Investigation on the Molecular Mechanism of 2OG-Oxygenases and Serine β-Lactamases through Computational Chemistry

---

Thesis submitted for the

Degree of Doctor of Philosophy

Department of Chemistry

Trinity 2016

Hwanho Choi



Hertford College

University of Oxford

Abstract

This thesis describes studies on the molecular mechanisms of 2OG-oxygenases and serine β-lactamases, and work towards the decomposition of the thermodynamics of solvation processes using Quantum Mechanics (QM), Molecular Dynamics (MD), and Free Energy Perturbation (FEP) methods.

The first chapter gives a brief introduction to quantum chemistry and free energy perturbation methods as applied to the concept of potential energy surface and transition state theory. Chapter 2 describes QM and MD calculations to investigate recyclisation and hydrolysis in avibactam mediated serine β-lactamase inhibition. β-Lactams inhibit penicillin-binding proteins (PBPs) and serine β-lactamases by acylation of a nucleophilic active site serine. Avibactam is approved for clinical use in combination with a cephalosporin, and is a breakthrough non β-lactam inhibitor, which also inhibits via serine acylation. QM and MD calculations on the avibactam-mediated inhibition of a clinically relevant cephalosporinase reveal recyclisation of the avibactam derived carbamoyl complex is favored over hydrolysis. In contrast, analogous recyclization in β-lactam mediated inhibition is disfavored. Avibactam recyclization is promoted by a proton shuttle, a 'structural' water protonating the nucleophilic serine, and stabilization of negative charge on the avibactam derived carbonyl oxygen.

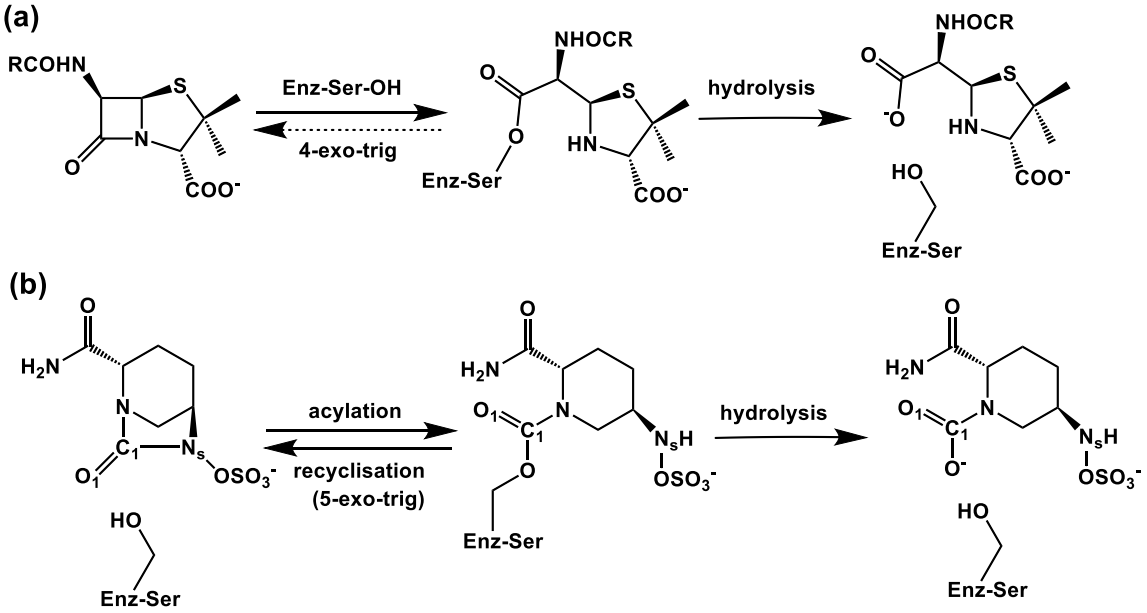


Figure 1. β-Lactams react irreversibly with penicillin binding proteins/β-lactamases to form an acyl-enzyme complex in contrast to avibactam which reacts more reversibly. C1 and N_S denote carbonyl carbon of ring-opened avibactam and the nitrogen of the avibactam derived hydroxylamine-O-sulfonate group, respectively.

The results reveal the potential of calculations for distinguishing between bifurcating pathways and in generating hypotheses for predicting resistance. The inability of β-lactams to undergo recyclization may be an Achilles heel, but one that might be addressed by reversibly binding inhibitors.

Chapter 3 describes work using DFT calculation on the selectivity of the hydroxylation of ankyrin repeat domain substrates as catalyzed by a 2-oxoglutarate (2OG) dependent oxygenase. Factor inhibiting hypoxia inducible factor (FIH) is a promiscuous protein hydroxylase that typically catalyses hydroxylations at the C-3 position of protein residues. Amongst residues hydroxylated by FIH, leucine has drawn special interest from a stereochemical perspective. Based on *ab initio* density functional calculations at the B3LYP/def2-TZVP level of theory, structural and energetic features of the FIH catalysed hydroxylation of L- and D-leucine were investigated. The DFT and experimental results show that L- and D-leucine are oxidized to (*3R*)-β-hydroxy-L-leucine and (*3S*)-β-hydroxy-D-leucine, respectively. D-, but not L, -leucine undergoes a double hydroxylation to form β,γ-dihydroxy-D-leucine. The results imply that the difference in the catalytic actions of FIH on L- and D-leucine stems from the different complexation modes in the active site caused by a change in C-α stereochemistry. X-ray crystallography analyses provide evidence for the proposed catalytic mechanisms.

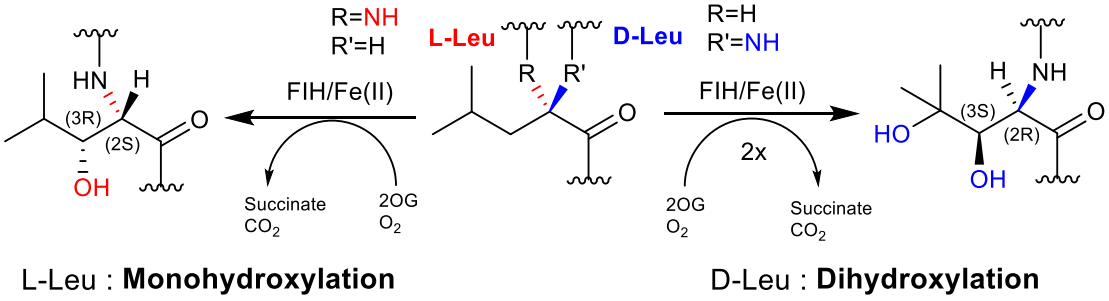


Figure 2. Single and double hydroxylations of L- and D-leucine, respectively, in FIH-catalyzed hydroxylation of an ankyrin substrate.

Despite the importance of molecular hydration entropy (ΔS_{hyd}) in chemical and biological processes, the accurate calculation of ΔS_{hyd} is difficult due to the complexities of solute-water interactions. The studies in chapter 4 concerned a novel method for accurately estimating ΔS_{hyd} based on a modified thermodynamic decomposition approach. The key feature of the method is that solute-water interactions are decomposed into hydrophobic and electrostatic parts to calculate their respective contributions to ΔS_{hyd}.

Although free energy perturbation (FEP) methods have been employed widely, the poor convergent behavior of the van der Waals interaction term in the potential function limits its accuracy and robustness. The new method combines the FEP approach and the scaled particle theory (or information theory) to separately calculate the electrostatic solute-water interaction term (ΔS_{elec}) and the hydrophobic contribution (the cavity formation) entropy (ΔS_{cav}). The method appears to be effective with a substantial accuracy enhancement in ΔS_{hyd} estimation compared to conventional FEP calculations. ΔS_{cav} appears to dominate over ΔS_{elec} in magnitude even in the case of polar solutes, implying that the major contribution to the entropic cost for hydration comes from the formation of a solvent-excluded volume. The method thus enhances the accuracy of ΔS_{hyd} prediction by complementing the conventional full FEP method.

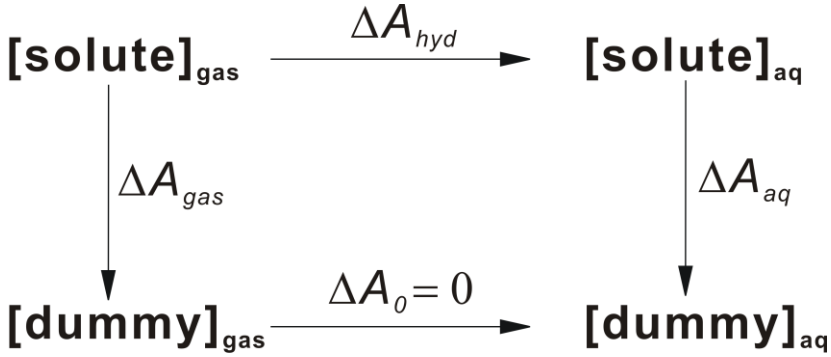


Figure 3. Thermodynamic cycle for calculating the hydration free energies of solute molecules. Note that $\Delta A_0 = 0$ because solute is not involved in the process.

Contents

Abstract ii
Contents v
Acknowledgements viii
Glossary of abbreviations and acronyms ix
Author's declaration xii

Chapter 1: Brief Introduction of Quantum Chemistry & Free energy Perturbation

1.1. Quantum Chemistry 2
1.1.1. Schrödinger equation 2
1.1.2. Born-Oppenheimer approximation 3
1.1.3. Self-Consistent Field (SCF) Theory 4
1.1.4. Electron Correlation 6
1.1.5. Basis sets 6
1.1.6. Density Functional Theory (DFT) 8
1.2. Statistical Mechanics 11
1.2.1. Molecular Dynamics (MD) 11
1.2.2. Free Energy Perturbation (FEP) 14
1.2.3. Alchemical Transformation 15
1.2.4. FEP of Computer-aided Molecular Design 17
1.2.5. Thermodynamic Integration (TI) 18
1.3. Potential Energy Surface (PES) 19
1.3.1. Reaction Coordinates 19
1.3.2. Transition State Theory 21
1.3.3. Rate Reaction Theory (Kramer's Theory) 22
1.4. Hydrophobicity 23
1.4.1. Overview of Scaled Particle Theory (SPT) 23
1.5. Homology Modeling 25

Chapter 1 References 27

Chapter 2: Investigations on Recyclisation and Hydrolysis in Avibactam Mediated Serine β-Lactamase Inhibition

2.1. Introduction 31
2.2. Computational methods 36
2.2.1. MD simulation 36
2.2.2. Quantum Mechanical Calculations 40
2.3. Results 41
2.3.1. MD simulation studies for the protonation state of Glu166 of CTX-M-15 β-lactamase in complex with avibactam 41
2.3.2. MD simulations on the possibility of recyclisation of the β-lactamase-avibactam complex 43
2.3.3. Avibactam recyclisation in the active site of CTX-M-15 47
52

2.3.4. Mechanism of hydrolysis in the CTX-M-15 active site
2.3.5. Putative recyclisation of oxacillin mechanism in the OXA-1 active site 55
2.3.6. Hydrolysis mechanism of desulfated avibactam in the active site of KPC-2 59
2.3.7. Energetic and entropic contributions to the enzymatic reaction 65
2.4. Conclusions 69

