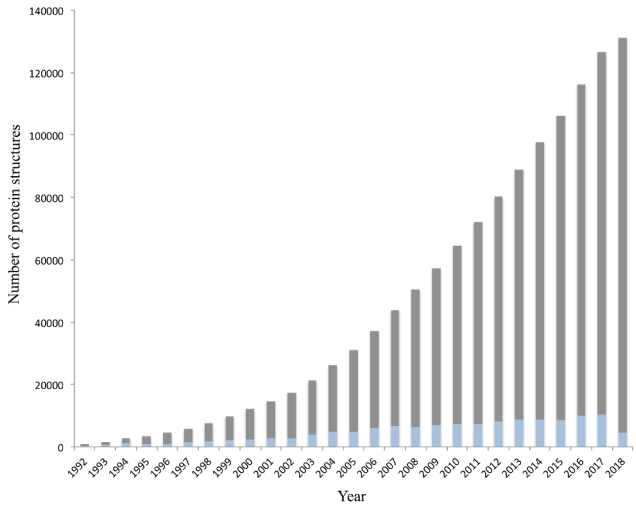


Figure 1.7 Growth of crystal structures available since 1992, in light blue crystal structures obtained during that year and in grey the accumulated number of PDB. Data obtained from PDB.[1]

Thanks to the advanced progression in X-ray and NMR techniques, collective or slow motions in proteins can be analyzed. X-ray spectroscopy has been applied to determine the conformational energy landscape of yeast ribosome, together with RNA translocation as a function of time.[100] NMR techniques, although lacking the resolution of X-ray crystallography, provide structural together with kinetic data in a timescale range from picoseconds to seconds.[101,102] Furthermore, both techniques have provided further evidence supporting the existence of enzymes as an ensemble of thermally accessible conformations,[103,104] as observed with computational studies based on MD simulations.[105] All these studies highlight the important role of the enzyme conformational dynamics for its function. It is also worth to mention that biophysical techniques such as fluorescence, circular dichroism, Raman spectroscopy, among others also give kinetic information complementary to these structural methods.[106,107]

Interactions after inhibition or substrate binding induce a redistribution of the populations in the accessible conformational states (i.e. population shift). [108] This is in accordance to the induced fit model (A in **Fig. 1.8**),[109] in which the binding of the substrate induces a conformational change from the *apo* (i.e. enzyme without the substrate in the active site) to the E-S complex state conformation of the enzyme. On the contrary, in the conformational selection concept the binding process does not induce a conformational change, but rather a redistribution of the populations of the conformational states that already exist occurred (B in **Fig. 1.8**). This hypothesis is the opposite to the induced fit model, which overlooked the fact that in solution the enzyme can pre-exist in multiple conformations.[110] Hence, there is an important question that is always open to discussion. Is it the substrate that induces the conformational change of the enzyme or is it the own enzyme that can sample different conformations in the *apo* state for the acceptance of a particular substrate? The population shift concept originated from the Monod–Wyman–Changeux model of allostery[111] has become more popular than the induced fit model in recent years (**Fig. 1.9**).

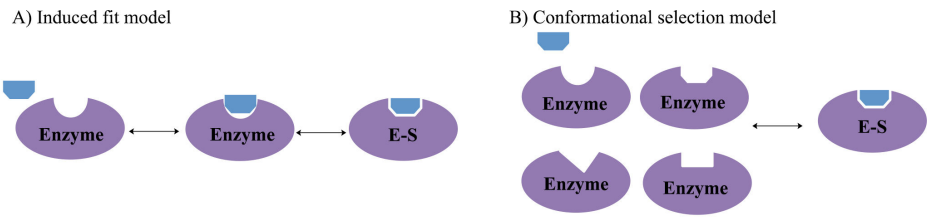


Figure 1.8 A) Representation of the Koshland induce fit model from *apo* to ES conformation of the enzyme. B) Representation conformational selection model of *apo* to E-S complex states. Enzyme in purple, substrate in blue, and the E-S is the enzyme-substrate complex state.

One recent example is reported by Kovermann and coworkers, providing evidence for a conformational selection pathway in the adenylate kinase (AdK) enzyme.[112] They have shown by X-ray crystallography how AdK adopts an open conformation in absence of ligands, whereas a closed conformation is required for catalysis. According to the conformational selection model, this high in energy closed conformational state should also be visited in the absence of ligand, but with a lower frequency. By introducing a disulfide bond, they succeeded in arresting AdK in a closed conformation in the *apo* state. The X-ray structure provided a key proof of the closed conformation of the enzyme being also sampled in the absence of any ligand, highlighting that high energy functionally important states are visited even in the *apo* state. (**Fig. 1.9**)

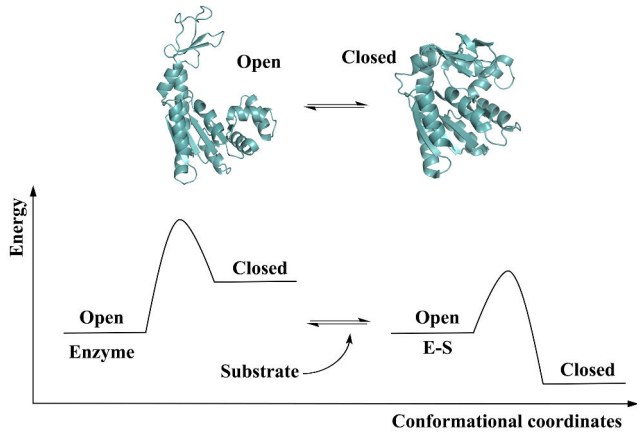


Figure 1.9 At the bottom left, Monod-Wyman-Changeux model for the free enzyme. At the bottom right, the E-S complex representation. Open (PDB 4AKE) and closed (PDB 1AKE) conformations of AdK enzyme are shown top as representative clear conformations.

1.3.1 The effect of mutations on the conformational landscape

External loops located in the periphery of the active site play an important role in enzyme catalysis,[113] but also in specificity[114] and stability.[115] Loops are a diverse class of secondary structures containing distinct turns and disordered random coils, which generally connect the main secondary structures, α-helices and β-pleated strands.[116] This type of secondary structures is quite flexible, and is often associated with sequence alternation across homologs, thus contributing to the diversification in terms of function within different families and superfamilies.[117] Many enzymes have important loops for catalysis, allowing different conformational changes that play a role in the catalytic cycle. For example, Palzkill and coworkers have nicely demonstrated through X-ray structures how a triple mutant in the active site loop of TEM-1 β-lactamase induced a large conformation change, thereby exhibiting a 400-fold increase in catalytic efficiency for a substrate that is not efficiently catalyzed by the wild-type natural enzyme.[118]

The effect of introducing mutations to the enzyme sequence for their evolution towards new functions and novel substrate scope has a high similarity to substrate binding. Substrate binding induces a shift in the populations of the pre-existing conformational states, but in the particular case of enzyme evolution, this population shift favors the catalytically competent conformational states for the new target reaction. This was recently demonstrated by Tokuriki and Jackson through a complete collection of X-ray structures.[103] They showed a gradual population shift of pre-existing conformational states to change in function from a phosphotriesterase into an arylesterase along the evolutionary pathway. Their study established that minor states that conferred the natural enzyme some arylesterase activity were gradually stabilized to become major states in the evolved arylesterases.[103] A similar finding was reported by Jackson in evaluating how ancestral binding proteins evolved into specialist binders.[119] An ancestral arginine-binding protein was crystallized in complex with L-arginine and L-glutamine revealing that the promiscuous binding of L-glutamine was possible due to alternative conformational states. These alternative conformational states were further populated along the evolution process to produce the contemporary L-glutamine specific protein binders. These studies support the idea underlying the principle that enzyme evolution lies in the population shift of the conformational states that pre-exist in solution.

Many examples have been provided in the literature demonstrating how mutations located at remote positions from the active site can have a large impact on the catalytic activity of the enzyme.[46] For instance, the single mutation of a tryptophan at >18Å away from the catalytic triad triggered significance shift towards inactive conformations states, losing the catalytic efficiency of the allosterically regulated human monoacylglycerol lipase (hMGL).[120] This study nicely demonstrated by experimental and computational analysis how a mutation of tryptophan by leucine decreased the efficiency of the enzyme 10^5-fold compared to the wild-type. NMR data and MD simulations showed the presence of important conformational changes, thus increasing the conformational flexibility and limiting the accessibility of the binding pocket with respect to the wild-type. They provided clear evidence that key residues far from the active site can induce a huge impact for the destabilization of important interactions of the enzyme, modifying the conformational dynamics and the allosteric communication.[120] The role of some distal mutations in recapitulating the allosteric regulation have been also demonstrated by an acylcarrier protein on the acyltransferase enzyme LovD by means of MD simulations.[105] The analysis of the conformational dynamics of the stand-alone LovD enzyme along the evolutionary pathway indicated that the introduced mutations induced a gradual population of the catalytically active conformational states. The key role exerted by remote mutations on the active site of the enzyme suggests that allostery might be an intrinsic characteristic of enzymes,[121] which might be exploited for designing enzymes.

Mimicking nature

Humanity has used Nature as inspiration to solve problems and to go one step forward in science and technology since the beginning. A clear old example is the observation of birds that prompted humans to create planes. Indeed, humans have used the experimentation to overcome challenges that affect human well-being. Specifically, chemists have used the shape and composition of metalloenzymes active site for the synthesis of organometallic complexes. The reason to mimic the function of enzymes could be related to understand and learn from Nature, how over billions of years of enzyme evolution has achieved their notably efficiency and efficacy. However, as said by Nobel Laureate Jean-Marie Lehn, "*the challenge for chemistry lies in the development of abiotic, non-*

natural systems, figments of the imagination of the chemist, displaying desired structural features and carrying out functions other than those present in biology with (at least) comparable efficiency and selectivity. Subject only to the constraints one chooses to impose and not to those of the living organism, abiotic chemistry is free to invent new substances and processes. The field of chemistry is indeed broader than that of the systems actually realized in Nature." [122,123]

1.4 Biomimetic catalysts

Biomimetic chemistry seeks to design novel efficient metal-base mimicking the structure-function from the enzyme's active site. The problem is that we still do not fully understand how enzymes work, which makes the design and synthesis of non-biological complexes as efficient as biocatalysts a challenging task. We can only be sure that we truly understand Nature's catalytic principles if we are able to design and synthetize new catalysts whose features are up to the standard of enzymes. The development of biomimetic catalysis therefore involves both an aim to gain knowledge and a test to achieve similar capabilities as biocatalysts. In this section, I am going to describe different examples of biomimetic complexes designed in the last century.[122]

Chemists have designed a large scope of different complexes mimicking enzymes. The clearest example relies on iron containing enzymes, where two main groups of enzymes are present: heme and non-heme iron containing enzymes. Although there are many other iron containing enzymes (see ref. 124 for more details), they will not be discussed in this thesis.

Heme-containing enzymes, normally cytochrome P450 enzymes, have an iron atom coordinated by the tetrapyrrole porphyrin with four nitrogen donors connected to a conjugated ring system. The axial position on the proximal site is normally occupied by histidine or cysteine, and the sixth coordination site is accessible for the corresponding substrate (**Fig. 1.10**). The conjugated porphyrin ligand is able to stabilize highly oxidized intermediates up to Fe^{IV} during the catalytic cycle of cytochrome P450. This active Fe^{IV}-oxo intermediate, so-called compound I (CpdI), is responsible for catalyzing alkane hydroxylation reactions.[125] However, there exist more intermediates during the

catalytic cycle that have been studied.[126,127] Many studies have been focused to understand the mechanism of heme proteins.[128-131] The first heme complex was synthesized in 1981 by Groves and coworkers,[132] the biomimetic complex was a competent oxidant in alkane hydroxylations and also in olefin epoxidations.[133,134] Furthermore, porphyrins involving other metals, such as Mn^{III}, have also been synthesized. These porphyrins using Mn^{III} metal atom are able to oxidize the indole in presence of hydrogen peroxide.[135]

Non-heme iron enzymes have also been extensively studied.[136,137] In this case, the iron center is normally coordinated by amino acid residues (or mimic thereof), water, and oxo-anions. The usual amino acids coordinated to the iron are histidine, cysteine, aspartate and glutamate. The remaining coordination sites are mostly occupied by water. Despite the diversity of this non-heme coordination, there is one frequently found motif in which the iron ion is coordinated by two histidines and one carboxylate (glutamate or aspartate), forming a triangular side at the octahedral coordination geometry (**Fig. 1.10**). This type of non-heme coordination is more flexible than the heme coordination, providing neighboring coordination sides for substrates and cofactors, which open a variety of additional reaction paths. The possible reactions catalyzed by non-heme iron enzymes covers oxidative cyclization, mono- and dioxygenation, hydroperoxidations and epoxidations.[138] The enzymatic cycle of these enzymes uses an oxygen molecule to generate a highly reactive ferryl intermediate, which then can act as hydrogen atom abstracting species.[139]

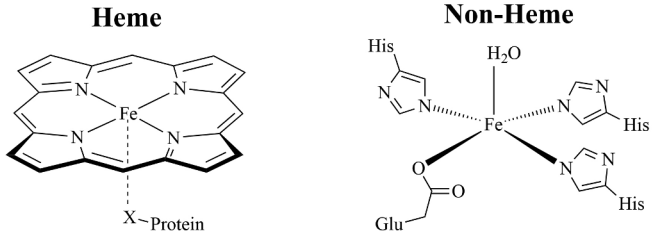


Figure 1.10 Heme and non-heme active site of a cytochrome c and dioxygenase enzyme respectively. Heme axial ligand (X) can be coordinated by S-Cys, S-Met, N-His, O-Tyr, among others.

The aim of biomimetic chemistry is to simulate the reactivity of these enzymes regarding reaction rates and substrate specificity with high efficiency or even overcome the capabilities of the natural enzyme. However, the enzyme 3D structure protects the active site, which otherwise might be vulnerable if exposed to the outside. In contrast, it is

difficult to protect synthetic compounds from undesired side-reactions. A comparison to organometallic chemistry shows that such complexes are very sensitive and demand a careful synthesis and storage. The ligands that are used in the synthesis have to fulfill certain conditions such as be compatible with the different oxidation states of the metal site. Nonetheless, the amount of knowledge obtained by understanding how heme and non-heme enzymes work have allowed the design of an enormous libraries of multidentate ligands (**Fig. 1.11**).[140] These ligands have been used to enhance the binding to the metal while introducing certain rigidity. The presence of steric properties of the ligands is important to protect the metal and create an asymmetric pocket.[141] Most of these ligands bind to the iron atom via nitrogen donor atoms, which do not correspond to the large part of oxygen donors found in the non-heme coordination sphere. However, the number of functional model complexes synthesized from polydentate nitrogen ligands with different topologies demonstrates the capability of these ligands, supporting highly oxidized metal centers with promiscuity towards different possible reactions.[142,143]

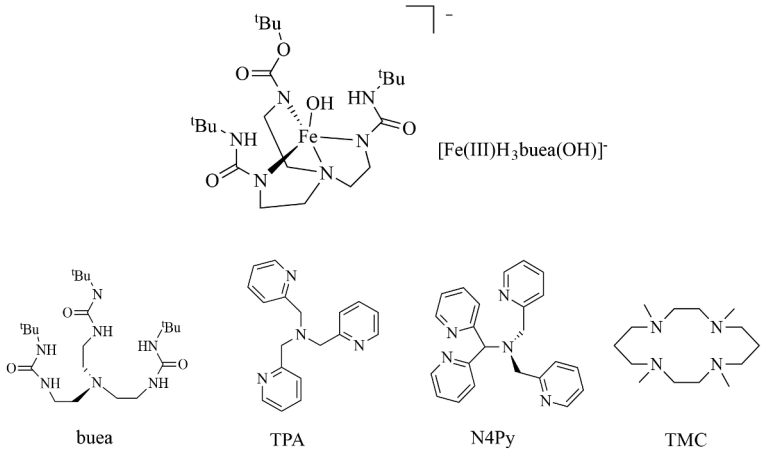


Figure 1.11 The ligands showed at the bottom are 1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane (TMC), N,N-bis(2-pyridylmethyl)-N-bis(2-pyridil)methylamine (N4Py), tris(5-phenylpyrrol-2-ylmethyl)amine (TPA), and 1,1,1-tris[(N'tert-butylureaylato)-N-ethyl]aminato (buea). Iron coordinated to the buea ligand on the top and 4 different kind of ligands synthesized to biomimic non-heme enzyme on the bottom.

Another group of enzymes used are catalases, present in many aerobic systems catalyzing the decomposition of toxic hydrogen peroxide into water and oxygen.[144] The crystal structures of some bacteria showed an active site containing two manganese metal ions triply bridged by a 1,3-carboxylate from a glutamate amino acid and two sol-

vent single-atoms bridges. Overall, both manganeses confer an octahedral coordination geometry composed by histidines, glutamates and water solvent molecules (**Fig. 1.12**).[145] Furthermore, biochemical and spectroscopic studies have revealed that the oxidation states of both manganese metals alternate between (II) and (III), while dismutation of H_2O_2 takes place. Further knowledge has been unraveled through the proposal reaction mechanism reported by computational chemistry calculations.[146,147]

Starting from the reduced form with an oxidation state of Mn^{II} at both metal centers, the initial complex is formed by the insertion of H_2O_2 thereby replacing a bridging water molecule, and forming a bridging peroxide intermediate. This position allows the activation of the oxygen-oxygen bond of the bound peroxide, facilitating the transfer of an electron from one of the Mn^{II} to H_2O_2 to further weaken the oxygen-oxygen bond. The bond-breaking step forms a hydroxyl free radical bound to a Mn^{II} changing the oxidation state of both manganese to Mn^{III} with the dissociation of a water molecule. The free radical formed in the last step is trapped by the metal complex, avoiding the Fenton reaction (i.e. formation of toxic free radicals). The rate-limiting step found is the oxidative reaction of the first hydrogen peroxide molecule.[147]

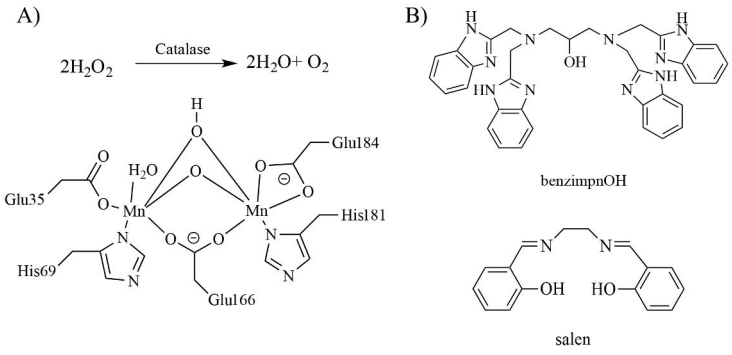


Figure 1.12 A) Catalase reaction and active site of a manganese catalase from *Lactobacillus plantarum* (PDB 1JKU). B) A biomimetic ligand to form a binucleating complex-N,N,N',N'-tetrakis(2-methylenebenzimidazolyl)-1,3-diaminopropan-2-ol (benzimpnOH) and a biomimetic ligand to form non-binucleating complex-1,2-bis(salicylideneamino)ethane (salen).

Biomimetic compounds have provided valuable structural information on the functional role of the bridging ligands that play a critical role in the hydrogen peroxide decomposition mechanism. A large number of functional biomimetic manganese compounds have been designed covering from binuclear compounds forming different type

of bridged structures to mononuclear complexes, both with a huge variety of possible ligands (**Fig. 1.12**).[148] The extensive research done for these complexes sheds light on different important features that are relevant to understand the dismutation of H_2O_2.

The catalytic rates are affected not only by the redox potential (i.e. tendency to acquire or release electrons) but also by the presence of an intramolecular base supporting the deprotonation of hydrogen peroxide. Moreover, important factors as the enhanced affinity towards the substrate have been found to happen due to the vacant position on the manganese metal ion and the oxidation change during the catalysis.[148] Indeed, studies suggested that mononuclear biomimetic complexes, specifically manganese salen compounds, are non-selective antioxidants able to be used as scavenger of hydrogen peroxide. The main drawback is the low stability of these compounds under the conditions needed for catalase reactions.[149] Interestingly, these compounds have been proposed to follow a mechanism involving Mn^V-oxo species[150,151] or through the formation of dimeric structures in solution,[152] totally different from the native enzymes.

There are many examples of biomimetic chemistry in the literature enclosing a huge range of possibilities.[148,153-155] These biomimetic complexes involve different metals (Cr, Mg, Zn,...) in similar structures as the examples mentioned above, in self-assembled structures, in nanocomposites, among others.[122] Despite all the studies related to this field with great and important advances catalyzing different reactions, understanding the reaction path mechanism and going beyond the limits of biocatalysts has been quite challenging. This fact underlies the difficult task to control and mimic systems that have evolved through billions of years to reach such efficiencies under mild environmental conditions.

1.5 Spin-states in transition-metal atoms

Biomimetic catalysts usually have at least one transition metal atom in its structure, which combined with the all-possible first-row transition metals on the periodic table gives a huge diversity of complexes. This diversity is expressed with a multitude of different reactivities and properties as we have seen before. The different reactivity arises from a wide range of ligands and the flexibility of the coordination around the metal.

The coordination in different environments leads to different distributions of electrons over the *d*-orbitals, invoking different oxidation and spin states (**Fig. 1.13**).[156]

Depending on the oxidation state of the metal and the coordination sphere, often several energetically accessible spin states are possible. Changes in those spin states come from different metal−ligand bond lengths and also in diverged coordination geometries. The metal oxidation states are usually well understood and controllable, however, spin states are less explored. In transition-metal complexes, the electronic spin can manifest itself as unpaired electrons at the metal, but further complexities arise in the presence of redox non-innocent ligands, thus making ferromagnetic or antiferromagnetic interactions with the spin of the metal center.[157] It is worth to mention that these properties are not only related to biomimetic catalysts, but also to all biocatalysts involving a metal, as are metalloenzymes.

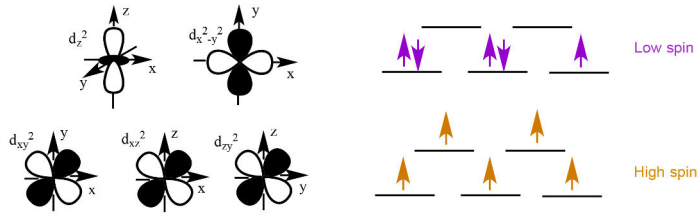


Figure 1.13 On the left, the transition metal *d* orbitals in an octahedral environment. On the right, the difference between low spin and high spin configuration.

The influence of spin states has an important role in the reactivity concerning reaction in metalloenzymes of cytochrome P450, in metal-oxo complexes, in spin-crossover compounds and even in catalysis mediated by organometallic compounds where different reactions can take place via different spin states.[158,159] The spin-forbidden reactions consist of the change of spin state from the reactant complex to product, thus producing increased activation energies compared to the same reaction where the spin state remains equal at the beginning and the end of the chemical reaction. However, the spin-forbidden reactions can be thermodynamically favored by the intermediate of the metal ions in low oxidation states as can be Fe^{II} and Mn^{II}. Those metals in low oxidation state are capable, for instance in the case of dioxygenases reactions to form peroxide species able to oxidize organic functionalities.[157]

Multiple spin states are usually the result of the different possibilities of accommodating electrons in *d*-orbitals, but sometimes ligands are redox non-innocent and can either offer ligand-based orbitals to accommodate electrons from the metal center or also transfer electrons to *d*-orbitals. P450 constitutes a clear example for this situation with an Fe^{IV}=O center and a porphyrin radical ligand. Occupation of the *d*-orbitals of the iron center and the porphyrin-based radical with the five electrons produces S=1/2 (one unpaired electron) and S=3/2 (3 unpaired electrons) systems, close in energy, exhibiting important differences in their reactivity.[160]

By combining experimental and computational studies it is possible to get knowledge on the role of spin state in reactivity for biomimetic catalysts and metalloenzymes, but this is often difficult to predict, interpret and/or to understand. Only through the utilization of a variety of techniques one can be certain that the interpretation of experimental and/or computational results are admissible to understand the reactivity together with its proposed reaction path. Many examples of this are present in the literature on oxidation states and/or spin states.[161,162] A variety of techniques have been involved in the reassignment of oxidation and spin states on iron catalytic cycles, such as X-ray, electron paramagnetic resonance (EPR), infrared (IR), ultraviolet-visible (UV-vis), Mössbauer spectroscopy techniques, and quantum chemistry.[161,163,164]

Chapter 2. Methodologies

Methods for *in silico* characterization

Computational techniques offer a broad range of possibilities to analyze chemistry at the atomistic level. Thus, depending on the process to study, one can select different techniques to describe the system. Those techniques used in this thesis are classified in two main categories: quantum mechanics (QM), based on density functional theory; and molecular mechanics (MM), based on force fields (FFs) and molecular dynamics (MD). The techniques used in this thesis differ in the level of accuracy and system size (**Fig. 2.1**). Moreover, the manner to describe the protein and/or organometallic complexes in both categories is completely different.

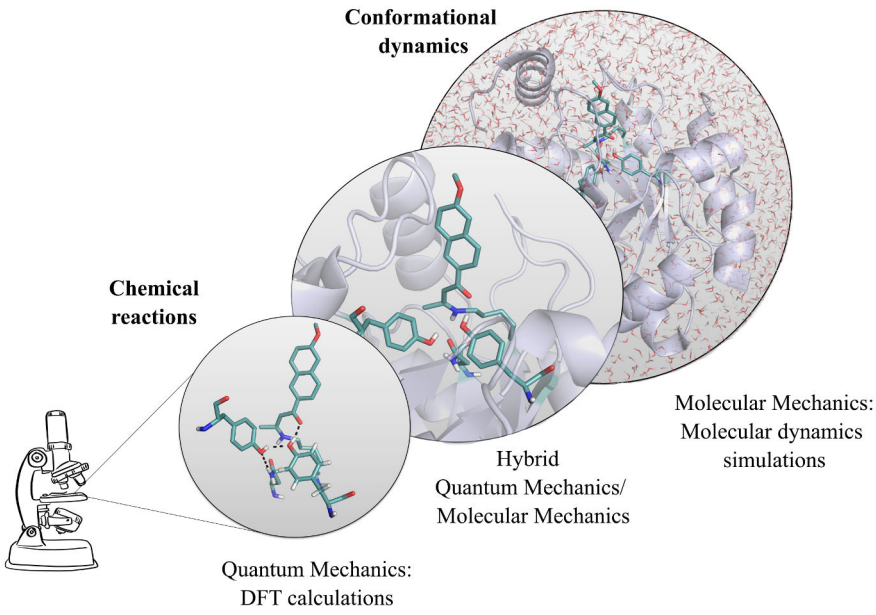


Figure 2.1 Representation of most used computational tools for evaluating chemical reactions and conformational dynamics of different systems.

It is worth to mention that exist more techniques to describe different processes that are not mentioned nor used in this thesis.[48]

2.1 Density functional theory

Density functional theory (DFT) is a computational quantum mechanical modeling approach, alternative to conventional *ab initio* methods, based on using electron density functions of the system to describe its corresponding properties.

The basis of DFT is formed by the Hohenberg-Kohn theorems: the 1st theorem considers that the electronic energy and any other property of the ground state can be calculated as a functional of the exact electron density of the system; and the 2nd theorem states that the exact electron density of the system is the one that minimizes the electronic energy of the ground state, following the variation principle.

$$E[\rho] \geq E[\rho_{exact}] = E_{\min} \tag{2.1}$$

A functional is any mathematical expression that is applied to the electron density in the case of DFT, to obtain a number. Since the electron density function ($\rho(r)$) only depends on three spatial coordinates of the system (x, y, and z), the DFT approach is able to reduce the computational cost of quantum many-electron calculations with respect to wavefunction-based methods. Hartree-fock (HF) and post-HF methods depend on $3N$ variable ($4N$ when taking into account spin), where N is the number of electrons. The final step in the development of DFT is to design the shape of the functional that connects the electron density of the system with its energy. Nevertheless, developments towards that end have not been an easy task, and are not being fully solved.

The total energy functional is divided into three terms: kinetic energy $T[\rho]$, attraction between the nuclei and electrons $E_{ne}[\rho]$, and electron-electron interaction $E_{ee}[\rho]$, while the nuclear-nuclear repulsion is assumed constant within the Born-Oppenheimer approximation. Furthermore, the $E_{ee}[\rho]$ term may be divided into Coulomb ($J[\rho]$) and exchange ($K[\rho]$) parts, implicitly including correlation energy all the terms.[165]

$$E[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.2}$$

The basic general idea of DFT methods in computational chemistry came in 1965, when Kohn and Sham found a practical method to calculate ρ and $E[\rho]$.[166] They introduced orbitals and defined the kinetic functional under the molecular-orbital approximation. Since the exact density matrix is not known, the approximate density is written as shown in equation 2.3 in terms of a set of auxiliary one-electron function, the Kohn-Sham (KS) orbitals, ϕ_i.

$$\rho(r) = \sum_{i=1}^{N} \left|\phi_i(r)\right|^2 \tag{2.3}$$

Kohn and Sham considered a fictitious reference system of non-interacting electron with the same electron density as the real system, moving through a local effective external potential, called the Kohn-Sham potential ($V_{eff}(r)$). The KS model is closely related to the HF method in the sense that HF uses an independent-electron wavefunction (Slater-determinant), whose electrons interact through the HF potential. Therefore, the energy functional can be described as:

$$\begin{aligned} E[\rho] = &-\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\int \phi_i^*(r)\nabla^2\phi_i(r)dr - \sum_{I=1}^{N}\int \frac{Z_i e^2}{4\pi\varepsilon_0\left|R_I - r\right|}\rho(r)\,dr \\ &+\frac{1}{2}\int\int \frac{\rho(r)\rho(r')e^2}{4\pi\varepsilon_0\left|r - r'\right|}drdr' + E_{xc}[\rho] \end{aligned} \tag{2.4}$$

Here the kinetic electronic energy of non-interacting electrons can be calculated by an equivalent expression to the expression used in HF theory, corresponding to the first term in equation 2.4. Other components that also have equivalent expressions to the ones used in HF theory are the second term corresponding to the classical electron-nucleus attraction, $E_{ne}[\rho]$, and the third term which is the classical Coulomb repulsion between charges, $J[\rho]$. Finally, the last term is the exchange-correlation term that corresponds to the difference between the exact kinetic energy of the real system and the kinetic energy of non-interacting electrons and all non-classical interactions between electrons, $E_{xc}[\rho]$. Nevertheless, the latter term is the only unknown energetic term in the KS formalism of DFT. The major problem in DFT is to derive accurate expressions to de-

scribe the exchange-correlation term. Assuming that the exact $E_{xc}[\rho]$ functional would exist, DFT-KS would be able to provide the exact total energy of the system with less computational cost as one needs for determining the uncorrelated HF energy.

The main difference between the large number of DFT functionals is the nature of the approximation for the exchange-correlation functional. The simplest approximation for the exchange-correlation energy used in DFT is the local density approximation (LDA),[165] which approximates the exchange-correlation energy density at a given position as a function of the electron density at that same local position. At the next level of sophistication comes the generalized gradient approximation (GGA),[165] for which the energy approximation also depends on the gradient of the density at that given position. A level above is meta-GGA, in which the energy density also depends on the local kinetic energy density of the calculated KS orbitals or the Laplacian of the density (i.e. second derivative). Higher-order approximations such as the so-called hybrid functionals, or fully non-local functionals, further improve the accuracy, but at the price of non-locality in the density dependence and higher computational cost.[167] In hybrid DFT methods, the exchange-correlation includes a fraction of the HF exchange; while in fully non-local functional or double hybrids functionals, the energy density approximation also depends on unoccupied orbitals.[165]

Despite that the accuracy of DFT functionals is in general quite good, the functional choice may depend on the property we want to predict. In this case we are interested in spin-states splitting energies calculations in transition metal containing systems. This issue is especially important in the last two chapters for the calculation of the mechanism for reactions and prediction of Mössbauer parameters. Hybrid functionals include HF exchange in the E_{xc} term, thus favoring high spin sates states because of a larger number of exchange interactions present when more electrons are with parallel spin. In contrast, the combination of the exchange functional (OPTX)[168] and PBEc[169] correlation functional gave excellent results regarding spin-state interactions, giving rise to the new OPBE functional.[170] Swart has shown an important study addressed to iron complexes suggesting the use of GGA type OPBE functional to properly describe spin state splitting energies and to obtain Mössbauer parameters.[171,172]

2.1.1 Application of DFT in reaction mechanisms and spin-states

Despite the common use of DFT to compute the mechanism of the chemical reaction, it is quite important to take computational considerations for properly describing the system and the environment within its own limitations. Extra terms and considerations are important to take into account in DFT for a better representation of the system explained below:

A. Solvent effects play an important role in the reaction mechanism, for which many different methods exist to evaluate the solvent in quantum chemical calculations. These methods can describe the solvent explicitly in the coordinate of the system or treat the solvent implicitly. The approximation used for the latter method in the Amsterdam Density Functional (ADF) package is the COnductor-like Screening MOdel (COSMO).[173-176] The COSMO model employs the Solvent Excluding Surface (SES), which consists of the path traced by the surface of a spherical solvent molecule rolling about the van der Waals surface. Thus, it employs the molecular shape cavities and represents the electrostatic potential by partial atomic charges.[165]

B. Weak dispersion interaction such as London dispersion and hydrogen bond interactions are not correctly described by DFT functionals, as they underestimate its value. For the different projects present in this thesis, hybrid and GGA based functionals have been used. The dispersion correction was included when necessary by using the methodology proposed by Grimme and coworkers.[177-179] Grimme's correction consists of an add-on term to standard KS density functional energy values to describe the total energy as:

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \tag{2.5}$$

Here $E_{KS\text{-}DFT}$ is the self-consistent KS energy from the density functional and the E_{disp} is the dispersion correction term.

C. Heavy element compounds are systems with many electrons, where the correct description of the more internal orbitals is needed using relativistic effects. The Zeroth-

Order Regular Approximation (ZORA) is implemented in ADF, allowing the correct treatment of the relativistic effects.[165,180]

D. The choice of the basis set is also an important point to take into account. Usually the expansion of atomic orbitals is done in terms of Slater type orbitals (STOs) or Gaussian type orbitals (GTOs). The charge distribution in a molecule is not uniform due the electronegativity properties of the atoms, for which the introduction of polarization terms in the basis sets describe more adequately the latter fact. Additionally, diffusion function can be introduced for better describing anion o dipole moments. Generally, the greater the size of the basis set (s-, p-, d-, and f-functions) to treat the different KS orbitals, the better the description will be, although it will also increase the computational cost of the calculations.

E. Within the calculation of spin states, the use of open-shell systems treated with unrestricted Hartree-Fock (UHF) method often show non-pure spin states (i.e. spin state contamination). A good descriptor to inspect the amount of spin-contamination is the expectation value of the S^2 average. This expectation value is given by $S\cdot(S+1)$ for pure spin states, with S being the total spin polarization. Thus, the expectation value leads to pure spin states as: 0 (singlet), 0.75 (doublet), 2.0 (triplet), 3.75 (quartet), 6.0 (quintet), and so on. Nevertheless, in many cases the value obtained is larger than the pure spin state, due to the admixture with states of higher multiplicity. For example, if the complex (singlet spin state) calculated gives an expectation value of 1.0, it indicates that there is a contamination of mixing singlet and triplet spin states. The correction energy calculation of the pure spin state (E_{pure}) can be done through the energy of the pure higher multiplicity (in that case triplet) and followed by the correction of the contaminated energy for the non-pure spin state as shown in equation 2.6 and 2.7.[181,182]

$$E_{pure} = \frac{E_{cont} - aE_{S+1}}{1-a} \tag{2.6}$$

$$a = \frac{\left\langle S^2 \right\rangle_{cont} - S(S+1)}{\left\langle S^2 \right\rangle_{S+1} - S(S+1)} \tag{2.7}$$

Here, it is considered the energy of the pure spin state as the sum of the non-pure spin state and a portion of energy with the higher multiplicity (*S*+1), obtaining the correction energy value of our system.

2.2 Mössbauer spectroscopy

^{57}Fe Mössbauer spectroscopy is a technique that provides information about the chemical nature of Fe systems, regardless to its spin and oxidation states. Because iron is an essential element in all living creatures and is incorporated into many enzymes and organometallic complexes, this technique has been exhaustively used to provide information along reaction mechanisms for a broad range of systems.[183] The Mössbauer effect is the recoilless resonance absorption and emission of nuclear γ irradiation in the ground state (I = 1/2) and its first excited spin state (I = 3/2) of ^{57}Fe.[184,185] This spectroscopic technique is mostly applied to studies of the ^{57}Fe isotope, a stable isotope with low γ-energy and long duration in its excited state, with a natural abundance of 2.2%.

Technically, a Mössbauer spectrum results in the counting of transmitted γ-photons as a function of the source's velocity (mm s^{-1}). Thus, the energy of the emitted γ-photon is modulated using the Doppler effect according to the following formula, in which *v* is the Doppler velocity and *c* is the speed of light:

$$\Delta E = \left(v / c\right) E\gamma \tag{2.8}$$

From the Mössbauer spectrum it is possible to measure different parameters: the isomer shift (δ, mm s^{-1}), quadrupole splitting (ΔE_Q, mm s^{-1}), and the magnetic hyperfine interaction.

The isomer shift is the measure of the *s*-electron density at the ^{57}Fe nucleus, and thus is affected in a major way by the *s*-electrons participating in the bonding and *d*-orbitals due to the shielding effect. For this reason the isomer shift (δ) is quite sensitive to its oxidation state, spin state, the number and the topology of the ligands. Hence, δ is directly related to geometry, coordination, and electronic properties of the system. A general trend that helps in the interpretation of Mössbauer isomer shifts (**Fig. 2.2**): (i) $\delta Fe^{IV} < \delta Fe^{III} < \delta Fe^{II}$, (ii) δ high spin < δ low spin state, (iii) soft ligands such as sulfur cause

lower isomer shifts due to the shorter Fe-S bond compared to nitrogen and oxygen-based ligands, and (iv) iron with a six coordination has higher isomer shifts than iron with four coordination, due to the less covalent bond behavior. Nevertheless, it is not always straightforward to obtain clear results just by looking at the δ. For instance, low spin states of Fe^{II} and Fe^{III} have close isomer shift values making it difficult to distinguish which is the correct spin state (**Fig. 2.2**).

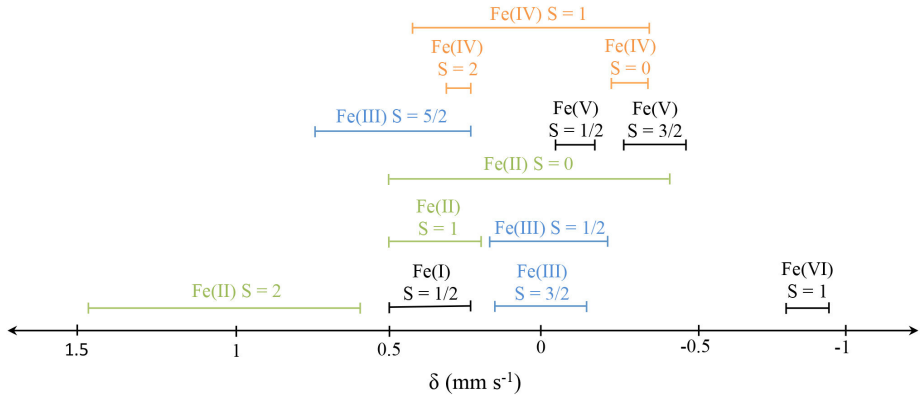


Figure 2.2 Representative isomer shift ranges in different oxidations and spin-states of ^{57}Fe systems.

Quadrupole splitting is measurable because the nuclear excited I = 3/2 state of ^{57}Fe produces a quadrupole moment that can interact with the electric field gradient (EFG). This interaction causes a degeneration into two sub-states with quantum numbers $M_I = \pm 1/2$ and ± 3/2, thus originating the ΔE_Q as the difference in energy between the separation of the two quadrupole lines and is expressed as shown in equation 2.9, where *e* is the elementary charge, Q is the quadrupole moment of the excited state (I = 3/2), V_{zz} is the z-component of the EFG, and η is the asymmetry parameter (see equation 2.30).

$$\Delta E_Q = \left(\frac{eQV_{ZZ}}{2}\right)\sqrt{1+\frac{\eta^2}{3}} \tag{2.9}$$

$$\eta = \frac{V_{xx} - V_{yy}}{V_{zz}} \tag{2.10}$$

The quadrupole splitting normally provides insights about the bond properties and local symmetry helping to establish the oxidation and spin state of the Fe ion in combination with δ.

Magnetic hyperfine interaction occurs by applying a magnetic field, because of the nuclear Zeeman effect (i.e. it determines the interaction of the electron spin with magnetic field, *e*). The magnitude of the splitting depends on the magnitude of the effective magnetic field at the nucleus, resulting from the addition of the applied magnetic field and the internal magnetic field, which is specific to a given Fe system. The analysis of such complex spectra can confirm the spin and oxidation states of the system established from the isomer shift and quadrupole splitting and determine important electronic parameters.[157,186]

The interpretation of Mössbauer spectra is not straightforward, making it important to have the support of theory for extracting all the relevant information from the measured data. A good agreement between experiment and theory can lead to a suitable method to understand and predict spectroscopic properties. DFT can provide acceptable results, however, it uses approximate exchange-correlation functionals that can lead to contradictory results, making the proper choice of DFT functional and basis set extremely important for spin state description.

Different methods and possibilities to calculate the Mössbauer parameters using DFT have been proposed.[187,188] Especially Noodleman and coworkers applied a methodology to predict the isomer shift and the quadrupole splitting of a large number of iron containing metalloenzymes and iron complexes.[189-191] There are many research studies focused on this topic during the last 20 years,[192-195] however, in this section just the methodology developed by Noodleman and coworkers will be explained. Specifically, how they used the DFT to predict the isomer shift and quadrupole splitting.

