ENZYMATIC AND BIOINSPIRED IRON OXIDATION CHEMISTRY: A COMPUTATIONAL STUDY

Lorenzo D'Amore

Per citar o enllaçar aquest document:
Para citar o enlazar este documento:
Use this url to cite or link to this publication:
http://hdl.handle.net/10803/674109

http://creativecommons.org/licenses/by-nc-sa/4.0/deed.ca

Aquesta obra està subjecta a una llicència Creative Commons Reconeixement-NoComercial-CompartirIgual

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike licence

DOCTORAL THESIS

Enzymatic and bioinspired iron oxidation chemistry: a computational study

Lorenzo D'Amore

2021

Doctoral program in Chemistry

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

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

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

WE DECLARE:

That the thesis entitled "Enzymatic and bioinspired iron oxidation chemistry: a computational study", presented by Lorenzo D'Amore to obtain the doctoral degree, has been completed under our supervision and meets the requirements to opt for International Doctorate.

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

Prof. Marcel Swart Prof. Sílvia Osuna

Girona, October 2021

Ad Aran,
a mio figlio.

Acknowledgements

My deepest gratitude to my partner Maite and my mother, for the continuous, unconditioned support. You have always encouraged me during my PhD journey, and this effort would not have been possible without your backing.

I first and foremost thank my supervisors, Prof. Marcel Swart and Prof. Sílvia Osuna. Your outstanding knowledge let me learn enourmously along my path, and your enduring fascination about science was always truly inspiring. I extend my recognition also to Prof. Manfred Reetz and Dr. Carlos Achevedo-Rocha for the great scientific discussions on P450; to Prof. Adrian Mulholland for hosting me as visiting PhD student at the University of Bristol and the great science going on there; to Prof. Paola Belanzoni and Dr. Leonardo Belpassi for our useful conversations about the projects we have been involved.

A warm thank you to the IQCC postdocs Ferran Feixas, Javier Iglesias, Marc Garcia-Borràs; PhD students Carla Calvò, Christian Curado, Eila Serrano, Fred Martins, Guillem Casadevall, Miguel Ángel María Solano; former members Abril Castro, Adrian Romero, Davide Angelone, Ouissam El Bakouri, Steven Roldán, Verònica Postils, QBIS PhDs Andrea Palone, Marco Cianfanelli and IT Dani Maso. Thanks all of you for the knowledge exchange and the nice afterwork time we have been through.

I would like also to acknowledge all the scientist and people that I had the pleasure to meet over these years: Giordano Amoruso, Christian "il King", Silvia Gervasoni and Francesco Di Michele during my research stay in Bristol; Filip Vlahović and Stepan Stepanović during their research stay here in Girona and the COST action meetings.

Finally thanks to Hodgkin House residents and management, who warmly welcomed me in Bristol.

This PhD has been possible with financial support granted by: Universitat de Girona (IF-UdG-2016 PhD fellowship and MOB-UdG-2018 mobility fellowship); European Research Council (ERC-2015-StG- 679001); Spanish MINECO (CTQ2017-87392-P) and computational resources from: Universitat de Girona; Barcelona Supercomputing Centre.

List of publications

Chapter 4-7 of this thesis are respectively based on the following publications:

1. L. D'Amore, L. Belpassi, J. E. M. N. Klein and M. Swart. Spin-resolved charge displacement analysis as intuitive tool for the evaluation of cPCET and HAT scenarios. *Chem. Commun.*, 2020, *56*, 12146.

2. L. D'Amore, K. Ray and M. Swart. Effect of ligand modification on the reactivity of bioinspired non-heme oxoiron(IV) complexes. *In preparation*.

3. C. G. Acevedo-Rocha,[#] A. Li,[#] L. D'Amore,[#] S. Hoebenreich, J. Sanchis, P. Lubrano, M. P. Ferla, M. Garcia-Borràs, S. Osuna and M. T. Reetz. Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. *Nat. Commun.*, 2021, *12*, 1621.

4. A. Li,[#] C. G. Acevedo-Rocha,[#] L. D'Amore,[#] J. Chen[#], Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou and M. T. Reetz. Regio- and stereoselective steroid hydroxylation at C7 by cytochrome P450 monooxygenase mutants. *Angew. Chem. Int. Ed.*, 2020, *59*, 12499.

Other publications not included in this thesis:

5. A. Brinkmeier,[#] K. E. Dalle,[#] L. D’Amore,[#] R. A. Schulz, S. Dechert, S. Demeshko, M. Swart and F. Meyer. Modulation of a μ-1,2-peroxo dicopper(II) intermediate by strong interaction with alkali metal ions. *J. Am. Chem. Soc.*, 2021, *143*, 17751.

6. R. Trammel, L. D'Amore, A. Cordova, P. Polunin, N. Xie, M. Siegler, P. Belanzoni, M. Swart and I. Garcia-Bosch. Directed hydroxylation of sp^2 and sp^3 C–H bonds using stoichiometric amounts of Cu and H_2O_2. *Inorg. Chem.*, 2019, *58*, 7584.

[#] These authors contributed equally.

List of abbreviations

Abbreviation	**Description**
α-KG	α-ketoglutarate
ACF	Adiabatic connection formula
ADF	Amsterdam density functional
AE	Arylesterase enzyme
aMD	Accelerated molecular dynamics
AO	Atomic orbital
BEP	Bell–Evans–Polanyi
Buea	1,1,1-tris[(N'-tert-butylureaylato)-N-ethyl]aminato
CAST	Combinatorial active-site saturation test
CDF	Charge displacement function
cMD	Conventional molecular dynamics
COSMO	Conductor-like screening model
cPCET	Concerted proton coupled electron transfer
CpdI	Compound I
CPR	Cytochrome P450 reductase
Cryo-EM	Cryogenic electron microscopy
CT	Charge transfer
Cychex	Cyclohexane
CYP	Cytochrome P450 enzyme
Cys	Cysteine residue
DNA	Deoxyribonucleic acid
DE	Directed evolution
DFA	Density functional approximation
DFT	Density Functional Theory
EAO	Electron accepting orbital
ECP	Effective core potential
EER	Exchange-enhanced reactivity
ET	Electron transfer
FAD	Flavin-adenine dinucleotide
FC	Frozen core approximation
FMN	Flavin mononucleotide
FMO	Frontier molecular orbital
GAFF	Generalized amber forcefield
GGA	Generalized gradient approximation

GPU	Graphics processing unit
GTO	Gaussian-type orbital
HAA	Hydrogen atom abstraction
HAT	Hydrogen atom transfer
His	Histidine residue
HOMO	Highest occupied molecular orbital
IBO	Intrinsic bond orbital
ISM	Iterative saturation mutagenesis
KDE	Kernel density estimation
KE	Kemp eliminase enzyme
KS	Kohn-Sham
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
L–J	Lennard-Jones
LUMO	Lowest unoccupied molecular orbital
MD	Molecular dynamics
ME	Magnitude epistasis
MeCN	Acetonitrile
Me_3NTB	tris[(N-methyl-benzimidazol-2-yl)methyl]amine
ML	Machine learning
MM	Molecular mechanics
MO	Molecular orbital
MOF	Metal organic framework
NAC	Near attack conformation
NAD(P)H	Nicotinamice adenine dinucleotide (phosphate); red. form
NMR	Nuclear magnetic resonance
NpT	Isobaric-isothermal ensemble
NVE	Microcanonical ensemble
NVT	Canonical ensemble
N4Py	N,N-bis(2-pyridylmethyl)-bis(2-pyridyl)methylamine
O_h	Octahedral coordination
OTf	Trifluoromethanesulfonate anion
PCA	Principal component analysis
PCM	Polarizable continuum model
Por	Porphyrin
PT	Proton transfer
PTE	Phosphotriesterase enzyme
QUILD	Quantum-regions interconnected by local descriptions

QM	Quantum mechanics
RC	Reactant complex
RESP	Restrained electrostatic potential
RMSD	Root-mean-square deviation
RNR	Ribonucleotide reductase
RSE	Reciprocal sign epistasis
SE	Sign epistasis
SES	Solvent-excluding surfaces
sMMO	Soluble methane monooxygenase
SPE	Single point energy
SPM	Shortest path map
STO	Slater-type orbital
TauD	Taurine dioxygenase
TBP	Trigonal bipyramidal coordination
^{t}Bu$_3$TACN	1,4,7-tri-*tert*-butyl-1,4,7-triazacyclononane
tICA	Time-structure independent component analysis
TMC	1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane
TMCO	4,8,12-trimethyl-1-oxa-4,8,12-triazacyclotetradecane
TMG$_3$tren	1,1,1-tris{2-[N^2-(1,1,3,3-tetramethylguanidino)]ethyl}amine
TPA	tris(2-pyridylmethyl)amine
tpa^{Ph}	tris(5-phenylpyrrol-2-ylmethyl)amine
TQA	tris(2-quinolylmethyl)amine
TS	Transition state
TSR	Two-state reactivity
TTF	Total turnover frequency
USER	Uracil-specific excision reagent
WT	Wild type
ZORA	Zeroth order regular approximation

Contents

List of figures

List of tables

Summary

Nature is an immense source of chemical reactions, and since immemorial time it has inspired us to imitate and harness the extraordinary catalytic efficiency of one of its most evolved (and still evolving) machineries: enzymes. Terms such as "biocatalysis" and "bioinspired catalysis" were coined to describe the use and mimicking of enzymes and their highly active intermediates to accelerate chemical transformations of readily available compounds into valuable industrial and pharmaceutical products. Although inspired by Nature's excellence, biomimetic synthetic complexes may however lack the outstanding activity of their biological counterparts. On the other hand, natural enzymes on their own do not always cover the widespread industrial demand and their function needs to be further engineered by laboratory evolution. Undoubtedly, incredible strides have been made over the past 50 years, nevertheless improving biomimetic catalysts, enhancing enzymatic traits or even enabling novel functions remain challenging tasks, which are often frustrated by our sometimes limited understanding of the chemical process at the molecular level. Within this framework, computational chemistry constitutes a powerful method to strengthen such understanding and to obtain data which are often difficult or even impossible to acquire and rationalize experimentally. Hence, this thesis is part of an ongoing effort carried out at the *Institut de Química Computational i Catàlisi* and *CompBioLab* of the University of Girona, and focuses on the application of various computational techniques and tools to study different aspects encompassed by the oxidative transformation of non-activated aliphatic bonds, performed by either enzymes (P450) and enzyme-mimicking complexes containing iron.

The activation of C(sp^3)–H bonds with introduction of hydroxy functions under ambient conditions represents a challenge to modern chemistry, nonetheless it is commonly operated in Nature by various oxygenases that cleave strong C–H bonds using molecular oxygen as the oxidant in key metabolic transformation and biological synthesis. Precisely pre-organized active site pockets containing powerful electrophilic high-valent heme and non-heme iron-oxo intermediates in oxygenases are capable of stabilizing the transition state, thereby enhancing the reaction rate.

In this regard, the spin state and metal coordination are pivotal, determining whether a reaction channel with lower (or higher) activation barrier is operative. However, besides ordered pockets, enzymes are inherently dynamic, and their function is usually connected to accessible conformational states that can be sampled in solution. By modifying the natural amino acid sequence, laboratory evolution can alter their conformational dynamics, populating active states which lead to enhanced catalytic traits or even enable novel functions.

Chapter 1 of this thesis offers an overview covering these backgrounds, with particular focus on the different scenarios invoked during C–H hydroxylation performed by oxoiron(IV) complexes and the implications of enzyme engineering on conformational dynamics and evolvability, where the phenotypic consequences of sequence mutagenesis may also depend on the genetic background (epistasis). Due to the varied scale of the systems considered, ranging from the molecular to the macromolecular level, different computational methods and tools are used, and therefore discussed in **Chapter 2** from the theoretical point of view. **Chapter 3** presents the objectives of the thesis, whose primary aims are (i) to characterize the reactivity principles and the structure-reactivity relationship underlying the C–H hydroxylation catalyzed by high-valent biomimetic iron-oxo complexes, and (ii) to investigate the connection between conformational dynamics, epistasis and enzymatic traits (activity and selectivity) in cytochrome $P450_{BM3}$ mutants.

The results of the different projects undertaken during this PhD research are reported in chapters 4-7. In **Chapter 4**, a direct and chemically intuitive methodology consisting of the analysis of the charge displacement function (CDF) is proposed, and applied to taurine dioxygenase and lipoxygenase model systems to infer whether a hydrogen atom transfer (HAT) or a concerted proton-coupled electron transfer (cPCET) is the operating mechanism in the first hydroxylation step. Subsequently, in **Chapter 5** the CDF analysis is further applied to a series of oxoiron(IV) complexes bearing biologically inspired ligands, in conjunction with a detailed assessment of the electronic structure variation in response to ligand modifications, and how this ultimately determines the preferred reaction channel. **Chapter 6** is dedicated to explore the connection within epistatic effects, conformational dynamics and selectivity for fine-tuning multiple functions of laboratory-evolved P450$_{BM3}$ variants for the selective hydroxylation of steroidal substrates. The same reaction catalyzed by a distinct P450$_{BM3}$ mutant at the pharmaceutically relevant C7β position is investigated in **Chapter 7** to decipher the origin of the regio- and stereoselectivity at the molecular level. From computer simulations, rational mutagenesis at key residues is finally proposed to design a new variant that could accommodate bulkier steroids retaining identical selectivity.

Resum

La natura és una enorme font de reaccions químiques, i des de temps immemorials ens ha inspirat i guiat cap a l'ús d'enzims, unes de les màquines més evolucionades (i encara en procés d'evolució) de les quals n'hem aprofitat les seves extraordinàries eficiències catalítiques. Termes com ara "biocatàlisi" i "catàlisis bioinspirada" es van crear per descriure tant l'ús com la imitació dels enzims i les seves espècies intermèdies altament actives per a l'obtenció de productes industrials i farmacèutics mitjançant la transformació de compostos fàcilment accessibles i comercials. Encara que inspirats per l'excel·lència de la natura, els complexos sintètics biomimètics no es caracteritzen per tenir les excepcionals activitats observades en els seus equivalents biològics. Per altra banda, els enzims naturals, de forma inherent no sempre abasteixen els amplis requisits industrials, de manera que per ser funcionals necessiten ser dissenyats i modificats al laboratori. Tot i que sense dubte durant els últims 50 anys s'han dut a terme increïbles avenços, tant la millora de catalitzadors biomimètics, de trets enzimàtics o inclús la creació de noves funcionalitats encara és tot un repte, una tasca desafiant que normalment es veu limitada pel nostre coneixement del procés químic a nivell molecular. En aquest context, la química computacional representa un mètode poderós que es pot emprar per ampliar aquest coneixement, per a obtenir i racionalitzar dades que sovint són difícils o fins i tot impossibles d'aconseguir fent ús de les tècniques experimentals. Per tant, aquesta tesi és part d'un esforç encara en procés, una tasca realitzada a l'Institut de Química Computacional i Catàlisi i el CompBioLab a la Universitat de Girona. Concretament, aquesta tesi es centra en l'aplicació de diferents tècniques i eines computacionals per tal d'englobar diferents aspectes de la transformació oxidativa d'enllaços alifàtics no actius, realitzada tant per enzims (P450) com per complexos bioinspirats en enzims que contenen ferro.

L'activació del enllaços de C(sp^3)–H amb la introducció de grups funcionals hidroxil sota condicions ambientals representa un dels reptes actuals per a la química moderna, encara que a la Natura s'observa de forma comuna com varies oxigenases s'encarreguen del trencament de l'enllaç C–H utilitzant una molècula d'oxigen com a oxidant en processos clau de transformació metabòlica i síntesis biològica. Els centres o cavitats actives dels enzims oxigenases presenten una pre-organització precisa i concreta que li ofereix crear un espai electrofílic amb

poderosos intermedis hemo i no-hemo de ferro-oxo d'alta valència capaços d'estabilitzar l'estat de transició, millorant així la velocitat de la reacció. Referent a això, l'estat d'espín i la coordinació del metall són essencials, ja que determinen si una reacció és operativa a través d'una via de reacció amb barrera d'activació inferior (o superior). Tot i així, a part del centres actius ordenats, els enzims són inherentment dinàmics i la seva funció també està connectada als diferents estats conformacionals que s'exploren i que són accessibles en solució. Modificant la seqüència d'aminoàcids natural al laboratori es poden alterar la dinàmica d'aquests estats conformacionals, de manera que es pot incrementar la població de conformacions actives que condueixen a estats de l'enzim amb trets catalítics millorats o inclús estats que presenten noves funcionalitats.

El **Capítol 1** d'aquesta tesi ofereix una visió general que cobreix aquests antecedents, amb especial atenció als diferents escenaris presents durant la hidroxilació C–H realitzada pels complexos de ferro-oxo (IV) i pel que fa les implicacions de l'enginyeria enzimàtica sobre la dinàmica conformacional i la seva capacitat d'evolució, on les conseqüències fenotípiques de la mutació de la seqüència poden també dependre dels antecedents genètics (epístasi). A causa de la variada escala dels sistemes considerats, tant des del nivell molecular com fins al nivell macromolecular, les diferents eines i mètodes computacionals emprats en aquesta tesi es discuteixen des del punt de vista teòric al **Capítol 2**. En el **Capítol 3** es presenten els objectius de la tesi, essent les finalitats principals (i) la investigació dels diversos aspectes subjacents al pas determinant de la reacció en la hidroxilació C–H catalitzada pels complexos ferro-oxo biomimètics amb valència elevada i (ii) l'estudi de la relació entre dinàmiques conformacionals, epístasi i selectivitat en les mutants del citocrom P450$_{BM3}$.

Els resultats dels diferents projectes realitzats durant la investigació del doctorat es presenten als Capítols 4-7. Al **Capítol 4**, es proposa una metodologia directa i químicament intuïtiva que consisteix en l'anàlisi de la funció de desplaçament de càrrega, que s'aplica als sistemes model de taurina dioxigenasa i lipoxigenasa per inferir si una transferència d'àtoms d'hidrogen (HAT) o una transferència concertada d'electrons acoblats a protons (cPCET) és el mecanisme d'operació en el primer pas d'hidroxilació. Posteriorment, al **Capítol 5**, l'anàlisi del desplaçament de càrrega s'aplica a una sèrie de complexos de ferro-oxo (IV) basats en lligands d'inspiració biològica, conjuntament amb una avaluació detallada de la variació de l'estructura electrònica en resposta a modificacions del lligand, i com això determina en última instància la preferència de la ruta de reacció. El **Capítol 6** està dedicat a explorar la connexió dins dels efectes epistàtics, la dinàmica conformacional i la selectivitat per afinar múltiples funcions de variants $P450_{BM3}$ desenvolupades al laboratori per a la hidroxilació selectiva de substrats esteroides. La mateixa reacció catalitzada per un altre mutant $P450_{BM3}$ a la posició C7β, la qual és rellevant dins de l'àmbit farmacèutic, s'investiga al **Capítol 7** per tal de desxifrar l'origen de la regió- i l'estereoselectivitat a nivell molecular. A partir de la simulacions computacionals, finalment es proposa la mutagènesi en residus clau per tal de dissenyar racionalment una nova variant que pugui acollir esteroides més voluminosos i que conservin una selectivitat idèntica.

Resumen

La Naturaleza es una fuente inmensa de reacciones químicas, y desde tiempos inmemoriales nos ha inspirado a imitarla y a aprovechar la extraordinaria eficiencia catalítica de una de sus más evolucionadas (y todavía en evolución) maquinarias: las enzimas. Términos como “biocatálisis” o “catálisis bioinspirada” fueron acuñados para describir el uso y la imitación de las enzimas y sus intermedios altamente activos para acelerar transformaciones químicas de compuestos fácilmente disponibles en productos de alto valor industrial y farmacéutico. A pesar de estar inspirados por la excelente Naturaleza, los complejos sintéticos bioinspirados pueden, sin embargo, no presentar de la sobresaliente actividad de sus contrapartes biológicas. Por otro lado, las enzimas naturales, por sí solas, no siempre son capaces de cubrir la demanda industrial y su función necesita ser diseñada mediante evolución de laboratorio. Indudablemente, en los últimos 50 años se han dado pasos enormes, no obstante, mejorar la catálisis biomimética, aumentar algunas características de las enzimas o incluso permitir nuevas funciones enzimáticas siguen siendo un reto complicado, que comúnmente se ve limitado por nuestro escaso entendimiento del proceso químico a nivel molecular. En este marco, la química computacional constituye un potente método que puede ser usado para fortalecer dicho entendimiento, así como para obtener datos que a nivel experimental serían a menudo extremadamente difícil, o incluso imposibles de obtener. Por tanto, esta tesis es parte de un esfuerzo que se está llevando a cabo en el Institut de Química Computacional i Catàlisis y en el CompBioLab de la Universitat de Girona, que se centra en la aplicación de diferentes técnicas y herramientas computacionales a diferentes aspectos de la transformación oxidativa de enlaces alifáticos no activados, que puede llevarse a cabo bien por enzimas (P450) o por complejos metálicos de hierro similares a las enzimas.

La activación de los enlaces C(sp^3)–H que introducen grupos funcionales hidroxilo bajo condiciones ambiente, representa un reto en la química moderna. No obstante, en la Naturaleza es comúnmente efectuado por varias oxigenasas que rompen los enlaces fuertes C–H utilizando oxígeno molecular como oxidante en transformaciones metabólicas clave y en síntesis biológica. Las cavidades del centro activo preorganizadas que contienen potentes grupos electrofílicos de alta valencia,

hemo y no-hemo con intermedios hierro-oxo, en oxigenasas son capaces de estabilizar el estado de transición, y, debido a esto, acelerar la velocidad de reacción. En este contexto, el estado de espín, y la coordinación del metal son esenciales, determinando si una ruta de reacción con una baja (o alta) barrera de activación es viable. Sin embargo, además de las cavidades ordenadas, las enzimas son intrínsecamente dinámicas, y su función está conectada a estados conformacionales accesibles que se pueden visitar en disolución. Modificando la naturaleza de la secuencia de amino ácidos, la evolución de laboratorio puede alterar la dinámica conformacional, poblando estados activos que pueden dar lugar a características potenciadas de la enzima, o incluso permitiendo la aparición de nuevas funciones.

El **Capítulo 1** de esta tesis ofrece una visión general que cubre este trasfondo, con particular énfasis en los diferentes escenarios de la hidroxilación de enlaces C–H llevados a cabo por complejos hierro-oxo(IV) y de las implicaciones de la ingeniería de proteínas en la dinámica conformacional y en su capacidad de evolución, donde las consecuencias fenotípicas de la mutagénesis en la secuencia pueden también depender del trasfondo genético (epistasis). Debido a la variedad de escalas de los sistemas que se han considerado, que varían desde el nivel molecular al macromolecular, se han utilizado diferentes métodos y herramientas, para así ser discutidos en el **Capítulo 2** desde un punto de vista computacional. En el **Capítulo 3** se presentan los objetivos de la tesis, cuyas finalidades primarias son (i) la investigación de varios aspectos que subyacen en el paso determinante de la velocidad de reacción en la hidroxilación de enlaces C–H catalizada por complejos hierro-oxo biomiméticos de alta valencia y (ii) estudiar la relación entre dinámica conformacional, epistasis y selectividad en mutantes del citocromo $P450_{BM3}$.

Los resultados de los diferentes proyectos llevados a cabo durante la investigación de doctorado se describen en los capítulos 4-7. En el **Capítulo 4**, se propone una metodología directa y químicamente intuitiva que consiste en el análisis de la función de desplazamiento de carga, aplicada a sistemas modelo de taurina dioxigenasa y lipoxigenasa para discernir si el mecanismo en el primer paso de hidroxilación se trata de una transferencia de átomo de hidrógeno (HAT) o una mecanismo de transferencia de electrones acoplada a protones (cPCET). Posteriormente, en el **Capítulo 5** el análisis de desplazamiento de carga se continúa aplicando a una serie de complejos hierro-oxo(IV) constituidos por ligandos bioinspirados, juntamente con una evaluación detallada de la variación en la estructura electrónica en respuesta a modificaciones del ligando, y como éste determina la ruta de reacción preferida. El **Capítulo 6** está dedicado a explorar la conexión con efectos epistáticos, dinámica conformacional y selectividad para ajustar múltiples funciones de variantes evolucionadas de P450$_{BM3}$ para la hidroxilación selectiva de sustratos esteroideos. La misma reacción catalizada por otra mutante de P450$_{BM3}$ en la posición farmacológicamente relevante C7β es investigada en el **Capítulo 7**, para descifrar el origen de la regio- y estereoselectividad a nivel molecular. A partir de simulaciones computacionales, finalmente se ha propuesto mutagénesis racional en residuos clave para el diseño de una nueva variante capaz de acomodar sustratos esteroideos más voluminosos manteniendo la misma selectividad.

Chapter 1: Introduction

1.1 C–H activation

The selective functionalization of relatively inert, saturated C(sp^3)–H bonds takes places in a myriad of diverse biological functions like drug[1] and xenobiotic[1c, 2] metabolism, hormone regulation,[3] antibiotic biosynthesis,[4] to name just a few. Although the cleavage of such strong bonds under ambient conditions represents an intrinsically difficult and challenging goal even in modern chemistry, this task has been routinely carried out with ease for billions of years by one of the most evolved (yet, still evolving) Nature's machineries: *enzymes*. The oxidative transformation of aliphatic bonds mediated by enzymes has attracted sustained attention over the past decades due to the appealing combination of mechanistic complexity and potential synthetic application in industrial processes.[5]

Among various oxidative processes, the direct oxidation of C–H bonds using O_2 constitutes an environment-friendly and efficient method to synthesize a variety of products starting from simple and readily available materials. However, molecular oxygen is kinetically stable towards reactions at room temperature because of its triplet ground state ($^3O_2$), which makes the two-electron reaction with closed-shell organic partners spin-forbidden,[6] and the reduction to superoxide anion unfavorable by 7.8 kcal mol^{-1} due to the highly negative one-electron reduction potential of triplet oxygen.[7] Nature elegantly overcomes the spin state barrier, harnessing the strong oxidizing capabilities of O_2 by employing enzymes containing 3d transition metals (such as Mn, Fe and Cu), which bind and activate oxygen in physiological conditions to carry out a countless number of essential biological processes,[8] including the challenging hydroxylation of the 105 kcal mol^{-1} C–H bonds of methane by soluble methane monooxygenase (sMMO) found in methanotrophs. Metal ions can supply the one-electron reductive activation of O_2 by metal coordination and at the same time serve as multielectron reductants to access thermodynamically more favorable two-electron, or even four-electron reduction pathways.[6d]

Among first-row transition metals, iron is one of the most commonly employed metalloenzymes, due to its bioavailability in Nature and its capability to access multiple oxidation states, in which different open-shell spin states may additionally be available. The different spin states arise because of the metal *d*-orbitals (Fig. 1.1) that can be occupied in a variety of ways, depending on the number of unpaired *d*-electrons present and the coordination around the metal, with in-plane tetra-coordination favoring intermediate spin (porphyrin ligands)[9] and trigonal bipyramidal (TBP) coordination favoring high spin.[10]

The high-spin iron(II) is the most relevant in binding and activating O_2 to generate high-valent oxoiron(IV) and formally oxoiron(V) active intermediates to carry out the C–H bond functionalization in the catalytic cycle of numerous enzymatic systems.[11]

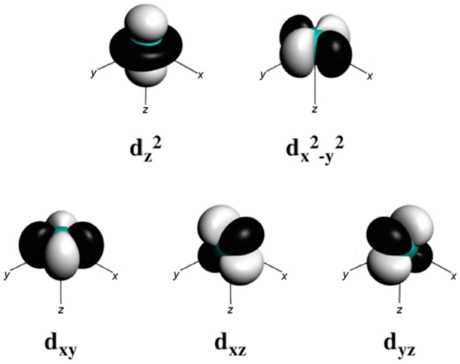


Figure 1.1. Metal d-orbitals (anti-bonding top, non-bonding bottom).

1.2 Iron-oxo biological intermediates

High-valent oxoiron active sites have been identified in both heme and non-heme enzymes (Fig. 1.2). The heme-containing oxygenases, peroxidases and catalases feature a mononuclear iron(II)-protoporphyrin IX active site coordinating a proximal cysteine (S), histidine (N) or tyrosine (O) residue as axial ligand, respectively.[12] The remaining sixth coordination site is usually occupied by a weakly bound water molecule that is displaced by the substrate upon binding, thereby generating highly oxidized intermediates up to the Fe(IV)-oxo in compound I (**CpdI**),[13] which is considered the major oxidant species in this class of enzymes. On the other hand, non-heme enzymes may comprise (i) mononuclear iron centers, generally coordinated to the characteristic two histidines and one carboxylate (2-His-1-carboxylate, where the carboxylate moiety is an aspartate or glutamate residue) facial triad[14] as in Rieske dioxygenase and α-ketoglutarate (α-KG) dependent Taurine dioxygenase (TauD),[8b] or even (ii) dinuclear iron clusters often coordinating two histidines and four carboxylates (2-His-4-carboxylate) structural motifs, as found in sMMO and ribonucleotide reductase (RNR) from *E. coli*.[8d]

Very often the enzymatic activation of O_2 is initiated by the Fe(II) state to carry out a variety of two-electron oxidation processes, then a co-substrate provides the remaining two reducing equivalents required for the four-electron reduction of O_2. In TauD the co-substrate is the α-KG, which delivers two electrons directly to the active site to form the Fe(IV)-oxo **J** intermediate (**TauD-J**)[15] whereas other enzymes such as sMMO, Rieske dioxygenase or cytochromes P450 normally use NAD(P)H as the electron donor, generating formally Fe(V)-oxo species[16]. Moreover, the redox equivalents of the formal Fe(V)-oxo species can be distributed either over the ligand, as in the Fe(IV)-oxo porphyrin π radical cation ($\mathrm{Por}^{\bullet +}\mathrm{Fe}^{IV}{=}O$) **CpdI** intermediate of P450 enzymes[17], or stored at the metal center in non-heme enzymes like the $(OH)\mathrm{Fe}^{V}{=}O$ intermediate of Rieske dioxygenase[18] or the $\mathrm{Fe}^{IV}_2(\mu\text{-}O)_2$ **Q** intermediate of sMMO.[8c, 8d, 19] In heme-based enzymes, the Fe(IV) center has a $S=1$ (triplet) spin state, which is coupled ferromagnetically or antiferromagnetically with the porphyrin radical cation ($S=1/2$), giving an overall $S=3/2$ (quartet) or $S=1/2$ (doublet) state in **CpdI** intermediates, respectively.[17, 20] In contrast, the spin state of the non-heme Fe(IV)-oxo within enzymes is usually high spin $S=2$ (quintet)[8b], although in dinuclear iron enzymes the diferrous centers can interact with each other via bridging ligands, either antiferromagnetically or ferromagnetically to afford an overall $S=0$ or $S=4$ ground state, respectively.[8d]

High valent Fe=O intermediates of enzymes like P450, sMMO, TauD and Rieske dioxygenase are known to perform even the most difficult C–H bond cleavage for a wide variety of organic compounds. Owing to their high reactivity, these biological intermediates are short-lived and often difficult to trap for chemical and physical characterization. Nevertheless, the investigation of suitable synthetic complexes inspired by enzyme intermediates can shed light, enhancing our basic understanding of the structure-function relationship in metalloenzymes and aid the development of bioinspired catalysts as efficient as enzymes.

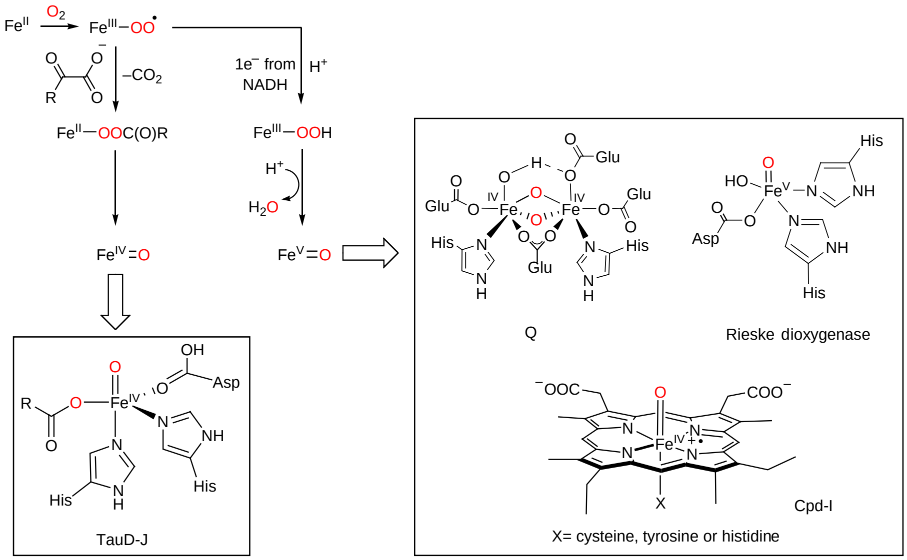
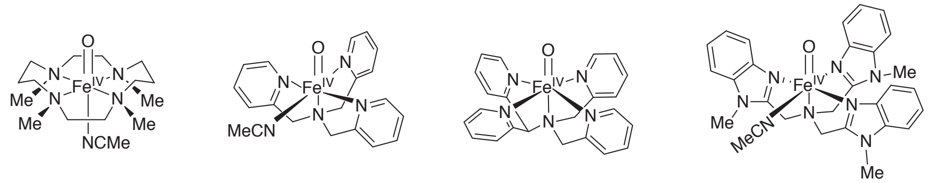


Figure 1.2. Mechanisms of dioxygen activation by iron-containing heme and non-heme oxygenases. Boxes depict high-valent intermediates of enzymatic reaction.

1.3 Iron-oxo model complexes

The first X-ray characterization of a synthetic $Fe^{IV}(O)$ model complex by Que, Nam, Münk and co-workers dates back to 2003[21], the same year that the TauD-J intermediate was also reported.[22] This complex, $[Fe^{IV}(O)(TMC)(NCMe)]^{2+}$ (**1**) was obtained using the macrocyclic 1,4,8,11-tetramethyl-1,4,8,11-tetraazacyclotetradecane ligand (TMC), which features a weak coordination site *trans* to the oxo unit, bearing a MeCN solvent molecule. The publication of the first crystal structure of a $Fe^{IV}(O)$ model complex was then followed by the synthesis and characterization of a multitude of similar high-valent iron-oxo compounds. The same year, a spectroscopically characterized $Fe^{IV}(O)$ model complex $[Fe^{IV}(O)(TPA)(NCMe)]^{2+}$ (**2**) was reported,[23] bearing a tripodal tetradentate tris(2-pyridylmethyl)amine ligand (TPA), and contrary to the TMC ligand, a weak coordination site *cis* to the oxo moiety occupied by MeCN. Subsequently, a third prototypical class of pentadentate N,N-bis(2-pyridylmethyl)-bis(2-pyridyl)methylamine supporting ligand (N4Py) with *no* open coordination site available was employed to synthesize[24] and characterize by X-ray[25] the $[Fe^{IV}(O)(N4Py)]^{2+}$ complex (**3**), exhibiting a remarkable thermal stability ($t_{1/2}$= 60h at 298 K). These early examples with weak *trans*, *cis*, or *no* coordination sites represent the three archetypes of almost all the $Fe^{IV}(O)$ complexes published afterwards in literature.[26] However, unlike the highly reactive S=2 high spin biological intermediates, the majority of the model complexes counterparts possess (pseudo) octahedral coordination (O_h) with nonbiological S=1 spin state, and rather sluggish oxidizing power.



$[Fe^{IV}(O)(TMC)(NCMe)]^{2+}$
1

$[Fe^{IV}(O)(TPA)(NCMe)]^{2+}$
2

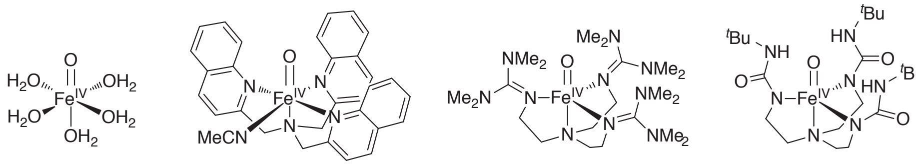
$[Fe^{IV}(O)(N4Py)]^{2+}$
3

$[Fe^{IV}(O)(Me_3NTB)(NCMe)]^{2+}$
4

Figure 1.3. Synthetic S=1 oxoiron(IV) model complexes.

Of particular note is the $[Fe^{IV}(O)(Me_3NTB)(NCMe)]^{2+}$ complex[27] (**4**) bearing a tris[(N-methyl-benzimidazol-2-yl)methyl]amine ligand (Me_3NTB), related to TPA but with all the pyridine donors replaced by N-methylbenzimidazoles. Interestingly, despite having a S=1 ground state in apparently an octahedral coordination environment, **4** exhibits a high reactivity capable of activating strong C–H bonds, such as those in cyclohexane (bond dissociation energy of 99.3 kcal mol^{-1}), even at a low temperature (k_2=0.25 M^{-1} s^{-1} at 233 K).

Much effort has been directed towards the synthesis of high-spin iron(IV)-oxo complexes mimicking their biological intermediates counterparts. In order to model the correct S=2 spin state and understand its effect on the reactivities of oxidation reactions, two different strategies have been used to achieve this goal: (i) weakening the equatorial ligand field donation in O_h geometry or (ii) enforcing TBP coordination at the iron(IV) center. Outstanding examples of oxoiron(IV) model compounds featuring the biological quintet ground state obtained using the first approach are represented by the $[Fe^{IV}(O)(H_2O)_5]^{2+}$ complex (**5**) synthesized by Bakac, Que, Münck and co-workers in 2005[28], which constitutes the first $Fe^{IV}(O)$ S=2 model compound identified, and by the $[Fe^{IV}(O)(TQA)(NCMe)]^{2+}$ complex (**6**) (TQA = tris(2-quinolylmethyl)-amine),[29] which exhibits the highest cyclohexane oxidation rate reported up to date (k_2=0.37 M^{-1} s^{-1} at 233 K) for an oxoiron(IV) complex. On the other hand, using bulky tripodal ligands to exploit the second strategy permitted to obtain the first crystallographically characterized S=2 oxoiron(IV) model compound, $[Fe^{IV}(O)(TMG_3tren)]^{2+}$ (TMG_3tren = 1,1,1-tris{2-[N^2-(1,1,3,3-tetramethyl-guanidino)]ethyl}amine)[30] (**7**) followed by $[Fe^{IV}(O)(H_3buea)]^-$ (buea = 1,1,1-tris[(N'-tert-butylureaylato)-N-ethyl]aminato)[31] in 2010 (**8**). In another example a phenylpyrrolydines (tpa^{Ph})-based tripodal ligand was also employed to achieve the S=2 state with TBP geometry.[32]



$[Fe^{IV}(O)(H_2O)_5]^{2+}$
5

$[Fe^{IV}(O)(TQA)(NCMe)]^{2+}$
6

$[Fe^{IV}(O)(TMG_3tren)]^{2+}$
7

$[Fe^{IV}(O)(H_3buea)]^-$
8

Figure 1.4. Synthetic S=2 oxoiron(IV) model complexes.

Depending upon the coordination symmetry at the oxoiron(IV) center, the ground spin state can be inferred from simple considerations of ligand-field theory. As discussed elsewhere in detail[33] the high-valent iron and oxo moieties interact covalently via Fe-$d_{xz/yz}$ and O-$p_{x/y}$ atomic orbitals in two half $\pi_{xz/yz}$ bonds, and via Fe-d_{z^2} and O-p_z in a $\sigma^*_{z^2}$ bond. The Fe-$d_{x^2-y^2}$ interacts with the equatorial ligand, while the Fe-d_{xy} is non-bonding. The spin ground-state is governed by the energetic separation of the $d_{x^2-y^2}$ and d_{xy} orbitals. Within the ligand field representation of the O_h geometry (C_{4v} symmetry), the energetic ordering of the Fe=O orbitals $d_{xy} < \pi^*_{xz,yz} < d_{x^2-y^2} < \sigma^*_{z^2}$ is a consequence of the equatorial compression associated with the strong Fe=O bond. The $d_{x^2-y^2}$ orbital is raised in energy by strong-donor equatorial ligands, resulting in a $d_{x^2-y^2}$ / d_{xy} energy gap that is higher than the spin-pairing energy, which favors a doubly occupied d_{xy} orbital and a S=1 state. In complex **5**, the weak field generated by the equatorial tetra-aqua ligand set permits a low-lying $d_{x^2-y^2}$ that can be occupied, resulting in a S=2 state. Similarly, complex **6** uses the TQA ligand (a variant of the TPA used in **2**). In TQA the replacement of the quinoline donors in place of the pyridines used in TPA elongates the Fe–N distance, generating a weaker ligand field. Consequently, in **6** the weak field resulting from TQA favor the stabilization and occupation of the $d_{x^2-y^2}$ leading to the S=2 ground state. In TBP coordination instead, the $d_{x^2-y^2}$ and d_{xy} orbitals become energetically degenerate, thus can be both occupied by two unpaired electrons, giving rise to the S=2 state[26] (Fig. 1.5)

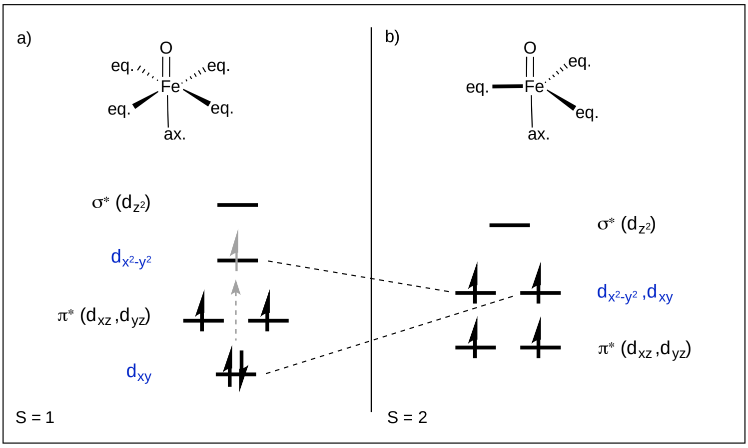


Figure 1.5. Simplified molecular orbital diagram for (a) S=1 octahedral and (b) S=2 trigonal bipyramidal geometry coordination. In (a) the spin-flip forbidden transition $d_{xy} \rightarrow d_{x^2-y^2}$ (gray dashed arrow) can give rise to a S=2 state if the orbital promotional energy is lower than the spin-pairing energy, as found for weak-donor equatorial ligands.

It is found abundantly[33a, 33b, 34] that S=2 is, in general, the most reactive state. However, sterically encumbered ligand such as TMG_3tren and buea used in complex **7** and **8** respectively, may prevent the access of the substrate to the Fe^{IV}(O) unit reducing their oxidizing power, thus, making these complexes poor functional models of enzymatic intermediates. Nevertheless, for complex **7** it has been demonstrated that the removal of one of the three bulky side-arms in the TMG_3tren ligand improves substrate accessibility, hence increasing significantly the oxidative reactivity.[35]

Another example that seems to contradict the general notion that the high-spin Fe^{IV}(O) should be considerably more reactive than the low spin is evidenced by comparing the cyclohexane oxidation rates of complexes **4** and **6**, which are very similar despite their respective S=1 and S=2 spin states. Despite **4** and **6** were proposed to be in O_h coordination, based on Density Functional Theory (DFT) calculations and spectroscopic characterization at 4 K[27, 29], further DFT studies[36] predicted that both species may lose one coordinated solvent ligand at 233 K, assuming a TBP structure and a quintet state (two state reactivity, section 1.5.2). Hence, reactions of **4** and **6** at 233K may follow the high-spin path in both cases, explaining the similar reactivity rates. Finally, a tripodal ligand ^{t}Bu$_3$TACN (1,4,7-tri-*tert*-butyl-1,4,7-triazacyclononane) was recently employed to synthesize and characterize a novel S=2 $Fe^{IV}(O)$ complex, exhibiting a pseudo-tetrahedral geometry distinct from the common O_h or TBP coordination previously reported, together with remarkable oxidizing power.[37]

1.4 C–H hydroxylation mechanism

In Nature, the oxidative hydroxylation of C–H bond is involved in many metabolic pathways and detoxification.[1-3] It faciliates the conversion of lipophilic compounds into water-soluble products that are removed by the liver or kidneys. Enzymes such as P450, sMMO, TauD and Rieske dioxygenase are known to realize even the most difficult hydroxylation of saturated carbon centers in a variety of organic compounds, and their reaction mechanism has been well established.[8, 13, 16, 19a, 20, 38]

Bioinspired oxoiron(IV) oxidants hydroxylate C–H bonds through enzyme-like stepwise mechanism, which is proposed to include a rate determining hydrogen atom abstraction (HAA) operated by the $Fe^{IV}(O)$ core. This yields a $Fe^{III}(O)H$ intermediate and a carbon centered alkyl radical $R_3C\bullet$, followed by different possibilities: (i) a rebound mechanism, where the alkyl radical is caged and it is captured by the OH moiety of the $Fe^{III}OH$ intermediate to give the hydroxylated product; (ii) a dissociative mechanism, where the alkyl radical reacts with an additional oxidant (O_2 or oxoiron) after escaping the solvent enclosure; or (iii) a desaturation mechanism, where the radical reacts with a second $Fe^{IV}(O)$ species to yield an alkene product (Fig. 1.6).