Chapter 2 References 72

Chapter 3: A Human Protein Hydroxylase that accept D-Amino acids

3.1. Introduction 76
3.2. Experimental Methods 79
3.2.1. Computational Methods 79
3.2.1.1. Quantum chemical investigations of catalytic hydroxylations of D-/L-leucine in the active site of FIH 79
3.3. Results and Discussion 81
3.3.1. Initial L-/D-allylglycine hydroxylation 81
3.3.2. Unexpected dihydroxylation of D-leucine substrate by FIH 82
3.3.2.1. Experimental confirmation of dihydroxylation of D-leucine by FIH 82
3.3.2.2. DFT investigations of the single hydroxylation reaction 83
3.3.2.3. DFT calculations on the second hydroxylation reactions 88
3.3.2.4. Amino acid analyses of stereochemical hydroxylated D-leucine 93
3.3.2.5. Chemical determination of diol 95
3.3.3. Crystallography of FIH with substrates 96
3.3.3.1. Overall structures of FIH in complex with D-leucine, D-β-(2R,3S)-hydroxyleucine, and D-allylglycine CAP 96
3.3.3.2. Crystallographic analysis of bifurcating monohydroxylation vs. dihydroxylation by stereochemical D-/L-leucine 98
3.4. Conclusions 100

Chapter 3 References 103

Chapter 4: Computational Prediction of Molecular Hydration Entropy with Hybrid Scaled Particle Theory and Free Energy Perturbation Method

4.1. Introduction 106
4.2. Theory and Computational methods 108
4.2.1. Hybrid Scaled Particle Theory and Free Energy Perturbation 108
4.2.2. Scaled Particle Theory 110
4.2.3. Information Theory 111

4.2.4. Implementation of electrostatic terms using Free Energy Perturbation 112
4.2.5. Comparison of hybrid SPT/FEP with full FEP 116
4.3. Results 119
4.3.1. Correlation for cavity formation of small molecules with SPT and IT 120
4.3.2. Calculated hydration entropies vs. Experimental hydration entropies 120
4.3.2.1. Calculated hydration entropies of hydrophobic molecules 123
4.3.2.2. Calculated hydration entropies of hydrophilic molecules 124
4.3.2.3. Correlation of calculated hydration entropies vs. experimental entropies 125
4.3.3. Results of Comparison of hybrid SPT/FEP with full FEP 127
4.3.4. Hydration free energy of thermodynamic decomposition 129
4.4. Inclusion of self-solvation terms in hydration free energy 132
4.4.1. Derivation of self-solvation free energies based on solvent-contact model 132
4.4.2. Optimization of atomic volume parameters with genetic algorithm 135
4.4.3. Correlation for calculated self-solvation free energies vs. experimental solvation free energies 136
4.5. Application of hydration free energies to virtual screening 139
4.5.1. BRAF kinase inhibitors 139
4.5.2. High throughput Docking Simulations 140
4.5.3. Docking Simulation results with experimental confirmation. 143
4.5.4. Analysis of binding modes 145
4.6. Conclusions 149

Chapter 4 References 150

Chapter 5: Summary Chapter

5.1. Reversibility of avibactam β-Lactamase inhibition 155

5.2 Hydroxylation of D-amino acids by FIH 158

5.3 Computational prediction of molecular hydration entropies 164

5.4 Overall conclusions and future directions 168

Chapter 4 References 170

Appendices 172

Chapter 3 Appendices 173
Chapter 4 Appendices 175

Acknowledgements

Above all, I would like to express my sincerest gratitude to Professor Christopher Schofield for providing me with the opportunity to pursue this research in his lab. He has endless patience for collaboration with theoretical research, gradually deepening my understanding of both experimental and theoretical methods throughout my D.Phil.

I am also grateful to Drs. Mike McDonough, Zhihong Zhang, Onur Atasoylu, Marios Markoulides, Robert Paton and Ian Clifton with whom I have shared many important discussions, and to all the members of the Schofield lab, especially afternoon tea members, John Scotti, Weishen Aik and Shoichiro Horita who have made me feel like family in United Kingdom.

Thanks also to my collaborators in Korea, in particular Prof. Hwangseo Park and Sungwoo Hong, they introduced me to the chemistry world between physics, computer science and medicine. I am also very thankful to Dr. Hongsuk Kang who has been through a lot with me over the 20 years.

Finally, I would like to thank my family. My Father (Kichoon), Brother (Geunbae), and my lovely wife, Hojeong. Without all of your unconditional love, understanding and support I could never have made it this far, I am forever indebted to you all, thank you.

I dedicate this D.Phil to my mother in heaven, *Sangsook Lee*.

Glossary of abbreviations and acronyms

2OG	2-Oxoglutarate
AEC	Acyl-enzyme complex
AMBER	Assisted model building with energy refinement
AO	Atomic orbital
AspH	Aspartyl/asparaginyl-β-hydroxylase
B3LYP	Becke exchange with Lee-Yang-Parr correlation
BDE	Bond dissociation energy
BLAs	β-Lactam antibiotics
BSSE	Basis set superposition error
CA	Consensus ankyrin (sequence : HLEVVKLLLEHGADVNAQDK)
CCSD	Coupled cluster with singles and doubles
CCSD(T)	Coupled cluster with perturbative estimate for connected triples
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CMD	Concerted metalation deprotonation
CGenFF	CHARMM General Force Field
COSMO	Conductor-like screening model
CP	Carbopalladation
CPCM	Conductor-like polarizable continuum model
CTX-M	Cefotaxime, first isolated in Munich'
DFT	Density functional theory
DZ	Double zeta
ECP	Effective core potential
ESP	Electrostatic potential map
FEP	Free energy perturbation
FES	Free energy surface
FIH	Factor inhibiting HIF
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
GAMESS	General atomic and molecular electronic structure system
GS2	Gonzalez-Schlegel second-order methods
GTO	Gaussian-type orbital
HF	Hartree-Fock
HIF	Hypoxia-inducible factor
HOMO	Highest occupied molecular orbital
HPLC	High performance liquid chromatography
HS	Hard Sphere
IPNS	Isopenicillin N synthase
IRC	Intrinsic reaction coordinate
IT	Information theory
K_D	Dissociation constant
KIE	Kinetic isotope effect
KPC-2	Klebsiella pneumoniae carbapenemase 2
KS	Kohn–Sham
LANL	Los Alamos National Laboratory
LC	Liquid chromatography
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation

LSDA	Local spin density approximation
LUMO	Lowest unoccupied molecular orbital
M06-2X	Minnesota 06 density functional of Truhlar
MALDI-ToF	Matrix-assisted laser desorption/ionosation-time of flight
MC	Monte Carlo method
MD	Molecular dynamics
MO	Molecular orbital
MPA	Mulliken population analysis
MP2	Møller-Plesset second order perturbation theory
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NAMD	Nanoscale molecular dynamics program
NBO	Natural bond orbital
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOG	N-oxalylglycine
OXA	Oxacillinase
PBE	Perdew–Burke–Ernzerhof exchange-correlation functional
PBPs	Penicillin-binding proteins
PCM	Polarizable continuum model
PDB	Protein data bank
PES	Potential energy surface
PHD	Prolyl hydroxylase domain
PW91	Perdew–Wang 91 correlation functional
QM/MM	Quantum mechanics/molecular mechanics
RE	Reductive elimination
RESP	Restrained electrostatic potential
RHF	Restricted Hartree–Fock
ROHF	Restricted open-shell Hartree–Fock
RMSD	Root mean squared deviation
RMSE	Root mean squared error
RT	Room temperature
SAMPL	Statistical assessment of modelling of proteins and ligands
SASA	Solvent-accessible surface area
SBLs	Serine β-lactamases
SCRF	Self-consistent reaction field polarizable continuum model
SMD	Universal continuum solvation model
SO	Spin–orbit effects
SPT	Scaled particle theory
STO	Slater-type orbital
SZ	Single zeta
TD-DFT	Time-dependent density functional theory
TI	Thermodynamic integration
TIP3P	3-site water model
TOF	Turnover frequency
TS	Transition state (first order saddle point)
TST	Transition state theory
TZ	Triple zeta
UFF	Universal force field
UHF	Unrestricted Hartree–Fock

WAV	Water-accessible volume
WHAM	Weighted histogram analysis method
WT	Wild type
ωB97XD	Dispersion corrected density functional of Head-Gordon
ZPE	Zero-point (vibrational) energy

Author's declaration

The thesis is an account of work carried out by the author in the Department of Chemistry, University of Oxford under the supervision of Professor Christopher J. Schofield. No part of this thesis has previously been submitted for a degree in this University or elsewhere. The work of other authors has been freely drawn and is duly acknowledged in the text. A list of references is given at the end of each chapter.

Some of the doctoral work and consequently description that is described in this thesis has been, or will be, reported in the academic literature.

Choi, H., Paton, R., Park, H., Schofield, C. J. Investigations on Recyclization and Hydrolysis in Avibactam mediated serine β-Lactamase inhibition. *Org. Biomol. Chem.* **2016**, *14*, 4116-4128. (Work described in Chapter 2)

Choi et al. A Human Protein Hydroxylase that Accepts D-Residues. *Manuscript in preparation*. (Chapter 3)

Choi, H., Kang, H., Park, H. Computational Prediction of Molecular Hydration Entropy with Hybrid Scaled Particle Theory and Free Energy Perturbation Method. *J. Chem. Theory Comput.* **2015**, *11*, 4933-4942. (Work described in Chapter 4)

van Berkel, S. S., Nettleship, J. E., Leung, I. K., Brem, J., Choi, H., Stuart, D. I., Claridge, T. D., McDonough, M. A., Owens, R. J., Ren. J., Schofield, C. J. Binding of (5S)-Penicilloic Acid to Penicillin Binding Protein 3. *ACS Chem. Biol.* **2013**, *8*, 2112-2116. (Chapter 4)

Aik, W., Scotti, J. S., Choi, H., Gong, L., Demetriades, M., Schofield, C. J., & McDonough, M. A. Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res.* **2014**, *42*, 4741-4754.

Schiller, R., Scozzafava, G., Tumber, A., Wickens, J. R., Bush, J. T., Choi, H., Yeh, T. L., Chan, M.C., Mott, B. T., McCullagh, J. S., Maloney, D. J., Schofield, C. J., Kawamura, A. A cell-permeable ester derivative of the JmjC histone demethylase inhibitor IOX1. *ChemMedChem*, **2014**, *9*, 566-571.

Scotti, J. S., Leung, I. K., Ge, W., Bentley, M. A., Paps, J., Kramer, H. B., Lee, J., Aik, W., Choi, H., Paulsen, S. M., Bowman, L. A., Loik, N. D., Horita, S., Ho, C. H., Kershaw, N. J., Tang, C. M., Claridge, T. D., Preston, G. M., McDonough, M. A., Schofield, C. J. Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13331-13336.

Daoud, H., Zhang, D., McMurray, F., Yu, A., Luco, S. M., Vanstone, J., Jarinova O, Carson N, Wickens J, Shishodia, S., Choi, H., McDonough, M. A., Schofield, C. J., Harper, M. E., Dyment, D. A., Armour, C. M. Identification of a pathogenic FTO mutation by next-generation sequencing in a newborn with growth retardation and developmental delay. *J. Med. Genet.*, **2016**, *53*, 200-207.