The δ is produced due to the interaction of the Fe nuclei with the *s*-electron density as mentioned before. The interaction affects the nuclear density, which induces a shift in the spectrum compared to a reference, usually α–Fe at 300 K. Nevertheless, this δ cannot be directly calculated through DFT, but the interaction produced can be expressed as:

$$\delta_S = \frac{2}{3}\pi Z e_0^2 \left(\left\langle R^2 \right\rangle^* - \left\langle R^2 \right\rangle \right) \left(\rho^S(0) - \rho^{Fe}(0) \right) \tag{2.11}$$

Here, Z is the atomic number of the iron nucleus, e_0 is the elementary charge, R^* is the nuclear radii of the excited state, while R is the radius of the ground state of iron. The ρ^S is the electron density at the nucleus of the iron system and the ρ^{Fe} is the same for the reference. Within equation 2.11 most of the terms are constant except the electron density of the iron system, simplifying the expression as:

$$\delta_S = A\left(\rho^S(0) - \rho^{Fe}(0) \right) + B \tag{2.12}$$

The electron density of the iron system and the reference value are possible to determine through DFT calculations. The coefficients A and B possibly can be obtained by a linear regression of the experimental δ versus the computed electron densities of a set of iron systems.[188,196] The electron densities vary depending on the functional and basis set used,[197] thus each combination of functional and basis set will need to be parameterized again to obtain the new coefficients for the desired method selected. Furthermore, two possible parameterizations have been done by Noodleman and coworkers: a general linear regression enclosing $Fe^{+2, +3, +4}$ and the other splitting Fe^{+2} and $Fe^{+3,+4}$ into two independent linear regressions. The global fit showed problems underestimating the isomer shift of Fe^{+2} and overestimating the isomer shift of Fe^{+3}, while the isomer shift is reasonably predicted using two different linear regression for Fe^{+2} and $Fe^{+3, +4}$.[190] Moreover, the application of solvent effects is also important to obtain reliable values for the correct comparison of isomer shift values from experiments.[193]

The quadrupole splitting can be obtained directly following equation 2.9, through the use of DFT calculations to retrieve the corresponding value. The only parameter that is not constant in equation 2.9 is the η, which describes the asymmetry of the charge distribution as mentioned before. Hence, DFT calculations are able to provide the charge distribution around the iron nucleus, which is then used in the quadrupole splitting calculation. As in the case of isomer shifts, the application of solvent effects enhances the DFT calculation of quadrupole splitting for a set of iron systems.[193]

Overall, no single universal method can provide the correct calculation of isomer shift and quadrupole splitting for all iron systems. Many studies recommend the use of hybrid functionals as they give better results due to the inclusion of nonlocal corrections. However, good results can be also obtained using GGA functionals as shown by Noodleman and Swart studies,[172,190] where the OPBE functional provides the correct calculation on the Mössbauer parameters for a large set of iron systems.

2.3 Force Fields

The bigger the system the more expensive and complex is the equation to be solved. This statement is the so-called many-body problem, which is a general category enclosing the physical problems to describe a large number of interacting particles. Moreover, the larger the number of atoms the more time consuming are the DFT calculations.[198] For these reasons different methods exist to treat large systems, being Molecular Mechanics (MM) one of the methods selected in this thesis for the treatment of large systems as enzymes.

In MM the Born-Oppenheimer method is assumed to be valid (i.e. assumption that the motion of atomic nuclei and electrons in a system can be separated) approximation and the potential energy is calculated as a function of the nuclear coordinates using force fields (FF, **Fig. 2.3**). The electrons are not explicitly considered in FF methods, making quantum features also neglected. Overall, the molecules are represented as "balls-on-spring" scheme, being each atom considered as a charged ball and the spring as a covalent bond.

The potential energy is formulated as a sum of different terms, each of them describing the energy required for distorting the molecule in a specific manner.

$$U_{FF} = U_{str} + U_{bend} + U_{torsion} + U_{vdw} + U_{el} \tag{2.13}$$

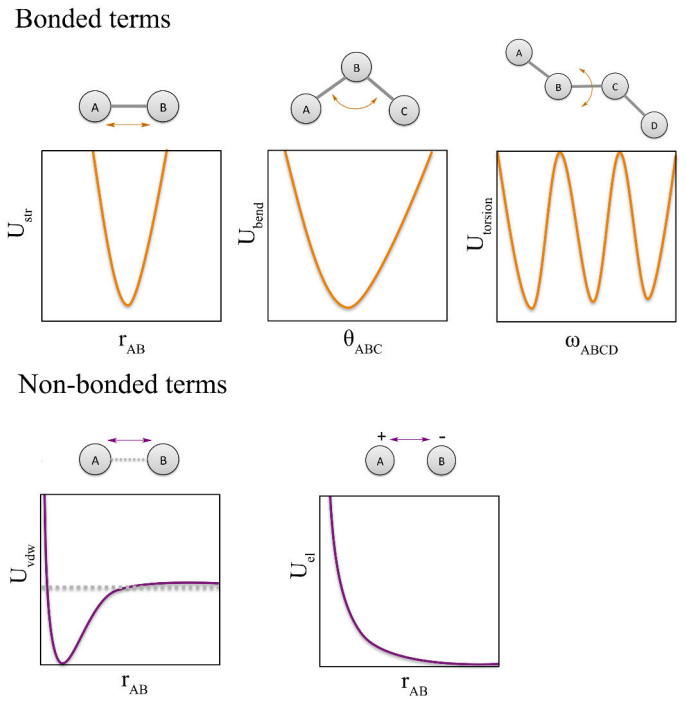


Figure 2.3 Bonded term models of bonds, angles, and dihedral torsions above. Non-bonded term models van der Waals (Lennard-Jones potential) and electrostatic interactions below.

The U_{str} is the energy function for stretching a bond between two atoms (A-B). Its expression is a quadratic displacement of the minimum as a harmonic oscillator.

$$U_{str} = \sum_{bonds} \frac{1}{2} k_{AB} \left(r_{AB} - r_{AB,eq} \right)^2 \tag{2.14}$$

Being k_{AB} the force constant of the bond, r_{AB} the bond length, and $r_{AB.eq}$ the equilibrium distance. The U_{bend} is the energy to describe an angle bend formed by three atoms connected by bonds (A-B-C), which is also described by a harmonic approximation as:

$$U_{bend} = \sum_{angles} \frac{1}{2} k_{ABC} \left(\theta_{ABC} - \theta_{ABC,eq} \right)^2 \tag{2.15}$$

Here k_{ABC} is the force constant of the bend angle, θ_{ABC} the angle, and $\theta_{ABC.eq}$ the equilibrium angle. In some cases the out-of-plane bending energy term (U_{oop}) is also includ-

ed in the potential energy calculation. This latter term is considered in the angle where the central atom is sp^2-hybridized (A-B-C).

$$U_{oop} = \sum_{impropers} \frac{1}{2} k_{ABC,oop} \left(\varphi_{ABC} - \varphi_{ABC,eq} \right)^2 \tag{2.16}$$

As showed before $k_{ABC,oop}$ is the force constant of the improper bend angle, φ_{ABC} the improper angle, and $\varphi_{ABC.eq}$ the equilibrium improper angle. The $U_{dihedral}$ describes the torsion energy to describe the rotation in four atoms connected by bonds (A-B-C-D), which is described as a Fourier series to enclose the periodicity represented as:

$$U_{torsion} = \sum_{dihedrals} \frac{1}{2} V_{ABCD} \left(1 + n\omega_{ABCD} - \gamma_{ABCD} \right) \tag{2.17}$$

Here, V_{ABCD} is the torsional force constant, n the multiplicity of the *cos* function, ω_{ABCD} the dihedral angle, and γ_{ABCD} the phase angle. All terms discussed so far correspond to the bonded interactions present in connected atoms, while the terms of the next paragraphs will correspond to the non-bonded potential energy terms of atoms.

The U_{vdw} is the van der Waals energy, which describes the attraction and repulsion between atoms not connected by bonds. At large interatomic distances the U_{vdw} is zero, while very short distances are very repulsive. Nevertheless, at intermediate distances there exists a slight attraction interaction. The force associated in this term is often referred to dipole-dipole, dispersion, London, and van der Waals interactions, being the latter the main interaction. A common function that fits with the behavior mentioned before is the Lennard-Jones (LJ) potential.

$$U_{vdw} = \sum_{A}^{N} \sum_{B=A+1}^{N} \left(4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right] \right) \tag{2.18}$$

Being ε the depth of the minimum, r_{AB} the distance between A and B, and σ the minimum energy distance. The last non-bonded term, U_{el}, describes the interaction between point charges given by a Coulomb potential.

$$U_{el} = \sum_{A}^{N} \sum_{B=A+1}^{N} \left(\frac{q_A q_B}{4\pi\varepsilon_0 r_{AB}} \right) \tag{2.19}$$

Here, the q_A and q_B are the fractional charges of the atoms respectively, ε_0 the dielectric constant, where the value one corresponds to vacuum. Normally, the value used is between one and five, although there is little theoretical justification for any specific value.[199] The r_{AB} is the distance between atoms.

The FF designed to treat large systems such as proteins are the so-called class I, where the energy terms are considered as simple as possible. However, it is worth to mention that FFs have also been developed to treat the explicit water molecules present in MD simulations such as TIP3P[200] and TIP4P-D.[201] All energy terms mentioned before have been used in the class I FFs,[202] in other words, only harmonic functions are used to calculate U_{str}, U_{bend}, and $U_{dihedral}$. LJ is used to describe U_{vdw} interactions, and U_{el} use fixed charges during the whole simulation. On the other hand, the so-called class II and III FF are more complex expressions trying to better represent the different motions, however, the inclusion of more terms increase the computational cost. Thus, usually both FF classes are used for small and medium size molecules due to the computational cost attributed. Class II FFs includes the cross energy term (U_{cross}), which consists of covering couplings between stretching, bending, and dihedrals terms. This latter term is able to correct interactions affecting others interactions, such as in water molecules, where a strongly bent H-O-H molecule tends to stretch its O-H bonds. Finally, the class III FFs considers, for example, electronic polarization effects, where the charges of the atoms are not fixed during the simulation.[165] In this thesis only class I FFs are used for all the projects related to conformational dynamics (Chapters 4, 5, and 6).

After the description of all the different terms of the FF, the question is how to determine the different force constants of the different terms. One of the issues is that ex-

perimentally it is complicated to abstract the desired information, being usually common the use of QM calculations to extract the parameters for the molecules, especially the bonded terms. Concerning to the non-bonded terms, the electrostatic data are obtained through fit parameters to the electrostatic potential obtained by sophisticated QM calculations, such as CHELPG[203] and RESP.[204] Once a set of charges has been found, one needs to optimize the van der Waals parameters. However, it is a difficult step due to the limited accuracy of QM calculations when dealing with dispersion interactions. Therefore, the assignment of values is usually based on fitting experimental data in solid and/or liquid phases.[165,205,206] These latter procedures can occur in an iterative process until the parameters converge. FF treats the molecules as atom types, which depends on the atomic number and the type of chemical bonding involved in. Furthermore, the FF selected will include different atom types for all the elements in the molecule, describing and putting together atoms with similar topology. This method decreases the dimensionality of the large number of atoms in a system. For instance, which type of carbon is selected depending on the neighbor atom describing its saturation bond; single bond, double bond, delocalized bond in aromatic rings, among others. After all these procedures, validation of FFs are done by evaluating how good they are in predicting geometries and relative energies of different conformations in a diverse range of systems.[165]

As we mentioned in the previous subsection, the correct QM description of metals is not straightforward. To further complicate this, the parametrization of metals in FF is a holy grail, and seeking which is the better method to describe all metals without taking into account the *d*-orbitals is not straightforward. Different approaches have been developed to consider the metal in class I FFs. The most important approaches are the bonded model, the non-bonded model, and cationic dummy model (CDM).[207-210]

The bonded model is a method that uses all energy terms mentioned before, normally combining QM calculations with the Seminario algorithm[207] to extract the information of stretching, bend, and torsions involving the metal. The Seminario algorithm is able to obtain the needed parameters of the metal and the surrounding atoms for the different terms, however, more protocols have been tested to obtain harmonic FFs.[208,211] The bonded model applied a covalent bond between the metal and the coordinated atoms permanently, making it also possible to include the charge transfer induced by the

presence of the metal using RESP algorithm.[212] Nevertheless, the main drawback is that the coordination number of the system cannot change during the whole simulation. This is due to the covalent bond specification, which makes it impossible to simulate the switch of ligands if that would be needed.

The non-bonded model only considers the non-bonded energy terms to model the interactions of the metal in the center of coordinated atoms. The values of the LJ parameters have been extensively adjusted for different metals and oxidation states to reproduce free energy of solvation and metal oxygen distances.[209] Furthermore, more terms have been included in the equation 2.18 to take in consideration the dipole-induced dipole interactions.[213-215] Here, the main advantage is the possibility to perform the simulation without covalent bonds on the metal, allowing more freedom between the metal and the coordinated atoms. Nevertheless, the principal drawback is that the charge transfer is not considered treating the metal as an integer charged ball.

Finally, CDM considers the metal center as a set of cationic dummy atoms, which are connected around a central atom in the specific coordination geometry to be attained.[210,216] The number of dummy sites can vary depending on the coordination number of the metal ion selected.[210,217,218] It is worth to mention that to favor experimental values of energy, the metal atom is usually negatively charged, while the dummies are positively charged. The general exception with the set-up is the copper metal atom, where different properties are taking into account to consider the Jahn-Teller effect (i.e. distortion of the molecules and ions because of the electronic configurations).[219] The CDM provides a powerful non-bonded description for different metals, considering the charges distributed in different regions of the sphere mimicking the coordination geometry of the metals. It is important to know that rather than fixed charges FFs there are the so-called polarizable FFs to model metal ions that have not been used in this thesis. For more details about the metal ion modeling using classical mechanics, Li and Merz reported a magnificent complete review with all tasks done within this topic.[220]

The clear advantage of FFs is the speed of the calculations performed, being able to treat large systems as proteins and DNA. For instance, modern all-atom fixed-charge protein FFs have been successfully used in protein folding and conformational dynam-

ics thanks to the improvements in the development of the different FFs,[221-223] which are mostly done in the correction of torsion potentials and non-bonded interactions. These improvements are achieved using different approaches: using only QM calculations, fitting experimental NMR or structural data, or by combining both latter approaches.[224] For systems with good parameters available, it is possible to perform very good predictions of geometries and its relative energies, obtaining the barrier for the interconversion between different conformations.[225,226] Despite these advantages, there are some important limitations to consider. It is normal to find molecules in which little or zero parameters are present. Thus, in some cases it is difficult to obtain the needed parameters to perform the simulation. Moreover, many deficiencies still exist in FFs, for example, in simulations of intrinsically disordered proteins,[227-229] protein folding equilibrium and its dependency on temperature,[230] and in correctly identifying the protein folding pathways/intermediate.[231]

2.4 Molecular Dynamics

Force fields mentioned in the previous section allow the treatment of large systems, specifically, the energy of the system selected can be calculated. Molecular Dynamics (MD) simulations calculate the motion of the atoms that compose the molecular assembly using Newtonian dynamics. By applying equation 2.20 the net force and acceleration experienced by each atom can be determined.

$$F_i(t) = -\frac{\partial}{\partial r_i} U\left(r_1, r_2, ..., r_N\right) = m_i \frac{d^2 r_i}{dt^2} \tag{2.20}$$

The $U(r_1,r_2,...,r_N)$ is the potential energy depending on the coordinates of N particles calculated by FFs, the m_i the mass of the particle, r_i the coordinates, and t time. Nonetheless, the latter equation can be only solved numerically, with the Taylor expansion algorithm (equation 2.21) as one possible solution.

$$r_{i+1} = r_i + \frac{\partial r}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 r}{\partial t^2}(\Delta t)^2 + \frac{1}{6}\frac{\partial^3 r}{\partial t^3}(\Delta t)^3 + ... \tag{2.21}$$

The first derivative (*dr/dt)* corresponds to the velocities (v_i), while the second derivative (d^2r/dt^2) to the acceleration (a_i), the third derivative (d^3r/dt^3) the hyperaccelerations (b_i), etc. The coordinate positions at a small time step earlier are derived by the substitutions of Δt by $-\Delta t$ in equation 2.21.

$$r_{i+1} = r_i - v_i(\Delta t) + \frac{1}{2}a_i(\Delta t)^2 - \frac{1}{6}b_i(\Delta t)^3 + \ldots \tag{2.22}$$

The combination of the two latter equations allows the calculation to predict the position a time step later from the current and previous coordination position, and also the current acceleration as shown in the next two equations:

$$r_{i+1} = \left(2r_i - r_{i-1}\right)a_i(\Delta t)^2 + \ldots \tag{2.23}$$

$$a_i = \frac{F_i}{m_i} = -\frac{1}{m_i}\frac{d}{dr_i}U\left(r_1, r_2, \ldots, r_N\right) \tag{2.24}$$

Equation 2.24 corresponds to the Verlet algorithm[232] for solving Newton's equation numerically, without considering the hyperacceleration. In order to determine the first position without previous coordinates position available, a first-order approximation is used.

$$r_{-1} = r_0 - v_0(\Delta t) \tag{2.25}$$

The acceleration is evaluated by the forces obtained by FFs at each time step, allowing the propagation of the particles along the time, generating a trajectory. There are some issues related to Velvet's algorithm. First, despite a small step will favor the correct trajectory of the system in a given total time, it increases inversely the computational effort. Second, the forces cannot be calculated with infinite precision. Third, there is a numerical disadvantage, since the time step is a small number and ($2r_i+r_{i-1}$) is a large number, thus will generate truncation errors to finite precision. The last disadvantage is that velocities are not explicitly taking into account, making the use the leap-frog algorithm[233] one of the possible solutions to overcome this drawback. The latter al-

gorithm consists of performing expansion analogous to equation 2.21 and 2.22 with half of the time steps, giving the analogous expansion of velocity (equation 2.25).

$$r_{i+1} = r + v_{i+\frac{1}{2}}(\Delta t) \tag{2.26}$$

$$v_{i+\frac{1}{2}} = v_{i-\frac{1}{2}} + a_i(\Delta t) \tag{2.27}$$

The main disadvantage of leapfrog algorithm is that coordinate positions and velocities are not known at the same time, being always out by half a time step. This inconvenience is solved using the propagate atoms equations.

$$r_{i+1} = r_i - v_i(\Delta t) + \frac{1}{2}a_i(\Delta t)^2 \tag{2.28}$$

$$v_{i+1} = v_i + \frac{1}{2}\left[a_i + a_{i+1}\right](\Delta t) \tag{2.29}$$

The choice of the proper time step is an important control parameter for a simulation. A too large time step choice may give instabilities due to high energy overlaps between particles, while a too small time step will increase the computational cost as mentioned above. The appropriate time step chosen will cover efficiently the phase space (i.e. 3N coordinates (r_i) and 3N momenta (v_i), 6N dimensional space) and the collisions will occur smoothly. Molecular motions usually require the order of femtoseconds (10^{-15} s) or less to occur, and thus it requires a total simulation time of one nanosecond (10^{-9} s) *ca.* 10^6 time steps to cover the full simulation time, while one microsecond (10^{-6} s) requires *ca.* 10^9 time steps. As mentioned in the conformational dynamics section, there are conformational changes that occur in the millisecond and second time scale, making it difficult to explore in a single trajectory the slow conformational changes of the systems. Although the size of the system influences the time needed to achieve the desired time scale, the billion steps required for obtaining microsecond time scale is rather difficult. Therefore, the accessibility to long time scale is another important issue regarding MD simulations, the so-called sampling problem.[234] Many simulations may be needed with different conditions to properly describe these slow dynamic processes.

The quickest motions occurring in molecules are the stretching vibrations, especially those involving hydrogen atoms. Nevertheless, these motions have a relatively low influence on many processes. Bond lengths involving hydrogen atoms can be considered rigid, which is advantageous, as it allows to consider longer time steps and longer simulation times at the same computational cost. As all atoms move individually according to Newton's equation, constraints must be applied for fixing hydrogen bond lengths, which is done using different methods such as SHAKE[235,236] and LINCS[237] algorithms. These methods allow the atoms to first move under the influence of the forces, and are subsequently forced to satisfy the constraints.

MD simulations generate a collection of points in phase space accomplishing certain thermodynamic conditions, also called ensembles. Those points generated as function of time belong to the same ensemble, corresponding to the different conformations of the system. The number of constraints to simulate thermodynamic ensembles are usually three. The standard in MD simulations is the microcanonical ensemble (*NVE*), which is characterized by fixed number of atoms *N*, fixed volume *V* and fixed energy *E*. In that case the total energy is almost constant, however, the contribution of kinetic and potential energy will contribute differently in every time step. The kinetic energy depends on the fluctuation of the temperature, while the potential energy depends on the position and mass of the particles of the system. Moreover, if the volume of the system is fixed the pressure will fluctuate, in some cases one can prefer to fix other parameters to better simulate experimental results. Fortunately, with MD simulations it is possible to generate canonical ensembles (*NVT*) and isobaric-isothermal ensembles (*NpT*). The number of atoms *N* and temperature *T* are fixed in both ensembles, while the main difference is in the fixed volume *V* or fixed pressure *p*, respectively. The achievement of such conditions will depend on the control of velocities or positions in each time step maintaining constant the selected desired conditions. To maintain the temperature almost constant there are two possible methods: multiplying the velocities at each time step by a factor of $(T_{desired}/T_{actual})^{1/2}$, or coupling the system to an external heat bath, which adds or extracts energy to/from the system to control the temperature. The last method mentioned is also used to maintain the pressure at a constant value.[165]

A realistic model of solutions requires the presence of a huge number of explicit solvent molecules without boundary effects. To do so, periodic boundary conditions (PBCs) allow the MD simulation to use a relatively small number of explicit solvent molecules experiencing forces as if the system was in a bulk fluid. The PBCs consist of placing the solvent molecules in any cell shape such as cube, truncated octahedral, and rhombic dodecahedron.[238] This cell shape is duplicated in all directions, providing the possibility that if a solvent molecule leaves the central cell shape through the bottom wall, its image will enter the cell shape through the above wall from the neighboring cell shape. This makes the solvent model quasi-periodic, with a periodicity equal to the dimension of the cell shape.

To further decrease the computational cost in MD simulations, a truncated cutoff is selected for the calculation of the non-bonded energies. These energies are the most expensive calculations, with the non-bonded interactions increasing as the square of the total number of particles in a system (N^2). Therefore, the usual approach consists of choosing a spherical truncation to only take into account the interaction between the particles separated by distances smaller than the selected sphere. Computationally efficient methods have also been developed for reducing the N^2 scale to calculate van der Waals interactions as the particle mesh ewald (PME)[239,240] and fast multiple moment (FMM).[241]

2.1.2 Conformational sampling

The time-dependent properties of MD simulations make it possible to sample a broad range of conformations that enzymes can adopt in solution. But they can only be connected with experimental observables if all relevant states or conformations of the system are explored (i.e. ergodic hypothesis).[242] However, certain sampling of conformations may not be possible since higher energy transitions are rarely visited. This is due to the small time steps used to perform the simulation, which are on the order of femtoseconds. Simulations with this small time step permit to compute a few nanoseconds in standard computers, making it rather difficult to compute the millisecond to second timescales needed to explore domain motions and allosteric transitions occurring in some enzymes, as mentioned in the introduction.[243]

The sampling problem, mentioned above, makes it difficult to compare directly with experimental data, encouraging the search for alternative approaches.[165,244] These alternative approaches can be classified in unbiased or bias methods. The unbiased methods consists of central processing unit (CPU) parallelization,[245] the Anton supercomputer,[246] graphic processing unit (GPU) based clusters, replica exchange and parallel tempering.[247] The biased methods consists of accelerated Molecular Dynamics (aMD),[248] metadynamics, [249] and umbrella sampling.[250] GPU based clusters and/or aMD have been used in the Chapters 4, 5, and 6 of this thesis to rationalize the conformational dynamics of enzymes.

The parallel power of GPU offers an affordable alternative to increase MD accessible timescales by running either single long and/or multiple short simulations of the same system. The multiple MD runs may allow the promotion of infrequent transitions or rare events by starting the simulations from different initial structures and combine them to recover the associated thermodynamics and kinetic of the system.[251] aMD becomes a really interesting method when no clear structural information is available about the transitions and conformational states important for enzymes. This method allows the exploration of biomolecular conformations without *a priori* structural knowledge. In aMD, a time-dependent bias potential is added to raise the energy minima while keeping the potential energy surface below a pre-chosen boost energy values unchanged (equation 2.30 and **Fig. 2.4**), and thus enhancing conformational exchanges.[248]

$$V_{aMD}(r) = \begin{cases} V(r) & ; \quad V(r) \geq E_{boost} \\ V(r) + U(r) & ; \quad V(r) \prec E_{boost} \end{cases} \tag{2.30}$$

Here the E_{boost} is the threshold energy or boost energy, $V(r)$ the new energy potential added by aMD and $U(r)$ the potential energy from the FFs.

$$U(r) = \frac{\left(E_{boost} - V(r)\right)^2}{\alpha + \left(E_{boost} - V(r)\right)} \tag{2.31}$$

Here the α parameter is the acceleration factor. The smaller the α value, the more flattened will be the potential energy surface and easier the transitions will occur from one conformation to the other. Nevertheless, if the value is too small the transitions to overcome conformations will be barrier-less, making it difficult to capture the important conformations of the system. One of the main advantages is that the boost potential can be applied to internal coordinates, for instance, only to the torsional bonded terms to enhance the conformational transitions. This method has been applied to fold a set of small proteins[252] and to study the conformational dynamics of biomolecules, such as the maltose binding protein[253] and HIV protease.[254]

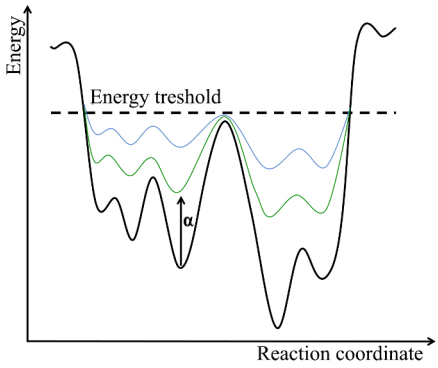


Figure 2.4 aMD diagram representation.

2.1.3 Conformational free energy calculation

The best way to represent the different conformations sampled in the MD simulations is mapping the so-called free energy landscape (FEL). In this FEL, the different conformational states in thermal equilibrium are mapped and populated following statistical thermodynamic distributions. The regions with high populations with similar conformers correspond to either local or global energy minima. The height of the barriers separating different conformational states indicate how fast or slow a conformational transition will occur. Thus, the conformational states separated by small energy barriers require picosecond–nanosecond timescales to exchange, while the conformations connected by high energy barriers are less likely to take place and the transition becomes slower.[243] A particular free energy is linked to a specific protein sequence and defined values of temperature, pressure, and solvent conditions. By manipulating these different parameters, for instance, introducing mutations or increasing temperature will result in dramatic changes in the relative conformational distributions, but also in the kinetics of the conformational state interconversions, as shown in the introduction.

The use of computational methods becomes quite useful for the reconstruction of the FEL of enzymes, commonly using equation 2.32.

$$\Delta G_{rp} \approx -k_B T \ln(P) \tag{2.32}$$

Here the free energy, ΔG_{rp}, is proportional to the negative logarithm of the relative population of different conformations (P), with Boltzmann constant (k_B) and temperature (T) units. A maximum in the population distribution corresponds to a minimum in the FEL. The relative populations may be estimated by going back and forth between states in equilibrium.[242] While the number of transitions increases, the error of the population estimation can be reduced. Despite that, more methods have also been used such as combining protein structure prediction algorithms and metropolis monte carlo techniques for exploring all-atom energy landscape called protein energy landscape exploration (PELE) developed by Guallar and coworkers.[255]

The problem for the FEL reconstruction is how to select the most representative global collective degree of freedom among the high dimensional space as a result of the large number of atoms present in the simulation. The selected parameters should correctly represent processes under study (i.e. the transitions from one conformation to another) within the amount of data generated by MD simulations. One solution is to specifically select a set of global or collective variables (CVs) to reduce the dimensionality. These CVs can be any explicit function of the coordinates of the enzyme, important for the transition of interest, such as backbone dihedral angles, distances between catalytic residues, or the root mean square deviation (RMSD) of loops, in other words, one is free to select whatever coordinates to describe correctly the desired transition of interest. The high dimensional data obtained in the MD trajectories can be projected onto the CVs selected, obtaining the probability distributions to reconstruct the associated FEL. However, reducing the dimensionality of our data to only a few CVs may omit relevant information of conformational transitions not included within the CVs selected. This fact makes the choice of an appropriate set of CVs the most important task, requiring a detailed knowledge of our system, since biased results may be obtained if the selection is not carefully done.

There are many computational approaches to automatically reduce the dimensionality of the data preserving as much information as possible. The most common method used is the principal component analysis (PCA, equation 2.33),[256,257] which performs a dimensionality reduction accounting for as much variance in the data set as possible. For instance, if one defines variance as the deviation of an atom from its mean position along the MD, then each principal component will be a linear combination of strongly correlated atomic motions with large oscillations.

$$Cr_i = \sigma_i r_i \tag{2.33}$$

Here, C is the covariance matrix of all the data, r_i is the matrix of eigenvectors in which the principal components are represented in the columns of the matrix, and σ_i is the eigenvalue diagonalized variance matrix. The dimension reduction (in this case PCA) consists in project the input data onto the principal components obtained from equation 2.33. The first principal component (PC1) will have the maximum variance of the input data, the second principal component (PC2) will have the second greatest variance, and so on so for. The reduced dimensional PCA space can be used to reconstruct the associated FEL, as has been done in this thesis and in several other studies of protein folding and allostery.[258,259] Nevertheless, there are more computational approaches such as the time-structure independent component analysis (tICA), which seeks to lower the dimensionality of the data considering the time correlation of the data instead of the variance.[257,260]

Chapter 3. Objectives

The aim of this thesis is the application of computational tools to evaluate different chemical and biological properties such as the conformational dynamics of enzymes, mechanistic reaction of biomimetic complexes, and Mössbauer prediction in oxo-complexes. The different computational methods applied allow to describe the different properties. For instance, molecular mechanics (MM) is capable of treating large systems for evaluating the conformational dynamics behavior, while quantum mechanics can be used to deeper understand the chemical reaction pathway, and even to predict parameters to unravel the properties of the electronic structure of compounds. This thesis encompasses both theoretical techniques to cover different properties in a broad range of systems. The goals can be summarized as:

Conformational dynamics: The main goal in the exploration of conformational dynamics of enzymes (Chapters 4, 5, and 6) consists of unravelling how enzymes explore different conformational states important for the substrate acceptance and product release (chapter 4 and 5) in natural enzymes and variants by means of MD simulations. Moreover, we also aim to explore how these conformational changes induce a higher activity of the enzyme variants towards novel functionality in the evolution path of a computed designed enzyme through DE using MD simulations (Chapter 6). We also tested the applicability of use a new tool based on correlated motions to identify important residues that can alter the population shift towards specific conformations of the enzyme to enhance novel functionality (Chapter 6).

Reaction mechanism: The main goal is to elucidate in detail the mechanism path for biomimetic complexes, e.g. EUK-8, for the decomposition of hydrogen peroxide using QM (Chapter 7). The proper description of the mechanism will elucidate an idea of which stationary points can be important to enhance the activity. Thus, the understanding of the mechanism and the description of the spin state splitting in the whole reaction pathway will give important knowledge to rationally design analogues of EUK-8, which will enhance the activity towards the decomposition of hydrogen peroxide.

Mössbauer prediction: The main goal in the last chapter is the recalibration of a training set for Mössbauer parameters prediction using functionals that correctly describe the spin state of iron systems. Hence, theoretical QM calculations will elucidate

the characterization of oxo-complexes and their respective spin state of iron metal atoms in which no X-ray structure is available through calculated Mössbauer parameters and computing the different possible structures (Chapter 8).

Chapter 4. Conformational dynamics of 2,4-dioxygenase

This chapter is based on the publication:

Romero-Rivera, A.; Ryde, U.; Feixas, F.; Osuna, S; Swart M. Role of a lid domain in the conformational dynamics of 2,4-Quercetinase dioxygenase. *In preparation*

4.1 State-of-the-art

As mentioned in the first chapter of this thesis, Koshland's induced fit model was proposed to explain conformational changes induced by the presence of the substrate.[261] Nevertheless, experimental and computational evidence suggest an alternative mechanism, the conformational selection model, which postulates that all protein conformations pre-exist and the ligand selects the most favorable conformation. After binding or substrate recognition, the ensemble undergoes a population shift, redistributing the population of the conformational states. There is still an open discussion to decide the correct model to describe the process of binding.[110,112,262] Generally, both conformational selection and induced fit models appear to play roles in molecular recognition and binding.

Enzymes are inherently dynamical and able to sample a vast ensemble of conformations in solution. Hence, thermally accessible conformational sub-states other than the native conformation are important for the processes previously explained. The conformational selection model suggests that weakly populated, higher in energy conformations are responsible for recognizing the substrate with a subsequent population shift towards these conformers after binding. To shed more light on the discussion of the different models, we extensively studied the conformational dynamics of 2,4-dioxygenase quercetinase enzyme and the role of an active site lid on substrate binding (**Fig. 4.1**) by means of long MD and aMD simulations. Many quercetinases have been obtained and crystallized in *apo* state and in presence of flavonols substrates (quercetin and kaempferol) for diverse species. Specifically, we focused our study on the quercetinase obtained from a fungi specie, *Aspergillus japonicius* (Que_A).[263,264] Que_A is a cupin dioxygenase enzyme, which catalyzes the degradation of flavonols in the presence of a molecule of oxygen to the corresponded diphenolic ester (DPE), thereby releasing carbon monoxide (**Fig. 4.1**). Furthermore, Que_A is a mononuclear copper dependent enzyme coordinated by three histidines and one glutamic residue. The chemical reaction mechanism was extensively studied through QM and quantum mechanics and molecular mechanics (QM/MM) calculations since more than 10 years ago.[265-269] Nevertheless, the conformational dynamics of this enzyme and the role of an active site lid on substrate binding have not been studied by means of long MD simulations.

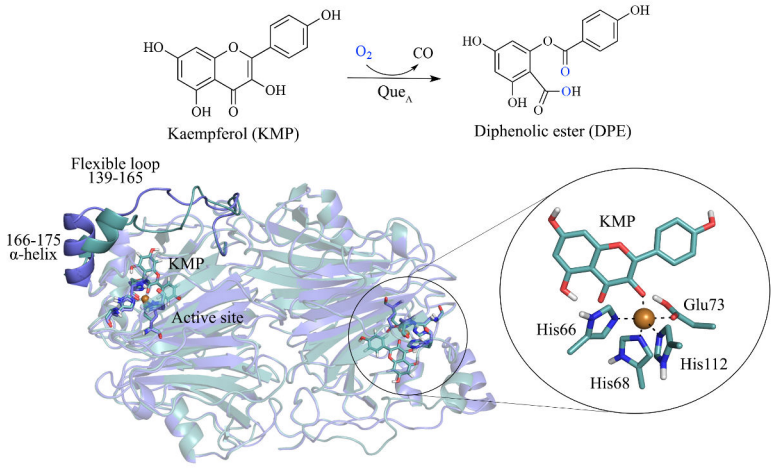


Figure 4.1 QueA dimer enzyme and the catalyzed reaction. *Apo* state (PDB 1JUH) in blue and E-S complex state (PDB 1H1M) in teal with its corresponded active site.

The X-ray structure contains kaempferol (KMP) substrate bound to the Cu^{II} in the active site (E-S complex), which shows van der Waals interaction between one of the rings of the substrate and Pro164, which is present in the flexible lid domain region (teal, **Fig. 4.1**). This interaction is supposed to stabilize a flexible loop region that becomes disordered in the absence of the substrate as shown in the X-ray structure of the *apo* state (blue, **Fig. 4.1**). Two studies revealed important aspects of the possible role of the flexible lid, the entrance of oxygen, the presence of water molecules in the cavity and the interaction of key residues for the binding in short MD simulations (on the order of picoseconds to nanoseconds).[270,271] Still, there are unknown important points from the conformational dynamics of Que_A in *apo* and E-S complex state perspective. For this reason, the present study thoroughly analyzes through microsecond MD simulations the different states of the Que_A (*apo* and E-S complex state), starting in open and closed conformations of the lid, to unravel its role in ligand binding. It is well accepted that conformational changes of open and closed conformations of flexible lid motions can occur in the order from nanoseconds to milliseconds.[98] Moreover, aMD simulations have also been performed to reconstruct the binding process (i.e. how the substrate is able to enter the active site cavity of the enzyme), but also the unbinding pathway of the product after catalysis (i.e. how the product leaves the inner cavity of the enzyme).

4.2 Computational details

MD simulations in explicit water were performed using the AMBER 16 package,[272] using as the open conformation the X-ray of the *apo* state (PDB 1JUH)[263] and the closed conformation the X-ray structure of the E-S complex state (PDB 1H1M).[264] Missing residues of the loop were included using ModLoop.[273,274] Parameters for KMP substrate and product (DPE) were generated within the antechamber module of AMBER 16 using the general AMBER force field (GAFF),[275] with partial charges set to fit the electrostatic potential generated at the B3LYP/6- 31G(d) level by the restrained electrostatic potential (RESP) model.[204] Amino acid protonation states were predicted using the PROPKA3.0 server (http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0). We have used the bonded model for Cu^{II} and the residues of the first coordination sphere, in particular we used the Seminario approach[207] to obtain the metal parameters needed for the simulation as implemented in Prof. Ryde program.[212] The optimization, frequencies and charge calculations to obtain the parameters were done at the TPSS/def2-SV(P) level using Turbomole software.[276] The different systems were solvated in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P water molecules using the AMBER16 leap module. The system was neutralized by the addition of explicit counterions (Na+ and Cl−). All calculations were done using the ff14SB amber force field.[277] A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on the 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.[278] An 8 Å cutoff was applied to LJ 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 maintained at 1 fs during the heating stages, allowing potential in homogeneities 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 a temperature of 300 K. After the systems were equili-

brated in the NPT ensemble, 5 independent one thousand nanosecond MD simulations were performed under the NVT ensemble and periodic boundary conditions. An accumulation of 5 μs of MD simulation of Que_A for different states was obtained: i) *apo* as in the X-ray (PDB 1JUH) with a water molecule bound to copper; ii) *apo* in the closed conformation of the lid with a water molecule bound to copper; iii) with KMP substrate bound to copper as in the X-ray of the E-S complex state (PDB 1H1M); and iv) with KMP substrate bound to copper in the open conformation of the flexible loop. All simulations were carried out with the dimer of the enzyme (**Fig. 4.1**). PCA analysis was done with the *pyEMMA* software.[279] More details are shown in the computational details section in appendix A.

aMD simulations were performed with four KMP substrate around the enzyme starting from the open conformation. Several replicas were done with a boost to the dihedral of the system with an energy threshold of 10695.60 kcal/mol and an α parameter value of 492.80. For the product release, we performed the simulation with two products molecules on their respective active site as calculated in the QM/MM study[267] of the dimer and we applied a boost to the dihedral of the system with an energy threshold of 10696.17 kcal/mol and an α parameter value of 490.00 for several replicas.

Substrate access tunnels were analyzed using the standalone version of CAVER 3.0 software.[280] We selected a structure from the different minima obtained in the PCA analysis in *apo* state. The starting point for the channel calculations was the water molecule bound to the copper metal atom. Transport tunnels were identified using a probe radius of 0.9 Å. Redundant tunnels were automatically removed from each structure, and the remaining tunnels were clustered using a threshold of 12 Å. We also perform the calculation of the binding pocket cavity using POVME software,[281] with limited radius of 5 Å and grid space of 1 Å. The cavity measurement where the substrate binds was done from the water molecule bound to the copper metal atom.

4.3 Results and discussion

4. 3. 1 Conformational dynamics in *apo* and E-S complex state

The Que_A enzyme has a flexible lid domain (140-164 residues, **Fig. 4.1**), which was not completely resolved in the X-ray structures due to the high fluctuation and the presence of lipids.[263,264] The X-ray structure with the KMP substrate in a closed conformation of the lid shows a distance between Leu64-Ser166 of 7.3 Å and an angle between Ala179-Phe175-Ser166 of 28.6°. In contrast, the open conformation in *apo* X-ray structure showed a distance of 13.9 Å and an angle of 44.0° in the lid (**Fig. 4.2**). Both features (distance and angle) allow us to classify the conformations of the lid in closed and open conformations for the analysis of the different sampled conformations by the enzyme along the trajectory. It is worth to mention that the *apo* state structure has more residues unresolved than the E-S complex state structure due to the high fluctuation without the presence of the substrate.

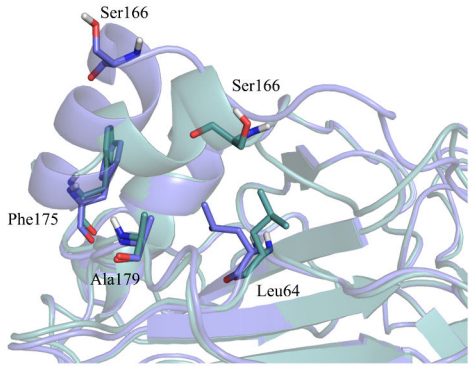


Figure 4.2 Both starting structures for MD simulation, with the angle between Ala179-Phe175-Ser166 of 44.0º for the open state (blue structure) and 28.6º for the closed state (green structure) and a distance between Leu64-Ser166 of 13.9 Å for open state (blue structure) and 7.3 Å for closed state (green structure).

The MD simulations in the *apo* state starting from its X-ray structure (open lid) showed open conformational states along the trajectory, and slightly closed conformations were explored only in short periods of time for one of the monomers (blue angles in Mon2 of *apo* open, **Fig. 4.3** blue angles *ca.* 25°). Thus, overall the average distance and angle are similar to those of the open X-ray structure (**Fig. 4.3** and **Fig. A.1**). Interestingly, only one of the monomers shows a widely open angle (*ca.* 80°) compared with the X-ray structure together to distances of *ca.* 20 Å, thus sampling widely open lid conformations

(blue line of Mon1 in *apo* open, **Fig. 4.3**) not explored before and may are important for the entrance of the substrate into the enzyme cavity. Starting from the closed conformation, only in three out of ten monomers the open states of the enzyme were populated, exploring some close-open-close transitions (dark yellow, **Fig. 4.3**). The analysis of the *apo* conformation suggests that longer timescales are needed to overcome the energy barrier to populate transitions from open-close and close-open conformation of the lid, while in the case of the native state mostly open state conformations are sampled.