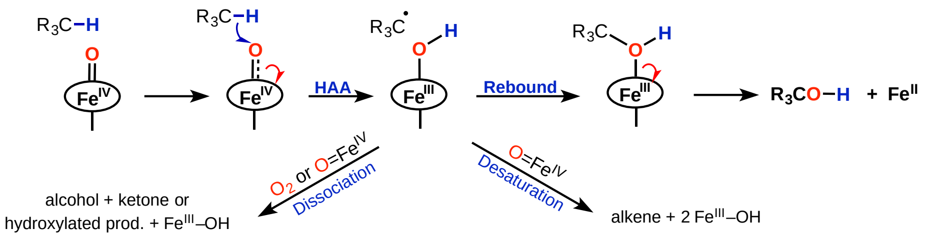


Figure 1.6. Schematic representation of possible reaction pathways of iron-containing oxygenases and model complexes.

It is speculated[39] that in enzymes the substrate is trapped within the active site cage and the rebound step is operative and fast, therefore the hydroxylated product is formed with high stereo-retention. Conversely, the active reaction center of synthetic model complexes is surrounded only by the solvent and it is not well-caged, thus radical rearrangement (i.e. dissociation or desaturation) may become prominent, especially if the corresponding activation barrier is smaller than the rebound barrier.[40]

1.4.1 cPCET vs. HAT

During the ordinarly rate-determining HAA reaction different scenarios may operate, depending on whether the proton (H^+) and its electron (e^-) are transferred from the substrate C(sp^3)-H orbital to the oxo moiety together in time and space. A useful general guidance[41] is to distinguish between (i) the stepwise mechanism where the electron transfer (ET) and the proton transfer (PT) take place in a sequential manner, and (ii) concerted PCET† (cPCET) in which the proton and the electron are transferred

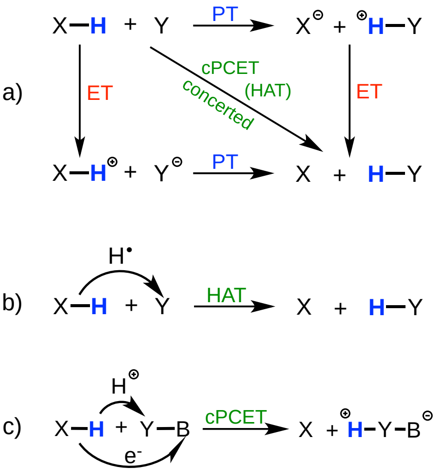


Figure 1.7. a) Concerted vs stepwise CPET mechanism. b) Hydrogen atom transfer (HAT). c) Concerted proton-coupled electron transfer (cPCET) that is not HAT.

---

† Various acronyms are used in the literature for this specific scenario, such as EPT and CPET. Unfortunately, at times simply PCET, the overarching term, is used to describe such scenarios.

simultaneously, but may travel to different endpoints. A variant of the latter, where the proton and the electron move together instead, is best described as the transfer of a genuine hydrogen atom (radical), which is usually denoted hydrogen atom transfer (HAT) reaction. Fig. 1.7 offers an overview of these representative cases.

1.5 Reactivity

In order to understand, modeling and predicting the reactivity of the high-valent transition-metal complexes, three views have been put forward in the past decade. These are based either on orbital energy, exchange interaction or the Hammond postulate, in all cases involving orbital overlap, electron flows and electronic reorganization.[34d]

1.5.1 Orbital picture

The key role played by orbital overlap and frontier molecular orbital (FMO) interactions on reactivity is well-known from the pioneering works on organic molecule by Woodward-Hoffmann[42] and Fukui.[43] In the HAA mechanism, the substrate (donor) transfers one σ_{CH} electron to an empty (or partially-empty) metal-oxo based orbital (acceptor). The electron accepting orbital (EAO) can be either the $\sigma^*_{z^2}$ (Fe d_{z^2} – O p_z) or $\pi^*_{xz/yz}$ (Fe $d_{xz/yz}$ – O $p_{x/y}$) antibonding orbital of the d-metal block bearing significant oxygen character, thereby giving rise to a σ or π reaction channel, respectively. In the reactivity of $Fe^{IV}(O)$ complexes, these two competing pathways are distinguished by the Fe-O-H angle, which is almost linear (165-175°) in the σ-channel and bent (120-140°) in the π-channel.[34b-d]

Conventionally, the most favored mechanism in HAA entails the transfer of an α-spin electron from the substrate to the $\sigma^*_{z^2}$ EAO on the S=2 surface ($^5\sigma$-pathway), due to the greatest extent of overlap between donor and acceptor orbitals.[44] Conversely, in the S=1 state the energy cost to shift an α electron to the $\sigma^*_{z^2}$ orbital is too high, making the $^3\sigma$ path less feasible,[33a, 34b-d] while the $\pi^*_{xz/yz}$ may accept a β electron from the C–H bond, making the $^3\pi$-pathway predominantly favored. In fact, Sautet and co-workers obtained[45] a remarkable linear correlation between the σ^* orbital energy and the HAA barrier for high-spin $[Fe^{IV}(O)(L)]^{2+}$ complexes (where L is a mono- or tri- to hexa-dentate ligand), showing that a relation between the π^* orbital energy and activation barrier also holds for the $^3\pi$ path in a series of neutral oxoiron(IV) complexes. In other words, the σ-path is the preferred reaction channel on the quintet surface, whereas the π-path becomes active in the triplet state.

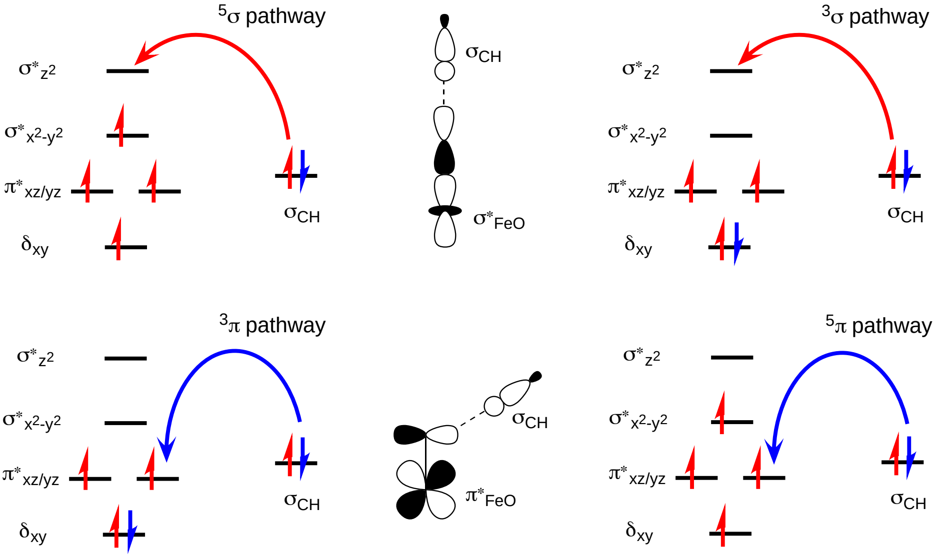


Figure 1.8. Electron transfer diagrams for the possible C–H activation pathways outlined in the chemistry of non-heme oxoiron(IV) complexes.

However, the less feasible $^3\sigma$ and $^5\pi$ pathways may also be operative[46] depending on the electronic and geometrical availability. As a proof of concept, taking into account the $[Fe^{IV}(O)(NH_3)_5]^{2+}$ model system, Neese and co-workers calculated by means of DFT that the barrier of ethane hydroxylation follows the order $^5\sigma<^3\pi\approx^5\pi<^3\sigma$ (Fig. 1.9).[34b] Noteworthy, the comparison of DFT with more expensive coupled cluster calculations on the same system showed that the latter were slightly biased in favor of the high-spin state of iron(IV)–oxo

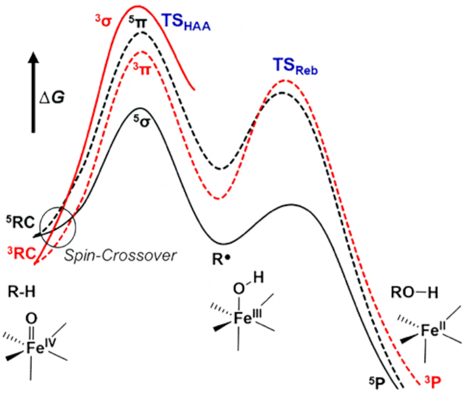


Figure 1.9. Gibbs energy profile for ethane hydroxylation by the $[Fe^{IV}(O)(NH_3)_5]^{2+}$ complex at B3LYP/def2-TZVPP level of theory. Redrawn from ref. [34b].

complexes, thus indicating that DFT may constitute a primary choise to investigate such systems at a reduced computational cost.

1.5.2 Exchange-enhanced reactivity

The concept of exchange-enhanced reactivity (EER) was introduced[47] by Shaik and co-workers to generalize Hund's rule of maximum multiplicity in the context of bioinorganic reactivity and kinetics; in EER, orbital interactions play a complementary role in completing the puzzle. The exchange interaction quantitatively measures the decrease of Pauli repulsion between n electrons of the same spin, such that the total exchange stabilization is ½[n(n-1)]. Therefore, transition states (TS) in which the number of unpaired electrons increases with respect to the reactant complexes will benefit of exchange stabilization, proceeding through low-barrier pathways. Indeed, *en route* to the TS (Fig. 1.8, top left), on the S=2 surface the shift of one electron to the unoccupied $\sigma^*_{z^2}$ orbital ($^5\sigma$-path) creates four new exchange interactions, whereas for S=1 one exchange interaction is lost during the shift of one electron to the half occupied $\pi^*_{xz/yz}$ orbital ($^3\pi$-path); hence, the HAA reaction will be faster for the $^5\sigma$ mechanism. Similar conclusions were drawn by Baerends[33a, 44a] and Solomon,[46b] who highlighted that in the S=2 state the $\sigma^*_{z^2}$ orbital is significantly stabilized by exchange interactions, allowing it to efficiently overlap with the σ_{CH} of the substrate. The EER principle also explains why the $^5\sigma$- is preferred over the alternative $^5\pi$-pathway on the S=2 surface, as in the latter case the transfer of the electron to the $\pi^*_{xz/yz}$ orbital reduces the exchange stabilization from six to three (Fig. 1.8 bottom right). However, it was recently argued that in the TauD-J intermediate these two pathways may exhibit similar activation energies for the HAA.[48]

Noteworthy, the d^4 (S=2) electronic configuration expresses the maximum attainable exchange enhancement, which is in line with Nature's preference for using metalloenzymes containing the $Fe^{IV}(O)$ core for C–H bond cleavage. Alternatively, the most reactive state can be reached through spin crossover from the S=1 ground state to the S=2 excited state, as in the two-state reactivity (TSR) mechanism described by Schröder, Shaik and Schwarz.[49] Essentially, the electronic configuration of the S=1 Fe=O core is $(d_{xy})^2(\pi^*_{xz})^1(\pi_{yz})^1(d_{x^2-y^2})^0(\sigma^*_{z^2})^0$ and the S=2 state is accessible occupying the low-lying $d_{x^2-y^2}$ orbital by (i) a one electron spin-flip forbidden transition from the $(d_{xy})^2$ to the empty $(d_{x^2-y^2})^0$ orbital or (ii) achieving a symmetry where the d_{xy} and $d_{x^2-y^2}$ orbitals are almost degenerate and can be both occupied, as previously seen for complex **4**.

1.5.3 Driving force principle

While the orbital picture and the exchange-enhanced reactivity discussed above represent two complementary explanations of the factors that determine the HAA reaction, the driving force principle constitutes an alternative, irrespective of orbitals and spin states. Based on the seminal work of Bell–Evans–Polanyi (BEP),[50] the driving force principle popularized by Mayer [51] correlates the HAA barrier height with the reaction enthalpy of the process. When comparing two similar reactions, by simply following the Hammond postulate one can assert *a priori* that the more exergonic reaction will exhibit an "early" TS (reactant-like), with lower energy barrier. As an example, in the HAA energy profile of the model system **3** / cyclohexane,[52] (Fig. 1.10) the greater exergonicity of the reaction on the S=2 surface (-14.6 kcal mol^{-1} vs. +1.4 kcal mol^{-1} for the S=1 state) accounts for the lower barrier of the quintet state and fully compensates the 4.6 kcal mol^{-1} energy cost needed to reach the high-spin reactant from the ground state triplet.

Furthermore, Mayer analysed[51] DFT results derived from 59 different examples of HAA ($Fe^{III/IV}(O)$- or $Mn^{V}(O)$-based complexes in combination with different substrates) remarking that in 52 cases the spin state with the greater driving force (ΔE^0) shows the lower energy barrier ($\Delta E^{\ddagger}$) and most frequently corresponds to the higher spin state. Nevertheless, in few cases, the lower barrier can be achieved with lower spin (even the singlet), thus demonstrating that high spin (and non-zero spin) is not necessarily required for HAA. The HAA energy barriers also correlate linearly with the bond dissociation energy of the MO–H bond being formed ($BDE_{O\text{–}H}$)[53] and the transition state MO–H bond distance:[54] the higher the $BDE_{O\text{–}H}$ (the longer the MO–H bond distance in the TS) the lower the HAA energy barrier. Similar structural–activity relationships related to the BEP principle have been applied successfully also to metal-organic framework (MOFs) demonstrating high correlations in predicting the barrier of methane C–H activation from a single parameter (i.e. the M–O formation energy of the MOF active site).[55]

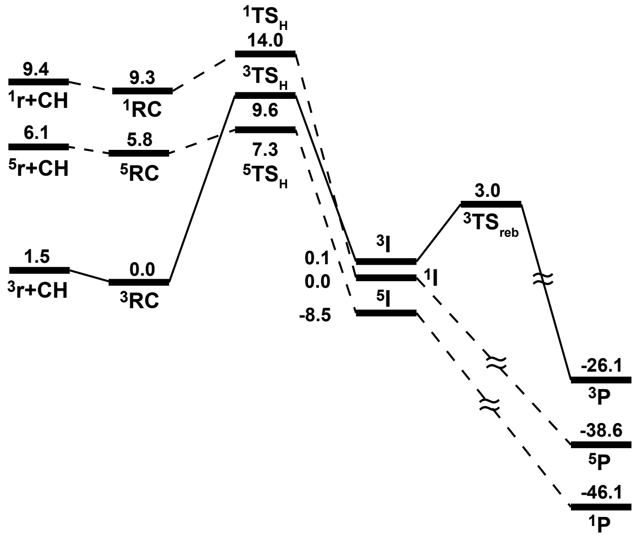


Figure 1.10. Gibbs energy profiles (kcal mol^{-1}) for cyclohexane hydroxylation by complex **3** at the B3LYP/LACV3P++**(Fe);6-311++G**(rest) level of theory. Redrawn from ref. [52].

1.6 Cytochrome P450

C–H hydroxylation is the predominantly occurring reaction mediated by cytochrome P450 (CYP) enzymes. This superfamily‡ of heme-containing monooxygenases are present in all domains of life and use molecular oxygen to hydroxylate an extraordinary broad array of endogenous and exogenous organic compounds as means of drug metabolism, detoxification, hormone regulation and biosynthesis.[1-3, 4e] The ubiquitous metonym "P450" derives from the spectrophotometric peak at the wavelength of the absorption maximum (450 nm) of the enzyme reduced state (CO-bound inactive form). P450 enzymes have a highly conserved tertiary structure and share a characteristic protein fold, but only one amino acid is invariant in the wild-

---

‡ To date, a total of over 300,000 CYPs have been sequenced by different genome projects. Nomenclature has been assigned to more than 41,000 CYPs, 57 of which are human.[56]

type (WT) enzyme: the cysteine residue that tethers the heme prosthetic group to the enzyme.[57]

The catalytic cycle[38d] (Fig. 1.11) starts with the P450 "resting" state (**A**) where a water molecule serves as sixth ligand to the low-spin ferric heme-iron (Fe^{3+}), which is coordinated by the axial cysteine thiolate (S) residue and by four equatorial nitrogen atoms of the porphyrin ring (Por). Substrate (R-H) binding displaces the water ligand inducing a subtle change in the iron position relative to the Por plane, favoring the high-spin state in the penta-coordinated ferric complex (**B**).[58] The transition can be monitored spectrophotometrically from the shift of the Soret band arising from the $\pi \rightarrow \pi^*$ electronic transition in the Por delocalized π system. Concurrently with the spin state shift, a change in the heme-iron redox potential initiates the reduction of the ferric to the ferrous (Fe^{2+}) state through a single ET process originated from NAD(P)H and shuttled to the heme iron *via* redox partner protein(s) (**C**). Next, molecular oxygen binds to the ferrous heme-iron to form a ferric-superoxo complex (**D**), which is reduced by a second ET process that yields the ferric-peroxo intermediate (**E**) with one formally negatively charged distal oxygen and the other negative charge delocalized over the cysteine thiolate ligand. This intermediate is quickly protonated to form the transient ferric-hydroperoxy (Cpd0) species (**F**), whose further protonation triggers water loss *via* heterolytic cleavage of the O–O bond leading to the formation of the porphyrin π radical cation ferryl-oxo species ($Por^{\bullet+}Fe^{IV}{=}O$) (**G**) CpdI. This highly reactive species is considered to be the major oxidant in P450 reaction and abstracts a hydrogen atom from RH to generate a substrate radical prior to "rebounding" the hydroxyl back to $R^{\bullet}$[39] (as discussed in section 1.4) forming the hydroxylated product ROH (**H**), which is released from the active site to restore the resting state of the enzyme. NAD(P)H consumption is not necessarily fully coupled to product formation, as different uncoupling pathways may occur.[59] Among the uncoupling pathways, the peroxide uncoupling that causes peroxide loss from Cpd0 (**F**) can be reverted by high peroxide concentration, in the so-called peroxide shunt pathway, which allows oxidation to be driven using peroxide rather than $NAD(P)H/O_2/H^+$.[38b]

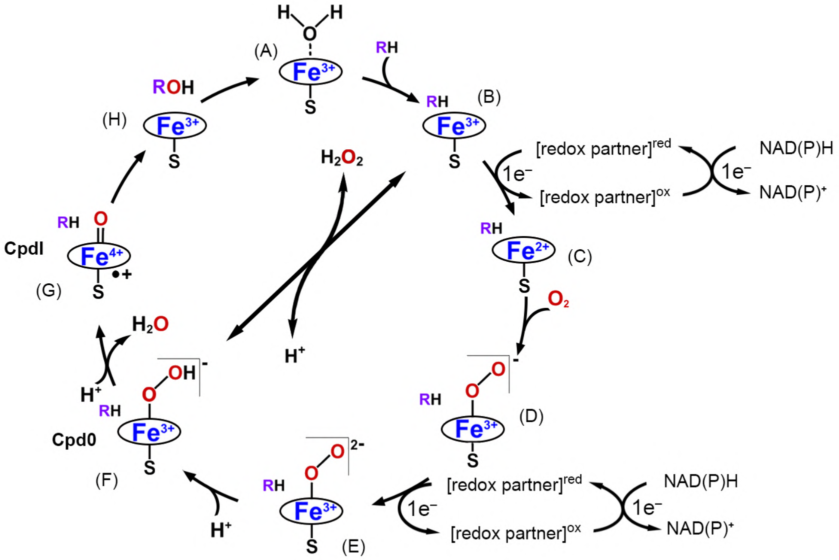


Figure 1.11. The cytochrome P450 catalytic cycle. The double headed arrow crossing the cycle between B and F illustrates the peroxide shut pathway.

Mammalian CYPs are membrane-bound enzymes that metabolize a wide array of substrates such as steroids, fatty acids, drugs, prodrugs, carcinogens, pesticide and herbicides.[60] It is estimated[1a] that just a dozen of these enzymes belonging to the CYP1–3 families metabolize 70-80% of all the drugs in clinical use, with human CYP3A4 alone being responsible of about half of this metabolism. Arguably, mammalian P450s could have tremendous potential applications as biocatalysts for metabolite generation to study drug pharmacokinetics, possible interactions and toxicity, which represent the chief reasons for the failure of drug candidates.[61] However, like any other P450 enzyme, mammalian CYPs require additional redox partners to generate the two electrons required to fuel the catalytic cycle. These redox partners are typically flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) flavoproteins, arranged in the cytochrome P450 reductase (CPR) domain, which in mammalian CYPs is separated from the heme-containing oxygenase domain. Separated reductase and oxygenase domains result in a poorly efficient ET process, causing low turnover rate and loss of catalytic efficiency, thereby making the use of mammalian CYPs impractical for large-scale applications.

In stark contrast, bacterial CYPs are soluble and have a self-sufficient ET system, i.e. the oxygenase domain and the diflavin reductase domain are fused together in a single polypeptide chain[62], resulting in a more efficient ET and higher turnover. Owing to catalytic self-sufficiency, high coupling efficiency and stability, bacterial CYP102A1 from *B. megaterium* (P450$_{BM3}$) is considered the most active P450 identified thus far.[63] Furthermore, it also exhibits high expression levels in common workhorses like *E. coli*, making it an ideal biocatalyst for the development of potential bioprocesses.

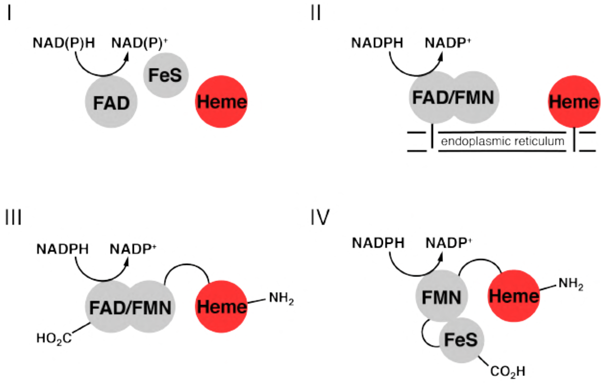


Figure 1.12. Typical topologies of cytochrome P450, all containing the common oxygenase domain (heme), but different redox partners (FAD = flavin adenine dinucleotide; FeS = iron-sulfur ferredoxin; FMN = flavin mononucleotide). Representative of I are P450$_{cam}$ (CYP101); of II human P450 involved in drug metabolism (CYP3A4); of III P450$_{BM3}$ (CYP102A1) and of IV P450 RhF. The latter two are of particular interest, being self-sufficient enzymes.

However, unlike the exquisite substrate promiscuity of human CYP3A4, P450$_{BM3}$ is a substrate-specific enzyme which naturally accepts only fatty acids,[58a, 64] whereas the majority of products that possess pharmaceutical value or (more in general) research interest have precursors that are non-natural substrates for the WT enzyme. Nevertheless, substrate specificity can be radically altered *via* mutagenesis (vide *infra*), leading to significant activity enhancement also towards non-natural substrates.[65] In particular, in P450$_{BM3}$ the introduction of less hindering residues than the naturally occurring Phe87 creates incremental space[66] within the active site in the proximity of the heme unit, allowing the binding of bulkier substrates. On the other hand, given that most of the organic substrates contain multiple oxidation sites, it should be considered that a more spacious active site may consequently allow a greater substrate mobility,[66a, 67] broadening inevitably the product distribution[68] where selectivity control could be desired, or even essential.

A striking example (section 1.7) is represented by BM3 variant F87A, which was reported to successfully accept bulky steroidal substrates, but gave a mixture of hydroxylated products.[69] A second round of mutagenesis based on the F87A template led to $P450_{BM3}$ variants capable of a highly regio- and stereoselective oxidation of the substrate.[69-70]

1.6.1 $P450_{BM3}$

Due to its potential applications as a biocatalyst, cytochrome $P450_{BM3}$ is one of the most investigated CYP enzymes and constitutes a popular platform for enzyme engineering.[38c, 61, 65, 71] Recent experimental and computational studies shed light on the machinery employed by BM3 during its catalytic cycle. The recently solved cryo-EM structure of the full-length enzyme[72] provided a clear picture of how a conformational shift of the dimeric enzyme brings the FMN cofactor in close proximity to the heme, presenting a mechanism by which electrons can be delivered efficiently[38c]. Shaik and co-workers indeed predicted that the ET event in penta-coordinated ferric complex (**B** in the P450 cycle, Fig. 1.11) is dramatically enhanced when the FMN-heme distance becomes shorter and speculated that such a conformational change might be driven by substrate binding.[38d, 73] Similar results were also obtained in a subsequent computational study[74] performed on the first resolved full-length crystal structure of a self-sufficient P450 system[75] ($P450_{TT}$), highlighting that the proximity between the FMN cofactor and the heme unit is crucial for efficient ET.

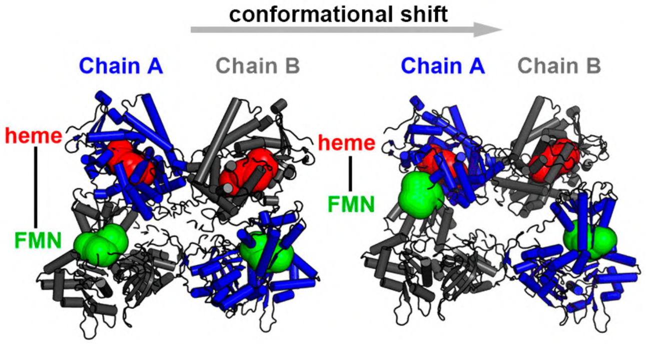


Figure 1.13. Conformational shift of the full-length $P450_{BM3}$ dimeric complex (model derived from cryo-EM maps).

The conformational choreography of BM3 has also been unveiled by computer simulations at different steps of the catalytic cycle, encompassing the gating mechanism of the enzyme during substrate binding in **A**, the O_2 binding in **C** and the two protonation steps in **E** and **F** (Fig. 1.11) which further led to the formation of the H-abstracting CpdI, overall demonstrating how computational studies can go well beyond X-ray data.[38d, 66b, 76]

In addition, the dynamic motion of the enzyme can cause opening/closing and changes in the relative dimension of channels that are essential for substrate access, product relase and the transit/egress of water and oxygen molecules involved in the reaction (Fig. 1.14). $P450_{BM3}$ shares a common pathway (2a) with other CYPs that appears to be the most likely route used by the substrate and product, which passes between the F-G loop, B' helix, B-B' loop, B-C loop and β1 sheet (tunnel in purple in Fig. 1.14).[77] Being dependent on the position of the F-G structural unit that is the most flexible region in all P450s, this 2a channel often merges with subchannels 2b and 2f (blue and red in Fig. 1.14, respectively). A contiguous water channel (W) from the active site to the bulk solvent is accessible in BM3 oxygenase domain also when the CPR domain is present, whereas channel S was computed by Pelletier and co-worker to be energetically most favorable for O_2 diffusion, exhibiting the lowest barrier in the free energy landscape along its pathway.[78]

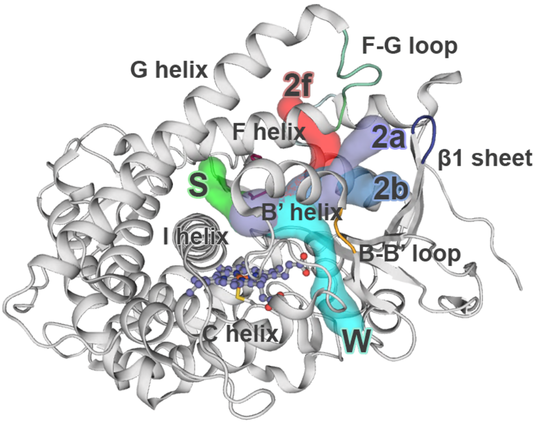


Figure 1.14. Channels mappend in $P450_{BM3}$. The nomenclature is based on ref. [77b]

1.7 Enzyme engineering

As mentioned, a large body of research on cytochrome P450$_{BM3}$ has been conducted to broaden substrate specificity in an effort to engineer a "human-like" P450 activity for generating chemical products of industrial and research interest. However, altering the substrate scope of the enzyme to accept non-natural substrates is only the first step towards converting BM3 to a valuable biocatalyst, since the fine-tuning of chemo-, regio- and stereoselectivity of the enzymatic reaction are other challenges that need to be commonly addressed in later stages of laboratory evolution.[71f]

Enzyme engineering has the primary aim to tackle such difficulties by modifying the original amino acidic sequence of the WT, evolving the enzyme towards desired catalytic traits. Mutations can be introduced in a rational way based on structural knowledge and mechanistic understanding. In practice, this is achieved via site-directed mutagenesis, which targets specific hot-spot residues known to influence the catalytic function, or in a non-rational way throughout iterative rounds of random mutagenesis followed by screening and (artificial) selection of the best variants with improved desired functions (Fig. 1.15). This latter technique called Directed Evolution[71f, 79] (DE) was recognized with the Nobel prize of Chemistry in 2018 reproduces in the lab the natural selection employed by Nature to evolve enzymes, and thus could be seen as the *in vitro* version of Darwinian evolution.[80]

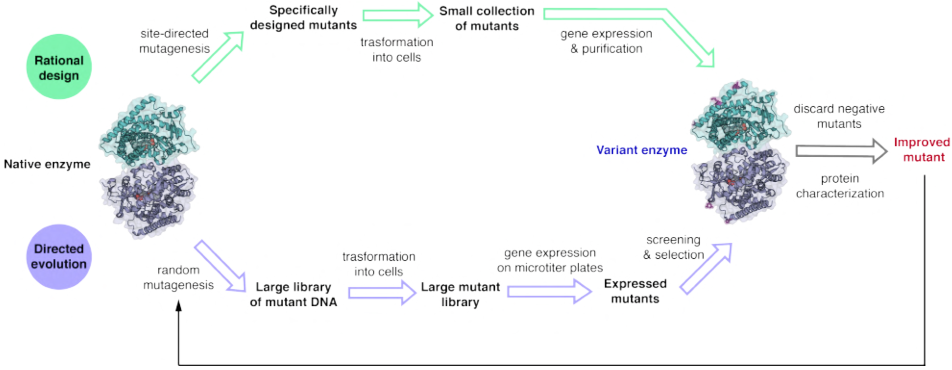


Figure 1.15. Comparison of the processes involved in rational design and directed evolution.

DE constitutes undoubtedly an extraordinary tool in enzyme engineering, but has the disadvantages of requiring initial activity and generating massive mutant libraries, especially when not combined with rational insights. These large libraries need to be tested for desirable properties, making the screening and selection step the bottleneck of the process. High-[71d, 71f] and ultrahigh[81] throughput assays are being developed to help speeding up the process, yet using DE alone has the drawback that only limited information can be extracted on how the induced mutations enhance enzyme proficiency. Due to the huge amount of data produced by DE, during the last few years machine learning (ML) also emerged as a tool to aid and accelerate this type of enzyme engineering.[82] Nevertheless, the molecular basis of any improved function still remain obscure using this approach, at least at its current stage of development.

Rational engineering has the advantage of reducing the number of variants generated, which in turn significantly decreases the screening effort, but it requires a well-characterized structure-activity relationship of the enzyme in order to be applied. As an example, Pande, Arnold and co-workers combined the results from large-scale simulations, high-resolution X-ray crystallography and rational mutagenesis to identify a single position in the F-G loop that controls the loop dynamics and acts as a regioselectivity switch across nitrating P450.[83] A convenient alternative approach consists of combining the advantages of DE and rational mutagenesis to perform semi-rational enzyme engineering, a strategy which has been successfully applied in several cases to create P450$_{BM3}$ variants capable of high regio- and stereoselectivity. For instance, in 2002 the Arnold group isolated, after five generations of random mutagenesis and screening, a BM3 variant (**139-3**) with the highest activity towards non-natural alkane substrates ever reported at the time, notwithstanding giving a rather broad distribution of alkane hydroxylated products.[84] The authors argued that the only active-site substitution V78A (among the other 10 mutations not included in the binding pocket of the **139-3** variant) created incremental space inducing loose binding of substrates. Subsequently, in 2003 the same group used the **139-3** template to perform other DE campaigns obtaining the **9-10A** variant, whose further evolution under random mutagenesis failed to produce improved mutants. The authors turned to site-directed mutagenesis, based on the X-ray structure they focused on decreasing the active site volume to favor substrate binding, conferring regioselectivity. In this way they obtained the **9-10A-A328V** mutant which gave primarily 2-alcohols with good degrees of stereoselectivity.[85]

In another study, Commandeur and co-workers reported the rationally-designed BM3 triple mutant R47L/F87V/L188Q with affinity for steroidal and drug-like compounds not accepted by the WT enzyme.[86] Further random mutagenesis based on the triple mutant generated the **M01** and **M11** variants[87] with activity improvement but still poor regioselectivity, resulting in multiple hydroxylated metabolites.[88] Subsequently, focusing on testosterone as substrate the authors enhanced significantly regio- and stereoselectivity in favor of the 16β-hydroxytestosterone product, by introducing the encumbering A82W mutation in the active-site of both variants *via* site-directed mutagenesis (**M01-W** and **M11-W** variants).[89] Furthermore, it was reported by the same group that in **M01-W** the single active-site mutation S72I (**M01-WI** variant) inverted the stereoselectivity hydroxylating the 16α- side of testosterone, although with a small activity drop.[90]

More recently, Reetz, Munro, Mulholland and co-workers combined exploratory experiments based on mutability landscapes and *in silico* guidance at appropriate points of the enzyme DE to evolve two BM3 mutants highly efficient in hydroxylating testosterone at position 16α and 16β, respectively. The strategy was particularly successful considering the contained screening effort, where both activity and selectivity were targeted at the same time. At first stage, iterative saturation mutagenesis (ISM, where a specific residue is mutated to all the other 19 possible amino acids) was used on a reduced set of residues lining the binding pocket (combinatorial active-site saturation test, CAST[71f, 91]) followed by another round of ISM to obtain the two most promising variants **LIFI-WC** and **WWV-Q**. In the following stage, the authors employed Molecular Dynamics (MD) simulations to gather information at the molecular level and guide mutagenesis focused at specific residues, rather than carrying out other rounds of saturation mutagenesis with inevitable increase of the screening effort. This approach led to further improvement of the variants, with the finally engineered mutants **WIFI-WC** and **WWV-QRS** allowing testosterone hydroxylation at position 16α and 16β with both activity and selectivity above 90%.[92]

As mentioned in section 1.6, the Reetz group found earlier[69] that BM3 variant F87A can accept testosterone, giving an unselective mixture of 2β- and 15β-hydroxylated regioisomers. Applying a second round of mutagenesis on the F87A template led to two highly active mutants **KSA-1** and **KSA-2** with enhanced selectivity for the 2β product. Interestingly, besides the space-creating F87A substitution, variant **KSA-1** contained an active-site hindering A330W mutation, which could control the regioselectivity by tightly binding the substrate, similar to the variants **9-10A-A328V**,[85] **M01-W** and **M11-W**[89] discussed above. Nonetheless, the 2β selective F87A-based variant **KSA-2** featured three mutations (R47I/T49I/Y51I) distal to the active site (>20 Å from the heme ferryl unit) to accomplish the same function, which is intriguing since these two variants seem to solve the same task in different ways.

Table 1.1. Summary of representative P450$_{BM3}$ variants-catalyzed hydroxylations achieved with different enzyme engineering strategies.

Starting enzyme	Mutant	Result	Strategy	Library size
WT BM3	139-3	alkane hydroxyl.	non-rational	10^4
139-3	9-10A-A328V	regiosel.	semi-rational	10^3
WT BM3	R47L/F87V/L188Q	steroidal hydroxyl.	rational	few colonies
R47L/F87V/L188Q	M01, M11	regiosel.	semi-rational	10^3
M01, M11	M01-W, M11-W	regiosel.,stereosel.	rational	few colonies
M01-W	M01-WI	regiosel.,stereosel.	rational	few colonies
WT BM3	WIFI-WC, WWV-QRS	regiosel.,stereosel.	semi-rational	10^3
F87A	KSA-1, KSA-2	regiosel.,stereosel.	non-rational	10^4

Detailed information of mutants: **139-3**
(V78A/H138Y/T175I/V178I/A184V/H36Q/E252G/R255S/A290V/A295T/L353V);
9-10A-A328V (**139-3** + Y138H/I178V/F205C/S226R/A290V–R47C/K94I/P142S–A328V);
M01 (R47L/F87V/L188Q–E267V/G415S); **M11** (**M01** + E64G/F81I/E143G);
M01-W (**M01** + A82W); **M11-W** (**M11** + A82W); **M01-WI** (**M01W** +S72I);
WIFI-WC (R47W/S72I/A82F/F87I–Y51W/L181C);
WWV-QRS (R47W/A82W/F87V–L181Q/T436R/M177S);
KSA-1 (F87A–A330W); **KSA-2** (F87A–R47I/T49I/Y51I)

1.8 Epistasis and fitness landscape

Understanding how to enhance desirable catalytic traits largely depends on the ability to decipher the contribution of mutations to the enzyme function.[93] This endeavor is far from trivial and can be furtherly complicated by the synergism/antagonism between mutations. Introducing a single mutation alone can enhance a given trait, nonetheless two mutations which are beneficial when introduced alone do not necessarily improve the same trait when they are combined.

The effect of a mutation in the presence/absence of another mutation[94] is known as epistasis. Epistasis is invoked in a broader sense[95] when the effect of two (or more) mutations is non-additive compared to the expected fitness value. Fitness here refers to a pursued enzyme function, e.g. activity, selectivity, stability etc. Additivity occurs instead when the sum of the individual effects of mutations A and B is equal to the value in double mutant AB. When the effect on fitness of two mutations is more severe than expected from their effects when alone, it is referred to as synergistic epistasis. The opposite situation, when the fitness difference of the double mutant from the wild type is smaller than expected from the effects of the two single mutations, it is referred to as antagonistic epistasis. Epistatic effects may arise in the form of:

(i) positive magnitude epistasis (+ME) if both the single mutations A and B are beneficial for the fitness and they act in synergism to produce a greater-than-additive fitness improvement (or a smaller-than-additive fitness drop if both are deleterious and they act in antagonism) when combined together in mutant AB;

(ii) negative magnitude epistasis (−ME) if both the single mutations A and B are beneficial for the fitness and they act in antagonism to produce a smaller-than-additive fitness improvement (or a greater-than-additive fitness drop if both are deleterious and they act in synergism) when combined into AB;

(iii) positive sign epistasis (+SE) if one mutation A is deleterious on its own but can enhance the beneficial effect of another mutation B, acting in synergism when combined into AB;

(iv) negative sign epistasis (−SE) if one mutation A is beneficial on its own but can enhance the deleterious effect of another mutation B, acting in antagonism when combined into AB;

(v) positive reciprocal sign epistasis (+RSE) if both mutations A and B are deleterious alone, but they produce a beneficial effect, acting in synergism when combined into AB;

(vi) negative sign epistasis (−RSE) if both are beneficial but they produce a deleterious effect, acting in antagonism when combined into AB.

Fig. 1.16 offers an overview of the different types of epistasis.

Due to pervasive epistatic effects, the result of adding multiple mutations can start varying stochastically upon the fitness, so that understanding the evolutionary process of the enzyme towards improved fitness may become challenging.[96] To tame such complexity, in 1932 Wright introduced in evolutionary biology the idea of fitness landscape,[97] a concept naturally embracing enzyme laboratory evolution.[94, 98] It has lately been abstracted to find application in many other areas like economics[99] or environmental modelling[100] where other kinds of complex adaptive systems are investigated.

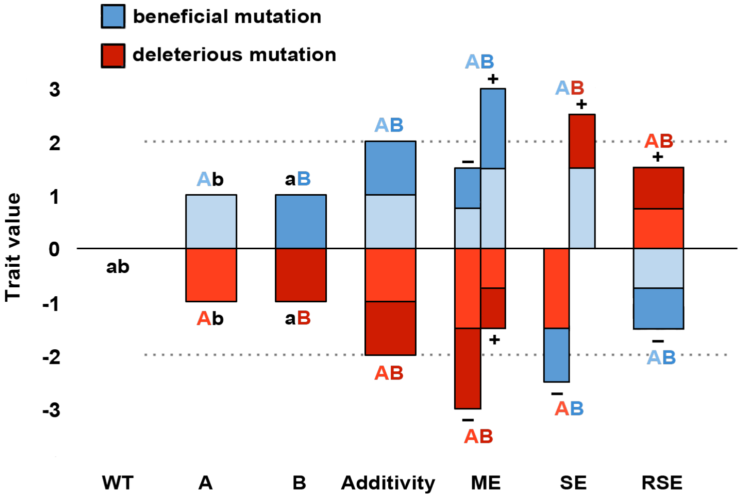


Figure 1.16. Illustration of additive and non-additive (epistatic) effects for the simple case of mutations A and B.

A fitness landscape depicts the shape of the fitness function for a particular objective (i.e. desired enzyme trait) with respect to the sequence space,[101] where the sequence space is the multidimensional representation of all possible enzyme sequences, each connected to its neighbours by edges representing a single amino acid mutation (Fig. 1.17a). Hence, laboratory (and natural) evolution can be envisioned as a trajectory of mutational steps from one functional enzyme to another in the space of all the possible sequences. Assigning the pursued enzyme function (fitness) to each sequence (mutant) yields the fitness landscape. In the absence of epistatic effects, any beneficial mutation results in a large number of possible pathways through the sequence space, all bringing to higher fitness (Fig. 1.17b). Conversely when present, epistasis introduces ruggedness to the fitness landscape, restricting the available trajectories towards optimal fitness by preventing some of them. It nevertheless may also open new paths to functions that would otherwise not have been accessible[102] (Fig. 1.17c).

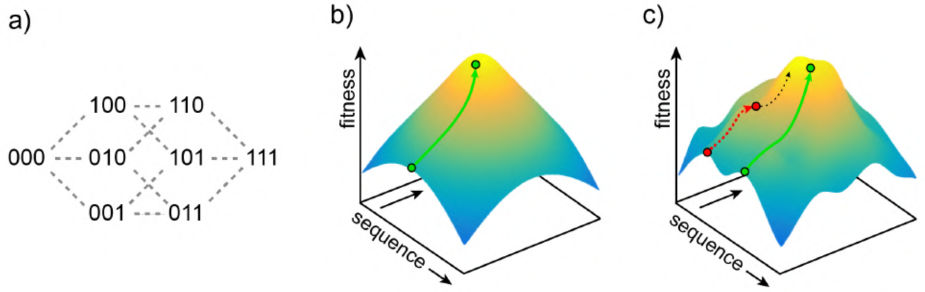


Figure 1.17. Enzyme fitness landscapes. a) Sequence space: mutants are constructed to carry all 2^N possible combinations of the N selected mutations (in this case, N = 3, and 0 and 1 indicate the absence and presence of the mutation, respectively). Each sequence differs from neighbours by edges representing a single amino acid mutation. b) Simple fitness landscape: any beneficial mutation brings the search close to higher fitness. c) Epistasis adds ruggedness to the landscape: along the red path the presence of a local optima might led a mutational step to the wrong direction (evolutionary trap) preventing fitness improvements. Fitness is the measure of how well an enzyme performs the required function.

In other words, epistasis enhances the predictability once evolution has committed to a particular peak,[96, 103] because it reduces *a priori* the combinatorial explosion of the possible pathways accessible during evolution. Contrarily, mapping the fitness landscape *a posteriori* by deconvoluting the best (evolved) mutant from the parent enzyme and analysing all the variants that differ by a single mutation allows to comprehensively characterise the epistatic interactions that affect the evolutionary process. Hence, fitness landscapes are particularly useful, since by dissecting the different steps of mutagenesis, they can operate outside the complexity of the evolutionary paradigm, providing important insights into the relationship between enzyme sequence and function.[104]

1.9 Conformational dynamics and conformational landscape

Nature shows us that evolvability, the ability of an organism to adapt in response to natural selection, is enriched by pre-existing diversity.[105] This obviously concerns proteins and enzymes as well. Yet, it is widely recognised that enzymes need highly preorganized catalytic residues within the active site in order to efficiently stabilize the TS(s) through well-defined interactions (e.g. electrostatic, covalent, acid/base, hydrogen bonding).[106] Enzymes solely conforming to this classical view of ordered and organized structures would however not likely respond to selection pressure, either by natural or laboratory evolution. Besides presenting precisely arranged active site pockets, enzymes are indeed highly flexible and dynamic, undergoing many conformational changes, which range from bond vibrations and side chain rotations at femtoseconds to microseconds timescales, up to slow domain and folding motions at microseconds to seconds timescales (Fig. 1.18).[107]

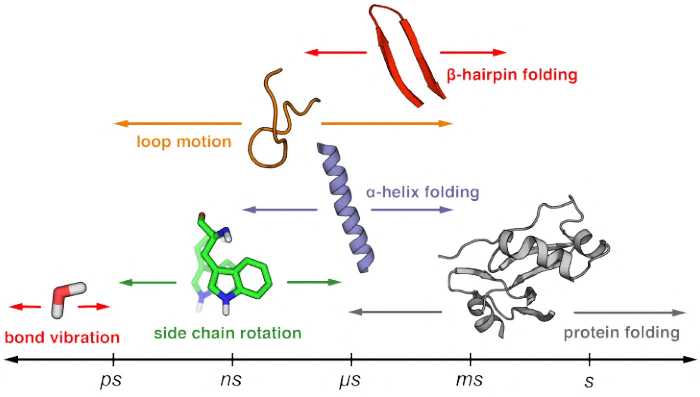


Figure 1.18. Timescales of different enzyme motions.