Choi, H., Min, M., Peng, Q., Kang, D., Paton, R. S., Hong, S. Unraveling innate substrate control in site-selective palladium-catalyzed C–H heterocycle functionalization. *Chem. Sci.* **2016**, *7*, 3900-3909.

Zhang, Z., Smart, T. J, Choi, H., Hardy, F., Lohans, C. T., Richardson, M. S. W., Paton, R. S., McDonough, M. A., Schofield, C. J. Structural and Stereoelectronic Insights into Oxygenase Catalysed Formation of Ethylene from 2-Oxoglutarate. *Manuscript for publication.*

Dr. Hwanho Choi M.D

Reprinted with permission from above references. Copyright © 2016 Royal Society of Chemistry & American Chemical Society

Chapter 1

Brief Introduction of Quantum Chemistry & Free Energy Perturbation

1.1 Quantum Chemistry

1.1.1 Schrödinger equation

Much progress has been made in recent decades in investigating the catalytic mechanism of enzymatic reactions by computational methods.[1] The advent of sophisticated computational methods, including *ab initio* quantum chemistry methods has made it possible to rationalize many biochemical reactions in terms of their electronic structure changes. The fundamental concept behind *ab initio* calculations is the Schrödinger equation for the time evolution of the molecular wavefunction (Ψ), which can be expressed as follows:

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \hat{H}\Psi(r,t) \qquad (1).$$

Where i, $\hbar$, and $\Psi(r,t)$ indicate the imaginary unit, Planck constant divided by 2π and the time-dependent wavefunction, respectively. $\hat{H}$ is the Hamiltonian operator that can be written in the following form:

$$\hat{H} = \frac{\mathbf{p}^2}{2m} + V(r) \qquad (2).$$

The Hamiltonian operator contains terms for the kinetic energy of the electrons and nuclei, and for electrostatic interactions involving electrons and nuclei. In general, the Schrödinger equation can be solved exactly only for simple systems such as harmonic oscillators, rigid rotors, hydrogen atoms, and some diatomic systems. In the case of complex systems including many electron and atoms, we need to make some approximations to make the Schrödinger equation solvable with numerical methods. For simplicity, Ψ can be separated into space and time parts in the absence of relativistic

effect.[2] If the Hamiltonian operator is time-independent, Ψ may be expressed as the product of space and time parts based on the separation of variables.

$$\Psi(r,t) = \Psi(r)\varphi(t) \qquad (3).$$

where Ψ and φ are only a function of space (r) and time (t), respectively. Scaling of relativistic effects can be assumed to be Z^4/c^2 where Z is the atomic number of the nucleus in Mendeleev's periodic tables and c = 137 au; the velocity of the electron is approximately close to the speed of light. For very heavy atoms such as I, Cs, Pt, Au, Hg, Pb, U, and so on, $(Z/c)^2$ reaches an appreciable magnitude, and relativistic effects may alter the chemical and physical behavior *qualitatively*. In particular, in the lower third of the periodic table, the chemical characteristics of the elements are strongly influenced by relativity.[3]

1.1.2 Born-Oppenheimer Approximation

For a system with N-particles, the time independent Schrödinger equation can be written as follows:

$$\widehat{H} = -\frac{\hbar^2}{2}\sum_\alpha \frac{1}{m_\alpha}\nabla_\alpha^2 - \frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_\alpha \sum_{\alpha>\beta i} \frac{Z_\alpha Z_\beta e'^2}{\gamma_{\alpha\beta}} - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{\gamma_{i\alpha}} + \sum_j \sum_{i>j} \frac{e'^2}{\gamma_{ij}} \qquad (4).$$

The first, second, third, fourth, and fifth terms indicate the kinetic energies of nuclei, those of electrons, nuclear-nuclear repulsion energies, attractive interactions between nuclei and electrons, and electron-electron repulsion energies, respectively. Since nuclei are much heavier than electrons, the former move even more slowly than the latter to the extent to be considered stationary. The Schrödinger equation can therefore be separated into two parts: one that describes the electronic wavefunction for a fixed nuclear geometry, and the other part which refers to the nuclear component which governed by the potential energy of the electronic part. This is called Born-Oppenheimer

approximation in which the electronic wavefunction depends only on the positions of the nuclei because the kinetic energy of the nuclei is factored out.[4] Within the framework of Born-Oppenheimer approximation, the total wavefunction is given by the product of the electronic wavefunction Ψ_{el} and the nuclear counterpart Ψ_N:

$$\Psi(R,r) = \Psi_{el}(r;R)\,\Psi_N(R) \qquad (5).$$

One can now obtain the total wavefunction by solving the electronic Schrödinger equation to describe a particular electronic state. On the basis of the Born-Oppenheimer approximation, the electronic Hamiltonian can be expressed in the following form:

$$\widehat{H}_{el} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_\alpha \sum_i \frac{Z_\alpha e'^2}{\gamma_{i\alpha}} + \sum_j \sum_{i>j} \frac{e'^2}{\gamma_{ij}} \qquad (6).$$

The eigenfunction of this electronic Hamiltonian is referred to as the potential energy surface (PES) using which the nuclear part of Schrödinger equation can also be solved with the following Hamiltonian operator.

$$\widehat{H}_N = -\frac{\hbar^2}{2}\sum_\alpha \frac{1}{m_\alpha}\nabla_\alpha^2 + U(R) \qquad (7).$$

With the nuclear wavefunction in hand, one can calculate the useful physical quantities such as the molecular vibrational frequencies and rotational energy levels.

1.1.3. Self Consistent Field (SCF) Theory

For many-electron systems, it is still difficult to exactly solve the Schrödinger equation due to the presence of many indistinguishable and interacting electrons. Besides the Born-Oppenheimer approximation, the Ψ_{el} wavefunction needs to be further approximated in such a way to be expanded using the Slater determinant.[5] Hartree suggested a method for solving the Schrödinger equation for an electron moving in the potential of nuclei.[1] One can use the variational principle to generate the best trial

function that has a number of adjustable parameters. The electronic energy should be minimized as a function of these parameters. Fock improved Hartree's method using the properly antisymmetrised wavefunction given by the Slater determinant instead of the one-electron wavefunctions in which the exchange interactions are excluded. The molecular orbitals comprising the Slater determinant can be obtained by numerically solving the following equation:

$$\varepsilon_i \Psi_i(r) = \left(-\frac{1}{2}\nabla^2 + \sum_j V(r)\right)\Psi_i(r) + \sum_j \int dr' \frac{|\Psi_j(r')|^2}{|r-r'|}\Psi_i(r) - \sum_j \delta_{\sigma_i\sigma_j} \int dr' \frac{\Psi_j^*(r')\Psi_i(r')}{|r-r'|}\Psi_j(r) \qquad (8),$$

where ε_i is an energy, ψ_i is the wavefunction, r is a position vector, σ_i is a spin function of ith electron, and δ is the Kronecker delta. The first term includes the kinetic energy and electrostatic potential energy of the ith electron itself interacting with the nuclei. The second term is the repulsive electrostatic potential energy from *all* electrons in the system including an unphysical self-interaction of the electrons. The third term which is also referred to as the exchange term, acts only on the electrons with the same spin. Because each electron is a fermion with a spin of eigenvalue $1/2$, the total electronic wavefunction must be antisymmetric with respect to the interchange of any two electron coordinates. Fock introduced Slater determinant to construct an antisymmetric wavefunction in order to it satisfied the Pauli exclusion principle that no two electrons can have the same set of quantum numbers:[6-8]

$$\Phi_0 = \frac{1}{\sqrt{n!}}\begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \dots & \phi_n(r_1) \\ \phi_1(r_2) & \phi_2(r_2) & \dots & \phi_n(r_2) \\ \cdots & \cdots & \ddots & \cdots \\ \phi_1(r_n) & \phi_2(r_n) & \dots & \phi_n(r_n) \end{vmatrix} = \left|\phi_1, \phi_2 \dots \phi_n\right| \qquad (9)$$

Here columns and rows indicate the molecular orbitals given by one-electron wavefunctions and the electron coordinates, respectively. An iterative and self-consistent method can be applied to solve Hartree-Fock equation to get molecular orbitals. A set of molecular orbitals obtained by solving this equation are called self-consistent field (SCF) orbitals.[9,10]

1.1.4. Electron Correlation

A drawback of the Hartree-Fock method is that it neglects instantaneous electron-electron interactions. In this regard, the electron correlation energy is defined as the difference between the energy calculated with exact wave-function and that obtained with Hartree-Fock wavefunction:

$$E_{corr} = E_{exact} - E_{HF} \qquad (10).$$

In order to account for electron correlation, configuration interaction (CI) uses a variational wave function that is a linear combination of Slater determinants (Φ_i). The expansion coefficients (d_i) are determined by imposing the limitation that the energy should be a minimum. The MOs used for building the excited determinants (Φ_i) are taken from Hartree–Fock calculation and held fixed:

$$\Psi_{el} = d_0\Phi_{HF} + \sum_{i=1} d_i\Phi_i \qquad (11).$$

Møller and Plesset[10] established perturbation theory for complex molecular systems to cope with electron correlation. In the second-order correction (MP2) method the wavefunction is expressed as a linear combination of the solutions for the Hartree-Fock equation. Thus, method is thus different from the CI method that assumes the exact molecular wavefunction as a linear combination of Slater determinants.[12,13] If all the excited states are taken into account to construct Φ_i determinants, the exact wavefunction for the given basis set can be obtained. Because such full CI calculations are not computationally feasible, the number of Φ_i should be limited in practice. In practice, the

CI methods that include all determinants of single- and double-excitations are most popular for the estimation of electron correlation.[14,15]

1.1.5. Basis sets

Hartree-Fock equations can be solved to obtain the molecular orbitals to construct the Slater determinant. Each molecular orbital need to be also expanded with the atomic orbitals that can in turn expressed with atomic basis sets. A basis set is a set of functions that are combined in linear combinations to produce the molecular orbitals. All expansion coefficients can be optimised using the variation principle. The variational principle states that 'If a system has a ground state energy of E_1, and ϕ is a normalized, well-behaved function of the system, then E_1 is no higher than the value of the variational integral W: $W = \int \phi^* \hat{H} \phi d\tau \geq E_1$. Therefore, for any guess ϕ, the ground state energy of the system E_1 is no higher than W, one should try to vary ϕ, and refine a form of it that minimises W. Each molecular orbital is thus expanded with a set of basis functions centered at the nuclei. The molecular orbital obtained in this way is called linear combination of atomic orbitals-molecular orbitals (LCAO-MO).[16]

$$\phi_i = \sum_k^M c_{ki} \chi_k \qquad (12)$$

The coefficients in the molecular orbital, c_{ki}, can be determined using the SCF procedure. The basis functions, χ_k, are atom-centered functions that mimic the solutions of the Schrödinger equation for the hydrogen atom. However, the exact atomic orbitals cannot be used as the basis functions even in the case of simple molecular systems due to the difficulty in calculating the two-center integrals. In the early stage of quantum chemistry, Slater type orbitals (STOs)[17] were used as the basis functions due to the similarity to the eigenfunctions of the hydrogen atom.

$$\phi^{STO}(r) = e^{-\zeta} \qquad (13)$$

The parameter ζ is Slater orbital exponent. STOs have an advantage in that they have direct physical interpretation and can thus serve as the basis for molecular orbitals. From a computational point of view, STOs have a drawback in that most of the required integrals in the course of the SCF procedure must be calculated numerically with a consequent computational burden. Therefore, most electronic structure calculations use Gaussian type orbital (GTO) functions (e^{-r^2}) because they are computationally much more efficient than STOs. Initially, the use of GTO to represent the basis functions was suggested by Sir John Pople and is as known as X-YZG series basis sets. Because only valence electrons are involved in most chemical processes, core and valence electrons are treated separately to save computational time and simultaneously to describe the valence electrons more rigorously than core ones.[18,19]

Correlation-consistent basis sets, known as cc-pVnZ series, were developed by Dunning and coworkers to describe the valence electrons.[20] These basis sets were designed in such a way that they have the unique property of *achieving* systematic convergence. By virtue of using a series of correlation consistent polarized valence n-zeta basis sets, the electronic structure calculations can become as accurate as those using the complete basis set.