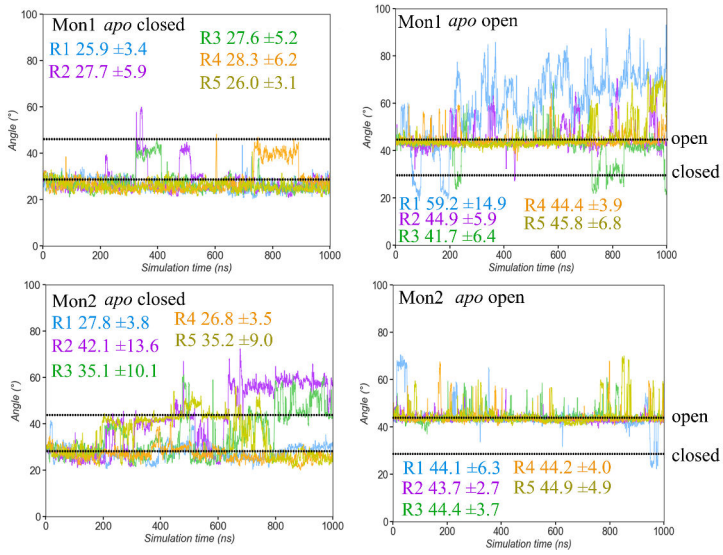


Figure 4.3 Angles measurement during 1000 ns for each replica in the first and second monomer of the *apo* state.

We have also analyzed the accumulated 5 µs MD simulations for both monomers of the *apo* state in a kde plot (**Fig. 4.4**). Kde plots are done representing the distance and the angle features mentioned previously. These plots show that the *apo* state starting from the open conformation populates longer distances and angles compared to the native structure (i.e. widely open conformations are explored). Additionally, in a few simulations a slight population of the closed conformations occurred. The other monomer has explored similar conformations as the native structure (shown in blue on the left, **Fig. 4.4**). On the other hand, starting from the closed conformation, the *apo* state of the enzyme has difficulties to populate open conformations (right, **Fig. 4.4**) and stays most of the simulation time in the closed conformation. This fact is probably due to the high en-

ergy needed to overcome the barrier from close to open conformation in the *apo* state, as shown above.

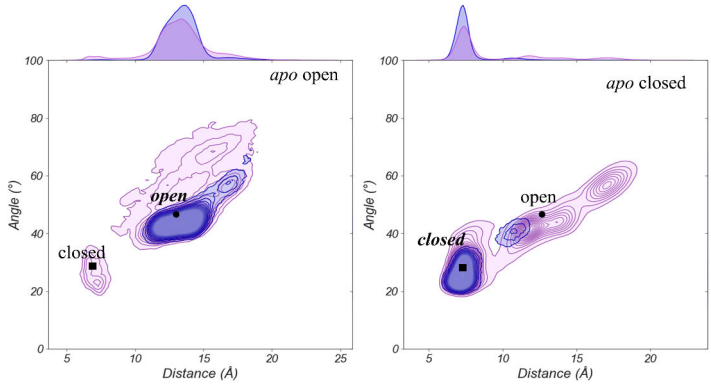


Figure 4.4 Kde plots for the different monomers in the *apo* state, on the left starting from the open conformation (black circle) and on the right starting from the closed conformation (black square).

When the enzyme is in the E-S complex state different behavior is observed in the MD simulations. The stable closed conformation of the native state in presence of KMP induces rigidity to the enzyme. This higher stabilization is responsible for allowing the crystallization of regions of the lid in the E-S complex compared to the *apo* state. During 1 μs of simulation time most of the monomers remain in at constant distances (*ca.* 7 Å) and angles (*ca.* 23-26°) similar to those shown in the native state, while only one monomer sampled transitions from closed to open conformations of the lid (E-S closed of Mon2, **Fig. 4.5**). On the other hand, MD simulations starting from the open conformation have shown that half of the monomers (five out of ten) sampled the transition from open to close conformation of the lid at different times scales (E-S open in **Fig. 4.5**). The transition occurred between 100-600 ns, which suggests that the presence of the substrate facilitates the transition to the closed conformation of the enzyme. Hence, the transition from open to closed conformation of the lid is more feasible thanks to interactions between the substrate and key residues of the flexible loop as Pro164 (**Fig. A.3**)

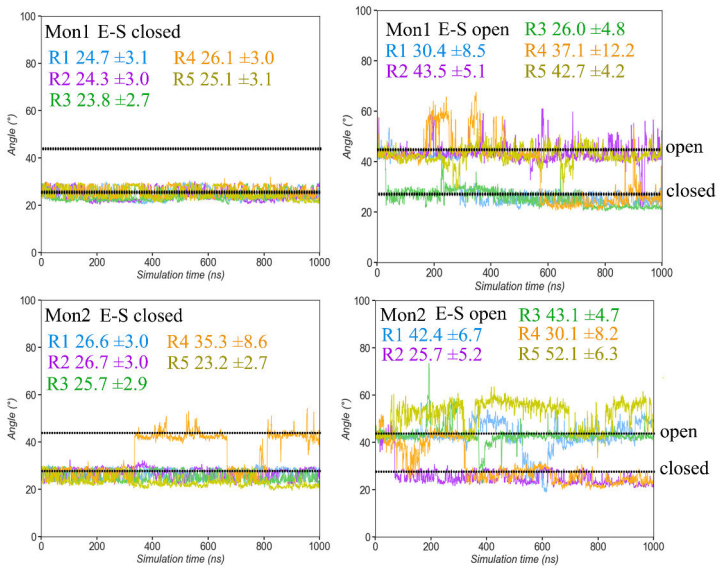


Figure 4.5 Angles measurement during 1000 ns for each replica in the second monomer of the E-S complex state.

Kde plots with the KMP substrate show that starting from the native state, only closed conformations of the lid are sampled with a poor population of the open conformations (**Fig. 4.6**). However, when the simulation started in the open conformation, transitions towards closed conformation are clearly populated. Therefore, the presence of the substrate in the active site facilitates the closure of the lid, which is being stabilized by an intrinsic network of interactions maintaining the closed state of the enzyme.

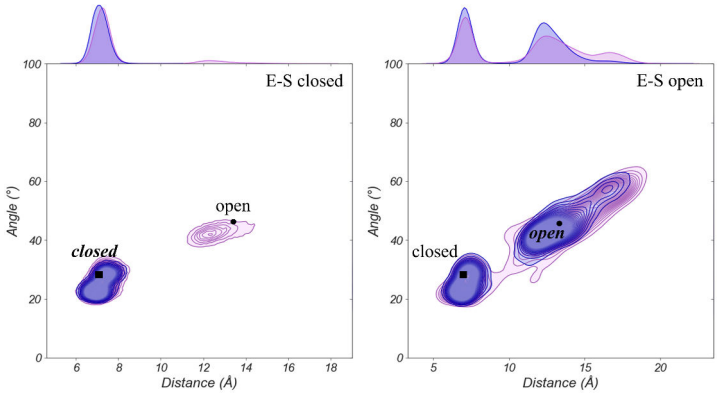


Figure 4.6 Kde plots for the different monomers in the E-S complex state, on the left starting from the closed conformation (black circle) and on the right starting from the open conformation (black square).

To shed light on the conformational dynamics of the enzyme, we have also performed PCA on the Cα of the whole enzyme in all monomers together along the MD simulations (**Fig. 4.7**). The first component, PC1, has higher contributions in residues of the lid indicating how important are those residues to describe the conformational dynamics of the enzyme. The PCA representation of the *apo* conformation shows a population of three different states clearly populated ($\mathbf{A}_{\mathbf{C}}^{apo}$, $\mathbf{B}_{\mathbf{O}}^{apo}$ and $\mathbf{C}_{\mathbf{O}}^{apo}$), while a less populated state is also sampled ($\mathbf{D}_{\mathbf{O}}^{apo}$). The ensemble of conformations sampled in $\mathbf{A}_{\mathbf{C}}^{apo}$ corresponds to closed conformations of the lid with average distances of 7.4 ±0.7 Å and average angles of 26.8 ±4.1°, similar to those values present in the closed conformation of E-S complex X-ray structure. The rest of states $\mathbf{B}_{\mathbf{O}}^{apo}$ (13.6 ±1.3 Å and 45.0 ±4.4°), $\mathbf{C}_{\mathbf{O}}^{apo}$ (14.7 ±2.4 Å and 47.7 ±8.2°), and $\mathbf{D}_{\mathbf{O}}^{apo}$ (15.1 ±2.3 Å and 65.1 ±9.0°) correspond to different open state conformations of the flexible region explored by the enzyme. From this PCA it becomes evident, how starting from the native *apo* state exclusively open state conformation are sampled without populating the closed state, while the PCA of the closed conformation of the lid is able to slightly populate open conformations ($\mathbf{B}_{\mathbf{O}}^{apo}$, $\mathbf{C}_{\mathbf{O}}^{apo}$, and $\mathbf{D}_{\mathbf{O}}^{apo}$ states). Hence, the PCA suggests that longer timescales are needed to overcome the high energy barrier, in order to increase the population towards open conformations when starting from a closed conformation in the *apo* state. However, we have seen that both conformations pre-exist in solution, but the transition from one conformation to the other is high in energy as suggested by the conformational selection hypothesis.

The PCA of the E-S complex state shows globally a higher population of the closed conformational states, which concurs with the X-ray structure that shows a stabilization of the closed lid. The PCA starting from the open conformation clearly shows how the enzyme is able to populate stable closed conformations, mainly thanks to the presence of the substrate inside the active site. The PCA suggests three main states ($\mathbf{A}_{\mathbf{C}}^{E\text{-}S}$, $\mathbf{B}_{\mathbf{O}}^{E\text{-}S}$ and $\mathbf{C}_{\mathbf{O}}^{E\text{-}S}$), with the $\mathbf{A}_{\mathbf{C}}^{E\text{-}S}$ clearly more populated. The latter state corresponds to closed conformations (7.2 ±0.4 Å and 25.0 ±3.2°), while the other states $\mathbf{B}_{\mathbf{O}}^{E\text{-}S}$ (12.6 ±0.9 Å and 42.2 ±3.7°) and $\mathbf{C}_{\mathbf{O}}^{E\text{-}S}$ 14.5 ±2.1 Å and 47.5 ±7.2°) correspond to open conformations of the flexible region in Que_A. These findings corroborate the hypothesis of the conformational selection, in which the presence of the substrate decreases the barrier to populate

the closed conformations. In the other case starting from the native closed state, mostly closed conformations are visited suggesting the high stabilization of the lid. We think that sufficiently long or several short MD simulations may reconstruct the complete free energy landscape of the enzyme, unraveling the timescales needed to overcome the different transitions with and without the presence of the substrate.

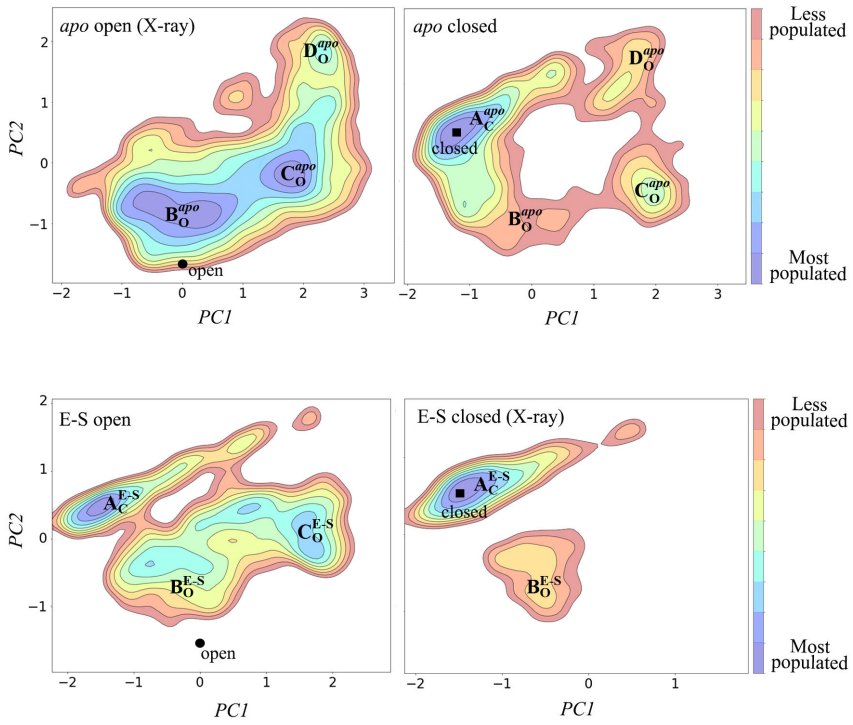


Figure 4.7 PCA for *apo* and E-S complex states with their respective states sampled in the MD simulations. Starting from the open conformation on the left side (black circle) and starting from the closed conformation of the enzyme on the right side (black square).

Despite all this information, we looked in greater detail at the possible role of the lid in this fungus Que_A. To do so, we selected representative structures from the different states of the *apo* simulations to measure the volume of the binding cavity. These calculations in closed and open conformations in the *apo* state suggest that there is not a big volume difference in the region where the substrate is supposed to bind (**Table 4.1**). The lid in open or close conformation is not drastically changing the residues of the binding pocket (left in **Fig. 4.8**).

Table 4.1 Volume calculation in the most populated clusters on the *apo* PCA representation.

States of *apo*	Volume (Å^3)
A^{apo}	68
B^{apo}	61
C^{apo}	61
D^{apo}	70
X-ray Closed	72
X-ray Open	95

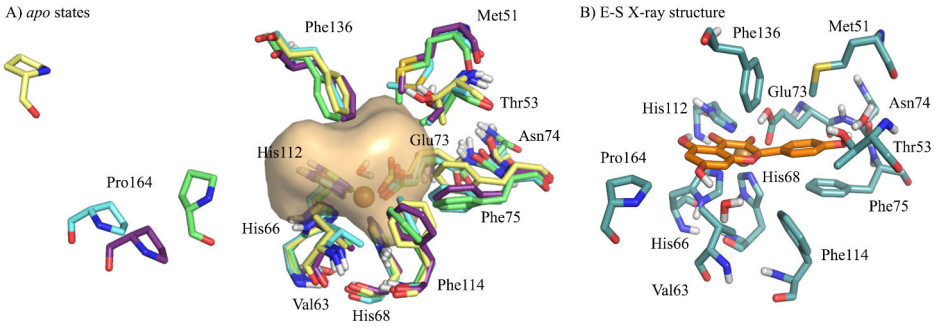


Figure 4.8 A) Volume in the active site for the different *apo* states found in the PCA: A^{apo} (green), B^{apo} (purple), C^{apo} (cyan), and D^{apo} (light yellow) . B) X-ray structure of the E-S complex state with KMP (orange) and binding residues (teal).

One can observe the difference is observed for Pro164, which is located in the flexible loop and is found far away from the active site in open conformations but closer in the binding region in the closed conformations. This fact suggested a high rigidity of the binding pocket towards flavonols substrates, forming key interactions between the aromatic rings of the binding pocket and the substrate (right in **Fig. 4.8**). The accessible tunnels analysis was done to elucidate the KMP substrate access (**Fig. 4.9**). Our calculations suggest the same principal entrance for all the *apo* states, however, a clear difference is found for the more open conformation states ($\mathbf{C}_{\mathbf{O}}^{apo}$, and $\mathbf{D}_{\mathbf{O}}^{apo}$) with a wider tunnel channel entrance than the other states ($\mathbf{A}_{\mathbf{C}}^{apo}$ and $\mathbf{B}_{\mathbf{O}}^{apo}$), suggesting that the lid opening may play an important role in substrate binding.

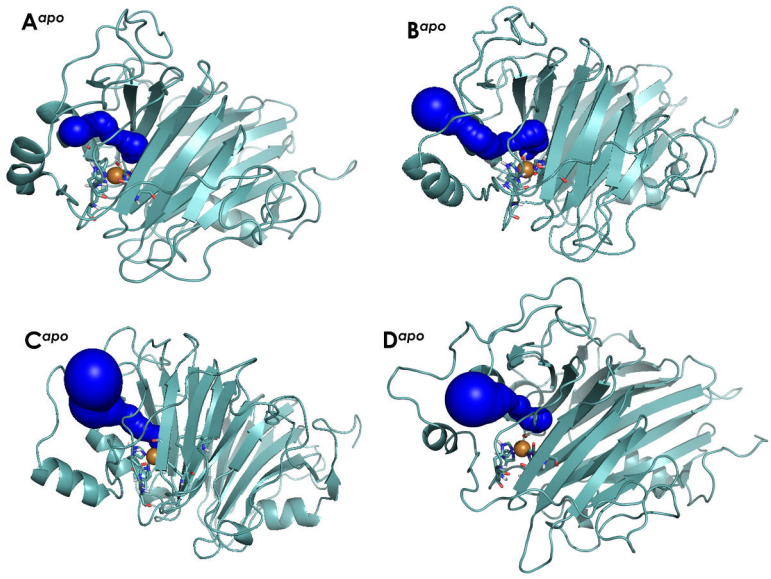


Figure 4.9 Access tunnels calculated with CAVER in the different *apo* states obtained in the PCA.

4. 3. 2 Substrate entrance and product release events

One can ask why in the *apo* state the enzyme is mostly found in an open conformation if through the MD simulation we found that closed conformations can also be stable. To answer this question, a complete reconstruction of the substrate entrance and product release has been performed by means of aMD simulations. Surprisingly, we were able to capture both processes obtaining important information of the conformational dynamics of the enzyme. Regarding the substrate entrance, the lid domain explored different open conformational states while the substrate is surrounding the Que_A, as we have seen before. However, only when the substrate is nearby the lid domain and the enzyme explore open conformations with angles *ca.* 80-90° the substrate enters into the cavity. Interestingly, substrate binding is accompanied by closure of the lid domain (**Fig. 4.10**). This is due to the inherent flexibility of the lid domain of the enzyme. This highlights the importance of the flexible region, however, another key point has been observed. In the wide-open conformation, Pro164 residue is located in a specific conformation not sampled before, allowing substrate entrance (**Fig. 4.10**). Thus, just at the precise moment that the substrate enters, the flexible region of the lid together with Pro164 change to a closed conformation pushing the substrate directly to the active site (right in **Fig. 10**). This allows sampling similar orientations as shown in the X-ray structure of the ES

complex (**Fig. A.4**). This fact is due to the hydrophobicity of the substrate, which forces the release of the water molecules present in the cavity allowing the interactions for closed stable conformations. However, since Glu71 is forced to be coordinated to the metal and may have to deprotonate a hydroxyl group of the KMP substrate to bind the copper metal atom and form the E-S complex, it has not achieved a fully E-S complex state.

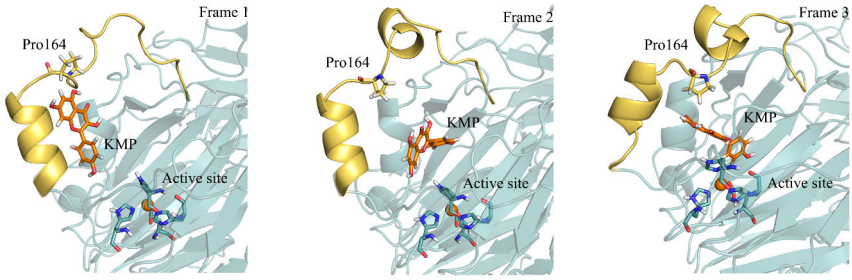


Figure 4.10 Representative snapshots of the aMD simulation for the KMP entrance within the enzyme cavity.

After substrate binding, Que_A catalyzes the reaction in the active site in the presence of the copper metal atom to form the product. This product loses the planarity of the KMP substrate incorporating new rotations in the newly formed molecule. The loss of planarity disrupts some of the interactions previously formed with the substrate, which at the same time produces the destabilization of the lid domain. In our aMD simulations, the product remains in the cavity until the lid domain is able to explore a sufficiently open conformation to release the product from the inner cavity of the enzyme (**Fig 4.11**). Theses simulations explain that in the presence of the substrate the lid is stabilized due to the stabilization of the closed conformation, while the release of the product induces the destabilization of the lid in the Que_A due to the unbinding process.

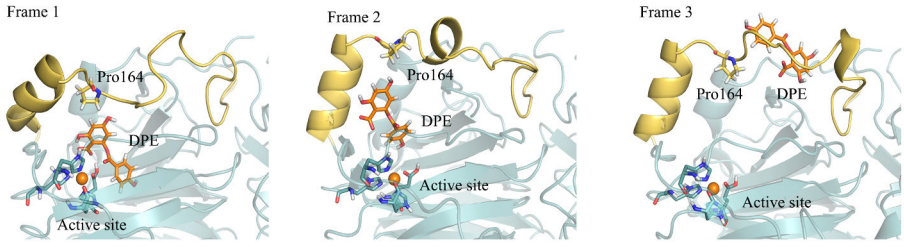


Figure 4.11 Representative snapshots of the aMD simulation for the product (DPE) release within the enzyme cavity.

Chapter 5. Conversion of D-sialic acid aldolase into L-KDO aldolase

This chapter is based on the publication:

Romero-Rivera, A; Iglesias-Fernández, J.; Osuna, S. Exploring the conversion of a D-sialic acid aldolase into a L-KDO aldolase. *Eur. J. Org. Chem.*, **2018**, 2603-2608.

5.1 State-of-the-art

One of the most fundamental reactions in organic chemistry and biology is the carbon-carbon bond formation or cleavage. In Nature it is possible to find aldolase enzymes in many pathways of central and secondary metabolism. These enzymes catalyze the addition of a ketone to an aldehyde to form a new carbon-carbon bond. The capability to perform the chemical reaction under mild conditions together with high stereoselectivity makes aldolases valuable for chemical synthesis.[282-285] Many protein engineering efforts have been devoted to improve the efficiency, reverse selectivity and broaden the substrate scope of aldolase enzymes.[282,286-288]

In this thesis in particular we were interested in D-sialic acid (N-acetyl-D-neuraminic acid) aldolase, which catalyzes the reversible aldol reaction of N-acetyl-D-mannosamine (ManNAc) and pyruvate to produce D-sialic acid. The enzyme is a class I aldolase (i.e. the enzyme forms a Schiff base intermediate by a lysine residue) that follows a multi-step sequential Bi-Uni kinetic mechanism: firstly, the binding of pyruvate to form an enamine Schiff base intermediate occurs, followed by ManNAc binding to generate the imine Schiff base intermediate, which after hydrolysis releases D-sialic acid product.[289,290]

Responsible for the Schiff base formation with the substrates in D-sialic acid aldolase is the catalytic Lys165, and a catalytic triad composed by Tyr137 and Ser47 from one monomeric unit, and Tyr110 from a second subunit. QM/MM-MD simulations together with site-directed mutagenesis confirmed that Tyr137 has a key role to act as a proton donor to the aldehyde oxygen of ManNAc for carbon-carbon formation.[290] The mutation of Ser47 to alanine or cysteine disrupts the hydrogen bond network, decreasing the activity drastically compared to the wild-type whereas the steric bulk of position 110 was crucial for maintaining Ser47 close to Tyr137 to stabilize its deprotonated form after catalysis.[290] The study of the reaction mechanism showed that the carbon-carbon bond formation step presents an activation barrier of *ca.* 6-7 kcal/mol when Tyr137, Ser47, and Tyr110 maintain the catalytically competent hydrogen bond network.[290] QM/MM-MD simulations also suggested that ManNAc binding mode is the main contributing factor to the high stereoselectivity of the enzyme.[290,291]

Many different crystallographic structures have been reported containing pyruvate,[292] pyruvate's analogues[293] or D-sialic acid,[294] and in the case of the inactive Tyr137Ala variant an imine D-sialic-pyruvate-Lys165 covalent Schiff base intermediate was found as a consequence of the reaction taking place during crystallization.[290] Directed evolution (DE) was applied to convert D-sialic acid aldolase into an efficient L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase (**Fig. 5.1**).[286] As observed in many DE studies,[46,295] the final variant (PDB 3LCW)[288] presented eight mutations occurring at distal positions from the active site (the mean distance between the alpha carbon of all mutations and the catalytic Lys165 is *ca.* 21 Å, **Fig. 5.1**). This evolved variant exhibited a >1000-fold improved ratio of the specificity constant ($[k_{cat}/K_M$(L-KDO)]/[k_{cat}/K_M(D-sialic acid)]). Several crystal structures of L-KDO aldolase (8 mutations including Val251), singly mutated variants (V251I, V251L, V251R, V251W), and the doubly mutated V251I/V265I enzyme were obtained.[288] Interestingly, mutation of position 251 located 19 Å from Lys165 was found to be key for accommodating substrates of varying size and still retaining stereoselectivity. Although QM/MM studies have been performed to elucidate the reaction mechanism of D-sialic acid aldolase,[290] the conformational dynamics of the enzyme (and variants) and how the introduced mutations affect the conformational landscape of the enzyme and its substrate specificity still remain unexplored.

The main goal of this study is the evaluation of the conformational dynamics of the natural D-sialic acid aldolase, and the DE-evolved variant that exhibits a high selectivity towards L-arabinose sugar. The summary of our results suggested that the change in specificity is due to a shift in the population of the conformational states, which present different active site volumes and access tunnels better suited for accommodating the different substrate's size.

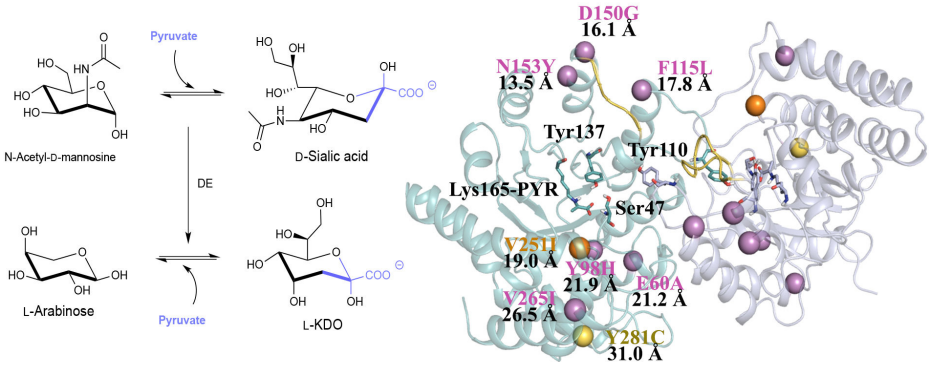


Figure 5.1 On the left, reaction representation of D-sialic acid aldolase to L-KDO aldolase. On the right, distal mutations introduced shown as spheres together with the distance between Cα of the mutation and the catalytic Lys165 in Å and loop 10 highlighted in yellow.

5.2 Computational details

MD simulations of the wild-type (PDB 3LBM) and the last evolved variant enzyme (PDB 3LCW) in explicit water were performed using the AMBER 16 package.[272] Production trajectories were accumulated for 1.5 μs, and were analyzed using the *cpptraj* module included in the AMBER 16 package. Amber 99SB force field (ff99SB) was used, and parameters for the Lys165-HYD were generated using the general AMBER force field (GAFF),[275] with partial charges set to fit the electrostatic potential generated at the B3LYP/6-31G(d) level by the restrained electrostatic potential (RESP) model.[204] Amino acid protonation states were predicted using the H++ server (http://biophysics.cs.vt.edu/H++). Water molecules were treated using the SHAKE algorithm, and long-range electrostatic effects were considered using the particle mesh Ewald method.[278] The Langevin equilibration scheme was used to control and equalize the temperature with a 2 fs time step at a constant pressure of 1 atm and temperature of 300 K, and an 8 Å cutoff was applied to LJ and electrostatic interactions. Production trajectories were accumulated for 1.5 μs, and were analyzed using the *cpptraj* module included in the AMBER 16 package. PCA analysis was done with the *pyEMMA* software.[279]

Substrate access tunnels were analyzed using the standalone version of CAVER 3.0 software.[280] 100 snapshots were randomly selected for each conformational state of D-sialic acid and L-KDO aldolases coming from the cumulative 1.5 μs MD simulations.

Transport tunnels were identified using a probe radius of 0.9 Å. Redundant tunnels were automatically removed from each structure, and the remaining tunnels were clustered using a threshold of 12.

CAVERDOCK a modified version of the AutoDock Vina molecular docking algorithm[296] was used to compute the substrate binding pathway and energy profile along access tunnels. A single enzyme structure for the D-sialic and L-KDO aldolases was used for the calculations, which was based on the most populated clustered conformation from the MD simulation. Access tunnels were calculated with the CAVER 3.0 software as detailed before. ManNAc and L-arabinose substrates in linear and cyclic forms were used as ligands. CAVERDOCK performs a constrained docking to search for a conformation in the vicinity of the previous docking structure, setting the maximal allowed movement between conformations to 0.8 Å.

5.3 Results and discussion

5.3.1 Conformational dynamics of D-sialic acid and L-KDO aldolases

The crystallographic structure of the L-KDO variant complexed with pyruvate (PDB 3LCW) was taken as initial structure, whereas for D-sialic acid aldolase pyruvate was docked in the *apo* state crystallographic structure (PDB 3LBM). Thereafter, we performed MD simulations to evaluate the conformational dynamics of the wild-type (D-sialic acid) and most evolved variant (L-KDO) with pyruvate covalently bound to the catalytic Lys165. Structural differences along the 1.5 μs MD simulations for the D-sialic acid and L-KDO aldolases were analyzed using PCA taking into account the distances between pairs of Cα for all residues (**Fig. 5.2**).

D-sialic acid aldolase sampled two main conformational states equally populated (**1^{D}** and **2^{D}**, **Fig. 5.2**). In **1^{D}**, the distances between the enamine Schiff base pyruvate intermediate and the catalytic Ser47, and Tyr137 in both active sites are short (3.0 ±0.4 Å and 4.8 ±1.6 Å, respectively for monomer 1 (A in **Fig. 5.3**). While, Tyr110'^{B} from the other subunit is located at *ca.* 7-9 Å from Tyr137 and Ser47 residues (B in **Fig. 5.3**). On the other hand, conformational state **2^{D}** maintains the hydrogen bond network between all catalytic residues Ser47, Tyr137, and Tyr110'^{B} (C and D in **Fig. 5.3**) as observed in

the X-ray structure. QM/MM studies by Mulholland and Berry showed that lower activation barriers of 6.2-7.7 kcal/mol are obtained when the latter hydrogen bond network takes place.[290] The transition of $\mathbf{1}^{\mathbf{D}}$ to $\mathbf{2}^{\mathbf{D}}$ requires the correct positioning of Tyr110' from *ca.* 7 to *ca.* 3 Å of Ser47. This conformational change of Tyr110' induces a major change in loop L10 (residues 144-148). Interestingly, the latter loop connects the catalytic Tyr137, which is part of a distorted β-sheet (residues 133-138) in monomer 1, with an α-helix (residues 149-156) where two of the DE mutations (N153Y and D150G) are located.

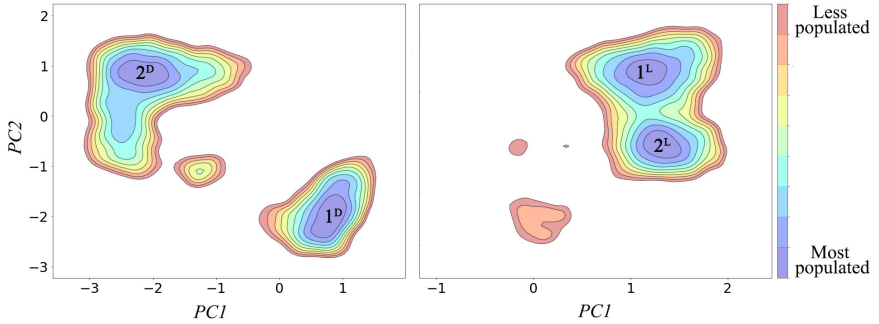


Figure 5.2 PCA representation of D-sialic acid aldolase ($\mathbf{1}^{\mathbf{D}}$ and $\mathbf{2}^{\mathbf{D}}$) and L-KDO aldolase ($\mathbf{1}^{\mathbf{L}}$ and $\mathbf{2}^{\mathbf{L}}$).

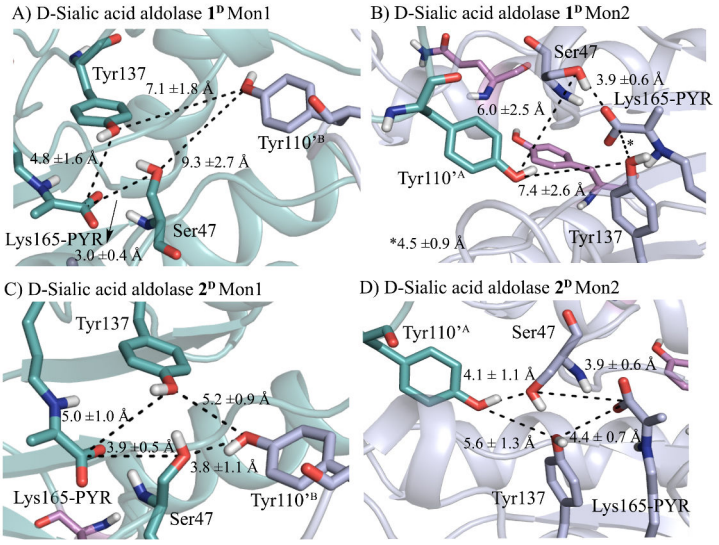


Figure 5.3 Representative snapshots of $\mathbf{1}^{\mathbf{D}}$ (A) and $\mathbf{2}^{\mathbf{D}}$ (B) with the active site residues average distances for D-sialic acid aldolase. Active site residues highlighted in blue and green, with mutated residues in pink.

L-KDO aldolase also displays two main conformational states that are equally populated in solution (**1**^{L} and **2**^{L}, **Fig. 5.2**). In **1**^{L}, Tyr110'^{A} and Tyr110'^{B} are completely displaced from their respectively active sites (at *ca.* 11 Å of Ser47) and are hydrogen bonded at the subunit interface. Nevertheless, Ser47 and Tyr137 remain close to the enamine Schiff base pyruvate intermediate in both active sites. The importance of Tyr110' for the catalytic activity of the enzyme was also evaluated by Mulholland and Berry.[290] The mutation of Y110A resulted in a 40-fold decrease of activity, while the activity was retained after mutation of Y110F, suggesting that the sterically bulkier residue is important for maintaining Ser47 close to Tyr137. Even though Tyr110'^{B} is distal to Ser47 in **1**^{L}, both Ser47 and Tyr137 are situated close to the enamine Schiff base pyruvate intermediate. In conformational state **2**^{L}, Tyr110'^{A} and Tyr110'^{B} are situated closer to the active site residues with a distance of *ca.* 5 Å of Ser47. Short distances are also observed between pyruvate and both Ser47 and Tyr137 (**Fig. 5.4**). Our MD simulations suggest that D-sialic acid and L-KDO aldolases display quite different conformational states, which might be responsible for their different specificity towards N-acetyl-D-mannosamine (ManNAc) and L-arabinose substrates, while maintaining the catalytically active pre-organization of active site residues for catalysis as shown by **2**^{D} and **2**^{L}.

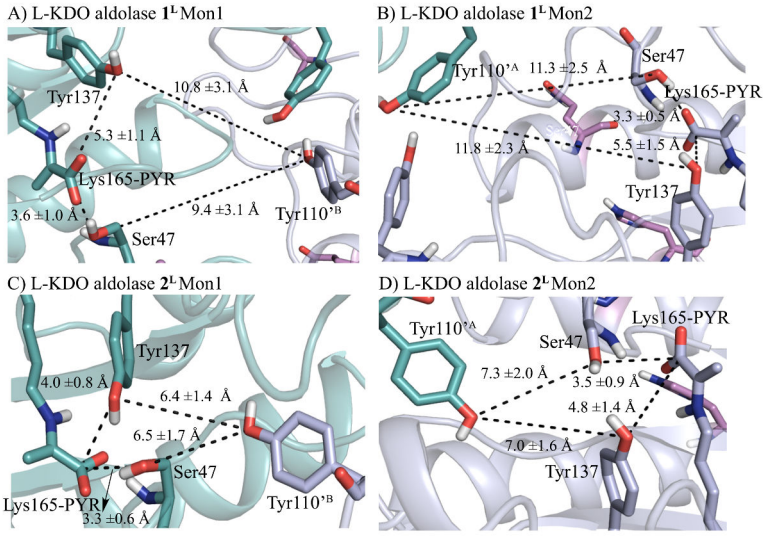


Figure 5.4 Representative snapshots of **1**^{L} (A) and **2**^{L} (B) with the active site residues average distances for L-KDO. Active site residues highlighted in blue and green, with mutated residues in pink.

5.3.2 Substrate access tunnels analysis

The mutations introduced through DE induce a shift in the conformational states that exist in solution for the most evolved aldolase, as compared to the wild-type enzyme. We have evaluated the active site shape to elucidate the reasons for the experimentally observed change in specificity, as well as substrate access tunnels of the different conformational states sampled by both enzymes (**Fig. 5.5** and **5.6**).

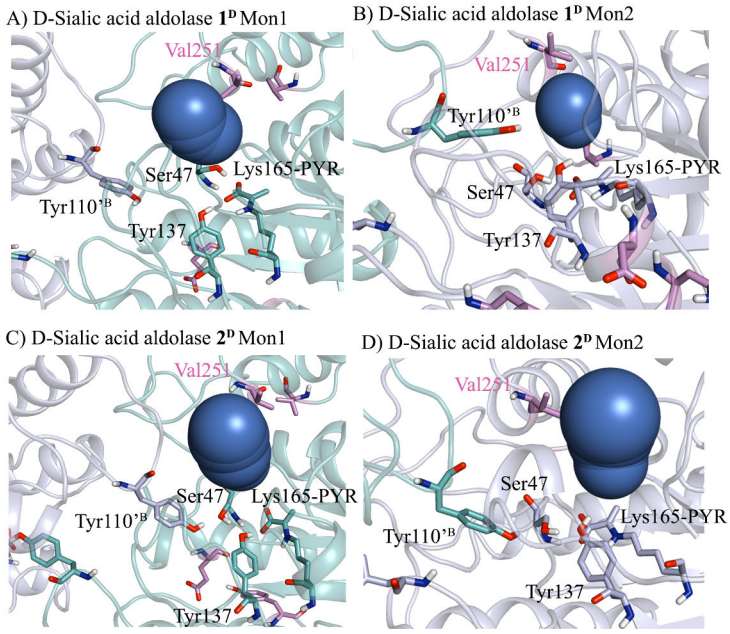


Figure 5.5 Principal substrate access tunnel (in blue) in the most populated cluster of $\mathbf{1^D}$ (A and B) and $\mathbf{2^D}$ (C and D). Active site residues highlighted in blue and green, with mutated residues in pink. Monomers 1 and 2 abbreviated as Mon1 and Mon2, respectively.

The analysis of the principal substrate access tunnel for the conformational state $\mathbf{1^D}$ of D-sialic acid aldolase indicates that the displacement of Tyr110' results in a wider active site, with a bottleneck radius of 3.5 ±0.5 Å for monomer 1 and 2.9 ±0.3 Å for monomer 2 (A and B in **Fig. 5.5**). The access tunnel is slightly narrower in $\mathbf{2^D}$ when Tyr110' is properly positioned in the active site (bottleneck radius is 3.4 ±0.4 Å and 2.2 ±0.5 Å for both monomers respectively, **Fig. 5.5**). However, Val251 is close to the narrow region of the substrate access tunnel, suggesting that the conformation of Val251 will have a greater effect on the bottleneck radius of the tunnels. Thus, amino acid changes at position 251 will modulate the bottleneck radius of the access tunnel, changing the prefer-

ence for the substrate specificity of the enzyme. Indeed, singly mutated variants at position 251 were found to accommodate substrates of varying size, while still retaining the same stereoselectivity.[288]

L-KDO aldolase conformational state $\mathbf{1}^{\mathbf{L}}$ presents both Tyr110' far away from their respective active sites. The Tyr110' conformational change also involves loop L10 (residues 144-148), which in turn can modify the active site shape and substrate access tunnel. Computed bottleneck radius for L-KDO monomer 1 and 2 are 2.2 ±0.4 Å and 2.1 ±0.3 Å, respectively (A and B in **Fig. 5.6**). Conformational state $\mathbf{2}^{\mathbf{L}}$ presents both Tyr110' close to L-KDO active site residues, and presents a bottleneck radius of 2.3 ±0.3 Å for monomer 1 and 1.9 ±0.3 Å for monomer 2 (C and D in **Fig. 5.6**). Since the influence concerning Tyr110' farther or closer to the active site only slightly changes the different tunnels, we suggested that mutation on residue 251 is the major contributor to modify the access tunnels without altering neither the active site machinery nor the stereoselectivity.[288]

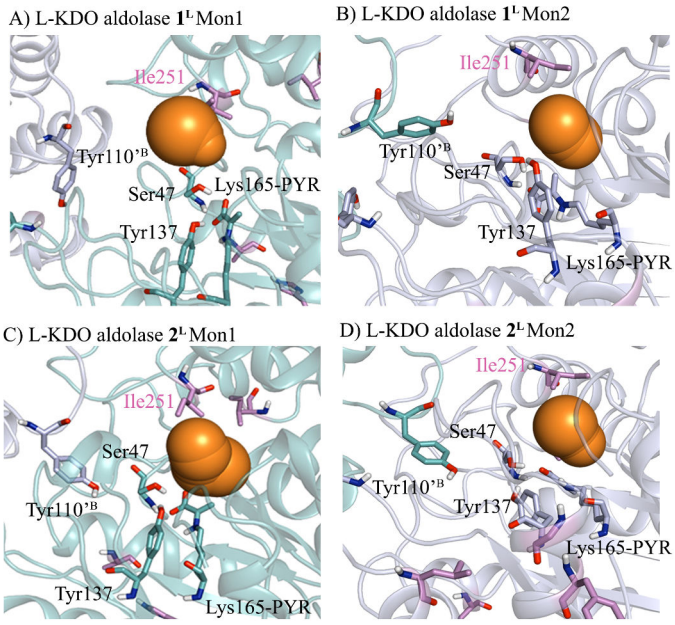


Figure 5.6 Principal substrate access tunnel (in orange) in the most populated cluster of $\mathbf{1}^{\mathbf{L}}$ (A and B) and $\mathbf{2}^{\mathbf{L}}$ (C and D). Active site residues highlighted in blue and green, with mutated residues in pink. Monomers 1 and 2 abbreviated as Mon1 and Mon2, respectively.

Our MD simulations coupled to active site analysis with CAVER indeed confirm that the L-KDO variant explores conformational states that present substrate access tunnels substantially narrower thus restricting its substrate accessibility to small substrates. These differences in the active site shape and accessibility arise from changes in the conformational landscape of the enzyme due to the introduced mutations located all over the protein, which is the main contributing factor for the alteration of enzyme specificity towards increasing the efficiency for smaller substrates.