Chemical reactions in enzymes take place in a short timescale within organized catalytic pockets, while structural motions at larger timescale create a conformational landscape of thermally accessible states, which are explored by the enzyme during the catalytic cycle to accomplish diverse functions spanning from fetching and binding the substrate to release the product. This conformational heterogeneity is believed to constitute the foundation of evolvability.[105, 108] Within this context, conformational shifts are often pivotal for enzymatic catalysis, as discussed for $P450_{BM3}$[38c] in section 1.6.1, and in some cases can be even rate limiting.[107a, 109] Highly debated is whether such changes in conformation are instigated by substrate binding, as hypothesized in the Koshland–Nemethy–Filmer model also known as *induced fit*,[110] or they result from the redistribution of the population of discrete conformational states that already pre-exist, as hypothesized in the Monod–Wyman–Changeux model known as *conformational selection*.[111]

In over five decades multiple evidences followed one another pointing out that, under certain conditions, (i) both scenarios may occur and can be distinguished,[112] (ii) one may prevail over the other[113] or even (ii) they gradually shift towards each other depending on the timescale of the conformational transition.[114] Nonetheless, in the recent years the preference for the conformational selection model reached consensus: whether multiple enzyme conformational states pre-exist as an ensemble, selection pressure stabilizes the productive sub-states over all the unproductive conformations to evolve a specific enzymatic function. This implication has been elegantly demonstrated by Tokuriki and Jackson through an impressive collection of X-ray structures, which showed that the change in function from a phosphotriesterase (PTE) into an arylesterase (AE) enzyme is achieved by gradually populating pre-existing conformational states along the laboratory evolution pathway. More specifically, minor states that conferred some arylesterase activity to the natural PTE enzyme were gradually stabilized by mutagenesis to become major states in the evolved AE.[115] Similar results were obtained by Hilvert and Kern, who demonstrated by NMR spectroscopy, crystallography and stopped-flow fluorescence how directed evolution gradually reshapes the conformational landscape of Kemp eliminase (KE) to populate a narrow, highly active conformational ensemble, achieving a nearly billionfold rate enhancement.[116] Analysis of the evolutionary trajectory revealed indeed how changes in the conformational sampling were critical to gradually supplant the inactive states rooted in the original enzyme scaffold with catalytically competent states, as directed evolution progressed.[116] Interestingly enough, in the evolved AE enzyme mutations that accumulated in the evolutionary trajectories involved also residues remote from the active site, exhibiting a rather small direct catalytic impact but a key role for optimizing and sampling conformational states which were essential for optimal catalysis.[115] Taken together, these results highlight the fundamental importance of conformational dynamics and the key role of mutations (including distal) in the population shift of the enzyme conformational states .

Notwithstanding the significance and the biological implications of the conformational selection model, particular attention should be taken to not overlook the importance of substrate binding for triggering specific events, which are essential for the catalytic cycle of certain enzyme classes. It is worth recalling that upon substrate binding, cytochrome P450 undergoes a conformational shift with concomitant change in the spin-state and reduction potential of iron to initiate the ET process from the reductase partner (section 1.6). In 2004, Munro and Leys reported the crystal structure of a substrate-free $P450_{BM3}$ A264E mutant exhibiting a conformation previously observed only in substrate-bound WT enzyme, but with the heme iron in the low-spin state.[117] This provides an evidence that although mutations can induce a conformational shift from the substrate-free to the substrate-bound enzyme form, the complete change in the spin state that P450 required for catalysis is not only determined by structural changes, but may depend on the presence of the substrate.[58b]

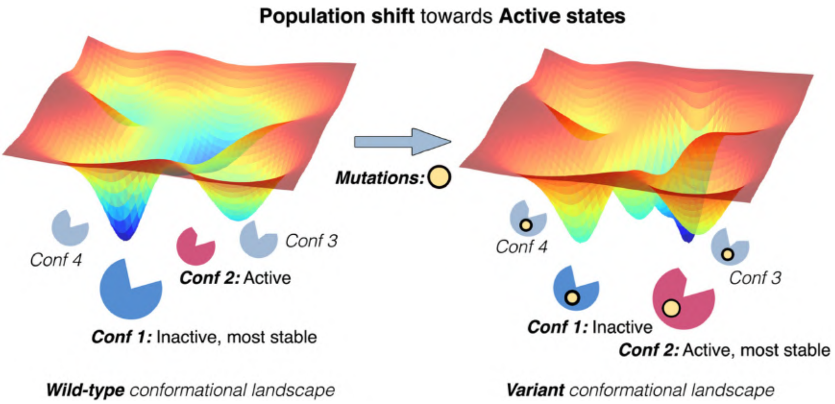


Figure 1.19. Representation of population shift induced by mutations. The conformational landscape of the wild-type enzyme (left) and evolved variant (right) differ in the relative stabilities of their conformational states, being stabilized due to mutations in the variant those that are functional for enzyme activity.

Chapter 2: *In silico* methods

As shown in the previous chapter, significant insights into diverse (bio)chemical systems, such as enzymes and metal complexes inspired by their reactive intermediates, can be obtained by carrying out computer simulations. Computational chemistry offers the opportunity to model and understand chemical systems of different varying molecular complexity at various levels of accuracy. Quantum mechanics (QM) methods are the most accurate, accounting for electronic effects and providing high resolution up to sub-atomistic level. Among QM methods, DFT is currently the most extensively used methodology[118] to study molecular systems; nevertheless, due to the computational cost its applicability is often restricted to a few hundreads of atoms. On the other hand, molecular mechanics (MM) approaches offer an alternative when a balance between accuracy and system size has to be considered. In MM, quantum effects are neglected while, in exchange, the size and feasibility of the computational model increase allowing the modeling of whole enzymes in explicit solvent. Multilevel approaches (i.e. QM/MM[119]) also exist, which consist of the treatment of a small portion of the chemical system (e.g. the enzyme active site) using QM, and the rest of the system is described using the less rigorous but computationally more efficient MM. In the case of biocatalysis, QM/MM methods allow the study of chemical reactivity by explicitly considering the effects of the protein environment and solvent molecules. QM/MM calculations are tipically focused on representative snapshots of the enzyme conformational states. However, as previously discussed in section 1.9, enzymes can undergo to large-scale structural motions exploring multiple conformational states, thereby limiting the generality and the applicability of the QM/MM protocol due to the large amount of calculations needed to account for each relevant state of the enzyme.

The Houk group has developed a simpler standard approach to understanding the mechanisms of (metallo)enzymes (particularly P450s) with great success.[120] This protocol consists of performing DFT calculations to study the transition states and mechanisms of a *theozyme*[121] (theoretical enzyme) model, comprising a truncated portion of the enzyme that includes catalytically relevant active site residues and cofactors along with the substrate.

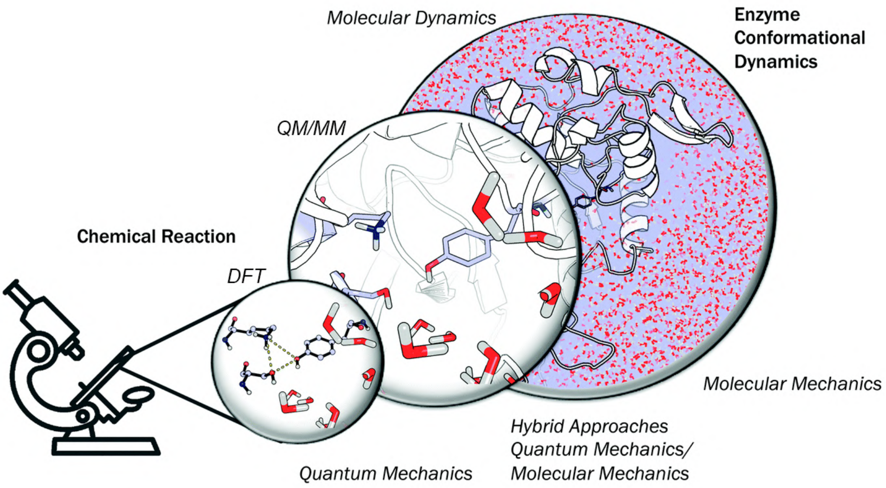


Figure 2.1. Representation of different level of accuracy in computational chemistry.

Moreover, the DFT calculations are coupled with molecular dynamics (MD) simulations, which allow us to study the substrate binding poses relative to the catalytic species during the enzyme conformational dynamics. Finally, the geometries resulting from MD simulations are compared to the ideal TS geometries derived from DFT to establish the enzyme's selectivity control.

2.1 Density Functional Theory

DFT is, in principle, an exact QM theory of the electronic ground-state structure, formulated in terms of the electronic density distribution $\rho(r)$ and based on the rigorous, simple lemma of Hohenberg and Kohn:[122] the specification of the ground-state density $\rho(r)$ determines the external potential $v(r)$ uniquely.

$$\rho(r) \Leftrightarrow v(r) \tag{2.1}$$

Since the integration of $\rho(r)$ yields the total number of N electrons of the system exactly, it is consequently sufficient to construct the full Hamiltonian $\hat{H}$ and then, implicitly, to obtain all the properties defined by $\hat{H}$. Thus, within this theory the expectation value of any ground state observable can be written as a functional of the ground state electron density $\rho(r)$. In particular, the electronic energy is written as follows:

$$E[\rho] = F_{HK}[\rho] + \int v_{ext}(r)\, \rho(r)\, dr \tag{2.2}$$

where v_{ext} is the external potential representing the electron-nuclei attraction and F_{HK} denotes the Hohenberg and Kohn functional for the electron density, which depends on the kinetic energy $T[\rho]$, the Coulombic ($J[\rho]$) and non-classical ($E_{NC}[\rho]$) electron-electron interaction energy, while the repulsion between nuclei is assumed constant within the Born-Oppenheimer approximation.

$$F_{HK}[\rho] = T[\rho] + J[\rho] + E_{NC}[\rho] \tag{2.3}$$

However, due to the complex nature of the many-electrons problem, the exact expression of $F_{HK}[\rho]$ connecting the energy of the system and $\rho(r)$ remains unknown. Therefore, suitable approximations have to be found.

In their original work,[123] Kohn and Sham proposed that for every ground state density, there exists an auxiliary non-interacting system described by an orbital Slater determinant Ψ_0, with the same ground state density, which conveniently mimics the true, many-electron system. In this way, the energy $E[\rho]$ can be divided into four parts:

$$E[\rho] = T_0[\rho] + \int v_{ext}(r)\rho(r)\, dr + J[\rho] + E_{XC}[\rho] \tag{2.4}$$

where,

- $T_0[\rho]$ is the kinetic energy of non-interacting electrons
- $E_{XC}[\rho]$ is the exchange-correlation energy, which includes the non-classical electronic interaction energy $E_{NC}[\rho]$ and the correction term for the kinetic energy $T[\rho]$ of the fully interacting system (Eq. 2.3)

The electronic density can be represented as a set of occupied one-electron orthonormal orbitals $\{\psi_i\}$. A functional of such orbitals can be defined as:

$$\Phi[\psi_i] = E[\rho] - \sum_i \sum_j \varepsilon_{ij} \int \psi_i^*(r)\psi_j\, dr \tag{2.5}$$

ε_{ij} are Lagrangian multipliers to ensure that the orbitals are orthonormal. The minimization of $\Phi[\psi_i]$ with respect to ψ_i^* finally gives the so-called Kohn-Sham (KS) equations:

$$\left[-\frac{1}{2}\nabla_i^2 + v_{eff}(r)\right]\psi_i(r) = \varepsilon_\iota\psi_i(r) \tag{2.6}$$

where the one-electron KS orbitals ψ_i determine the exact density *via*

$$\rho(r) = \sum_{i=1}^{N}|\psi_i(r)|^2 \tag{2.7}$$

and v_{eff} is the effective potential that gives the exact ground state density $\rho(r)$:

$$v_{eff}(r) = \ v_{ext}(r) + \ v_H + v_{xc}(r) \tag{2.8}$$

with v_H Hartree potential:

$$v_H = \int \frac{\rho(r')}{|r - r'|}dr'$$

and v_{xc} exchange-correlation potential:

$$v_{xc}(r) = \frac{\delta E_{XC}}{\delta\rho(r)} \tag{2.9}$$

The one-electron KS orbital (Eq. 2.7) equations must be solved self-consistently. Despite their single particle form, they are formally exact equations, hence they provide the exact ground-state density and energy. Nevertheless, E_{XC} remains unknown, and several density functional approximations (DFAs) have been proposed and are routinely used. DFT studies have shown intrinsic spin-state preferences upon the approximation employed.[124] This led to to a class of specially designed DFAs (e.g. S12g by Swart[125]) for the appropriate description of the complicated spin state preferences in transition metals.

2.1.1 Exchange-correlation approximation

The simplest but surprisingly serviceable DFA for E_{XC} consists of the *local density approximation* (LDA), which assumes that although molecules possess nonhomogeneous electron densities, they can be treated locally as a uniform electron gas. Using LDA, the exchange-correlation energy can be expressed as:

$$E_{XC}^{LDA} = \int \rho(r)\, e_{xc}^{unif}(r)\, dr \tag{2.10}$$

where e_{xc}^{unif} stands for the exchange-correlation energy per particle of an uniform gas of density ρ. However, in most practical cases the density of a molecule is not constant, and the LDA approximation produces considerable errors. To improve the density calculation, a non-uniform electron gas should be considered. A convenient way to achieve this, consists of making the exchange and the correlation energies dependent not only on the electron density, but also on its derivatives. This method is known as *generalized gradient approximation* (GGA), and can be described as follows:

$$E_{XC}^{GGA} = \int \rho(r)\, e_{xc}^{unif} F_{XC}[\rho, \nabla\rho]\, dr \tag{2.11}$$

where $\nabla\rho$ is the gradient of the density and F_{XC} is dimensionless.

The exchange-correlation energy and the corresponding potential can be connected by linking the non-interacting KS system with the true interacting system. The resulting equation is known as the *adiabatic connection formula* (ACF):[126]

$$E_{XC} = \int_0^1 E_{XC}^{\lambda}\, d\lambda \tag{2.12}$$

$\lambda=0$ represents the non-interacting KS reference system and $\lambda=1$ the true interacting system. Eq 2.12 can be approximated with a parametrical equation, where the parameters are semi-empirical and are found by fitting the heat of formation of a standard set of molecules. One popular example is the 3-parameter equation used by the B3LYP[127] functional.

$$E_{XC}^{B3LYP} = (1-a)E_X^{LDA} + aE_X^{HF} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LDA} \tag{2.13}$$

where,

- E_X^{LDA} is the local density approximation to the exchange energy
- E_X^{HF} is the exact Hartree-Fock exchange energy

- E_X^{B88} is the GGA part of Becke (B88)[128] exchange energy
- E_C^{LYP} is the Lee-Yang-Parr (LYP)[129] correlation energy
- E_C^{LDA} is the local density approximation to the correlation energy
- $a = 0.20, b = 0.72, c = 0.81$

2.1.2 Relativistic DFT

The Einstein's special theory of relativity[130] is based on two postulates:

- The physics laws should be the same in all inertial frames
- The velocity of light c in vacuum is the same in all the inertial frames

As a direct consequence of these two postulates, the Schrödinger equation is no longer valid when the speed of a particle is a substantial fraction of the speed of light and that is the case of the inner electrons of heavy elements. The unification of the Einstein's special theory of relativity with Schrödinger theory of quantum mechanics was accomplished by Dirac[131], using the time-independent Dirac equation:

$$\hat{H}^D \psi^D = E\psi^D \tag{2.14}$$

where $\hat{H}^D$ is the Dirac Hamiltonian and the eigenfunction ψ^D is a four-component vector containing two "large" components, which pass to the corresponding relativistic wave functions in the limit $c \to \infty$, and two "small" components, which vanish in the nonrelativistic limit. Nonetheless, the Dirac equation is not very often used in the original four-component form because of its computational cost. It is therefore convenient to reduce the four-components Dirac equation to a two-components form, ideally without a significant loss of accuracy. This can be achieved with an approximate transformation into the two-component (scaled) Zeroth Order Regular Approximation (ZORA) equation, initially proposed by Chang *et al.*[132] and Heully *et al.*[133] and generalized by Lenthe, Snijders and Baerends.[134] In the ZORA method, the large and small components are decoupled using a Foldy–Wouthuysen transformation[135] and subsequently expanded in $E(2mc^2 - V)^{-1}$, with the resulting equation to zeroth order representing the ZORA equation. The ZORA equation is a two-component method if spin-orbit coupling is included, while if only scalar-relativistic effects are taken into account, it is reduced to one-component with the Hamiltonian essentially as expensive to use as its non-relativistic counterpart.

It should be noted that for many chemical properties only the valence electrons are of importance, and the electron density near the nucleus does not change much on molecule formation. This insight is used in the *frozen core* (FC) *approximation* (*vide infra*) and in the treatment of relativistic effects using the so-called *effective core potential* (ECP) method.[136] In the ECP method the core electrons are not taken into account explicitly in the Hamiltonian, but their effect is modeled by a set of parameters whose values are based on relativistic atomic reference calculations. This method is a convenient approximation to treat relativistic effects since the number of electrons (of heavy elements) is drastically reduced and besides the extra parameters, which only affect the one-electron part of the Hamiltonian, the non-relativistic Hamiltionian can be used. The ECP method may be a useful method in geometry optimizations, calculation of electric properties like dipole moments, polarizabilities, etc. However, for properties heavily determined by the shape of the wavefunction in the vicinity of the nucleus, like NMR parameters, the use of ECPs is unsuitable.

A detailed descriptio of relativistic DFT can be found in ref. [137]

2.1.3 Basis set

In DFT, the molecular orbitals (MOs) used to solve the KS equations (Eq. 2.6) are expanded in terms of linear combination of atomic orbitals (LCAO). The set of basis functions used to describe atomic orbitals (AOs) are typically of Slater-type orbitals (STOs) or Gaussian-type orbitals (GTOs),

$$\varphi^{STO} = N\, r^{n-1} e^{-\zeta r}\, Y_{l m_l} \tag{2.15}$$

$$\varphi^{GTO} = N\, r^{2n-2-l} e^{-\zeta r^2}\, Y_{l m_l} \tag{2.16}$$

where,

- N is a normalization constant
- n is a natural number that plays the role of the *principal quantum number*
- l is the *orbital angular momentum quantum number*
- r is the distance of the electron from the atomic nucleus

- ζ is a variational parameter of the radial function that fixes the extent of diffusion of the function
- Y_{lm_l} are spherical harmonics functions (with m_l magnetic component of l)

Despite the only difference between STOs and GTOs is the quadratic dependence with r in the exponential term, this has dramatic implications for the behavior of each type of orbital. STOs are hydrogen-like AOs and suitable for expanding MOs because they reproduce the correct electronic behavior both close the nucleus ($r \rightarrow 0$ region) and far away (asymptotic long-range behavior $e^{-\zeta r}$). On the other hand, GTOs enable a more facile calculation of molecular multi-center integrals by an analytical procedure, nevertheless a linear combination of at least three GTOs is required to roughly model a single STO[138] (Fig. 2.2).

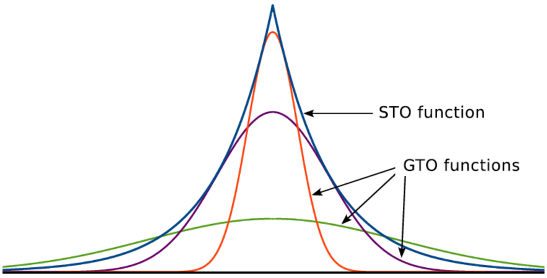


Figure 2.2. Approximation of a Slater-type orbital with three Gaussian-type orbital.

Depending on the number of functions used to describe each AO, the basis set can be classified in terms of its quality. Usually, more functions are used to describe the valence shell orbitals, which contribute the most to chemical reactivity. The exponential factor ζ denotes the number of functions used to represent valence orbitals with respect to the minimal basis set. Hence, in case the double(triple…nth) of the functions is used, it will be referred as double(triple..nth)-ζ basis set. In addition, functions with higher angular momentum (polarization functions) or with small ζ exponent (diffuse functions) can be added to further improve the description of inhomogeneously charged atoms (e.g. electronegative) and the flexibility of the valence orbitals to account for atoms possessing loosely bound electrons (e.g. anions, metals), respectively. The notation of STOs typically takes the first letter of the basis set quality and adds a "P" whether polarization function are added, therefore DZP will denote a double-ζ basis set with one added polarization functions, TZ2P triple-ζ plus two polarization functions, etc. Otherwise, for GTOs in the Pople-style, the first number refers to the number of Gaussians (or primitives) that describe the basis functions associated to inner core electrons, and the following two (or three) numbers indicate how many Gaussian (primitives) are in each valence orbitals. Additionally, a single (or double) "*" is added to the notation when polarization functions are included for heavy (or heavy and H atoms), whereas single (or double) "+" denotes the inclusion of diffuse functions for heavy (or heavy and H atoms), respectively. Therefore, a 6-311*+ indicates 6 primitives describing each basis set function for core electrons, a triple split valence shell with 3,1,1 primitives for each of the valence basis functions plus polarization and diffuse function added on heavy elements.

Core electrons are generally less important than valence shell electron, which contribute to chemistry. For systems involving heavy elements (generally from the third row) the computational cost to treat core electrons can be reduced using the FC approximation, in which core electrons are first treated with accurate single-atom calculations using large STOs and then explicitly orthogonalized against valence orbitals. The resulting frozen core density in subsequently included in the molecular calculation.

2.1.4 Dispersion correction

Weak inter- and intramolecular forces due to the interaction between temporary induced multipoles in molecules which do not have permanent dipole moment (London dispersion forces) or between specific permanent dipoles (hydrogen bond) are usually underestimated by DFT, and the resulting interaction energy turns smaller than the real one. For this reason, some DFT functionals already contain a built-in dispersion correction (e.g. S12g[125]). For all the other functionals, a correction term can be added to the standard KS-DFT energy (eq. 2.4) to calculate the total energy, as proposed by Grimme[139]:

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \tag{2.17}$$

were E_{KS-DFT} is the self-consistent Kohn-Sham energy and E_{disp} is the dispersion correction term.

2.1.5 Charge displacement function analysis

A theoretical bonding analysis is of prime importance for the deep understanding of the various chemical interactions, covalent or not. The electron density changes occurring upon bond formation between two fragments A and B to form the AB complex can be analyzed through the charge-displacement function (CDF):[140]

$$\Delta q(z) = \int_{-\infty}^{z} dz' \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta\rho(x, y, z')\, dx\, dy \tag{2.18}$$

In this equation $\Delta\rho(x, y, z')$ is the electron density difference between the density of the AB complex and that of the two noninteracting fragments A and B in the geometries they occupy in the overall complex. A partial progressive integration of $\Delta\rho(x, y, z')$ along a suitable chosen axis yields the charge-displacement function $\Delta q(z)$ (eq. 2.18). This function measures at each point z along the selected axis (typically that joining A and B fragments) the amount of electrons that, upon bond formation, moves across a plane perpendicular to this axis passing through z. Positive (negative) values of $\Delta q(z)$ correspond to electrons flowing in the direction of decreasing (increasing) z, while a positive (negative) slope indicates regions of charge accumulation (depletion). The analysis of the CDF can thus provide a thorough, spatially detailed picture of the total charge flow occurring upon any change of the A–B bond analysed. A quantitative estimation of such change can be derived from the amount of charge flowing across a plane perpendicular to the integration axis and placed at any given z value (Fig. 2.3).

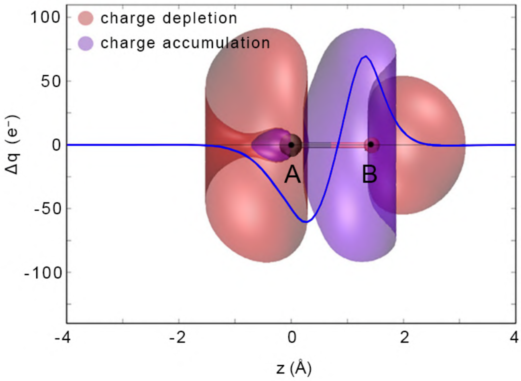


Figure 2.3. Representation of the CDF for a generic AB complex. Substantial charge accumulation is found in the intramolecular region, while on the A atom and the outer right region of the B atom charge depletion is observable.

2.2 Solvation models

Chemical and biological reactions usually take place in presence of solvent. Consequently, the solvent environment influences the molecular properties and therefore requires an appropriate description to account for the solvation free energy. Based on their nature, solvation effects can be classified as *non-specific*, comprising the long-range effects due to the presence of the solute such as solvent polarization and solvent-solvent electric multipole moments, and *specific*, which account for short-range solvent-solute interactions such as van der Waals interactions, hydrogen bonds, solvent-solute dynamics and charge transfer, hydrophobic and entropic effects.[141]

Depending on the strategy used to treat the energetic contribution of the solvent, solvation models can be divided into: *(i) explicit models*, including discrete solvent molecules and all the interaction between them and *(ii) continuum models*, describing solvent molecules implicitly by placing the solute into a cavity and considering the rest of the environment by its average effect, i.e. as a *dielectric continuum*.

2.2.1 Explicit models

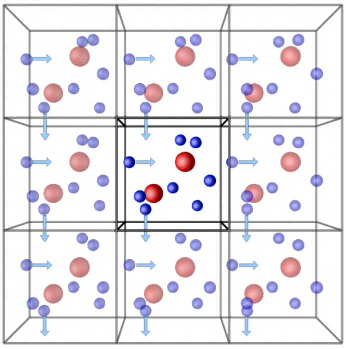


Figure 2.4. Periodic boundary condition representation.

Explicit solvation models provide an accurate, spatially resolved physical representation of a solute embedded into a solvent medium to describe all the corresponding solute-solvent interactions. Although under certain circumstances including the solvent explicitly is required in QM methods (if it participates in the chemical reaction, by e.g. accepting or donating a proton), depending upon the number of solvent molecule to take into account the use of explicit solvent could become computational demanding and only feasible using MM approaches, where up to millions of molecules can be treated at a reasonable computational cost. Since biological reaction take place in aqueous condition, simple rigid water models like the three-point charge TIP3P model[142] have been developed to reduce the degrees of freedom while evaluating water's contributions to energy and to describe the liquid phase properties. The solvated system is typically placed at the center of a water box which is repeated in all directions, thereby constituting a quasi-periodic system whose periodicity is equal to the box dimension (*periodic boundary conditions*, Fig. 2.4). Despite that the rigid model approximation only relies on non-bonded interactions, their calculation can be demanding when dealing with a large number of solvent molecules. However, this effort is not always required since some non-bonded interaction like van der Waals forces rapidly decay with distance, hence a cut-off distance can be defined for taking into account only atoms lying inside an interaction sphere. Conversely, Coulombic electrostatic interaction are long-ranged and would result unjustly represented by a cut-off approximation, hence they are often splitted into "near-field" and a "far-field" potential contributions (*Ewald sum method*[143]).

2.2.2 Continuum models

The preferred method in QM to account for solvent effect is the use of *dielectric continuum* models. This class of methods is approximated by embedding the solute S into a cavity of a dielectric continuum with dielectric constant ε, which defines the degree of polarizability of the solvent (Fig. 2.5). The charge distribution of the solute (S) polarizes the dielectric medium on the surface cavity, which consequently responds and influences the solute charge distribution (polarization response of the solute in presence of the solvent), that is ultimately calculated iteratively with *ab initio* methods until getting self-consistency and is determined by the amount of perturbation that the solvent exerts on the QM Hamiltonian:

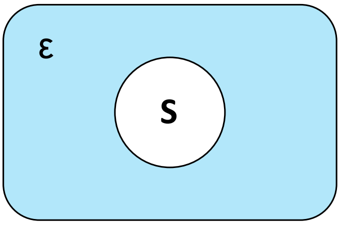


Figure 2.5. Continuum solvation model.

$$\hat{H}^{tot}(r_s) = \hat{H}^{mol}(r_s) + V^{mol+solv}(r_s) \tag{2.19}$$

where $V^{mol+solv}$ is the term including the interaction operators required to describe the system response resulting from placing the solute from the gas phase into the continuum medium. This approach called *Self-Consistent Reaction Field* (SCRF) comprehends different methods which vary on how (i) size and shape of the cavity are defined, (ii) dispersion contributions are calculated, (iii) the charge distribution of S is represented, (iv) the solute S is treated with semiempirical or *ab initio* methods and (v) the dieletric medium is described. For instance, within the *Polarizable Continuum Model*[144] (PCM) a van der Waals surface type cavity is used and the continuum polarizability is taken into account with the exact dielectric, whereas the *COnductor-like Screening MOdel*[145] uses a scaled-conductor approximation to derive the continuum polarization charges and usually employes *Solvent Excluding Surfaces* (SES) to approximate the cavities.

2.3 Molecular Mechanics

Biological macromolecules like proteins and enzymes contain from thousand up to millions of atoms, so that their description based on full QM methods is hardly computationally feasible thus far. On the other hand, molecular mechanics (MM) approaches do not consider electrons explicitly and treat atoms as classical particles subjected uniquely to Newton's second law, making inevitably quantum features (e.g. bond breaking and formation) neglected, yet simplifying computational calculations sufficiently to allow the simulation of large systems. In MM molecules are represented by the "ball and spring" model, with each atom taken as charged sphere of certain mass and radius, and its covalent bonds the springs. The interactions between atoms is then determined by the potential energy, formulated as a sum of different terms which represent the total energy required to distort a molecule in specific manners, and that can be grouped in *bonded* and *non-bonded* terms (Fig. 2.6).

2.3.1 Force fields

The functional form of all the different *bonded* and *non-bonded* terms of the potential energy represents the so-called force field (FF):

$$U_{FF} = U_{str} + U_{bend} + U_{torsion} + U_{VdW} + U_{el} \tag{2.20}$$

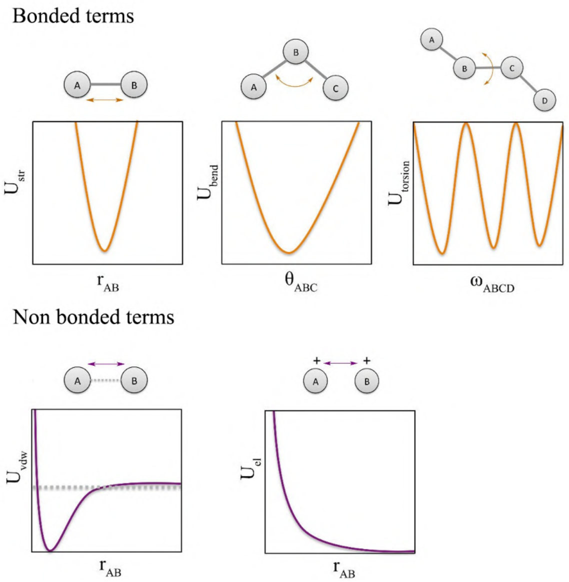


Figure 2.6. Bonded (top) and non-bonded (bottom) terms for the potential energy of the "ball and spring" model.

The *bonded* terms account for connected atoms, including (i) the energy function U_{str} for stretching a bond between the A-B atoms

$$U_{str} = \sum_{bonds} k_{AB}(r_{AB} - r_{AB,eq})^2 \tag{2.21}$$

(ii) the energy function U_{bend} for bending and angle formed by the A-B-C atoms

$$U_{bend} = \sum_{angles} k_{ABC}(\theta_{ABC} - \theta_{ABC,eq})^2 \tag{2.22}$$

where both functions related to bonds and angles usually follow a harmonic approximation in which the energy varies with the square of the displacement from the equilibrium values ($r_{AB,eq}$ and $\theta_{AB,eq}$, respectively) and k_{AB}, k_{ABC} are the corresponding force constants of the bond and bend angle; (iii) the torsion energy $U_{torsion}$ to describe the dihedral rotation

$$U_{torsion} = \sum_{dihedrals} k_{ABCD}\,[1 + \cos(n\omega - \delta)] \tag{2.23}$$

where k_{ABCD} is the torsional force constant, n the periodicity of the cos function, ω the dihedral angle formed by the A-B-C-D atoms and δ its phase. Often, an out-of-plane bending energy term U_{oop} is also included in the FF to account for the energy penalty associated with the pyramidalization of planar sp^2 atoms (improper torsion).

The *non-bonded* terms describe "through space" interactions between atoms that are not directly connected through *bonded* interactions. They include (iv) van der Waals interactions U_{VdW} usually described as a Lennard-Jones (L-J) potential

$$U_{VdW} = \sum_{A}\sum_{B>A} 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right] \tag{2.24}$$

with r_{AB} the distance between the interacting atoms, $\varepsilon_{AB}$the L-J well depth and σ_{AB} the distance at which the interaction becomes zero; and (v) electrostatic interaction U_{el} described as Coulomb potential between A and B atoms with partial charges q_A and q_B, interatomic distance r_{AB} and ε_0 dielectric constant

$$U_{el} = \sum_{A}\sum_{B>A} \frac{q_A q_B}{4\pi\varepsilon_0 r_{AB}} \tag{2.25}$$

FFs including the described terms are the simplest (class I) and typically designed to treat large systems. Increased accuracy is achied using higher-order functions and a cross-energy term (U_{cross}) to account for coupling between *bonded* terms (class II) or using FFs which also describe quantum effects such as electronic polarization (class III), where atom charges are not fixed during the simulation so that the charge distribution can vary in response to the environment.

Because all FFs terms extensively use force constants, reference bond lengths, angles and torsions, partial charges and L-J potentials, it is clear that the accuracy of these parameters is crucial to correctly reproduce the properties of the systems that are to be modeled by MM. Depending upon the specific FF, parameters may be derived from a variety of empirical procedures comprising spectroscopy, X-ray crystallographic structures, gas- and solution-phase properties, or alternatively obtained *ab initio* from QM calculations. All these procedures are settled in an iterative approach, where several rounds of optimization and refinement are performed until the final model is achieved. Recently, ML has emerged as a promising approach for fitting QM data and generate bespoke FFs from active-learning-based potentials.[146] It should be noted that reference data used for force field fitting (either empirical or *ab initio*) are collected from systems that are not necessarily identical to that simulated, i.e. considering small molecular fragments (aminoacids/peptides) to simulate full proteins. Therefore, FF parameters must be transferable across a wide range of compounds to encompass the extent of structural and chemical diversity. In this context, many FFs are optimized for nucleic acids to simulate correctly protein environments; hence, to obtain the parameters for organic and drug-like molecules often binding proteins active site, generic FFs as the Generalized AMBER Force Field[147] (GAFF) have been developed that incorporate both empirical and heuristic models to cover a broad chemical space. The general protocol to parametrise organic compounds consists of a single point energy (SPE) calculation on top of the optimized geometry to estimate atom charges with the *restrained electrostatic potential* (RESP) method,[148] which generates atom-centered point charges based on a charge fitting procedure to an electrostatic potential generated at the HF/6-31G* level of theory. The RESP charges and atom connectivity are then used to search within GAFF for those force constants matching the atoms involved. However, generic FFs applicable to biological systems do not include parameters to describe metal centers or metal-based cofactor present in metalloenzymes. In this regard, a commonly employed strategy is the Seminario method,[149] which consists of a QM geometry optimization of the metal and the coordinated moiety followed by a frequency calculation to obtain the force constants for bonds, angles, dihedrals and improper torsions from the Hessian matrix. Covalent bonds between the metal and its coordination sphere permit to account for charge transfer, whose values are calculated typically with the RESP[148] approach. Notwithstanding, the main limitation of this approach is that by including predefined bonds, the metal coordination number cannot change during the simulation, therefore any oxidation state and coordination geometry has to be parametrised and studied independently. On the other hand, a convenient alternative is the cationic dummy model[150] (CDM), where the metal center is described within a non-bonded approach by a set of cationic dummy atoms connected around a central atom in its specific coordination, making it a transferable model that can be used when changes in the metal coordination occur without further parametrization.

2.3.2 Molecular docking

Molecular docking is a structure-based method to predict the preferred orientation of a ligand molecule (e.g. an enzyme substrate) bount to a receptor (e.g. the enzyme active site).[151] The preferred orientation can be used to assess the binding affinity between the ligand and the receptor using a *scoring function*. Docking, therefore, can be envisioned as the computational simulations of an enzyme molecular recognition process.

In a docking simulation, the conformational sampling of the ligand is performed within the receptor. This conformational search incorporates changes to the ligand structure as torsion angles, translations and rotations of the entire molecule. Each transformation in the ligand conformational space induces a total energetic cost of the system, and the system's total energy is calculated after every change. In this way, ligand flexibility is incorporated within the docking process by selecting energetically reasonable conformations in the presence of the receptor by FF energy evaluations. Additionally, receptor flexibility can also be introduced by searching into rotamer libraries of amino acid side chains surrounding the active site. At the end of this process, the docking algorithm generates a number of potential ligand poses, which are evaluated by the scoring function to find favorable binding interactions. Most of the scoring functions are based on FFs, that estimate the energy of the pose within the binding site, which contribution can be written as

$$\Delta G_{bind} = \Delta G_{solv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/t} + \Delta G_{vib} \tag{2.26}$$

The components consist of solvent effects (ΔG_{solv}), conformational changes in the protein and ligand (ΔG_{conf}), free energy due to protein-ligand interactions (ΔG_{int}), internal rotations (ΔG_{rot}), association energy of ligand and receptor to form a single complex ($\Delta G_{t/t}$) and free energy due to changes in vibrational modes (ΔG_{vib}). A low (negative) energy indicates a stable system and thus a likely binding interaction.

2.3.3 Molecular Dynamics

Molecular dynamics (MD) simulations permit to study dynamical processes occurring in molecular systems like enzyme conformational changes and substrate binding, or allosteric transitions in proteins. The time evolution of a molecular assembly of N atoms, can be calculated from the differential form of Newton's second law of motion:

$$F_i(t) = m_i a_i(t) = -\frac{dU}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \; ; \; i = 1,2 \dots N \tag{2.27}$$

where $(dU/\, dr_i)$ represents the gradient of the potential energy and relates to the force exerted on any atom i of mass m_i and with r_i its atomic coordinates, at a given time t. Although the latter equation has no analytical solution, the position of the atoms at a later (Δt) or earlier $(-\Delta t)$ time step can be approximated by a Taylor expansion:

$$r_{i+1} = r_i + v_i\,(\Delta t) + \frac{1}{2}\,a_i\,(\Delta t)^2 + \frac{1}{6}\,b_i\,(\Delta t)^3 + \cdots \tag{2.28}$$

$$r_{i+1} = r_i - v_i\,(\Delta t) + \frac{1}{2}\,a_i\,(\Delta t)^2 - \frac{1}{6}\,b_i\,(\Delta t)^3 + \cdots \tag{2.29}$$

in which velocities v_i, accelerations a_i and hypervelocities b_i are respectively the first (dr_i/dt), second (d^2r_i/dt^2) and third (d^3r_i/dt^3) derivative of the r_i position with respect to time t. Combining the two latter equations truncated at the second order and calculating the acceleration from the forces F_i at each time step leads to the *Verlet* algorithm,[152] which allows the propagation of the position to generate the trajectory:

$$r_{i+1} = (2r_i - r_{i-1})\,a_i(\Delta t)^2 + \cdots \tag{2.30}$$

$$a_i = \frac{F_i}{m_i} = -\frac{1}{m_i}\frac{dU}{dr_i} \tag{2.31}$$

However, since the time step is typically small and $(2r_i - r_{i-1})$ a larger number, this would be generating truncation errors, given that the velocity term does not appear explicitly. Such drawbacks are overcome by the *leapfrog* algorithm,[153] which uses half time steps to calculate velocities, and then uses these to obtain the positions at time $t + 1$

$$v_{i+\frac{1}{2}} = v_{i-\frac{1}{2}} + a_i \Delta t \tag{2.32}$$

$$r_{i+1} = r_i + v_{i+\frac{1}{2}} \Delta t \tag{2.33}$$

with the only disadvantage that velocities and positions cannot be computed at the same time step since they are out of phase by half time step. This inconvenience is finally solved by the *velocity Verlet* algorithm[154]

$$r_{i+1} = r_i - v_i(\Delta t) + \frac{1}{2} a_i (\Delta t)^2 \tag{2.34}$$

$$v_{i+1} = v_i + v_i + \frac{1}{2}[a_i + a_{i+1}](\Delta t) \tag{2.35}$$

Selecting the proper time step while setting up a MD simulation is fundamental: too large time steps may result in instability due to high energy overlaps between particles, while too small time steps may lead to trajectories covering only a limited phase space. A useful general guide to ensure that the phase space is covered efficiently and collisions occur smoothly is to select a time step approximately one-tenth of the fastest motion timescale of the studied system, which is typically the bond vibration involving hydrogen atoms (e.g. O–H bonds vibration has *ca* 10 fs period).[155] Therefore, a 1 fs (10^{-15}s) timestep would be required to treat a water-solvated system accurately and 10^6 timestep to converge a total simulation time of 1 ns (10^{-9} s). Nevertheless, the use of larger time steps (*ca* 2 fs) at a reduced computational cost is feasible by imposing constraints on hydrogen bond lengths as done by the SHAKE[156] or the LINCS[157] algorithms.

MD simulations generate a collection of points in the phase space accomplishing specific thermodynamic conditions known as ensembles, which depends upon the constrain used. Standard MD simulations sample the *microcanonical ensemble* (NVE), characterized by a fixed number of atoms *N*, fixed volume *V* and fixed energy *E*. However, experiments are usually performed at constant temperature and to reproduce such conditions the *canonical ensemble* (NVT) or the *isobaric-isothermal ensemble* (NpT) may be used, in which the volume *V* or the pressure *p* are conserved by exchanging the energy of the system with a thermostat or a barostat plus a thermostat, respectively.

2.3.4 Accelerated Molecular Dynamics

Time-dependent thermodynamic properties obtained through MD simulations can only be connected to experimental observables if all relevant states (conformations) of the system are explored. Within the ergodic hypothesis[158] if the system evolves in time indefinitely, all the accessible microstates are equally likely to be visited, and the time-averaged conformational sampling is equal to the probability distribution of all microstates in thermodynamic equilibrium. Nonetheless, in practice, this is not the case, as some domain motions and allosteric transitions occur in the millisecond to second timescale. Consequently, the gap with respect to the nanosecond timescale achievable by ordinary GPU-computed MD simulations frustrates a direct comparison with experiments. Massive parallel (e.g. Anton supercomputer) or distributed (e.g. Folding@home project) computing enables the sampling of several milliseconds of simulation per day without introducing any bias. On the other hand, it is possible to promote infrequent transitions and to increase the frequency of crossing the barriers of rarely explored states by introducing a biased external potential into the MD simulation. Among biased methods, accelerated molecular dynamics (aMD) allows the exploration of conformational states without *a priori* information about the specific transition, by adding a non-negative boost potential to the original potential of the system when it is lower than an energy threshold (eq. 2.36 and Fig. 2.7). This raises the energy minima while keeping the transition states almost unaffected, thus smoothing the free energy landscape and enhancing the conformational sampling.

$$V_{aMD}(r) = \begin{cases} U(r) & ; \quad U(r) \geq E_{thres} \\ U(r) + \Delta V(r) & ; \quad U(r) < E_{thres} \end{cases} \tag{2.36}$$

where $\Delta V(r)$ is the boost potential added by aMD and $U(r)$ is the system potential obtained from the FF. In its simplest form, the boost potential is given by:

$$\Delta V(r) = \frac{[E_{thres} - V(r)]^2}{\alpha + [E_{thres} - V(r)]} \tag{2.37}$$

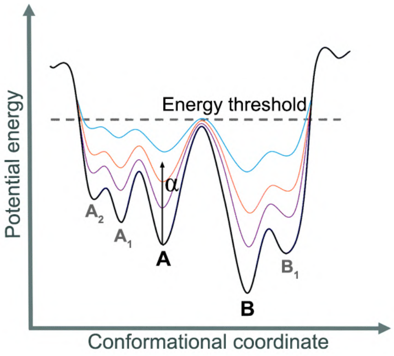


Figure 2.7. Schematics of the aMD method.

where the *a* parameter is the accelerating factor. As it decreases, the potential energy surface gets flattened and the transition from one state to another occurs easier. Excessively small *a* values should be avoided to prevent the transition to become barrierless, which would make it difficult to capture.

One of the main advantages of aMD is that the boost potential can be applied to internal coordinates, e.g. solely to the torsional term to enhance the confor-mational transition.