Split valence basis sets can be improved not only by adding the orbitals with different shapes such as the polarization functions, but also by augmenting the orbitals with slow decaying diffuse functions. Polarization functions allow for the space occupied by the valence electrons to be described properly in calculating molecular orbitals by affording the flexibility in the basis set.[21] Diffuse functions tend to decay more slowly as become distant from the nucleus than the normal basis functions, which has the effect of expanding the electron density. Therefore, the use of diffuse functions is necessary for

dealing with anions, lone pairs, and the excited states that accompany the expansion of the volume occupied by the valence electrons.[22]

1.1.6. Density functional theory

Density functional theory (DFT) has contributed to quantum chemistry in the sense that it reduces the computational burden for electronic structure considerably compared the other post-HF methods using the same basis set. In essence, DFT assumes that sufficient information for the electronic energy and the wavefunction of a system can be obtained from the electron density. Hohenberg and Kohn formulated the fundamental theorem that the unique electron density ($\rho(\vec{r})$) of the ground-state one can determines all the physical properties of a system.[23] According to the first theorem, the ground-state energy of a system comprising many particles is a unique functional of $\rho(\vec{r})$ because both the external potential $V_{ext}(\vec{r})$ and Hamiltonian operator ($\widehat{H}$) are a functional of $\rho(\vec{r})$. In other words, $\rho(\vec{r})$ determines $V_{ext}(\vec{r})$ and all the physical properties of the ground state including kinetic energy $T[\rho]$, potential energy $V[\rho]$, and the total energy $E[\rho]$. We can write the total energy as follows:

$$E[\rho] = E_{Ne}[\rho] + T[\rho] + E_{ee}[\rho] = \int \rho(\vec{r})\, V_{Ne}(\vec{r})\, d\vec{r} + F_{HK}[\rho]$$

$$T[\rho] + E_{ee}[\rho] = F_{HK}[\rho] \qquad (14).$$

This functional $F_{HK}[\rho]$ is the key element of density functional theory. If $F_{HK}[\rho]$ were known, one can solve the Schrödinger equation exactly. The second Hohenberg-Kohn theorem states that the functional that evaluates the ground state energy of the system ($F_{HK}[\rho]$) can generate the lowest energy if and only if the input electron density is that of the true ground state:

$$E_0 \leq E[\tilde{\rho}] = E_{Ne}[\tilde{\rho}] + T[\tilde{\rho}] + E_{ee}[\tilde{\rho}] \qquad (15).$$

In other words, this means that for any trial density $\tilde{\rho}(\vec{r})$, which satisfy the necessary boundary conditions such as $\tilde{\rho}(\vec{r}) \geq 0$, $\int \tilde{\rho}(\vec{r})\, d\vec{r} = N$, and which is associated with some external potential $\tilde{V}_{ext}(\vec{r})$, the energy obtained from the functional of Eq. (14) represents an upper bound to the true ground state energy E_0.

Therefore, one needs to derive the exact form of the functional $F_{HK}[\rho]$. Kohn and Sham suggested that calculating the kinetic energy of non-interacting reference system with the same density with the assumption that the electron density corresponds to a wavefunction given by a single Slater determinant.[24] They described the exchange-correlation potential as follows:

$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho])$$

$$E[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \qquad (16).$$

Here, only the $E_{XC}[\rho]$ term does not have an explicit mathematical form. The exchange-correlation potential (V_{XC}) is defined as the functional derivative of E_{XC} with respect to ρ, $V_{XC}[\rho] = {\partial E_{XC}[\rho]}\big/{\partial \rho}$. It is noteworthy that the Schrödinger equation can be solved easily in the presence of the exact forms of E_{XC} and V_{XC}. Accordingly, much effort has been devoted to making good approximations for the exchange-correlation functional. The first successful exchange-correlation functional is the local density approximation (LDA) that is based on the exchange-correlation energy per particle of a uniform electron gas with density $\rho(\vec{r})$.[25] However, the applicability of LDA is limited because the electron density can vary rapidly over a small region of space. A reasonable way to improve the LDA method was to impose the functional depend on both the density and the gradient of the density. This leads to the development of the gradient-corrected generalized gradient approximation (GGA) functionals.[26] As a consequence of the

continuous efforts to further improve the accuracy of the exchange-correlation energy, various hybrid functionals have been developed including Becke-Lee-Yang-Parr's three parameter method (B3LYP),[27] the PBE0 functional,[28] and Meta hybrid GGA method such as M06-2X developed by Truhlar and Zhao.[29]

1.2. Statistical Mechanics

In a molecular dynamics (MD) or Monte Carlo simulations, one wants to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change. The connection between microscopic simulations and macroscopic properties is made via *statistical mechanics* which provides mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system. An ensemble is a collection of all possible systems that have different microscopic states but have an identical macroscopic or thermodynamic state.

1.2.1. Molecular dynamics (MD)

Molecular dynamics (MD) is a technique for simulation of complex systems including biological molecules. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids, ions are transported through membranes, enzymes catalysing chemical reactions. MD methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes.[30,31]

There are two types of MD simulation methods, which can be distinguished according to the model chosen to represent a physical system. In the 'classical' mechanics approach to MD simulations molecules are treated as classical molecules, assigning partial charges

and bond lengths, bond angles based on empirical potential. The laws of classical mechanics define the dynamics of the system. The 'quantum' MD simulations, which started in the 1980s with the pioneering work of Car and Parinello, explicitly take into account the quantum nature of the chemical bond.[32] Other methods that are currently in use are for instance Car-Parrinello, Hellmann-Feynman, first principles and quantum chemical molecular dynamics. The electron density function for the valence electrons that determine bonding in the system is computed using quantum equations. At present, the classical MD methods are of practical use for simulations of biomolecular systems comprising many thousands of atoms over time scales of nanoseconds. Henceforth, classical MD will simply be referred to as MD.

In a molecular dynamics simulation, one often wishes to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change. The connection between microscopic simulations and macroscopic properties is made via statistical mechanics which provides rigorous expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the *N*-body system; molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas.

MD methods are based on Newton's second law or the equation of motion. If the particles of interest are atoms, and if there are a total of *N* interacting atoms in the system, the force acting on the *i*th atom at a given time can be obtained from the interatomic potential $U(r_1, \dots, r_N)$ that, in general, is a function of the positions of all the atoms:

$$F_i = -\nabla_{r_i} U(r_1, \dots, r_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right) \qquad (17)$$

Atomic force field models and classical MD are based on empirical potentials with a specific functional form, representing the physics and chemistry of the systems of interest. A simulation package force-field will specify the precise form of Eq. (5), with the adjustable parameters. Quantum mechanical calculations may give a guide to the best molecular force-field. Further, comparison of simulation results with experimental data and vibration frequencies is essential in force-field development and refinement. A separate family of force-fields, such as AMBER,[33,34] CHARMM[35] and OPLS[36] are geared more to larger molecules (proteins, polymers) in condensed phases; their functional form is simpler, closer to that of Eq. (18),

$$U(r_1, \ldots, r_N) = \sum_{bonds} \frac{a_i}{2}(l_i - l_{i0})^2 + \sum_{angles} \frac{b_i}{2}(\theta_i - \theta_{i0})^2 + \sum_{torsions} \frac{c_i}{2}[1 + \cos(n\omega_i - \gamma_i)] + \sum_{atompairs} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}{}_{i}\right] + \sum_{atompairs} k\frac{q_i q_j}{r_{ij}} \qquad (18)$$

In the first three terms summation indices include all the bonds, angles and torsion angles defined by the covalent structure of the system, whereas in the last two terms summation indices include all the pairs of atoms.

Protein-ligand interactions, molecular recognition and signalling commonly occur as a result of intermolecular interactions between receptor molecules and inhibitors or peptides. Binding of a ligand to a specific binding site of the receptor is known as the docking problem. Finding the geometry and strength of ligand binding (ligand affinity) is the main goal of computational studies of docking.[37] This goal may be achieved using global optimization and free energy methods coupled with MD or MC sampling. Docking techniques are based either on the simple key-lock model, in which binding partners are regarded commonly as rigid, or on the more computationally demanding

induced fit model, which takes into account the flexibility of ligands and ligand-induced conformational changes in host proteins.[38]

1.2.2. Free Energy Perturbation (FEP)

The free energy of a system is the most important thermodynamic quantity, and is usually expressed as the Helmholtz's (A) or Gibb's (G) free energy. The free energy difference between the two states A and B can be given by the ensemble average of the perturbation term with respect to the reference system. In the early years of atomistic simulations for chemical and biological systems, its applicability was limited to small solutes dissolved in water. The perturbation approach is now used extensively even in the investigation of protein–ligand interactions and *in silico* protein engineering.[39,40] The statistical mechanical approach to calculate free energy changes was introduced by Kirkwood and Zwanzig.[41] More specifically, the free energy difference between two states *A* and *B* can formally be obtained from the Zwanzig's formula:[42]

$$\Delta G = G_B - G_A = -\beta^{-1} \ln\langle \exp(-\beta \Delta V) \rangle_A \qquad (19).$$

Here, $\beta = 1/kT$ and $\langle \cdots \rangle_A$ denotes the ensemble average of $\Delta V = V_B - V_A$ calculated from molecular dynamics (MD) or Monte Carlo (MC) simulations using the potential (V_A). In Eq. (19), it is assumed that the configurational sampling is carried out under constant temperature and pressure conditions, which is known as isothermal-isobaric (NPT) ensemble while the simulations in canonical (NVT) ensemble yield the corresponding Helmholtz free energy. To obtain a precise estimate there is a requirement for sufficient sampling of the equilibrium distribution of the thermodynamic states during the simulation to obtain a precise estimate (convergence problem), a multistep procedure is usually adopted in calculating the free energy change using Eq. (19). A thermodynamic path between the states A and B is defined by introducing a set of intermediate potential

energy functions that are constructed as linear combinations of those of the initial (*A*) and final (*B*) states.

$$V_m = (1 - \lambda_m)V_A + \lambda_m V_B \qquad (20)$$

State A can be transformed to B with the change of the coupling parameter λ_m from 0 to 1. Molecular dynamics (MD) simulations should therefore be performed at each intermediate step characterized by the discretized λ_m values $(\mathrm{m} = 1, 2, \cdots, \mathrm{n})$. This segmentation of the coupling parameters is allowed because the free energy difference depends only on the initial and final states in the path independent manner.[43] The free energy change can thus be computed along any reversible path connecting the two states. The total change of Gibbs free energy can be obtained by summing over the intermediate states along the λ variables.