5.3.3 Cyclic and lineal substrate binding energies

To further confirm the hypothesis of enzyme active site accessibility as the main factor driving substrate specificity, we computed the substrate binding pathway and energy profile for D-sialic acid and L-KDO aldolases with both substrates in the cyclic and lineal shape using Caverdock. The natural substrate ManNAc binds more tightly to the active site than L-arabinose for D-sialic acid aldolase with an energy difference of *ca.* 1.3 kcal/mol (**Fig. 5.7**). The more favorable binding energy is in agreement with the smaller K_M values reported experimentally for the reaction with ManNAc as compared to L-arabinose.[288] Ser208 makes a hydrogen bond with the amide group of ManNAc and contributes to the proper positioning of the natural substrate in the wide active site (**Fig. 5.7**).

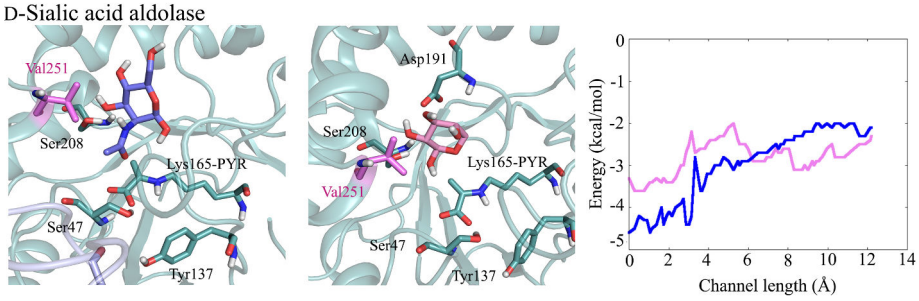


Figure 5.7 Representation of the most relevant hydrogen bond interaction of the natural ManNAc (left, blue) and L-arabinose (middle, light pink) substrate, and their respective binding energy (in kcal/mol) along the access tunnel (in channel length, Å). Active site residues highlighted in blue and green, with mutated residues in pink.

The L-KDO variant shows statistically identical values of the K_M parameters for both substrates, in agreement with the computed energy difference of only 0.3 kcal/mol. It is worth to mention that for the evolved variant the natural substrate ManNAc is not able

to completely access the active site as a consequence of the reduced size and narrower bottleneck radius of the access channel, thereby explaining its preferred selectivity towards the smaller L-arabinose substrate. The mutation V251I in the L-KDO variant forces the ring of L-arabinose to adopt a different conformation due to the decreased size of the bottleneck. This allows the sugar to interact with Ser208, thus slightly decreasing the binding energy (**Fig. 5.8**).

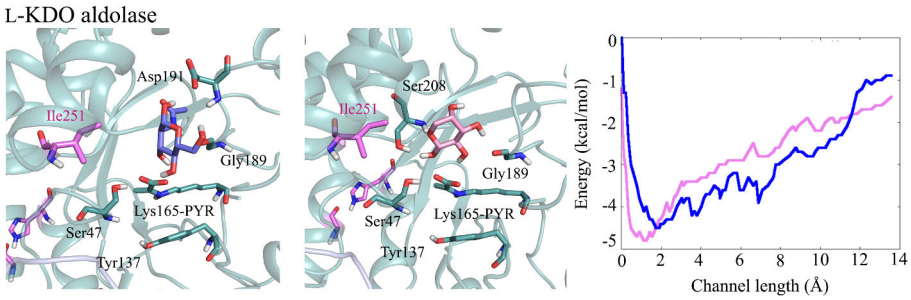


Figure 5.8 Representation of the most relevant hydrogen bond interaction of ManNAc (left, blue) and L-arabinose (middle, light pink), and their respective binding energy along the access tunnel. Active site residues highlighted in blue and green, with mutated residues in pink.

The first step in the aldol reaction is the ring-opening of sugar substrates that was proposed to occur in the enzyme active site, favoring the linear open form.[294] Energy profiles computed for the open-ring sugars using the same methodology, also match the experimentally reported catalytic activities. The linear open ManNAc substrate displays a more favorable binding energy in D-sialic acid aldolase as compared to open L-arabinose, as a result of the large number of interactions formed that position the substrate for catalysis (**Fig. 5.9**). This concurs with the higher k_{cat} of the natural enzyme towards ManNAc. Calculations also indicate that the linear form of the sugar structures is not affected by the narrow L-KDO aldolase tunnel, as both open L-arabinose and open ManNAc molecules can access the active site. In this case, specificity of L-KDO aldolase towards L-arabinose can be explained by the slightly higher stability of the latter, by approximately 0.5 kcal/mol (**Fig. 5.9**).

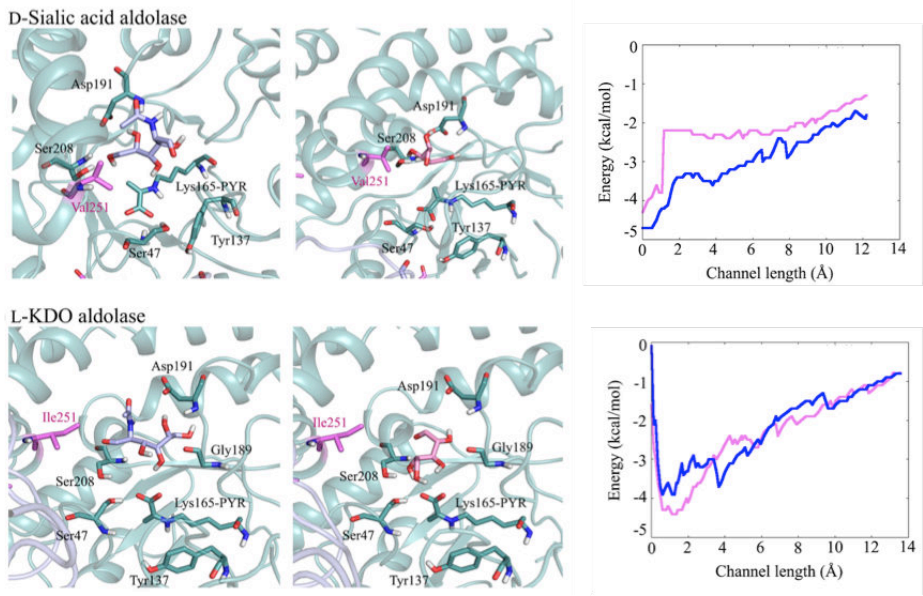


Figure 5.9 Representation of the most relevant hydrogen bond interaction of lineal ManNAc (left, light blue) and L-arabinose (middle, light pink) substrates, and their respective binding energy within the access tunnel for D-sialic acid (top) and L-KDO aldolase (bottom). Active site residues highlighted in blue and green, and mutated residues in pink.

Chapter 6. Conformational dynamics in the evolution of retro-aldolase activity

This chapter is based on the publication:

Romero-Rivera, A; Garcia-Borràs, M.; Osuna, S. Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catal.*, **2017**, *7(12)*, 8524-8532.

6.1 State-of-the-art

Retro-aldolases (RA) that cleave the methodol substrate (**Fig. 6.1**) are, from a mechanistic perspective, the most complex computationally designed enzymes to date,[43,60,87] as-mentioned in the introduction. They follow a multi-step reversible pathway involving covalent Schiff base intermediates to catalyze the carbon-carbon cleavage. Many experimental and computational analyses of initial computational designs revealed that the low efficiency was due to the lack of precision and specificity towards accomplishing good interactions with the methodol substrate (**Fig. 1.4 and 6.1**).[295,297] Specific activity increased more than 4400-fold by further DE engineering of the computationally designed RA95.[89] As explained in sub-section 1.2.2, the increase in catalytic efficiency was mainly due to an improvement in k_{cat}, which was principally achieved via mutations located all around the enzyme structure (**Fig. 6.1**). During the evolution path towards the most efficient variant these mutations located all over the protein caused pronounced molecular changes in the active site and in the conformational dynamics of the enzyme variants.

One of the important changes that occurred during the DE evolution was the abandoning of Lys210 in favor of a new lysine residue on the opposite site of the binding pocket.[89] The variant RA95.5-8 presents a 7200-, 60-, 2-fold higher activity than RA95.0, RA95.5, and RA95.5-5, respectively (**Table 6.1**). Notwithstanding, further mutations performed by Hilvert and coworkers evolved RA95.5-8 into a highly active RA95.5-8F variant using a new ultrahigh-throughput droplet-based microfluidic screening platform.[90] This high activity was achieved thanks to the emergence catalytic tetrad composed by Tyr51, Lys83, Asn110, and Tyr180 along the DE pathway.

Interestingly, the generated RA95.5-8 variant showed promiscuous activity for carbon-carbon bond formation catalyzing asymmetric Michael additions of carbanions to unsaturated ketones,[93] Knoevenagel condensations of electron-rich aldehydes and activated methylene donors,[94] and synthesis of γ-nitroketones.[298] One of the important aspects that make enzymes evolvable is its dynamic nature.[299] The capability to explore different thermally accessible conformations, which are important for substrate binding, regulation, inhibition, and product release, allows the enzymes the ability to catalyze

different reactions granting the promiscuity. Thus, the inherent flexibility of RA95.5-8 may be the main reason behind the promiscuity of this variant toward certain types of reactions, in this case, the carbon-carbon bond formation reactions. Hilvert and coworkers evolved RA95.5-8 into a more efficient variant for the Michael addition reaction for unsaturated ketones with only two mutations.[93] They even evolved the RA95.5-8 variant to control the stereoselectivity of the Michael addition reaction. Those mutations to control the stereoselectivity were located in the binding pocket, specifically, the I133G mutation is supposed to induce a conformational change of the Schiff base intermediate to favor the opposite stereoselectivity found in RA95.5-8 for the Michael addition reaction.[92]

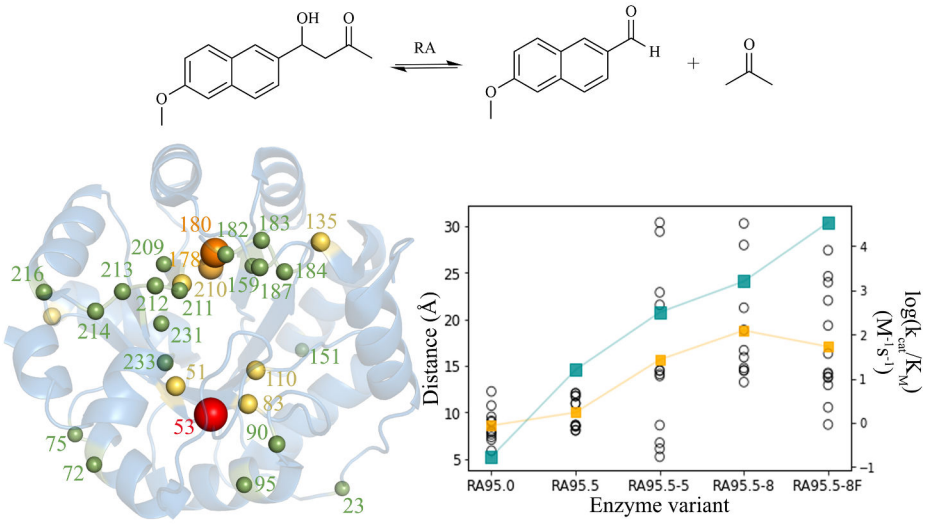


Figure 6.1 On the left, representation of the mutation sites introduced by DE. Sphere sizes and color are weighted according to the number of times the position was mutated along the evolutionary pathway. On the right, representation of the catalytic efficiencies of the RA95 variants (in teal, right axis in log kcat/K_M $M^{-1}s^{-1}$), together with the distances between the Cα of the different mutations (black spheres in Å) and the catalytic lysine (nitrogen atom of side-chain). The mean mutation distance is marked with a dark yellow square.

The goal of this study is the evaluation of the active site conformational dynamics of the different RA variants and the Schiff base intermediate for the methodol cleavage and the Michael addition reaction for RA95.5-8 and I133G RA95.5-8. Summarizing the key points of this chapter, we demonstrate how the mutations introduced evolved the enzyme towards better pre-organizations of the active site to perform the catalysis of

methodol cleavage during the DE pathway. For the Michael addition promiscuous reaction, we were able to rationalize the change of conformation induced of the covalent intermediate due to the single mutation of I133 by glycine. Interestingly, we also demonstrate that MD simulations coupled to residue-by-residue correlation can identify the residues involved in the population shift toward the catalytically active states for the evolved variants, which coincide with DE mutation points.

6.2 Computational details

MD simulations in explicit water were performed using the AMBER 16 package.[272] Production trajectories were run for 1000 ns and were analyzed using the cpptraj module included in the AMBER 16 package. The Amber 99SB force field (ff99SB) was used, and parameters for the Schiff base intermediates (**Int^{RA}** and **Int^{MA}**) were generated using the general AMBER force field (GAFF),[275] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model.[204] Amino acid protonation states were predicted using the H++ server (http://biophysics.cs.vt.edu/H++). Water molecules were treated using the SHAKE algorithm, and long-range electrostatic effects were considered using the particle mesh Ewald method.[278] The Langevin equilibration scheme was used to control and equalize the temperature with a 2 fs time step at a constant pressure of 1 atm and temperature of 300 K, and an 8 Å cutoff was applied to LJ and electrostatic interactions. PCA analysis was done with the pyEMMA software.[279]

The shortest path map (SPM) analysis performed in this chapter relies on the construction of a graph based on the computed mean distances and correlation values, in a similar way as has been done in allostery for previous studies.[300,301] The graph obtained is then further simplified by making use of the Dijkstra algorithm as implemented in the igraph module[302] to identify the shortest path lengths. The algorithm goes through all nodes of the graph and identifies which is the shortest length path to go from the first until the last protein residue. The method therefore identifies which edges of the graph are shorter (i.e. more correlated) and are more central for the communication pathway. Only those edges having a higher contribution are represented, and they are weighted according to their contributions.

6.3 Results and discussion

6.3.1 Evolution of active site conformational dynamics

Our analysis begins with the exploration of the RA95 active site conformational dynamics to evaluate the molecular changes induced by mutations located nearby and far from the active site.[89,90] We analyzed the key catalytic distances in both *apo* and Schiff base intermediate states sampled along the long-time-scale MD simulations, together with representative overlays of the structures visited during the simulation (**Fig. 6.2**). The available X-ray structures of the RA95 variants with a diketone inhibitor bound are also shown to compare the different conformations sampled.

The computational design RA95.0, created from a $(\beta\alpha)_8$ barrel scaffold (16 mutations on PDB 1LBL), contains a Glu53 residue that may be responsible for activating a water molecule promoting the carbon-carbon bond cleavage at the Schiff base intermediate. Our MD simulations show that *ca.* 11 water molecules are present around the catalytic Lys210 in the *apo* state (**Fig. B.1**). We observe a rather long distance between Glu53 and the catalytic Lys210 (7.1 ±2.8 Å at the Schiff base intermediate; **Fig. 6.2**), suggesting a minor role of Glu53 in positioning a pre-organized water molecule ready for catalysis. Indeed, our observations are in agreement with the reported low activity of the computational RA95.0 variant and the fact that full activity is retained after mutating E53A.[89]

The initial RA95.0 computational design was further evolved through DE to RA95.5 with a total of six mutations introduced.[89] In this case a second lysine appeared at position 83, which in the case of the most evolved RA variants was found to be more effective than Lys210 as a catalytic group under normal turnover conditions.[89] Glu53 was replaced by serine (E53S), and a new tyrosine residue (V51Y) appeared for promoting the β-alcohol deprotonation (**Fig. 6.2**). The MD simulations of the Schiff base intermediate in the catalytic Lys83 show two clear conformations of the L6 loop (residues 180−190), forcing two different orientations of the intermediate (**Fig. 6.2**). The positions of these conformations coincide with the two possible binding modes of the inhibitor (one bound to position 83 and another to position 210) observed in the RA95.5 X-ray structure. The high flexibility of the Schiff base intermediate leads to a poorly pre-organized active

site, with a catalytic distance of *ca.* 5.0 ±1.4 Å between the base (Y51) and the β-alcohol (**Fig. 6.2**). The mean hydrogen bond angle measured is *ca.* 74 ±52° (**Fig. B.3**); therefore, the ideal hydrogen bond conformation (180°) is hardly sampled along time. The active site of the enzyme is solvent exposed with *ca.* 10 water molecules around Lys83 (**Fig. B.2**), which indicates that water could also play a crucial role in assisting the reaction for the catalysis on RA95.5 variant.

RA95.5-5, generated by additional rounds of mutagenesis and plate screening, has six additional mutations, three of them at the enzyme active site (E53T, S110N, and G178S), and the rest on the protein surface (R23H, R43S, and T95M). The X-ray of this variant in the presence of a diketone inhibitor confirmed that it is exclusively bound at Lys83 (**Fig. 6.2**). Our MD simulations revealed that the covalent intermediate is still not in a proper position for catalysis (distance between the base and the β-alcohol is *ca.* 5.4 ±1.1 Å and the angle is *ca.* 73 ± 30°, **Fig. 6.2** and **B.3**).

A further evolved variant, RA95.5-8, contains five additional mutations: three near the active site (S178 V, K135N, and G212D), and two distal (S43R, F72Y). This new variant is *ca.* 60-fold more active than RA95.5, mainly due to an increase in k_{cat} rather than K_M. The MD simulations performed indicate that the distance between Lys83 and Tyr51 is *ca.* 4.6 ±0.8 Å in the *apo* state (**Fig. B.1**) but *ca.* 5.5 ±1.6 Å for the Schiff base intermediate (**Fig. 6.2**). The hydrogen bond angles for deprotonating the β-alcohol are quite broad, but the average angle (*ca.* 98 ± 36°) is larger than in the other variants (**Fig. B.3**), thus it shows more consistency with the higher catalytic efficiency observed for this enzyme. The Schiff base intermediate is still quite floppy and can occupy two different pockets (**Fig. 6.2**). One of the conformations sampled is similar to that observed for RA95.5-5, as well as in the catalytically productive binding pose observed for the highly evolved RA95.5-8F. However, the latter state is hardly visited along the MD simulation of this variant.

For the highly evolved RA95.5-8F variant, which provides a $>10^9$ rate enhancement, k_{cat} value of 10.8 s^{-1} and a k_{cat}/K_M value of 34000 M^{-1} s^{-1}, shows comparable kinetic values to those of natural class I aldolases.[90] RA95.5-8F contains 13 additional mutations located throughout the whole protein. This enzyme has a catalytic tetrad composed

of the RA95.5-8 triad Lys83, Tyr51, and Asn110, with an additional Tyr180 (due to the F180Y mutation). It is worth mentioning that the original catalytic Lys210 is mutated to leucine. The MD simulations, similar to like the X-ray structure, suggest that the catalytic tetrad is perfectly arranged to bind the substrate and adapt to geometric and electrostatic changes occurring during the mechanistically complex reaction pathway. In the *apo* state, the distance between Tyr51 and Lys83 is *ca.* 4.6 ± 0.9 Å but *ca.* 7.2 ± 1.4 Å between Tyr180 and Lys83 (**Fig. B.1**). For the Schiff base intermediate, we also investigated the possibility of either Tyr51 or Tyr180 acting as the base for deprotonating the β-alcohol. The results indicate that Tyr180 is in a better position for deprotonating the alcohol, as shown by *ca.* 3.1 ±0.8 Å (*ca.* 5.9 ±1.2 Å for Tyr51) distances (**Fig. 6.2**) with an angle close to the ideal 180° (*ca.* 143 ±38° for Tyr51, **Fig. B.3**). This stable and catalytically competent conformation of the Schiff base intermediate is mainly due to the formation of a hydrogen bond between the oxygen atom of the β-alcohol and Tyr51. In fact, when the Y180F mutation is introduced in RA95.5-8F, slightly longer distances (*ca.* 3.6 ±1.3 Å between Tyr51 and the β-alcohol, **Fig B.5**) are visited. However, a more drastic effect is observed in the Y51F variant (*ca.* 4.6 ±2.1 Å for Tyr180 and covalent intermediate, **Fig B.5**). In Y51F, the key hydrogen bond with Tyr51 for maintaining the catalytically competent conformation of the Schiff base intermediate is lost. This is in line with the 4- and 90-fold decreases in k_{cat} observed experimentally for Y180F and Y51F, respectively.[90] The importance of both Tyr51 and Tyr180 for catalysis was also found experimentally, as replacing both residues resulted in a 17000-fold decrease in k_{cat}.[90]

The overlay of some representative MD snapshots with the X-ray structure reveal a highly pre-organized active site, with the Schiff base intermediate properly positioned for catalysis most of the simulation time because of the enhanced hydrogen bond network due to the presence of Tyr180. These results are in line with the recent finding by Hilvert and coworkers that the rate-limiting step of the process in the most evolved RA95.5-8 and RA95.5-8F variants is not the carbon-carbon bond scission (as in RA95.0), but rather the product release step.[91] The Schiff base conformation adopted in RA95.5-8F matches that observed in the X-ray structure, but interestingly, the binding pose is well matched with that of the computational design RA95.0. The decrease in the pKa of the catalytic lysine observed experimentally is in line with the decrease in the

water solvation shell of Lys83 (from 11 to 6), which favors its deprotonated form, and thus facilitates Schiff base formation (**Fig. B.2**).

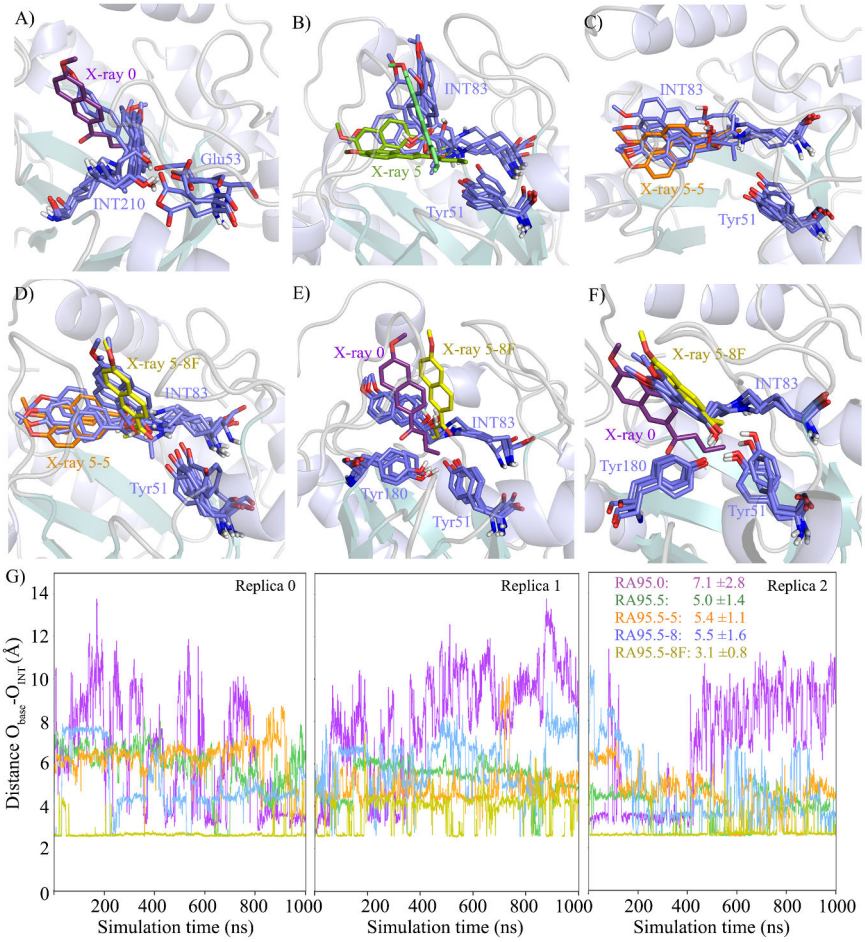


Figure 6.2 Representation of the enzyme active site conformational dynamics: (A) RA95.0; (B) RA95.5; (C) RA95.5-5; (D) RA95.5-8; (E) RA95.5-8F with Tyr51 acting as the base; (F) RA95.5-8F with Tyr180 deprotonated; (G) Distance's plot between the supposed base and the β-alcohol that will be deprotonated along the three 1 μs MD trajectories. Catalytic residues are represented by blue sticks (for visualization purposes Asn110 has not been included). X-ray structures with the diketone inhibitor bound are displayed for: RA95.0 (PDB 4A29, in purple), RA95.5 (PDB 4A2S, in two types of green, lime green for the inhibitor bound to position 83 and light green for position 210), RA95.5-5 (PDB 4A2R, orange), RA95.5-8F (PDB 5AN7, in yellow).

6.3.2 Evolution of conformational dynamics for enhancing RA activity

The evaluation of the enzyme active site conformational dynamics has revealed that distal mutations progressively stabilized the catalytically competent arrangement for catalysis. This is especially true for the highly evolved RA95.5-8F enzyme, thanks to a catalytic tetrad that stays in the catalytically competent arrangement for the entire simulation. The analysis of the structural differences observed along the MD simulations using PCA indicates that a population shift occurred along the evolutionary pathway (**Fig. 6.3**). The first principal component (PC1) is able to distinguish inactive states (those presenting a long distance between the base and the β-alcohol, in red in **Fig. 6.3**) from the catalytically competent states (active, marked in green in **Fig. 6.3**). The main difference between both states arises from the different binding modes of the Schiff base intermediate in the enzyme active site and conformational changes in the flexible loops L1 (residues 52−66), L2 (82−89), L6 (180−190), and L7 (211−215).

The least efficient enzyme (RA95.0) explores conformational sub-states that hardly sample catalytically proficient distances (**Fig. 6.3**). This is observed for all variants from RA95.0 to RA95.5-8. DE mutations progressively stabilize catalytically competent substates (**Fig. 6.3**). The major conformational states in RA95.5-8F have the catalytic Tyr180 properly positioned to deprotonate the alcohol of the Schiff base intermediate. Even more important is that all conformational states show catalytically competent distances for catalysis. Our microsecond time scale MD simulations have thus been able to capture significant differences in the conformational sub-states sampled by the evolved DE variants. Nevertheless, we are aware that much longer or multiple short MD simulations will be needed to capture both kinetics and thermodynamic aspects of this system.[303]

RA95.0 is substantially more flexible than the rest of the variants (**Fig. B.4**), and a gradual reduction in the conformational flexibility of loops L1 (residues 52−66) and L6 (residues 180−190) is observed, RA95.5-8F being substantially less flexible than the other variants (**Fig. B.4**). DE mutations decreased enzyme flexibility and thus increased its thermostability, as observed experimentally.[90]

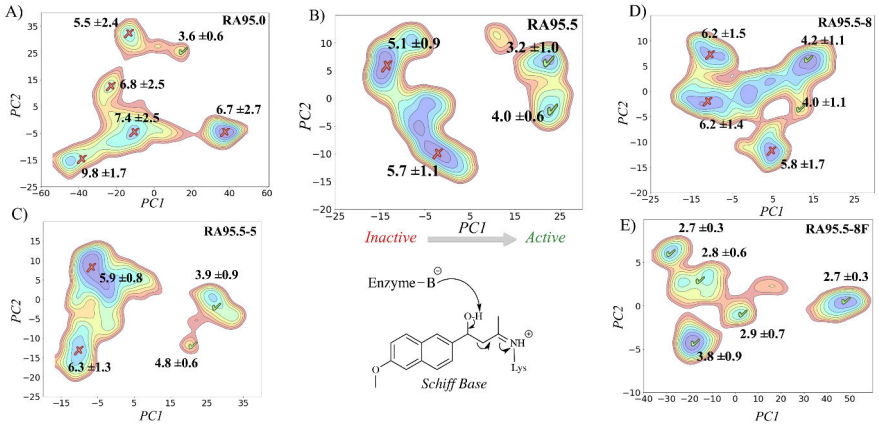


Figure 6.3 PCA representation projected into the two principal components (PC1, PC2) based on Cα contacts for (A) RA95.0, (B) RA95.5, (C) RA95.5-5, (D) RA95.5-8, and (D) RA95.5-8F Schiff base intermediates. For each sub-state, the mean distance between the heteroatom of the base and the oxygen of the Schiff base β-alcohol is represented together with the standard deviation (in Å). Those states exploring distances in the 2.0−4.0 Å range are shown in green (catalytically competent), while the other states are shown in red. PC1 (x axis) differentiates inactive states (low PC1 values, pink structure in (B)) that present long catalytic distances from those properly oriented for the catalysis (high PC1 values, green structure in (B)).

6.3.3 Schiff base intermediate conformational dynamics on RA and Michael addition reactions.

The most evolved variants, RA95.5-8 and RA95.5-8F, have been shown and demonstrated to be highly catalytic for the cleavage of methodol, suggesting enhancement of the catalytic distances between the Schiff base intermediate and its catalytic residues.[304] However, to further understand the catalytic promiscuity exhibited by RA95.5-8, PCA has been done by explicitly taking into account the distances between all atoms of the Schiff base intermediate with all Cα of the protein (**Fig. 6.4**). This PCA clearly shows that the Schiff base intermediate in RA95.5-8 is more flexible than in RA95.5-8F. RA95.5-8 prefers the extended conformation of the covalent intermediate, which is very similar to the conformation shown in the X-ray structure of RA95.5-5 (**A**^{8} in **Fig. 6.4**). This may be due to the nearby residue N135, which facilitates the extended conformation of the Schiff base intermediate avoiding the pointing up conformations as shown in the RA95.5-8F X-ray structure (**C**^{8} in **Fig. 6.4**). Another important feature is that Thr53 residue can form a hydrogen bond interaction with Tyr51 avoiding catalytically competent distances with the β-alcohol of the Schiff base intermediate at the extended conformation in RA95.5-8 (B in **Fig. 6.4**). The 13 mutations introduced in RA95.5-8F

makes it possible to mostly populate conformations where the Schiff base is pointing up as in the RA95.5-8F X-ray structure. Among the 13 residues introduced, we found that some important contributions are made by the Leu53 avoiding the hydrogen bond interaction with Tyr51 and properly positioning of the tyrosine due to a CH···π interaction. The introduced Tyr180 allows a better active site pre-organization for properly stabilizing the TS for the methodol cleavage as suggested above. It is worth to mention that Glu135 may also play an important role in not allowing extended conformations of the Schiff base intermediate (C in **Fig 6.4**).

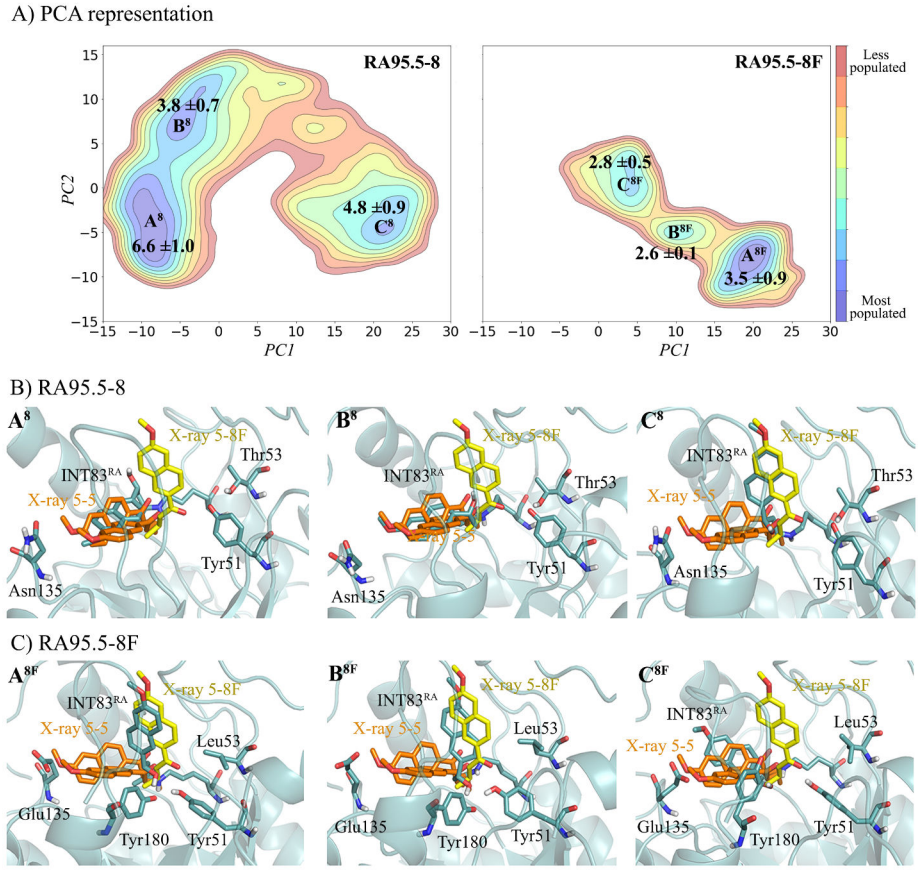


Figure 6.4 A) PCA representation projected into the two principal components (PC1, PC2) based on the atoms of the intermediate compound for RA95.5-8 and RA95.5-8F Schiff base intermediates. For each substate, the mean distance between the heteroatom of the base and the oxygen of the Schiff base β-alcohol is represented together with the standard deviation (in Å). Those states exploring distances in the 2.0−4.0 Å range are shown in green (catalytically competent), while the other states are shown in red. B) Representative

structures of the different minima in the PCA representation for RA95.5-8. C) Representative snapshots of the PCA representation for RA95.5-8F.

We also evaluated the promiscuity observed in RA95.5-8 variant for Michael addition reaction (**Fig. 6.5**). We analyzed the Schiff base intermediate of the Michael addition reaction to shed more light on how key mutations control the conformational sampling of the covalent intermediate. RA95.5-8 is presumed to perform the reaction with high stereoselectivity towards a specific face of the substrate, while a single I133G mutation is responsible for switching the enzyme selectivity towards the opposite stereoselectivity of the Michael addition reaction.[92] PCA has been done on the accumulated 3 μs MD simulations using the covalent intermediate for the Michael addition reaction by taking into account the distances between all atoms of the Schiff base intermediate with all Cα of the protein to differentiate the conformations sampled during the MD simulation (A in **Fig. 6.6**). RA95.5-8 explores a heterogeneous landscape thanks to the inherent flexibility of this variant, while mostly maintaining extended conformations of the Schiff base intermediate for the Michael addition reaction. This extended conformation is also observed in an X-ray structure of an evolved variant in which the same stereoselectivity but higher activity was achieved (B in **Fig. 6.6**). On the other hand, the presence of the I133G mutation liberates the steric space produced by the alkyl chain of isoleucine allowing the new conformation for the opposite stereoselectivity. PCA for I133G shows two main conformations states (C in **Fig 6.6**) that corresponds to the extended Schiff base intermediate conformation pointing towards Gly133 slightly different to the X-ray structure of the other variant, demonstrating how the new space generated by the small amino acid facilitates to adopt a new conformation not explored before due the lack of space ($\mathbf{A}^{\mathbf{8\text{-}MA}}$ and $\mathbf{A}^{\mathbf{I133G\text{-}MA}}$ of **Fig. 6.6** to see the difference orientation). Thus, both RA and Michael addition reactions suggest that the residues enclosed in residue 130-135 of the $(\beta\alpha)_8$ barrel enzyme may control the possible conformation sampled by the Schiff base intermediate of the computational designed enzyme.

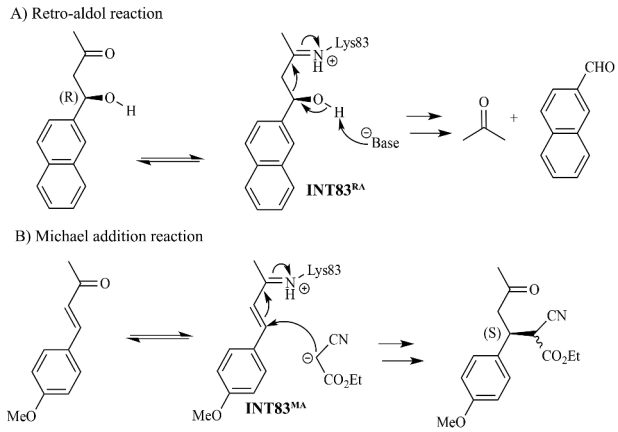


Figure 6.5 Promiscuous reactions observed in RA95.8 variant.

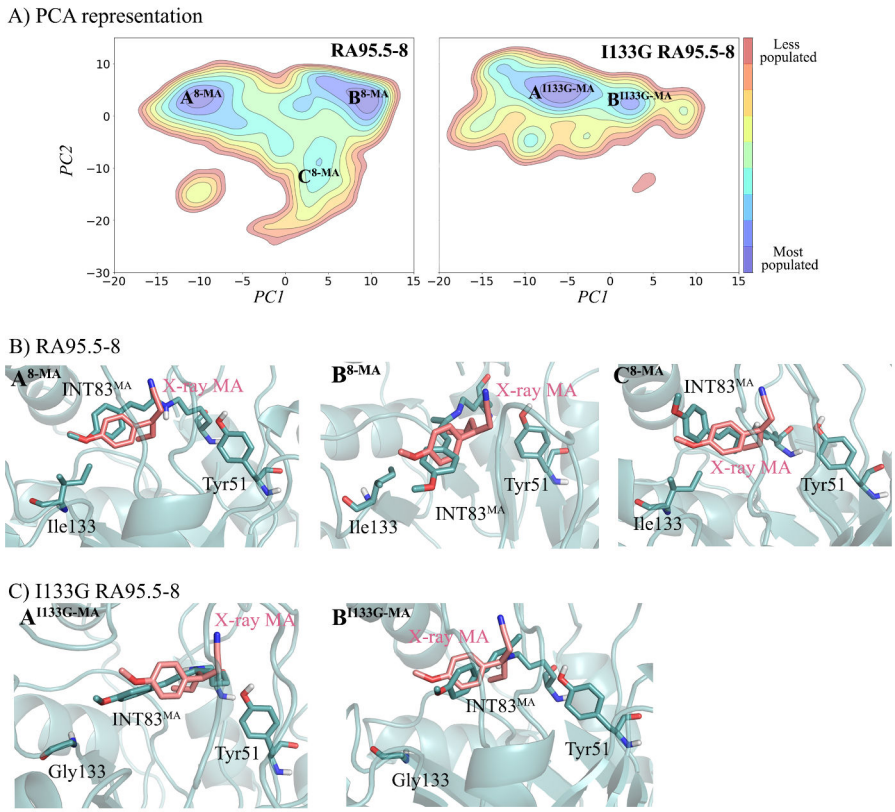


Figure 6.6 PCA representation projected into the two principal components (PC1, PC2) based on the atoms of the intermediate compound for the Michael addition reaction for RA95.5-8 and I133G RA95.5-8 Schiff base intermediates.

6.3.4 Identification of residue pathways for novel function

The abovementioned results have demonstrated that the conformational landscape along evolutionary pathways is significantly altered, with the catalytically active conformational sub-states progressively populated to an increasing extent. This study, together with work on other enzymes,[105,244,305] demonstrates the power of MD for rationalizing DE evolutionary pathways. However, the main question is whether MD can be used as a predictive tool to determine, *a priori*, which changes might enable novel function. By analogy to DE, the latter changes could be located either close to the enzyme reaction center or at distal positions. Although some studies have used MD to propose mutations in the past,[60,67,303] the predictions were always restricted to the active site. As shown in the previous sections, RA enzymes sample different conformational states that are not equally populated and these populations shift upon introduction of mutations. The population shift concept has been used previously to rationalize substrate binding mechanisms and allosteric regulation as discussed in the introduction of this thesis.[7,300]

From a computational perspective, allosteric processes have been evaluated via MD to generate molecular ensembles coupled with community network analysis, identifying highly connected clusters residues based on residue-by-residue correlation and proximity.[300,306] Given the high similarity of both processes, we hypothesized that correlation-based measures might be useful in the enzyme design field as well. For this reason, our group developed a new tool for enzyme design that explores residue-by-residue correlated movements and inter-residue distances as shown in previous allosteric studies.[300,301,307] This generates a complex graph based on proximity and correlation. Further evaluation of the latter graph is done making use of the Dijkstra algorithm as implemented in the igraph module[302] to identify the shortest path lengths, and thus generating a map that we call the shortest path map (SPM). The main purpose is the identification of pair residues that have a higher contribution to the communication pathway. We applied this tool to the accumulated 3 μs MD simulations performed for the studied enzyme variants and compared the SPMs obtained with the DE mutation sites (**Fig. 6.7** and **6.8**).

When the structures were compared, it was found that the residues identified by SPM and the positions mutated experimentally during DE are strikingly similar. Interestingly, SPM can be constructed along the DE process to track which positions could be mutated for pursuing novel RA activity. In RA95.0, five out of six DE mutations are included in the SPM; the missing site is located at an adjacent position (D183 in orange; **Fig. 6.7**). The same happens in the RA95.5 variant, where the exact mutation sites or adjacent positions are predicted by SPM (only T95 is displaced five positions from SPM). The same holds for RA95.5-5 (four out of five predicted, one out of five in an adjacent position, see S43R in **Fig. 6.7**). The most evolved RA95.5-8F variant was generated from the previous RA95.5-8 enzyme. Surprisingly, the SPM of RA95.5-8 highlights that the catalytic Lys83, Try51, and Asn110 are all contained in the graph, and more importantly Phe180, which was subsequently mutated to Tyr180 to complete the catalytic tetrad, appear also in the path. This evidences the correlated motions of these positions. In RA95.5-8, seven mutations are predicted, four are in adjacent positions, and only two are displaced six positions from SPM (see S151G and N90D in **Fig. 6.8**).

The analysis of the most conserved hydrogen bonds (H-bond) along the MD trajectories in RA95.5-8 indicates that some of the residues included in the SPM also participate in the protein H-bond network (**Fig. 6.8**). We also find that there is not a direct correlation between the residues included in the path and the most flexible parts of the protein due to the threshold on the distance considering less than 6 Å (**Fig. 6.8**). However, most of the residues included in SPM are located adjacent to flexible regions of the enzyme (**Fig. 6.8**). Interestingly, a detailed analysis of the evolutionary conservation of the positions included in the SPM on the original scaffold 1LBL (**Fig. 6.7** and **6.8**) reveals that many of them are quite conserved. [308] SPM successfully identifies many residues involved in the conversion between the sampled conformational states, which seem to be the target points for DE for enhancing the novel function.