2.4 Conformational population analysis

The accurate description of the conformational ensemble sampled by biomolecular systems is of primary interest. As seen in section 1.8, natural and laboratory evolution are committed to evolve enzymes exploiting their inherent conformational dynamism and heterogeneity by populating conformational active states which confers to the enzyme a desired function or enhance pre-existing traits. MD simulations constitute a primary tool to describe the structural and dynamical changes that drive such a process at the atomistic level. Due to the large number of atoms present while simulating an enzyme in solvent, a flood of data is generated and the atomic population distribution is defined in an high dimensional space, thereby making the task of identifying discrete conformational states not trivial. A possible solution to reduce the degrees of freedom of the system is to identify and select few variables, which collectively represent the transitions of interest in the system's state, and then project the MD trajectory onto these variables to obtain a probability distribution. Such *collective variables* can be any differentiable function of the system coordinates (e.g. distance between catalytic residues, backbone dihedral angles), nevertheless they are seldom known *a priori*, therefore, they require detailed knowledge of the system and may introduce a bias or omit important information if not selected properly. A possible solution to reduce the dimensionality of the data while preserving information is to perform the *Principal Component Analysis*[159] (PCA), which achieves dimensionality reduction, accounting for as much variance in the dataset as possible. In a nutshell, by defining variance as the deviation of an atom from its mean position along the MD trajectory, then each principal component results in the linear combination of strongly correlated atomic motions with large oscillations.

$$Cr_i = \sigma_i r_i \tag{2.38}$$

where C is the covariance matrix of the dataset, r_i is the eigenvector matrix in which the principal components are represented as columns, and σ_i is the eigenvalue diagonalized variance matrix. The resulting lower dimensional PCA space can be used to project the MD trajectory and reconstruct the conformational landscape (Fig. 2.8) explored within the simulation. For instance, PCA has been applied to study the coupling of redox and conformational states in cytochrome P450 redox partners.[160] However, since transitions with the highest variance do not necessarily correlate with the slowest (i.e. kinetically relevant) motions, an alternative approach consists of considering the time-correlation of the data instead of the variance, as done in the *time-lagged independent component analysis*[161] (tICA), which lowers the data dimensionality while minimising the loss of kinetic information.

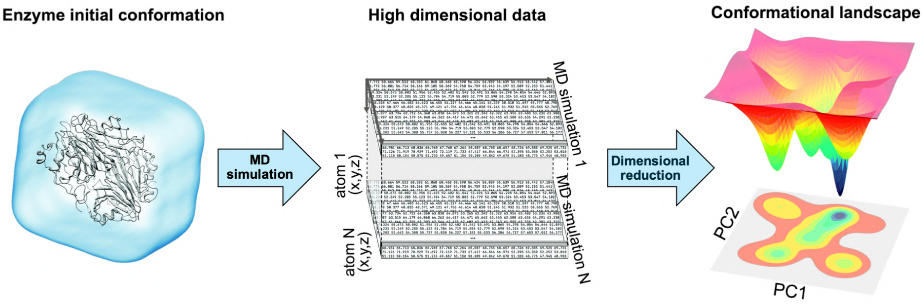


Figure 2.8. Schematic view of the dimensional reduction process and conformational landscape generation by PCA.

2.5 Shortest Path Map

Identifying key regions regulating the enzyme conformational dynamics is a crucial aspect when studying biological systems. These regions might not only be involved in the allosteric regulation of proteins but could also exhibit high dynamic coupling to the catalytic site even if present at distal position.[162] In many examples where DE has been successfully applied in enzyme design, dramatic fold increases were obtained after introducing mutations situated far away from the active site.[163] This suggests that the coupling of distal residues can have a direct impact on the catalytic activity and long-range dynamic effects may operate via communication networks within the enzyme scaffold. Correlation-based methods appear a valuable strategy to uncover the connection of different dynamic regions in proteins, providing a fingerprint of the motion along the MD simulations. Among these methods, the Shortest Path Map (SPM) developed in our group was recently applied to several systems to identify key positions, including residues outside the catalytic pocket, which were targeted by DE and shift the conformational landscape of the studied enzymes, thus influencing their catalytic activity.[164] The first step to compute the SPM is to gather the distance and the correlation matrix from the MD data, to estimate how the Cα of each residue deviate from its average position and the correlation between it and the Cα of any *i,j* residue along the MD trajectory, which is given by:

$$C_{ij} = \frac{< \Delta r_i \cdot \Delta r_j >}{\sqrt{< \Delta r_i^2 >< \Delta r_j^2 >}} \tag{2.39}$$

where Δr_i and Δr_j are the displacement of the Cα of the *i, j* residues with respect to their positions in the most populated cluster (as a reference). A node is created on each Cα and edges connecting pair of nodes (residues) are defined if their Cα distance is shorter than a cutoff (e.g. < 6 Å) and weighted according to the correlation coefficient:

$$d_{ij} = -\log\left|C_{ij}\right| \tag{2.40}$$

Thus, edges connecting pairs of highly correlated residues ($C_{ij} \rightarrow 1$) will be rather short ($d_{ij} \rightarrow 0$), whereas residues whose movement is noncorrelated ($C_{ij} \rightarrow 0$) will be connected by rather long edges ($d_{ij} \rightarrow \infty$). At this point, a graph with nodes and edges based on proximity-correlations is generated, and the Dijkstra algorithm is then applied to identify which edges are shorter, i.e. more correlated. The algorithm finally weights the width of each edge and the size of each node by counting the number of shortest paths passing through during the calculation (Fig. 2.9). The SPM provides a view of the enzyme pathways that contribute most to the enzyme dynamics correlated motions.

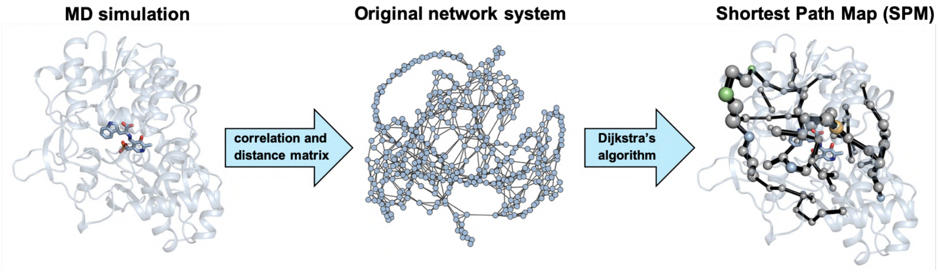


Figure 2.9. Schematic representation of the Shortest Path Map (SPM) construction from MD dataset. Speres represent the graph nodes connected by the edges.

Chapter 3: Objectives

The main purpose of the research presented in this thesis is the computational investigation of C–H oxidative hydroxylation catalysed by enzymes and biologically inspired non-heme iron-oxo complexes. This includes inherently orthogonal aspects owing to the wide range of sizes and features of the systems studied, embracing the intrinsic chemical reactivity of C–H activation, as well as the conformational dynamics and its implications on catalytic traits of laboratory-evolved P450-BM3 monooxygenase mutants. As in all disciplines of science, the convergence of computer simulations and experiments into a common goal permits a more comprehensive understanding of the mechanistic details, which could drive improvements both in biocatalysis and bioinspired homogeneous catalysis. A large part of the computational results reported in the following chapters are part of projects deriving from collaborative efforts with experimental groups. Depending on the nature and the properties of the investigated system, the principal aim of this work can be divided into two main blocks encompassing specific objectives:

- **Biomimetic non-heme iron-oxo complexe**s

Propose and apply a new computational strategy based on the analysis of the CDF, to determine whether hydrogen atom transfer (HAT) or concerted proton-coupled electron transfer (cPCET) occur during the first step of the C–H hydroxylation. Rationalize the electronic rearrangement of the iron-oxo and substrate moieties taking place during the oxidative cleavage of the C–H bond (Chapter 4).

Perform an in-depth analysis of the electronic structure and orbital picture variation in response to ligand modifications of oxoiron(IV) bioinspired complexes and how it involves distinct reaction channels. Apply the CDF analysis to quantify the degree of C–H cleavage during activation and rationalize the reactivity of the oxoiron compounds bearing different ligand arrangements (Chapter 5).

- **Laboratory evolved P450 monooxygenase variants**

Explore computationally the fitness landscape of a deconvoluted P450-BM3 mutant to unravel the intrinsic connection between epistasis and conformational dynamics and how epistatic effects can propagate from distal residues *via* a long range communication network to reshape the active site and fine-tune multiple catalytic functions (Chapter 6).

Decipher the molecular basis of regio- and stereoselectivity of a newly-designed BM3 mutant which hydroxylate steroids targeting position C7β. Based on the gathered results, offer a computational guidance to further mutate active site residues to engineer a novel BM3 variant capable of accepting bulkier steroidal substrates and retain the selectivity at C7β (Chapter 7).

Chapter 4: HAT vs. cPCET in biological non-heme oxoiron intermediates

4.1 State-of-the-art

Proton-coupled electron transfer (PCET)[41, 165] is a ubiquitous process in chemistry and biology, covering areas such as oxidation catalysis,[166] enzymatic reactions[167] and photosynthesis.[168] The PCET concept is used ambiguously to describe processes involving the movement of a proton and an electron.[165b] However, in the literature the use of the term PCET is far from consistent, following definitions that differ depending on whether the proton and the electron move together in space (same or different endpoint) and time (stepwise or concerted mechanism), as discussed in section 1.4.1.

Whether a stepwise or a concerted mechanism is in operation, can be inferred by the presence of a discrete intermediate that can in some cases be isolated experimentally. However, such a rigorous distinction between stepwise and concerted is not always straightforward, as it may relate back to the arbitrary definition of a lifetime (of an intermediate), which in turn may depend on the (choice of) experimental apparatus.[165c] Because of this, the Hammes-Schiffer group has developed a protocol to distinguish cPCET from HAT, based on the electron-proton nonadiabaticity, which reflects the change in the charge distribution taking place during the reaction.[169] Nevertheless, the interplay between experimental and theoretical methods might be still required,[165c] making this a rather a complex procedure to apply where often chemical intuition is not appealed to directly. Other methods have also been proposed, e.g. based on the evaluation of the deformation energies at the TS of the active complex[166] or the analysis of the intrinsic bond orbitals (IBOs) along the reaction pathway.[170] The latter tool identifies very effectively whether a cPCET or HAT mechanism is operating, by displaying the orbital change occurring upon bond reorganization along the reaction coordinate.

In this study an alternative approach is proposed, which relies on the analysis of the CDF[140] that accompanies the formation of a new chemical bond, in terms of the electron density rearrangement that is taking place during the generation of an adduct from two constituting fragments (section 2.1.5). As any PCET event is governed by a significant charge rearrangement, this approach addresses the underlying changes to the electronic structure directly and provides an intuitive way of interpreting results.

In particular, the charge displacement function, Δq(z), measures at each point z along a chosen axis (typically chosen along the bond(s) to be formed) the amount of electrons that move across a plane perpendicular to this axis passing through z (Fig. 4.3b). This flow of electrons is based on the difference between the final density of the total system and the sum of fragment densities. Positive values of Δq(z) correspond to electrons flowing in the direction of decreasing z, and negative values to electrons moving to increasing z. A positive slope indicates regions of charge accumulation, and *vice versa* a negative slope indicates depletion.[140] The CDF analysis has been widely applied in closed-shell systems to study e.g. charge transfer (CT) contribution in weak interactions[171] or the π coordination/activation by gold complexes.[172]

Here, the open-shell CDF (osCDF) will be applied using unrestricted DFT, showing that it can successfully discriminate between cPCET and HAT in a chemically intuitive way. To this end, the same case studies investigated before using IBOs,[170b] will be used, including model systems for the well-studied reactions of lipoxygenase[167b, 169c] and the intermediate TauD-J[14c, 15b, 22b] of taurine dioxygenase. The active species in lipoxygenase is an Fe(III)–OH, and for TauD-J an Fe(IV)–oxo, which cleave C(sp^3)–H bonds with cPCET and HAT mechanisms, respectively. For consistency, the osCDF was applied directly on the available TS coordinates of the two model systems:[170b] (i) a high-spin (S = 5/2) Fe^{III}–OH moiety coordinated to an amide, a carboxylate and three imidazoles, with a 2,5-diene as substrate to model the cPCET reaction of lipoxygenase; and (ii) a high-spin (S = 2) Fe^{IV}=O moiety coordinating to two imidazoles and a carboxylate (facial triad[14]) with an acetate to mimic the decarboxylated a-ketoglutarate coordination environment, and ethane as substrate to model the HAT reaction of the TauD-J intermediate (Fig. 4.1).

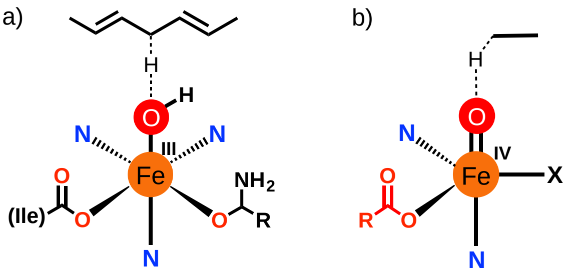


Figure 4.1. Transition State model structures used in this study: a) lipoxygenase; b) TauD-J.

4.2 Computational details

All DFT calculations were performed using the ADF2017.113 program package[173] under the unrestricted formalism. Molecular orbitals (MOs) were expanded in an uncontracted set of Slater-type orbitals (STOs) of triple-ζ quality basis, containing diffuse functions and two sets of polarization functions (TZ2P). Core electrons were treated by the frozen core (FC) approximation.[174] Scalar relativistic corrections were included self-consistently using the Zeroth Order Regular Approximation[134] (ZORA). Single point calculations were carried out, on top of the available optimized TS geometries,[170b] using the S12g functional[125] with TZ2P basis set (S12G/TZ2P), which was previously demonstrated to provide a consistent description of the spin state and electronic structure in transition-metal complexes.[175] The coordinates of the TS structures taken from ref. [170b] were used without further adjustment.

The electron density changes occurring upon the chemical bond formation between two fragments A and B can be analyzed through the CDF[140] (eq. 2.18). In this study, the overall complex corresponds to the TS structure of (i) the TauD-J + ethane substrate to model the HAT reaction, and (ii) the lipoxygenase + 2,5-heptadiene substrate to model the cPCET reaction. In each case, the two A and B fragments are represented by the high valent iron complex and its substrate; the z reference axis was chosen to lie along the Fe=O/Fe–O(H) bond, with the O atom at z = 0 and the iron complex at negative z. The CDF provides a spatially detailed picture of the total charge flow between the iron complex and the substrate fragments. The CDF analysis has proven to be a stable method with respect to computational settings such as basis set quality and exchange-correlation functional.[176]

4.3 Results and discussion

As a starting point, the HAT mechanism in TauD-J will be studied, as it can be considered the simpler case. As previously mentioned in section 1.5.1-2, it is generally accepted[33a, 34b-d] that when the C–H bond is cleaved via HAT, the $^5\sigma$-channel is the most favorable pathway (Fig. 1.8 top left panel and Fig. A1, Appendix A). In this pathway the α-spin electron moves from the C–H bond to the σ^* (Fe d_{z^2}–O p_z) antibonding orbital of the Fe=O unit to form the new O–H bond, whereas the other electron proceeds from the β-spin manifold of the :O(Fe) lone pair, as depicted in Fig. 4.2. Fig. 4.3 shows the 3D contour plots of the α/β/total electron density change when the fragments are brought together from infinite separation to the structure they attain at the TS.

A substantial decrease of α-spin density (σ-shaped) is observed in the C–H region (Fig. 4.3a), with a less pronounced increase in the O–H region, which is accompanied by a simultaneous charge rearrangement in the Fe=O unit: a decrease in the O(p_z)-shaped α-density is noticed, with a concurrent increase in the Fe(d_{z^2})-shaped α-density. On the other hand, the β-spin difference shows the rearrangement of the Fe=O unit with a shift from Fe(p_z) to the O–H moiety. This is accompanied by a decrease of β-density in the H atom region and an increase around the C atom, consistent with Figure 4.2.

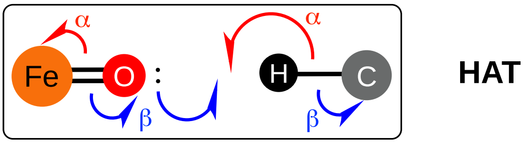


Figure 4.2. Representation of the electron flow in the HAT mechanism from C(sp^3)–H bond to the Fe^{IV}=O acceptor of the TauD-J intermediate. α- and β-spin electrons are drawn with red and blue arrows, respectively.

The total electron density difference 3D contour plots is also shown in Figure 4.3a. Here, a substantial depletion from the C–H region with a concomitant increase in the O–H region can be appreciated, concurrently with a decrease in the Fe–O bond and a slight accumulation on the Fe atom. The osCDF plot (Fig. 4.3b) finally gives us a quantitative picture of the charge fluxes associated with α-, β- and total-density rearrangement. The osCDF associated with the α-density rearrangement is positive in the whole molecular region (red curve), thus unambiguously indicating a net charge flux from the right to the left in the direction going from the C(sp^3)–H bond to the Fe^{IV}=O acceptor. On the contrary, the osCDF associated to the β-spin polarization (blue curve) is negative and the associated charge flux goes in the opposite direction. Noteworthy, by evaluating the CDF in the region of the Fe–O bond, it can be easily quantified that about 0.28 electrons of α-spin accumulate at the Fe site while about the same amount of charge (0.22 *e*) of opposite spin polarization (β-spin) moves from the same region. This can be interpreted as the swapping of an β-electron on Fe with an α-electron, where the β-electron moves onto OH, antiferromagnetically coupled to Fe^{III} in an exchange-coupled Fe^{III}–OH system.[170b]

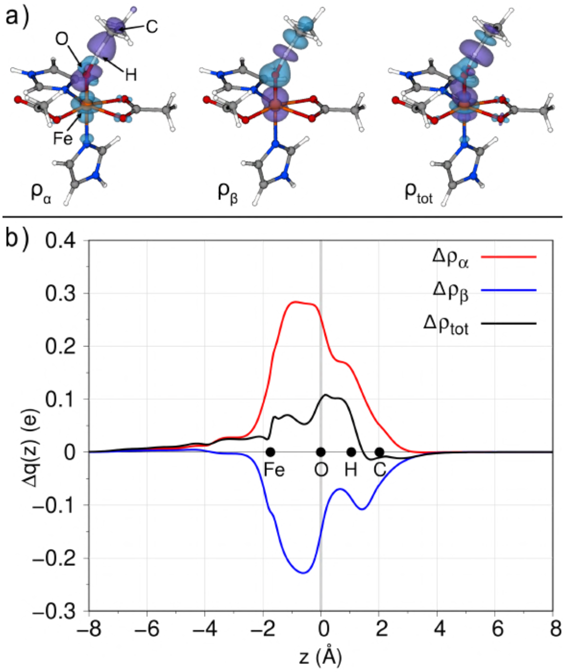


Figure 4.3. a) 3D contour plots (± 0.035 e au^{-3}) of the Δρ (α, β and total spin-density) for the FeO–H(substrate) bond in the TS of the TauD-J model complex; b) CDF curves of the corresponding density differences in a). Red, blue and black curves corresponds to α, β and total density, respectively.

The cPCET reaction of lipoxygenase can now be considered, following the same scheme as before. In this case, the d-shell manifold of the Fe–OH complex is fully occupied with 5 electrons of α-spin, so that a β-electron is transferred from the $\sigma_{C\text{–}H}$ to the $\pi^*_{xy/yz}$ (Fe $d_{xz/yz}$–O $p_{x/y}$) antibonding orbital of the Fe–OH unit via a $^6\pi$-channel (Fig. A1, Appendix A). However, typical for PCET events, this time the β-electron from the cleaved C–H bond populates the metal-based d-orbitals, whereas the newly formed O–H bond is created with both electrons proceeding from the :OH(Fe) lone pair, as depicted in Figure 4.4.

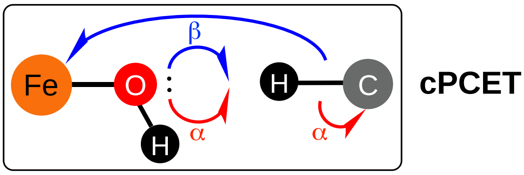


Figure 4.4. Representation of the electron flow in the cPCET mechanism from C(sp^3)–H bond to the Fe^{III}–OH acceptor of the lipoxygenase intermediate. α- and β-spin electrons are drawn with red and blue arrows, respectively.

The α-spin contour plot (Fig. 4.5a) shows a decrease of α-density on oxygen and an increase in the O–H region of the newly formed bond; simultaneously, α-density is moving within the CH-moiety from the hydrogen to carbon; a small depletion of α-electron on Fe is also appreciable. For the change in β-spin density, generally an accumulation on the O atom can be noted, with depletion on oxygen to form a lobe in the direction of the to-be formed O–H bond, resulting in an accumulation in the O–H region. A substantial decrease of β-spin density (σ-shaped) in the C–H region is present, together with an accumulation on Fe resembling a $d_{xz/yz}$ shaped orbital. However, a certain amount of β-spin density depletion is also appreciable around Fe. Notewhorthy, a depletion at oxygen in both the α- and β-spin density was not present in the TauD-*J* intermediate; this is a clear indication that in lipoxygenase both the α- and β-spin electrons are transferred to form the new O–H bond. Overall, in the total spin density difference, we note on the O atom a lobe of accumulating charge density pointing towards the to-be formed O–H bond, with a consequent increase at the O–H region. This is accompanied by a significant decrease in density on the transferring H atom, coupled with an accumulation on the C atom of the substrate. The total density accumulated and depleted on the iron is mainly coming from the β-spin density in Figure 4.5a, since the change in α-spin density on iron was negligible. Therefore, the flow of electrons in lipoxygenase is consistent with the arrows of Figure 4.4.

The difference between the HAT mechanism of TauD-*J* (Fig. 4.3b) and the cPCET mechanism of lipoxygenase (Fig. 4.5b) is easily seen by focusing on the Fe–O region. In the HAT mechanism, the total density rearranged is low around the Fe atom (about 0.06 *e*) and maximum in the O–H region. However, for cPCET, the total density rearrangement across the Fe-O unit assumes its maximum value at about 0.11 *e*. Overall, a net gain of charge density is only observed for cPCET, but not for HAT, which would be consistent with the expected flow of electrons for these two mechanisms (Fig. 4.2 and 4.4).

In summary, when applied to the TS structures of typical HAT or cPCET reactions, the osCDF analysis is capable to capture the different electron flow and rearrangement occurring during the HAT or cPCET mechanisms, allowing to identify which of the two scenarios is operating during the chemical reaction. In the HAT case, one electron travels together with the proton purely as hydrogen atom from the C–H bond of the substrate to the acceptor (Fe=O), while the other electron proceeds from the :O(Fe) lone pair to form the new O–H bond. In contrast, when a cPCET is operating, one electron moves from the C–H bond of the substrate to the anti-bonding orbital of the acceptor (Fe–OH), populating the metal-based $\pi^*_{xy/yz}$ (Fe $d_{xz/yz}$–O $p_{x/y}$) orbital, whereas the newly formed O–H bond is created with both electrons proceeding from the :OH(Fe) lone pair. In both cases, the electron flow perfectly matches the chemical intuition, which is represented by the arrows depicted in Figures 4.2 and 4.4.

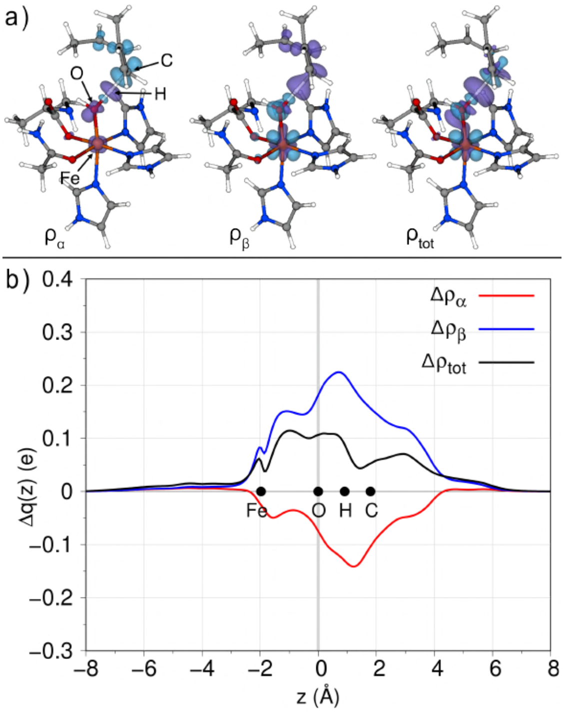


Figure 4.5. a) 3D contour plots (± 0.035 e au^{-3}) of the Δρ (α, β and total spin-density) for the FeO–H(substrate) bond in the TS of the lipoxygenase model complex; b) CDF curves of the corresponding density differences in a). Red, blue and black curves corresponds to α, β and total density, respectively.

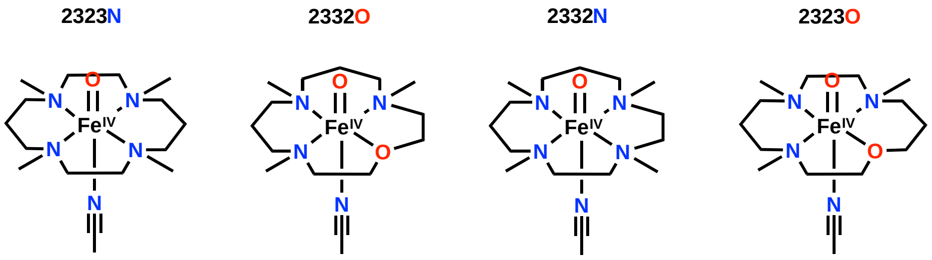
Chapter 5: Effect of ligand modification on the reactivity of bioinspired non-heme oxoiron(IV) complexes

5.1 State-of-the-art

Since the late 1950s, the controlled functionalization of non-activated C(sp^3)–H bonds has become an ubiquitous theme in transition-metal chemistry. Until the beginning of this century, the field of direct C–H activation was largely dominated by second- or third-row transition metal catalysts (Pt, Pd, Rh, Ru), limiting industrial applications due to their high cost.[177] As discussed earlier in this thesis, enzymes can efficiently and selectively functionalize aliphatic bonds under physiological conditions, and the use of these biocatalysts for oxidative C–H transformations is in principle possible. Yet, in practice, such an approach is rather limited to small-scale production of selected chemicals, and existing data demonstrate that the performance of exciting new enzymes and reactions has not yet been quite satisfactory for larger-scale industrial processes.[178]

During the past decade, various efforts have been directed towards enzyme-mimetic chemistry, since bio-inspired metal-oxo complexes may mimic the highly reactive intermediates found in metalloproteins, overcoming the limits imposed by small scale enzymatic reactions. Within this picture, biomimetic non-heme oxoiron(IV) complexes have attracted attention, as iron-containing enzymes like TauD, Rieske dioxygenase and sMMO can oxidize strong aliphatic bonds in a variety of organic molecules. To the best of our knowledge, over 80 non-heme oxoiron(IV) synthetic complexes have been reported to date,[26] but only few of them were identified as containing the reactive $S=2$ state[33a, 33b, 34] typical of non-heme enzymatic intermediates.[8b] Noteworthy, most of the biological $Fe^{IV}(O)$ units contain weak donor oxygen-based (often carboxylate) equatorial ligands, which as discussed in section 1.3 play a significant role in modulating the d_{xy} / $d_{x^2-y^2}$ energy gap, therefore stabilizing the more reactive high-spin state. Nonetheless, while $S=2$ $Fe^{IV}(O)$ complexes bearing oxygen-based equatorial ligands in O_h coordination have been successfully synthetized, their instability has prevented detailed investigations of their reactivity[28]. Moreover, efforts to stabilize the reactive $S=2$ species using carboxylate-rich ligands resulted in a low-lying quintet excited state and only a small enhancement (5-10 fold) of the HAT reaction[179] with respect to the common $Fe^{IV}(O)$ complexes supported by classical N-donor based scaffolds.

Recently Ray, Swart, Nam and co-worker reported the synthesis, characterization and DFT studies[180] of a $Fe^{IV}(O)$ complex bearing a weaker-field N_3O modification of the popular N_4-donor TMC ligand used in the $[Fe^{IV}(O)(TMC)(NMe)]^{2+}$ ($\mathbf{1}_{\mathrm{R}}$) complex. Notably, this novel $[Fe^{IV}(O)(TMCO)(OTf)]^{+}$ complex (TMCO = 4,8,12-trimethyl-1-oxa-4,8,12-triazacyclotetradecane; OTf = trifluoromethanesulfonate) exhibits a remarkable cyclohexane oxidation reactivity ($k_2 = 1 \cdot 10^{-2}$ M^{-1} s^{-1} at 233K) only one order of magnitude lower than the two most reactive oxoiron(IV) compounds reported to date, $[Fe^{IV}(O)(Me_3NTB)(NCMe)]^{2+}$ and $[Fe^{IV}(O)(TQA)(NCMe)]^{2+}$ (section 1.3). In comparison, the closely related $\mathbf{1}_{\mathrm{R}}$ supported by TMC is a sluggish oxidant, not reacting at all with cyclohexane even at higher temperatures. Computational studies at the OLYP/TDZP level of theory on $[Fe^{IV}(O)(TMCO)(NCMe)]^{2+}$ ($\mathbf{2}_{\mathrm{R}}$) and $\mathbf{1}_{\mathrm{R}}$ revealed[180] that while $\mathbf{1}_{\mathrm{R}}$ has an S=1 ground state at all temperatures, $\mathbf{2}_{\mathrm{R}}$ was predicted to be a degenerate S=1/S=2 state. Hence, despite being spectroscopically characterized as S=1 at 4K, $\mathbf{2}_{\mathrm{R}}$ may exhibit a TSR[49] at the reaction temperature of 233K (as discussed in section 1.5.2). DFT electronic energies at the same level of theory also showed a significant reduction of the HAA barriers going from $\mathbf{1}_{\mathrm{R}}$ to $\mathbf{2}_{\mathrm{R}}$ in both high- and low-spin states, contributing to the enhanced oxidizing power of $\mathbf{2}_{\mathrm{R}}$. However, it should be noted that the TMCO ligand employed differs from TMC, not only for the replacement of an $N\text{-}CH_3$ donor group in TMC by an O atom, but also for the ethylene and propylene spacers in the macrocycle placed in a different position (Fig. 5.1).



$\mathbf{1}_{\mathrm{R}}$ $\mathbf{2}_{\mathrm{R}}$ $\mathbf{3}_{\mathrm{R}}$ $\mathbf{4}_{\mathrm{R}}$

Figure 5.1. Structures of the oxoiron(IV) complexes (reactant form) considered in this study; note that $\mathbf{1}_{\mathrm{R}}$ corresponds to TMC and $\mathbf{2}_{\mathrm{R}}$ to TMCO, respectively. At the top, it is indicated the number of equatorial C atoms found counting clockwise from the N (O) atom replaced within the ligand environment.

Thus, the aim of this study is to extend previous DFT calculations taking into account all the different ligand spacer variations and/or N-CH_3 to O replacement, including complex $\mathbf{3}_R$ and $\mathbf{4}_R$. A frontier orbital perspective on the effect of the ligand spacers variation and the atom replacement is proposed here to elucidate how such ligand modifications affect the electronic structure of the different complexes and, as a consequence, how the electronic structure influences the HAA reaction profile. Finally, the driving force principle (section 1.5.3) is invoked and investigated by means of the charge displacement function (CDF) analysis (section 2.1.5) to disclose the diverse reactivity of the complexes, where the orbital picture alone cannot account for the predicted reactivities.

5.2 Computational details

All the DFT calculations were performed with the ADF2017.113 program package[173] using the OLYP functional[129] under the unrestricted formalism. Molecular orbitals (MOs) were expanded using two different uncontracted sets of Slater-type orbitals (STOs) of (i) triple-ζ quality basis containing diffuse functions and two sets of polarization functions (TZ2P) or (ii) triple-ζ quality basis on the Fe metal combined with double-ζ quality basis on all the other atoms, both containing one set of polarization functions (TDZP). Core electrons were treated by the frozen core (FC) approximation.[174] Scalar relativistic corrections were included self-consistently using the Zeroth Order Regular Approximation[134] (ZORA). Solvent effects were taken into account with the Conductor-like Screening Model[145a, 181] (COSMO). Geometry optimizations and subsequent frequency calculations to verify stationary points were performed with QUILD[182] (included within the ADF package) with the TDZP basis set (OLYP/TDZP). Single point calculations were carried out, on top of the optimized geometries with TZ2P basis set (OLYP/TZ2P).[183]

For the CDF analysis (section 2.1.5), the overall systems corresponds to the transition state (TS) structure occurring in the HAA reaction between the $Fe^{IV}(O)$ complexes and the cyclohexane substrate. In each case, two fragments A and B were used for respectively the high valent iron-oxo complex and the substrate, and the z reference axis was chosen to lie along the Fe=O bond, with the H atom of the C–H bond undergoing cleavage fixed at z = 0, and the Fe^{IV}(O) complex lying at negative z values.

5.3 Results and discussion

To start, let us compare the frontier molecular orbitals (FMOs) picture of the S=1 and S=2 states in $\mathbf{1_R}$ and $\mathbf{2_R}$ (Fig. 5.2), which can be considered two extreme cases featuring both the N-CH_3 to O replacement and the different position of the ethylene and propylene spacers in the ligand environment. The N_4 donor arrangement of the TMC ligand in $\mathbf{1_R}$ yields a strong equatorial ligand field, whereas the N_3O macrocycle in $\mathbf{2_R}$ represents the weak-field situation. The result of the weaker ligand field on the molecular orbitals will be discussed first, moving then to the effect of the different ligand spacers by taking into account the whole series of complexes $\mathbf{1_R}$-$\mathbf{4_R}$.

Whether the S=2 or S=1 is the ground state is dictated by the relative energy of the unoccupied $d_{x^2-y^2}$ with respect to the doubly occupied d_{xy} orbital. Occupying a low-lying $d_{x^2-y^2}$ instead of doubly occupying d_{xy} would result in exchange stabilization of the four unpaired *d* electrons, but if the $d_{x^2-y^2}$ energy is too high the penalty of occupying it will not be compensated by the exchange stabilization, and spin pairing in d_{xy} will prevail resulting in a S=1 ground state. A qualitative but practical expression to assess the interplay between these two factors was proposed by Shaik,[47b] which considers the relative energy of two states ($\Delta E_{1,2}$) as a balance between the exchange interactions ($\Delta K_{1,2}$) within the *d* block and the corresponding orbital-promotion energy difference ($\Delta E_{orb(1,2)}$).

In the S=1 state, the weaker field exerted by the TMCO ligand in $\mathbf{2_R}$ stabilizes the $d_{x^2-y^2}$ α LUMO orbital by 0.33 eV with respect to that in $\mathbf{1_R}$ (bearing TMC), and to a lesser extent (0.19 eV) the d_{xy} β orbital, overall resulting in an orbital promotional energy of 1.14 eV in $\mathbf{2_R}$ with respect to the 1.28 eV found for $\mathbf{1_R}$. In the S=2 state the $d_{x^2-y^2}$ α becomes occupied (HOMO), and in $\mathbf{2_R}$ it lies 0.31 eV lower than in $\mathbf{1_R}$ (Fig 5.2 and Table B1, Appendix B for all the orbital energies).

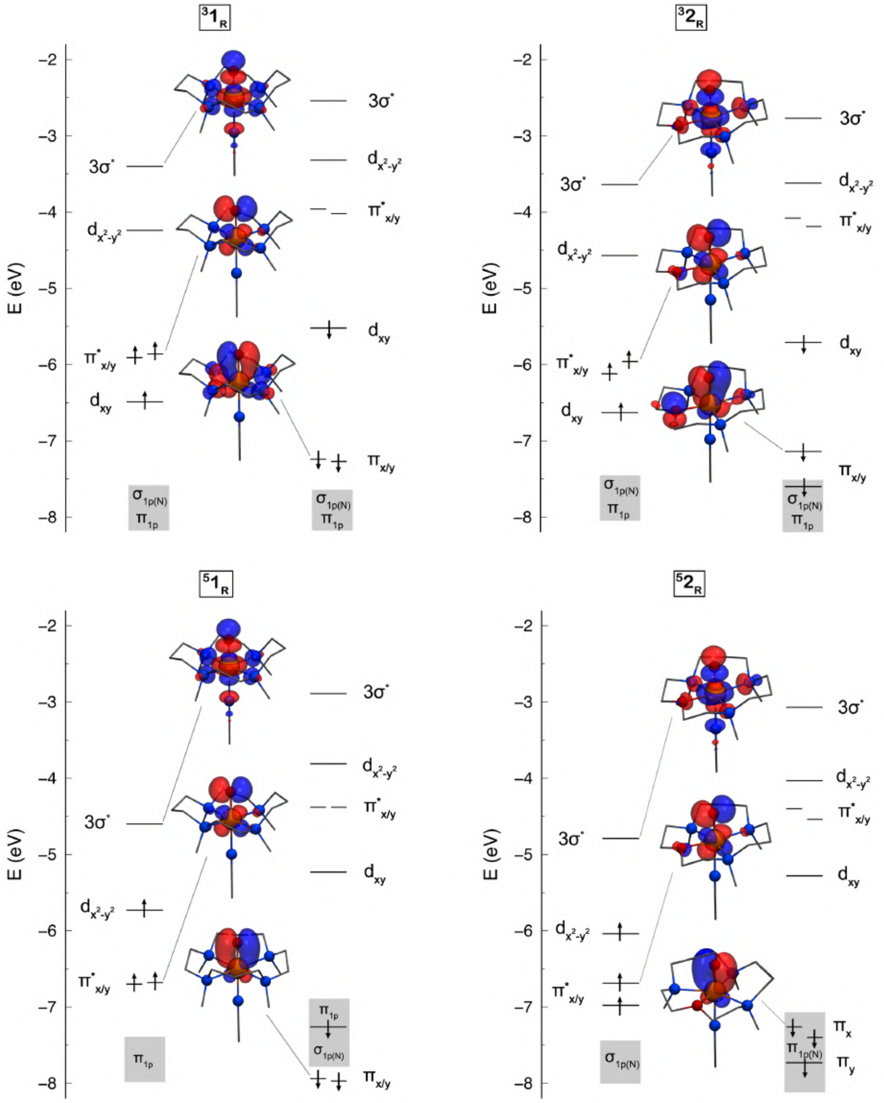


Figure 5.2. Frontier molecular orbitals of complex $\mathbf{1}_R$ and $\mathbf{2}_R$ in the S=1 and S=2 states. In each panel the α and β spin orbitals are to the left and right, respectively. Isodensity value ± 0.015 e au^{-3}.

Indeed, by looking at the electronic energies, the quintet state of $\mathbf{2}_R$ results 0.4 kcal mol^{-1} more stable than the triplet, whereas for $\mathbf{1}_R$ the triplet is predicted as ground state, 3.5 kcal mol^{-1} more stable than the quintet (Fig. 5.3). Moreover, in section 1.5.1-2 the key role played by the $\sigma^*_{z^2}$ orbital in the HAA (rate determining step) has been discussed along the more reactive S=2 surface ($^5\sigma$-pathway). In the reactive S=2 state, **2_R** exhibits a low-lying $\sigma^*\alpha$ orbital stabilized by 0.19 eV with respect to **1_R** thanks to the weaker field of the TMCO ligand. The σ^* α orbital (electron accepting orbital, EAO) of **2_R** in the quintet state can readily accept a σ_{CH} α electron from the substrate (Fig. 1.8 top-left panel), leading to a lower activation barrier of 22.3 kcal mol^{-1} compared to the 32.3 kcal mol^{-1} required by **1_R** in the same state.

Conversely, in the S=1 state the $^3\pi$-pathway is favored, where a β electron is transferred from σ_{CH} to the $\pi^*_{xy/yz}$ orbital (Fig. 1.8 bottom-left panel), which in **2_R** is stabilized by 0.12-0.17 eV and gives rise to a lower HAA TS barrier of 35.3 kcal mol^{-1} compared to the 44 kcal mol^{-1} in **1_R**.

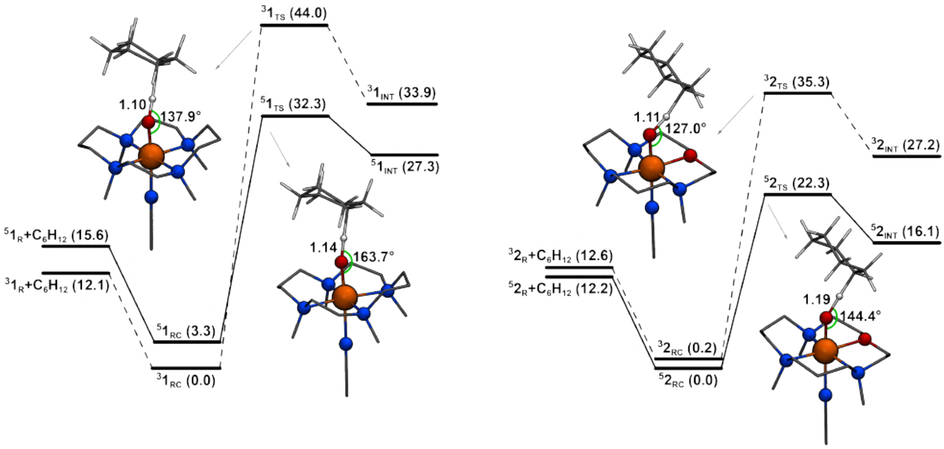


Figure 5.3. Energy profiles (in kcal mol^{-1}) for cyclohexane H-abstraction by complex **1_R** (left) and **2_R** (right) in the S=1 and S=2 states at the OLYP/TZ2P//OLYP/TDZP level of theory. O··H distances and Fe-O-H angles of the corresponding transition state structures are reported in Å and °, respectively.

Hence, it may be concluded that the remarkable oxidizing capabilities of **2_R** are due to the weaker field generated by the TMCO ligand, which pushes the iron-based *d*-orbitals energy levels much less, thereby improving the electron accepting abilities of the σ^* and π^* orbitals. It makes at the same time the more reactive S=2 path more accessible, therefore ensuring the reaction to proceed along the $^5\sigma$-channel.

Having established (i) the central role of the FMOs on both triplet-to-quintet energy gap ($\Delta E_{T,Q}$) and HAA activation barrier (related to the low-lying σ^* or π^* EAO depending on the given spin state) and (ii) how the stronger (or weaker) equatorial ligand field can upshift such orbitals, the effect of the macrocyclic ligand spacer arrangement (2323 or 2332, Fig. 5.1) was considered, taking into account the whole series of ligands in **1_R**-**4_R** and monitoring the FMOs. By comparing complexes **1_R** vs. **4_R** (or **2_R** vs **3_R**) the ligand field effect of replacing N-Me by oxygen is assessed; alternatively, comparing complexes **1_R** vs. **3_R** (or **2_R** vs. **4_R**) one purely identifies the effect of the different arrangement of the ethylene and propylene spacers (either 2323 or 2332). Within the same coordination environment around the iron (either N_4 or N_3O), it should expected that the electron donating atoms push the Fe d-orbitals to a different extent depending on their distances with respect to Fe. Interestingly, the diverse positioning of the ethylene and propylene spacers featured by the two types of ligands employed (2323 and 2332) results in different Fe–N and Fe–O distances (Table 5.1) and consequently in a stronger (weaker) ligand field that pushes the Fe d-orbitals energy levels to a greater (smaller) exent when the Fe–N and Fe–O distance becomes shorter (longer).

Table 5.1. Ligand arrangement and relevant bond distances (Å) for the series of Fe(IV)-oxo complexes **1_R**-**4_R**.

Spin	**Complex**	**Ligand**[a]	**Fe–N_1(O)**	**Fe–N_2**	**Fe–N_3**	**Fe–N_4**	**Fe–N_{ax}**	**Fe–N(O)$_{avg}$**[b]
S=1	**1_R**	2323N	2.12	2.16	2.16	2.12	2.09	2.13
	2_R	2332O	2.01	2.16	2.09	2.17	2.13	2.11
	3_R	2332N	2.14	2.14	2.22	2.14	2.09	2.15
	4_R	2323O	2.07	2.12	2.05	2.11	2.11	2.09
S=2	**1_R**	2323N	2.21	2.28	2.28	2.21	2.06	2.21
	2_R	2332O	2.18	2.26	2.13	2.26	2.08	2.18
	3_R	2332N	2.48	2.18	2.23	2.18	2.07	2.23
	4_R	2323O	2.23	2.18	2.16	2.11	2.07	2.15

[a] The numbering refers to the equatorial C atoms counted clockwise starting from the N_1 (or O) atom.

[b] Average calculated as (Fe-N_1(O) + Fe-N_2 + Fe-N_3 + Fe-N_4 + Fe-N_{ax}) / 5

Therefore, considering the set **1_R**-**4_R**, the ligand field exerted on the Fe atom will depend upon both the N-Me to O replacement (referred as atom replacement) and the ethylene and propylene spacers position (referred as spacer arrangement), where the 2323 arrangement yields a stronger ligand field (shorter Fe–N(O) distances) and the 2332 a weaker ligand field (larger Fe–N(O) distances).

The FMO levels of the whole series of complex **1_R**-**4_R** are depicted in Fig. 5.4 (for the FMO representation of **2_R** and **3_R** and energies see Fig. B1 and Table B1, Appendix B). The $\pi^*_{xy/yz}$ and $d_{x^2-y^2}$ / d_{xy} are reported for the S=1 as these orbitals are the EAO of the $^3\pi$-pathway and HOMO/LUMO responsible of the spin-spitting, respectively. The σ_z^* is reported for the S=2, being the EAO of the $^5\sigma$-path. At first glance, it is clear that the orbital energies of the $\pi^*_{xy/yz}$ and d_{xy} orbitals are not affected by the propylene and ethylene spacer positioning, as these orbitals have comparable energies in **1_R**/**3_R** and **2_R**/**4_R**. On the other hand, the N-Me to O replacement (i.e. comparing **1_R** vs. **4_R** and **3_R** vs. **2_R**) lowers the energy of these orbitals thanks to the weaker equatorial ligand field of the O atom. The $d_{x^2-y^2}$ orbital energy is instead lowered by both the atom replacement (comparing **1_R** vs. **4_R** and **3_R** vs. **2_R**) and to a larger extent by the spacer arrangement (**1_R** vs. **3_R** and **4_R** vs. **2_R**).