$$\Delta G(\lambda_m \rightarrow \lambda_{m+1}) = -k_B T \, ln\langle \exp(-\beta \Delta V_m) \rangle \qquad (21)$$

This approach is generally referred to as the free energy perturbation (FEP) method.[44] Within the FEP framework, the total Gibbs free energy change is obtained using the following relation:

$$\Delta G = G_B - G_A = -\beta^{-1} \sum_{m=1}^{n-1} ln\langle \exp[(-\beta V_{m+1} - V_m)] \rangle_m \qquad (22)$$

1.2.3. Alchemical Transformation

The application of FEP to alchemical transformations requires the construction of an efficient thermodynamic path through intermediate states. Because free energy is a state function, it is often better to select the most efficient or convenient path.[43] This approach would be useful for describing the protein–ligand binding, mutational effects on protein structure and function, and solvent-solute interactions. The estimation of solvation free energies by FEP has been successful because the solute-solvent interactions can be described by adding the perturbation term in the Hamiltonian. In practice, the solvation free energy of a solute can be expressed as the difference between the free energies of

the solute in solution and as an ideal gas. We use Helmholtz free energy (ΔA) because the systems are simulated under approximately constant volume conditions. The solvent molecules undergo a substantial reorganization in the presence of solutes to an extent that affects the conformational equilibria in flexible solute molecules. ΔA_{hyd} refers to the change in free energy for a solute molecule to leave the gas phase and to enter the liquid water. In this case, a thermodynamic cycle may be useful to calculate the ΔA_{hyd} values at varying temperatures as illustrated in Figure 1.1. Because A is thermodynamically a state function, the sum of all free energy changes in the cycle must be zero. ΔA_{hyd} can therefore be expressed as follows:

$$\Delta A_{hyd} = \Delta A_{gas} - \Delta A_{aq} \qquad (23).$$

ΔA_{aq} and ΔA_{gas} correspond to the difference between A_q and A_0 values of the solute in aqueous solution and that in the gas phase, respectively. The ΔA_{hyd} values of all the solute molecules can be obtained through the FEP calculations of ΔA_{aq} and ΔA_{gas} based on the independent MD simulations in water and in the gas phase, respectively.[45,46]

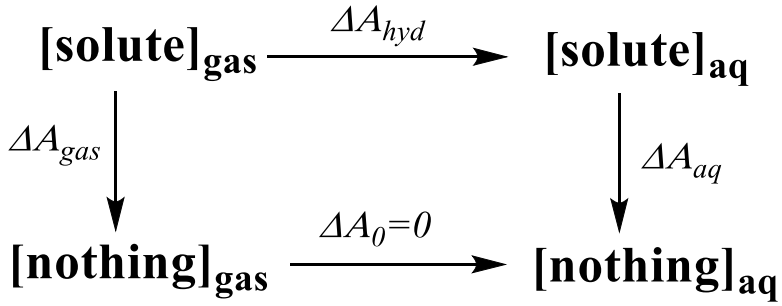


Figure 1.1. Thermodynamic cycle for calculating the hydration free energies of solute molecules. Note that $\Delta A_0 = 0$ because solute is not involved in the process.

To determine the ΔA_{aq} and ΔA_{gas} values, the solute molecule with atomic charges (state A) was changed gradually to a dummy molecule in which all the atoms were uncharged (state B) by perturbing the Hamiltonian of the system. This transformation could be

accomplished by scaling the electrostatic interaction potential with the change of a state variable (λ_{elec}) that maps the initial and the final states onto 0 and 1, respectively.

1.2.4. FEP approach to computer-aided molecular design

The FEP approach can serve as a useful computer-aided molecular design tool for the synthesis of derivatives from starting compounds to enhance the biochemical potency with respect to a target protein, or other macromolecule. We can use Gibbs free energy (ΔG) as biochemical reactions are usually carried out under constant pressure conditions.[47] MD-FEP calculations can be performed using a thermodynamic cycle approach as depicted in Figure 1.2.[48]

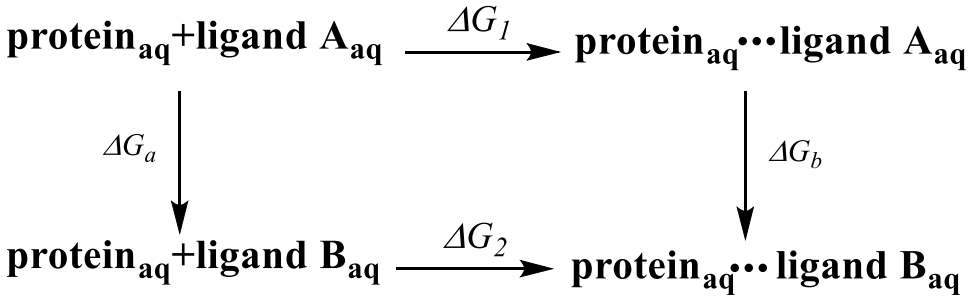


Figure 1.2. Thermodynamic cycle used in calculating the free energy change in aqueous solution (ΔG_a), that in the binding site of a target protein (ΔG_b), and the difference in binding free energies ($\Delta\Delta G_{bind}$) between the two inhibitors (ligand A and ligand B).

The change in binding free energy ($\Delta\Delta G_{bind}$) between the two inhibitors can be computed through the nonphysical paths connecting the desired initial and final states. More specifically, this approach enables the calculation of $\Delta\Delta G_{bind}$ between two structurally similar inhibitors by computationally simulating the transformation of one into the other. Because the free energy is a state function and, therefore, the sum of all energies in the thermodynamic cycle in Figure 1.2 is zero, $\Delta\Delta G_{bind}$ can be expressed as follows:

$$\Delta\Delta G_{bind} = \Delta G_2 - \Delta G_1 = \Delta G_b - \Delta G_a \qquad (24).$$

Here, ΔG_1 and ΔG_2 refer to the binding free energies of inhibitors ligand A and ligand B, respectively, and ΔG_a and ΔG_b to the free energies associated with the structural transformation of ligand A into ligand B in aqueous solution and in the binding site of the target protein. [49,50]

Hence, to determine the $\Delta\Delta G_{bind}$ value between ligand A and ligand B, the free energy changes in transforming ligand A into ligand B should be calculated both in aqueous solution and in the binding site of the target protein. The free energy change for converting ligand A into ligand B was computed by perturbing the Hamiltonian of ligand A (initial state) into that of ligand B (final state). This transformation can be accomplished through the parametrization of the terms comprising the interaction potentials of the system with the change of a state variable (λ) that mapped onto reactant ($\lambda = 0$) and product ($\lambda = 1$) states. The total free energy changes for the mutation from the initial to the final state is then computed by summing the incremental free energy changes contributions coming from simulations at different values of λ as it changes from 0 to 1.[51]

1.2.5. Thermodynamic Integration

Thermodynamic integration (TI) is a conceptually simple and widely used method to calculate free energy differences from MC or MD simulations. [52-57] In TI, free energy differences are calculated by defining thermodynamic paths between the states and integrating over the ensemble-averaged enthalpy changes along the path. Such paths can either be real chemical processes or alchemical processes. Due to the free enthalpy being a state function, i.e., it is path independent, the dependence on the coupling parameter λ can be chosen freely. By varying λ, the system can be transformed from state A ($\lambda_A = 0$) to state B ($\lambda_B = 1$). The free enthalpy difference ΔG_{BA} is then given by

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \langle \frac{\partial H(\lambda)}{\partial \lambda} \rangle_\lambda \, \partial\lambda \qquad (25)$$

where the ensemble average for the system $H(\lambda)$ is denoted by $\langle \cdots \rangle_\lambda$. In practice, ΔG_{BA} is obtained by performing a finite number, N_λ of simulations at discrete λ-values ranging from 0 to 1 and subsequent numerical integration using an interpolation formula. Note, the simulation time must be sufficiently long for the system to be in equilibrium at every λ-value and for $\langle \frac{\partial H(\lambda)}{\partial \lambda} \rangle_\lambda$ to have converged. In principle, the path along which the system is perturbed should be chosen such that sampling problems are minimized. This choice, however, is not always trivial and may strongly influence the convergence of $\langle \frac{\partial H(\lambda)}{\partial \lambda} \rangle_\lambda$at particular λ-values. Further improvement in trajectory sampling may be achieved with the metadynamics method,[58,59] that enables filling of the already visited regions of the phase space with Gaussians in potential energy, as well as with the adaptive bias force method[60,61] that allows the optimization of biasing force to obtain the uniform sampling along the reaction coordinate.

1.3. Potential Energy Surface (PES)

1.3.1 Reaction Coordinates

One purpose of computational chemistry lies in the explanation of experimental results based on the theoretical potential energy surface (PES).[62] The potential energy is a function that gives the energy of a molecule as a function of the nuclear geometry that varies during chemical reactions. A typical PES has the same dimensionality as the number of vibrational degrees of freedom of a molecule (3*N*–6) comprising *N* atoms. Since the geometrical changes are smooth in some chemical reactions, the PES concept creates a smooth energy 'landscape'.[62] A variety of computational frameworks can be applied to calculate the potential energies including quantum mechanics (QM), molecular dynamics (MD), and free energy perturbation (FEP). QM methods would be

useful for chemical reactions involving bond-breaking and bond-creating processes. MD simulations are based on the parametrized potentials for macromolecular systems such as protein stability and transmembrane simulations. FEP, ab initio MD and classical MD are based on the statistical mechanics to calculate the free energy changes in protein-ligand interactions with sampling methods.

The characteristic feature that discriminates the PES approach from the others is that it allows the identification of the stationary points on which the gradients with respect to all internal coordinates are zero. The energy values corresponding to the equilibrium and rate constants can be calculated by locating the stationary-state structures including the energy minima and transition states encountered along the reaction coordinates. The energy minima correspond to stable or quasi-stable species such as reactants, products or intermediates whereas saddle points represent the structure of energy maximum on the minimum energy path.[62] Mathematically, all saddle points should have a single imaginary vibrational frequency due to a negative force constant with respect to the vibrational motion along the reaction coordinate. PES is thus a multidimensional function that represents the variation of potential energy with the change of the nuclear geometry caused by chemical reactions. The minimum energy path, which is also referred to as the reaction coordinate, is the one-dimensional projection of PES to connect the reactants and products. To describe a chemical reaction, one needs to locate at least three stationary points: two energy minima for reactants and products and one saddle point for the transition state.