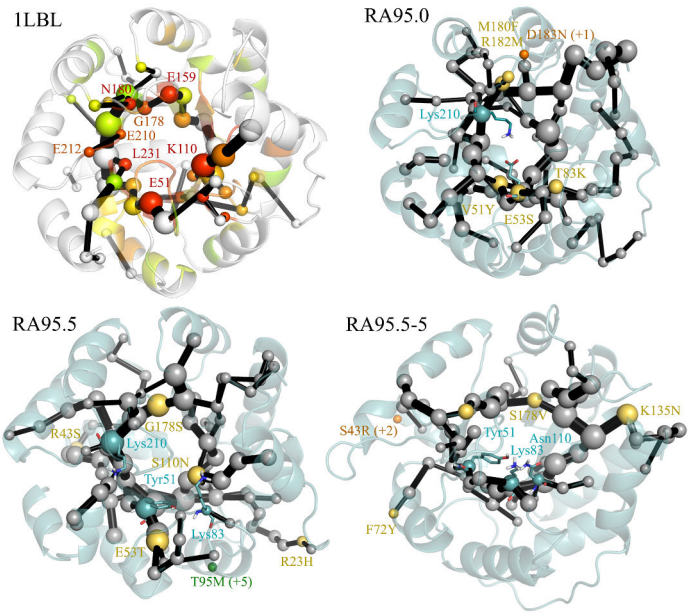


Figure 6.7 Representation of the shortest path map (SPM) along the evolutionary pathway. The size of the sphere is indicative of the importance of the position, and black edges represent the communication path (i.e. how the different residues are connected). Those points mutated via DE are marked in dark yellow (if they are included in the SPM), in orange if they are located in adjacent positions of the SPM (in parentheses is shown how far in the sequence from the closest residue included in SPM), and in green if the mutation is located at more than five positions away in sequence from the SPM. In 1LBL, the positions have been colored according to their evolutionary conservation using Evolutionary Trace Server (most conserved in red; less conserved in gray).[308]

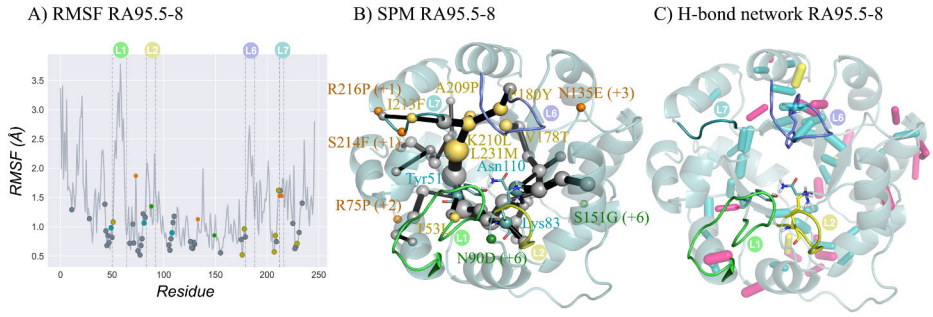


Figure 6.8 (A) Root mean square fluctuation (RMSF, in Å) for RA95.5-8 variant along the microsecond time scale MD simulations. (B) Amino acids identified with the Shortest Path Map (SPM) for each enzyme are indicated using gray dots. DE mutations are marked with dots in dark yellow (if they are included in the SPM), orange (if displaced by a few positions from SPM), and green (if located more than five positions from the path). The locations of the most mobile loops L1 (residues 52−66, green), L2 (residues 82−89, yellow), L6 (residues 180−190, blue), and L7 (residues 211− 215, cyan) are marked. The catalytic residues, also included

in the SPM, are marked with blue dots. (C) Analysis of the H-bond network in RA95.5-8. Those hydrogen bonds that have been maintained at least half of the simulation time are represented by sticks: in blue those hydrogen bonds that occur between backbone atoms, in pink those contacts between backbone and side-chain positions, and finally in yellow hydrogen bonds between side-chains. The weight of the H-bond stick indicates how frequently the H-bond is observed.

The good results obtained from SPM in this particular enzyme could be related to the fact that the original scaffold (1LBL) is an indole-3-glycerol phosphate synthase.[309] This enzyme has 30% sequence identity to imidazole-glycerol phosphate synthase (IGPS), which is a known allosterically regulated enzyme.[301,310] Nussinov and coworkers argued that allostery is an intrinsic property of all dynamic (non-fibrous) proteins, which suggests that the SPM tool might be applied to other cases as well.[121] Although we have restricted our study to the RA evolution case, this tool can be applied to other unrelated systems. Another important observation is that both 1LBL and IGPS are $(\beta\alpha)_8$ barrel enzymes, which is the most common enzyme fold in the PDB, as described in the very beginning of the introduction.[2] Moreover, many of the residues included in SPM are quite conserved, especially those located at the end of the β-sheets. This is of importance, as the catalytic groups in $(\beta\alpha)_8$ barrel enzymes are usually located at the ends of these different β-strands.[311]

Chapter 7. Catalase manganese complex

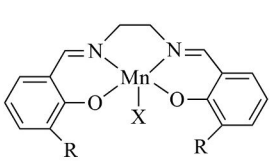
This chapter is based on the publication:

Romero-Rivera, A.; Swart, M., Unraveling the reaction pathway of EUK-8 catalase complex. *In preparation*.

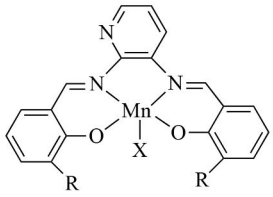
7.1 State-of-the-art

The dioxygen molecule (O_2) is highly beneficial for the organisms, but the same molecule can be converted into toxic metabolites forming reactive oxygen species (ROS) under aerobic conditions as hydrogen peroxide (H_2O_2), superoxide ($O_2^{\cdot-}$), and hydroxyl radicals ($OH^{\cdot}$). Overproduction of ROS formation is associated to different pathologies such as inflammation and infection.[312] Thus, protection towards this overproduction occurs thanks to the enzymes, specifically, catalase enzymes.[313] These enzymes are used in the majority of organisms as antioxidant defense detoxifying hydrogen peroxide.[314] Nevertheless, if the catalase activities cannot scavenge the excess of hydrogen peroxide produced, then the substrate can participate in the Fenton reaction forming hydroxyl radicals ($OH\cdot$),[315] producing the oxidative stress process. Hence, one solution for protection of organisms against excess of hydrogen peroxide formation is the synthesis of biomimetic organometallic complexes capable to mimic the antioxidant catalase functions.

Some catalases incorporate a dinuclear manganese atom in their active site to perform the catalysis, as we mentioned in the introduction. For this reason, many studies have been devoted to synthesize biomimetic complexes mimicking the active site of these enzymes.[148] Notwithstanding, in this study, we are going to focus on EUK-8 Mn^{III}(salen) complexes, which showed catalase and peroxidase activity.[316] However, EUK-134, and EUK-161 are some of the analogues with greater catalase activity than their ancestor EUK-8 (**Fig. 7.1**).[149] These complexes are small and water soluble molecules, which is favorable for medical purposes. Thus, the antioxidant complexes composed of Mn^{III}(salen) compounds have been found to be possible candidates for therapeutic treatments.[317,318] This suggests the possibility to manipulate salen compounds to improve their catalase activity for hydrogen peroxide scavenging. The understanding and knowledge gained about the catalase mechanistic reaction is essential to rationally design better manganese salen complexes for improving its catalase activity.



EUK-8: R = H, X = Cl
EUK-134: R = OCH_3, X = Cl



EUK-161: R = H, X = OAc

Figure 7.1 Salen complexes structures. OCH_3 methoxy and OAc acetoxy.

Only few theoretical examples studies have been done for understanding the catalase reaction by biomimetic complexes.[150,319,320] Specially, Burt and coworkers reported the first theoretical study about the catalase mechanism using salen complexes.[150,319] They considered the general "ping-pong" mechanism[321] to describe decomposition of hydrogen peroxide. This mechanism is structured in two dismutation reactions: the first dismutation reaction consists of the metal oxidation and formation of a water molecule by one hydrogen peroxide, while the second dismutation reaction consists of the conversion of a second hydrogen peroxide molecule into water and dioxygen molecules, recovering the initial salen manganese complex. Their results suggested that on the first dismutation reaction one-step process and the **intra**molecular proton transfer occurred.[319] While for the second dismutation reaction, they studied two possible pathways: (i) **inter**molecular transfer of hydrogen peroxide protons to the axial oxygen atom of the oxo complex; and (ii) **intra**molecular transfer of hydrogen peroxide protons with the formation of oxygen-oxygen bond with the axial oxygen atom of the salen manganese-oxo compound. They showed two possible initial reactions for the second dismutation reaction: the top and side approach. Their calculations suggested that the side approach is the bottleneck for the whole process. Furthermore, the catalyst could be in a deactivated form due to a possible formation of kinetically stable intermediate,[150] which was corroborated experimental studies by Nocera and coworkers.[151] Despite that, key points related to the influence of the spin states into the mechanism and the behavior of analogues of the EUK-8 with better rate reaction are still unknown.

In this work, our aim is to propose a detail mechanism path for the catalase reaction of hydrogen peroxide using the principle salen compound, EUK-8. The main goal is to provide a good description of the reaction barriers providing information about which stationary-points have an important role in the catalase activity and the rate-determining

step calculation of the reaction for two analogues with greater catalase activity (EUK-134 and EUK-161). These salen manganese compounds are potential complexes for therapeutical applications[322,323] and the control of the catalase activity is an important point to achieve better complexes. As a second goal, we are also interested to provide the spin-state splitting of the whole reaction pathways, including the three spins states of Mn^{III} (singlet, triplet, and quintet), corroborating which of them is the most favorable, and analyzing possible crossing points during the whole reaction path.

Our results showed that the rate-determining step of the first dismutation reaction is the oxygen-oxygen bond cleavage, and the second dismutation reaction has different possibilities for undergoing the reaction path with triplet and quintet spin state crossovers. We also found that water molecules could also play a role during the reaction pathway suggesting further calculations including explicit water molecules.

7.2 Computational details

All DFT calculations were performed with the ADF program.[173] A basis set can roughly be characterized by its size (single-, double-, triple-zeta; with or without polarization) and by the level of frozen core approximation. The basis set chosen is TZ2P, which is triple-zeta with two polarization functions. Scalar relativistic corrections were included self-consistently using ZORA.[180]

The election of DFT functional for geometry optimization has chosen between the functional PBE[169]-D_3 and S12g[324], which includes Grimmes's (D_3) dispersion contribution[178]. The calculations of the whole reaction path were performed with the functional S12g, which is a good alternative to describe well both the spin state splittings and the weak interactions. The geometry optimizations and TS searches were performed with the Quantum-regions Interconnected by Local Descriptions[325] (QUILD) program, which works as a wrapper around ADF program. The calculations were done in implicit solution with COSMO[174,326,327] (with water as a solvent), while explicit water molecule were used to unravel the rate-determining step of EUK-8 and two analogues with greater catalase activity (EUK-134 and EUK-161).

7.3 Results and discussion

The optimization of the EUK-8 complex for the different spin-states of Mn^{III} (singlet, triplet, and quintet) was done using both functionals (PBE-D_3 and S12g) to select which can be the better one. The results suggested that the quintet state is the ground state with the triplet and singlet state higher in energy by 20.1 (triplet) and 44.1 (singlet) kcal/mol at S12g/TZ2P level of theory, and 13.0 (triplet) and 33.3 (singlet) kcal/mol at PBE-D_3/TZ2P level of theory, respectively. The comparisons of bond distances between ligands coordinated to the manganese metal atom are shown in **Table C.2**. Our calculations showed that similar bond distances were obtained for both functionals with less than 0.06 Å differences compared to the X-ray structure. Therefore, both functionals are feasible to be used for optimization geometries in the whole catalase reaction mechanism. However, we decided to choose S12g due to its properties to analyze properly the different spin states on the mechanism, as well as the weak interactions, which are thought to play an important role during the reaction of the first and second dismutation reaction of hydrogen peroxide for this salen complex.

7.3.1 First dismutation reaction

The first dismutation reaction consists of the conversion of a hydrogen peroxide into water with formation of a Mn^{V}-oxo complex (equation 7.1).

$$Mn^{III}\text{-salen} + HOOH \rightarrow Mn^{V}\text{=O-salen} + H_2O \qquad (7.1)$$

The study of this first step in the ping-pong mechanism suggests similar mechanistic pathway for the different spin states (**Fig. 7.2**), except for the reaction complex (RC) for quintet state (RC^{q}, **Fig. 7.2**). As one can see, the difference in the RC for both singlet and triplet states with the quintet state are based on the position of the hydrogen peroxide in the salen complex. In the case of the quintet state two different orientations are found close in energy (RC_a 0.5 kcal/mol lower in energy than RC_b): one with hydrogen bond interactions between one oxygen atom of the salen ligand and the axial chloride atom of the complex (RC_a, **Fig. 7.2**), while for the other orientation the hydrogen bond interaction is formed between the two oxygen atoms of the salen ligand (RC_b, **Fig. 7.2**). On the other hand, in the singlet and triplet spin state the hydrogen peroxide is posi-

tioned above the manganese metal atom in an end-on orientation (Mn-O1 = 2.16 Å in both cases). It is worth to mention that the spin density calculated in all reactant complexes are mostly localized at the manganese metal atom, 3.8 for RC_a^q corresponding to the four unpaired electrons in quintet state, 2.2 for RC^t corresponding to twice unpaired electron in triplet state, and 0.1 for RC^s corresponding to zero unpaired electrons singlet state. Our calculations show that the quintet state is the ground state in the first step of the reaction until the first transition state. The binding of the hydrogen peroxide at first sight seems to be most favorable for the singlet state with a binding energy of -21.1 kcal/mol, compared to -10.2 and -7.5 kcal/mol for the triplet and quintet state, respectively (**Table C.1**). However, this is deceptive because the quintet state remains the spin ground state as well for this initial complex. The large binding energy should therefore be regarded as resulting from a highly destabilized EUK-8 singlet state.

The next step found is the hemolytic bond breaking of the oxygen atoms in the hydrogen peroxide leading to the first transition state (TS1). This TS1 essentially describes the bond breaking of O1-O2 with a bond distance of 1.86 Å, 0.41 Å larger than in the RC of quintet state (1.45 Å). The broken O1-O2 bond leads to the formation of an intermediate (Int) structure with a metal-hydroxo bond (Mn-OH), the Multiple Derived Charge (MDC)[173] analysis suggest a change of the oxidation state from Mn^{III} to Mn^{IV} due to the shared electron from *d*-orbitals of the metal atom and the electron of the hydroxyl radical. The Int is followed by a hydrogen transfer from the metal-hydroxo to the hydroxyl radical, which corresponds to the second transition state two (TS2). The last step of the first dismutation reaction is the formation of a water molecule and the manganese oxo-salen complex (Mn^V=O), where the distance between Mn-O1 is 1.85 Å, which corresponds to the product complex (PC_{oxo}) of the first dismutation reaction. MDC analysis suggests another change of the oxidation state from Mn^{IV} to Mn^V due to the same fact as the Int structure commented before.

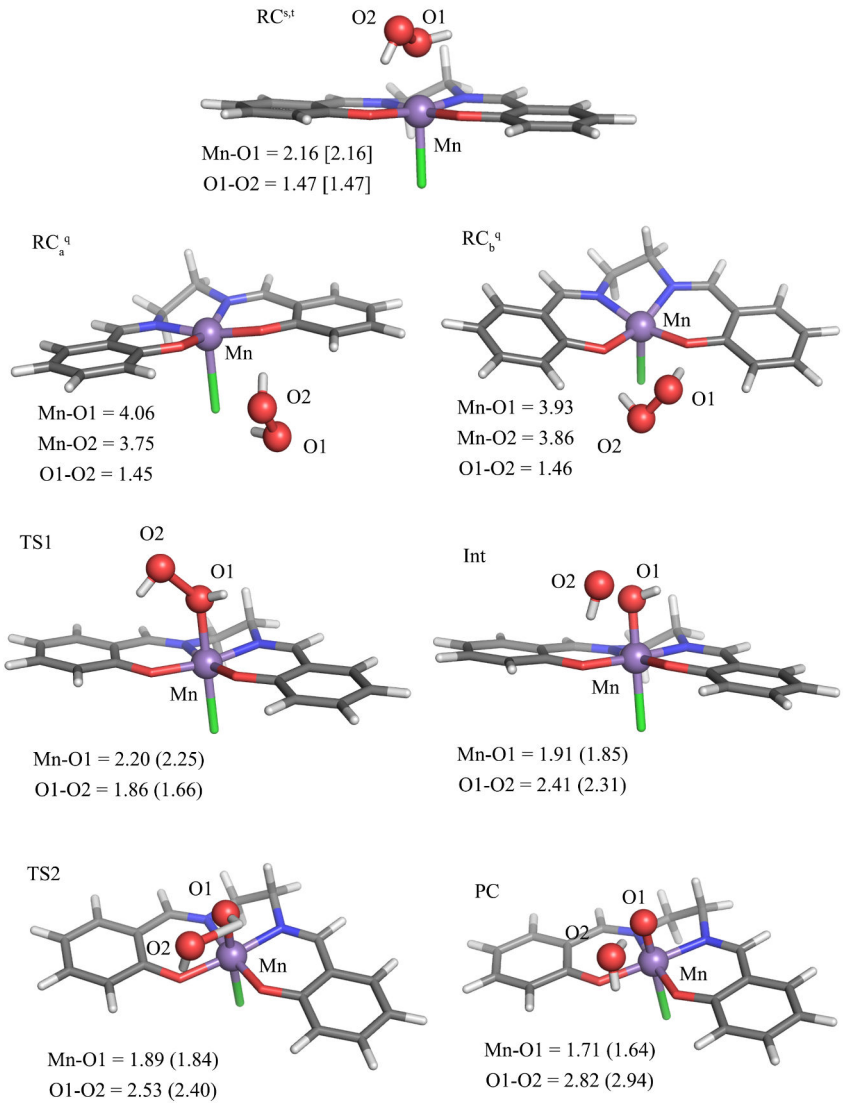


Figure 7.2 Optimized structures for the first dismutation reaction: Reaction complex (RC) of singlet and triplet state, two different reaction complexes for quintet state ($RC_{a,b}$), first transition state (TS1), intermediate (Int), second transition state (TS2), and product complex (PC). Distances for triplet state in blank, singlet state showed in brackets [ ], and for quintet state in parenthesis ( ).

The energetic profile (**Fig. 7.3**) shows that the quintet state is the lowest in energy for the initial steps suggesting an energy barrier of 13.7 kcal/mol for the oxygen-oxygen cleavage. However, the triplet state has a lower energy barrier of 2.7 kcal/mol, being 11.0 kcal/mol lower in energy than the quintet state, this fact is due to the stabilization

and orientation of the hydrogen peroxide, where an important influence in the reaction barrier is clear, making an exchange-enhanced reactivity when the hydrogen peroxide is stabilized above the manganese metal atom with shorter distances of Mn-O1. Afterwards, a spin "flip" from quintet to triplet state occurred on going from the TS1 to Int, with the triplet as the ground state until the formation of the oxo compound. The reaction barriers for the TS2, the hydrogen-transfer from the hydroxyl group of the salen complex to the substrate radical, are 0.2 and 0.5 kcal/mol for triplet and quintet state respectively, which is a very low barrier easy to surpass. The calculation of the single state is performed through single point calculations of the triplet state geometries (TS1, Int, TS2 and PC) due to the impossibility to optimize the structures. The high energy showed during the first dismutation reaction path, it prompted us to suggest that the singlet state can be unimportant for ping-pong mechanism with energies much higher (>30 kcal/mol) than the energies for the RC of quintet state.

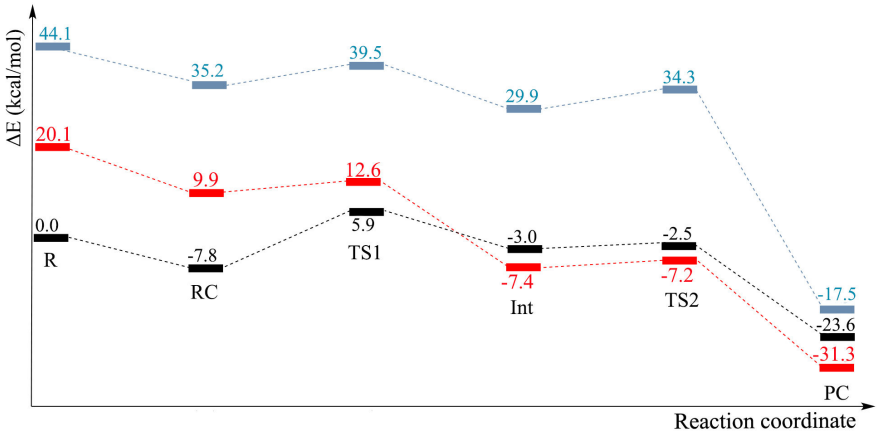


Figure 7.3 Energy profile (kcal/mol) for first dismutation reaction of hydrogen peroxide catalyze by the salen manganese complex. Stationary points represented are: Reactants at infinite distance (complex and hydrogen peroxide, R); reaction complex (RC), first transition state (TS1), intermediate (Int), second transition state (TS2) and product complex (PC). Also, the different spin-states for Mn^{III}: singlet state (blue line), triplet state (red line), and quintet state (black line)**.**

7.3.2 Second dismutation reaction

The second dismutation reaction converts another hydrogen peroxide into a water molecule and dioxygen molecule (eq. 7.2).

$$Mn^{V}=O\text{-salen} + HOOH \rightarrow Mn^{III}\text{-salen} + H_2O + O_2 \quad (7.2)$$

Two possible conformations are found to initialize the second dismutation reaction (RC_i and RC_{ii}, **Fig. 7.4**) for quintet and triplet state, in this case we have not calculated the energies of the singlet state due to the same reasons commented in the first dismutation reaction. In the first reaction complex (RC_i) one hydrogen bond is formed between one hydroxyl group of the hydrogen peroxide and the axial oxygen of the oxo-manganese complex and another hydrogen bond is formed between the other hydroxyl group of the hydrogen peroxide and one oxygen of the salen ligand (left, **Fig. 7.4**). On the contrary, in the RC_{ii} only one hydrogen bond is found between the one hydroxyl group of the hydrogen peroxide and the axial oxygen of the oxo-manganese complex (right, **Fig. 7.4**). The difference in energy between both is less than 1 kcal/mol, making both possibilities compatible as starting point for the second dismutation reaction. For this reason the study of the second part of the ping-pong mechanism is performed using both possible conformations.

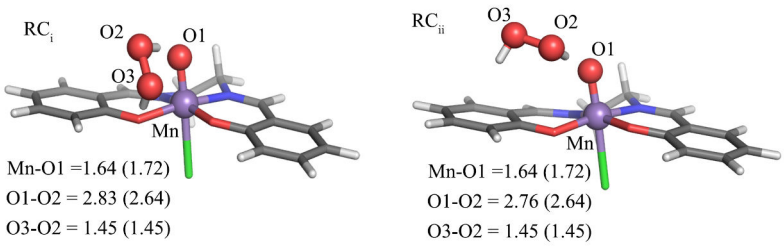


Figure 7.4 Two possible conformations for the reaction complex (RC) in the second dismutation reaction. Bond length in Å, distances for quintet in parenthesis ( ) and distances for triplet in blank**.**

The mechanism starting from RC_i and following the **inter**molecular hydrogen transfer mechanism shows a hydrogen transferred from the hydrogen of O2 in the hydrogen peroxide to the oxo atom (O1) of the oxo-manganese salen complex ($TS1_a^i$), followed by the formation of a hydroxo manganese complex ($Int1_a^i$). Afterwards, a rearrangement of the hydrogen peroxide moiety ($TS2_a^i$) towards the hydroxo of the complex occurred to get the second intermediate structure ($Int2_a^i$). This intermediate facilitates the second hydrogen transfer ($TS3_a^i$) to form the final water molecule, dioxygen molecule, and to recover the initial EUK-8 salen complex (PC_a^i). The bond distances of the O1 and Mn

of the oxo-complex increased step by step during the reaction path from 1.64 Å to 2.73 Å because of the formation of the water molecule. On the other hand, the reaction followed by **intra**molecular proton transfer is performed only by one transition state (TS_b^i, **Fig. C.2**). The transition state consists of breaking the oxygen-oxygen bond of the hydrogen peroxide and a simultaneous proton transfer and thus it forms the product complex with the oxygen molecule oriented above the Mn metal atom and the water molecule (PC_b^i, **Fig. C.2**).

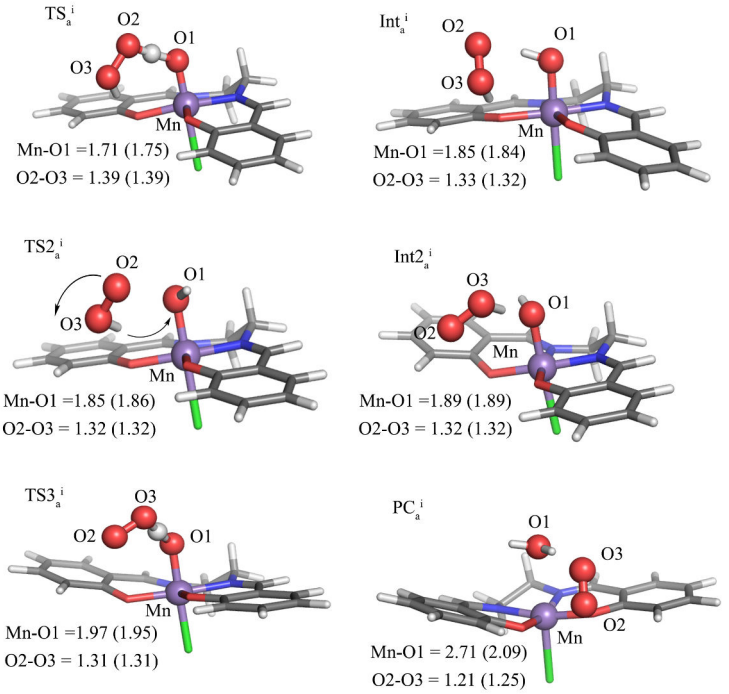


Figure 7.5 Optimized structures of the second dismutation reaction of the intermolecular transfer of hydrogen through the RC_i. The whole path contains the first transition state (TS1), first intermediate (Int1), second transition state (TS2), second intermediate (Int2), third transition state (TS3); and product complex (PC). Bond length in Å, distances for quintet in parenthesis ( ) and distances for triplet in blank**.**

The energetic profile (**Fig. 7.6**) of the **inter**molecular pathway suggests the triplet as the ground state until the first proton transfer (TS_a^i). Then, from the first to the second intermediate structures ($Int1_a^i$ and $Int2_a^i$) both spin states are closer in energy making it impossible to differentiate the preferred one. Just in the product complex it is possible to see a higher stabilization for the triplet state. However, the sum of the different compo-

nents of the product shows the quintet state as ground state. The low energy barriers make the **inter**molecular pathway easier to follow than the **intra**molecular pathway with an energetic barrier of >11 kcal/mol for both triplet and quintet spin states (**Fig. C.3**).

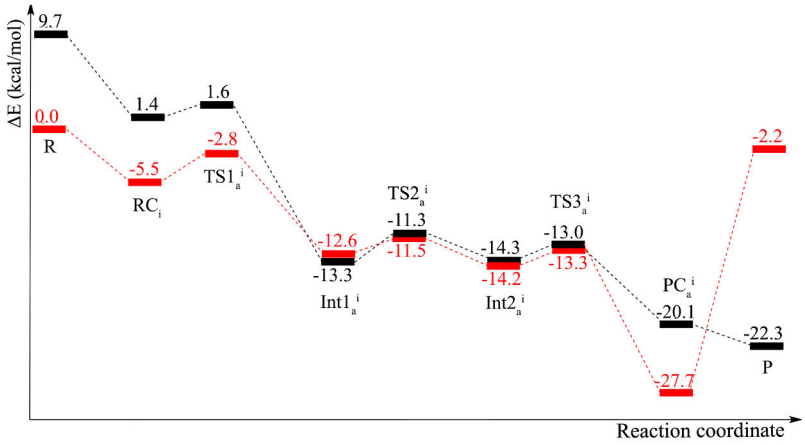


Figure 7.6 Energy profile (kcal/mol) of the second dismutation reaction following the intermolecular proton transfers from the reaction complex i). Stationary points represented are: reactants at infinite distance (oxo-complex and hydrogen peroxide, R), reaction complex (RC_i), first transition state ($TS1_a^i$), first intermediate ($Int1_a^i$), second transition state ($TS2_a^i$), second intermediate ($Int2_a^i$), third transition state ($TS3_a^i$), product complex (PC_a^i), product at infinite distance (complex, water and oxygen molecule, P). Also, the different spin-states for Mn^{III}: triplet state (red line) and quintet state (black line).

The **inter**molecular proton transfer mechanism starting from RC^{ii} begins similar to the previous mechanism, with the first hydrogen transfer from the O2 of the hydrogen peroxide to the oxo (O1) atom of the Mn-oxo complex ($TS1_a^{ii}$, **Fig. 7.7**) leading to an intermediate structure (Int_a^{ii}). We can see during these steps that the bond distance between the Mn and the O1 of the complex also increased (Mn-O1, **Fig. 7.7**). This intermediate undergoes a rearrangement of the hydrogen peroxide moiety ($TS2_a^{ii}$), with a distortion of the salen ligand, converging to the first intermediate ($Int1_a^i$) of the previous reaction mechanism, and thus it follows the same mechanism mentioned before until the end of the reaction. For the **intra**molecular pathway, a similar transition state (TS_b^{ii}) is found with high reaction barrier due to the oxygen-oxygen breaking bond.

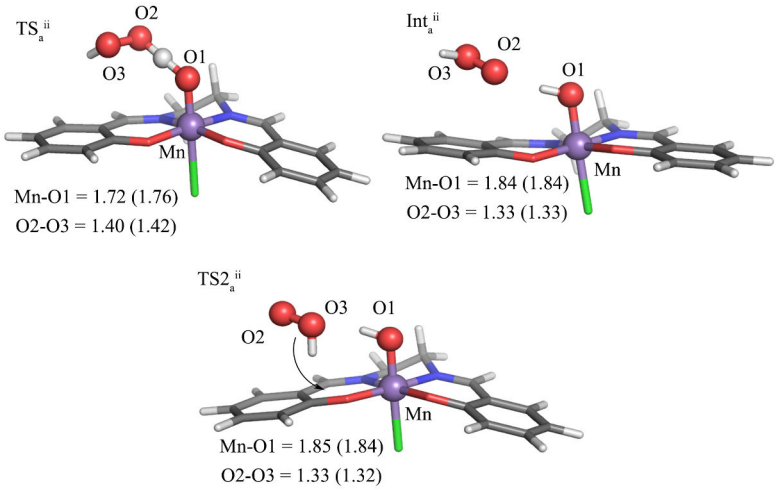


Figure 7.7 Optimized structures of the second dismutation reaction of the intermolecular transfer of hydrogen through the RC^{ii}. The first steps of the pathway contain the first transition state (TS1$_a^{ii}$), the intermediate structure (Int), second transition state (TS2$_a^{ii}$). Bond length in Å, distances for quintet in parenthesis ( ) and distances for triplet in blank.

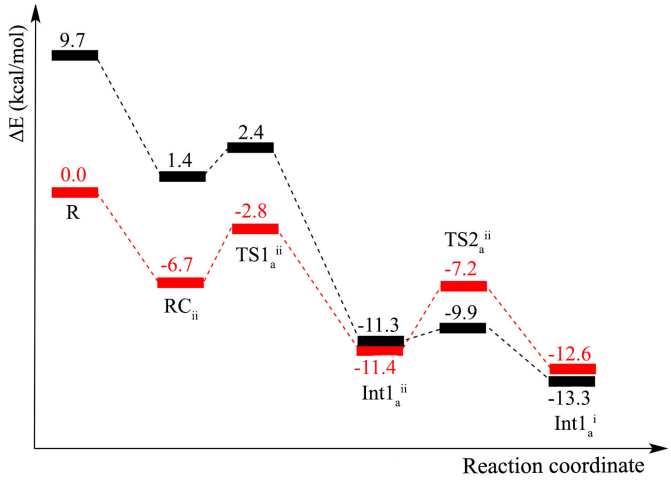


Figure 7.8 Energy profile (kcal/mol) of the second dismutation reaction following the intermolecular proton transfers from the RC$_{ii}$. Stationary points represented are: reactants at infinite distance (oxo-complex and hydrogen peroxide, R), reaction complex (RC$_{ii}$), first transition state (TS1$_a^{ii}$), first intermediate (Int1$_a^{ii}$), second transition state (TS2$_a^{ii}$), first intermediate of the previous reaction pathway (Int1$_a^{i}$). Also, the different spin-states for Mn^{III}: triplet state (red line) and quintet state (black line).

The initial steps of this **inter**molecular pathway in the triplet state are the ground state (**Fig. 7.8**). However, our calculations suggest a very similar situation as before with closer energy between both spin states in the case of the Int_a^{ii}, while the rearrangement in the second transition state ($TS2_a^{ii}$) has the quintet state as the lowest in energy, and a lower energetic barrier compared with the triplet state. From the latter intermediate the mechanism is exactly the same as before, starting from the Int_a^{i}. The energetic barrier of the transition state in the **intra**molecular proton transfer pathway is >25 kcal/mol for both spin states. It clearly shows that in both cases the energetic barrier is higher than the **inter**molecular pathway. Thus, it makes the latter pathway the preferred one for following the second hydrogen peroxide decomposition in water and dioxygen molecule.

Interestingly, we have also found another interesting mechanism for the **inter**molecular hydrogen transfer, which consists of the attack of the hydrogen peroxide behind the Mn^{V}-oxo complex. The first step is the first hydrogen transfer from O2 of the hydrogen peroxide to the oxo (O1) component of the complex ($TS1_a^{behind}$) as in the previous situations, but in this case with a simultaneous rearrangement of the hydrogen peroxide moiety (Int_a^{behind}). Then, a second hydrogen transfer occrus ($TS2_a^{behind}$) from the H_2O_2 to the hydroxo moiety of the complex, followed by the formation of a water molecule, dioxygen molecule and recovering the initial complex, EUK-8.

The energetic profile (**Fig. 7.10**) shows an energetic barrier for the first proton transfer ($TS1_a^{behind}$) of 6.3 kcal/mol for the triplet state and 0.1 kcal/mol for the quintet state, with the triplet state lower in energy than the quintet state. Afterwards, the intermediate (Int_a^{behind}) is quite stabilized energetically, followed by the second proton transfer ($TS2_a^{behind}$), in which the energetic barrier is 2.3 and 3.0 kcal/mol for triplet and quintet state respectively. In this case the rearrangement occurs in $TS1_a^{behind}$, which induces less steps for accomplishing the second dismutation reaction.

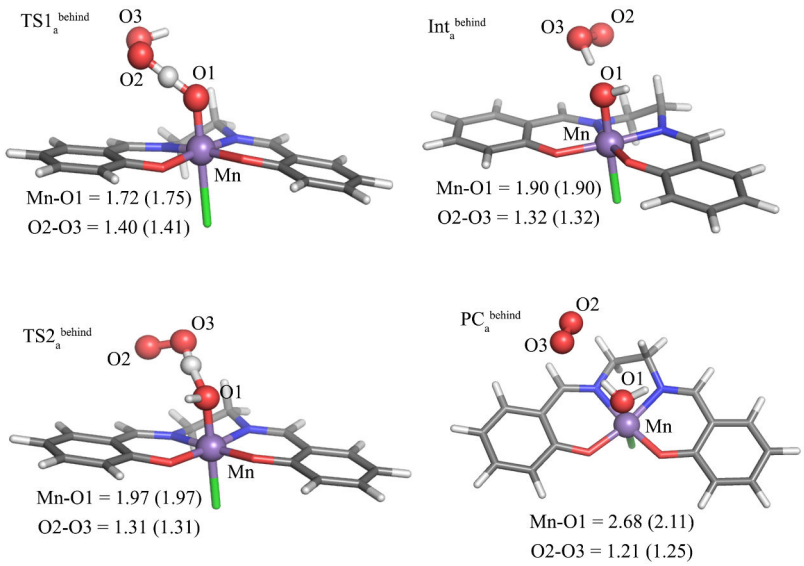


Figure 7.9 Optimized structures of the second dismutation reaction of the **inter**molecular pathway through the RC_{ii} where the hydrogen peroxide attacks behind the complex. The stationary point are the first transition state ($TS1_a^{behind}$), the intermediate structure (Int_a^{behind}), second transition state ($TS2_a^{behind}$), and product complex (PC_a^{behind}). Bond length in Å, distances for quintet in parenthesis ( ) and distances for triplet in blank**.**

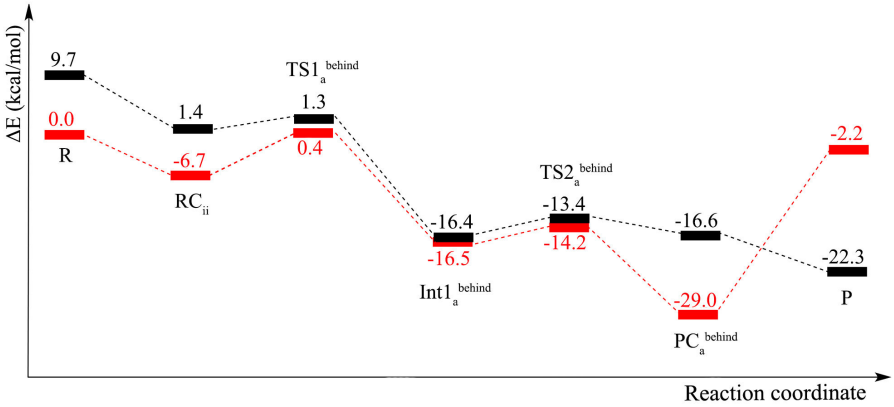


Figure 7.10 Energy profile (kcal/mol) of the second dismutation reaction following the intermolecular proton transfers from the RC_{ii}. Stationary points represented are: reactants at infinite distance (oxo-complex and hydrogen peroxide, R); reaction complex (RC_i), first transition state ($TS1_a^{behind}$), first intermediate ($Int1_a^{behind}$), second transition state ($TS2_a^{behind}$), product complex (PC_a^{behind}), and product at infinite distance (complex, water and oxygen molecule, P). Also, the different spin-states for Mn^{III}: triplet state (red line) and quintet state (black line).

Our calculations suggest that the rate-determining step for the ping-pong mechanism is the breaking oxygen-oxygen bond on the first dismutation reaction. To shed more light towards the catalase activity in the manganese salen compounds, we performed calculations on other analogs of EUK-8[149] (EUK-134 and EUK-161 using chlorine as the axial ligand) with greater catalase activity and including explicit water molecules to stabilize the hydrogen peroxide above the manganese metal (**Fig. 7.11**). The water molecules stabilize the hydrogen peroxide in the end-on orientation above the manganese metal atom suggesting that water molecules may play an important for activating the oxygen-oxygen cleavage bond, contrary to the calculation without explicit water molecules in which the hydrogen peroxide is positioned forming the hydrogen bond interaction with the oxygen of the ligand in the quintet state (**Fig. C.5**). All transition states have similar distances for Mn-O1 and O1-O2, while the reactant complexes have a difference between Mn-O1 distance of EUK-161 and the EUK-8 and EUK-134. The main difference is because of the hydrogen bond interaction of the water molecule and the nitrogen in the aromatic ring behind the complex, this interaction forces another end-on orientation of the hydrogen peroxide, which may facilitate the oxygen-oxygen bond breaking and increase the rate as shown in the experiments.[149]

The energy profile for the rate-determining step for the three complexes with and without explicit water molecules suggest that the EUK-8 has the higher energetic barrier to overcome the oxygen-oxygen bond cleavage than the EUK-134 and EUK-161 (**Fig. 7.12**). However, EUK-134 does not show a good trend because in the case with explicit water molecules the energy barrier is almost equal to EUK-161. Despite that, it is clear that EUK-161 is the most active salen compound in agreement with the experiments. The explicit water molecules showed lower energy barriers than without waters and thus it suggests that water molecule may have to be explicitly included during the first and second dismutation reaction for proposing detailed mechanism pathways for the ping-pong mechanism of these salen complexes.

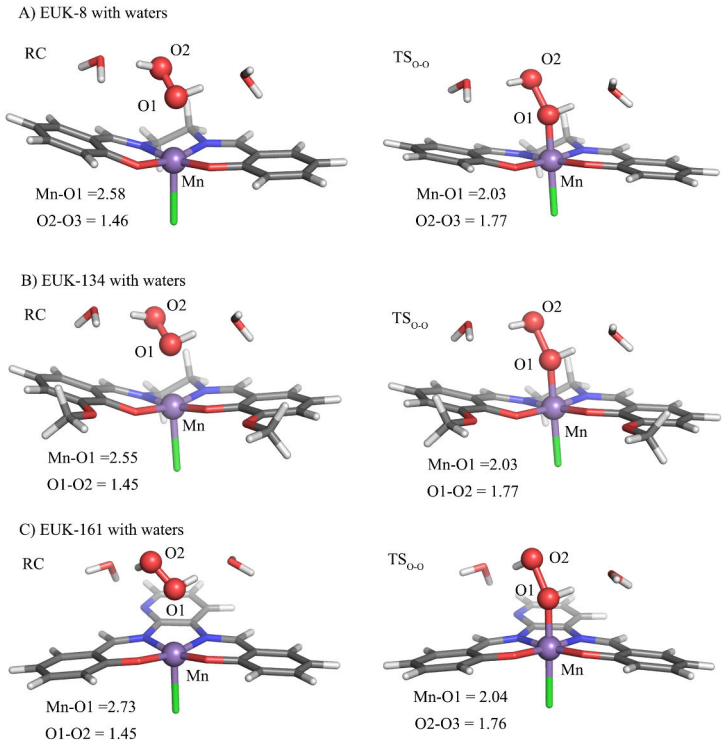


Figure 7.11 Optimized structures of the rate-determining state of A) EUK-8, B) EUK-134, and C) EUK-161 in presence of explicit water molecules for the quintet state. All distances are in Å.