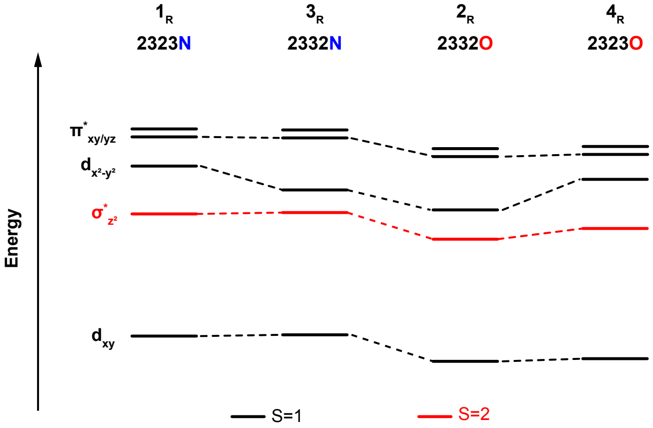


Figure 5.4. Ligand effect on the splitting of the frontiers orbitals of the complex series **1_R**-**4_R**.

The latter effect is not surprising since the $d_{x^2-y^2}$ is the only orbital pointing directly along the Fe–N(O) bonds, therefore it is highly influenced by small Fe–N(O) distance variations. Within this picture, complex **3_R** benefits from a favorable situation in which the $d_{x^2-y^2}$ orbital is stabilized by the spacer arrangement, but no atom replacement stabilizes d_{xy}, leading to a small $\Delta E_{T,Q}$ of 1.14 eV. The difference of 0.14 eV between the $\Delta E_{T,Q}$ of $\mathbf{1_R}$ and $\mathbf{2_R}$ corresponds to 3.23 kcal mol^{-1}, close to the shift of 3.9 kcal mol^{-1} in spin state splittings (from -3.5 kcal mol^{-1} to +0.4 kcal mol^{-1}) between $\mathbf{1_R}$ and $\mathbf{2_R}$. This stabilization of the $d_{x^2-y^2}$ orbital is therefore the primary reason why the reactant form $\mathbf{2_R}$ is predicted to have the quintet ground spin state (Fig. 5.5) and was experimentally characterized as high-spin. In general, the $\Delta E_{T,Q}$ qualitatively reflects the trends of the predicted ground spin state (Table 5.2).

Table 5.2. Triplet-to-quintet orbital energy gap ($\Delta E_{T,Q}$) and spin-state energies.

Complex	$\mathbf{4_R}$	$\mathbf{1_R}$	$\mathbf{2_R}$	$\mathbf{3_R}$
$\Delta E_{T,Q}$ (eV)	1.35	1.28	1.14	1.09
S=1 (kcal mol^{-1})	0.0	0.0	0.4	1.6
S=2 (kcal mol^{-1})	3.3	3.5	0.0	0.0

However, despite its high-spin S=2 ground state, $\mathbf{3_R}$ is a weak oxidant comparable to $\mathbf{1_R}$ (k_2=4.0·10^{-3} and 1.3·10^{-3} M^{-1} s^{-1} respectively, for the oxidation of thioanisole at 233K), which can be expected considering their similar S=2 barriers and σ^* EAO energies (Fig. 5.3 and 5.5). Likewise, in the S=1 state $\mathbf{1_R}$ and $\mathbf{3_R}$ also exhibit comparable $^3\pi$-pathways (i.e. similar TS barriers and π^* orbital energies). Counterintuitively, an analogous HAA barrier is also found for the triplet of complex $\mathbf{4_R}$ (Fig. 5.5), which could appear too high considering that the π^* EAO of $\mathbf{4_R}$ is 0.16 eV more stable than in $\mathbf{1_R}$ and $\mathbf{3_R}$ and lies at the same energy as in $\mathbf{2_R}$, thereby a barrier similar to this latter would be expected.[34b-d] Nevertheless, the TS geometry of $\mathbf{4_{TS}}$ reveals an Fe-O-H angle of 137.3°, notably larger than in $\mathbf{2_{TS}}$ (127.0°) and comparable to those found in $\mathbf{1_{TS}}$ and $\mathbf{3_{TS}}$, indicating a less efficient overlap between the π^* (acceptor) and the σ_{CH} (donor) orbitals (section 1.5.1), which can explain the higher activation energy of these $^3\pi$-pathways. In $\mathbf{1_R}$ and $\mathbf{3_R}$ the σ^* orbital is located at the same energy, indeed the two complexes on the quintet surface exhibit analogous reaction barriers of 44.0 and 41.4 kcal mol^{-1}, respectively.

The σ^* orbital is stabilized by the atom replacement (**1**$_\mathrm{R}$ vs. **4**$_\mathrm{R}$ and **3**$_\mathrm{R}$ vs. **2**$_\mathrm{R}$) and slightly lowered by the spacer arrangement only in TMCO-based ligands (**4**$_\mathrm{R}$ vs. **2**$_\mathrm{R}$), having the same energy in TMC-based ligands (**1**$_\mathrm{R}$ and **3**$_\mathrm{R}$).

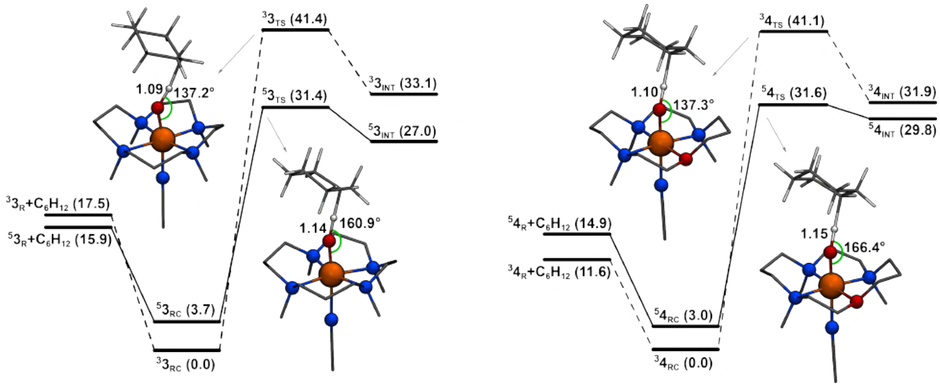


Figure 5.5. Energy profiles (in kcal mol^{-1}) for cyclohexane H-abstraction by complex **3**$_\mathrm{R}$ (left) and **4**$_\mathrm{R}$ (right) in the S=1 and S=2 states at the OLYP/TZ2P//OLYP/TDZP level of theory. O··H distances and Fe-O-H angles of the corresponding transition state structures are reported in Å and °, respectively.

In **4**$_\mathrm{R}$, although the σ^* is stabilized by 0.11 eV with respect to **1**$_\mathrm{R}$ and **3**$_\mathrm{R}$ by the weaker ligand field (either due to atom replacement and/or spacer arrangement), the activation energy of the S=2 state ($^5\sigma$-pathway) is unexpectedly high (31.6 kcal mol^{-1}) and similar to those of **1**$_\mathrm{R}$ and **3**$_\mathrm{R}$. For comparison, in **2**$_\mathrm{R}$ a further stabilization by 0.08 eV of the same orbital (with respect to the σ^* energy in **4**$_\mathrm{R}$) yelds the low S=2 barrier of 22.3 kcal mol^{-1}. In this case the orbital overlap outlined by the Fe-O-H angle in the TS structures cannot account for the reactivity. Indeed, despite **1**$_\mathrm{TS}$, **3**$_\mathrm{TS}$ and **4**$_\mathrm{TS}$ showi rather similar angles and analogues barriers, the Fe-O-H angle in **2**$_\mathrm{TS}$ is counterintuitively smaller, since an higher value (closer to 180°) should be expected[34b-d] for an efficient overap between the σ^* and the σ_CH orbitals along the $^5\sigma$-path.

In this case, the TS geometries of the quintet states of **1**$_\mathrm{TS}$, **3**$_\mathrm{TS}$ and **4**$_\mathrm{TS}$ show a rather comparable (Fe)O··H(C) distance (1.14-1.15 Å) whereas for **2**$_\mathrm{TS}$ it is longer (1.19 Å), indicating an incipient O··H bond that is not yet formed. Likewise, the distance of the C–H bond that undergo cleavage in **1**$_\mathrm{TS}$, **3**$_\mathrm{TS}$ and **4**$_\mathrm{TS}$ ranges between 1.42-1.44 Å, while for **2**$_\mathrm{TS}$ it is shorter (1.35 Å) indicating a C–H bond that is less cleaved. All data point to an earlier TS that resembles the reactant complex (RC) more for $\mathbf{2}_{TS}$ than for $\mathbf{1}_{TS}$, $\mathbf{3}_{TS}$ and $\mathbf{4}_{TS}$. The BEP principle discussed in section 1.5.3, which correlates the Hammond postulate with the activation energy of a given reaction, could elegantly explain: (i) how although $\mathbf{4}_{R}$ possesses a stabilized σ^* EAO compared to $\mathbf{1}_{R}$ and $\mathbf{3}_{R}$, it exhibits an analogous barrier of the 5σ-path to these latter, and (ii) why $\mathbf{2}_{TS}$ features a lower barrier with respect $\mathbf{4}_{TS}$ despite having a low Fe-O-H angle. As a proof of concept, a correlation is found by plotting the activation energies versus the (Fe)O··H(C) and the C–H distances for the series of complexes $\mathbf{1}_{TS}$-$\mathbf{4}_{TS}$ (Fig. 5.6). In the plot, the O··H (C–H) distance in $\mathbf{2}_{TS}$ is the longest (shortest) in both the S=1 and S=2 spin states, although in this latter the difference with respect to $\mathbf{1}_{TS}$, $\mathbf{3}_{TS}$ and $\mathbf{4}_{TS}$ is more pronounced.

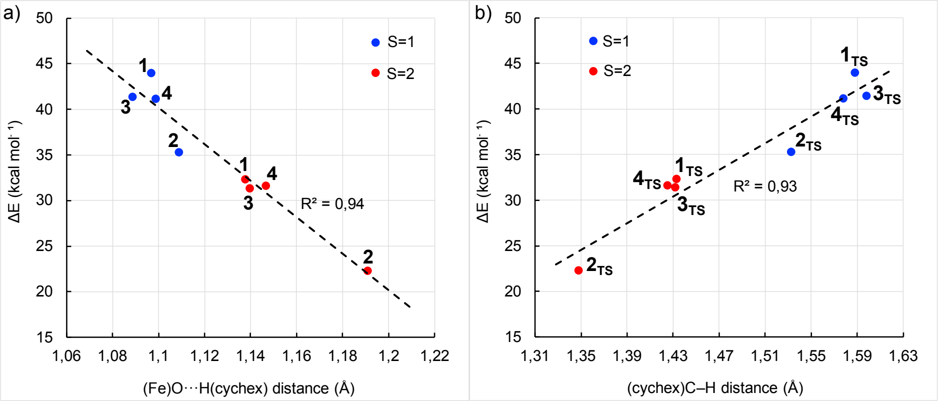


Figure 5.6. Correlation of the energy barriers with respect to the distance of the new formed O···H (a) and the cleaving C–H bond (b) in the TS structure.

Obviously, early or late TSs implying different degrees of C–H bond cleavage and O–H bond formation should also encompass different amounts of charge transferred from the substrate to the Fe=O acceptor. To further investigate this hypothesis, the previously introduced (Chapter 4) open-shell CDF analysis was applied, which constitutes a useful tool for studying the charge rearrangement taking place upon bond formation. In Fig. 5.7 the result obtained for complex $\mathbf{2}_{TS}$ is reported as example. Analogous results were obtained for all the other complexes (Appendix B).

In the S=2 state, the osCDF plot associated with the α-density rearrangement is found to be positive in the whole molecular region (red curve), thus unambiguously indicating a net charge flux from the C(sp^3)–H bond (right) to the Fe^{IV}=O acceptor (left). This result is in agreement with the postulated $^5\sigma$-path during the HAA, in which an α-spin electron is transferred from the σ_{CH} orbital of the substrate to the low-lying $\sigma^*_{z^2}$ (Fe d_{z^2}–O p_z) EAO of the Fe=O moiety (Fig. 1.8, top-left panel). On the contrary, in the S=1 state the β-density flow (blue curve) is positive (hence flowing from the C–H bond to the Fe=O unit), indicating a β-spin electron traveling to the half-occupied $\pi^*_{xy/yz}$ (Fe $d_{xz/yz}$–O $p_{x/y}$) EAO of Fe=O, which is fully consistent with the postulated $^3\pi$-pathway (Fig. 1.8, top-right panel). Furthermore, for each spin state an electron flow of opposite spin (i.e. β-spin in the quintet and α-spin in the triplet) originates from the Fe=O moiety towards the substrate (thus a negative density flow is observed) to participate in the formation of the newly O··H bond. Similar to the TauD-J intermediate studied in the previous chapter, the total rearranged density (black curve) is relatively small around the Fe atom and is mainly observed in the O–H region, indicating that no electrons are transferred directly to the Fe *d*-orbitals; this has as consequence that the operating mechanism in the rate determining H-abstraction is an HAT scenario, rather than a cPCET.[184] Qualitatively analogous results were obtained for the other complexes (Appendix B), evidencing that the concepts so far discussed regarding $\mathbf{2}_{TS}$ may serve as a paradigm to infer that in all cases the reaction is feasible along the $^5\sigma$- and $^3\pi$-pathway and takes place as the transfer of a genuine hydrogen atom.

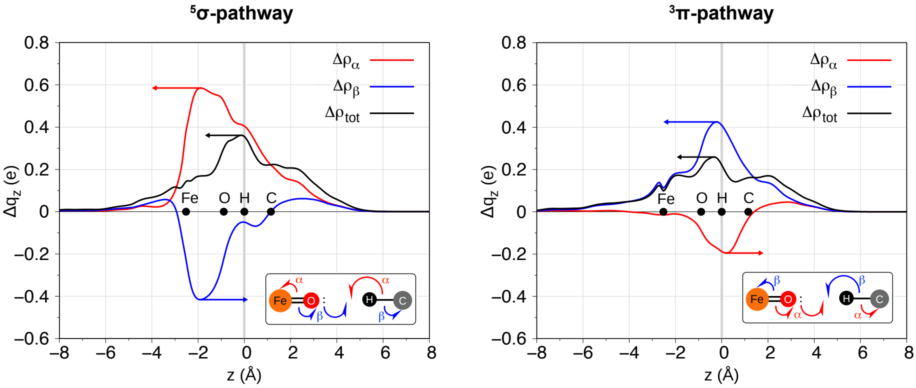


Figure 5.7. CDF plots of the FeO···H(cychex) bond in $\mathbf{2}_{TS}$. S=2 (left) and S=1 (right) spin state. Inset: representation of the electron flow in the HAT mechanism from C(sp^3)–H bond to the Fe^{IV}=O acceptor. Red and blue curves/arrows correspond to α and β spin density flows, respectively. Black curves correspond to the total density flow.

On the other hand, in the HAT the proton travels together with its electron towards the Fe=O acceptor during the C–H bond lengthening. Within the HAt scenario, the value of the charge transfer (CT) measured at the H atoms in the TS is of particular interest, as it offers a quantitative assessment of the degree of C–H bond cleavage. In the S=2 state the $^5\sigma$-pathway is operative, and the degree of C–H cleavage can be inferred through the α-spin density difference on H, which is related to the transfer of an α-electron from the σ_{CH} orbital to the empty σ^* EAO. Conversely, in the S=1 state, the $^3\pi$-pathway prevails and the β-spin density difference on H is related to the transfer of a β-electron to the half-occupied π^* EAO (for the CT values see Appendix B). A correlation is found between the values of these spin-density flows and the activation energy of the reaction in the corresponding spin state (Fig. 5.8), thereby confirming that a strict relationship between the degree of C–H cleavage and the energy barrier exists.

Likewise what has been discussed regarding the O··H distance, complex $\mathbf{2}_{TS}$ reaches a TS structure that exhibits the lowest degree of cleavage of the C–H bond in both the triplet and quintet states, in agreement with the low energy barrier predicted. Therefore, the remarkable oxidizing capability of $\mathbf{2}_{TS}$ does not originate only from the high electrophilicity of the FeO unit (low-lying EAO) but is in fact the result of the interplay of both the FMOs and the driving force. In other words, when evaluating the HAT reactivity, both the electronic and the thermodynamic factors must be taken into account.

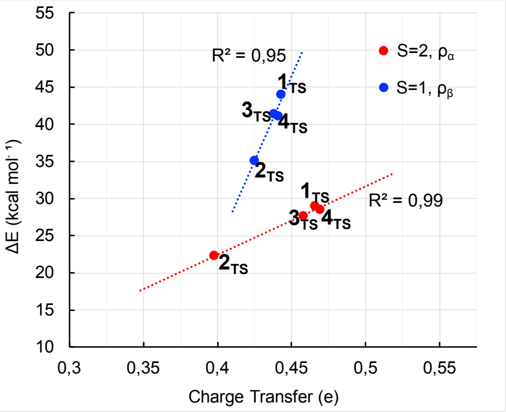


Figure 5.8. Correlation of the energy barriers with respect to the spin-density flows at the H atom in the TS.

Indeed, although the orbital picture may be sufficient enough to account for the enhanced reactivity of $\mathbf{2}_\mathrm{R}$ based solely on its stabilized EAO (either σ^* in the S=2 or π^* is the S=1 states) compared to $\mathbf{1}_\mathrm{R}$, $\mathbf{3}_\mathrm{R}$ and $\mathbf{4}_\mathrm{R}$, it cannot explain why although $\mathbf{4}_\mathrm{R}$ possesses in the S=2 state a lower energy σ^* acceptor than $\mathbf{1}_\mathrm{R}$ and $\mathbf{3}_\mathrm{R}$, the reaction barrier of these complexes are all comparable. Instead, by invoking the BEP principle, a correlation between activation energies, spin states and rate of progression of the TS along the (C)H···O(Fe) reaction coordinate is observed. According to the BEP principle, a lower HAA barrier corresponds to an earlier TS and *vice versa*. By quantitatively estimating the charge transferred from the C–H bond to the EAOs, the CDF analysis reconnects the orbitals with the degree of cleavage of the C–H bond (which obviously depends upon an earlier or later TS) and the activation barrier of the HAT reaction. The CDF analysis demonstrates that the high electrophilicity of low-lying EAOs, which is commonly believed to correlate with the HAT capability of $\mathrm{Fe^{IV}(O)}$ species, does not necessarily involve a lower HAT barrier, that is instead expected when the TS is reached earlier and the C–H bond is less cleaved.

In summary, the results presented in this chapter indicate that both the electronic and thermodynamic factor play a role in lowering the energy barrier of the HAT reaction, thereby defining the optimal properties of a $\mathrm{Fe^{IV}(O)}$ complex for activating C–H bonds. Equatorial ligands with a weak ligand field will favor both the reactive high-spin state and electrophilicity of the Fe=O unit (low-lying EAO), determining the preference for the $^5\sigma$ channel which relates to lower reaction barriers. On the other hand, TSs achieved with less cleaved C–H bonds will be reached earlier along the reaction coordinate, contributing to reduce the activation energy of the reaction according to the BEP principle.

Chapter 6: Pervasive epistasis emerge *via* conformational dynamics in laboratory-evolved $P450_{BM3}$ monooxygenase mutants

6.1 State-of-the-art

Directed evolution (DE) constitutes undoubtedly a powerful tool to enable novel enzymatic functions and optimize existing ones like substrate scope, stability, activity and selectivity, to name a few. By applying iterative rounds of mutagenesis, expression and screening/selection, the original aminoacidic sequence of the wild-type (WT) enzyme can be modified to evolve desired biocatalytic traits.[71f, 79] In DE the inherent structural heterogeneity and flexibility due to enzyme conformational dynamics are exploited to reshape the landscape of thermally accessible conformational states, populating those that are active toward novel functions or improving pre-existing traits, leading to higher fitness.[115-116, 164a, 185]

Yet, when multiple mutations are introduced into the enzyme sequence to obtain the fittest variant, their non-additive synergism/antagonism (i.e. positive/negative epistasis, respectively) may constrain the combinatorial explosion of mutationally-available evolutionary pathways, increasing the evolution of new and enhanced functions by allowing the access to mutational steps that avoid deleterious downfalls (evolutionary trap).[103-104]

Epistatic effects can arise between mutations located far away from the active site, indirectly interacting with the substrate *via* long-range interactions mediated by changes in the enzyme conformational dynamics.[186] For fundamental and practical reason, it is thus important to determine the existence, type and molecular basis of epistasis and its connection with conformational dynamics, which in turn will increase our understanding of protein evolution. To date, such connection has been studied in few protein families as proline isomerase (cyclophilin A),[187] phosphotriesterase[188] and β-lactamases[93, 189] providing fascinating insights, nevertheless only with a single protein trait (usually activity) as a limited measure of fitness, in contrast with DE where often two or more traits (e.g. activity and selectivity, or stability) are sought for practical purposes.[71f, 92]

Recently, Reetz and co-workers applied DE to the $P450_{BM3}$ monooxygenase F87A variant, targeting both activity and selectivity to achieve the stereo- and regio-selective hydroxylation of testosterone (**1**) into 2β-hydroxytestosterone (**2**, Fig. 6.1). Combinatorial saturation mutagenesis at the randomization site R47I/T49I/Y51I allowed the evolution of mutant **III** (F87A–R47I/T49I/Y51I, also reported as **KSA-2** variant in a previous study, section 1.7) displaying 94% 2β-selectivity and 67% conversion of **1** in a 24h whole cell reaction.[69] Interestingly, the three mutated residues are located next to each other (distances of Cα is ~6 Å) lining the large binding pocket but relatively far away (~15-20 Å) from the Fe=O heme cofactor. Subsequently, the complete deconvolution of variant **III** starting from parental F87A enabled the exploration of all the 3!=6 step-wise mutational paths, thus determining the full multidimensional fitness landscape based on activity and selectivity as a measure of fitness. As mentioned in Section 1.8, fitness landscape are particulary useful to address key questions as (i) how mutated residues impact evolved enzyme traits and (ii) whether mutations interact epistatically.

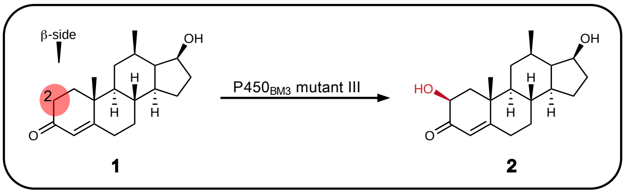
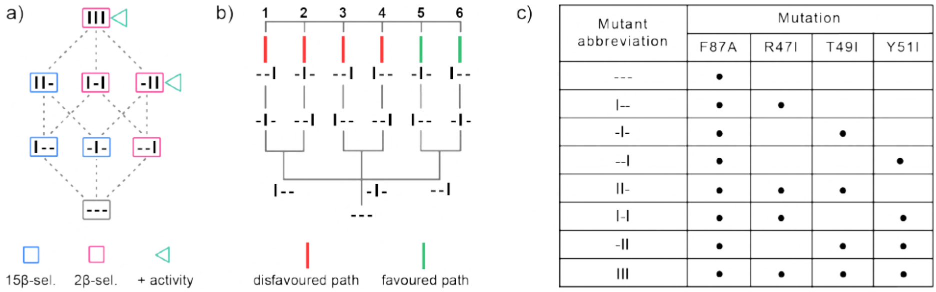


Figure 6.1. Enzymatic reaction modeled in this study. Testosterone (**1**) is selectively hydroxylated to 2β-hydroxystestosterone (**2**) by P450$_{BM3}$ monooxygenase mutant **III** (F87A–R47I/T49I/Y51I).

6.1.1 Multidimensional fitness landscape

To explore the step-wise accessibility in the evolution of parental F87A (indicated as **---** since no mutations were made) towards mutant **III**, the corresponding fitness landscape was constructed by generating all the intermediate variants connecting **---** to **III** by single amino acid mutations at each evolutionary step, for a total of 2^3=8 mutants as sequence space (Fig. 6.2a) and fitness defined by both activity and 2β-selectivity. Mutant **---** provided in a whole-cell system only ~20% conversion with formation of a 1:1 mixture of **2** and 15β-hydroxytestosterone (**3**), whereas the most 2β-selective variants (~67-91%) all contain mutation Y51I (**--I**, **I-I**, **-II** and **III**) while the remaining ones proved to be 15β-selective, with the total turnover frequency (TTF) being highest in mutants **III** and **-II** (158 and 72, respectively), and lower in all the others (for all the enzymatic parameters see Fig. C1, Appendix C).



Mutant abbreviation	Mutation			
	F87A	R47I	T49I	Y51I
---	•			
I--	•	•		
-I-	•		•	
--I	•			•
II-	•	•	•	
I-I	•	•		•
-II	•		•	•
III	•	•	•	•

Figure 6.2. Deconvolution of mutant **III** starting from the parent enzyme F87A (---). a) sequence space of the fitness landscape illustrating the 3!=6 possible evolutionary pathways (dashed lines) between parental F87A (---) variant and triple mutant **III**, involving all the single and double intermediate mutants. Each of the 2^3=8 mutants differ by a single mutation from neighbours connected through dashed lines. 15β- and 2β-selective mutants are depicted in blue and red frames, respectively. Mutant with enhanced activity are marked with a green triangle. b) Representation of all the possible combinations of mutations to construct the 6 different evolutionary pathways occurring in the step-wise DE of mutant **III**. Favored paths leading to higher fitness (i.e. enhanced activity and 2β-selectivity) and disfavored paths are indicated with green and red lines, respectively. c) Mutant abbreviations.

The fitness landscape was able to identify epistatic effect arising between the mutated residues, affecting all the enzymatic traits (for detailed data of all epistatic interactions and pathway accessibility see Table C1-C3, Appendix C). As an example, the introduction of mutation **-I-** alone in the first step has no or deleterious effect on 2β-selectivity and activity (the corresponding mutant are 15β-selective and poorly active) whereas mutation **--I** alone has a beneficial effect on both parameters. However, in the second step the combinations of **-I-** (which alone has deleterious effect) with **--I** (beneficial) yielding mutant **-II** has overall a beneficial effect on both activity and selectivity, indicating positive epistasis. In the last step combining **I--** (deleterious alone) with **-II** to obtain the **III** mutant further enhances all parameters. Additionally, epistasis was found to be pervasive and occurring in all the evolutionary pathways, which characterize the fitness landscape. Noteworthy, the mutated residues are all distal from the active site and thus could rarely interact directly with the substrate, therefore the effect of epistasis impacting the enzyme functions could be transmitted *via* long-range interactions modulated by enzyme conformational dynamics.

To investigate this hypothesis, this chapter presents a detailed study of the conformational changes occurring along the evolutionary trajectory going from parent **---** towards **III**, and how enzyme conformational dynamics is exploited during the step-wise evolution to fine-tuning protein functions such activity and selectivity at a molecular level.

6.2 Computational details

6.2.1 Quantum Mechanical calculations

In the computational model used for compound I (CpdI), the P450 protein-bound porphyrin complex coordinated with iron and containing the $Fe^{IV}(O)$ moiety, substituents at the periphery of the porphyrin ring were replaced with hydrogen atoms. Additionally, the axial Fe−SCys bond in CpdI was modeled as a $Fe-SCH_3$ bond. 2β- and 15β-hydroxylation by P450 consists of the hydrogen atom abstraction (HAA) respectively from the 2C–H and 15C–H bond of **1** by CpdI (which is the rate limiting step), followed by a radical rebound mechanism. In agreement with previous work[190] the HAA transition-state (TS) in the $S=3/2$ spin state was calculated to be consistently lower in energy than in the $S=5/2$ and $S=1/2$ states. Hence, all the reported Gibbs free-energies ($\Delta G^{\ddagger}$) correspond to the HAA on the quartet potential energy surface.

A comprehensive computational Density Functional Theory (DFT) study was performed with the Gaussian09(D.01)[191] program package using the B3LYP[127] hybrid functional under the unrestricted formalism (except for **1**, restricted singlet) and including Grimme's D_3 formalism to account for dispersion energy corrections (B3LYP-D_3).[139] Solvent effects were taken into account using the Polarizable Continuum Model[144] (PCM) with parameters for dichloromethane to simulate the enzyme dielectric permittivity and accounting for electronic polarization and small backbone fluctuations in the enzyme active site.[192]

Molecular orbitals (MOs) were expanded using three different sets of Gaussian-type orbitals (GTOs) consisting of (i) double-ζ quality split-valence basis plus one set of polarization functions (6-31G*); (ii) triple-ζ quality split-valence basis plus two sets of polarization functions (6-311G**) and (iii) double-ζ quality basis with associated effective core potential (ECP) by the Los Alamos National Laboratory (LANL2DZ).[193] In all calculations, the Fe metal and relativistic effects were described with the LANL2DZ basis set and relative ECP, whereas the other atoms were treated with 6-31G* basis set during geometry optimization and subsequent frequency calculation, or 6-311G** basis set for final energy single point calculation. This computational protocol has been extensively applied with success by the Houk Lab.[120]. Table C4 (Appendix C) summarizes electronic energies, thermal corrections and Gibbs free energies of the QM structures calculated for the HAA catalyzed by model CpdI.

6.2.2 Molecular Dynamics simulations

Long-timescale conventional molecular dynamics simulations (MD) in explicit water were performed using AMBER16 package[194] at our in-house GPU Galatea cluster. Testosterone (**1**) substrate parameters for the MD simulations were generated with the antechamber module of AMBER16 using the general AMBER force field (GAFF),[147] with partial charges set to fit the electrostatic potential generated at the HF/6-31G* level by the restrained electrostatic potential (RESP) model.[148] The charges were calculated according to the Merz-Singh-Kollman scheme[195] using Gaussian 09(D.01).[191] Parameters for the heme resting state, CpdI and axial Cys were taken from ref. [196]. Amino acid protonation states were predicted using the H++ server (http://biophysics.cs.vt.edu/H++).[197] Then, the enzyme was solvated in a pre-equilibrated truncated hexagonal box with a 10-Å buffer of TIP3P[142] water molecules using the AMBER16 leap module, resulting in the addition of ~13,000 solvent molecules. The systems were neutralized by addition of explicit counterions (Na^+ and Cl^-). All subsequent calculations were done using the widely tested Lindorff-Larsen modification of the Amber99 force field (ff99SBildn).[198]

For all the mutants, the apo structures used were generated from the P450-BM3 original variant (PDB: 1FAG[64]), by removal of the palmitoleic acid bound to the protein and the introduction of the corresponding mutations using the Rosetta-Backrub web-server (https://kortemmeweb.ucsf.edu/backrub).[199] Substrate-bound structures were produced using the former mutants generated with Rosetta-Backrub as template, by placing **1** molecule in the position originally occupied by the palmitoleic acid. Pose 2 and pose 15 were obtained by orienting **1** with its C2 or C15 atom towards the heme-moiety, respectively. The stability of all the aforementioned apo and substrate-bound structures was tested by conventional MD (cMD) simulations. A two-stage geometry optimization approach was performed. The first stage minimizes the positions of solvent molecules and ions imposing positional restraints on solute by a harmonic potential with a force constant of 500 kcal mol^{-1} Å^{-2}, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems are gently heated using six 50-ps steps, incrementing the temperature 50 K each step (0–300 K) under constant volume and periodic boundary conditions (NVT). Water molecules were treated with the SHAKE algorithm[156] such that the angle between the hydrogen atoms is kept fixed. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.[200] An 8-Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 10 kcal/mol were applied to the solute, and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated in the NpT ensemble without restraints for 2 ns with a 2-fs timestep at a constant pressure of 1 atm and temperature of 300 K. After the systems were equilibrated, for both apo and substrate-bound structures, three independent 600 ns cMD simulations (i.e. 1.8 µs accumulated simulation time) were performed under the NVT ensemble and periodic-boundary conditions.

Long-timescale accelerated Molecular Dynamics simulations (aMD)[201] were used to explore the binding trajectory of **1** in the different mutants. To allow the substrate to diffuse freely until being spontaneously recognized by the enzyme surface and finally getting in the access channel, we started with four different molecules of **1** randomly placed in the bulk solvent, performing three independent replicas, each one of 250 ns unconstrained cMD simulations (*vide supra*), from which the acceleration parameters were determined, followed by 750 ns of dual-boost aMD. A total boost potential was applied to all atoms in the system in addition to a more aggressive dihedral boost (i.e. E_{dih}, a_{dih}; E_{tot}, a_{tot}), within the dual-boost aMD approach. The acceleration parameters used were the following:

$$E_{\mathrm{dih}} = V_{\mathrm{dih\_avg}} + 2.0 \cdot N_{\mathrm{res}},\ a_{\mathrm{dih}} = 2.0 \cdot N_{\mathrm{res}} / 5$$

$$E_{\text{tot}} = V_{\text{tot\_avg}} + 0.16 \cdot N_{\text{atoms}}, a_{\text{tot}} = 0.16 \cdot N_{\text{atoms}}$$

Where N_{res} is he number of protein residues, N_{atoms} is the total number of atoms, and $V_{\text{dih\_avg}}$ and $V_{\text{tot\_avg}}$ are the average dihedral and total potential energies calculated from the first 250 cMD simulation, respectively.

The robustness of the acceleration parameters was ensured by checking that the investigated system conserved the correct folding along the aMD simulation.

6.2.3 Additional packages

The principal component 2 (PC2) shown in Figs. 6.4d, 6.5b and C4,C8 (Appendix C) was computed from the cumulated apo trajectory of the different mutants using the Principal Component Analysis (PCA) available in the Bio3D package,[202] considering Cα distances. The volumes of (i) the active site in mutants **---** and **III** (Fig C5, Appendix C) and (ii) the area of the access channel 2 in mutant **-I-**, **III** (Fig. 6.6c) were calculated with the POVME2 software,[203] with a 1.0 Å grid spacing and 1.35 Å distance cutoff. In (i) the calculation was done on the most populated clusters of the cumulated apo trajectories of the two mutants, with a sphere of 6 Å as the including region, whereas in (ii) the structures used were clustered from the conformational population analysis (*vide infra*) with a sphere of 4 Å as the including region. The conformational population analysis displayed in Figs 6.5c, 6.6a and C9 (Appendix C) was obtained by applying the dimensionality reduction technique Principal Component Analysis (PCA) on the whole dataset of the apo trajectories (three replicas of 600 ns cMD for each mutant, i.e. an accumulated simulation time of 14.4 μs) using the PyEMMA[204] software and considering the distances between all Cα. The structural conformational changes involved in PC1 of such population analysis were found clusterizing the trajectories of the two mutants exhibiting the greatest variation of PC1 (namely **-I-** and **III**). The Shortest Path Map (SPM) (section 2.5, *in silico* methods) in Fig. 6.6d was computed using the in-house Dynacomm.py code.[164a]

6.3 Results and discussion

6.3.1 Active site conformational changes required for regioselectivity are enabled by distal mutations

To gain insights about the origin of selectivity and activity, a suite of combined DFT and MD computational studies were performed. Given the identification of comparable reaction barriers for HAA from C2 and C15 of **1** using DFT calculations on heme compound I (CpdI) truncated models (ΔG differ <1.0 kcal mol^{-1}, Table C4, Appendix C), MD simulations were carried out for all mutants starting either from pose 2 and from pose 15 (i.e., positioning C2 or C15 of **1** closer to the Fe=O of CpdI, respectively) to analyse whether the binding pose of **1** in the active site determines the experimentally observed selectivity (Fig. 6.3 and Figs. C2-3, Appendix C).

The DFT-predicted (**1**)2/15C···O(Fe-heme) distance and (heme-Fe)O–H_β–C2/15(**1**) angle (where the C2/15 atom was considered for pose 2 or pose 15, respectively) of the TS were taken as ideal values (upper Fig. 6.3) and further monitored along the MD trajectories to generate the Kernel Density Estimation (KDE) plots and assessing the proficiency of the binding pose of **1** in all the mutants for undergoing 2C–H_β or 15C–H_β cleavage selectively. A substrate binding pose can be considered proficient and defined as near attack conformation (NAC)[205] when the reacting atoms assume angle and distance reseambling the bond to be formed in the TS (computed with QM using a reduced model, upper Fig 6.3).

Starting from the parent **---**, both pose 2 and pose 15 generated from manual dockings are possible. Further analysis of these binding poses along MD simulations in **---** indicate that substrate **1** in pose 15 explores NACs closer to the DFT-predicted ideal TS geometry for HAA, rather than in pose 2 (Fig. 6.3a), thus making pose 15 more productive towards 15β-hydroxylation. Introducing mutations R47I and/or T49I (i.e. in variants **I--** and **-I-**, respectively) does not have any effect on selectivity (Fig. 6.3b-c, e), i.e., the selectivity is retained due to the catalytically competent conformation inherent in pose 15 along MD simulations (pose 2 adopts a reduced number of catalytically competent conformations). However, the picture completely changes when mutation Y51I is introduced: the substrate bound in pose 15 becomes unstable in mutant **I--** and leaves the active site in 1 out of 3 replicas (ca. >15 Å C2··O distances explored), whereas pose 2 is highly stabilized and explores short C2··O distances for the incipient C–H eventually leading to 2β-hydroxytestosterone in 2 out of 3 replicas (Fig. 6.3d and Fig. C2-3, Appendix C). As experimentally determined, 2β-selectivity is retained in variants **I-I**, **-II** and **III** that contain mutation Y51I (Fig. 6.3f-h). This is even more dramatic in variant **III**, in which pose 15 is highly unstable and **1** rapidly rotates to position C2 close to the catalytic Cpd I for 2β-hydroxylation. At first, **1** is bound in pose 15, perpendicularly with respect the heme plane and interacting with T260 and A330 through its hydroxyl and carbonyl moieties, respectively. Gradually, **1** assumes a lying pose over its α-side, and the subsequent (**1**)CO-T327 and (**1**)OH-A87 interactions start driving a counterclockwise rotation of **1** inside the heme pocket. Afterwards, (**1**)CO-G265 followed by (**1**)OH-T327 interactions complete the 180° rotation of **1**, which ultimately assumes the final binding pose 2 perpendicular to the heme plane (Fig. 6.4a). On the other hand, starting the MD simulation of variant **III** with **1** bound in pose 2 indicates that this pose is stable along the whole trajectory, adopting NACs in all MD replicas (Fig. 6.3h and Fig C2-3, Appendix C).

Notwithstanding, mutant **-II** and **III** only differ for the R47I mutation in the latter case, yet the re-orientation of the substrate from pose 15 to pose 2 is observed only during the MD simulation of mutant **III**. To further investigate the specific effect of R47I mutation on substrate rotation inside the heme pocket, a Principle Component Analysis (PCA) was performed on the substrate-bound MD trajectories of mutant **-II** and **III**, finding that PC2 indeed describes an increased flexibility of residues A87, T260, G265 and T327 in mutant **III**, as compared to **-II** (Fig. C4, Appendix C). Thus, R47I may modulate via long-range conformational dynamic effect the flexibility of such residues, which have been shown to be instrumental to promote substrate re-orientation in mutant **III**. Moreover, mutant **III** presents a substantially wider active site pocket as compared to the other variants: the active site volume in the **---** variant is 89 Å^3, which is expanded to 235 Å^3 in **III** (Fig. C5, Appendix C). It can be hypothesised that in all variants, except **III**, selectivity must be determined by the orientation adopted by the substrate while accessing the heme cavity.

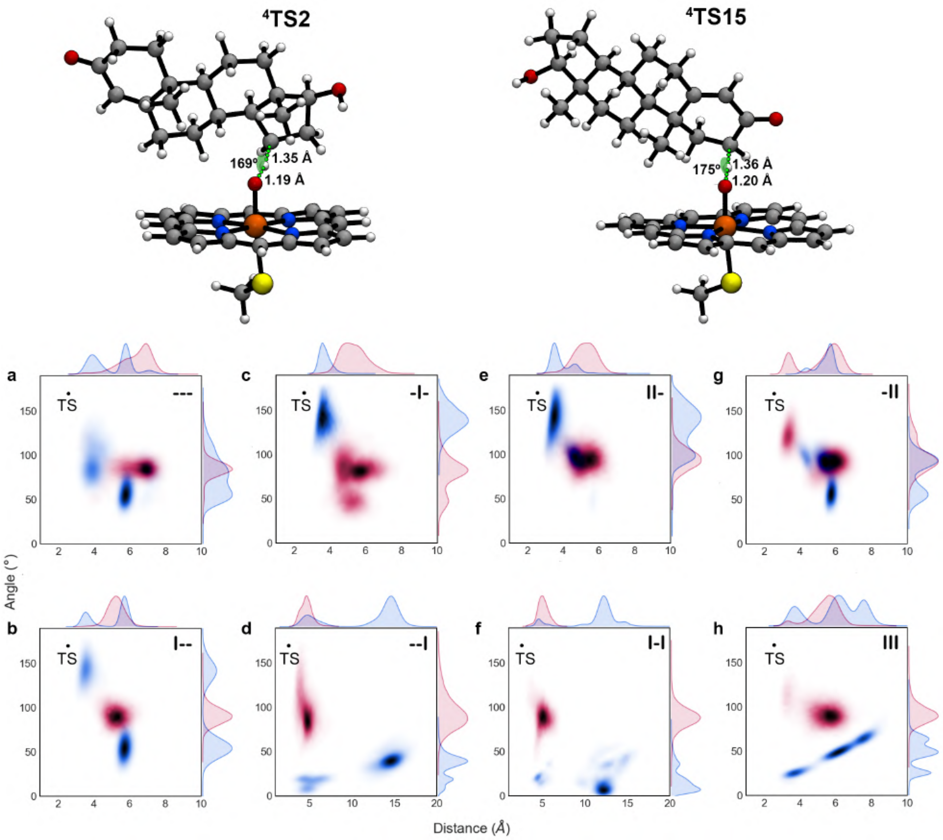


Figure 6.3. Upper figure: DFT transition states for the H-abstraction in **1** from C2 (left) and C15 (right). Lower figure: KDE plots of key geometric parameters for hydroxylation. Distances determined between the oxygen atom of heme-Fe=O and the C-atom of **1** (x-axis) and angles formed by (heme-Fe)O–H_β–C2/15(**1**) (C2 in pose 2 or C15 in pose 15) of **1** (y-axis) from the first replica of the MD dataset of all mutants (see Fig C3-C4 for the two additional replicas). Geometric parameters measured for C2 and C15 are shown in red and blue, respectively. The ideal distance and angle for the TS (black dot) corresponds to the DFT-optimized geometry of the transition state (TS) for the C–H abstraction by heme-Fe=O using a truncated computational model.

Recently, Mondal and co-workers characterized the substrate recognition and binding pathway in related P450cam using MD simulations, showing the formation of a single key channel in which the substrate needs to reside in a long-lived intermediate state before reaching the catalytic iron-oxo species.[206] To reconstruct the substrate binding process in $P450_{BM3}$, substrate **1** was placed in the bulk solvent to subsequently start 250 ns unbiased MD simulations followed by 750 ns accelerated Molecular Dynamics (aMD) simulations. Among independent MD and aMD trajectories for all variants, only a single trajectory by mutant **I--** was observed to be productive where the substrate reached the heme active site (Fig. 6.4b). A two-step binding mechanism can be observed in this trajectory: first, the carbonyl moiety of **1** enters channel 2a (Fig. 6.4c) and stays above the β1-2 strand, where residues R47I, T49I, Y51I are located, forming a long-lived substrate-enzyme bound intermediate. There, substrate **1** can reorient, although its access to the active site is restricted by the β4 sheet that acts as a gate. Second, a network of coupled conformational changes occur simultaneously: the G helix adopts a bend conformation, which impacts the F helix, F-G loop and β1 sheet conformation, and in turn retreats the β4 sheet, allowing **1** progression towards the catalytic center. A similar 2-step mechanism was observed Mondal and co-worker while mapping the substrate binding pathway in P450cam.[206]

The aMD simulations indicate that the orientation of the substrate when accessing the catalytic site during the second step of the binding pathway dictates selectivity. Once inside the heme pocket, substrate rotation was not observed in mutant **I--**, thus predicting that the orientation assumed by the substrate when entering the heme site ultimately governs selectivity. In fact, previous substrate-bound MD simulations suggest that only variant **III** has a sufficiently wide active site pocket for allowing substrate rotation. These findings indicate that in all variants except **III**, selectivity is determined by the orientation adopted by the substrate while entering the heme pocket. In the productive substrate binding trajectory corresponding to mutant **I--**, **1** accesses the heme with the correct orientation for 15β-hydroxylation (Fig. 6.4b). In this case, residue Y51 establishes a hydrogen bond with the carbonyl group of **1** (Fig. C6, Appendix C), constraining the substrate in a such way that it can only progress into the active site pocket pointing its C15 ahead towards heme-Fe=O.(51) Thus, Y51 is instrumental in promoting the observed C15-selectivity in **---**, and in **I--**, **-I-** and **II-** variants. It should be noted that in previous studies, R47 and especially Y51 were found to interact with the terminus end of long-chain fatty acids while bound at the P450$_{BM3}$ active site[64]. Such direct interaction with testosterone and Y51 is only possible at the pre-binding pocket, which is lost after the retreat of β4 sheet, allowing substrate access to the heme pocket. Additionally, the higher C2-selectivity observed in variant **III** occurs due to the flipping and motion of the β4 sheet, destabilizing pose 15 while favoring pose 2 (Fig. 6.4a).