1.3.2. Transition State Theory

The constant *k* is useful for understanding the kinetics of chemical reactions. The simple relation between the rate constant of a chemical reaction and temperature is given by the empirical Arrhenius equation:[63]

$$k = Ae^{\frac{-\Delta G^*}{k_B T}} \quad (26).$$

Here, k_B and ΔG^* denote the Boltzmann constant and the activation energy, respectively, while A is the pre-exponential factor. The attempt to define the pre-exponential factor A led to the establishment of transition state theory (TST) to describe the chemical reactions in which the reactants with enough energy cross a single energy barrier to arrive at a new state. TST also stipulates that the transition state cannot turn backwards to reform the reactants. Under these assumptions, the rate constant can be rewritten by the following Eyring equation:[64]

$$k = \kappa \frac{k_B T}{h} e^{\frac{-\Delta G^*}{k_B T}} \qquad (27).$$

Here, h and κ are Plank's constant and the transmission coefficient, respectively. Many researchers have investigated how to determine κ from basic physical principals as opposed to empirical observations. The pre-exponential has a strong effect, which is largely dictated by the nature of the medium that the reaction is occurring in. The local environment can do this by affecting the transmission coefficient κ. One question is how the solvent affects the transmission coefficient. In 1940, Hendrik Kramers understood well the mechanism of climbing the reaction barrier in solution as governed by the thermal (Brownian) motions of the solvent.

1.3.3. Rate Reaction Theory (Kramer's Theory)

To determine the transmission coefficient of chemical reactions, Kramer[65] used stochastic dynamics employing the Langevin equation.[66]

$$m\ddot{x} = -\nabla U(x) - \gamma\dot{x} + \sqrt{2\gamma k_B T f(t)} \quad (28)$$

Here, *m* and x are mass of the reactant the reaction coordinate on a potential energy surface defined by *U(x)*, respectively, while γ is the solvent viscosity. $f(t)$ represents a delta-correlated stationary Gaussian process with zero-mean, which is referred to as Brownian motion. In the Langevin equation, the presence of buffer gas or solvent can be reflected by viscosity. As indicated in Eq. (28), the viscosity acting on the reactant is compensated by the Gaussian form of random force $f(t)$. After manipulation of Langevin equation, Kramers was able to express the transmission coefficient κ as a function of solvent viscosity, which is illustrated in Figure 1.3.

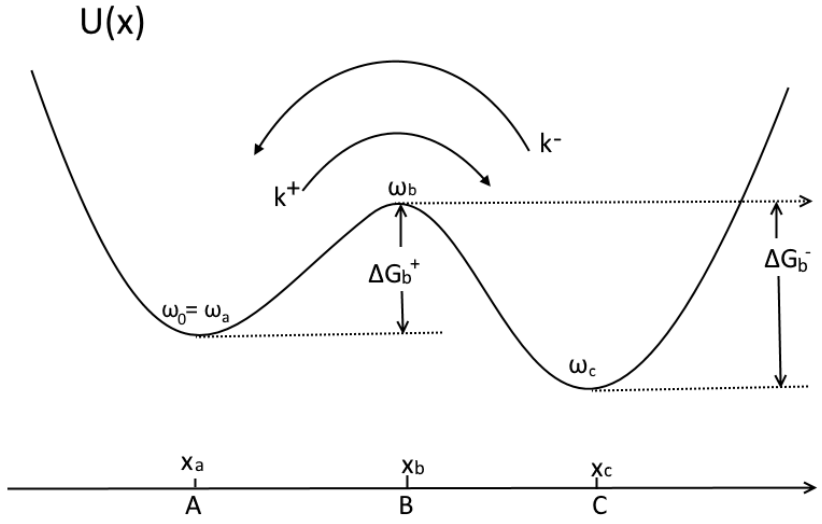


Figure 1.3. Potential *U(x)* with two states A and C. Escape occurs via forward (k^+) and backward (k^-) rate constants. $\Delta G_b^{\pm}$ indicates the activation energies. x_a and x_c denote the reactant and product states, the maximum of *U(x)* at x_b separating these states corresponds to the transition state, respectively. $\omega_a$and ω_c are the angular frequencies inside the metastable minima and $\omega_b$denotes the angular frequency at the transition state.

$$k_{A\to C} = \frac{\lambda_+}{\omega_b}\left[\frac{\omega_0}{2\pi}\exp(-\beta\Delta G^*)\right] = \frac{\left[\frac{\gamma^2}{4}+\omega_b^2\right]^{1/2}-\frac{\gamma}{2}}{\omega_b}\frac{\omega_0}{2\pi}\exp(-\beta\Delta G^*) \quad (29)$$

This equation describes the spatial-diffusion-controlled rate of escape under moderate-to-strong friction.[65] The $\left[\frac{\omega_0}{2\pi}\exp(-\beta E_b)\right]$ term in Eq. (29) represents the simple transition state of theory k_{TST} as shown in Eq. (26). In the presence of a finite-damping rate γ, the diffusive transmission factor $\kappa \equiv \frac{\lambda_+}{\omega_b}$ is always less than one. When the solvent viscosity is very high, the transmission coefficient can be simplified to that of the rate constant in the overdamped regime as follows.

$$k^+ = \frac{\omega_0\omega_b}{2\pi\gamma}\, e^{-\beta\Delta G^*} \qquad (\gamma \gg \omega_B) \qquad (30)$$

Here, ω_0 and ω_b are related to the curvature of the potential energy curve in the vicinity of the reactants and transition state, respectively.

1.4. Hydrophobicity

1.4.1. Overview of Scaled Particle Theory

Hydrophobic interactions are observed in a variety of chemical and biological phenomena such as protein folding and the ordering of amphiphilic chain molecules. In this regard, the hydrations of hydrophobic solute molecules have been widely investigated with theoretical and computational methods. The structure of water clusters around hydrophobic solutes is related to the entropic cost for achieving the ordering of water molecules which can be evaluated by the expansion of pair correlation function.[67-70] The occurrence of a molecular cavity in a liquid stems from the density fluctuations at equilibrium. The creation of a cavity in liquid with a constant number of molecules means the selection of the molecular configurations including the desired cavity from the initial

liquid configurations.[69] The excess chemical potential is therefore central to resolving the aqueous solubility of a solute. In case of the impenetrable hard sphere (HS) solutes, the solute excess chemical potential in water can be defined:[71]

$$\mu_A^{ex}(aq) = -kTln[p_0(R)] \quad (31).$$

The insertion probability, $p_0(R)$, is the probability that a solute-sized stencil randomly placed in water is devoid of water oxygen centers and *R* is the radius of the solute. An alternative relationship between the chemical potential and to the structure of water molecules in the vicinity of the solute can be expressed in the following form:[72-74]

$$\mu_A^{ex}(aq) = kT \int_0^R \rho_W G(r) 4\pi r^2 dr \quad (32)$$

Here, $\rho_W G(r)$ is the density of water in contact with a hard-sphere solute with radius *r*. The product $kT\rho_W G(r))$ has the dimension of force/area, and is the pressure due to the collisions of water molecules with the hard surface of the solute. In the limiting case of a macroscopically large cavity, the contact correlation function can be represented as an asymptotic expansion of *1/R*:

$$G(R) \sim \sum_{j \geq 0} \frac{G_j}{R^j} \quad (33)$$

Retaining the contributions up to j = 2 yields an expression equivalent to that derived from classical thermodynamics for the force acting on the cavity surface:[70,73,74]

$$kT\rho_W G(r) \sim p + \frac{2\gamma_\infty}{R} - \frac{4\gamma_\infty}{R^2} \quad (34)$$

Here, p and γ_∞ are the liquid saturation pressure p_{sat} and the surface tension of a flat interface denoted by γ while δ (Tolman length) describes the initial curvature correction to the surface tension.[75] G_j can be expressed as Eq. (33) and expanded as Eq. (34). After

suitable algebraic manipulations one can derive the following equation to be useful for estimating molecular hydration entropy of solute molecules:

$$\frac{\Delta\mu_A^{ex}}{k_BT} = -\ln(1-\varphi) + 3\left(\frac{\varphi}{1-\varphi}\right)R + \left[\left(\frac{3\varphi}{1-\varphi}\right) + 4.5\left(\frac{\varphi}{1-\varphi}\right)\right]R^2 + \frac{\varphi P}{\rho k_BT}R^3 \qquad (35)$$

Here, φ and ρ represent the volume fraction of water molecules with respect to the total volume and the number density of water, respectively. *R* denotes the ratio of the radius of solute to that of water molecule and *P* stands for the pressure of water.

1.5. Homology Modeling

Homology modeling is used to predict a protein 3D structure using an experimentally determined structure of a homologous protein as a template. The evidence is that often during protein evolution, the overall protein fold is more conserved rather than its sequence.[76] Proteins which share similar sequences often have very similar structures and distantly related proteins by sequence can fold into similar structures.[77]

During homology modeling, the coordinates of the model are generated based on alignments between the model target and template amino acid sequences, which define the correspondence between residues in both proteins. Similar sequences can be searched for using BLAST[78] or fold recognition[79] methods and aligned with the known structures in the protein data bank (PDB).[80] The PDB which is a large database that contains only experimentally resolved structures, principally but not exclusively of proteins. BLAST enables structure based comparison of sequence.

Given a template structure and an alignment with a target for modeling, the information contained therein can be used to generate a three-dimensional structural model of the target, represented as a set of Cartesian coordinates for each atom in the protein. In practice, this step is made up of backbone generation, loop modeling and side chain

modeling. The backbone generation from the aligned regions can be done using modelling tools such as Modeller or CASP.[81,82] In the majority of cases, the alignment between model and template sequence contains gaps. There are two main approaches to loop modeling. First, knowledge based loop modeling searches the PDB for known loops with endpoints that match the residues between which the loop has to be inserted, and simply copies the loop conformation.[82] All major molecular modeling programs and servers employ this approach. Second, energy based loop modeling employs *ab initio* loop fold prediction to generate an energy function that is used to judge the quality of a loop.[83] This function is minimized, using Monte Carlo or MD techniques. The analysis of side chain conformations is partially knowledge based libraries of rotamers extracted from X-ray crystal structures.[84] To build a position-specific rotamer library, one can use high-resolution protein structures and combine information of fragment sequences with a given amino acid at the center of the sequence. The prediction accuracy can be quite high for residues in the hydrophobic core, however, is not accurate enough to get complicated electrostatic interactions on the surface of protein.[82] Comparative protein modelling methods have been automated in recent years, and several Internet servers offer protein modelling services which are easy to use. MODELLER,[81] SWISS-MODEL,[85] ROBETTA[86] and Yasara,[87] are widely used web servers based methods for protein structure homology modeling.