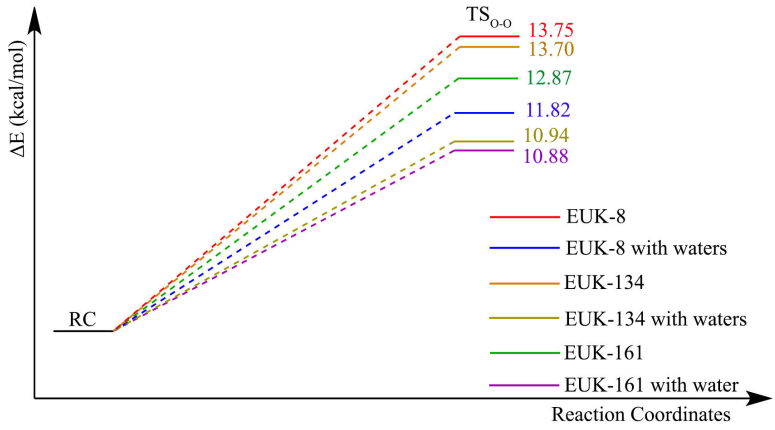


Figure 7.12 Energetic profile of the rate-determining step of the three different salen compounds with and without explicit water molecules.

Chapter 8. Mössbauer parameters prediction

This chapter is based on the publication:

Romero-Rivera, A.; Swart, M., New fit parameters for Mössbauer parameters predictions. *In preparation*.

Hill, E. A.; Weitz, A. C.; Onderko, E.; Romero-Rivera, A.; Guo, Y.; Swart, M.; Bominaar, E. L.; Green, M. T.; Hendrich, M. P.; Lacy, D. C.; Borovik, A. S. Reactivity of an Fe^{IV}-Oxo complex with protons and oxidants. *J. Am. Chem. Soc.* **2016**, *138*, 13143-13146.

8.1 State-of-the-art

Mössbauer spectroscopy using the ^{57}Fe nucleus has become a common technique in both bioinorganic and coordination chemistry due to the ability to selectively probe Fe environments.[328] As mentioned before, from the Mössbauer spectrum it is possible to measure two important parameters. The isomer shift, δ, which reflects the electron density at the iron nucleus, is often used to interpret the oxidation state and spin state of the iron center. However, many features influence the isomer shift, and thus it is difficult to predict the correct value. On the other hand, the quadrupole splitting reflects the asymmetry in electron distribution around the nucleus and therefore can offer insight into the occupancy of the 3*d* orbitals. For high-symmetry iron centers with octahedral and tetrahedral coordination environments, interpretation of the quadrupole splitting can be straightforward, but in more complex systems with lower symmetry and orbital mixing, interpretation is usually more complicated.[329]

The complexity of interpreting the experimental spectral to understand the electronics has prompted chemists to use DFT computations to predict the Mössbauer parameters for a given structure.[187,188,193] The accuracy of calculations compared to the experimental values has also become a valuable way to validate a computational model.[191,195,330] Calculation of isomer shift values generally depends on a correlation between the calculated electron density and the experimental isomer shift. These fits are typically linear, but the fit parameters only apply when using a specific set of computational parameters (functional, basis set, relativistic corrections, among others).[193,331]

The good results shown by OPBE and S12g functionals for spin state description prompted us to mix both functionals to calculate Mössbauer parameters.[156,172] We used DFT calculations to generate new correlation from a representative subset of iron complexes reported by Noodleman as a training set,[190] and therefore we tested the effectiveness of the fit parameters to predict Mössbauer parameters using a test set of different complexes and Fe-oxygen complexes.[172,195,332] The power of the same computations for prediction of quadrupole splitting values is also evaluated for all complexes. Thus, we describe the effects of ligand and the effect of the spin state on predicted isomer shift

and quadrupole splitting values in many examples. We anticipate that the new fit parameters can be used to predict Mössbauer isomer shift in iron-oxygen species.

8.2 Computational details

All DFT calculations were performed with the ADF program.[173] We used OPBE[170]/DZP level of theory for the geometry optimization of the training set complexes. For the correlating the Mössbauer parameters, the S12g[324]/TZ2P and OPBE/DZ, TZP, and TZ2P level of theory were used. The calculations were done in implicit solution with COSMO[174,326,327] and methanol or acetonitrile as a solvent, depending on the experimental procedure used for the synthesis and Mössbauer measurements. For all geometries optimization scalar relativistic corrections were included self-consistently using ZORA.[180] For the Fe^{IV}-oxo[332] geometry optimizations and frequency calculations were performed using PBE[169]-D_2/TZ2P. Predictions of Mössbauer parameters were obtained using S12g/TZ2P and OPBE/TZP level of theory, depending on each case of studied.

8.3 Results and discussions

8.3.1 Linear fit parameters for isomer shifts

As we mentioned in the introduction, the isomer shift is proportional to the electron density (ρ(0)) difference at the iron nuclei between the system under study ($\rho^S(0)$) and a reference system ($\rho^{Fe}(0)$) that can be described as:

$$\delta_S = A\left(\rho^S(0) - \rho^{Fe}(0)\right) + B \tag{8.3.1}$$

Here, we obtain A and B by a linear regression between the calculated electron density (ρ(0)) and the experimental isomer shifts (δ_S). We used the same linear fit procedures as Noodleman and coworkers, which consists on two sets of iron compounds, $Fe^{2+,\,2.5+}$ and $Fe^{2.5+,\,3+,\,3.5+,\,4+}$.[189,190] These set are composed of 17 $Fe^{2+,\,2.5+}$ with 31 Fe sites and 19 $Fe^{2.5+,\,3+,\,3.5+,\,4+}$ with 30 iron sites.[189] The difference relies in using as a reference value ($\rho^{Fe}(0)$) the electron density of a single iron atom with the corresponded oxidation state (Fe^{II} for $Fe^{2+,\,2.5+}$ set and Fe^{III} or Fe^{IV} for $Fe^{2.5+,\,3+,\,3.5+,\,4+}$) and the lowest-energy spin state. We used S12g/TZ2P and OPBE/DZ, TZP, and TZ2P levels of theory to calculate

the electron density. All of the calculated electron densities of the reference values and the training set iron complexes are shown in **Table D.1-8**.

Table 8.1 Fit parameters and correlation values for all levels of theory used in the training set and the average value for the different training set.

	$Fe^{2+, 2.5+}$			**$Fe^{2.5+, 3+, 3.5+, 4+}$**		
Level theory	A	B	Correl.	A	B	Correl.
S12g/TZ2P	-0.452	2.305	0.886	-0.349	-0.044	0.961
OPBE/DZ	-0.440	2.243	0.868	-0.356	-0.099	0.965
OPBE/TZP	-0.426	2.266	0.891	-0.346	-0.047	0.960
OPBE/TZ2P	-0.430	2.288	0.894	-0.345	-0.033	0.961
Average	-0.437	2.275		-0.349	-0.056	

The new linear regressions for both training sets $Fe^{2+, 2.5+}$ and $Fe^{2.5+, 3+, 3.5+, 4+}$ are shown in **Table 8.1**. Similar coefficient correlations are found for the different levels of theory used with *ca.* 0.89 for $Fe^{2+, 2.5+}$ and *ca.* 0.96 for $Fe^{2.5+, 3+, 3.5+, 4+}$. These coefficient correlations have a similar value to those found by Noodleman and coworkers.[190] Therefore, using the new fitted linear parameters, A and B, for each level of theory and the average of those parameters, we calculated the isomer shifts for both training set of $Fe^{2+, 2.5+}$ and $Fe^{2.5+, 3+, 3.5+, 4+}$ complexes (**Table D.1-8** and **Fig. 8.2**, only the linear fit and isomer shift calculation plots of S12g/TZ2P is illustrated because the same trend is found in all cases). The results show similar coefficient correlations of *ca.* 0.98 for the calculated isomer shift and for the average parameters. Overall, the accuracy of the fit correlations are very similar to those found by Noodleman and coworkers[189,190] and are in good agreement with the experimental values. Hence, we decided to use the linear fit **average** parameters values for calculating isomer shifts in all iron complexes shown later. Regarding quadrupole splitting, the plot between experimental and calculated quadrupole splitting shows a good correlation of *ca.* 0.95 for the complexes encompassed in both training set. However, the values are more dispersed than those found for the isomer shift calculations, indicating that depending on the system a larger range of errors will be found for the quadrupole splittings.

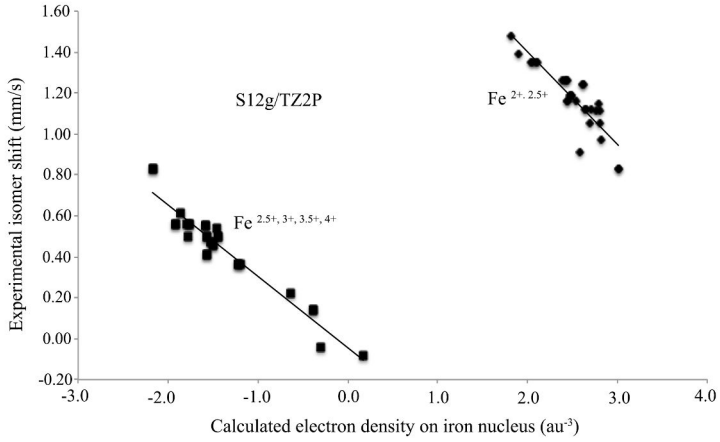


Figure 8.1 Correlation between the S12g/TZ2P calculated electron densities at Fe nuclei and the experimental isomer shifts of both training set iron complexes.

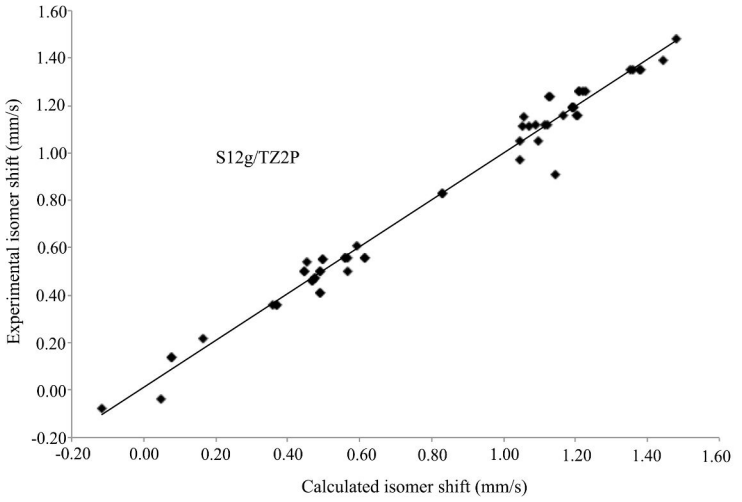


Figure 8.2 Correlation between the calculated and the experimental isomer shifts for all the compound encompassed in both training sets obtained from S12g/TZ2P.

8.3.2 Mössbauer parameters prediction

We carried out an analysis of our new fit parameters on different iron complexes. Most of the selected iron complexes are well described in previous studies, those complexes selected have 2+, 3+, and 4+ oxidation state and diverse spin states.[195] In this case we also use the fit parameters in a single training set $Fe^{2+.\ 2.5+,\ 3+,\ 3.5+,\ 4+}$ using as reference

the three different oxidation states of isolated iron atoms (+2, +3, and +4 with their respective lowest-energy spin state). The latter fit is supposed to give bad results for predicting the isomer shift of iron complexes despite the good coefficient correlation found (*ca.* 0.97 in our case) in the linear regression.[189] We used the optimized structures obtained by Neese and coworkers.[195] All the isomer shift calculations can be found in **Table 8.2** and **D.9**. These results show that fit parameters separated in two training set ($Fe^{2+,\ 2.5+}$ and $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$) give much better results than using only one training set ($Fe^{2+,\ 2.5+,\ 3+,\ 3.5+,\ 4+}$), which supports the studies performed by Noodleman and coworkers.[189] These results also show similar results for the isomer shift calculation using both level of theory OPBE and S12g, only in one case ($[FeCN_6]^{-4}$) are differences observed. Moreover, the fit parameters of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ show almost exactly the same results using iron 3+ or 4+, which suggest that any of both these oxidation states are good reference values for isomer shift calculations.

Table 8.2 Isomer shift calculation of a different set of iron complexes with different oxidation states (2+, 3+ (in parenthesis), and 4+) and spin states using OPBE/TZP and S12g/TZ2P level of theory.

Complex	Ox. State	S_{total}	δ_{exp}	δ_{OPBE}	δ_{S12g}
$[FeCN_6]^{-4}$	2+	0	-0.02	-0.14	-0.05
[a][FePor(OAc)]$^{-}$	2+	2	1.05	0.93	0.96
$[Fe(SR)_3]^{-}$	2+	2	0.56	0.55	0.57
$[FeAz]^{+}$	3+	1/2	0.29	0.25 (0.26)	0.26 (0.26)
$[FeCN_6]^{3-}$	3+	1/2	-0.13	-0.17 (-0.16)	-0.14 (-0.14)
[a][FePor(O_2)]$^{-}$	3+	5/2	0.67	0.67 (0.68)	0.67 (0.67)
$[FeMAC]^{-}$	4+	2	-0.02	0.07 (0.07)	0.07 (0.07)
$[FePorO]^{+}$	4+	3/2	0.08	-0.02 (-0.02)	0.05 (0.06)
$[FeTMCO]^{2+}$	4+	1	0.17	0.10 (0.10)	0.10 (0.10)

[a]S_{Fe} = 1, the metal is ferromagnetically coupled to a porphyrin S = 1/2 radical

Then, we have also compared our new fit parameters with old ones in the same iron-oxygen complexes studied by Swart.[172] This set of complexes encompasses Fe^{IV}-oxo with crystal structures available[333-337] and well characterized Fe^{IV}-oxo, Fe^{III}-oxo, Fe^{III}-hydroxo, Fe^{III}-peroxo, Fe^{III}-hydroperoxo complexes.[338-340] We performed the calculation using S12g/TZ2P level of theory with the new fit parameters obtained before for the isomer shift calculation, but in this case we also compared the quadrupole splitting ob-

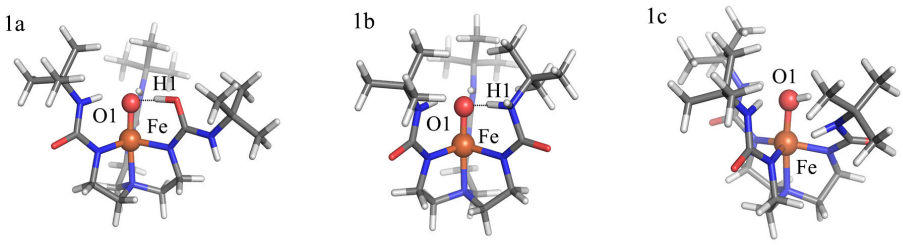
tained in that study between the ones obtained with S12g/TZ2P. The results shown in **Table 8.3** suggest that some of the calculated isomer shifts got values closer to the experimental data, while others shifted to larger values, specially, ligands containing 1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane (TMC). However, the isomer shift calculation has slightly improved (considering an error <0.1 mm/s between the difference of the experimental and the calculated value) using the same optimized structures calculated in that study. Thus, these results show that the new fit parameters are good enough for predicting correctly the isomer shift of iron-oxygen species. Regarding the quadrupole splitting, the results are similar to those calculated by Swart,[172] which still deviate from the experimental value but some cases show good results.

Table 8.3 Experimental and calculated Mössbauer parameters for a set of iron-oxygen species.

		Exp.		OPBE/TZP[a]		S12g/TZ2P	
Complex	S_{total}	**δ**	ΔE_Q	**δ**	ΔE_Q	**δ**	ΔE_Q
$[Fe^{III}(O)H_3buea]^{2-}$	5/2	0.30	0.71	0.42	-1.61	0.36	-1.56
$[Fe^{III}(OH)H_3buea]^{-}$	5/2	0.32	0.92	0.40	0.92	0.34	0.66
$[Fe^{IV}(O)H_3buea]^{-}$	2	0.02	0.43	0.15	0.65	0.06	0.72
$[Fe^{IV}(O)(N4Py)]^{2+}$	1	-0.04	0.93	0.03	0.80	-0.07	0.89
$[Fe^{IV}(O)(TMC)(NCCH_3)]^{2+}$	1	0.17	1.24	0.18	0.96	0.09	0.94
$[Fe^{IV}(O)(TMC^{i})(NCCH_3)]^{2+}$ inverted	1	0.14	0.78	0.18	1.21	0.09	1.20
$[Fe^{IV}(O)(TMC)(NCS)]^{+}$	1	0.18	0.55	0.17	0.40	0.09	0.36
$[Fe^{IV}(O)(TMC\text{-}Py)]^{2+}$	1	0.18	1.08	0.19	1.00	0.10	1.00
$[Fe^{IV}(O)(TMCS)]^{+}$	1	0.19	-0.22	0.22	-0.21	0.14	-0.20
$[Fe^{IV}(O)(TMCSO_2)]^{+}$	1	0.19	1.28	0.22	0.61	0.14	0.60
$[Fe^{IV}(O)(TMG_3tren)]^{2+}$	2	0.09	-0.29	0.14	-0.22	0.04	0.11
$[Fe^{IV}(O)(tpa^{Ph})]^{-}$	2	0.09	0.51	0.13	0.43	0.03	0.54
$[Fe^{III}(OO)(TMC^{i})]^{+}$	5/2	0.58	-0.92	0.66	-1.24	0.61	-1.18
$[Fe^{III}(OOH)(TMC^{i})]^{2+}$	5/2	0.51	0.20	0.49	1.16	0.56	1.30

[a] Results from Swart study at OPBE/TZP level of theory.

Intermediates involving metal-oxo/hydroxo species are important in a variety of oxidative transformations in biology.[341] The transition between high-valent Fe-oxo and Fe-OH complexes is an important function of heme-containing enzymes.[342] There are recent discussions to determine if the protonation occurs at the Fe^{IV}-oxo species or at a proximal site that then forms a H-bond to the oxido ligand.[343,344] However, there is no experimental evidence for protonated non-heme Fe^{IV}-oxo complexes and most of the synthetic systems lack intramolecular H-bond networks that are vital in regulating the site of protonation in proteins with Fe^{IV}-oxo units.[343,344] In collaboration with the experimental group of Borovik, we shed more light onto this issue applying our new fit parameters to calculate the Mössbauer parameters to a Fe^{IV}-oxo species containing a tris-urea tripodal ligand. These species change in a formation of a new **intra**molecular H-bond with the oxido ligand that stabilizes the protonated adduct, however, there is no X-ray structure to determine the correct H-bond interaction formed. Therefore, we used DFT to understand the structural properties, in which three different optimized structured were suggested with different localization of the proton in Fe-oxo and Fe-OH species (**Fig. 8.3**). The Fe-O1 bond distance 1.696 and 1.703 Å in 1a and 1b are quite similar to bond distance of 1.680(1) Å of the experimental compound using EXAFS.[332] The long O1···H1 distances of 1.450 Å in 1a and 1.549 Å in 1b suggest that the apical ligand is better described as a H-bonded oxido than a hydroxido. For 1c, the Fe–O1 bond length is increased to 1.812 Å that reflects the formation of the hydroxido ligand. The experimental Mössbauer parameters are ΔE_Q=0.87 mm/s and δ=0.04 mm/s and our results showed that the lowest in energy structured (1a) showed a ΔE_Q=0.81 mm/s and δ=0.045 (**Table 8.4**), even we were able to differentiate between all spin states possibilities in Fe^{IV}. Thus, the experimental and calculated Mossbauer parameters suggest that the Fe-oxo with an intramolecular H-bond is the preferred iron-oxygen species. These results showed really accurate values providing an opportunity for using these new fit parameters to characterize unknown Fe-oxygen intermediate species with possible different conformations when no X-ray structure is obtained.



Fe-O1 = 1.703
O1-H1 = 1.450

Fe-O1 = 1.696
O1-H1 = 1.549

Fe-O1 = 1.812

Figure 8.3 Geometry optimizations and Fe-O1 distances in 1a-c with different hydrogen bond interactions for the high spin state.

Table 8.4 Computed Mössbauer parameters for 1a-c from DFT calculations for low, intermediate, and high spin state.

	1a			**1b**			**1c**		
Spin	S=0	S=1	S=2	S=0	S=1	S=2	S=0	S=1	S=2
δ	-0.062	0.012	0.045	-0.059	-0.052	0.056	-0.060	-0.013	0.061
ΔE_Q	3.58	2.85	0.81	3.43	1.98	0.68	3.40	2.61	1.49

Chapter 9. Conclusions

The main conclusions of this thesis are compiled and organized in three main groups:

Conformational dynamics of enzymes

In Chapter 4 we have explored the conformational dynamics of the quercetinase metalloenzyme in the *apo* and E-S complex state. MD simulations reveal the behavior of the lid domain present in this enzyme for substrate binding and product release. The simulations show destabilization of the lid domain in the *apo* state due to the lack of interactions to stabilize the lid, while in the case of the E-S complex state we explored how the lid domain is most of the time stabilized either starting from open conformation or closed conformation (native state), suggesting that the substrate is responsible for further stabilizing closed conformations. Although essential information was extracted from the conformational dynamics of the metalloenzyme, a more in-depth analysis was done using aMD simulations for reconstructing the binding process of the substrate entrance and the product release. These simulations starting with the substrate compound far away from the active site of the enzyme clearly show how the lid domain in the absence of substrate is destabilized and sample wide open conformation (angle >80º), which allow substrate entrance. Then, the hydrophobicity of the substrate displaces water molecules located in the cavity, and thus generating the interactions between Pro164 and the substrate in which the lid domain confers a stable closed conformation. We demonstrated how MD and aMD simulations are valuable techniques for exploring the conformational dynamics of the enzyme, specifically, the important conformations responsible for substrate binding and product release in enzymes.

In Chapter 5 our microseconds MD simulations on the natural D-sialic acid aldolase and the laboratory-engineered L-KDO aldolase, which contains eight distal mutations, reveal dramatic changes in the enzyme's conformational dynamics. For D-sialic acid aldolase, the conformational states sampled involve short displacements of Tyr110. This residue was described previously to be important for properly positioning the catalytic Ser47 close to the Schiff base intermediate. In L-KDO aldolase, Tyr110 from both monomers can be completely displaced from the active site and form a H-bond at the dimer interfaces. It is worth mentioning that Tyr110 connects the loop containing the catalytic Tyr137 and α-helix 6, which define the active site pocket of the enzyme. Thus, its conformational plasticity can be an important factor for modulating the substrate specificity

of the enzymes. We analyzed the shape of both enzyme pockets by access tunnel calculations, indicating dramatic changes in the shape of the bottleneck radius of substrate access channels, with differences of approximately 1 Å between variants. Interestingly the distal mutation found to be more important for substrate specificity is located in the bottleneck radii of the tunnel. We demonstrated how the effect of distal mutations in the enzyme specificity can be rationalized.

In Chapter 6 our MD studies evaluated the conformational landscapes of the *de novo* computational design retro-aldolases and identify the different conformational sub-states sampled during the DE pathway. We identified how mutations located near the active site and at distal positions induced a population shift from inactive to catalytically active conformations enhancing new catalytic activities, in the case of the last variant (RA95.5-8F), to similar orders of magnitude to those found in natural class I aldolases. We found that MD simulations coupled to correlation-based tools similar to those used to investigate processes such as allosteric regulation and molecular recognition can be successfully applied in the enzyme design field. We developed the shortest path map (SPM), which analyzes the different conformation sub-states sampled along the MD trajectory and identifies which residues are important for the substrate interconversion. Therefore, if catalytically competent states are sample in the MD simulation, the new tool facilitates identification of residues that contribute to the inactive-to-active interconversions. SPM thus identifies not only active site but also distal residues that could lead to a population shift towards the catalytically competent conformations for enhancing activity. This is totally unprecedented and thus opens the door to new computational paradigms that are not restricted to active site alterations.

Mechanism description of biomimetic complexes

Chapter 7 describes a computational study using the S12g density functional on the mechanism of hydrogen peroxide dismutation reaction catalyzed by a manganese salen complex, EUK-8. In the first dismutation reaction, the rate-determining step is found to occur during the cleavage of the oxygen-oxygen bond of hydrogen peroxide, following the quintet state energy profile. This is followed by a spin crossover from quintet to triplet in generating the Mn^{IV}-OH intermediate state with a hydrogen transfer forming the Mn^{V}-oxo compound. In the second dismutation reaction different pathways are found to

catalyze the second hydrogen peroxide into water and oxygen. Generally, the second dismutation reaction consists of the first hydrogen transfer to the oxo compound, followed by a rearrangement of the hydrogen peroxide moiety for the second hydrogen transfer. Even starting from two different possible reactant complexes in the second dismutation reaction the rearrangement of both $TS2_a$ leads to the same intermediate, and thus it follows the same pathway to form the water and oxygen molecules. Interestingly, we also found another possible pathway behind the salen compound, which consists of fewer steps for converting the second hydrogen peroxide into the corresponding products. Since the rate-determining step of the ping-pong mechanism is the oxygen-oxygen bond cleavage of the first dismutation reaction, we performed the rate-determining step calculation in analogues of EUK-8, EUK-134 and EUK-161, with greater catalase activities and including explicit water molecules to stabilize the hydrogen peroxide above the manganese metal atom. Our results concur with the experiments in which both analogues EUK-134 and EUK-161 have smaller energy barrier for the oxygen-oxygen bond cleavage thus allowing for greater catalase activities. The explicit water molecules included suggest that the ping-pong mechanism pathways should also be done in presence of explicit water molecules to unravel the complete description of the dismutation reaction of hydrogen peroxide.

Mössbauer parameters prediction

In Chapter 8 we have generated new fit parameters using DFT OPBE and S12g functionals, which showed good results for spin state description, for Mössbauer prediction in iron containing complexes. One can use as a reference value Fe^{II} or $Fe^{IV(III)}$ depending on the complex under study. Our tests suggested a good agreement between the experimental and calculated isomer shift values for a set of different iron and iron-oxygen complexes. These nice results prompted us to collaborate with experimental groups to determine Mössbauer parameters for an Fe^{IV}-oxo species containing a tris-urea tripodal ligand. This compound changes in a formation of a new intramolecular H-bond stabilizing a protonated adduct, but there was no X-ray structure to determine which is the correct hydrogen bond interaction formed. Different possible structures were optimized based on different plausible hydrogen bond interactions. The experimental Mössbauer parameters show ΔE_Q=0.87 mm/s and δ=0.04 mm/s and our results suggested that the lowest-energy structure had a ΔE_Q=0.81 mm/s and δ=0.045 using the new fit parameters

at the S12g/TZ2P level of theory. Thus, we were able to elucidate that the Fe-oxo with an intramolecular H-bond is the preferred iron-oxygen species. These results showed really accurate values providing an opportunity to use these new fit parameters to characterize unknown Fe-oxygen intermediate species in which different possible conformations can exist.

Following the computational microscope idea, in this thesis we showed how diverse techniques allow us to solve different problems. Depending on the accuracy needed and the size of the system, the used techniques give us the information needed to understand conformational dynamics of natural and mutated enzyme, the mechanism of biomimetic complexes, and even to characterize iron-oxygen structures where no X-ray structure is available. The combination of experimental and computational techniques allow the gaining of important knowledge for understanding and improving (bio)catalysts for novel functionality.

References

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

(2) Wierenga, R. K. *FEBS Lett.* **2001**, *492*, 193-198.

(3) Dunn, M. F.; Niks, D.; Ngo, H.; Barends, T. R. M.; Schlichting, I. *Trends Biochem. Sci* **2008**, *33*, 254-264.

(4) Emil, F. *Ber. Chem. Gesellsch.* **1894**, *27*, 2985-2993.

(5) Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. *J. Mol. Biol.* **2010**, *396*, 1025-1042.

(6) Tawfik, O. K.; Dan, S. *Annu. Rev. Biochem* **2010**, *79*, 471-505.

(7) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789-796.

(8) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. *Biochemistry* **2011**, *50*, 10422-10430.

(9) Cleland, W.; Kreevoy, M. *Science* **1994**, *264*, 1887-1890.

(10) Purich, D. L. In *Enzyme Kinetics: Catalysis & Control*; Purich, D. L., Ed.; Elsevier: Boston, 2010, p 1-51.

(11) Pauling, L. *Chemical & Engineering News Archive* **1946**, *24*, 1375-1377.

(12) Copeland, R. A. *Enzymes a practical introduction to structure, mechanism, and data analysis*; Wiley-VCH: New York, 2002.

(13) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210-3235.

(14) Plattner, N.; Noé, F. *Nat. Commun.* **2015**, *6*, 7653.

(15) P., J. W. *Protein Sci.* **1994**, *3*, 2459-2464.

(16) Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 139-148.

(17) Jencks, W. P. *Catalysis in chemistry and enzymology*; Dover Publications: United States, 1987.

(18) Villà, J.; Štrajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11899-11904.

(19) Schowen, R. L. In *In Mechanistic Principles of Enzyme Activity by J. F. Liebman and A. Greenberg (Eds.): VCH Publishers: New York* 1989; Vol. 9, p 119-168.

(20) Tousignant, A.; Pelletier, J. N. *Chem. Biol.* **2004**, *11*, 1037-1042.

(21) Watney, J. B.; Agarwal, P. K.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2003**, *125*, 3745-3750.

(22) Stec, B.; Hehir, M. J.; Brennan, C.; Nolte, M.; Kantrowitz, E. R. *J. Mol. Biol.* **1998**, *277*, 647-662.

(23) Nagel, Z. D.; Klinman, J. P. *Nat. Chem. Biol.* **2009**, *5*, 543-550.

(24) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186-195.

(25) McDonald, A. G.; Boyce, S.; Tipton, K. F. *Nucl. Acids Res.* **2009**, *37*, D593-D597.

(26) Bairoch, A. *Nucl. Acids Res.* **2000**, *28*, 304-305.

(27) Loewus, F. A.; Westheimer, F. H.; Vennesland, B. *J. Am. Chem. Soc.* **1953**, *75*, 5018-5023.

(28) Keinan, E.; Hafeli, E. K.; Seth, K. K.; Lamed, R. *J. Am. Chem. Soc.* **1986**, *108*, 162-169.

(29) McPhalen, C. A.; Vincent, M. G.; Jansonius, J. N. *J. Mol. Biol.* **1992**, *225*, 495-517.

(30) Lindskog, S. *Pharmacol. Ther.* **1997**, *74*, 1-20.

(31) Wurtz, A. *Comptes rendus* **1880**, *91*, 787.

(32) Brown, A. J. *J. Chem. Soc. Trans.* **1902**, *81*, 373-388.

(33) Henri, V. *Lois générales de l'action des diastases, Hermann, Paris* **1903**.

(34) Michaelis, L.; Menten, M. *Biochemistry* **1913**, 333-369.

(35) Cornish-Bowden, A. *Perspect. Sci.* **2015**, *4*, 3-9.

(36) Briggs, G. E.; Haldane, J. B. S. *Biochem. J* **1925**, *19*, 338-339.

(37) Fersht, A. *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*; World Scientific Publishing: New Jersey, 2017.

(38) Francis, J. C.; Hansche, P. E. *Genetics* **1972**, *70*, 59-73.

(39) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012**, *485*, 185-194.

(40) Turner, N. J. *Nat. Chem. Biol.* **2009**, *5*, 567.

(41) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. *Protein Eng. Des. Sel.* **2014**, *27*, 273-280.

(42) Addington, T. A.; Mertz, R. W.; Siegel, J. B.; Thompson, J. M.; Fisher, A. J.; Filkov, V.; Fleischman, N. M.; Suen, A. A.; Zhang, C.; Toney, M. D. *J. Mol. Biol.* **2013**, *425*, 1378-1389.

(43) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387-1391.

(44) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. *Nature Chem.* **2016**, *9*, 50.

(45) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W. *Nat. Biotechnol.* **2007**, *25*, 338.

(46) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. *Chem. Soc. Rev.* **2015**, *44*, 1172-1239.

(47) Wedge, D. C.; Rowe, W.; Kell, D. B.; Knowles, J. *J. Theor. Biol.* **2009**, *257*, 131-141.

(48) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. *Chem. Commun.* **2017**, *53*, 284-297.

(49) Bornscheuer, U. T.; Pohl, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 137-143.

(50) Dalby, P. A. *Curr. Opin. Struct. Biol.* **2003**, *13*, 500-505.

(51) Renata, H.; Wang, Z. J.; Arnold, F. H. *Angew. Chem. Int. Ed.* **2015**, *54*, 3351-3367.

(52) Zeymer, C.; Hilvert, D. *Annu. Rev. Biochem* **2018**, *87*, 131-157.

(53) Glieder, A.; Farinas, E. T.; Arnold, F. H. *Nat. Biotechnol.* **2002**, *20*, 1135.

(54) Li, X.; Zhang, Z.; Song, J. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201209007.

(55) Pavelka, A.; Chovancova, E.; Damborsky, J. *Nucl. Acids Res.* **2009**, *37*, W376-W383.

(56) Robert, K.; Helge, J.; Sebastian, B.; Remko, K.; Kumar, P. S.; Markus, G.; Dominique, B.; Henk‐Jan, J.; T., B. U. *ChemBioChem* **2010**, *11*, 1635-1643.

(57) Lutz, S. *Curr. Opin. Biotechnol.* **2010**, *21*, 734-743.

(58) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15*, 2785-2794.

(59) Pavlidis, I. V.; Weiß, M. S.; Genz, M.; Spurr, P.; Hanlon, S. P.; Wirz, B.; Iding, H.; Bornscheuer, U. T. *Nature Chem.* **2016**, *8*, 1076.

(60) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew. Chem. Int. Ed.* **2013**, *52*, 5700-5725.

(61) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190-195.

(62) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309-313.

(63) Nosrati, G. R.; Houk, K. N. *Protein Sci.* **2012**, *21*, 697-706.

(64) Malisi, C.; Schumann, M.; C Toussaint, N.; Kageyama, J.; Kohlbacher, O.; Höcker, B. *PLoS ONE* **2012**, *7*, e52505.

(65) Fazelinia, H.; Cirino, P. C.; Maranas, C. D. *Protein Sci.* **2009**, *18*, 180-195.

(66) Huang, X.; Han, K.; Zhu, Y. *Protein Sci.* **2013**, *22*, 929-941.

(67) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A* **2012**, *109*, 3790-3795.

(68) Kiss, G.; Röthlisberger, D.; Baker, D.; Houk, K. N. *Protein Sci.* **2010**, *19*, 1760-1773.

(69) Lindert, S.; Meiler, J.; McCammon, J. A. *J. Chem. Theory Comput.* **2013**, *9*, 3843-3847.

(70) Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B. *Angew. Chem. Int. Ed.* **2015**, *54*, 3726-3730.

(71) Chica, R. A.; Doucet, N.; Pelletier, J. N. *Curr. Opin. Biotechnol.* **2005**, *16*, 378-384.

(72) Li, R.; Wijma, H. J.; Song, L.; Cui, Y.; Otzen, M.; Tian, Y. e.; Du, J.; Li, T.; Niu, D.; Chen, Y.; Feng, J.; Han, J.; Chen, H.; Tao, Y.; Janssen, D. B.; Wu, B. *Nat. Chem. Biol.* **2018**, *14*, 664-670.

(73) Kawata, Y.; Tamura, K.; Kawamura, M.; Ikei, K.; Mizobata, T.; Nagai, J.; Fujita, M.; Yano, S.; Tokushige, M.; Yumoto, N. *Eur. J. Biochem.* **2001**, *267*, 1847-1857.

(74) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Gruetter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. *Nature* **2013**, *503*, 418-421.

(75) Acevedo-Rocha, C. G.; Gamble, C. G.; Lonsdale, R.; Li, A.; Nett, N.; Hoebenreich, S.; Lingnau, J. B.; Wirtz, C.; Fares, C.; Hinrichs, H.; Deege, A.; Mulholland, A. J.; Nov, Y.; Leys, D.; McLean, K. J.; Munro, A. W.; Reetz, M. T. *ACS Catal.* **2018**, *8*, 3395-3410.

(76) Reetz, M. T. *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174.

(77) Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A. *Angew. Chem. Int. Ed.* **2005**, *44*, 4192-4196.

(78) Sun, Z.; Wikmark, Y.; Backvall, J. E.; Reetz, M. T. *Chemistry* **2016**, *22*, 5046-5054.

(79) Xia, Y.; Cui, W.; Cheng, Z.; Peplowski, L.; Liu, Z.; Kobayashi, M.; Zhou, Z. *ChemCatChem* **2018**, *10*, 1370-1375.

(80) Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A* **2011**, *108*, 6823-6827.

(81) Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U.S.A* **2012**, *109*, 10358-10363.

(82) Merski, M.; Shoichet, B. K. *Proc. Natl. Acad. Sci. U.S.A* **2012**, *109*, 16179-16183.

(83) Seebeck, F. P.; Hilvert, D. *J. Am. Chem. Soc.* **2005**, *127*, 1307-1312.

(84) Risso, V. A.; Martinez-Rodriguez, S.; Candel, A. M.; Krüger, D. M.; Pantoja-Uceda, D.; Ortega-Muñoz, M.; Santoyo-Gonzalez, F.; Gaucher, E. A.; Kamerlin, S. C. L.; Bruix, M.; Gavira, J. A.; Sanchez-Ruiz, J. M. *Nat. Commun.* **2017**, *8*, 16113.

(85) Isom, D. G.; Cannon, B. R.; Castañeda, C. A.; Robinson, A.; García-Moreno E, B. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17784-17788.

(86) Pey, A. L.; Rodriguez-Larrea, D.; Gavira, J. A.; Garcia-Moreno, B.; Sanchez-Ruiz, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 1218-1219.

(87) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; Hilvert, D.; Baker, D. *Protein Sci.* **2012**, *21*, 717-726.

(88) Bjelic, S.; Kipnis, Y.; Wang, L.; Pianowski, Z.; Vorobiev, S.; Su, M.; Seetharaman, J.; Xiao, R.; Kornhaber, G.; Hunt, J. F.; Tong, L.; Hilvert, D.; Baker, D. *J. Mol. Biol.* **2014**, *426*, 256-271.

(89) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. *Nat. Chem. Biol.* **2013**, *9*, 494-U449.

(90) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. *Nat. Chem.* **2017**, *9*, 50-56.

(91) Zeymer, C.; Zschoche, R.; Hilvert, D. *J. Am. Chem. Soc.* **2017**, *139*, 12541-12549.

(92) Xavier, G.; Stuart, M. D.; M., W. B. I.; Donald, H. *Angew. Chem. Int. Ed.* **2018**, *57*, 5288-5291.

(93) Garrabou, X.; Beck, T.; Hilvert, D. *Angew. Chem. Int. Ed.* **2015**, *54*, 5609-5612.

(94) Garrabou, X.; Wicky, B. I. M.; Hilvert, D. *J. Am. Chem. Soc.* **2016**, *138*, 6972-6974.

(95) Garrabou, X.; Macdonald Duncan, S.; Hilvert, D. *Chem. Eur. J.* **2017**, *23*, 6001-6003.

(96) Osuna, S.; Jiménez-Osés, G.; Noey, E. L.; Houk, K. N. *Acc. Chem. Res.* **2015**, *48*, 1080-1089.

(97) Handley, L. D.; Fuglestad, B.; Stearns, K.; Tonelli, M.; Fenwick, R. B.; Markwick, P. R. L.; Komives, E. A. *Sci. Rep.* **2017**, *7*, 39575.

(98) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hübner, C. G.; Kern, D. *Nature* **2007**, *450*, 838.

(99) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666

(100) Frank, J.; Ourmazd, A. *Methods* **2016**, *100*, 61-67.

(101) Neu, A.; Neu, U.; Fuchs, A.-L.; Schlager, B.; Sprangers, R. *Nat. Chem. Biol.* **2015**, *11*, 697-704.

(102) Baldwin, A. J.; Kay, L. E. *Nat. Chem. Biol.* **2009**, *5*, 808-814.

(103) Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J. *Nat. Chem. Biol.* **2016**, *12*, 944-950.

(104) Axe, J. M.; Yezdimer, E. M.; O'Rourke, K. F.; Kerstetter, N. E.; You, W.; Chang, C.-e. A.; Boehr, D. D. *J. Am. Chem. Soc.* **2014**, *136*, 6818-6821.

(105) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. *Nat. Chem. Biol.* **2014**, *10*, 431-436.

(106) Diez, M.; Zimmermann, B.; Borsch, M.; Konig, M.; Schweinberger, E.; Steigmiller, S.; Reuter, R.; Felekyan, S.; Kudryavtsev, V.; Seidel, C. A.; Graber, P. *Nat. Struct. Mol. Biol.* **2004**, *11*, 135-141.

(107) Myong, S.; Stevens, B. C.; Ha, T. *Structure* **2006**, *14*, 633-643.

(108) Gianni, S.; Dogan, J.; Jemth, P. *Biophys. Chem.* **2014**, *189*, 33-39.

(109) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A* **1958**, *44*, 98-104.

(110) Vogt, A. D.; Di Cera, E. *Biochemistry* **2012**, *51*, 5894-5902.

(111) Monod, J.; Wyman, J.; Changeux, J.-P. *J. Mol. Biol.* **1965**, *12*, 88-118.

(112) Kovermann, M.; Grundstrom, C.; Sauer-Eriksson, A. E.; Sauer, U. H.; Wolf-Watz, M. *Proc. Natl. Acad. Sci. U.S.A* **2017**, *114*, 6298-6303.

(113) Malabanan, M. M.; Amyes, T. L.; Richard, J. P. *Curr. Opin. Struct. Biol.* **2010**, *20*, 702-710.

(114) Azzi, A.; Clark, S. A.; Ellington, W. R.; Chapman, M. S. *Protein Sci.* **2004**, *13*, 575-585.

(115) Balasco, N.; Esposito, L.; Simone, A. D.; Vitagliano, L. *Protein Sci.* **2013**, *22*, 1016-1023.

(116) Nestl, B. M.; Hauer, B. *ACS Catal.* **2014**, *4*, 3201-3211.

(117) Furnham, N.; Sillitoe, I.; Holliday, G. L.; Cuff, A. L.; Laskowski, R. A.; Orengo, C. A.; Thornton, J. M. *PLOS Comput. Biol.* **2012**, *8*, e1002403.

(118) Stojanoski, V.; Chow, D.-C.; Hu, L.; Sankaran, B.; Gilbert, H. F.; Prasad, B. V. V.; Palzkill, T. *J. Biol. Chem.* **2015**, *290*, 10382-10394.

(119) Clifton, Ben E.; Jackson, Colin J. *Cell Chem. Biol.* **2016**, *23*, 236-245.

(120) Tyukhtenko, S.; Rajarshi, G.; Karageorgos, I.; Zvonok, N.; Gallagher, E. S.; Huang, H.; Vemuri, K.; Hudgens, J. W.; Ma, X.; Nasr, M. L.; Pavlopoulos, S.; Makriyannis, A. *Sci. Rep.* **2018**, *8*, 1719.