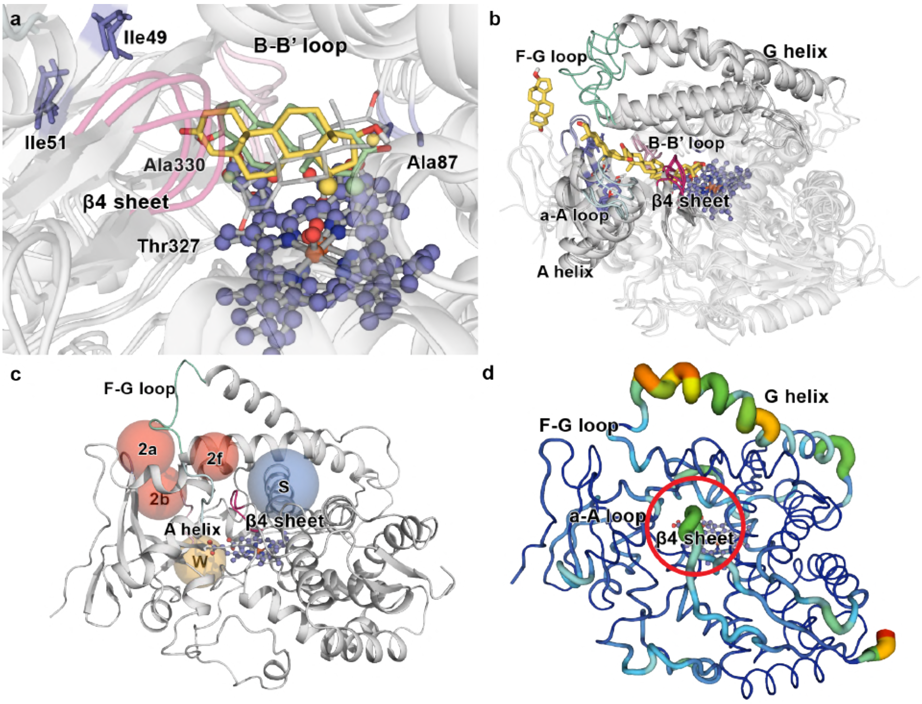


Figure 6.4. Secondary structural elements determining regioselectivity and activity. a) Rotation of **1** from pose15 to pose2 in the active site of mutant **III**. β Hydrogens belonging to C2 and C15 atoms are depicted in pale green and yellow color, respectively. b) Spheres indicate the channels observed in the WT crystal structure (PDB: 1FAG) and in the mutants of our MD simulations (red and blue color). c) Trajectory of **1** towards the active site of mutant **I--** and binding of **1** above the heme. d) Principal Component Analysis (PC2) of mutant **III** (apo). The thickness of the line is proportional to the motion and the color scale varies from blue (minimum motion) to red (maximum motion). The β4 sheet is highlighted with a red circle. Standard nomenclature for channels[77b] and secondary structure elements[207] is used.

These findings suggest that favoring a more efficient substrate gating in the active site and binding in a catalytically competent pose increases enzyme TTF, while NADPH leak is reduced due to a more efficient interaction between the substrate and the catalytically active Fe=O species once generated.

6.3.2 Conformational dynamics shape the evolution of the fitness landscape

The multidimensional fitness landscape explores the step-wise evolutionary accessibility of parent **---** towards **III** (Fig 6.5a). Therein, two kinds of trajectories can be distinguished: those lacking local minima (favored) and those characterized by at least one local minimum (disfavored). Pathways 1-4 are characterized by a decrease in both selectivity (measured as $\Delta\Delta G^{\ddagger}$, Additivity terms and equations, Appendix C) and activity (TTF) in the first step, indicating that they are evolutionary disfavored. On the other hand, pathways 5-6 are favored because mutation **--I** has a beneficial effect on both activity and selectivity (TTF = 11 → 21; $\Delta\Delta G^{\ddagger}$ = -2.3 → 1.8), whereas in the second step pathway 6 increases significantly both parameters towards **-II** and **III** (TTF = 21 → 72 → 158; $\Delta\Delta G^{\ddagger}$ = 1.8 → 4.4 → 5.6). Interestingly, when all other parameters are considered, pathways 5 and 6 are the only ones that remain accessible (Fig. 6.5a and Fig. C1, Tables C1-C2, Appendix C).

In the previous section (*vide infra*) the crucial role of the β4 sheet for gating the substrate access into the heme pocket was emphasised. Therefore, the mobility of this particular region directly impacts the activity of the enzyme, favoring the gating mechanism. Fig. 6.5b reports the progression of the β4 flexibility along the fitness landscape by means of PCA (PC2) showing that indeed only mutant **-II** and **III** feature higher flexibility compared to all other variant, with **III** exhibiting the highest.

To identify the most important conformational changes in all evolutionary pathways, and to describe how distal mutations influence selectivity, extensive MD simulations were performed in the absence of substrate of each variant and the dimensionality reduction technique PCA applied to the whole dataset (three replicas of 600 ns MD for each mutant, i.e. an accumulated simulation time of 14.4 μs) considering the distances between all Cα. The conformational population analysis resulting from all the accumulated simulation data was projected in the principal components space outlined by PC1 and PC3, which describe the first and third most important conformational differences among all variants (for PC1/PC2 space see Fig. C9, Appendix C). Notably, a clear distinction between 2β- and 15β-selective mutants is revealed (Fig. 6.5c and 6.6a) through their separation with respect to PC1 (x-axis), suggesting that changes in selectivity are linked to the impact that the introduced mutations have on the enzyme conformational dynamics. Similarly to how the β4 sheet correlates to activity, Fig. 6.5c reports the conformational population progression of the different variants at each step mutational step of the fitness landscape, where 10 out of 12 total replicas of 2β-selective mutants (**--I**, **I-I**, **-II** and **III**) lying at positive values of PC1, while the remaining 15β-selective mutants lie at negative PC1 values in 9 out of 12 total replicas. These two "theoretical fitness landscapes" (Fig. 6.5b-c) nicely show how the conformational changes in the enzyme are committed to the evolution of activity and 2β-selectivity.

To further evaluate the physical meaning of the conformational changes involved in PC1 in Fig. 6.5c and 6.6a, the following procedure was followed: (i) the trajectory of the two mutant **-I-** (replica2) and **III** (replica3) were chosen as they exhibit the major variation in conformation along PC1, being located at its minimum and maximum values, respectively; (ii) the PCA dataset of each of the two mutant were clustered into 200 clusters; (iii) the cluster located closer to the minimum of the conformational populational analysis was chosen as representative structure for the overlay. These conformational changes related to selectivity mainly involve the G helix, the F-G loop, the β1 hairpin and the B' helix (located at the entrance of the 2a/b channels) as well as the a-A loop and the β4 sheet (located at the entrance of the 2f channel) (Fig. 6.6b). In variant **-I-**, the channel 2a has a narrower substrate access entrance due to a closed state of the F-G loop (*ca.* 9.3 Å determined between the Cα of R47 and N192).

Conversely, the combination of mutations introduced in **III** favors an open conformational state of the same F-G loop (*ca.* 12.4 Å measured between the Cα of I47 and N192), enlarging the access channel 2a, which is mainly responsible for allowing access to the enzyme binding pocket (Fig. 6.6c). Indeed, the area surrounding the access channel 2a in mutant **III** is calculated to have a volume of 140 Å^3 with respect to 44 Å^3 in -I-.

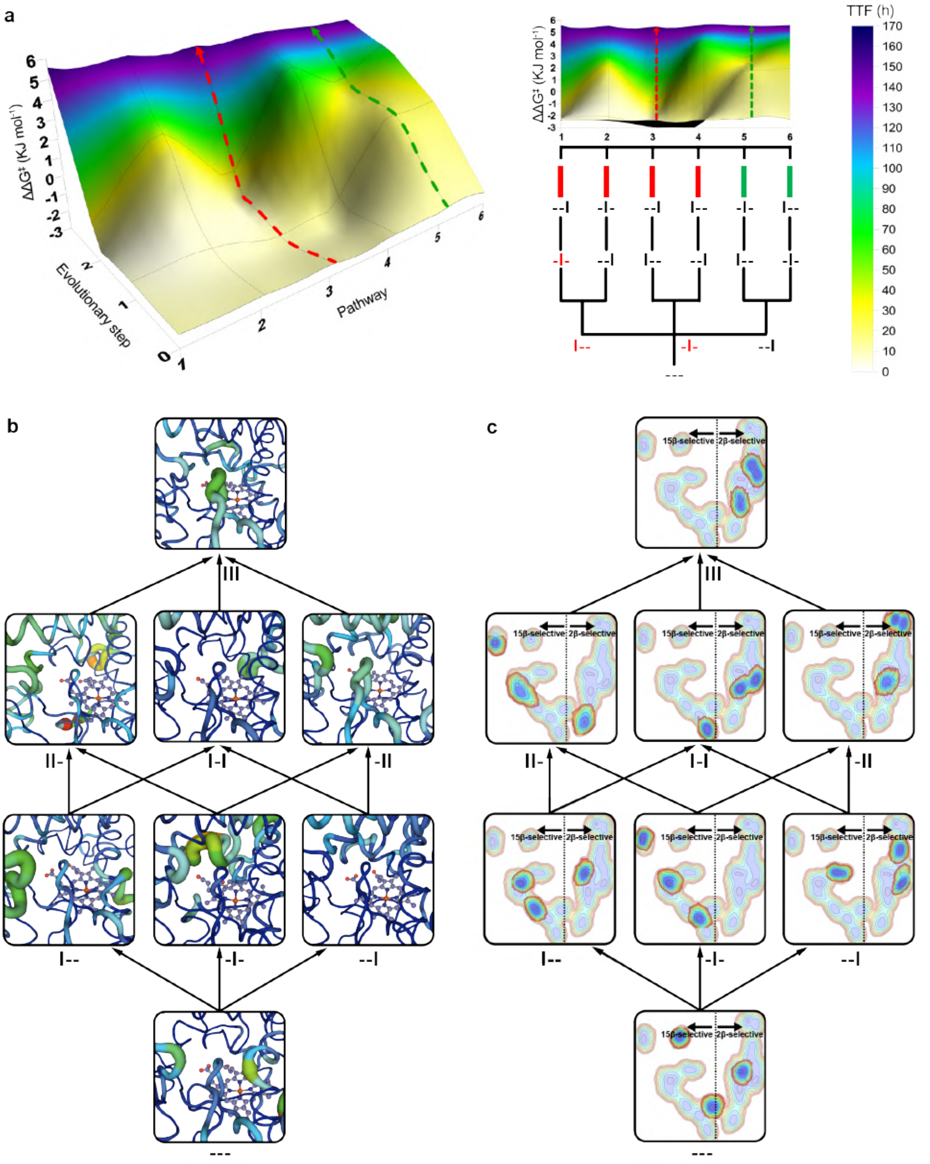


Figure 6.5. Stepwise evolution of multiple functions and conformational dynamics. a) Multiparametric fitness pathway landscapes of the 6 evolutionary pathways leading from parent mutant --- upward to mutant **III** are shown in 3D (left) and frontal (right) format (green and red arrows indicate examples of favorable and unfavorable pathways, respectively). Fitness is defined by activity as total turnover frequency (TTF) displayed as heat-maps from 0 (white) to 170 (blue) as well as 2β-selectivity as $\Delta\Delta G^{\ddagger}$ in kJ per mol (y-axis) at each evolutionary step (z-axis) for each pathway (x-axis). Green and red bars indicate favored and disfavored pathways, respectively. The mutant in red represents steps with disfavored energy, i.e. the point where the pathway is blocked. The data represent the →

→ average of two independent experiments (n=2) not part of the work carried out in this thesis, and are presented only for reference. b) Progression of the β4 sheet flexibility along the 6 pathways as revealed by PC analysis (PC2) of the substrate-free simulations of mutated enzymes analysed separately. The thickness of the line is proportional to the motion and the color scale varies from blue (minimum motion) to red (maximum motion). c) Evolution of the conformational dynamics along the 6 pathways and its connection to 2β- or 15β-selectivity. The analysis of the global conformational dynamics of the substrate-free simulations of mutated enzymes, as shown by PC1/PC3, indicate that 2β- and 15β-selective mutants explore conformations lying at positive and negative values of PC1, respectively. Color scale varies from red (less populated) to blue (more populated).

To finally unravel the link between epistasis and conformational dynamics, the SPM analysis[164a] is applied using the accumulated 1.8 μs MD simulation performed on parent **---** in the absence of substrate. SPM considers the different conformations that the enzyme samples along the MD simulation, and identifies which residues are those that are more important for the observed conformational changes, which in this case are associated with different selectivities and activities. In the parent **---** enzyme, the generated SPM identifies residue Y51 (as well as V78 and A330, known from earlier studies[69]) to be important for enzyme activity and selectivity, and to be interconnected in terms of Cα correlated movements, thus highly contributing to the enzyme inactive-to-active conformational interconversion. This highlights why these three distal positions are found to be key during the evolutionary pathway for improving catalysis, in line with what we observed for the laboratory-evolved retro-aldolases.[164a] Importantly, the SPM also describes strong connections between all five-stranded β1 sheet hosting the mutated residues with the β4-2 strand and the B' helix, which we showed to be crucial for substrate binding and gating (Fig. 6.6d). This long-distance communicating pathway between β1 and β4 sheets directly relates the mutated positions on β1-2 strand (positions R47, T49 and Y51) and the increased flexibility of the β4 sheet. This shows how evolutionary pathways take advantage of networks of residue-residue interactions to fine-tune the conformational dynamics along the evolutionary pathways for improving enzyme function.

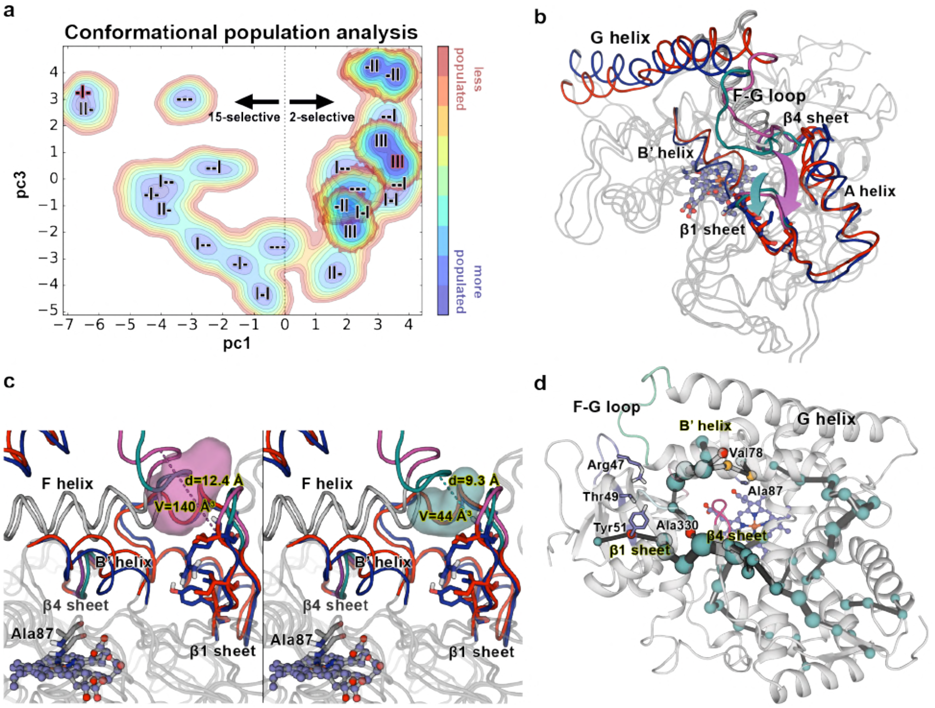


Figure 6.6. Analysis of the conformational dyamics of deconvoluted mutants. a) Conformational population analysis built from the combined PCA of substrate-free simulations of all mutants. The conformational populations of 2β-selective mutants **-II** and **III** are highlighted. All replicas (3/3) of **III** and **-II**, 2/3 replica of **I-I** and **--I** lie on positive values of PC1 (10 out of 12 replicas of 2β-selective mutants). All replicas (3/3) of **-I-**, 2/3 replicas of **---**, **I--** and **II-** show negative values of PC1 (9 out of 12 replicas of all 15β-selective mutants). The replica of **III** and **-I-** selected to show the conformational changes involved in PC1 (b) are outlined in red. b) Overlay of the conformational changes involved in PC1. c) Zoom of (b) showing the channel 2a cavity of mutant **III** (left panel) and **-I-** (right panel). In (b), (c), Mutant **-I-** is shown in blue, whereas the evolved mutant **III** is colored in red. The F-G loop, β1 sheet and channel 2a cavity are highlighted in teal and magenta for **-I-** and **III**, respectively. d) Analysis of the most important correlated motions of mutant **---** by means of the shortest path map (SPM). Mutational hotspot used in this or precedent work[69] that appear in the SPM are highlighted with red spheres, whereas positions adjacent to mutational hotspot are highlighted with orange spheres.

Chapter 7: Regio- and stereoselective steroid hydroxylation at C7 position by cytochrome P450$_{BM3}$ monooxygenase mutants

7.1 State-of-the-art

Steroidal C7β alcohols and their respective esters have attracted particular interest, showing significant promise as neuroprotective and anti-inflammatory agents[208] against acute and chronic neuronal damage caused by adverse events such as stroke, brain trauma and cerebral ischaemia. Unfortunately, the C7 group of steroids is spatially distant from any functional group that could direct C–H activation, and the synthesis of these compounds requires tedious multistep chemical processes,[209] making this class of therapeutic drugs not readily accessible using modern organic synthesis techniques.

Biocatalytic routes to steroidal C7β alcohols thus represent an appealing alternative, considering that novel biocatalysts are obtainable with enzyme engineering, which can ideally enable the targeted steroid hydroxylation at any desired position with high stereoselectivity and activity. Yet, in practice when focusing on the steroidal B ring, this approach has not been expedient to date because of low yields[71a, 90, 208a-c, 210] or the unfortunate formation of mixture of regioisomers.[208a]

Using testosterone (**1**) as a model substrate in different DE campaigns, cytochrome P450$_{BM3}$ monooxygenase mutants were obtained capable of oxidizing **1** with high activity and selectivity at position C2β, C15β[69] and C16α/β[92] in a targeted manner. The amount of screening effort involved about 18,000 bacterial colonies for selecting two active variants displaying either 2β- and 15β-selectivity, and 3,000 colonies using advanced DE approaches for two active variants with C16α- and C15β-selectivity. In a successive study,[211] Reetz and co-workers found that the triple mutant F87G/A328G/A330W was capable to hydroxylate **1** at position 7β (**2**) with about 3% selectivity and reasonable activity. This triple mutant was subsequently used as a template for further mutagenesis with the aim of improving both activity and 7β-selectivity. Focused mutagenesis on 15 residues near or at the active site was performed with a novel protein library construction method based on DNA assembly and USER (Uracil-Specific Excision Reagent), which produced a reasonable library-size (2^{15}) that otherwise would have resulted astronomical (15^{32}) if created with traditional NNK-based saturation mutagenesis.

Screening of only 1,600 variants (ca. 5% library coverage) led to the identification of the **LG23** variant, showing an outstanding C7β-selectivity improvement from 3% to 83% and total diastereoselectivity. Scaling up the reaction (50 mL) allowed further improvement of substrate conversion (>99% within 5h) and selectivity (90%). Although displaying high activity for the hydroxylation of **1** at position C7β, variant **LG23** was not able to accept the cognate substrate progesterone (**1b**). In addition, the X-ray structure of the monooxygenase domain harboring **1** could not offer structural insights on the regioselectivity, bearing an imidazole molecule from the protein purification process coordinated to the heme iron at the sixth position, and consequently pushing **1** away from the iron atom.

Therefore, the aim of this study is to provide at first a solid computational model to unraveling the molecular basis of the selective hydroxylation of **1** at the C7β position in mutant **LG23**, to subsequently offer a computational guide for further rational mutagenesis. The final aim is to achieve the 7β-hydroxylation also of substrate **1b**.

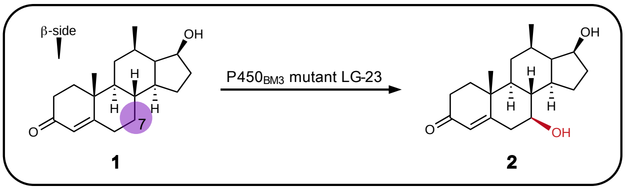


Figure 7.1. Enzymatic reaction modeled in this study. Testosterone (**1**) is selectively hydroxylated to 7β-hydroxystestosterone (**2**) by P450$_{BM3}$ monooxygenase mutant **LG23** (F87G/A328G/A330W–R47W/S72W/F77Y/V78L/F81I/A82L/T88S/M177T/M185Q/L188Q/I209T

7.2 Computational details

Density Functional Theory (DFT) calculations were performed using the same software, computational details and CpdI model discussed in section 6.2, except that in this case the modeled reaction was the CpdI-mediated hydrogen atom abstraction (HAA) from the 7C–H bond of **1** yielding the corresponding transition state (TS) structure. Energies, thermal correction and Gibbs free energies resulting from Density Functional Theory (DFT) calculations for the reactant complex (RC) CpdI + **1** were taken from the previous study (Chapter 6, Table C4, Appendix C).[212]

Molecular Dynamics (MD) simulations were carried out using the same software, computational details and methodology discussed in section 6.2, except that in this case the starting pose 7 was obtained by orienting **1** (**1b**) with its C7 atom towards the heme-moiety. For **1** bound to mutant **LG23**, three independent 1000 ns cMD simulations (i.e. 3 µs accumulated simulation time) were performed, whereas for the substrate-free **LG23-MGS** variant (**LG23** + Q185M/L437G/T438S) one 500 ns cMD simulation was performed, in all cases under the NVT ensemble and periodic-boundary conditions.

Docking studies to evaluate the binding pose of **1b** in the computationally designed **LG23-MGS** variant were done with AutoDock Vina[213] and performed in the most populated cluster structure resulting from the MD trajectory of **LG23-MGS** using a rigid-receptor approach for the residues surrounding the active site.

7.3 Results and discussion

To investigate the molecular basis of the improved activity and 7β selectivity of the LG-23 mutant, combined DFT and MD simulations with **1** bound in pose 7 (presenting C7 close to the heme) were performed. The DFT transition state (TS) of the HAA from C7 was identified at $\Delta G = 15.5$ kcal mol^{-1} (Table D1, Appendix D), only slightly higher in energy than those found for cleaving the $2C\text{–}H_\beta$ and $15C\text{–}H_\beta$ bonds[212] (Chapter 6, Table C4, Appendix C), indicating that the cleavage of the $7C\text{–}H_\beta$ bond is in principle feasible. The DFT-predicted (**1**)7C···O(Fe-heme) distance (2.5 Å) and (heme-Fe)O–H_β–C7(**1**) angle (167°) of the TS (Fig. D1, Appendix D) were taken as ideal values and further monitored along the MD trajectories (*vide infra*) to evaluate the proficiency of the binding pose of **1** in the **LG23** mutant for undergoing to 7C–H cleavage selectively.

MD simulations with **1** bound in **LG-23** were first compared with previous results obtained for the F87A variant (Chapter 6).[212] Mutant F87A was shown to accept **1**, leading to about 20% conversion in a similar whole-cell system, with formation of a 1:1 mixture of 2β- and 15β-hydroxytestosterone. MD studies on the F87A variant have shown that **1** can effectively explore two different binding poses, in which the substrate is oriented alternatively with its hydroxyl group towards the β1-4 strand and the carbonyl group pointing along the I helix (pose 2, presenting the C2 carbon close to the heme) or vice versa, with the hydroxyl group facing the I helix and the carbonyl towards the β1-4 strand (pose 15, presenting C15 close to the heme). On the other hand, the MD trajectories of **1** bound in the LG-23 variant reveal a different binding pose that allows the substrate to selectively undergo 7β hydroxylation (Fig. 7.2). Indeed, in this case **1** is placed perpendicularly with respect to pose 2 or 15, that is, the carbonyl and the hydroxyl group are pointing to the B'C loop and the β4 sheet, respectively (Fig. D2, Appendix D).

Interestingly, in the **LG-23** mutant the replacement of alanine by the bulkier tryptophan (A330W mutation) would result in a steric clash with **1,** that is consequently reoriented along the β1-4 strand. At the same time on the B'C loop, the mutation of phenylalanine to glycine F87G (rather than to alanine as in mutant F87A), combined with the T88S mutation, removes the steric hindrance because of the presence of the methyl groups in F87A and T88, thus allowing **1** to be oriented along the B'C loop. Moreover, MD simulations show that in **LG-23**, both the G87 and S88 amino groups are involved in hydrogen bonds with the carbonyl group of **1** ((**1**)CO···HN(G87) = 2.3±0.4 Å and (**1**)CO···HN(S88) = 2.6±0.5 Å mean distances, Fig. D3 a and b, Appendix D), whereas on the β4 sheet, the T438 amino group interacts with the hydroxy group of **1** ((**1**)HO···HN(T438) = 3.4±0.9 Å mean distance, Fig. D3c, Appendix D). Hence, hydrogen bond interactions with the B'C loop and the β4 sheet hold the substrate in pose 7, which is stable along the whole MD trajectory and shows a mean distance (**1**)7C···O(Fe-heme) of 3.8±0.3 Å, and a mean angle formed by (heme-Fe)O–H$_\beta$–C7(**1**) of 147±18° (Fig. 7.2).

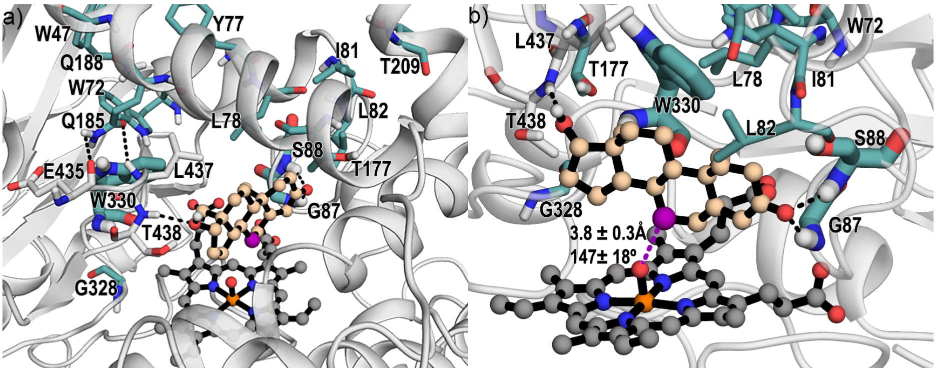


Figure 7.2. Representative snapshot of the MD simulation of **LG23** variant. a) Active site of LG23 with **1** bound (show in light orange). The most important polar interactions are marked with a black dashed line. C7 of **1** is colored in violet, whereas the mutated residues are shown in teal color sticks. b) Zoom of the active site of **LG23** showing the H-bond between **1** and G87 and T438 residues (black dashed lines) together with the mean value and standard deviation for the (**1**)7C···O(Fe-heme) distance and the (heme-Fe)O–H$_\beta$–C7(**1**) angle along the whole simulation time.

These values are close to the predicted DFT ideal TS geometry for the HAA from C7 (Fig. D1b, Appendix D). Remarkably, pose 7 is also the only catalytically competent pose, exploring NAC[205] compared to the closest adjacent C6 and C15, which show similar (**1**)C···O(Fe-heme) distances but non-optimal (heme-Fe)O–H–C(**1**) angles to undergo C−H abstraction (Fig. 7.3).

Noteworthy, the β4 sheet in mutant LG-23 is more displaced inwards to the active site with respect to mutant F87A (Fig. D2, Appendix D), facilitating the interaction between T438 and the hydroxyl group of **1**. Such displacement may be triggered by the polar interactions between i) the carbonyl group of Q185 and the amino group of leucine L437 ((Q185)CO···HN(L437)=2.8±1.0 Å mean distance, Fig. D4a, Appendix D); and ii) the amino group of Q185 and the carbonyl group of E435 ((Q185)NH···OC(E435)=4.5±2.2 Å mean distance, Fig. D4b, Appendix D). In fact, mutant F87A presents a methionine at position M185, whose apolar methyl sulfide group is unable to establish the same interactions and, consequently, to drag the β4 sheet inwards as mutant **LG-23** does in which M185 is mutated to glutamine (M185Q).

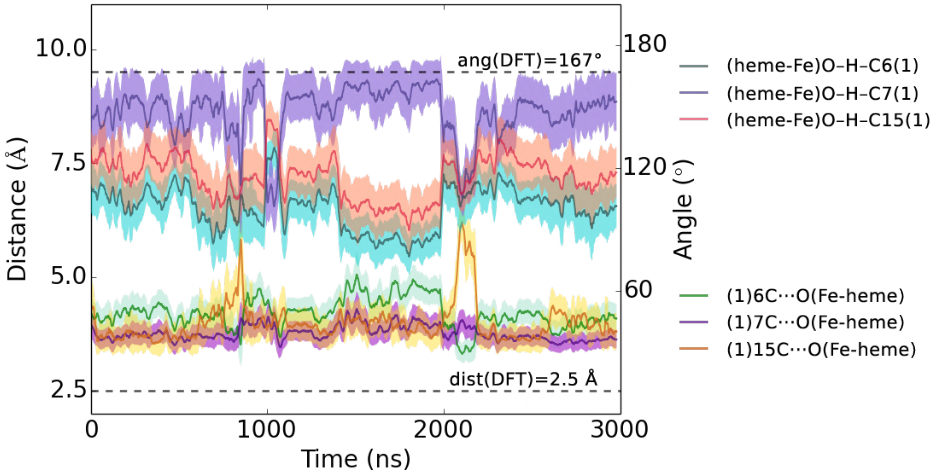


Figure 7.3. Plot of the (**1**) C6,C7,C15··O=Fe(heme) distances (y primary axis) and the (heme) Fe=O–H–C6,C7,C15 (**1**) angles (y secondary axis) along the simulation time (x axis) for the three MD replicas. Ideal DFT distances (Å) and angles (°) corresponding to the TS structure are shown with dashed lines.

Overall, the MD trajectories of mutant **LG23** reveal how substrate **1** is constrained by hydrogen bond interactions within the B'C loop and the β4 sheet, and is thereby enforced in pose 7 which is required to selectively hydroxylate at position C7β. Moreover, the hydrogen bonds between **1** and the β4 sheet are facilitated by the inward displacement of this latter within the active site due to the participation in polar interactions with Q185 on the F helix.

The computational model of **LG23** and the comparison with that of variant F87A additionally offer important insights for guiding further mutagenesis to broaden the substrate scope of **LG23** in order to accept the cognate substrate **1b** and to achieve its C7β hydroxylation. Noteworthy, **1** and **1b** only differ for the presence in the latter of an acetyl group at C17, rather than the hydroxyl group of **1** (Fig. 7.4 a and b). Visual inspection of **1b** bound in pose 7 within the active site of mutant **LG23** evidences that the acetyl group hindrance may prevent the correct binding in pose 7, as that would result in steric clashes with the β4 sheet (Fig. 7.4b). In this particular case the inward position of the β4 sheet in the active site by anchoring to the F helix will likely increase the steric encumbrance against the acetyl group of **1b**. Therefore, in order to remove the anchoring of the β4 sheet and to reduce the congestion at the active site towards **1b** acetyl group, a computationally designed variant based on **LG23** was conceived, where the polar Q185 residue was recovered to the original methionine hydrophobic sidechain (mutation Q185M) and residues L437 and T438 were mutated to less hindering glycine and serine (mutation L437G and T438S, respectively).

The MD simulation of this novel, computationally designed **LG23-MGS** variant revealed as expected that the hydrophobic sulfide group in M185 cannot establish polar interactions with the β4 sheet, which is ultimately released from anchoring to the F helix and consequently displaced outward the active site (Fig. 7.4c). This creates the required space to accommodate **1b** while preserving other important residues which are instrumental to bind correctly the steroidal substrate in pose 7 to undergo C7β hydroxylation. Final docking of **1b** in the most populated cluster resulting from the MD trajectory of **LG23-MGS** gave the substrate placed correctly in pose 7 as best scoring pose, with a binding affinity of -8 kcal mol^{-1}, confirming that **LG23-MGS** may be able to accept progesterone for converting it selectively to 7β-hydroxyprogesterone.

This new mutant is currently being tested in the Laboratory of Biocatalysis and Enzyme Engineering of the Hubey University by Prof. A. Li, showing promising results and demonstrating how, where possible, the use of computational models at key steps of the DE process workflow and the lessons learnt by computer simulations can guide and aid the development of novel enzymes with desired functions, reducing the number of libraries generated in sequential rounds of random mutagenesis and thereby relieving the screening effort.

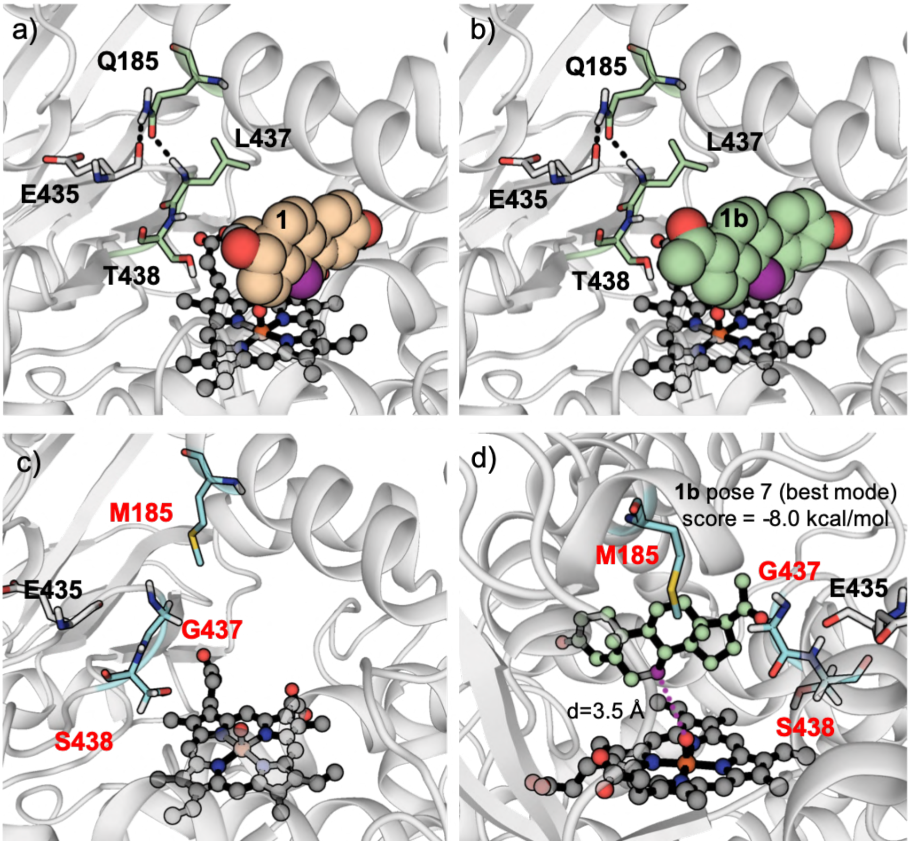


Figure 7.4. (a) Representative snapshot of the MD trajectory of **LG23** with **1** (light orange spheres) bond in pose 7; (b) representation of **1b** pose 7 in **LG23**, showing the steric clashes of **1b** acetyl group and β4 sheet residues (L437, T438); (c) most populated cluster of the MD trajectory of the computationally designed **LG23-MGS** mutant (substrate-free); (d) docking of **1b** in the active site of **LG23-MGS**.

Conclusions

Nature's catalytic machinery has inspired the scientific community for centuries. Leveraging our understanding of enzymes and biomimetic homogeneous catalysis through the combination of theory and wet labs offers key advantages, markedly towards the functionalisation of unactivated C–H bonds, which occur ubiquitously in organic molecules and remain largely inaccessible through synthetic chemical methods. In particular, the research illustrated in this thesis emphasizes that computational chemistry constitutes a powerful approach, especially when coupled to experiments. Principles identified herein, by characterising the reactivity of bioinspired iron-oxo species and the regio- and stereoselectivity of P450 laboratory variants, could be further used to fine-tune new metaloxo species and engineering novel enzymes towards desired functions for biocatalysis. The main conclusions that can be drawn from this thesis work are organized per chapters as it follows:

Chapter 4: the analysis of the charge displacement function (CDF) applied to lipoxygenase and taurine dioxygenase (TauD-J) model systems clearly demonstrates that cPCET and HAT mechanisms can be distinguished during the first step of hydroxylation in a chemically intuitive manner. The analysis matches perfectly the expected reaction channels and electronic rearrangements taking place at the transition state (TS) during the hydrogen atom abstraction steps. In TauD-J, an α-spin electron is transferred from the σ_{CH} orbital to the σ^* (Fe d_{z^2}–O p_z) anti-bonding orbital of the Fe=O unit to form the new O–H bond via a $^5\sigma$-channel, while the other electron proceeds from the β-spin manifold of the :O(Fe) lone pair. Overall, the electron travels together with the proton purely as a hydrogen atom, thereby constituting a HAT reaction. In contrast, in lipoxygenase a β-spin electron moves from the σ_{CH} orbital to the $\pi^*_{xy/yz}$ (Fe $d_{xz/yz}$–O $p_{x/y}$) anti-bonding orbital of the Fe–OH unit via a $^6\pi$-channel, populating the metal-based d-orbitals as typically seen for PCET events, whereas the newly formed O–H bond is created with both electrons proceeding from the :OH(Fe) lone pair.

Chapter 5: the different reactivity of the oxoiron(IV) complex series bearing TMC and TMC-based ligands can be mainly accounted for by the diverse extent of ligand field exerted by equatorial atoms. Either the weaker or stronger donating power of these latter and their distance with respect to the Fe=O moiety play a role, by shifting the orbitals which determine the spin ground state and those accepting the electron (electrophilicity) from the C–H bond undergoing cleavage, thereby ultimately influencing the preferred reaction channel. Nevertheless, considering only the favoured reaction channel and orbital overlap does not explain the reactivity difference in some particular cases. Conversely, the CDF analysis shows a clear correlation between HAT barriers and the amount of charge displaced on the H atom that is abstracted from the C–H bond, which overall can be related to the degree of C–H bond cleavage. Less cleaved C–H bonds indicate earlier TSs that resemble the reactant complexes (as indicated by the Fe=O···H–C distances of the compounds studied) which according to the Bell–Evans–Polanyi principle correspond to lower reaction barriers. Therefore, the HAT reactivity of oxoiron(IV) complexes does not originate only from the high electrophilicity of the FeO unit (low-lying EAO) but is rather the result of an interplay of factors which embrace both the frontier molecular orbitals and the driving force principle.

Chapter 6: the molecular basis of the targeted activity, regio- and stereoselectivity by laboratory evolution originates from pervasive epistasis occurring along the evolutionary trajectory of the P450-BM3 mutant F87A–R47I/T49I/Y51I (**III**) from parental variant F87A (**---**). Epistatic effects between the mutated residues that are distal from the active site, propagate via a long-range communication network to optimize the targeted function through the fine-tuning of the conformational dynamics of loops, helices and β-strands that gate the substrate access channel improving the enzymatic activity and reshape the active site determining the selectivity. The computational exploration of the fitness landscape successfully unravels that (i) only the two most active mutants **III** and F87A–T49I/Y51I (**-II**) exhibit higher flexibility of the gating β-sheet, therefore leading to higher activity and (ii) the conformational population of 2β-selective mutants is distinct from that of 15β-selective ones and it is characterized by positive values of the first principal component (PC1) corresponding to an open state of the F-G loop and channel 2a, whereas the latter lie at negative values and features a closed conformation of the same elements.

Chapter 7: the molecular basis of regio- and stereoselectivity at C7β targeted in the laboratory-evolved P450-BM3 variant **LG-23** results from specific mutated residues within the active site that allow correct positioning of the testosterone substrate and stabilize its binding pose. In particular, the M185Q mutation is found to be instrumental, as it promotes the inwards displacement of the β4-sheet which consequently contributes, among other pivotal interactions, to the catalytically competent pose of testosterone. Nevertheless, the inward displacement of the β4-sheet concurrently hinders the binding of bulkier steroids as progesterone. Therefore, as rational strategy, the recovery of the original methionine residue at position 185 is proposed (Q185M mutation), in conjunction with mutations to less hindering residues at key positions of the β4-sheet. Docking studies of the resulting, computationally designed **LG23-MG**S variant, show the capability of this novel mutant to correctly bind progesterone retaining the same regio- and stereoselectivity for leading to C7β-hydroxylation.

Overall, the studies reported in this thesis show two sides of the coin when investigating metalloenzymes, where an accurate description is needed for: (i) the electronic structure of the active species, and (ii) the dynamical correlations of amino acid residues' movements coupled to catalysis. At present it is impossible to combine these two approaches within one overarching set of calculations, because of: (i) the computational cost needed for the electronic structure, and (ii) the long-term dynamics simulations needed for describing the dynamical correlations. Conclusions can be drawn for each of these two sides of the coin, i.e. from chapters 4-5 for the electronic structure, and from chapters 6-7 for the molecular dynamics simulations. The results of this thesis may lead to future studies where an attempt might be made to combine these two sides, if either a simplified yet accurate method is obtained for the electronic structure, or if the available computational power has advanced to such an extent that a full QM/MM molecular dynamics approach is feasible with long-term simulations (microseconds).

Bibliography

1. (a) U. M. Zanger, M. Schwab, *Pharmacology & Therapeutics* **2013**, *138*, 103; (b) F. P. Guengerich, *Toxicological Research* **2020**; (c) F. P. Guengerich, *Chem. Res. Toxicol.* **2001**, *14*, 611.

2. Z. Fu, J. Chen. Xenobiotic Metabolism by Cytochrome P450 Enzymes: Insights Gained from Molecular Simulations. In: *Advances in Computational Toxicology: Methodologies and Applications in Regulatory Science*. Springer International Publishing (2019).

3. M. Szaleniec, A. M. Wojtkiewicz, R. Bernhardt, T. Borowski, M. Donova, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 8153.

4. (a) P. L. Roach, I. J. Clifton, C. M. H. Hensgens, N. Shibata, C. J. Schofield, J. Hajdu, J. E. Baldwin, *Nature* **1997**, *387*, 827; (b) K. Valegård, A. C. T. van Scheltinga, M. D. Lloyd, T. Hara, S. Ramaswamy, A. Perrakis, A. Thompson, H.-J. Lee, J. E. Baldwin, C. J. Schofield, J. Hajdu, I. Andersson, *Nature* **1998**, *394*, 805; (c) N. I. Burzlaff, P. J. Rutledge, I. J. Clifton, C. M. H. Hensgens, M. Pickford, R. M. Adlington, P. L. Roach, J. E. Baldwin, *Nature* **1999**, *401*, 721; (d) Z. Zhang, J. Ren, D. K. Stammers, J. E. Baldwin, K. Harlos, C. J. Schofield, *Nat. Struct. Biol.* **2000**, *7*, 127; (e) J. R. Cupp-Vickery, T. L. Poulos, *Nat. Struct. Biol.* **1995**, *2*, 144.

5. S. B. Wedde, Marc; Choi, Ji Eun; Oike, Keiko; Zumbrägel, Nadine; Gröger, Harald. The Recent Developments of Enzymatic Oxidation. In: *Green Oxidation in Organic Synthesis*. John Wiley & Sons, Ltd (2019).

6. (a) R. D. Jones, D. A. Summerville, F. Basolo, *Chem. Rev.* **1979**, *79*, 139; (b) M. Y. M. Pau, J. D. Lipscomb, E. I. Solomon, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18355; (c) E. G. Kovaleva, J. D. Lipscomb, *Nat. Chem. Biol.* **2008**, *4*, 186; (d) S. Sahu, D. P. Goldberg, *J. Am. Chem. Soc.* **2016**, *138*, 11410.

7. P. M. Wood, *Biochem. J* **1988**, *253*, 287.

8. (a) W. Nam, *Acc. Chem. Res.* **2007**, *40*, 465; (b) M. Costas, M. P. Mehn, M. P. Jensen, L. Que, *Chem. Rev.* **2004**, *104*, 939; (c) C. E. Tinberg, S. J. Lippard, *Acc. Chem. Res.* **2011**, *44*, 280; (d) A. J. Jasniewski, L. Que, *Chem. Rev.* **2018**, *118*, 2554.

9. H. Fujii, *Coord. Chem. Rev.* **2002**, *226*, 51.

10. A. R. McDonald, L. Que, *Coord. Chem. Rev.* **2013**, *257*, 414.

11. J. Hohenberger, K. Ray, K. Meyer, *Nat. Commun.* **2012**, *3*, 720.

12. H.-P. Hersleth, U. Ryde, P. Rydberg, C. H. Görbitz, K. K. Andersson, *J. Inorg. Biochem.* **2006**, *100*, 460.

13. (a) I. Schlichting, J. Berendzen, K. Chu, A. M. Stock, S. A. Maves, D. E. Benson, R. M. Sweet, D. Ringe, G. A. Petsko, S. G. Sligar, *Science* **2000**, *287*, 1615; (b) J. Rittle, M. T. Green, *Science* **2010**, *330*, 933.