Chapter 1 References

1. a) Levine, I. N. *Quantum Chemistry*; 7th edition, Pearson Higher Ed. USA, **2013**; (b) Atkins, P.; Friedman, R. *Molecular Quantum Mechanics*; Oxford University Press, **2005**; (c) McQuarrie, D. A. *Statistical Mechanics*; University Science Books, **2000.**
2. Sakurai, J. J. *Modern Quantum Mechanics*; Revised edition, Addison-Wesley Publishing Company: Boston, **1993**.
3. Pyykko, P. Relativistic effects in structural chemistry. *Chem. Rev*. **88**, 563-594 (1988).
4. Born, M., Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **84**, 457-484 (1927).
5. Slater, J.; Verma, H. C. The Theory of Complex Spectra. *Phys. Rev*. **34**, 1293-1322 (1929).
6. Fock, V. A. Selfconsistent field "mit Austausch für Natrium. *Z. Phys.* **62**, 795-805 (1930).
7. Fock, V. A. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* **61**, 126-148 (1930).
8. Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Phys.* **31**, 765-783 (1925).
9. Slater, J. C. A Simplification of the Hartree-Fock Method. *Phys. Rev*. **81**, 385-390 (1951).
10. Slater, J. C. Magnetic Effects and the Hartree-Fock Equation. *Phys. Rev*. **82**, 538-541 (1951).
11. Møller, C., Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev*. **46**, 618-622 (1934).
12. Krishnan, R., Schlegel, H. B., Pople, J. A. Derivative studies in configuration–interaction theory. *J. Chem. Phys.* **72**, 4654-4655 (1980).
13. Pople, J. A., Head-Gordon, M., Krishnan, R. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **87**, 5968-5975 (1987).
14. Head-Gordon, M., Rico, R. J., Oumi, M., Lee, T. J. A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chem. Phys. Lett.* **219**, 21-29 (1994).
15. Head-Gordon, M., Maurice, D., Oumi, M. A perturbative correction to restricted open shell configuration interaction with single substitutions for excited states of radicals. *Chem. Phys. Lett.* **246**, 114-121 (1995).
16. Hückel, E. Quantentheoretische beiträge zum benzolproblem. *Z. Physik*, **70**, 204-286 (1931).
17. Slater, J. C. Atomic shielding constants. *Phys. Rev.* **36**, 57-64 (1930).
18. Ditchfield, R., Hehre, W. J., Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724-728 (1971).
19. Hehre, W. J., Ditchfield, R., Pople, J. A. Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **56**, 2257-2261 (1972).
20. Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007-1023 (1989).
21. Frisch, M. J., Pople, J. A., Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys*. **80**, 3265-3269 (1984).
22. Clark, T., Chandrasekhar, J., Spitznagel, G. W., Schleyer, P. v. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *J. Comp. Chem.* **4**, 294-301 (1983).
23. Hohenberg, P., Kohn, W. Inhomogeneous electron gas. *Phys. Rev*. **136**, B864-B871 (1964).
24. Kohn, W., Sham, L. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev*. **140**, A1133 (1965).
25. Perdew, J. P., Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems *Phys. Rev. B* **23**, 5048-5079 (1981).
26. Perdew, J. P., Burke, K., Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).
27. Evans, D. A., Siska, S. J., Cee, V. J. Resurrecting the Cornforth Model for Carbonyl Addition: Studies on the Origin of 1,2-Asymmetric Induction in Enolate Additions to Heteroatom-Substituted Aldehydes. *Angew. Chem. Int. Ed.* **42**, 1761-1765 (2003).
28. Perdew, J. P., Ernzerhof, M., Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982-9985 (1996).
29. Zhao, Y., Truhlar, D. Density functionals with broad applicability in chemistry. *Theor. Chem. Acc.* **120**, 157-167 (2008).

30. M. J. Field. *A Practical Introduction to the Simulation of Molecular Systems*. Cambridge University Press, Cambridge, **1999**.
31. A. R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2nd edition, **2001**.
32. Car, R., Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, **55**, 2471 (1985).
33. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, *106*, 765-784 (1984).
34. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M. & Kollman, P. A. A second-generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197 (1995).
35. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187-217 (1983).
36. Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **118**, 11225-11236 (1996).
37. Rosenfeld, R., Vajda, S., DeLisi, C. Flexible docking and design. *Ann. Rev. Biophys. Biomol. Struct.*, **24**, 677-700 (1995).
38. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.*, **49**, 534-553 (2006).
39. Seeliger, D., De Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophysical journal*, **98**, 2309-2316 (2010).
40. Damborsky, J., Brezovsky, J. Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.*, **13**, 26-34 (2009).
41. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300-313 (1935).
42. Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* **22**, 1420-1426 (1954).
43. Chipot, C., Pohorille, A. *Free Energy Calculations*, Springer: Berlin, **2007**.
44. Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94**, 8021-8024 (1990).
45. Pohorille, A., Jarzynski, A.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **114**, 10235-10253 (2010).
46. Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., Sherman, W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **6**, 1509-1519 (2010).
47. Dill, K., Bromberg, S. *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. Garland Science, **2010**.
48. Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **21**, 150-160 (2011).
49. Woo, H. J., Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA.* **102**, 6825-6830 (2005).
50. Dixit, S. B., Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **105**, 9795-9799 (2001).
51. Durrant, J. D., McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology* **9**, 71-79 (2011).
52. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300-313 (1935).
53. Straatsma, T. P., & McCammon, J. A. Multiconfiguration thermodynamic integration. *J Chem. Phys*. **95**, 1175-1188 (1991).
54. Tironi, I. G., Van Gunsteren, W. F. A molecular dynamics simulation study of chloroform. *Mol. Phys.* **83**, 381-403 (1994).
55. Mitchell, M. J., McCammon, J. A. Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value. *J. Comput. Chem.* **12**, 271-275 (1991).
56. Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596-3607 (1999).
57. Frenkel, D., Smit, B. Understanding molecular simulation: from algorithms to applications. *Academic Press*, California, *1*, 1-638. **2002**.

58. Ensing, B., De Vivo, M., Liu, Z., Moore, P., Klein, M. L. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* **39**, 73-81 (2006).
59. Laio, A., Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, **71**, 126601 (2008).
60. Darve, E., Rodríguez-Gómez, D., Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **128**, 144120 (2008).
61. Chipot, C., Pohorille, A. *Free energy calculations*. Springer-Verlag Berlin Heidelberg, **2007**.
62. Truhlar, D. *Potential Energy Surfaces and Dynamics Calculations: For Chemical Reactions and Molecular Energy Transfer*. Springer Science & Business Media, **2013**.
63. Arrhenius, S.A. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. *Z. Phys. Chem*. **4**, 96 (1889).
64. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys*. **3**, 107-115 (1935).
65. Hänggi, P., Talkner, P., Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys*. **62**, 251-341 (1990).
66. Langevin, P. Sur la théorie du mouvement brownien. *C. R. Acad. Sci*. **146**, 530-533 (1908).
67. Henderson, J. R. Solvation of a solvophobic sphere. *J. chem. Phys.* **116**, 5039-5045 (2002).
68. Hummer, G., Garde, S., Garcia, A. E., Pohorille, A., Pratt, L. R. An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci. USA* **93**, 8951–8955 (1996).
69. Pohorille, A., & Pratt, L. R. Cavities in molecular liquids and the theory of hydrophobic solubilities. *J. Am. Chem. Soc.* **112**, 5066-5074 (1990).
70. Pierotti, R. A. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.* **76**, 717–726 (1976).
71. Widom, B. Potential-distribution theory and the statistical mechanics of fluids. *J. Phys. Chem.* **86**, 869-872 (1982).
72. Stillinger, F. H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Sol. Chem.* **2**, 141-158 (1973).
73. Reiss, H. *Statistical Mechanics and Statistical Methods in Theory and Application*, edited by U. Landman Plenum, New York, pp. 99–138, **1977**.
74. Reiss, H. Scaled particle methods in the statistical thermodynamics of fluids. *Adv.Chem. Phys.* **9**, 1-84 (1965).
75. Tolman, R. C. The effect of droplet size on surface tension. *J. chem. Phys.* **17**, 333-337 (1949).
76. Chothia, C., Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, **5**, 823-826 (1986).
77. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., Šali, A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* **29**, 291-325 (2000).
78. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.*, *215*(3), 403-410 (1990).
79. Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics*, **21**, 951-960 (2005).
80. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Bourne, P. E. The protein data bank. *Nucleic Acids Res.*, *28*(1), 235-242 (2000).
81. Fiser, A., & Šali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461-491 (2003).
82. Moult, J., Pedersen, J. T., Judson, R., & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins*, **23(3)**, (1995).
83. Chivian, D., Baker, D. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res*, **34**, e112-e112 (2006).
84. Dunbrack, R. L., Karplus, M. Backbone-dependent rotamer library for proteins application to side-chain prediction. *J. Mol. Biol.*, **230**, 543-574 (1993).
85. Schwede, T., Kopp, J., Guex, N., Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381-3385 (2003).
86. Kim, D. E., Chivian, D., Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, **32**(suppl 2), W526-W531 (2004).
87. Krieger, E., Koraimann, G., Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*, **47**, 393-402 (2002).

Chapter 2

Investigations on Recyclisation and Hydrolysis in Avibactam Mediated Serine β-Lactamase Inhibition

2.1. Introduction

The penicillins were first isolated and applied ~70 years ago, yet the family of β-lactam containing antibiotics to which they were the progenitors is probably still the most important of all pharmaceuticals.[1] β-Lactam antibiotics (BLAs) target transpeptidases/carboxypetidases or penicillin binding proteins that are crucially involved in bacterial cell wall biosynthesis by a mechanism that involves (at least normally) essentially irreversible acylation (irreversible in terms of the time it takes for a cycle of bacterial cell division to occur) of a catalytically important and nucleophilic serine residue.[2-5] The penicillins were followed by successive generations of BLAs, including the cephalosporins, monobactams, and carbapenems, all of which are characterized by the presence of a β-lactam ring. The development of new generations of β-lactams was in part driven by a desire to combat resistance due to β-lactamase enzymes, which catalyze the hydrolysis of BLAs to give β-amino acids which are inactive or much less active as transpeptidase inhibitors.[6] An alternative strategy is to combat β-lactamase mediated resistance by combination of the β-lactam antibiotics with a β-lactamase inhibitor. Three β-lactamase inhibitors have been developed for clinical use in combination with penicillins.[7,8] However, whilst they can protect against Class A β-lactamases ('penicillinases') they do not protect against the 3 other classes of β-lactamases (B, C, D).

There are 4 classes of β-lactamases, the Class A, C, and D (serine β-lactamases, SBLs) enzymes and the zinc-dependent Class B β-lactamases.[9] The SBLs operate via mechanisms related to those of the penicillin binding proteins (PBPs). A key general

Abbreviations. N_S : the nitrogen of the avibactam derived hydroxylamine-O-sulfonate group, C1: the carbonyl carbon of ring-opened avibactam, NZ : the nitrogen of the side chain Lys73, OE2: oxygen of side chain carboxylic acid group Glu166, O_W : Wat411 are close to the carbamoyl group linked to the nucleophilic Ser70 from PDB entry : 4HBU.

difference, however, is that the acyl-enzyme complexes formed by reaction of the SBLs with their β-lactam antibiotic substrates are much more labile to hydrolysis than the analogous 'inhibitory' PBP complexes.[10,11] The clinically used β-lactam SBL inhibitors react to form acyl-enzyme complexes, which can fragment or further rearrange to form species that are resistant to hydrolysis in part because they provide an alternative 'sink' (i.e. a –C(R)=O/NR– group) that can 'trap or block' a potentially 'hydrolytic' water molecule.[8,11] Remarkably, like all clinically used PBP inhibitors, the three clinically used Class A SBL inhibitors (clavulanic acid, tazobactam and sulbactam) are themselves β-lactams.[8,11] The complete dominance of β-lactam-containing compounds as clinically useful PBP and β-lactamase inhibitors led to the proposal that they may be sacrosanct in this regard.[12]

Attempts to replace β-lactams began soon after the determination of the penicillin structure,[13] with the synthesis of γ-lactam and other analogues.[14] Successes were achieved when unsaturated bicyclic γ-lactams were found to be antibacterials.[14] The isolation of the cycloserine containing natural product lactivicin, further revealed the potential for useful non-β-lactam acylating agents.[15,20,52] Model studies led to the idea that an advantage of β-lactam inhibition is effectively irreversible acylation. In contrast inhibition by 5- or 6-membered lactams is compromised by kinetically favoured recyclisation of the acyl-enzyme complex to reform the lactam (Figure 2.1).[16-18]; this reversibility can be countered by features that hinder recyclisation, e.g. ring strain, steric, or electronic factors.[19-21]