(121) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433-443.

(122) Hanks, T. W.; Swiegers, G. F. In *Introduction: The Concept of Biomimicry and Bioinspiration in Chemistry by Gerhard F. Swiegers (Eds.)* 2012, p 1-15.

(123) Lehn, J.-M. *Resonance* **1996**, *1*, 39-53.

(124) Roat-Malone, R. M. *Bioinorganic chemistry : a short course*; John Wiley & Sons: Chichester, 2007.

(125) Ortiz de Montellano, P. R. *Cytochrome P450 : structure, mechanism, and biochemistry*; Springer International Publishing, 2015.

(126) Fertinger, C.; Hessenauer‐Ilicheva, N.; Franke, A.; van Eldik, R. *Chem. Eur. J.* **2009**, *15*, 13435-13440.

(127) Franke, A.; Fertinger, C.; van Eldik, R. *Angew. Chem. Int. Ed.* **2008**, *47*, 5238-5242.

(128) Guallar, V.; Wallrapp, F. H. *Biophys. Chem.* **2010**, *149*, 1-11.

(129) Simonneaux, G.; Bondon, A. *Chem. Rev.* **2005**, *105*, 2627-2646.

(130) Huang, X.; Groves, J. T. *Chem. Rev.* **2018**, *118*, 2491-2553.

(131) Guallar, V.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 6998.

(132) Groves, J. T.; Haushalter, R. C.; Nakamura, M.; Nemo, T. E.; Evans, B. J. *J. Am. Chem. Soc.* **1981**, *103*, 2884-2886.

(133) Groves, J. T. *J. Inorg. Biochem.* **2006**, *100*, 434-447.

(134) Fujii, H. *Coord. Chem. Rev.* **2002**, *226*, 51-60.

(135) Linhares, M.; Rebelo, S. L. H.; Simões, M. M. Q.; Silva, A. M. S.; Neves, M. G. P. M. S.; Cavaleiro, J. A. S.; Freire, C. *Appl. Catal., A* **2014**, *470*, 427-433.

(136) Solomon, E. I.; Decker, A.; Lehnert, N. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 3589-3594.

(137) Solomon, E. I.; Goudarzi, S.; Sutherlin, K. D. *Biochemistry* **2016**, *55*, 6363-6374.

(138) Solomon, E. I.; Brunold, T. C.; Davis, M. I.; Kemsley, J. N.; Lee, S.-K.; Lehnert, N.; Neese, F.; Skulan, A. J.; Yang, Y.-S.; Zhou, J. *Chem. Rev.* **2000**, *100*, 235-350.

(139) Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M.; Krebs, C. *Biochemistry* **2003**, *42*, 7497-7508.

(140) Nam, W. *Acc. Chem. Res.* **2007**, *40*, 522-531.

(141) Kryatov, S. V.; Rybak-Akimova, E. V.; Schindler, S. *Chem. Rev.* **2005**, *105*, 2175-2226.

(142) Andrea, D.; I., S. E. *Angew. Chem. Int. Ed.* **2005**, *44*, 2252-2255.

(143) Nam, W.; Lee, Y.-M.; Fukuzumi, S. *Acc. Chem. Res.* **2014**, *47*, 1146-1154.

(144) Dismukes, G. C. *Chem. Rev.* **1996**, *96*, 2909-2926.

(145) Barynin, V. V.; Whittaker, M. M.; Antonyuk, S. V.; Lamzin, V. S.; Harrison, P. M.; Artymiuk, P. J.; Whittaker, J. W. *Structure* **2001**, *9*, 725-738.

(146) E. M. Siegbahn, P. *Theor. Chem. Acc.* **2001**, *105*, 197-206.

(147) Whittaker, J. W. *Arch. Biochem. Biophys.* **2012**, *525*, 111-120.

(148) Signorella, S.; Hureau, C. *Coord. Chem. Rev.* **2012**, *256*, 1229-1245.

(149) Doctrow, S. R.; Huffman, K.; Marcus, C. B.; Tocco, G.; Malfroy, E.; Adinolfi, C. A.; Kruk, H.; Baker, K.; Lazarowych, N.; Mascarenhas, J.; Malfroy, B. *J. Med. Chem.* **2002**, *45*, 4549-4558.

(150) Abashkin, Y. G.; Burt, S. K. *Inorg. Chem.* **2005**, *44*, 1425-1432.

(151) Liu, S.-Y.; Soper, J. D.; Yang, J. Y.; Rybak-Akimova, E. V.; Nocera, D. G. *Inorg. Chem.* **2006**, *45*, 7572-7574.

(152) Maneiro, M.; R. Bermejo, M.; Garcia, M. I.; Gómez-Fórneas, E.; González-Noya, A.; M. Tyryshkin, A. *New J. Chem.,* **2003**, *27*, 727-733.

(153) Costas, M.; Chen, K.; Que, L. *Coord. Chem. Rev.* **2000**, *200-202*, 517-544.

(154) Borovik, A. S. *Acc. Chem. Res.* **2005**, *38*, 54-61.

(155) Rebilly, J.-N.; Colasson, B.; Bistri, O.; Over, D.; Reinaud, O. *Chem. Soc. Rev.* **2015**, *44*, 467-489.

(156) Swart, M.; Gruden, M. *Acc. Chem. Res.* **2016**, *49*, 2690-2697.

(157) Swart, M.; Costas, M. *Spin states in biochemistry and inorganic chemistry : influence on structure and reactivity*; John Wiley & Sons: Chichester, 2016.

(158) P. Shaver, M.; E. N. Allan, L.; Rzepa S., H.; Gibson C., V. *Angew. Chem. Int. Ed.* **2006**, *45*, 1241-1244.

(159) Poli, R.; N Harvey, J. *Chem. Soc. Rev.* **2003**, *32*, 1-8.

(160) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279-2328.

(161) Lancaster, K. M.; Roemelt, M.; Ettenhuber, P.; Hu, Y.; Ribbe, M. W.; Neese, F.; Bergmann, U.; DeBeer, S. *Science* **2011**, *334*, 974-977.

(162) Saouma, C. T.; Mayer, J. M. *Chem. Sci.* **2014**, *5*, 21-31.

(163) Scepaniak, J. J.; Vogel, C. S.; Khusniyarov, M. M.; Heinemann, F. W.; Meyer, K.; Smith, J. M. *Science* **2011**, *331*, 1049-1052.

(164) Brazzolotto, D.; Gennari, M.; Queyriaux, N.; Simmons, T. R.; Pécaut, J.; Demeshko, S.; Meyer, F.; Orio, M.; Artero, V.; Duboc, C. *Nature Chem.* **2016**, *8*, 1054-1060.

(165) Jensen, F. *Introduction to computational chemistry*, 2017.

(166) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133-A1138.

(167) Car, R. *Nature Chem.* **2016**, *8*, 820.

(168) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403-412.

(169) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.

(170) Swart, M.; Ehlers, A. W.; Lammertsma, K. *Mol. Phys.* **2004**, *102*, 2467-2474.

(171) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479-5483.

(172) Swart, M. *Chem. Commun.* **2013**, *49*, 6650-6652.

(173) T. te Velde, G.; Bickelhaupt, F. M.; Baerends, E.; Fonseca Guerra, C.; Gisbergen, S.; G. Snijders, J.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931-967.

(174) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perk. Trans. 2* **1993**, 799-805.

(175) Pye, C. C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396-408.

(176) Swart, M.; Rösler, E.; Bickelhaupt, F. M. *Eur. J. Inorg. Chem.* **2007**, *2007*, 3646-3654.

(177) Antony, J.; Grimme, S. *PCCP* **2006**, *8*, 5287-5293.

(178) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(179) Stefan, G. *J. Comput. Chem.* **2006**, *27*, 1787-1799.

(180) Lenthe, E. v.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597-4610.

(181) In *Quantum Biochemistry: Electronic structure and biological activity*; Matta, C. F. E., Ed.; Wiley, 2010; Vol. 2, p 551-583.

(182) Swart, M.; Güell, M.; Luis, J. M.; Solà, M. *J. Phys. Chem. A* **2010**, *114*, 7191-7197.

(183) Greenwood, N. N.; Gibb, T. C.; Springer Netherlands: Dordrecht, 1971.

(184) Mössbauer, R. L. *Z. Phys.* **1958**, *151*, 124-143.

(185) Mössbauer, R. L. *Naturwissenschaften* **1958**, *45*, 538-539.

(186) Pandelia, M.-E.; Lanz, N. D.; Booker, S. J.; Krebs, C. *Biochim. Biophys. Acta* **2015**, *1853*, 1395-1405.

(187) Liu, T.; Lovell, T.; Han, W.-G.; Noodleman, L. *Inorg. Chem.* **2003**, *42*, 5244-5251.

(188) Neese, F. *Inorg. Chim. Acta* **2002**, *337*, 181-192.

(189) Han, W. G.; Liu, T.; Lovell, T.; Noodleman, L. *J. Comput. Chem.* **2006**, *27*, 1292-1306.

(190) Han, W.-G.; Noodleman, L. *Inorg. Chim. Acta* **2008**, *361*, 973-986.

(191) Hopmann, K. H.; Ghosh, A.; Noodleman, L. *Inorg. Chem.* **2009**, *48*, 9155-9165.

(192) Hedegard, E. D.; Knecht, S.; Ryde, U.; Kongsted, J.; Saue, T. *PCCP* **2014**, *16*, 4853-4863.

(193) Pápai, M.; Vankó, G. *J. Chem. Theory Comput.* **2013**, *9*, 5004-5020.

(194) Bochevarov, A. D.; Friesner, R. A.; Lippard, S. J. *J. Chem. Theory Comput.* **2010**, *6*, 3735-3749.

(195) Römelt, M.; Ye, S.; Neese, F. *Inorg. Chem.* **2009**, *48*, 784-785.

(196) Zhang, Y.; Mao, J.; Oldfield, E. *J. Am. Chem. Soc.* **2002**, *124*, 7829-7839.

(197) Bochevarov, A. D.; Friesner, R. A. *J. Chem. Phys.* **2008**, *128*, 034102.

(198) Houk, K. N.; Cheong, P. H.-Y. *Nature* **2008**, *455*, 309.

(199) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400-417.

(200) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

(201) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113-5123.

(202) Halgren Thomas, A. *J. Comput. Chem.* **1996**, *17*, 490-519.

(203) Breneman Curt, M.; Wiberg Kenneth, B. *J. Comput. Chem.* **1990**, *11*, 361-373.

(204) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Chem. Phys.* **1993**, *97*, 10269-10280.

(205) González, M. A. *Collection SFN* **2011**, *12*, 169-200.

(206) Berweger, C. D.; van Gunsteren, W. F.; Müller-Plathe, F. *Chem. Phys. Lett.* **1995**, *232*, 429-436.

(207) Seminario, J. M. *Int. J. Quantum Chem.* **1996**, *60*, 1271-1277.

(208) Hu, L.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 2452-2463.

(209) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. *J. Chem. Theory Comput.* **2013**, *9*, 2733-2748.

(210) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Åqvist, J.; Kamerlin, S. C. L. *J. Phys. Chem. B* **2014**, *118*, 4351-4362.

(211) Li, P.; Merz, K. M. *J. Chem. Inf. Model.* **2016**, *56*, 599-604.

(212) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2003**, *59*, 274-289.

(213) Li, P.; Merz, K. M. *J. Chem. Theory Comput.* **2014**, *10*, 289-297.

(214) Li, P.; Song, L. F.; Merz, K. M. *J. Phys. Chem. B* **2015**, *119*, 883-895.

(215) Li, P.; Song, L. F.; Merz, K. M. *J. Chem. Theory Comput.* **2015**, *11*, 1645-1657.

(216) Aaqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860-2868.

(217) Lu, S. Y.; Huang, Z. M.; Huang, W. K.; Liu, X. Y.; Chen, Y. Y.; Shi, T.; Zhang, J. *Proteins* **2012**, *81*, 740-753.

(218) Jiang, Y.; Zhang, H.; Feng, W.; Tan, T. *J. Chem. Inf. Model.* **2015**, *55*, 2575-2586.

(219) Liao, Q.; Kamerlin, S. C. L.; Strodel, B. *J. Phys. Chem. Lett.* **2015**, *6*, 2657-2662.

(220) Li, P.; Merz, K. M. *Chem. Rev.* **2017**, *117*, 1564-1686.

(221) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. *J. Am. Chem. Soc.* **2014**, *136*, 13959-13962.

(222) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795-808.

(223) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A* **2018**.

(224) Nerenberg, P. S.; Head-Gordon, T. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129-138.

(225) Aytenfisu, A. H.; Spasic, A.; Seetin, M. G.; Serafini, J.; Mathews, D. H. *J. Chem. Theory Comput.* **2014**, *10*, 1292-1301.

(226) Tzanov, A. T.; Cuendet, M. A.; Tuckerman, M. E. *J. Phys. Chem. B* **2014**, *118*, 6539-6552.

(227) Henriques, J.; Cragnell, C.; Skepö, M. *J. Chem. Theory Comput.* **2015**, *11*, 3420-3431.

(228) Robustelli, P.; Piana, S.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A* **2018**.

(229) Huang, J.; MacKerell, A. D. *Curr. Opin. Struct. Biol.* **2018**, *48*, 40-48.

(230) Best, R. B.; Mittal, J. *Proteins* **2011**, *79*, 1318-1328.

(231) McKiernan, K. A.; Husic, B. E.; Pande, V. S. *J. Chem. Phys.* **2017**, *147*, 104107.

(232) Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.

(233) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press, 1989.

(234) Bottaro, S.; Lindorff-Larsen, K. *Science* **2018**, *361*, 355.

(235) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(236) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1988**, *89*, 5115-5127.

(237) Hess, B.; Bekker, H.; Berendsen Herman, J. C.; Fraaije Johannes, G. E. M. *J. Comput. Chem.* **1998**, *18*, 1463-1472.

(238) Bekker, H. *J. Comput. Chem.* **1998**, *18*, 1930-1942.

(239) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

(240) Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. *J. Chem. Theory Comput.* **2015**, *11*, 2938-2944.

(241) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325-348.

(242) Zuckerman, D. M. *Annu. Rev. Biophys.* **2011**, *40*, 41-62.

(243) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964-972.

(244) Maria-Solano, M. A.; Serrano-Hervas, E.; Romero-Rivera, A.; Iglesias-Fernandez, J.; Osuna, S. *Chem. Commun.* **2018**.

(245) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437-449.

(246) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341-346.

(247) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140-150.

(248) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919-11929.

(249) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562-12566.

(250) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187-199.

(251) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.

(252) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. *J. Comput. Chem.* **2015**, *36*, 1536-1549.

(253) Markwick, P. R.; McCammon, J. A. *Phys Chem Chem Phys* **2011**, *13*, 20053-20065.

(254) Huang, Y.-m. M.; McCammon, J. A.; Miao, Y. *J. Chem. Theory Comput.* **2018**, *14*, 1853-1864.

(255) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304-1311.

(256) Ringner, M. *Nat. Biotechnol.* **2008**, *26*, 303-304.

(257) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. *J. Chem. Phys.* **2013**, *139*, 015102.

(258) Ernst, M.; Sittel, F.; Stock, G. *J. Chem. Phys.* **2015**, *143*, 244114.

(259) Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z. *Structure* **2017**, *25*, 276-286.

(260) Naritomi, Y.; Fuchigami, S. *J. Chem. Phys.* **2013**, *139*, 215102.

(261) Koshland, D. E. *Angew. Chem. Int. Ed.* **1994**, *33*, 2375-2378.

(262) Thayer, K. M.; Lakhani, B.; Beveridge, D. L. *J. Phys. Chem. B* **2017**, *121*, 5509-5514.

(263) Fusetti, F.; Schröter, K. H.; Steiner, R. A.; van Noort, P. I.; Pijning, T.; Rozeboom, H. J.; Kalk, K. H.; Egmond, M. R.; Dijkstra, B. W. *Structure* **2002**, *10*, 259-268.

(264) Steiner, R. A.; Kalk, K. H.; Dijkstra, B. W. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 16625.

(265) Antonczak, S.; Fiorucci, S.; Golebiowski, J.; Cabrol-Bass, D. *PCCP* **2009**, *11*, 1491-1501.

(266) Siegbahn, P. E. M. *Inorg. Chem.* **2004**, *43*, 5944-5953.

(267) Saito, T.; Kawakami, T.; Yamanaka, S.; Okumura, M. *J. Phys. Chem. B* **2015**, *119*, 6952-6962.

(268) Wang, W.-J.; Wei, W.-J.; Liao, R.-Z. *PCCP* **2018**, *20*, 15784-15794.

(269) Li, H.; Wang, X.; Tian, G.; Liu, Y. *Catal. Sci. Technol.* **2018**.

(270) Fiorucci, S.; Golebiowski, J.; Cabrol-Bass, D.; Antonczak, S. *Proteins* **2007**, *67*, 961-970.

(271) van den Bosch, M.; Swart, M.; van Gunsteren, W. F.; Canters, G. W. *J. Mol. Biol.* **2004**, *344*, 725-738.

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

(273) Fiser, A.; Do, R. K.; Sali, A. *Protein Sci.* **2000**, *9*, 1753-1773.

(274) Fiser, A.; Sali, A. *Bioinformatics* **2003**, *19*, 2500-2501.

(275) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(276) *TURBOMOLE V70 (2015) A development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH.*

(277) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

(278) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(279) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. *J. Chem. Theory Comput.* **2015**, *11*, 5525-5542.

(280) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. *PLOS Comput. Biol.* **2012**, *8*, e1002708.

(281) Durrant, J. D.; Votapka, L.; Sorensen, J.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 5047-5056.

(282) Bolt, A.; Berry, A.; Nelson, A. *Arch. Biochem. Biophys.* **2008**, *474*, 318-330.

(283) Clapés, P.; Fessner, W.-D.; Sprenger, G. A.; Samland, A. K. *Curr. Opin. Chem. Biol.* **2010**, *14*, 154-167.

(284) Dean, S. M.; Greenberg, W. A.; Wong, C.-H. *Adv. Synth. Catal.* **2007**, *349*, 1308-1320.

(285) Fesko, K.; Gruber-Khadjawi, M. *ChemCatChem* **2013**, *5*, 1248-1272.

(286) Hsu, C.-C.; Hong, Z.; Wada, M.; Franke, D.; Wong, C.-H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9122-9126.

(287) Güclü, D.; Szekrenyi, A.; Garrabou, X.; Kickstein, M.; Junker, S.; Clapés, P.; Fessner, W.-D. *ACS Catal.* **2016**, *6*, 1848-1852.

(288) Chou, C.-Y.; Ko, T.-P.; Wu, K.-J.; Huang, K.-F.; Lin, C.-H.; Wong, C.-H.; Wang, A. H. J. *J. Biol. Chem.* **2011**, *286*, 14057-14064.

(289) Groher, A.; Hoelsch, K. *J. Mol. Catal. B: Enzym.* **2012**, *83*, 1-7.

(290) Daniels, A. D.; Campeotto, I.; van der Kamp, M. W.; Bolt, A. H.; Trinh, C. H.; Phillips, S. E. V.; Pearson, A. R.; Nelson, A.; Mulholland, A. J.; Berry, A. *ACS Chem. Biol.* **2014**, *9*, 1025-1032.

(291) Lin, C. H.; Sugai, T.; Halcomb, R. L.; Ichikawa, Y.; Wong, C. H. *J. Am. Chem. Soc.* **1992**, *114*, 10138-10145.

(292) Campeotto, I.; Bolt, A. H.; Harman, T. A.; Dennis, C.; Trinh, C. H.; Phillips, S. E. V.; Nelson, A.; Pearson, A. R.; Berry, A. *J. Mol. Biol.* **2010**, *404*, 56-69.

(293) Joerger, A. C.; Mayer, S.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 5694.

(294) Barbosa, J. A. R. G.; Smith, B. J.; DeGori, R.; Ooi, H. C.; Marcuccio, S. M.; Campi, E. M.; Jackson, W. R.; Brossmer, R.; Sommer, M.; Lawrence, M. C. *J. Mol. Biol.* **2000**, *303*, 405-421.

(295) Obexer, R.; Studer, S.; Giger, L.; Pinkas, D. M.; Gruetter, M. G.; Baker, D.; Hilvert, D. *ChemCatChem* **2014**, *6*, 1043-1050.

(296) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2009**, *31*, 455-461.

(297) Lassila, J. K.; Baker, D.; Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4937-4942.

(298) Garrabou, X.; Verez, R.; Hilvert, D. *J. Am. Chem. Soc.* **2017**, *139*, 103-106.

(299) Tokuriki, N.; Tawfik, D. S. *Science* **2009**, *324*, 203-207.

(300) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6620-6625.

(301) Rivalta, I.; Sultan, M. M.; Lee, N.-S.; Manley, G. A.; Loria, J. P.; Batista, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1428-E1436.

(302) Csárdi, G.; Nepusz, T. *InterJournal, Complex Systems* **2006**, *1695*, 1695-1704.

(303) Dodani, S. C.; Kiss, G.; Cahn, J. K.; Su, Y.; Pande, V. S.; Arnold, F. H. *Nat. Chem.* **2016**, *8*, 419-425.

(304) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. *ACS Catal.* **2017**, *7*, 8524-8532.

(305) Kiss, G.; Röthlisberger, D.; Baker, D.; Houk, K. N. *Protein Sci.* **2010**, *19*, 1760-1773.

(306) Guo, J.; Zhou, H.-X. *Chem. Rev.* **2016**, *116*, 6503-6515.

(307) VanWart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. *J. Chem. Theory Comput.* **2012**, *8*, 2949-2961.

(308) Lichtarge, O.; Bourne, H. R.; Cohen, F. E. *J. Mol. Biol.* **1996**, *257*, 342-358.

(309) Hennig, M.; Darimont, B. D.; Jansonius, J. N.; Kirschner, K. *J. Mol. Biol.* **2002**, *319*, 757-766.

(310) Lisi, G. P.; Loria, J. P. *Chem. Rev.* **2016**, *116*, 6323-6369.

(311) Nagano, N.; Orengo, C. A.; Thornton, J. M. *J. Mol. Biol.* **2002**, *321*, 741-765.

(312) Bayani, U.; Ajay, V. S.; Paolo, Z.; Mahajan, R. T. *Curr. Neuropharmacol.* **2009**, *7*, 65-74.

(313) Nicholls, P.; Fita, I.; Loewen, P. C. In *Adv. Inorg. Chem.*; Academic Press: 2000; Vol. 51, p 51-106.

(314) Zámocký, M.; Koller, F. *Prog. Biophys. Mol. Biol.* **1999**, *72*, 19-66.

(315) Gutteridge, J. M. C.; Halliwell, B. *Free Radical Biol. Med.* **1992**, *12*, 93-94.

(316) Doctrow, S. R.; Huffman, K.; Marcus, C. B.; Musleh, W.; Bruce, A.; Baudry, M.; Malfroy, B. In *Advances in Pharmacology*; Sies, H., Ed.; Academic Press: 1996; Vol. 38, p 247-269.

(317) Melov, S.; Ravenscroft, J.; Malik, S.; Gill, M. S.; Walker, D. W.; Clayton, P. E.; Wallace, D. C.; Malfroy, B.; Doctrow, S. R.; Lithgow, G. J. *Science* **2000**, *289*, 1567.

(318) Gonzalez, P. K.; Zhuang, J.; Doctrow, S. R.; Malfroy, B.; Benson, P. F.; Menconi, M. J.; Fink, M. P. *J. Pharmacol. Exp. Ther.* **1995**, *275*, 798.

(319) Abashkin, Y. G.; Burt, S. K. *J. Phys. Chem. B* **2004**, *108*, 2708-2711.

(320) Merlini, M. L.; Britovsek, G. J. P.; Swart, M.; Belanzoni, P. *ACS Catal.* **2018**, *8*, 2944-2958.

(321) Boelrijk, A. E. M.; Dismukes, G. C. *Inorg. Chem.* **2000**, *39*, 3020-3028.

(322) Rahman, I. *Biochim. Biophys. Acta* **2012**, *1822*, 714-728.

(323) Melov, S.; Doctrow, S. R.; Schneider, J. A.; Haberson, J.; Patel, M.; Coskun, P. E.; Huffman, K.; Wallace, D. C.; Malfroy, B. *J. Neurosci.* **2001**, *21*, 8348.

(324) Swart, M. *Chem. Phys. Lett.* **2013**, *580*, 166-171.

(325) Swart, M.; Matthias Bickelhaupt, F. *Int. J. Quantum Chem.* **2006**, *106*, 2536-2544.

(326) Klamt, A. *J. Chem. Phys.* **1995**, *99*, 2224-2235.

(327) Klamt, A.; Jonas, V. *J. Chem. Phys.* **1996**, *105*, 9972-9981.

(328) Krebs, C.; Martin Bollinger, J. *Photosynth. Res.* **2009**, *102*, 295.

(329) Gutlich; Philipp; Bill; Eckhard; Trautwein; X., A. *Mössbauer spectroscopy and transition metal chemistry: fundamentals and applications*; Springer-Verlag Berlin AN: [S.l.], 2016.

(330) Sandala, G. M.; Hopmann, K. H.; Ghosh, A.; Noodleman, L. *J. Chem. Theory Comput.* **2011**, *7*, 3232-3247.

(331) McWilliams, S. F.; Brennan-Wydra, E.; MacLeod, K. C.; Holland, P. L. *ACS Omega* **2017**, *2*, 5973-5973.

(332) Hill, E. A.; Weitz, A. C.; Onderko, E.; Romero-Rivera, A.; Guo, Y.; Swart, M.; Bominaar, E. L.; Green, M. T.; Hendrich, M. P.; Lacy, D. C.; Borovik, A. S. *J. Am. Chem. Soc.* **2016**, *138*, 13143-13146.

(333) Rohde, J.-U.; In, J.-H.; Lim, M. H.; Brennessel, W. W.; Bukowski, M. R.; Stubna, A.; Münck, E.; Nam, W.; Que, L. *Science* **2003**, *299*, 1037.

(334) Klinker, E. J.; Kaizer, J.; Brennessel, W. W.; Woodrum, N. L.; Cramer, C. J.; Que, L. *Angew. Chem. Int. Ed.* **2005**, *44*, 3690-3694.

(335) Thibon, A.; England, J.; Martinho, M.; Young, V. G.; Frisch, J. R.; Guillot, R.; Girerd, J.-J.; Münck, E.; Que, L.; Banse, F. *Angew. Chem. Int. Ed.* **2008**, *47*, 7064-7067.

(336) England, J.; Guo, Y.; Van Heuvelen, K. M.; Cranswick, M. A.; Rohde, G. T.; Bominaar, E. L.; Münck, E.; Que, L. *J. Am. Chem. Soc.* **2011**, *133*, 11880-11883.

(337) Lacy, D. C.; Gupta, R.; Stone, K. L.; Greaves, J.; Ziller, J. W.; Hendrich, M. P.; Borovik, A. S. *J. Am. Chem. Soc.* **2010**, *132*, 12188-12190.

(338) MacBeth, C. E.; Golombek, A. P.; Young, V. G.; Yang, C.; Kuczera, K.; Hendrich, M. P.; Borovik, A. S. *Science* **2000**, *289*, 938.

(339) Hersleth, H.-P.; Ryde, U.; Rydberg, P.; Görbitz, C. H.; Andersson, K. K. *J. Inorg. Biochem.* **2006**, *100*, 460-476.

(340) de Visser, S. P.; Rohde, J.-U.; Lee, Y.-M.; Cho, J.; Nam, W. *Coord. Chem. Rev.* **2013**, *257*, 381-393.

(341) Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L. *Chem. Rev.* **2004**, *104*, 939-986.

(342) Gumiero, A.; Metcalfe, C. L.; Pearson, A. R.; Raven, E. L.; Moody, P. C. E. *J. Biol. Chem.* **2011**, *286*, 1260-1268.

(343) Chreifi, G.; Baxter, E. L.; Doukov, T.; Cohen, A. E.; McPhillips, S. E.; Song, J.; Meharenna, Y. T.; Soltis, S. M.; Poulos, T. L. *Proc. Natl. Acad. Sci. U.S.A* **2016**, *113*, 1226.

(344) Casadei, C. M.; Gumiero, A.; Metcalfe, C. L.; Murphy, E. J.; Basran, J.; Concilio, M. G.; Teixeira, S. C. M.; Schrader, T. E.; Fielding, A. J.; Ostermann, A.; Blakeley, M. P.; Raven, E. L.; Moody, P. C. E. *Science* **2014**, *345*, 193.

Appendices

Supplementary Information Chapter 4

Appendix A

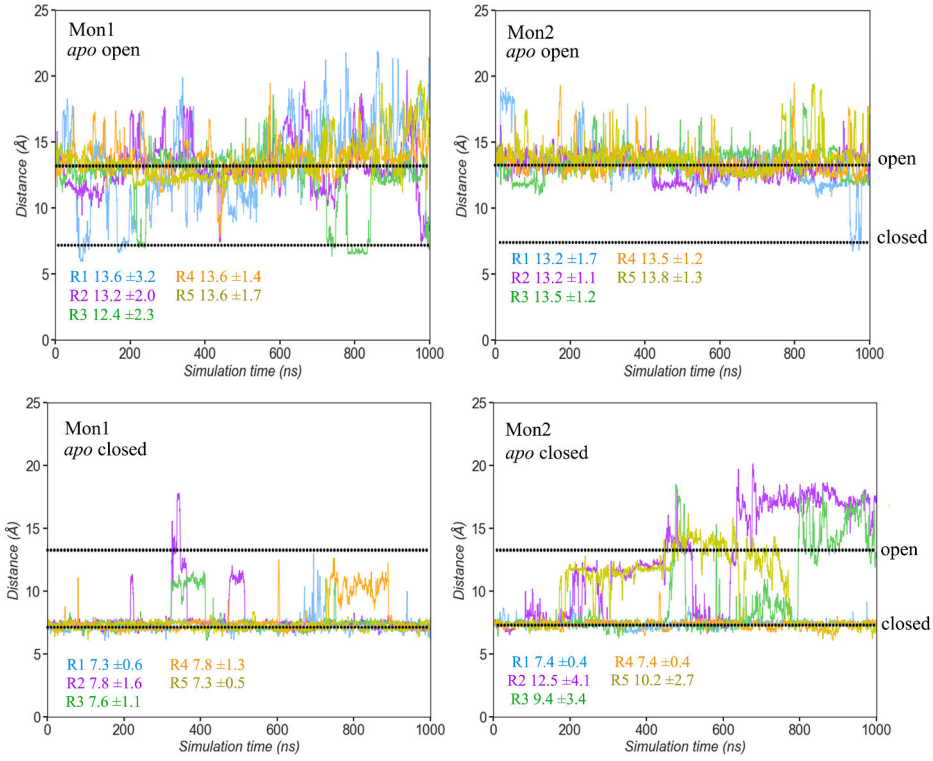


Figure A.1 Distances of Leu62-Ser164 in Å along time (ns) for both monomers starting in the *apo* open (native state) and closed conformation.

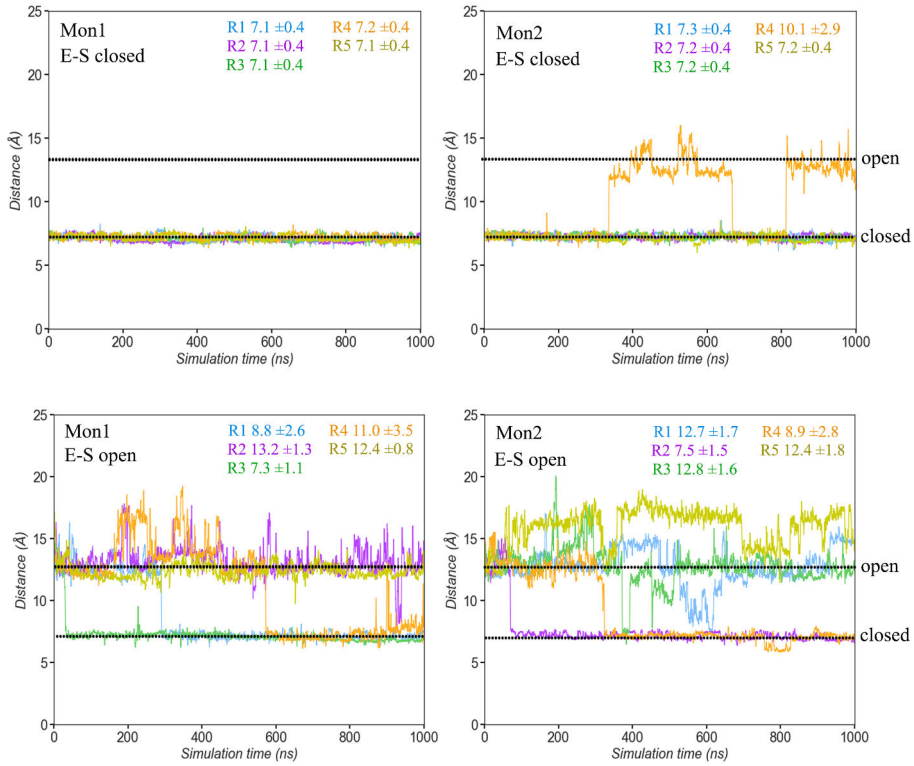


Figure A.2 Distances of Leu62-Ser164 in Å along time (ns) for both monomers starting in the E-S closed and open.

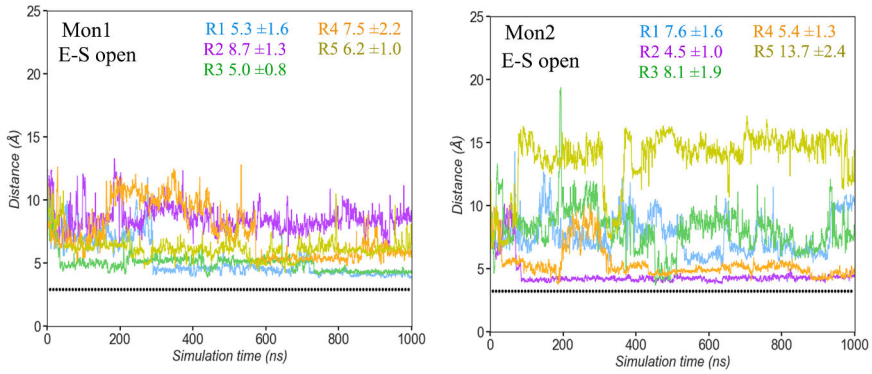


Figure A.3 Pro164 and KMP distances for both monomers starting from E-S complex in open conformation and the black dot line is the distance observed in the X-ray structure of the E-S complex state.

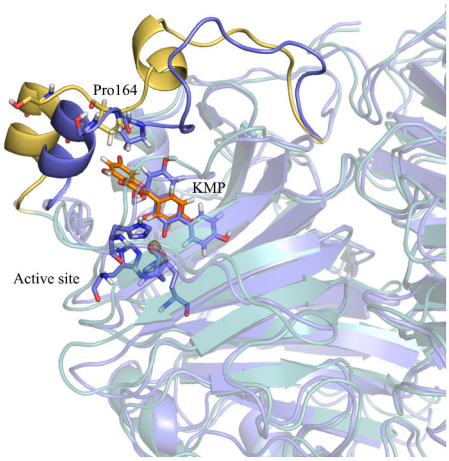


Figure A.4 X-ray structure (blue) and the frame (dark yellow for the lid domain and steal rest of the protein) where the KMP is closer to the ES complex X-ray structures, highlighting the flexible lid domain and active site.

Supplementary Information Chapter 6

Appendix B

Table B. 1 Mutations and kinetic properties of the RA variants.

Enzyme	N° of mutations	Mutations	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
RA95.0[87]	16	K10E, F22V, E51V, K53E, S70A, L83T, K110S, E159L, N180S, L184F, L187G, E210K, S211L, G233S, F246L, L247E	$5x10^{-5}$	300	0.17
RA95.5[90]	6	V51Y, E53S, T83K, M180F, R182M, D183N	$4.3x10^{-4}$	270	16
RA95.5-5[89]	6	R23H, R43S, E53T, T95M, S110N, G178S,	0.17	230	320
RA95.5-8[90]	5	S43R, F72Y, K135N, S178V, G212D	0.36	230	1600
RA95.5-8F[90]	13	T53L, R75P, N90D, N135E, S151G, V178T, F180Y, A209P, K210L, I213F, S214F, R216P, L231M	10.8 ±0.6	320 ± 36	34000

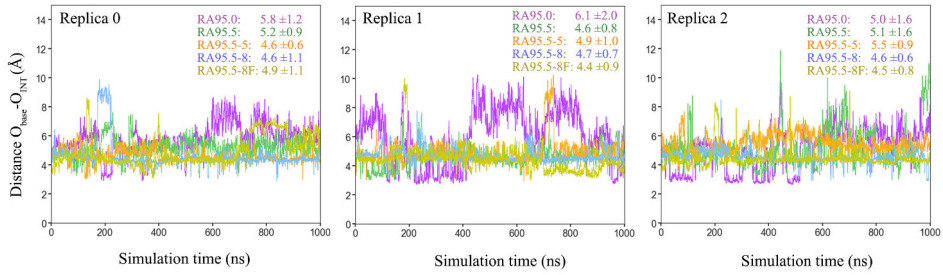


Figure B.1 Plot of the distance between the base and the catalytic lysine that will be involved in the Schiff base intermediate formation along the three 1 microsecond MD trajectories for RA95.0 (purple), RA95.5 (green), RA95.5-5 (orange), RA95.5-8 (blue), and RA95.5-8F (yellow). All distances are in Å.

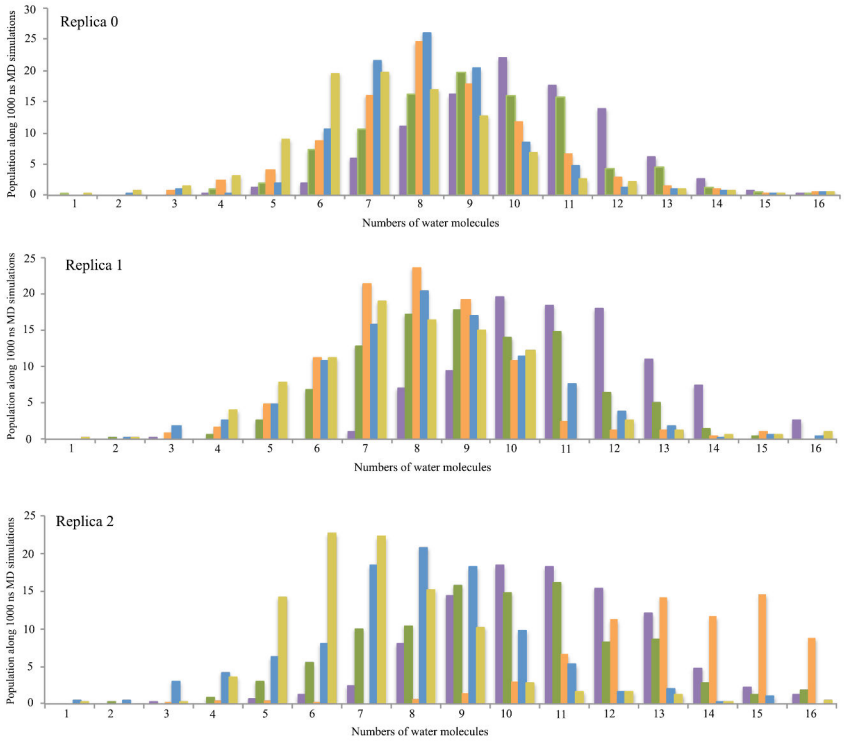


Figure B.2 Water-shell estimation along the MD trajectories for the different RA variants in their *apo* state. Number of water molecules surrounding the catalytic lysine, which suggest a change of acidity from RA95.0 to RA95.5-8F. RA95.0 (purple), RA95.5 (green), RA95.5-5 (orange), RA95.5-8 (blue), and RA95.5-8F (yellow).

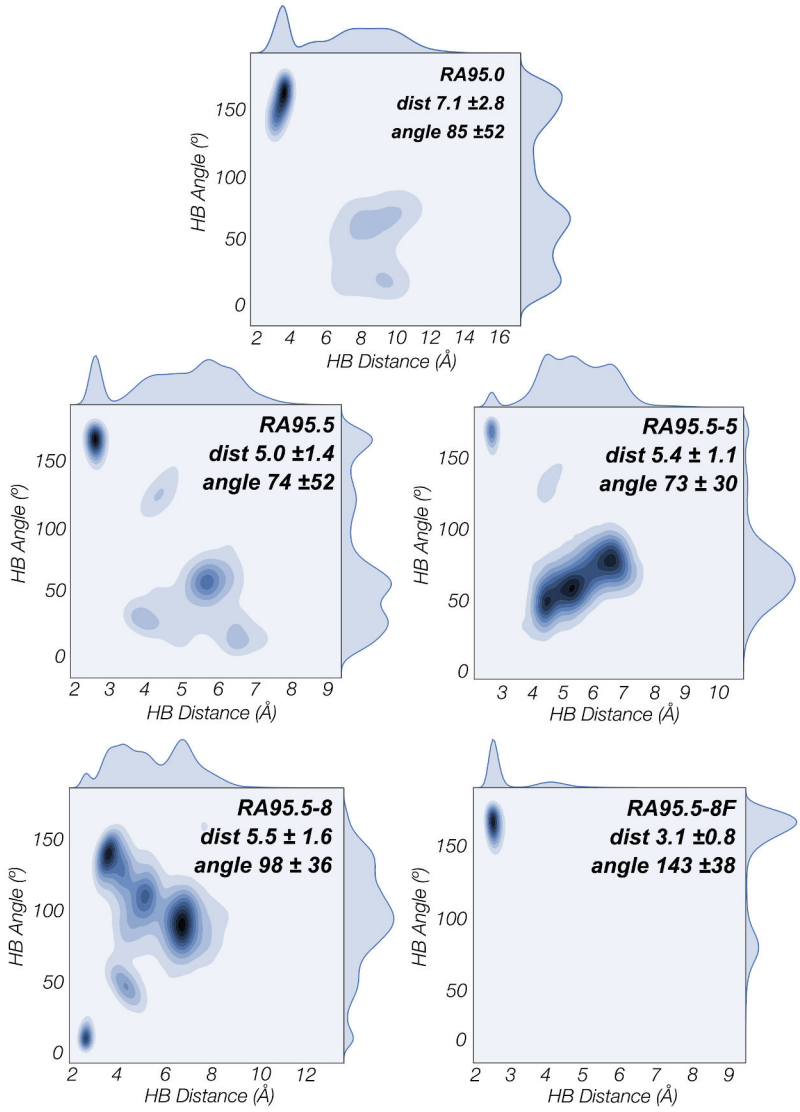


Figure B.3 Schiff base intermediate distances and angles. Representation of the distance between the deprotonated base and the oxygen of the Schiff base β-alcohol is represented (in Å) and the angle (in °) of the hydrogen from the oxygen of the Schiff base β-alcohol along the MD trajectories. The representations show the different H-bond states explored for the deprotonation step of the reaction.