14. (a) L. Que, *Nat. Struct. Biol.* **2000**, *7*, 182; (b) K. D. Koehntop, J. P. Emerson, L. Que, *J. Biol. Inorg. Chem.* **2005**, *10*, 87; (c) S. Kal, L. Que, *J. Biol. Inorg. Chem.* **2017**, *22*, 339.

15. (a) P. J. Riggs-Gelasco, J. C. Price, R. B. Guyer, J. H. Brehm, E. W. Barr, J. M. Bollinger, C. Krebs, *J. Am. Chem. Soc.* **2004**, *126*, 8108; (b) C. Krebs, D. Galonić Fujimori, C. T. Walsh, J. M. Bollinger, *Acc. Chem. Res.* **2007**, *40*, 484; (c) J. M. Bollinger Jr, W.-c. Chang, M. L. Matthews, R. J. Martinie, A. K. Boal, C. Krebs. CHAPTER 3 Mechanisms of 2-Oxoglutarate-Dependent Oxygenases: The Hydroxylation Paradigm and Beyond. In: *2-Oxoglutarate-Dependent Oxygenases*. The Royal Society of Chemistry (2015).

16. S. P. K. de Visser, Devesh. *Iron-containing enzymes: Versatile catalysts of hydroxylation reactions in nature*. Royal Society of Chemistry (2011).

17. X. Huang, J. T. Groves, *Chem. Rev.* **2018**, *118*, 2491.

18. S. Chakrabarty, R. N. Austin, D. Deng, J. T. Groves, J. D. Lipscomb, *J. Am. Chem. Soc.* **2007**, *129*, 3514.

19. (a) M.-H. Baik, M. Newcomb, R. A. Friesner, S. J. Lippard, *Chem. Rev.* **2003**, *103*, 2385; (b) L. Shu, J. C. Nesheim, K. Kauffmann, E. Münck, J. D. Lipscomb, L. Que, *Science* **1997**, *275*, 515.

20. S. Shaik, S. Cohen, Y. Wang, H. Chen, D. Kumar, W. Thiel, *Chem. Rev.* **2010**, *110*, 949.

21. J.-U. Rohde, J.-H. In, M. H. Lim, W. W. Brennessel, M. R. Bukowski, A. Stubna, E. Münck, W. Nam, L. Que, *Science* **2003**, *299*, 1037.

22. (a) J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger, C. Krebs, *Biochemistry* **2003**, *42*, 7497; (b) J. C. Price, E. W. Barr, T. E. Glass, C. Krebs, J. M. Bollinger, *J. Am. Chem. Soc.* **2003**, *125*, 13008.

23. M. H. Lim, J.-U. Rohde, A. Stubna, M. R. Bukowski, M. Costas, R. Y. N. Ho, E. Münck, W. Nam, L. Que, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3665.

24. J. Kaizer, E. J. Klinker, N. Y. Oh, J.-U. Rohde, W. J. Song, A. Stubna, J. Kim, E. Münck, W. Nam, L. Que, *J. Am. Chem. Soc.* **2004**, *126*, 472.

25. E. J. Klinker, J. Kaizer, W. W. Brennessel, N. L. Woodrum, C. J. Cramer, L. Que Jr, *Angew. Chem. Int. Ed.* **2005**, *44*, 3690.

26. J. E. M. N. Klein, L. Que. Biomimetic High-Valent Mononuclear Nonheme Iron-Oxo Chemistry. In: *Encyclopedia of Inorganic and Bioinorganic Chemistry*. John Wiley & Sons, Ltd (2016).

27. M. S. Seo, N. H. Kim, K.-B. Cho, J. E. So, S. K. Park, M. Clémancey, R. Garcia-Serres, J.-M. Latour, S. Shaik, W. Nam, *Chem. Sci.* **2011**, *2,* 1039.

28. O. Pestovsky, S. Stoian, E. L. Bominaar, X. Shan, E. Münck, L. Que Jr, A. Bakac, *Angew. Chem. Int. Ed.* **2005**, *44*, 6871.

29. A. N. Biswas, M. Puri, K. K. Meier, W. N. Oloo, G. T. Rohde, E. L. Bominaar, E. Münck, L. Que, *J. Am. Chem. Soc.* **2015**, *137*, 2428.

30. J. England, Y. Guo, E. R. Farquhar, V. G. Young Jr, E. Münck, L. Que Jr, *J. Am. Chem. Soc.* **2010**, *132*, 8635.

31. D. C. Lacy, R. Gupta, K. L. Stone, J. Greaves, J. W. Ziller, M. P. Hendrich, A. S. Borovik, *J. Am. Chem. Soc.* **2010**, *132,* 12188.

32. J. P. Bigi, W. H. Harman, B. Lassalle-Kaiser, D. M. Robles, T. A. Stich, J. Yano, R. D. Britt, C. J. Chang, *J. Am. Chem. Soc.* **2012**, *134*, 1536.

33. (a) L. Bernasconi, M. J. Louwerse, E. J. Baerends, *Eur. J. Inorg. Chem.* **2007**, *2007*, 3023; (b) A. Decker, J.-U. Rohde, L. Que, E. I. Solomon, *J. Am. Chem. Soc.* **2004**, *126*, 5378; (c) A. Decker, J.-U. Rohde, E. J. Klinker, S. D. Wong, L. Que, E. I. Solomon, *J. Am. Chem. Soc.* **2007**, *129*, 15983.

34. (a) S. Shaik, H. Hirao, D. Kumar, *Acc. Chem. Res.* **2007**, *40*, 532; (b) C. Geng, S. Ye, F. Neese, *Angew. Chem. Int. Ed.* **2010**, *49*, 5717; (c) S. Ye, F. Neese, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1228; (d) S. Ye, C.-Y. Geng, S. Shaik, F. Neese, *Phys. Chem. Chem. Phys.* **2013**, *15*, 8017.

35. J. England, Y. Guo, K. M. Van Heuvelen, M. A. Cranswick, G. T. Rohde, E. L. Bominaar, E. Münck, L. Que, *J. Am. Chem. Soc.* **2011**, *133*, 11880.

36. N. Y. Lee, D. Mandal, S. H. Bae, M. S. Seo, Y.-M. Lee, S. Shaik, K.-B. Cho, W. Nam, *Chem. Sci.* **2017**, *8*, 5460.

37. K. Warm, A. Paskin, U. Kuhlmann, E. Bill, M. Swart, M. Haumann, H. Dau, P. Hildebrandt, K. Ray, *Angew. Chem. Int. Ed.* **2020**, *n/a*.

38. (a) P. R. Ortiz de Montellano, *Chem. Rev.* **2010**, *110*, 932; (b) O. P. R. de Montellano. *Cytochrome P450 : structure, mechanism, and biochemistry*, 4th Edition edn. Springer Science (2016); (c) N. P. Dunham, F. H. Arnold, *ACS Catal.* **2020**, *10*, 12239; (d) K. D. Dubey, S. Shaik, *Acc. Chem. Res.* **2019**, *52*, 389.

39. (a) J. T. Groves, G. A. McClusky, *J. Am. Chem. Soc.* **1976**, *98*, 859; (b) X. Huang, J. T. Groves, *JBIC* **2017**, *22*, 185.

40. K.-B. Cho, H. Hirao, S. Shaik, W. Nam, *Chem. Soc. Rev.* **2016**, *45*, 1197.

41. J. J. Warren, T. A. Tronic, J. M. Mayer, *Chem. Rev.* **2010**, *110*, 6961.

42. R. B. Woodward, R. Hoffmann, *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 781.

43. K. Fukui, T. Yonezawa, H. Shingu, *J. Chem. Phys.* **1952**, *20*, 722.

44. (a) A. Kazaryan, E. J. Baerends, *ACS Catal.* **2015**, *5*, 1475; (b) M. J. Louwerse, E. Jan Baerends, *Phys. Chem. Chem. Phys.* **2007**, *9*, 156; (c) S. Fukuzumi, Y. Morimoto, H. Kotani, P. Naumov, Y.-M. Lee, W. Nam, *Nat. Chem.* **2010**, *2*, 756; (d) R. Hoffmann, S. Alvarez, C. Mealli, A. Falceto, T. J. Cahill, T. Zeng, G. Manca, *Chem. Rev.* **2016**, *116*, 8173; (e) H. Hirao, D. Kumar, L. Que, S. Shaik, *J. Am. Chem. Soc.* **2006**, *128*, 8590.

45. P. C. Andrikopoulos, C. Michel, S. Chouzier, P. Sautet, *ACS Catal.* **2015**, *5*, 2490.

46. (a) Y. Wang, K. Han, *JBIC* **2010**, *15*, 351; (b) M. L. Neidig, A. Decker, O. W. Choroba, F. Huang, M. Kavana, G. R. Moran, J. B. Spencer, E. I. Solomon, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12966.

47. (a) D. Janardanan, Y. Wang, P. Schyman, L. Que Jr, S. Shaik, *Angew. Chem. Int. Ed.* **2010**, *49*, 3342; (b) S. Shaik, H. Chen, D. Janardanan, *Nat. Chem.* **2011**, *3*, 19.

48. M. Srnec, S. R. Iyer, L. M. K. Dassama, K. Park, S. D. Wong, K. D. Sutherlin, Y. Yoda, Y. Kobayashi, M. Kurokuzu, M. Saito, M. Seto, C. Krebs, J. M. Bollinger, E. I. Solomon, *J. Am. Chem. Soc.* **2020**, *142*, 18886.

49. D. Schröder, S. Shaik, H. Schwarz, *Acc. Chem. Res.* **2000**, *33*, 139.

50. M. G. Evans, M. Polanyi, *T. Faraday Soc.* **1938**, *34*, 11.

51. C. T. Saouma, J. M. Mayer, *Chem. Sci.* **2014**, *5*, 21.

52. D. Kumar, H. Hirao, L. Que, S. Shaik, *J. Am. Chem. Soc.* **2005**, *127*, 8026.

53. (a) D. Kumar, R. Latifi, S. Kumar, E. V. Rybak-Akimova, M. A. Sainna, S. P. de Visser, *Inorg. Chem.* **2013**, *52*, 7968; (b) D. Kumar, B. Karamzadeh, G. N. Sastry, S. P. de Visser, *J. Am. Chem. Soc.* **2010**, *132*, 7656; (c) S. P. de Visser, *J. Am. Chem. Soc.* **2010**, *132,* 1087.

54. V. Postils, A. Company, M. Solà, M. Costas, J. M. Luis, *Inorg. Chem.* **2015**, *54*, 8223.

55. (a) A. S. Rosen, J. M. Notestein, R. Q. Snurr, *ACS Catal.* **2019**, *9*, 3576; (b) M. Barona, S. Ahn, W. Morris, W. Hoover, J. M. Notestein, O. K. Farha, R. Q. Snurr, *ACS Catal.* **2020**, *10*, 1460.

56. D. R. Nelson, *Biochim. Biophys. Acta* **2018**, *1866*, 141.

57. S. Rupasinghe, M. A. Schuler, N. Kagawa, H. Yuan, L. Lei, B. Zhao, S. L. Kelly, M. R. Waterman, D. C. Lamb, *FEBS Lett.* **2006**, *580*, 6338.

58. (a) D. C. Haines, D. R. Tomchick, M. Machius, J. A. Peterson, *Biochemistry* **2001**, *40*, 13456; (b) W.-C. Huang, A. C. G. Westlake, J.-D. Maréchal, M. G. Joyce, P. C. E. Moody, G. C. K. Roberts, *J. Mol. Biol.* **2007**, *373*, 633.

59. P. J. Loida, S. G. Sligar, *Biochemistry* **1993**, *32*, 11530.

60. M. J. Coon, *Annu. Rev. Pharmacol. Toxicol.* **2004**, *45*, 1.

61. E. O'Reilly, V. Köhler, S. L. Flitsch, N. J. Turner, *Chem. Commun.* **2011**, *47*, 2490.

62. M. B. Murataliev, M. Klein, A. Fulco, R. Feyereisen, *Biochemistry* **1997**, *36*, 8401.

63. L. O. Narhi, A. J. Fulco, *J. Biol. Chem.* **1986**, *261*, 7160.

64. H. Li, T. L. Poulos, *Nat. Struct. Biol.* **1997**, *4*, 140.

65. C. J. C. Whitehouse, S. G. Bell, L.-L. Wong, *Chem. Soc. Rev.* **2012**, *41*, 1218.

66. (a) D. A. Rock, A. E. Boitano, J. L. Wahlstrom, D. A. Rock, J. P. Jones, *Bioorg. Chem.* **2002**, *30*, 107; (b) K. D. Dubey, B. Wang, S. Shaik, *J. Am. Chem. Soc.* **2016**, *138*, 837.

67. C. F. Oliver, S. Modi, M. J. Sutcliffe, W. U. Primrose, L.-Y. Lian, G. C. K. Roberts, *Biochemistry* **1997**, *36*, 1567.

68. P. C. Cirino, F. H. Arnold, *Adv. Synth. Catal.* **2002**, *344*, 932.

69. S. Kille, F. E. Zilly, J. P. Acevedo, M. T. Reetz, *Nat. Chem.* **2011**, *3*, 738.

70. W. Chen, M. J. Fisher, A. Leung, Y. Cao, L. L. Wong, *ACS Catal.* **2020**, *10*, 8334.

71. (a) X. Zhang, Y. Peng, J. Zhao, Q. Li, X. Yu, C. G. Acevedo-Rocha, A. Li, *Bioresour. Bioprocess.* **2020**, *7*, 2; (b) E. Romero, B. S. Jones, B. N. Hogg, A. Rué Casamajo, M. A. Hayes, S. L. Flitsch, N. J. Turner, C. Schnepel, *Angew. Chem. Int. Ed. n/a*; (c) C. G. Acevedo-Rocha, F. Hollmann, J. Sanchis, Z. Sun, *ACS Catal.* **2020**, *10*, 15123; (d) R. Fasan, *ACS Catal.* **2012**, *2*, 647; (e) S. T. Jung, R. Lauchli, F. H. Arnold, *Curr. Opin. Biotechnol.* **2011**, *22*, 809; (f) G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59*, 13204.

72. M. Su, S. Chakraborty, Y. Osawa, H. Zhang, *J. Biol. Chem.* **2020**, *295*, 1637.

73. K. D. Dubey, S. Shaik, *J. Am. Chem. Soc.* **2018**, *140*, 683.

74. Z. Wang, S. Shaik, B. Wang, *J. Am. Chem. Soc.* **2021**, *143*, 1005.

75. C.-C. Chen, J. Min, L. Zhang, Y. Yang, X. Yu, R.-T. Guo, *ChemBioChem* **2020**, *n/a*.

76. S. Kalita, S. Shaik, H. K. Kisan, K. D. Dubey, *ACS Catal.* **2020**, *10*, 11481.

77. (a) P. J. Winn, S. K. Lüdemann, R. Gauges, V. Lounnas, R. C. Wade, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5361; (b) V. Cojocaru, P. J. Winn, R. C. Wade, *Biochim. Biophys. Acta* **2007**, *1770*, 390.

78. M. C. C. J. C. Ebert, S. L. Dürr, A. A. Houle, G. Lamoureux, J. N. Pelletier, *ACS Catal.* **2016**, *6*, 7426.

79. (a) F. H. Arnold, *Angew. Chem. Int. Ed.* **2019**, *58*, 14420; (b) G.-D. Roiban, M. T. Reetz, *Chem. Commun.* **2015**, *51*, 2208; (c) C. Zeymer, D. Hilvert, *Annu. Rev. Biochem* **2018**, *87*, 131.

80. A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, *Chem. Commun.* **2017**, *53*, 284.

81. H. A. Bunzel, X. Garrabou, M. Pott, D. Hilvert, *Curr. Opin. Struct. Biol.* **2018**, *48*, 149.

82. (a) K. K. Yang, Z. Wu, F. H. Arnold, *Nature Methods* **2019**, *16*, 687; (b) G. Li, Y. Dong, M. T. Reetz, *Adv. Synth. Catal.* **2019**, *361*, 2377; (c) B. J. Wittmann, K. E. Johnston, Z. Wu, F. H. Arnold, *Curr. Opin. Struct. Biol.* **2021**, *69*, 11.

83. S. C. Dodani, G. Kiss, J. K. B. Cahn, Y. Su, V. S. Pande, F. H. Arnold, *Nat. Chem.* **2016**, *8*, 419.

84. A. Glieder, E. T. Farinas, F. H. Arnold, *Nat. Biotechnol.* **2002**, *20*, 1135.

85. M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442.

86. B. M. A. van Vugt-Lussenburg, M. C. Damsten, D. M. Maasdijk, N. P. E. Vermeulen, J. N. M. Commandeur, *Biochem. Biophys. Res. Commun.* **2006**, *346*, 810.

87. B. M. A. van Vugt-Lussenburg, E. Stjernschantz, J. Lastdrager, C. Oostenbrink, P. E. Vermeulen, J. N. M. Commandeur, *J. Med. Chem.* **2007**, *50*, 455.

88. (a) E. Vottero, V. Rea, J. Lastdrager, M. Honing, N. P. E. Vermeulen, J. N. M. Commandeur, *JBIC* **2011**, *16*, 899; (b) J. S. B. de Vlieger, A. J. Kolkman, K. A. M. Ampt, J. N. M. Commandeur, N. P. E. Vermeulen, J. Kool, S. S. Wijmenga, W. M. A. Niessen, H. Irth, M. Honing, *J. Chromatogr. B* **2010**, *878*, 667.

89. V. Rea, A. J. Kolkman, E. Vottero, E. J. Stronks, K. A. M. Ampt, M. Honing, N. P. E. Vermeulen, S. S. Wijmenga, J. N. M. Commandeur, *Biochemistry* **2012**, *51*, 750.

90. H. Venkataraman, S. B. A. de Beer, L. A. H. van Bergen, N. van Essen, D. P. Geerke, N. P. E. Vermeulen, J. N. M. Commandeur, *ChemBioChem* **2012**, *13*, 520.

91. M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem. Int. Ed.* **2005**, *44*, 4192.

92. C. G. Acevedo-Rocha, *et al.*, *ACS Catal.* **2018**, *8*, 3395.

93. M. M. González, L. A. Abriata, P. E. Tomatis, A. J. Vila, *Mol. Biol. Evol.* **2016**, *33*, 1768.

94. P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866.

95. R. A. Fisher, *Transactions of the Royal Society of Edinburgh* **1919**, *52*, 399.

96. J. A. G. M. de Visser, S. F. Elena, I. Fragata, S. Matuszewski, *Heredity* **2018**, *121*, 401.

97. S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the sixth international congress of Genetics*) (1932).

98. M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260.

99. T. Khraisha, *Structural Change and Economic Dynamics* **2020**, *52*, 390.

100. H. R. Maier, S. Razavi, Z. Kapelan, L. S. Matott, J. Kasprzyk, B. A. Tolson, *Environ. Model. Softw.* **2019**, *114,* 195.

101. J. Maynard Smith, *Nature* **1970**, *225*, 563.

102. T. N. Starr, J. W. Thornton, *Protein Sci.* **2016**, *25*, 1204.

103. (a) A. C. Palmer, E. Toprak, M. Baym, S. Kim, A. Veres, S. Bershtein, R. Kishony, *Nat. Commun.* **2015**, *6*, 7385; (b) C. M. Miton, K. Buda, N. Tokuriki, *Curr. Opin. Struct. Biol.* **2021**, *69*, 160.

104. E. C. Hartman, D. Tullman-Ercek, *Curr. Opin. Syst. Biol.* **2019**, *14*, 25.

105. N. Tokuriki, D. S. Tawfik, *Science* **2009**, *324*, 203.

106. (a) S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196; (b) M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186; (c) S. Martí, M. Roca, J. Andrés, V. Moliner, E. Silla, I. Tuñón, J. Bertrán, *Chem. Soc. Rev.* **2004**, *33*, 98; (d) A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, M. H. M. Olsson, *Chem. Rev.* **2006**, *106*, 3210.

107. (a) K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, D. Kern, *Nature* **2007**, *450*, 838; (b) S. Hammes-Schiffer, S. J. Benkovic, *Annu. Rev. Biochem* **2006**, *75*, 519; (c) S. Osuna, G. Jiménez-Osés, E. L. Noey, K. N. Houk, *Acc. Chem. Res.* **2015**, *48*, 1080.

108. D. Petrović, V. A. Risso, S. C. L. Kamerlin, J. M. Sanchez-Ruiz, *J. R. Soc. Interface* **2018**, *15*, 20180330.

109. (a) A. Neu, U. Neu, A.-L. Fuchs, B. Schlager, R. Sprangers, *Nat. Chem. Biol.* **2015**, *11*, 697; (b) N. Tokuriki, Colin J. Jackson, *Chemistry & Biology* **2014**, *21*, 1259.

110. D. E. Koshland, G. Némethy, D. Filmer, *Biochemistry* **1966**, *5*, 365.

111. J. Monod, J. Wyman, J.-P. Changeux, *J. Mol. Biol.* **1965**, *12*, 88.

112. (a) G. G. Hammes, Y.-C. Chang, T. G. Oas, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 13737; (b) S. Gianni, J. Dogan, P. Jemth, *Biophys. Chem.* **2014**, *189*, 33; (c) E. A. Galburt, J. Rammohan, *Biochemistry* **2016**, *55*, 7014.

113. (a) A. D. Vogt, N. Pozzi, Z. Chen, E. Di Cera, *Biophys. Chem.* **2014**, *186*, 13; (b) A. D. Vogt, E. Di Cera, *Biochemistry* **2013**, *52*, 5723; (c) A. D. Vogt, E. Di Cera, *Biochemistry* **2012**, *51*, 5894.

114. H.-X. Zhou, *Biophys. J.* **2010**, *98*, L15.

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

116. R. Otten, R. A. P. P·dua, H. A. Bunzel, V. Nguyen, W. Pitsawong, M. Patterson, S. Sui, S. L. Perry, A. E. Cohen, D. Hilvert, D. Kern, *Science* **2020**, *370*, 1442.

117. M. G. Joyce, H. M. Girvan, A. Munro, D. Leys, *J. Biol. Chem.* **2004**, *279*, 23287.

118. C. J. Cramer, W. John, Sons. *Essentials of computational chemistry: theories and models*. John Wiley & Sons (2004).

119. A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227.

120. K. Caddell Haatveit, M. Garcia-Borràs, K. N. Houk, *Frontiers in Chemistry* **2019**, *6*, 663.

121. D. J. Tantillo, C. Jiangang, K. N. Houk, *Curr. Opin. Chem. Biol.* **1998**, *2*, 743.

122. P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.

123. W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

124. M. Swart. Dealing with Spin States in Computational Organometallic Catalysis. In: *New Directions in the Modeling of Organometallic Reactions*. Springer International Publishing (2020).

125. (a) M. Swart, *Chem. Phys. Lett.* **2013**, *580*, 166; (b) M. Swart, M. Gruden, *Acc. Chem. Res.* **2016**, *49*, 2690.

126. J. Harris, *Phys. Rev. A* **1984**, *29*, 1648.

127. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

128. A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

129. C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.

130. A. Einstein, *Annalen der Physik* **1905**, *322*, 891.

131. P. A. M. Dirac, R. H. Fowler, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1928**, *117*, 610.

132. C. Chang, M. Pelissier, P. Durand, *Phys. Scr.* **1986**, *34*, 394.

133. J. L. Heully, I. Lindgren, E. Lindroth, S. Lundqvist, A. M. Martensson-Pendrill, *J. Phys. B* **1986**, *19*, 2799.

134. (a) E. v. Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1993**, *99*, 4597; (b) E. van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1994**, *101*, 9783; (c) E. van Lenthe, R. van Leeuwen, E. J. Baerends, J. G. Snijders, *Int. J. Quantum Chem* **1996**, *57*, 281.

135. L. L. Foldy, S. A. Wouthuysen, *Phys. Rev.* **1950**, *78*, 29.

136. M. Dolg. Chapter 14 - Relativistic Effective Core Potentials. In: *Theoretical and Computational Chemistry*. Elsevier (2002).

137. M. Reiher, A. Wolf. *Relativistic quantum chemistry: the fundamental theory of molecular science*. Wiley (2009).

138. D. C. Young. *Using Existing Basis Sets* (2001).

139. (a) S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104; (b) S. Grimme, A. Hansen, J. G. Brandenburg, C. Bannwarth, *Chem. Rev.* **2016**, *116*, 5105.

140. L. Belpassi, I. Infante, F. Tarantelli, L. Visscher, *J. Am. Chem. Soc.* **2008**, *130*, 1048.

141. F. Jensen. *Introduction to computational chemistry*. Wiley (2017).

142. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

143. (a) P. P. Ewald, *Annalen der Physik* **1921**, *369*, 253; (b) D. Frenkel, B. Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press (2012).

144. J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.

145. (a) A. Klamt, *WIREs Comput. Mol. Sci.* **2011**, *1*, 699; (b) A. Klamt, G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2* **1993**, 799.

146. T. Young, T. Johnston-Wood, V. Deringer, F. Duarte, *ChemRxiv* **2021**.

147. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.

148. C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269.

149. J. M. Seminario, *Int. J. Quantum Chem* **1996**, *60*, 1271.

150. (a) J. Aaqvist, A. Warshel, *J. Am. Chem. Soc.* **1990**, *112*, 2860; (b) F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist, S. C. L. Kamerlin, *J. Phys. Chem. B* **2014**, *118*, 4351.

151. T. Lengauer, M. Rarey, *Curr. Opin. Struct. Biol.* **1996**, *6*, 402.

152. L. Verlet, *Phys. Rev.* **1967**, *159*, 98.

153. M. P. Allen, D. J. Tildesley, P. Oxford University. *Computer simulation of liquids*. Oxford University Press (2017).

154. W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.* **1982**, *76*, 637.

155. T. Schlick, E. Barth, M. Mandziuk, *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 181.

156. (a) J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327; (b) D. J. Tobias, C. L. Brooks, *J. Chem. Phys.* **1988**, *89*, 5115.

157. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463.

158. D. M. Zuckerman, *Annual Review of Biophysics* **2011**, *40*, 41.

159. M. Ringnér, *Nat. Biotechnol.* **2008**, *26*, 303.

160. (a) M. Iijima, J. Ohnuki, T. Sato, M. Sugishima, M. Takano, *Scientific Reports* **2019**, *9*, 9341; (b) S. A. Hollingsworth, D. Batabyal, B. D. Nguyen, T. L. Poulos, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 8723.

161. (a) Y. Naritomi, S. Fuchigami, *J. Chem. Phys.* **2011**, *134*, 065101; (b) Y. Naritomi, S. Fuchigami, *J. Chem. Phys.* **2013**, *139*, 215102.

162. S. Osuna, *WIREs Comput. Mol. Sci.* **2021**, *11*, e1502.

163. A. Currin, N. Swainston, P. J. Day, D. B. Kell, *Chem. Soc. Rev.* **2015**, *44*, 1172.

164. (a) A. Romero-Rivera, M. Garcia-Borràs, S. Osuna, *ACS Catal.* **2017**, *7*, 8524; (b) M. A. Maria-Solano, J. Iglesias-Fernández, S. Osuna, *J. Am. Chem. Soc.* **2019**, *141*, 13049; (c) C. Curado-Carballada, F. Feixas, J. Iglesias-Fernández, S. Osuna, *Angew. Chem. Int. Ed.* **2019**, *58*, 3097.

165. (a) J. M. Mayer, *Annu. Rev. Phys. Chem.* **2004**, *55*, 363; (b) M. H. V. Huynh, T. J. Meyer, *Chem. Rev.* **2007**, *107*, 5004; (c) S. Hammes-Schiffer, *J. Am. Chem. Soc.* **2015**, *137*, 8860; (d) D. R. Weinberg, C. J. Gagliardi, J. F. Hull, C. F. Murphy, C. A. Kent, B. C. Westlake, A. Paul, D. H. Ess, D. G. McCafferty, T. J. Meyer, *Chem. Rev.* **2012**, *112*, 4016.

166. D. Usharani, D. C. Lacy, A. S. Borovik, S. Shaik, *J. Am. Chem. Soc.* **2013**, *135*, 17090.

167. (a) J. Stubbe, D. G. Nocera, C. S. Yee, M. C. Y. Chang, *Chem. Rev.* **2003**, *103*, 2167; (b) M. J. Knapp, K. Rickert, J. P. Klinman, *J. Am. Chem. Soc.* **2002**, *124*, 3865.

168. (a) C. W. Hoganson, G. T. Babcock, *Science* **1997**, *277*, 1953; (b) A. Magnuson, M. Anderlund, O. Johansson, P. Lindblad, R. Lomoth, T. Polivka, S. Ott, K. Stensjö, S. Styring, V. Sundström, L. Hammarström, *Acc. Chem. Res.* **2009**, *42*, 1899; (c) J. H. Alstrum-Acevedo, M. K. Brennaman, T. J. Meyer, *Inorg. Chem.* **2005**, *44*, 6802; (d) T. J. Meyer, M. H. V. Huynh, H. H. Thorp, *Angew. Chem. Int. Ed.* **2007**, *46*, 5284.

169. (a) J. H. Skone, A. V. Soudackov, S. Hammes-Schiffer, *J. Am. Chem. Soc.* **2006**, *128*, 16655; (b) A. Sirjoosingh, S. Hammes-Schiffer, *J. Phys. Chem. A* **2011**, *115*, 2367; (c) A. V. Soudackov, S. Hammes-Schiffer, *J. Phys. Chem. Lett.* **2014**, *5*, 3274.

170. (a) G. Knizia, *J. Chem. Theory Comput.* **2013**, *9*, 4834; (b) J. E. M. N. Klein, G. Knizia, *Angew. Chem. Int. Ed.* **2018**, *57*, 11913.

171. (a) F. Pirani, D. Cappelletti, S. Falcinelli, D. Cesario, F. Nunzi, L. Belpassi, F. Tarantelli, *Angew. Chem. Int. Ed.* **2019**, *58*, 4195; (b) F. Nunzi, D. Cesario, L. Belpassi, F. Tarantelli, L. F. Roncaratti, S. Falcinelli, D. Cappelletti, F. Pirani, *Phys. Chem. Chem. Phys.* **2019**, *21*, 7330; (c) F. Nunzi, D. Cesario, F. Pirani, L. Belpassi, F. Tarantelli, *ChemPhysChem* **2018**, *19*, 1476; (d) G. Ciancaleoni, L. Belpassi, *J. Comput. Chem.* **2020**, *41*, 1185; (e) L. Belpassi, M. L. Reca, F. Tarantelli, L. F. Roncaratti, F. Pirani, D. Cappelletti, A. Faure, Y. Scribano, *J. Am. Chem. Soc.* **2010**, *132*, 13046; (f) D. Cappelletti, E. Ronca, L. Belpassi, F. Tarantelli, F. Pirani, *Acc. Chem. Res.* **2012**, *45*, 1571.

172. (a) G. Bistoni, P. Belanzoni, L. Belpassi, F. Tarantelli, *J. Phys. Chem. A* **2016**, *120*, 5239; (b) L. D'Amore, G. Ciancaleoni, L. Belpassi, F. Tarantelli, D. Zuccaccia, P. Belanzoni, *Organometallics* **2017**, *36*, 2364; (c) L. Gregori, D. Sorbelli, L. Belpassi, F. Tarantelli, P. Belanzoni, *Inorg. Chem.* **2019**, *58*, 3115; (d) G. Bistoni, L. Belpassi, F. Tarantelli, *Angew. Chem. Int. Ed.* **2013**, *52*, 11599.

173. (a) E. J. Baerends¸ *et al.* ADF2017.). ADF2017.113 edn. SCM, Theoretical Chemistry, Vrije Universiteit (2017); (b) G. te Velde, F. M. Bickelhaupt, E. J.

Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.

174. E. Van Lenthe, E. J. Baerends, *J. Comput. Chem.* **2003**, *24*, 1142.

175. (a) S. K. Padamati, D. Angelone, A. Draksharapu, G. Primi, D. J. Martin, M. Tromp, M. Swart, W. R. Browne, *J. Am. Chem. Soc.* **2017**, *139*, 8718; (b) K. Rajabimoghadam, Y. Darwish, U. Bashir, D. Pitman, S. Eichelberger, M. A. Siegler, M. Swart, I. Garcia-Bosch, *J. Am. Chem. Soc.* **2018**, *140*, 16625; (c) S. Banerjee, A. Draksharapu, P. M. Crossland, R. Fan, Y. Guo, M. Swart, L. Que, *J. Am. Chem. Soc.* **2020**, *142*, 4285.

176. K. M. Azzopardi, G. Bistoni, G. Ciancaleoni, F. Tarantelli, D. Zuccaccia, L. Belpassi, *Dalton Trans.* **2015**, *44*, 13999.

177. T. Gensch, M. N. Hopkinson, F. Glorius, J. Wencel-Delord, *Chem. Soc. Rev.* **2016**, *45*, 2900.

178. B. Hauer, *ACS Catal.* **2020**, *10*, 8418.

179. A. R. McDonald, Y. Guo, V. V. Vu, E. L. Bominaar, E. Münck, L. Que, *Chem. Sci.* **2012**, *3*, 1680.

180. I. Monte Pérez, X. Engelmann, Y.-M. Lee, M. Yoo, E. Kumaran, E. R. Farquhar, E. Bill, J. England, W. Nam, M. Swart, K. Ray, *Angew. Chem. Int. Ed.* **2017**, *56*, 14384.

181. C. C. Pye, T. Ziegler, *Theor. Chem. Acc.* **1999**, *101*, 396.

182. M. Swart, F. M. Bickelhaupt, *J. Comput. Chem.* **2008**, *29*, 724.

183. (a) J. E. M. N. Klein, B. Dereli, L. Que, C. J. Cramer, *Chem. Commun.* **2016**, *52*, 10509; (b) B. B. Averkiev, D. G. Truhlar, *Catalysis Science & Technology* **2011**, *1*, 1526.

184. L. D'Amore, L. Belpassi, J. E. M. N. Klein, M. Swart, *Chem. Commun.* **2020**, *56*, 12146.

185. M. A. Maria-Solano, E. Serrano-Hervás, A. Romero-Rivera, J. Iglesias-Fernández, S. Osuna, *Chem. Commun.* **2018**, *54*, 6622.

186. (a) H. Yu, P. A. Dalby, *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E11043; (b) C. M. Miton, N. Tokuriki, *Protein Sci.* **2016**, *25*, 1260.

187. R. Otten, L. Liu, L. R. Kenner, M. W. Clarkson, D. Mavor, D. S. Tawfik, D. Kern, J. S. Fraser, *Nat. Commun.* **2018**, *9*, 1314.

188. G. Yang, N. Hong, F. Baier, C. J. Jackson, N. Tokuriki, *Biochemistry* **2016**, *55*, 4583.

189. (a) M. Patel, L. Hu, C. Brown, Z. Sun, C. J. Adamski, V. Stojanoski, B. Sankaran, B. Prasad, T. Palzkill, *J. Biol. Chem.* **2018**, *293*, 17971; (b) E. Dellus-Gur, M. Elias, E. Caselli, F. Prati, M. L. M. Salverda, J. A. G. M. de Visser, J. S. Fraser, D. S. Tawfik, *J. Mol. Biol.* **2015**, *427*, 2396; (c) J. Z. Chen, D. M. Fowler, N. Tokuriki, *bioRxiv* **2021**, 2021.04.14.439889.

190. A. R. H. Narayan, G. Jiménez-Osés, P. Liu, S. Negretti, W. Zhao, M. M. Gilbert, R. O. Ramabhadran, Y.-F. Yang, L. R. Furan, Z. Li, L. M. Podust, J. Montgomery, K. N. Houk, D. H. Sherman, *Nat. Chem.* **2015**, *7*, 653.

191. M. J. Frisch, *et al.* Gaussian09 Revision D.01.). G09D.01 edn. Gaussian Inc. (2009).

192. (a) C. N. Schutz, A. Warshel, *Proteins* **2001**, *44*, 400; (b) L. Li, C. Li, Z. Zhang, E. Alexov, *J. Chem. Theory Comput.* **2013**, *9*, 2126.

193. P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 270.

194. D. A. Case, *et al.* Amber 2016. In: *AMBER 2016*). AMBER 2016 edn. University of California (2016).

195. (a) U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129; (b) B. H. Besler, K. M. Merz Jr, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*, 431.

196. K. Shahrokh, A. Orendt, G. S. Yost, T. E. Cheatham III, *J. Comput. Chem.* **2012**, *33*, 119.

197. R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537.

198. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins* **2010**, *78*, 1950.

199. (a) C. A. Smith, T. Kortemme, *J. Mol. Biol.* **2008**, *380*, 742; (b) E. L. Humphris, T. Kortemme, *Structure* **2008**, *16*, 1777; (c) I. W. Davis, W. B. Arendall, D. C. Richardson, J. S. Richardson, *Structure* **2006**, *14*, 265.

200. T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.

201. (a) Y. Miao, F. Feixas, C. Eun, J. A. McCammon, *J. Comput. Chem.* **2015**, *36*, 1536; (b) D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, *120*, 11919; (c) D. Hamelberg, C. A. F. de Oliveira, J. A. McCammon, *J. Chem. Phys.* **2007**, *127*, 155102.

202. (a) L. Skjærven, X.-Q. Yao, G. Scarabelli, B. J. Grant, *BMC Bioinformatics* **2014**, *15*, 399; (b) B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, *Bioinformatics* **2006**, *22*, 2695.

203. J. D. Durrant, L. Votapka, J. Sørensen, R. E. Amaro, *J. Chem. Theory Comput.* **2014**, *10*, 5047.

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

205. H. J. Wijma, R. J. Floor, S. Bjelic, S. J. Marrink, D. Baker, D. B. Janssen, *Angew. Chem. Int. Ed.* **2015**, *54*, 3726.

206. (a) N. Ahalawat, J. Mondal, *J. Am. Chem. Soc.* **2018**, *140*, 17743; (b) B. R. Dandekar, N. Ahalawat, J. Mondal, *Biophys. J.* **2021**.

207. C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, J. Deisenhofer, *Structure* **1995**, *3*, 41.

208. (a) W. Lu, J. Feng, X. Chen, Y.-J. Bao, Y. Wang, Q. Wu, Y. Ma, D. Zhu, *Appl. Environ. Microbiol.* **2019**, *85*, e01182; (b) M. Bureik, R. Bernhardt, *Modern Biooxidation* **2007**, 155; (c) M. V. Donova. Steroid Bioconversions. In: *Microbial Steroids: Methods and Protocols*. Springer New York (2017); (d) B. Dudas, I. Hanin, M. Rose, E. Wülfert, *Neurobiology of Disease* **2004**, *15*, 262; (e) A. K. Pringle, W. Schmidt, J. K. Deans, E. Wulfert, K. G. Reymann, L. E. Sundstrom, *Eur. J. Neurosci.* **2003**, *18*, 117.

209. (a) H. Pellissier, M. Santelli, *Org. Prep. Proced. Int.* **2001**, *33*, 1; (b) H. J. E. Loewenthal, *Tetrahedron* **1959**, *6*, 269.

210. (a) K. E. Smith, S. Latif, D. N. Kirk, *Journal of Steroid Biochemistry* **1990**, *35*, 115; (b) B. Ma, Q. Wang, H. Ikeda, C. Zhang, L.-H. Xu, *Appl. Environ. Microbiol.* **2019**, *85*, e01530.

211. A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59*, 12499.

212. C. G. Acevedo-Rocha, A. Li, L. D’Amore, S. Hoebenreich, J. Sanchis, P. Lubrano, M. P. Ferla, M. Garcia-Borràs, S. Osuna, M. T. Reetz, *Nat. Commun.* **2021**, *12*, 1621.

213. O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455.

Appendices

APPENDIX A

Supplementary information Chapter 4

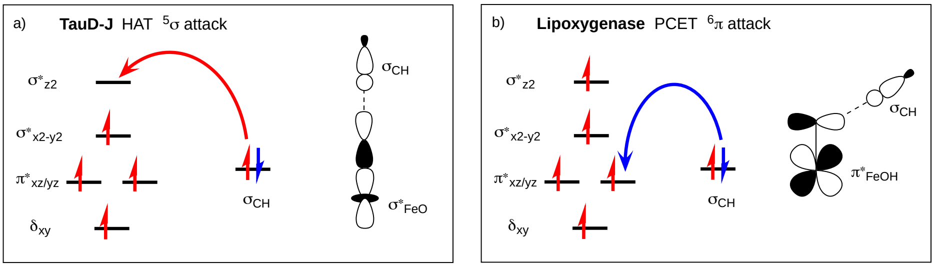


Figure A1. Favored reaction pathway for: a) TauD-J intermediate; b) Lipoxygenase.

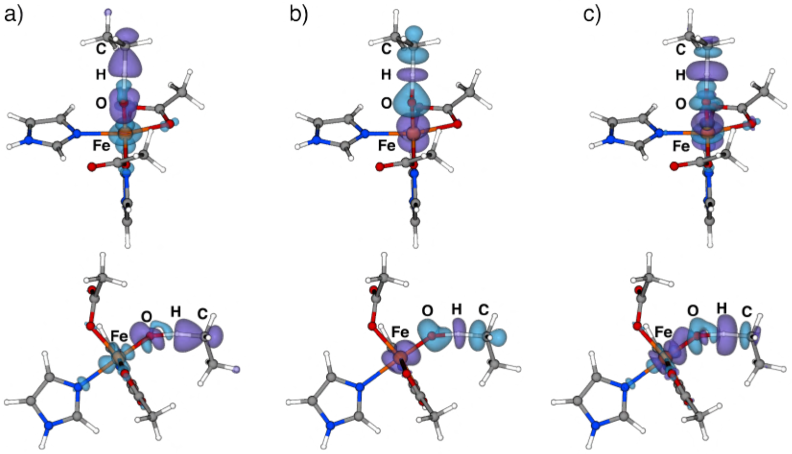


Figure A2. Alternative views of the 3D contour plots (± 0.035 e au^{-3}) of the Δρ for the FeO–H(substrate) bond in the TS of the TauD-J model complex. a) α spin-density difference; b) β spin-density difference; c) total density difference cyan (violet) ribbon color corresponds to regions of charge accumulation (depletion).

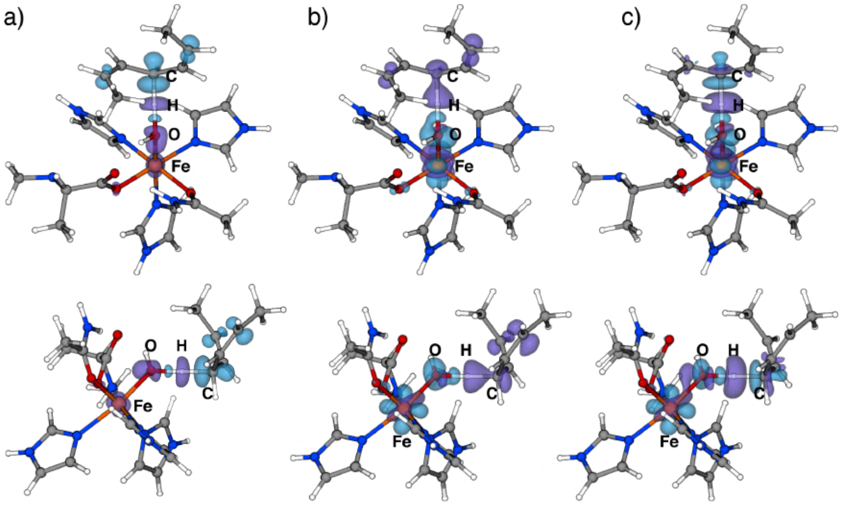


Figure A3. Alternative views of the 3D contour plots (± 0.035 e au^{-3}) of the Δρ for the FeO–H(substrate) bond in the TS of the lipoxygenase model complex. a) α spin-density difference; b) β spin-density difference; c) total density difference cyan (violet) ribbon color corresponds to regions of charge accumulation (depletion).

Effect of the functional on the open-shell charge displacement function

In addition to the S12g (GGA) results, the same calculations were carried out using the B3LYP (hybrid) and LDA functionals. In all cases, the effect of the functional on the open-shell charge displacement function was found to be negligible.

Breaking of the C-H bond

The relevant changes for the breaking of the C-H bond are not immediately obvious from the total osCDF in Figures 4.3(b) and 4.5(b), which is influenced by the choice of substrates that we used. The rearrangement of electrons during the HAT/cPCET process, resulting in the radical formation on carbon involves movement of charge from the substrate towards the carbon. This can be seen directly by focusing on the region to the right of the carbon, where in the ethane of Fig. 4.3(b) there is hardly any change, while in the case of the heptadiene of Fig. 4.5(b) there is a significant flow of electrons within the substrate. The origin of this is most easily seen in the alternative view from Figures A4 and A5 (*vide infra*). For TauD-*J*, the flow of electrons in the substrate involves exclusively the C–H region, while in the case of lipoxygenase also other parts of the substrate are affected. This additional flow in the substrate makes that the total osCDF at the CH region includes not just changes in the CH bond region, but also from other parts of the substrate.

For understanding/visualizing the breaking of the C-H bond, one should take into account not so much the value of the osCDF, but instead its slope. As reported in Section 2.1.5 ("*Positive values of Δq(z) correspond to electrons flowing in the direction of decreasing z, and negative values to electrons moving to increasing z. A positive slope indicates regions of charge accumulation, and vice versa a negative slope indicates depletion*"), this slope determines if there is density accumulation or density depletion. By looking carefully at the region around the C–H bond, there is indeed in the middle of the C–H bond a turning point with overall density accumulation going to the right towards C and overall depletion to the left towards H.

In TauD-J in the C–H region there exists an α-electron flow (red arrow) from the right to the left and a β-electron flow (blue arrow) from the left to the right, in accordance with scheme 2. However, the same scheme does not take into account the total electron flow (Δq_{tot}), which is the sum of the α and the β electron flows (black curve of the CD function plot). Here, in the proximity of the H atom we observe an electron depletion (negative slope) with $\Delta q_{tot}>0$, meaning that electrons are moving from right to the left (green arrow), whereas in the proximity of the C atom we observe an opposite trend i.e. electron accumulation (positive slope) and $\Delta q_{tot}\leq 0$ (electrons moving from the left to the right, green arrow).

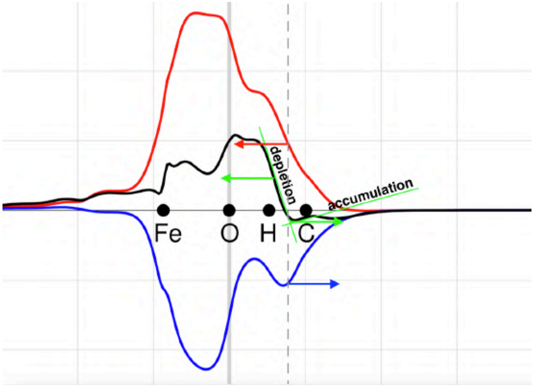


Figure A4. Modified version of Figure 4.3(b) with trendlines (green) that indicate the slope of the total osCDF in the H··C region.

Likewise, for lipoxygenase, the α-electron flow (red arrow) and β-electron flow (blue arrow) is in agreement with scheme 3 and it's worth noting that the total electron accumulation and depletion around the H··C region follows the same trend as for TauD-J (i.e. a depletion around H followed by an accumulation around C). However, here, $\Delta q_{tot}>0$ over the whole plot, including the CH region.

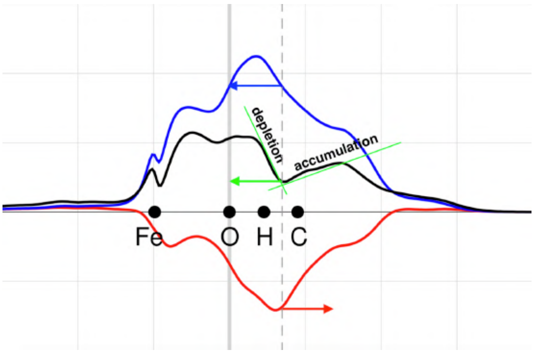
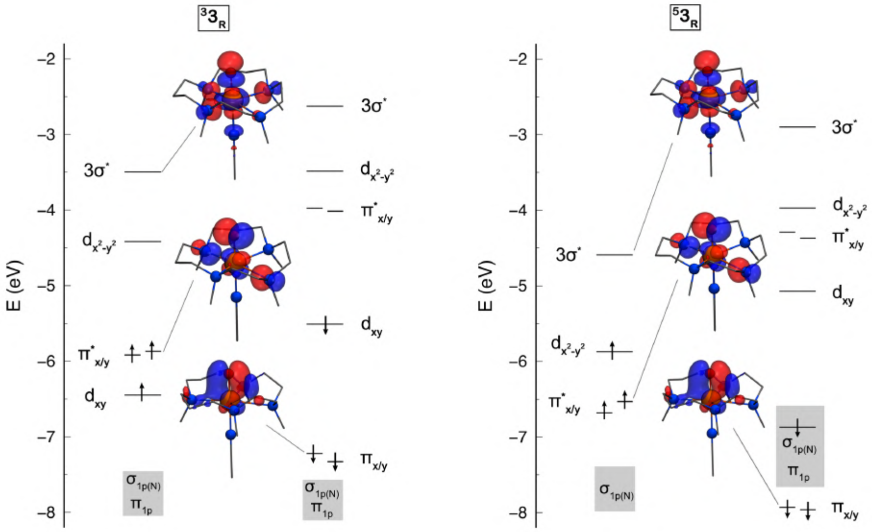


Figure A5. Modified version of Figure 4.5(b) with trendlines (green) that indicate the slope of the total osCDF in the H··C region.

APPENDIX B

Supplementary information Chapter 5



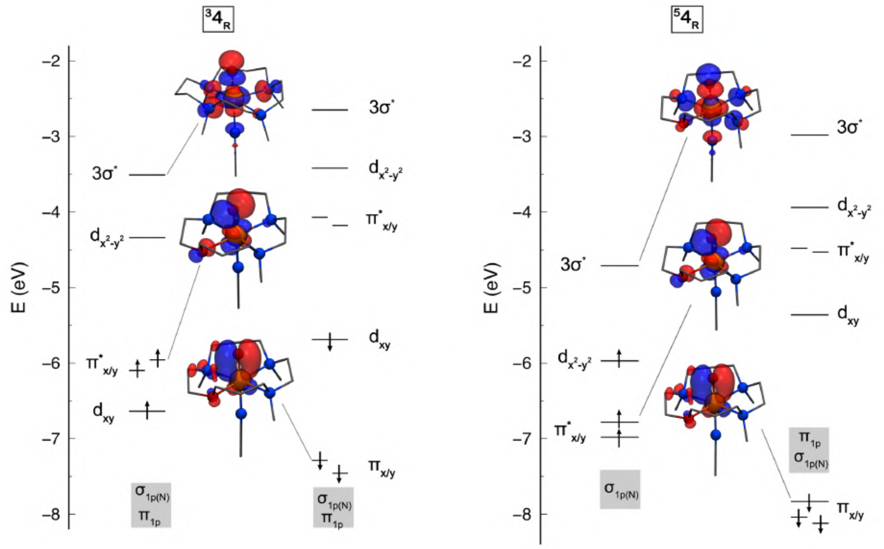


Figure B1. Frontier molecular orbitals of complex **3**$_{\mathrm{R}}$ and **4**$_{\mathrm{R}}$ in the S=1 and S=2 states. In each panel the α and β spin orbitals are to the left and right, respectively. Isodensity value ± 0.015 e au^{-3}.

Table B1. Energies (in eV) of the frontier molecular orbitals. Bold and regular text correspond to occupied and unoccupied orbitals, respectively.

S=1	**1**$_{\mathrm{R}}$		**4**$_{\mathrm{R}}$		**3**$_{\mathrm{R}}$		**2**$_{\mathrm{R}}$	
	α	β	α	β	α	β	α	β
σ*		-2.54		-2.65		-2.63		-2.77
d_{x2-y2}		-3.32		-3.42		-3.49		-3.62
σ*	-3.4		-3.51		-3.5		-3.64	
π*		-3.96		-4.07		-3.98		-4.08
π*		-4.02		-4.18		-4.02		-4.19
d_{x2-y2}	-4.24		-4.34		-4.42		-4.57	
d_{xy}		**-5.52**		**-5.69**		**-5.51**		**-5.71**
π*	**-5.86**		**-5.96**		**-5.87**		**-5.96**	
π*	**-5.91**		**-6.1**		**-5.92**		**-6.12**	
d_{xy}	**-6.49**		**-6.64**		**-6.45**		**-6.63**	

<u>*S=2*</u>	**1_R**		**4_R**		**3_R**		**2_R**	
	α	β	α	β	α	β	α	β
σ*		-2.89		-2.98		-2.9		-3.07
d_{x2-y2}		-3.81		-3.94		-3.97		-4.03
π*		-4.38		-4.48		-4.29		-4.4
π*		-4.38		-4.53		-4.37		-4.54
σ*	-4.6		-4.71		-4.59		-4.79	
d_{xy}		-5.23		-5.36		-5.07		-5.28
d_{x2-y2}	**-5.73**		**-5.97**		**-5.87**		**-6.04**	
π*	**-6.68**		**-6.78**		**-6.53**		**-6.69**	
π*	**-6.7**		**-6.98**		**-6.68**		**-6.98**	

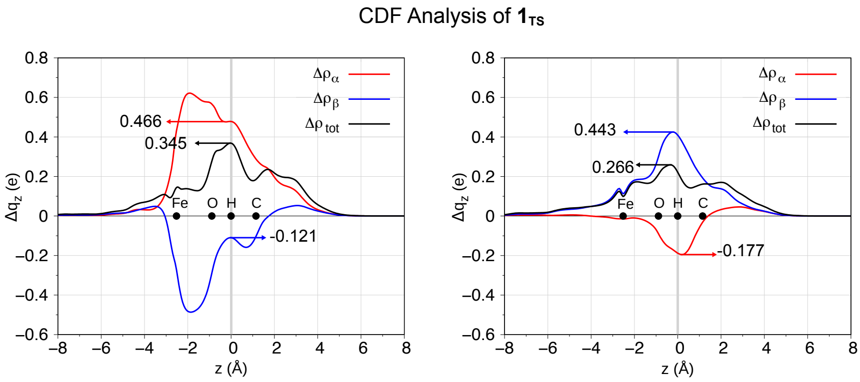


Figure B2. CD curves of the FeO···H(cychex) bond in **1_{TS}**. S=2 (left) and S=1 (right) spin state. Red, blue and black curves correspond to the α-, β-, and total-spin densities, respectively. The value of the charge transfer at the H atom region is indicated.

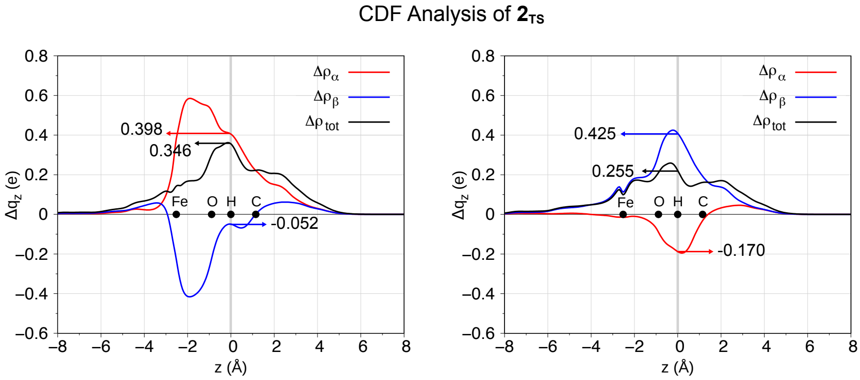


Figure B3. CD curves of the FeO···H(cychex) bond in **2_{TS}**. S=2 (left) and S=1 (right) spin state. Red, blue and black curves correspond to the α-, β-, and total-spin densities, respectively. The value of the charge transfer at the H atom region is indicated.

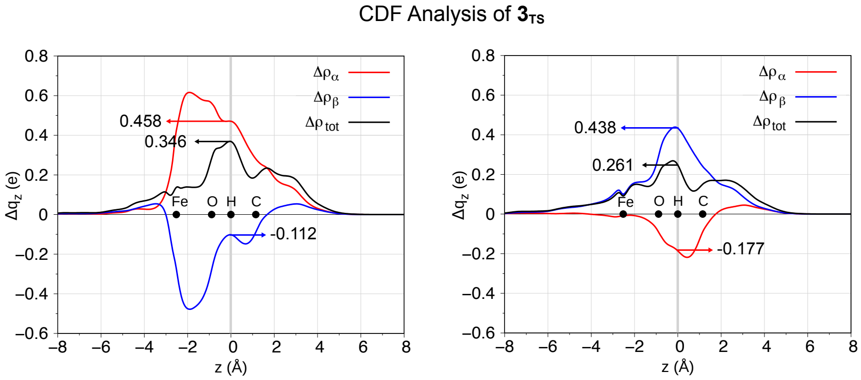


Figure B4. CD curves of the FeO···H(cychex) bond in $\mathbf{3_{TS}}$. S=2 (left) and S=1 (right) spin state. Red, blue and black curves correspond to the α-, β-, and total-spin densities, respectively. The value of the charge transfer at the H atom region is indicated.

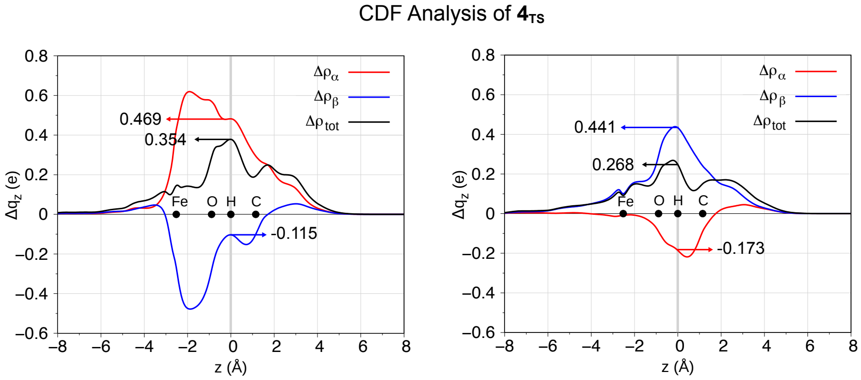


Figure B5. CD curves of the FeO···H(cychex) bond in $\mathbf{4_{TS}}$. S=2 (left) and S=1 (right) spin state. Red, blue and black curves correspond to the α-, β-, and total-spin densities, respectively. The value of the charge transfer at the H atom region is indicated.

APPENDIX C

Supplementary information Chapter 6

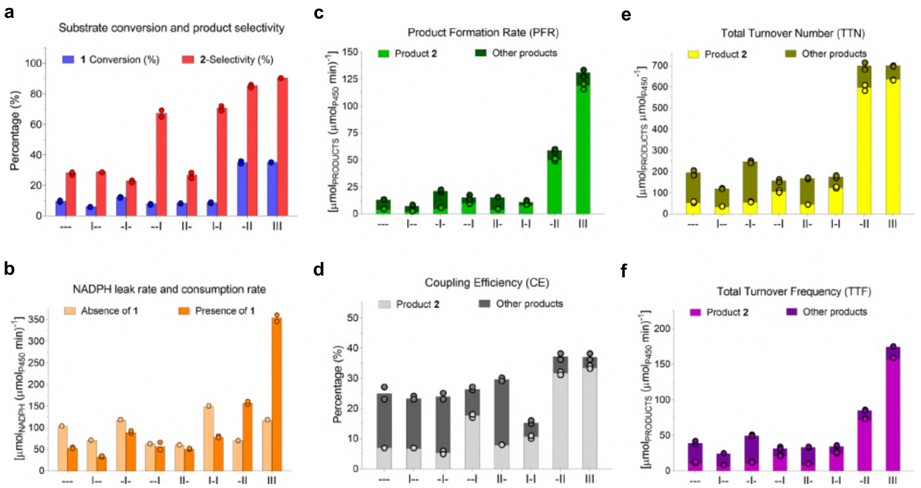


Figure C1. Multiple enzymatic parameters of deconvolution mutants. a) Selectivity and conversion data are shown in percentages. b) NADPH leak and consumption rate were measured in the absence and presence of **1**, respectively. c) Product formation rate (PFR) is calculated by multiplying the NADPH consumption rate by coupling efficiency. d) Coupling efficiency is the ratio between NADPH consumption and production formation, and it is reported in percentages. e) TTN describes the total moles of products per moles of enzyme after NADPH depletion. f) TTF normalizes TTN by time after NADPH depletion. See Fig. 6.2c for mutants abbreviation. Other products mainly include the 15β-alcohol and other regioisomers. The data represent the average of two independent experiments (n = 2) not part of the work carried out in this thesis, and are presented only for reference.

Table C1. Pathway accessibility analysis based on selectivity towards 2β-hydroxytestosterone (**2**)

Pathway	**Mutation**				**ΔΔG‡ (kJ mol-1)**			
	0	**1st**	**2nd**	**3rd**	**0**	**1st**	**2nd**	**3rd**
1	---	I--	II-	III	-2.3	-2.2	-2.5	5.6
2	---	I--	I-I	III	-2.3	-2.2	2.2	5.6
3	---	-I-	II-	III	-2.3	-3.0	-2.5	5.6
4	---	-I-	-II	III	-2.3	-3.0	4.4	5.6
5	---	--I	I-I	III	-2.3	1.8	2.2	5.6
6	---	--I	-II	III	-2.3	1.8	4.4	5.6

Table C2. Pathway accessibility analysis based on Total Turnover Frequency (TTF) towards 2β-hydroxytestosterone (**2**)

Pathway	Mutation				TTF			
	0	1st	2nd	3rd	0	1st	2nd	3rd
1	---	I--	II-	III	11	7	9	158
2	---	I--	I-I	III	11	7	24	158
3	---	-I-	II-	III	11	11	9	158
4	---	-I-	-II	III	11	11	72	158
5	---	--I	I-I	III	11	21	24	158
6	---	--I	-II	III	11	21	72	158

A green or red pathway/value means that there is respectively increase or decrease of fitness in each step of the evolutionary process from parent mutant --- (F87A) towards triple mutant **III** (F87A–R47I/T49I/Y51I). Values in orange indicate that there is neither increase nor decrease of fitness between the precedent and subsequent mutation, thus this pathway turns red. Conversely to red, green pathways are accessible to stepwise evolution. The data represent the average of two independend experiments (n=2) not part of the work carried out in this thesis, and are presented only for reference.

Table C3. Epistatic analysis of all possible mutational combination on multiple parameters towards formation of product **2**.

Type[a]	Mutation combin.[b]	Mutant result	Parameter[c,d,e]						
			Conv.	Sel.	NCR	PFR	CE	TTN	TTF
B	I-- + -I-	II-	+SE 0.4	+ME 2.8	-SE -1.3	+SE 1	+RSE 3	+SE 9	+SE 2
B	-I- + --I	-II	+ME 25	+SE 5.6	+ME 91	+ME 39	+SE 16	+ME 490	+ME 51
B	I-- + --I	I-I	+SE 2	+SE 2.6	+SE 23	ADD -0.2	-SE -7	+SE 38	+SE 7
B	I-- + -II	III	+SE 3	+SE 3.5	+SE 192	+SE 70	+SE 2	+SE 58	+SE 90
B	II- + --I	III	+SE 27	+SE 6.3	+SE 284	+ME 108	+ME 15	+SE 539	+SE 139
B	I-I + -I-	III	+ME 25	+SE 6.4	+ME 260	+ME 109	+SE 24	+ME 509	+ME 133
T	I-- + -I- + --I	III	+SE 27	+SE 9.0	+SE 283	+SE 109	+SE 18	+SE 548	+SE 140

[a] Binary (B) and tertiary (T) combinations.

[b] See Fig. 6.2c for nomenclature of mutation and mutants abbreviation.

[c] The parameters are: substrate conversion (Conv.), 2β-selectivity (Sel.), NADPH consumption rate (NCR), product formation rate (PFR), coupling efficiency (CE), total turnover number (TTN), total turnover frequency (TTF).

[d] The types of epistatic effects are: sign epistasis (SE), magnitude epistasis (ME), reciprocal sign epistasis (RSE), which can be positive (+) or negative (-). Additivity (ADD) means absence of epistatic effects.

[e] The data represent the average of two independend experiments (n=2) not part of the work carried out in this thesis, and are presented only for reference.

Table C4. Energies, thermal corrections and Gibbs free energies of the QM structures calculated for H-abstraction catalyzed by heme Compound I (CpdI).

Struct.[a]	E_{el}[b]	E_{el}[c]	G correction[d]	G[e]	ΔG
	(Hartree)[f]	(Hartree)[f]	(Hartree/particle)[f]	(Hartree)[f]	(kcal/mol)
^{4}CpdI	-1625.181930	-1625.492613	0.267701	-1625.224912	
1	-891.402149	-891.636974	0.397895	-891.239079	
^{4}RC[g]	- 2516.584079	-2517.129587	0.665596	-2516.463991	
^{4}TS2	-2516.560005	-2517.123777	0.681106	-2516.442671	13.4
^{4}TS15	-2516.559394	-2517.123139	0.680743	-2516.442396	13.6

[a] All calculations were performed in the quartet state under the unrestricted formalism except for **1** (singlet, restricted).

[b] Energy obtained from geometries optimized at the B3LYP-D3 / 6-31G* + LANL2DZ(Fe) level of theory.

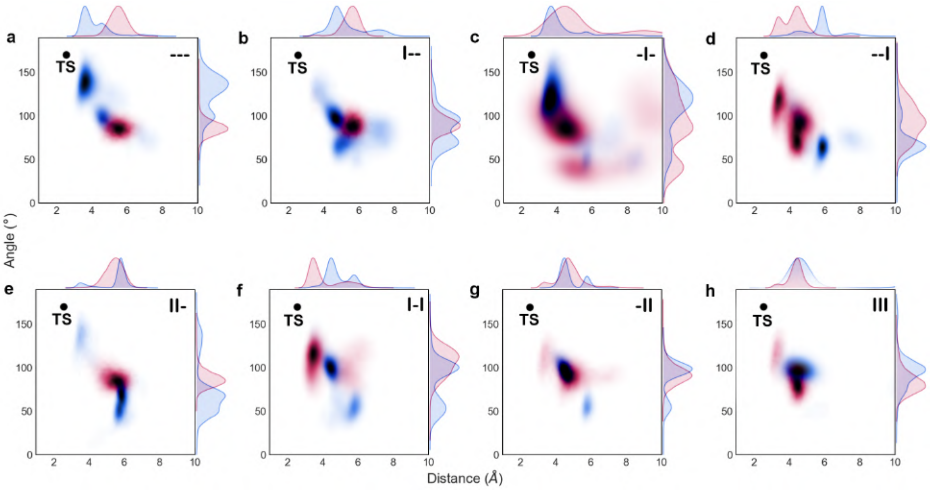
[c] Energy obtained from single point calculation at the B3LYP-D3 / 6-311G** + LANL2DZ(Fe) level of theory on top of the optimized geometry.

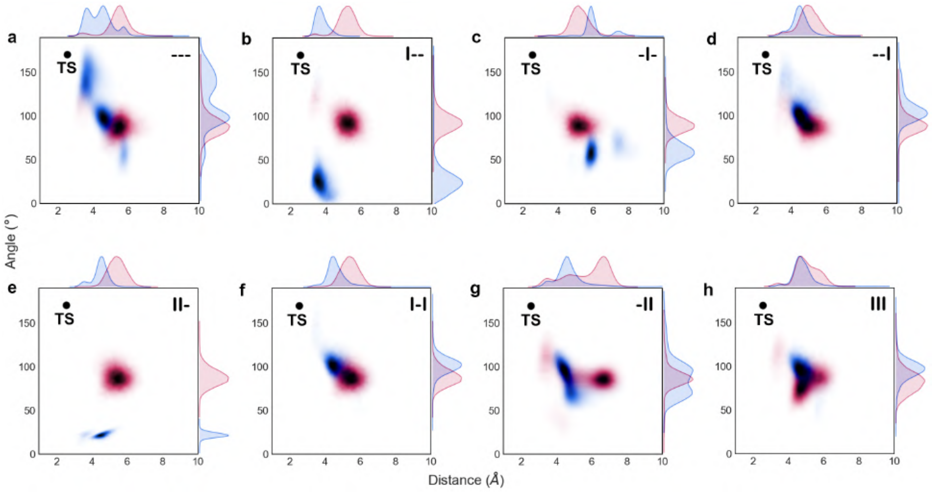
[d] Thermal correction at 298.15 K obtained from frequency calculation on top the optimized geometry at the same level of theory.

[e] Energy calculated as E_{el} + G correction.

[f] 1 Hartree = 627.5 kcal/mol.

[g] Energy and thermal correction calculated as ^{4}CpdI + 1.





Figures C2-C3. KDE plots of key geometric parameters for hydroxylation. Distances determined between the oxygen atom of heme-Fe=O and the C-atom of **1** (x-axis) and angles formed by (heme-Fe)O–H_β–C2/15(**1**) (C2 in pose 2 or C15 in pose 15) of **1** (y-axis) from the third replica of the MD dataset of all mutants. Geometric parameters measured for C2 and C15 are shown in red and blue, respectively. The ideal distance and angle for the TS (black dot) corresponds to the DFT-optimized geometry of the transition state (TS) for the C–H abstraction by heme-Fe=O using a truncated computational model.

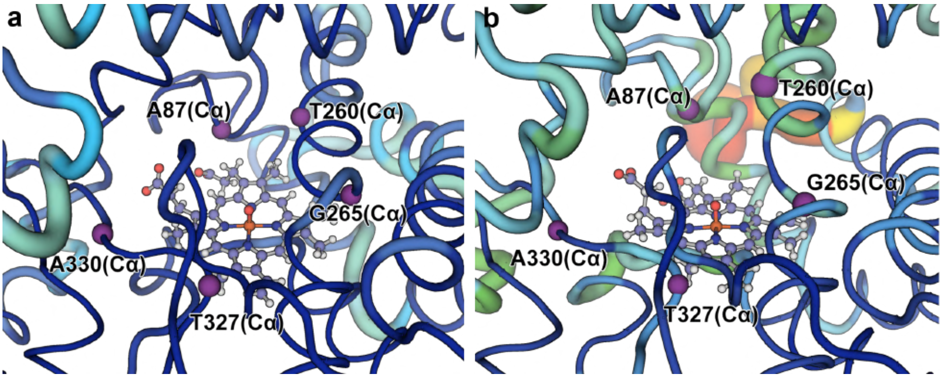


Figure C4. PC Analysis (PC2) of the substrate-bound mutant **-II** (a) and **III** (b). The thickness of the line is proportional to the motion and the color scale varies from blue (minimum motion) to red (maximum motion). Cα of selected residues are shown in purple spheres.

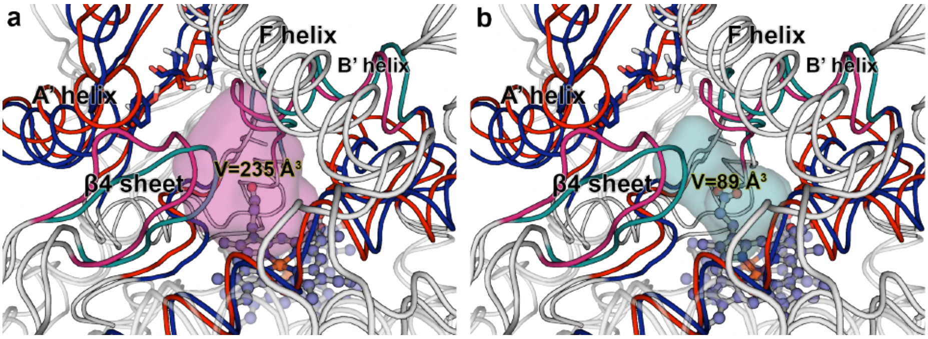


Figure C5. Computed active site volume for mutant **III** (a) and --- (b). Mutant --- is shown in blue whereas mutant III is shown in red. The F-G loop and the β1 sheet / active site volume are highlighted in teal and magenta color for --- and **III**, respectively. The volumes were calculated on the most populated cluster which resulted joining the 3 different replicas of the apo trajectories.

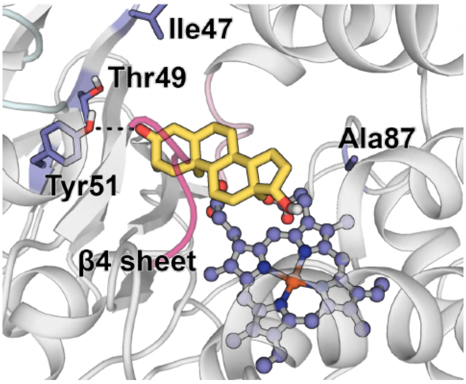


Figure C6. (**1**)CO–Y51 polar interaction in the substrate binding trajectory of mutant **I--**.

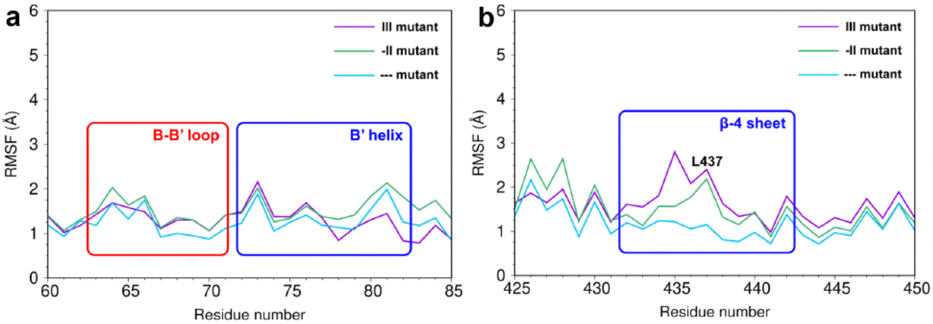


Figure C7. Root-mean square fluctuation (RMSF) of selected backbone elements in mutant ---, **-II** and **III**. RMSF measured from the simulation in the absence of substrate for variant --- (cyan). **-II** (green) and **III** (purple) for: a) B-B' loop region (highlighted with a red box) and B'-helix (blue box), and b) β4 sheet (blue box). Variants III and -II show higher RMSF values of the β4 sheet region compared to ---, whereas the RMSF is rather comparable for all variants over the B-B' loop and B' helix regions.

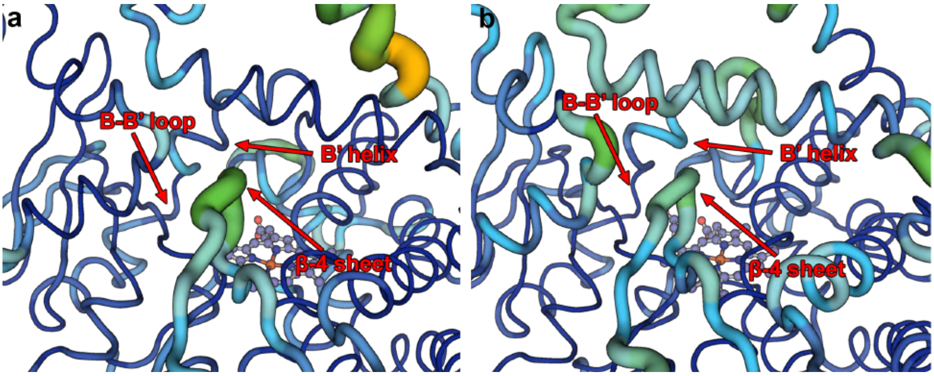


Figure C8. PC Analysis (PC2) of substrate-free variant **III** (a) and **-II** (b). The thickness of the line is proportional to the motion and the color scale varies from blue (minimum motion) to red (maximum motion). Variants **III** and **-II** show higher flexibility of the β4 sheet region compared to ---, with little or no impact on the flexibility of the the B-B' loop and B' helix regions.

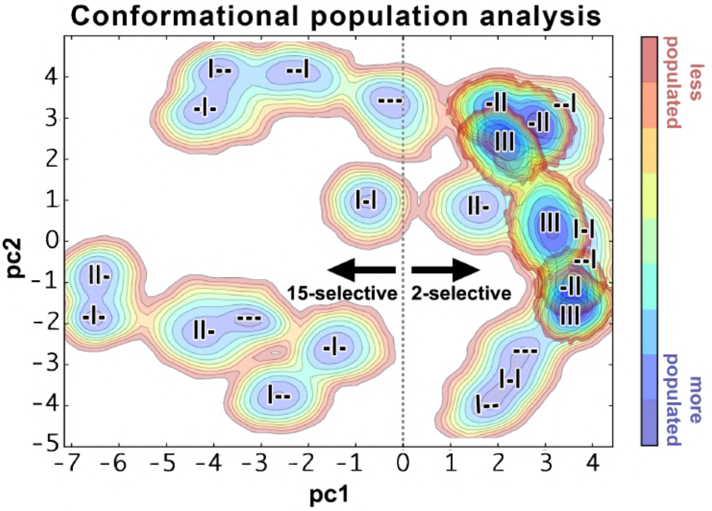


Figure C9. Conformational population analysis (PC1, x-axis; PC2, y-axis) of deconvolution mutants. The population of the 2β-selective variant **-II** and **III** are highlighted.

Additivity terms and equations

Mathematically, the addition between A and B can be determined according to the trait in quest.

In the case of stereoselectivity or enantioselectivty, the free energy of interaction ($\Delta G^{\neq}_{AB}$) between two mutations (or sets of mutations) A and B can be calculated using equation C1:

$$\Delta G^{\neq}_{AB} = \Delta\Delta G^{\neq}_{exp} - (\Delta\Delta G^{\neq}_{A} + \Delta\Delta G^{\neq}_{B}) \quad \text{(C1)}$$

Where $\Delta\Delta G^{\neq}_{exp}$ is the difference in activation energy with respect to regioselectivity (**2** versus **3** formation and all the rest of the stereo- and regioisomers) experimentally obtained for the binary combination, while $\Delta\Delta G^{\neq}_{A}$ and $\Delta\Delta G^{\neq}_{B}$ are the experimental energies obtained for each mutant (or sets of mutations) independently.

In the case of conversion (or any other trait), herein we introduce equation C2:

$$\Delta C^{\%}_{AB} = \Delta C^{\%}_{exp} - \left(\Delta C^{\%}_{A} + \Delta C^{\%}_{B}\right) \quad \text{(C2)}$$

Where $\Delta C^{\%}_{exp}$ is the difference in conversion for **2** and the parent mutant F87A experimentally obtained for the binary combination, while $\Delta\%C_A$ and $\Delta\%C_B$ are the differences in the experimental conversion values between each mutant (or sets of mutations) separately and that of the parent mutant F87A (28 ± 1%): The conversion values are not logarithmic (in contrast to the selectivity ones) and thus the value of the parental mutant F87A has to be subtracted from the single-point mutants and the double mutant.

In any case, additivity for selectivity (conversion) occurs when $\Delta G^{\neq}_{AB} = 0$ ($\Delta C^{\%}_{AB} = 0$), while positive (+ME) and negative (-ME) magnitude epistasis correspondingly pertains when $\Delta G^{\neq}_{AB} > 0$ ($\Delta C^{\%}_{AB} > 0$) and $\Delta G^{\neq}_{AB} < 0$ ($\Delta C^{\%}_{AB} < 0$). Positive (+SE) and negative (-SE) sign epistasis occurs when $\Delta G^{\neq}_{AB} > 0$ ($\Delta C^{\%}_{AB} > 0$) if $A < 0$ *or* $B < 0$ and $\Delta G^{\neq}_{AB} < 0$ ($\Delta C^{\%}_{AB} < 0$) if $A > 0$ *or* $B > 0$, respectively. Finally, positive (+RSE) and negative (-RSE) reciprocal sign epistasis occurs when $\Delta G^{\neq}_{AB} > 0$ ($\Delta C^{\%}_{AB} > 0$) if $A < 0$ *and* $B < 0$ or $\Delta G^{\neq}_{AB} < 0$ ($\Delta C^{\%}_{AB} < 0$) if $A > 0$ *and* $B > 0$, respectively[28].

To exemplify the various cases of additivity/non-additivity, if the desired fitness trait is **2β-hydroxytestosterone selectivity**, the following theoretical examples are given:

1. **Additivity:** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 79% 2β sel. (3.3 kJ/mol). Mutant AB is 90% 2β sel. (5.4 kJ/mol).
2. **Positive Magnitude Epistasis (+ME):** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 63% 2β sel. (1.3 kJ/mol). Mutant AB is 90% 2β sel. (5.4 kJ/mol).
3. **Positive Sign Epistasis (+SE):** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 63% 15β sel. (-1.3 kJ/mol). Mutant AB is 90% 2β sel. (5.4 kJ/mol).

4. **Positive Reciprocal Sign Epistasis (+RSE):** Mutant A is 70% 15β sel. (-2.1 kJ/mol). Mutant B is 63% 15β sel. (-1.3 kJ/mol). Mutant AB is 90% 2β sel. (5.4 kJ/mol).
5. **Negative Magnitude Epistasis (-ME):** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 63% 2β sel. (1.3 kJ/mol). Mutant AB is 60% 2β sel. (1.0 kJ/mol).
6. **Negative Sign Epistasis (-SE):** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 63% 15β sel. (-1.3 kJ/mol). Mutant AB is 90% 15β sel. (-5.4 kJ/mol).
7. **Negative Reciprocal Sign Epistasis (-RSE):** Mutant A is 70% 2β sel. (2.1 kJ/mol). Mutant B is 63% 2β sel. (1.3 kJ/mol). Mutant AB is 90% 15β sel. (-5.4 kJ/mol).

Using equations C1 and C2, the degree of interaction (if any) between two mutations (or sets of mutations) can be calculated.

APPENDIX D

Supplementary information Chapter 7

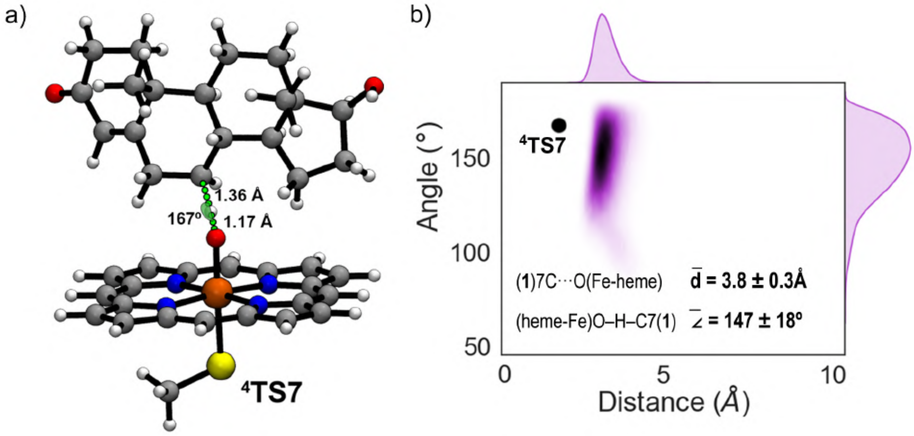


Figure D1. (a) DFT transition-states for the H-abstraction in testosterone (**1**) from C7. (b) KDE plot of key geometric parameters for hydroxylation. Distances determined between the oxygen atom of the Fe=O and the C7 atom of **1** (x axis) and angles formed by (heme-Fe)O–H–C7(**1**) (y axis) along the whole simulation time (3 joint MD replicas). The ideal distance and angle for the ^{4}TS (black dot) corresponds to the Density Functional Theory (DFT) optimized geometry for the 7C–H abstraction by CpdI using a truncated computational model. Mean distance (Å), mean angle (°) and corresponding standard deviations are shown.

Table D1. Energies, thermal corrections and Gibbs free energies of the QM structures calculated for H-abstraction catalyzed by heme Compound I (CpdI).

Struct.[a]	**E_{el}**[b]	**E_{el}**[c]	**G correction**[d]	**G**[e]	**ΔG**
	(Hartree)[f]	**(Hartree)**[f]	**(Hartree/particle)**[f]	**(Hartree)**[f]	**(kcal/mol)**
^{4}CpdI	-1625.181930	-1625.492613	0.267701	-1625.224912	
1	-891.402149	-891.636974	0.397895	-891.239079	
^{4}RC[g]	- 2516.584079	-2517.129587	0.665596	-2516.463991	
^{4}TS7	-2516.555582	-2517.119657	0.680317	-2516.439340	15.5

[a] All calculations were performed in the quartet state under the unrestricted formalism except for **1** (singlet, restricted).

[b] Energy obtained from geometries optimized at the B3LYP-D3 / 6-31G* + LANL2DZ(Fe) level of theory.

[c] Energy obtained from single point calculation at the B3LYP-D3 / 6-311G** + LANL2DZ(Fe) level of theory on top of the optimized geometry.

[d] Thermal correction at 298.15 K obtained from frequency calculation on top the optimized geometry at the same level of theory.

[e] Energy calculated as E_{el} + G correction.

[f] 1 Hartree = 627.5 kcal/mol.

[g] Energy and thermal correction calculated as ^{4}CpdI + 1.

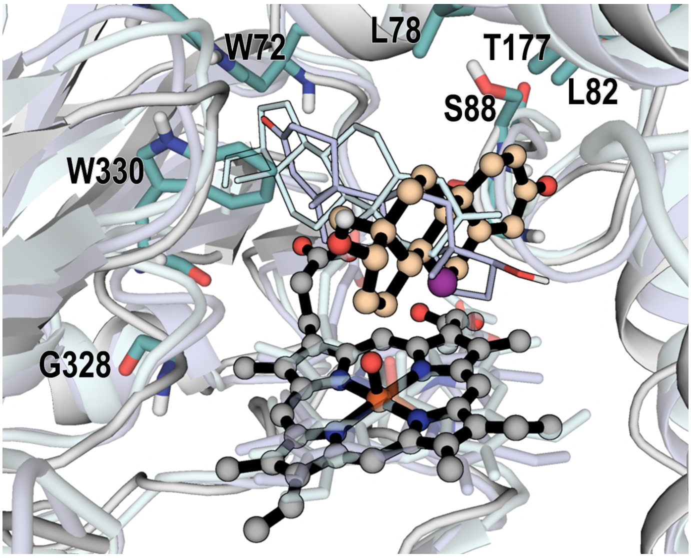


Figure D2. Overlay of representative snapshots from the MD simulations of mutant **LG23** with **1** bound in pose 7 (black sticks and light orange spheres) and mutant F87A with **1** bound in pose 2 (pale cyan sticks) and pose 15 (violet sticks). In mutant **LG23** pose7, the 7C atom of **1** is highlighted in violet color and the mutated residues are shown in teal color sticks.

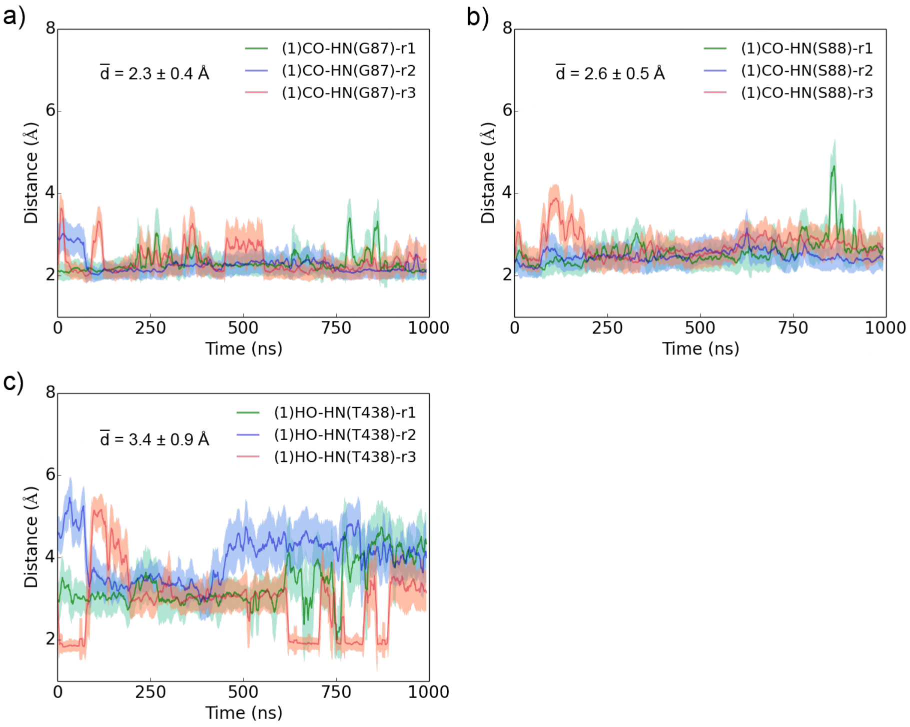


Figure D3. Plot of the distances (a) (**1**)CO···HN(G87); (b) (**1**)CO···HN(S88) and (c) (**1**)CO···HN(T438) of the three MD replica dataset. Mean values (in Å) and standard deviations for the whole MD replica dataset are shown.

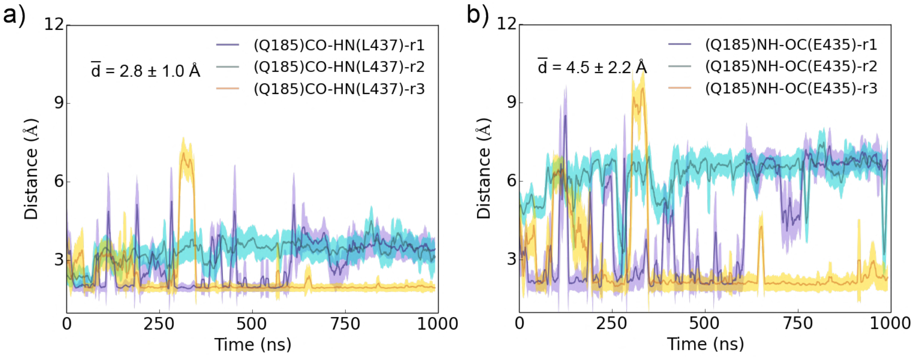


Figure D4. Plot of the distances (a) (Q185)NH···OC(E435) and (b) (Q185)CO···HN(L437) for the three MD replica dataset. Mean values (in Å) and standard deviations for the whole MD replica dataset are shown.

Data availability

Input files and scripts used to carry out the projects in this thesis are available here: https://github.com/ldamore