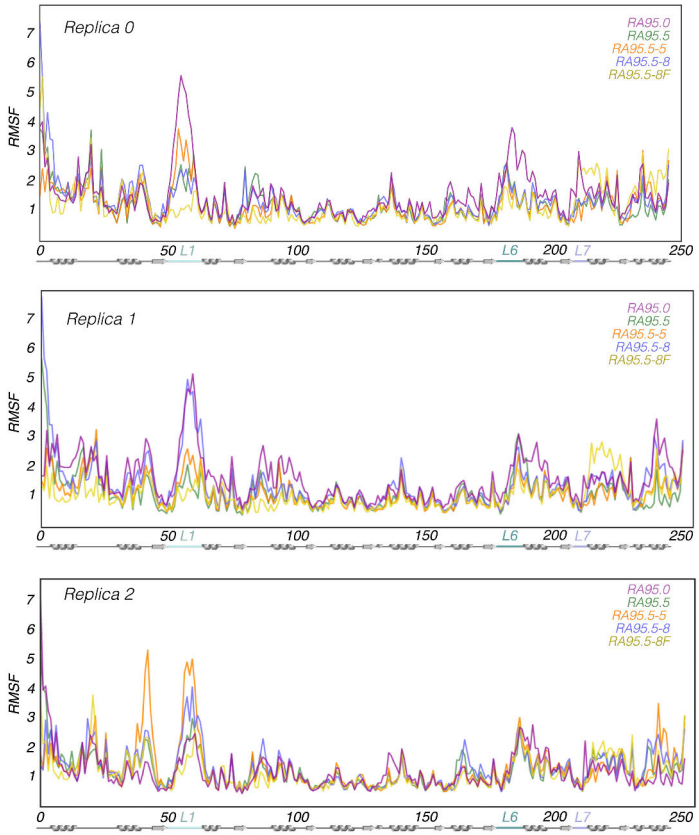


Figure B.4 Root Mean Square Fluctuation (RMSF, in Å) for all RA Schiff base intermediate variants along the microsecond timescale MD simulations for all replicas.

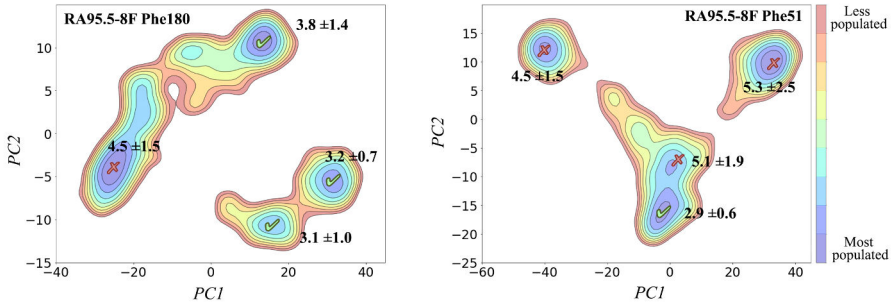


Figure B.5 Projection of the MD trajectories in the Schiff base intermediate into the two most important principal components (PC1, PC2) based on Cα contacts. Those states exploring distances in the 2.0−4.0 Å range are shown in green (catalytically competent), whilst the other states are shown in red.

Supplementary Information Chapter 7

Appendix C

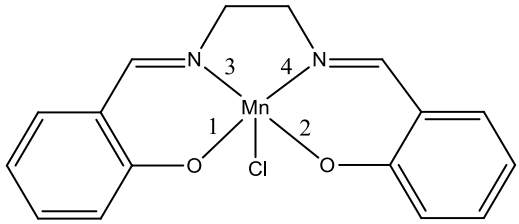


Figure C.1 Salen manganese complex, EUK-8.

Table C.1 Relative spin state energies of EUK-8, HOOH···EUK-8 and binding energy of HOOH towards EUK-8 (singlet, triplet, and quintet; kcal/mol) at (COSMO)S12g/TZ2P.

	Singlet	**Triplet**	**Quintet**
EUK-8	44.1	20.1	0.0
EUK-8 (PBE-D)	23.4	13.0	0.0
EUK-8···HOOH	30.6	17.4	0.0
$E_{binding,HOOH}$	-21.1	-10.2	-7.5

Table C.2 Bond distances (Å) in optimization geometries using both functionals and in the X-ray structure of EUK-8 for ligands linked on the metal atom (Mn). In parenthesis ( ) the difference of bond distances between the X-ray structure and the optimized geometry.

		PBE-D			S12g		
EUK-8	$R_{X\text{-}ray}$	$R_{Singlet}$	$R_{Triplet}$	$R_{Quintet}$	$R_{Singlet}$	$R_{Triplet}$	$R_{Quintet}$
Mn-O1	1.905	1.821 (-0.085)	1.863 (-0.043)	1.897 (-0.009)	1.814 (-0.092)	1.868 (-0.037)	1.901 (-0.004)
Mn-O2	1.879	1.876 (-0.002)	1.870 (-0.009)	1.888 (0.009)	1.880 (0.001)	1.871 (-0.008)	1.895 (0.016)
Mn-N3	1.993	1.968 (-0.024)	1.934 (-0.059)	1.975 (-0.017)	1.958 (-0.035)	1.926 (-0.066)	1.977 (-0.016)
Mn-N4	1.981	1.853 (-0.129)	1.932 (-0.050)	1.987 (0.005)	1.836 (-0.146)	1.925 (-0.056)	1.988 (0.007)
Mn-Cl	2.461	2.217 (-0.244)	2.236 (-0.224)	2.444 (-0.016)	2.228 (-0.232)	2.246 (-0.215)	2.472 (0.012)

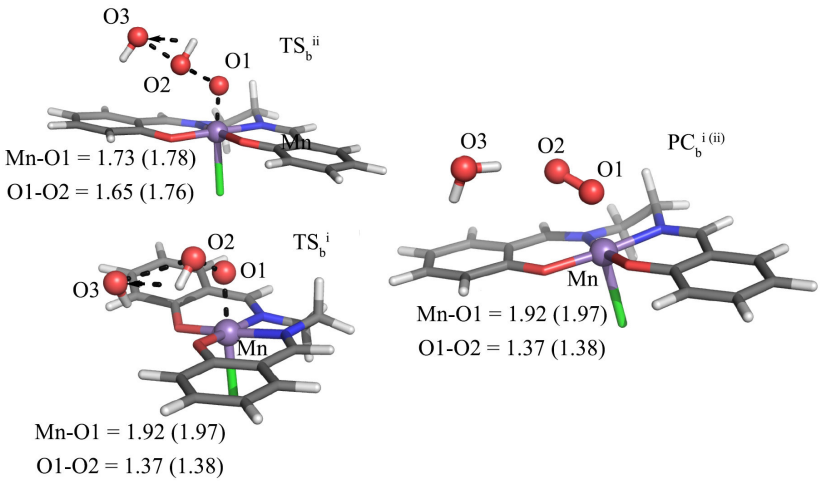


Figure C.2 Optimized structures of the second dismutation reaction of the intramolecular pathway through the RC^{ii} and RC^{i}. Transition state (TS_b for i and ii) and product complex (PC_b) for the intramolecular pathway. Bond length in Å, distances for quintet in parenthesis ( ) and distances for triplet in blank**.**

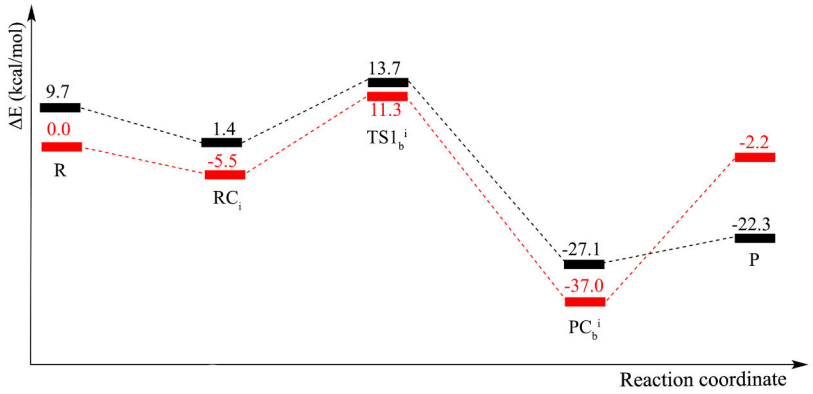


Figure C.3 Energy profile (kcal/mol) of the second dismutation reaction following the intramolecular proton transfers from the reaction complex i). Stationary points represented are: R) reactants at infinite distance (oxo-complex and hydrogen peroxide); RC_i) reaction complex; TS_b^i) transition state; PC_b^i) product complex; P) product at infinite distance (complex, water and oxygen molecule). Also, the different spin-states for Mn^{III}: triplet state (red line) and quintet state (black line).

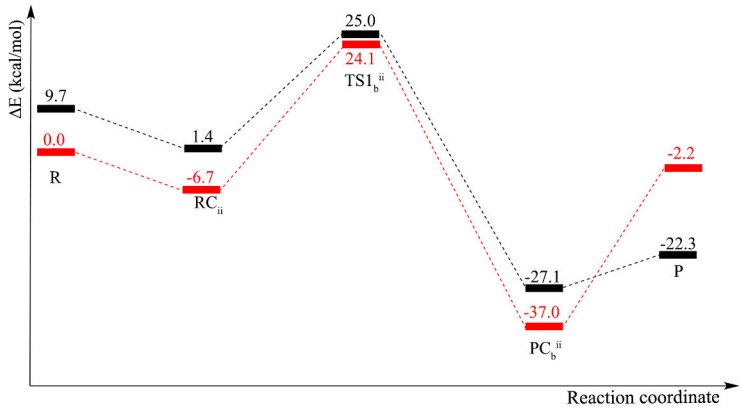


Figure C.4 Energy profile (kcal/mol) of the second dismutation reaction following the intramolecular proton transfers from the reaction complex ii). Stationary points represented are: R) reactants at infinite distance (oxo-complex and hydrogen peroxide); RC_{ii}) reaction complex; TS_b^{ii}) transition state; PC_b^{ii}) product complex; P) product at infinite distance (complex, water and oxygen molecule). Also, the different spin-states for Mn^{III}: triplet state (red line) and quintet state (black line).

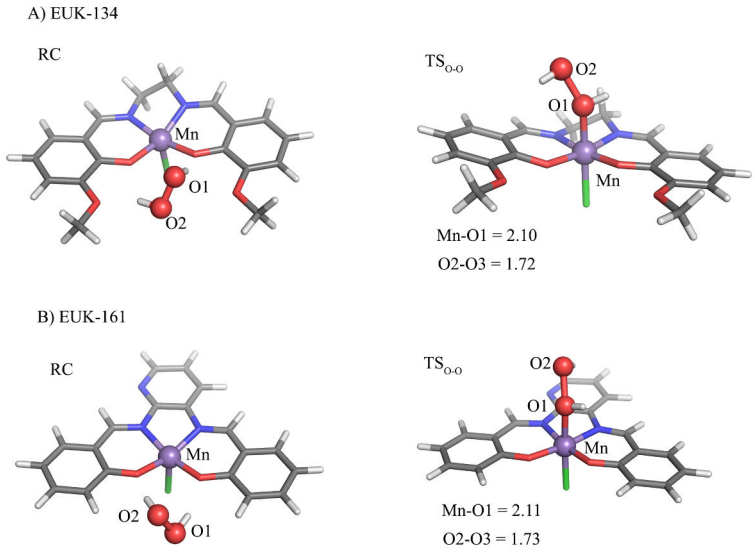


Figure C.5 Optimized structures of the rate-determining state of A) EUK-134 and B) EUK-161 without water molecules. All distances are in Å.

Supplementary Information Chapter 8

Appendix D

Table D.1 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2+, 2.5+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using S12g/TZ2P) electron density of $Fe^{2+, 2.5+}$ nuclei, the reference value of Fe^{II} metal atom, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox-State	Experiment		S12g/T2ZP			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q(cal)}$	η_{cal}
$Fe(II)_{ref.}$	2	+2	-	-	11835.135	-	-	-
FeF_2 $(FeF_6)^{4-}$	2	+2	1.48	2.85	11836.957	1.48	3.22	0.0004
$FeCl_4^{2-}$	2	+2	1.05	3.27	11837.830	1.10	3.51	0.0473
$FeBr_4^{2-}$	2	+2	1.12	3.23	11837.848	1.09	3.52	0.2265
$Fe(NCS)_4^{2-}$	2	+2	0.97	2.83	11837.952	1.04	3.32	0.0077
$Fe(salmp)_2^{2-}$	0	+2	1.11	2.24	11837.893	1.07	1.66	0.2015
		+2	1.11	2.24	11837.935	1.05	1.68	0.2078
$Fe(H_2O)_6$	2	+2	1.39	3.38	11837.040	1.44	3.49	0.0174
$Fe(phen)_2Cl_2$	2	+2	1.05	3.15	11837.947	1.05	2.85	0.2672
$Fe(opda)_2Cl_2$	2	+2	0.91	3.17	11837.719	1.15	3.62	0.2003
$Fe(Py)_4Cl_2$	2	+2	1.16	3.14	11837.675	1.17	3.20	0.0073
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_4(C_5H_5N)_2$	0	+2	1.12	3.05	11837.787	1.12	2.89	0.4060
		+2	1.12	3.05	11837.773	1.12	2.94	0.3746

$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(THF)_2$	4	+2	1.26	2.9	11837.573	1.21	2.83	0.0813
		+2	1.26	2.9	11837.548	1.22	2.71	0.2950
	0	+2	1.26	2.9	11837.571	1.21	3.22	0.0565
		+2	1.26	2.9	11837.529	1.23	3.07	0.2464
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(NH_2CH_2CH_3)_2$	4	+2	1.19	2.9	11837.609	1.19	2.64	0.5288
		+2	1.19	2.9	11837.617	1.19	2.66	0.5284
	0	+2	1.19	2.9	11837.617	1.19	3.26	0.2668
		+2	1.19	2.9	11837.617	1.19	3.27	0.2656
$Fe_2(\mu\text{-}OH_2)_2(\mu\text{-}O_2C\text{-}CH_3)_2\text{-}(O_2C\text{-}CH_3)_3(THF)_2(OH_2)$	4	+2	1.35	3.26	11837.178	1.38	3.18	0.1979
		+2	1.35	3.26	11837.247	1.35	3.23	0.4279
	0	+2	1.35	3.26	11837.191	1.38	3.20	0.1531
		+2	1.35	3.26	11837.228	1.36	3.28	0.4215
$Fe_2BPMP(OPr)_2^+$	0	+2	1.24	2.72	11837.752	1.13	2.81	0.3886
		+2	1.24	2.72	11837.760	1.13	2.82	0.3760
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+2	1.15	2.69	11837.924	1.06	3.11	0.1484
$Fe_2(OH)(Oac)_2(Me_3TACN)_2^+$	0	+2	1.16	2.83	11837.582	1.21	3.00	0.0335
		+2	1.16	2.83	11837.585	1.20	2.97	0.0560
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11838.147	0.96	0.80	0.9171

		+2.5	0.83	1.08	11838.146	0.96	0.81	0.8937

Table D.2 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2+,\,2.5+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/TZ2P) electron density of $Fe^{2+,\,2.5+}$ nuclei, the reference value of Fe^{II} metal atom, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox-State	Experiment		OPBE/T2ZP			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q(cal)}$	η_{cal}
$Fe(II)_{ref.}$	2	+2	-	-	11816.207	-	-	-
FeF_2 $(FeF_6)^{4-}$	2	+2	1.48	2.85	11818.114	1.44	3.24	0.0004
$FeCl_4^{2-}$	2	+2	1.05	3.27	11819.004	1.05	3.70	0.1216
$FeBr_4^{2-}$	2	+2	1.12	3.23	11818.973	1.07	3.66	0.0771
$Fe(NCS)_4^{2-}$	2	+2	0.97	2.83	11819.124	1.00	3.38	0.0102
$Fe(salmp)_2^{2-}$	0	+2	1.11	2.24	11819.093	1.01	1.68	0.1745
		+2	1.11	2.24	11819.137	0.99	1.70	0.1755
$Fe(H_2O)_6$	2	+2	1.39	3.38	11818.179	1.41	3.53	0.0208
$Fe(phen)_2Cl_2$	2	+2	1.05	3.15	11819.129	1.00	2.91	0.3283
$Fe(opda)_2Cl_2$	2	+2	0.91	3.17	11818.894	1.10	3.63	0.1888
$Fe(Py)_4Cl_2$	2	+2	1.16	3.14	11818.823	1.13	3.22	0.0067
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_4(C_5H_5N)_2$	0	+2	1.12	3.05	11818.951	1.08	2.94	0.3961
		+2	1.12	3.05	11818.937	1.08	2.98	0.3600
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(THF)_2$	4	+2	1.26	2.9	11818.704	1.18	3.06	0.1175
		+2	1.26	2.9	11818.695	1.19	2.96	0.2394

	0	+2	1.26	2.9	11818.714	1.18	3.25	0.0834
		+2	1.26	2.9	11818.676	1.20	3.14	0.2302
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(NH_2CH_2CH_3)_2$	4	+2	1.19	2.9	11818.749	1.16	3.08	0.4406
		+2	1.19	2.9	11818.760	1.16	3.10	0.4204
	0	+2	1.19	2.9	11818.759	1.16	3.35	0.1633
		+2	1.19	2.9	11818.765	1.16	3.37	0.1643
$Fe_2(\mu\text{-}OH_2)_2(\mu\text{-}O_2C\text{-}CH_3)_2\text{-}(O_2C\text{-}CH_3)_3(THF)_2(OH_2)$	4	+2	1.35	3.26	11818.316	1.35	3.25	0.1755
		+2	1.35	3.26	11818.373	1.33	3.30	0.4323
	0	+2	1.35	3.26	11818.323	1.35	3.25	0.1531
		+2	1.35	3.26	11818.361	1.33	3.33	0.4202
$Fe_2BPMP(OPr)_2^+$	0	+2	1.24	2.72	11818.923	1.09	2.86	0.3901
		+2	1.24	2.72	11818.931	1.08	2.86	0.3675
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+2	1.15	2.69	11819.083	1.02	3.19	0.1544
$Fe_2(OH)(Oac)_2(Me_3TACN)_2^+$	0	+2	1.16	2.83	11818.750	1.16	3.02	0.0988
		+2	1.16	2.83	11818.754	1.16	3.00	0.0560
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11819.365	0.90	0.82	0.9621
		+2.5	0.83	1.08	11819.363	0.90	0.83	0.9792

Table D.3 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2+,\,2.5+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/TZP) electron density of $Fe^{2+,\,2.5+}$ nuclei, the reference value of Fe^{II} metal atom, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox-State	Experiment		OPBE/TZP			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q(cal)}$	η_{cal}
$Fe(II)_{ref.}$	2	+2	-	-	11816.202	-	-	-
FeF_2 $(FeF_6)^{4-}$	2	+2	1.48	2.85	11818.092	1.45	3.28	0.0005
$FeCl_4^{2-}$	2	+2	1.05	3.27	11818.959	1.07	3.72	0.1203
$FeBr_4^{2-}$	2	+2	1.12	3.23	11818.926	1.08	3.69	0.0771
$Fe(NCS)_4^{2-}$	2	+2	0.97	2.83	11819.113	1.00	3.39	0.0105
$Fe(salmp)_2^{2-}$	0	+2	1.11	2.24	11819.057	1.03	1.70	0.1674
		+2	1.11	2.24	11819.099	1.01	1.72	0.1682
$Fe(H_2O)_6$	2	+2	1.39	3.38	11818.109	1.44	3.57	0.0173
$Fe(phen)_2Cl_2$	2	+2	1.05	3.15	11819.085	1.02	2.95	0.3620
$Fe(opda)_2Cl_2$	2	+2	0.91	3.17	11818.859	1.11	3.68	0.1823
$Fe(Py)_4Cl_2$	2	+2	1.16	3.14	11818.789	1.14	3.28	0.0068
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_4(C_5H_5N)_2$	0	+2	1.12	3.05	11818.913	1.09	2.98	0.3978
		+2	1.12	3.05	11818.899	1.10	3.02	0.3617
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(THF)_2$	4	+2	1.26	2.9	11818.679	1.19	3.09	0.0939
		+2	1.26	2.9	11818.669	1.20	2.96	0.2325
	0	+2	1.26	2.9	11818.690	1.19	3.30	0.1409
		+2	1.26	2.9	11818.650	1.21	3.16	0.2186

$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(NH_2CH_2CH_3)_2$	4	+2	1.19	2.9	11818.719	1.18	3.10	0.4295
		+2	1.19	2.9	11818.729	1.17	3.11	0.4108
	0	+2	1.19	2.9	11818.728	1.17	3.38	0.1472
		+2	1.19	2.9	11818.735	1.17	3.40	0.1485
$Fe_2(\mu\text{-}OH_2)_2(\mu\text{-}O_2C\text{-}CH_3)_2\text{-}(O_2C\text{-}CH_3)_3(THF)_2(OH_2)$	4	+2	1.35	3.26	11818.272	1.37	3.28	0.1502
		+2	1.35	3.26	11818.336	1.34	3.33	0.4111
	0	+2	1.35	3.26	11818.280	1.37	3.28	0.1252
		+2	1.35	3.26	11818.322	1.35	3.36	0.3997
$Fe_2BPMP(OPr)_2^+$	0	+2	1.24	2.72	11818.888	1.10	2.89	0.3857
		+2	1.24	2.72	11818.896	1.10	2.90	0.3635
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+2	1.15	2.69	11819.048	1.03	3.22	0.1424
$Fe_2(OH)(Oac)_2(Me_3TACN)_2^+$	0	+2	1.16	2.83	11818.710	1.18	3.06	0.1066
		+2	1.16	2.83	11818.714	1.18	3.04	0.0632
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11819.325	0.91	0.83	0.9658
		+2.5	0.83	1.08	11819.323	0.91	0.83	0.9836

Table D.4 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2+, 2.5+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/DZ) electron density of $Fe^{2+, 2.5+}$ nuclei, the reference value of Fe^{II} metal atom, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox-State	Experiment		OPBE/DZ			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q(cal)}$	η_{cal}
$Fe(II)_{ref.}$	2	+2	-	-	11815.939	-	-	-
FeF_2 $(FeF_6)^{4-}$	2	+2	1.48	2.85	11817.884	1.43	3.40	0.0005
$FeCl_4^{2-}$	2	+2	1.05	3.27	11818.566	1.13	3.64	0.0308
$FeBr_4^{2-}$	2	+2	1.12	3.23	11818.592	1.12	3.80	0.0802
$Fe(NCS)_4^{2-}$	2	+2	0.97	2.83	11818.654	1.09	3.48	0.0095
$Fe(salmp)_2^{2-}$	0	+2	1.11	2.24	11818.626	1.10	1.71	0.1286
		+2	1.11	2.24	11818.665	1.08	1.72	0.1080
$Fe(H_2O)_6$	2	+2	1.39	3.38	11817.657	1.52	3.68	0.0158
$Fe(phen)_2Cl_2$	2	+2	1.05	3.15	11818.695	1.07	2.79	0.2910
$Fe(opda)_2Cl_2$	2	+2	0.91	3.17	11818.457	1.17	3.81	0.1389
$Fe(Py)_4Cl_2$	2	+2	1.16	3.14	11818.415	1.19	3.26	0.0065
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_4(C_5H_5N)_2$	0	+2	1.12	3.05	11818.523	1.15	3.06	0.3762
		+2	1.12	3.05	11818.510	1.15	3.11	0.3269
$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(THF)_2$	4	+2	1.26	2.9	11818.286	1.25	3.13	0.1787
		+2	1.26	2.9	11818.288	1.25	3.07	0.3036
	0	+2	1.26	2.9	11818.296	1.25	3.37	0.1036
		+2	1.26	2.9	11818.264	1.26	3.31	0.3054

$Fe_2(\mu\text{-}O_2C\text{-}CH_3)_2(O_2C\text{-}CH_3)_2\text{-}(NH_2CH_2CH_3)_2$	4	+2	1.19	2.9	11818.351	1.22	3.05	0.5684
		+2	1.19	2.9	11818.362	1.22	3.06	0.5487
	0	+2	1.19	2.9	11818.363	1.22	3.48	0.1569
		+2	1.19	2.9	11818.369	1.21	3.50	0.1626
$Fe_2(\mu\text{-}OH_2)_2(\mu\text{-}O_2C\text{-}CH_3)_2\text{-}(O_2C\text{-}CH_3)_3(THF)_2(OH_2)$	4	+2	1.35	3.26	11817.874	1.43	3.45	0.2893
		+2	1.35	3.26	11817.946	1.40	3.42	0.4834
	0	+2	1.35	3.26	11817.878	1.43	3.45	0.2743
		+2	1.35	3.26	11817.934	1.40	3.44	0.4648
$Fe_2BPMP(OPr)_2^+$	0	+2	1.24	2.72	11818.484	1.16	2.89	0.4041
		+2	1.24	2.72	11818.492	1.16	2.89	0.3828
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+2	1.15	2.69	11818.634	1.10	3.25	0.2216
$Fe_2(OH)(Oac)_2(Me_3TACN)_2^+$	0	+2	1.16	2.83	11818.278	1.25	3.09	0.0881
		+2	1.16	2.83	11818.282	1.25	3.06	0.0457
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11818.870	0.99	0.83	0.9989
		+2.5	0.83	1.08	11818.868	1.00	0.83	0.9849

Table D.5 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using S12g/TZ2P) electron density of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ nuclei, the reference value of $Fe^{III,\ IV}$ metal atoms, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox. state	Experiment		S12g/TZ2P			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q\ cal}$	η_{cal}
$Fe(III)_{ref.}$	5/2	+3	-	-	11836.985	-	-	-
$Fe(IV)_{ref.}$	2	+4	-	-	11840.319	-	-	-
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11838.147	0.70	0.80	0.9171
		+2.5	0.83	1.08	11838.146	0.70	0.81	0.8937
$Fe(bipy)_2Cl_2^+$	5/2	+3	0.54	0.24	11838.864	0.45	0.33	0.7089
FeF_6^{3-}	5/2	+3	0.61	0.00	11838.461	0.59	0.02	0.7492
$FeCl_6^{3-}$	5/2	+3	0.56	0.04	11838.534	0.57	0.05	0.6891
$FeCl_4^-$	5/2	+3	0.36	0.00	11839.134	0.36	0.03	0.8370
$Cl_3FeOFeCl_3^{2-}$	0	+3	0.36	1.24	11839.100	0.37	1.05	0.0259
		+3	0.36	1.24	11839.100	0.37	1.05	0.0376
$Fe_2O(OAc)_2(HBpz_3)_2$	0	+3	0.55	1.60	11838.734	0.50	1.58	0.6181
		+3	0.55	1.60	11838.733	0.50	1.56	0.6169
$Fe_2(OH)(OAc)_2(HBzp_3)_2^+$	0	+3	0.5	0.25	11838.752	0.49	0.34	0.3845
		+3	0.5	0.25	11838.757	0.49	0.38	0.3851
$Fe_2O(OAc)_2(Me_3TACN)_2^{2+}$	0	+3	0.47	1.50	11838.801	0.47	1.48	0.5734
		+3	0.47	1.50	11838.797	0.48	1.49	0.5680
$Fe_2O(OAc)_2(bipy)_2Cl_2$	0	+3	0.41	1.80	11838.751	0.49	1.51	0.5221

	0	+3	0.41	1.80	11838.756	0.49	1.47	0.4472
$Fe_2(salmp)_2$		+3	0.56	0.88	11838.551	0.56	1.10	0.5855
		+3	0.56	0.88	11838.552	0.56	1.10	0.5849
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+3	0.50	0.50	11838.536	0.57	0.74	0.2468
$Fe_2O_2(6TLA)_2^{2+}$	0	+3	0.50	1.93	11838.879	0.45	1.79	0.8874
		+3	0.50	1.93	11838.878	0.45	1.79	0.8833
$Fe_2O(Me_3TACN)_2(Cl_4cat)_2$	0	+3	0.46	1.41	11838.820	0.47	1.39	0.9973
		+3	0.46	1.41	11838.820	0.47	1.39	0.9796
$Fe_2(Cat)_4(H_2O)_2^{2-}$	0	+3	0.56	0.90	11838.401	0.61	1.33	0.8395
		+3	0.56	0.90	11838.403	0.61	1.32	0.8403
$Fe_2O_2(5\text{-}Et_3\text{-}TPA)_2^{3+}$	3/2	+3.5	0.14	0.49	11839.941	0.08	0.50	0.2784
		+3.5	0.14	0.49	11839.934	0.08	0.50	0.2375
Fe(OEC)Cl	1	+4	0.22	2.99	11839.684	0.17	2.94	0.0308
$Fe(OEC)C_6H_5$	1	+4	-0.08	3.72	11840.491	-0.12	3.28	0.0593
$FeCl(\eta^4\text{-}MAC^*)^-$	2	+4	-0.04	-0.89	11840.018	0.05	-0.85	0.6273

Table D.6 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/TZ2P) electron density of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ nuclei, the reference value of $Fe^{III,\ IV}$ metal atoms, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox. state	Experiment		OPBE/TZ2P			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q\ cal}$	η_{cal}
$Fe(III)_{ref.}$	5/2	+3	-	-	11818.171	-	-	-
$Fe(IV)_{ref.}$	2	+4	-	-	11821.530	-	-	-
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11819.365	0.70	0.82	0.9621
		+2.5	0.83	1.08	11819.363	0.70	0.83	0.9792
$Fe(bipy)_2Cl_2^+$	5/2	+3	0.54	0.24	11820.100	0.44	0.34	0.6836
FeF_6^{3-}	5/2	+3	0.61	0.00	11819.685	0.59	0.02	0.9388
$FeCl_6^{3-}$	5/2	+3	0.56	0.04	11819.735	0.57	0.05	0.7054
$FeCl_4^-$	5/2	+3	0.36	0.00	11820.354	0.35	0.03	0.2319
$Cl_3FeOFeCl_3^{2-}$	0	+3	0.36	1.24	11820.317	0.37	1.06	0.0223
		+3	0.36	1.24	11820.315	0.37	1.06	0.0396
$Fe_2O(OAc)_2(HBpz_3)_2$	0	+3	0.55	1.60	11819.967	0.49	1.57	0.5504
		+3	0.55	1.60	11819.967	0.49	1.55	0.5530
$Fe_2(OH)(OAc)_2(HBzp_3)_2^+$	0	+3	0.5	0.25	11819.993	0.48	0.35	0.3119
		+3	0.5	0.25	11819.997	0.48	0.39	0.3258
$Fe_2O(OAc)_2(Me_3TACN)_2^{2+}$	0	+3	0.47	1.50	11820.029	0.47	1.48	0.5084
		+3	0.47	1.50	11820.025	0.47	1.50	0.5038
$Fe_2O(OAc)_2(bipy)_2Cl_2$	0	+3	0.41	1.80	11819.982	0.48	1.49	0.4940

	0	+3	0.41	1.80	11819.991	0.48	1.46	0.4168
$Fe_2(salmp)_2$		+3	0.56	0.88	11819.795	0.55	1.12	0.5526
		+3	0.56	0.88	11819.796	0.55	1.12	0.5540
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+3	0.50	0.50	11819.776	0.56	0.69	0.3610
$Fe_2O_2(6TLA)_2^{2+}$	0	+3	0.50	1.93	11820.097	0.44	1.81	0.9155
		+3	0.50	1.93	11820.097	0.44	1.80	0.9120
$Fe_2O(Me_3TACN)_2(Cl_4cat)_2$	0	+3	0.46	1.41	11820.048	0.46	1.40	0.8954
		+3	0.46	1.41	11820.048	0.46	1.40	0.9150
$Fe_2(Cat)_4(H_2O)_2^{2-}$	0	+3	0.56	0.90	11819.622	0.61	1.36	0.8281
		+3	0.56	0.90	11819.623	0.61	1.36	0.8289
$Fe_2O_2(5\text{-}Et_3\text{-}TPA)_2^{3+}$	3/2	+3.5	0.14	0.49	11821.205	0.06	0.46	0.3212
		+3.5	0.14	0.49	11821.199	0.06	0.46	0.2798
Fe(OEC)Cl	1	+4	0.22	2.99	11820.909	0.16	2.93	0.0242
$Fe(OEC)C_6H_5$	1	+4	-0.08	3.72	11821.717	-0.12	3.28	0.0626
$FeCl(\eta^4\text{-}MAC^*)^-$	2	+4	-0.04	-0.89	11821.254	0.04	0.89	0.4997

Table D.7 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/TZP) electron density of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ nuclei, the reference value of $Fe^{III,\ IV}$ metal atoms, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox. state	Experiment		OPBE/TZP			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q\ cal}$	η_{cal}
$Fe(III)_{ref.}$	5/2	+3	-	-	11818.168	-	-	-
$Fe(IV)_{ref.}$	2	+4	-	-	11821.516	-	-	-
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11819.325	0.71	0.83	0.9658
		+2.5	0.83	1.08	11819.323	0.71	0.83	0.9836
$Fe(bipy)_2Cl_2^+$	5/2	+3	0.54	0.24	11820.042	0.46	0.33	0.6736
FeF_6^{3-}	5/2	+3	0.61	0.00	11819.648	0.60	0.02	0.9371
$FeCl_6^{3-}$	5/2	+3	0.56	0.04	11819.671	0.59	0.04	0.7039
$FeCl_4^-$	5/2	+3	0.36	0.00	11820.289	0.37	0.03	0.2231
$Cl_3FeOFeCl_3^{2-}$	0	+3	0.36	1.24	11820.258	0.38	1.05	0.0228
		+3	0.36	1.24	11820.255	0.38	1.05	0.0400
$Fe_2O(OAc)_2(HBpz_3)_2$	0	+3	0.55	1.60	11819.926	0.50	1.57	0.5443
		+3	0.55	1.60	11819.926	0.50	1.55	0.5468
$Fe_2(OH)(OAc)_2(HBzp_3)_2^+$	0	+3	0.5	0.25	11819.946	0.49	0.35	0.2898
		+3	0.5	0.25	11819.950	0.49	0.39	0.3032
$Fe_2O(OAc)_2(Me_3TACN)_2^{2+}$	0	+3	0.47	1.50	11819.980	0.48	1.49	0.5116
		+3	0.47	1.50	11819.977	0.48	1.50	0.5067
$Fe_2O(OAc)_2(bipy)_2Cl_2$	0	+3	0.41	1.80	11819.934	0.50	1.49	0.4917

	0	+3	0.41	1.80	11819.941	0.49	1.46	0.4152
$Fe_2(salmp)_2$		+3	0.56	0.88	11819.750	0.56	1.13	0.5524
		+3	0.56	0.88	11819.751	0.56	1.13	0.5537
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+3	0.50	0.50	11819.727	0.57	0.69	0.3753
$Fe_2O_2(6TLA)_2^{2+}$	0	+3	0.50	1.93	11820.059	0.45	1.80	0.9322
		+3	0.50	1.93	11820.059	0.45	1.80	0.9285
$Fe_2O(Me_3TACN)_2(Cl_4cat)_2$	0	+3	0.46	1.41	11820.002	0.47	1.40	0.8883
		+3	0.46	1.41	11820.002	0.47	1.41	0.9078
$Fe_2(Cat)_4(H_2O)_2^{2-}$	0	+3	0.56	0.90	11819.567	0.62	1.37	0.8290
		+3	0.56	0.90	11819.569	0.62	1.37	0.8294
$Fe_2O_2(5\text{-}Et_3\text{-}TPA)_2^{3+}$	3/2	+3.5	0.14	0.49	11821.145	0.07	0.43	0.3328
		+3.5	0.14	0.49	11821.138	0.08	0.43	0.2894
Fe(OEC)Cl	1	+4	0.22	2.99	11820.859	0.17	2.90	0.0257
$Fe(OEC)C_6H_5$	1	+4	-0.08	3.72	11821.675	-0.11	3.24	0.0682
$FeCl(\eta^4\text{-}MAC^*)^-$	2	+4	-0.04	-0.89	11821.200	0.05	0.89	0.5034

Table D.8 Calculated and the experimental isomer shifts and quadrupole splitting (mm/s) of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ training set complexes were used in the fitting linear parameters for getting the linear correlation between calculated (using OPBE/DZ) electron density of $Fe^{2.5+,\ 3+,\ 3.5+,\ 4+}$ nuclei, the reference value of $Fe^{III,\ IV}$ metal atoms, and the corresponding experimental isomer shifts values at 4.2 K.

Iron Complex	S_{total}	Ox. state	Experiment		OPBE/DZ			
			$\delta_{4.2K}$	ΔE_Q	$\rho(0)$	δ_{cal}	$\Delta E_{Q\ cal}$	η_{cal}
$Fe(III)_{ref.}$	5/2	+3	-	-	11817.885	-	-	-
$Fe(IV)_{ref.}$	2	+4	-	-	11821.174	-	-	-
$Fe_2(salmp)_2^-$	9/2	+2.5	0.83	1.08	11818.870	0.75	0.83	0.9989
		+2.5	0.83	1.08	11818.868	0.75	0.83	0.9849
$Fe(bipy)_2Cl_2^+$	5/2	+3	0.54	0.24	11819.570	0.50	0.32	0.6971
FeF_6^{3-}	5/2	+3	0.61	0.00	11819.316	0.59	0.02	0.8945
$FeCl_6^{3-}$	5/2	+3	0.56	0.04	11819.287	0.60	0.04	0.7148
$FeCl_4^-$	5/2	+3	0.36	0.00	11819.838	0.41	0.03	0.1835
$Cl_3FeOFeCl_3^{2-}$	0	+3	0.36	1.24	11819.813	0.42	1.06	0.0179
		+3	0.36	1.24	11819.811	0.42	1.06	0.0330
$Fe_2O(OAc)_2(HBpz_3)_2$	0	+3	0.55	1.60	11819.458	0.54	1.53	0.5363
		+3	0.55	1.60	11819.456	0.54	1.52	0.5374
$Fe_2(OH)(OAc)_2(HBzp_3)_2^+$	0	+3	0.5	0.25	11819.501	0.53	0.40	0.2769
		+3	0.5	0.25	11819.507	0.53	0.43	0.2948
$Fe_2O(OAc)_2(Me_3TACN)_2^{2+}$	0	+3	0.47	1.50	11819.541	0.51	1.47	0.5071
		+3	0.47	1.50	11819.537	0.52	1.49	0.5016
$Fe_2O(OAc)_2(bipy)_2Cl_2$	0	+3	0.41	1.80	11819.478	0.54	1.42	0.4990

	0	+3	0.41	1.80	11819.485	0.53	1.39	0.4132
$Fe_2(salmp)_2$		+3	0.56	0.88	11819.272	0.61	1.08	0.5727
		+3	0.56	0.88	11819.274	0.61	1.07	0.5740
$Fe(II)Fe(III)BPMP(OPr)_2^{2+}$	1/2	+3	0.50	0.50	11819.299	0.60	0.71	0.1661
$Fe_2O_2(6TLA)_2^{2+}$	0	+3	0.50	1.93	11819.580	0.50	1.73	0.9975
		+3	0.50	1.93	11819.579	0.50	1.73	0.9939
$Fe_2O(Me_3TACN)_2(Cl_4cat)_2$	0	+3	0.46	1.41	11819.537	0.52	1.41	0.8652
		+3	0.46	1.41	11819.537	0.52	1.41	0.8799
$Fe_2(Cat)_4(H_2O)_2^{2-}$	0	+3	0.56	0.90	11819.146	0.65	1.34	0.8162
		+3	0.56	0.90	11819.148	0.65	1.34	0.8167
$Fe_2O_2(5\text{-}Et_3\text{-}TPA)_2^{3+}$	3/2	+3.5	0.14	0.49	11820.633	0.13	0.40	0.5399
		+3.5	0.14	0.49	11820.627	0.14	0.40	0.4893
Fe(OEC)Cl	1	+4	0.22	2.99	11820.447	0.20	2.94	0.0249
$Fe(OEC)C_6H_5$	1	+4	-0.08	3.72	11821.183	-0.06	3.31	0.0603
$FeCl(\eta^4\text{-}MAC^*)^-$	2	+4	-0.04	-0.89	11820.775	0.08	0.84	0.7548

Table D.9 Isomer shift calculation using the single training set of iron complexes using OPBE/TZP and S12g/TZ2P level of theory. Values in blank using as reference Fe^{II}, values in ( ) using as reference Fe^{III}, and values in [ ] using as reference Fe^{IV}.

Complex	Ox. State	S_{total}	δ_{exp}	δ_{OPBE}	δ_{S12g}
$[FeCN_6]^{-4}$	2+	S=0	-0.02	-0.36 (-0.34) [-0.33]	-0.26 (-0.30) [-0.30]
$[FePor(OAc)]^{-}$	2+	S=2	1.05	0.86 (0.88) [0.89]	0.90 (0.86) [0.89]
$[Fe(SR)_3]^{-}$	2+	S=2	0.56	0.43 (0.44) [0.45]	0.44 (0.40) [0.46]
$[FeAz]^{+}$	3+	S=1/2	0.29	0.19 (0.21) [0.22]	0.26 (0.22) [0.22]
$[FeCN_6]^{3-}$	3+	S=1/2	-0.13	-0.40 (-0.39) [-0.38]	-0.31 (-0.35) [-0.35]
$[FePor(O_2)]^{-}$	3+	S=5/2	0.67	0.79 (0.81) [0.81]	0.84 (0.81) [0.89]
$[FeMAC]^{-}$	4+	S=2	-0.02	-0.07 (-0.05) [-0.04]	-0.01 (-0.05) [-0.05]
$[FePorO]^{+}$	4+	S=3/2	0.08	-0.20 (-0.18) [-0.18]	-0.03 (-0.07) [-0.07]
$[FeTMCO]^{2+}$	4+	S=1	0.17	-0.03 (-0.01) [-0.004]	0.03 (-0.01) [-0.007]