Avibactam ((2*S*,5*R*)-7-oxo-6-(sulfoxy)-1,6-diazabicyclo[3,2,1]octane-2-carboxamide) is a breakthrough in β-lactamase inhibition, because it inhibits class A, C, and some class D enzymes and is the first non-β-lactam β-lactamase inhibitor to complete clinical trials.[22-25] The efficacy of avibactam is due to its unusual nitrogen-bridged bicyclic

structure containing a strained cyclic urea that was developed to enable efficient acylation by optimization of the interactions with the key components of the catalytic machinery.[24,25] Avibactam inhibits class A, C, and some class D β-lactamases including the class A extended-spectrum β-lactamase (CTX-M-15) and carbapenemase (KPC-2), and class C AmpC-type enzymes with potencies in the nanomolar range.[26,27]

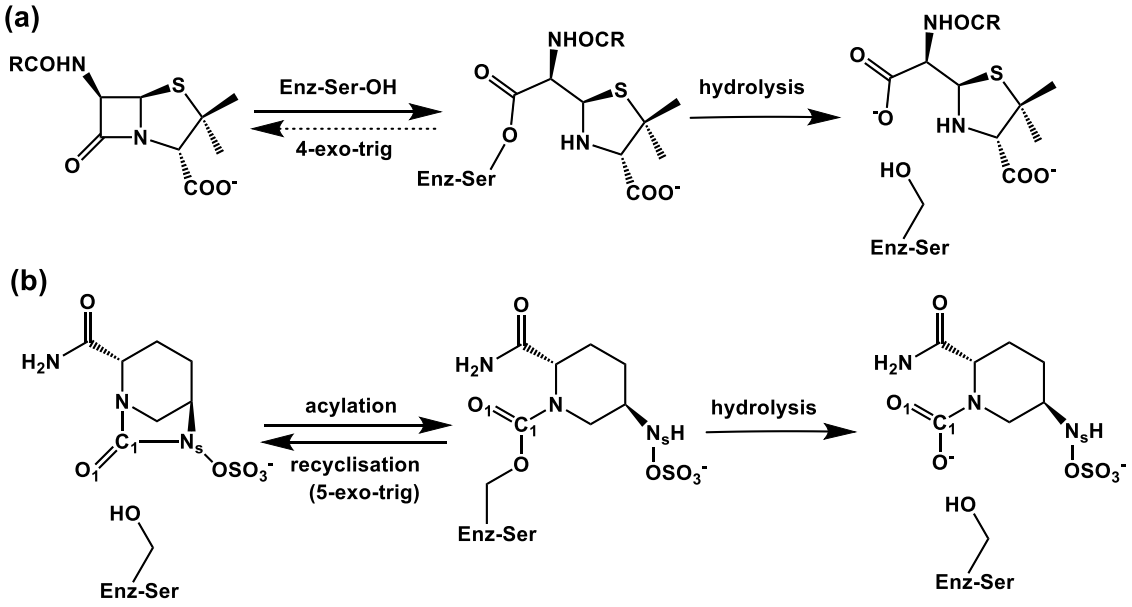


Figure 2.1. β-Lactams react irreversibly with penicillin binding proteins/β-lactamases to form an acyl-enzyme complex in contrast to avibactam which reacts reversibly. (a) Irreversible reaction of a β-lactam with a nucleophilic serine residue in the enzyme as exemplified by reaction of a penicillin with a penicillin binding protein (transpeptidase) to give a stable acyl-enzyme complex, which only undergoes slow hydrolysis; (b) Schematic representation of the acylation, recyclisation, and hydrolysis reactions between a serine β-lactamase and avibactam. C1 and N_S denote carbonyl carbon of ring-opened avibactam and the nitrogen of the avibactam derived hydroxylamine-O-sulfonate group, respectively. Figure taken from Choi et al, *Org. Biomol. Chem*. *14*, p.4117. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

Avibactam inhibits serine β-lactamases via reaction of its urea carbonyl with the nucleophilic serine of the enzyme (Figure 2.1).[30,31] However, in contrast to reaction with lactivicin (a naturally occurring non β-lactam inhibitor)[32] and β-lactam mediated acylation,[8] avibactam mediated serine β-lactamase acylation is reversible (Figure

2.1(b)).[27] Hydrolysis of the avibactam derived acyl-enzyme can occur, likely via initial loss of a sulfate ion, as shown for the KPC-2 β-lactamase, where hydrolysis is faster than for CTX-M-15.[33] Understanding the factors that regulate the balance between reversibility/irreversibility and recylisation/hydrolysis of the inhibitory complexes is important because enzymatic hydrolysis of its cyclic urea is irreversible and hence inactivates avibactam. Consequently, optimization of the hydrolysis step is one path by which bacteria may evolve resistance to avibactam.

Several crystal structures for avibactam-acylated SBLs have been reported.[34-36] In the CTX-M-15 serine β-lactamase avibactam complex,[34] the nitrogen of the hydroxylamine-O-sulfonate group of avibactam and a structural water (Wat411) are close to the carbamoyl group linked to the nucleophilic Ser70, and are apparently stabilized by the interactions with the Lys73-Ser130 and Glu166-Asn170 'catalytic dyads', respectively. Although the structures suggest recyclisation to avibactam is feasible, they do not provide information on the parameters determining its rate relative to hydrolysis.[36]

Computational studies have been used to scrutinize the mechanisms for the acylation of β-lactam antibiotics and β-lactamase inhibitors in the active site of various β-lactamases. Tripathi and Nair have reported studies on the mechanism of the formation of the covalent acyl-enzyme adduct from the non-covalent enzyme-inhibitor complex formed between class C β-lactamases and β-lactam antibiotics.[37] Based on molecular dynamics (MD) simulation studies, they were able to identify residues that play roles in general base and electrophilic catalysis in the acylation step. More recently, Sgrignani et al. recently addressed the mechanism of the formation of the carbamoyl-complex between the TEM-1 class A β-lactamase TEM-1 and avibactam.[38] The results of hybrid quantum mechanics/molecular mechanics (QM/MM) simulations indicate that the rate-

limiting process in acylation is water-assisted deprotonation of Glu166 by the nucleophilic residue Ser70 in order to form a tetrahedral intermediate. It was also concluded that the N^{ε}-amino group of Lys73 plays a key role in providing acid base catalysis, including via proton transfer to Ser-130, which in turn protonates the avibactam derived urea nitrogens.[38]

The scope of previous computational investigations of the reaction between β-lactamases and ligands has been limited to possible mechanisms of covalent complex formation. The key mechanistic question of potential recyclisation in the enzymatic active site via addition-elimination mechanisms has not been addressed even for β-lactam antibiotics (although non-enzymatic β-lactam synthesis has been studied by quantum chemical calculations.[39]). Hence, the work described in this chapter used MD simulations and *ab initio* quantum chemical calculations to investigate the recyclisation and hydrolysis of the avibactam derived acyl-enzyme complex with a clinically important SBL; comparison was made with the potential recyclization of β-lactam (oxacillin) derived acyl-enzyme complex. On the basis of the structural features of transition states and reaction intermediates identified, I investigate the molecular driving forces that render the recyclisation of avibactam from the acyl-enzyme complex favorable. In addition, the possibility and mechanism of hydrolysis of the avibactam derived complex by KPC-2 was investigated.[33] To examine the effect of the presence of the sulfate moiety in avibactam on the dissociation rate of the ring-opened avibactam covalently bound to KPC-2 enzyme, I calculated free energies and structures of the transition states and reaction intermediate on the reaction pathway for the hydrolysis of the desulfated avibactam by KPC-2. The results should serve as useful information in designing new potent inhibitors for broad-spectrum inhibition of β-lactamases. They will provide information that will be useful not only in the design of new and more potent β-lactamase

inhibitors following from avibactam, but are also of interest in predicting the properties of clinically observed substitutions of β-lactamases as they emerge in response to use of avibactam and related β-lactamase inhibitors.

2.2. Computational Methods

2.2.1. MD simulations of the of CTX-M-15 in complex with avibactam

MD simulations of CTX-M-15 in complex with avibactam were performed using the AMBER program (version 12) and the standard force field parameters.[40] To obtain the potential parameters for the avibactam molecule which are not available in the force field database, the procedure of Fox and Kollman was followed[41] to be consistent with the standard AMBER force field. The starting coordinates for MD simulations were derived from a crystal structure of CTX-M-15 in complex with avibactam (PDB entry: 4HBU).[34] To obtain an all-atom model including hydrogens for CTX-M-15, the protonation states of ionizable residues were assigned by inspecting their hydrogen bonding patterns in the crystal structure. For example, the side chains of Asp and Glu residues were assigned as neutral if either of their carboxylate oxygens was directed toward a hydrogen-bond accepting group within the distance of 3.5 Å (generally accepted distance limit for hydrogen bonds of moderate strength).[42] Similarly, the side chains of lysine residues were assumed to be protonated unless the nitrogen of side chain (NZ) atom is in proximity to a hydrogen-bond donating group. The same procedure was also applied for determining the protonation states of His residues.

Because Glu166 plays an important role in the catalytic action of CTX-M-15[34] (and most, but not all, other serine β-lactamases), the knowledge of its protonation state is crucial in mechanistic investigations. Despite the presence of X-ray crystal structures of CTX-M-15,[34] it is, however, difficult to assign the protonation state of Glu166 in an

unambiguous manner due to the presence of a 'structural' water molecule (Wat411) in its vicinity (Figure 2.2(c)). In contrast to the preference of the ionized form in the apo CTX-M-15 structure,[34] the neutral form of Glu166 was predicted to be favored in the X-ray crystal structure of CTX-M-15-avibactam complex.[34] However, the role of deprotonated Glu166 in the deacylation mechanism of other class A enzyme cannot be generalized because of the possibility that Glu166 plays a role as a general base to activate a structurally observed water nearby Wat411, (Figure 2.2) to act as a nucleophile to hydrolyze the avibactam acyl-enzyme complex; such hydrolysis occurs, albeit slowly, with KPC-2 whose active-site residues occupy very similar positions to those of CTX-M-15.[33,43]

On the basis of the crystallographically observed interatomic distances between Wat411, Glu166, Asn170, and Ser70 in the structure of CTX-M-15-Avibactam complex,[34] two possibilities for the hydrogen-bonding patterns should be taken into account according to the protonation state of Glu166 (Figure 2.2). The first involves the deprotonated form of Glu166 with Wat411 playing the roles of hydrogen bond donor with respect to Glu166 and Ser70 (Figure 2.2(a)). Alternatively, Glu166 may exist in its neutral form in CTX-M-15-avibactam complex in which Wat411 receives and donates the hydrogen bonds from Glu166 and to Asn170 and Ser70 (Figure 2.2(b)), respectively. I investigated the preferred protonation state of Glu166 by comparing stabilities of the hydrogen bond between Glu166 and Wat411 based on MD simulations of the CTX-M-15-avibactam complex in aqueous solution.

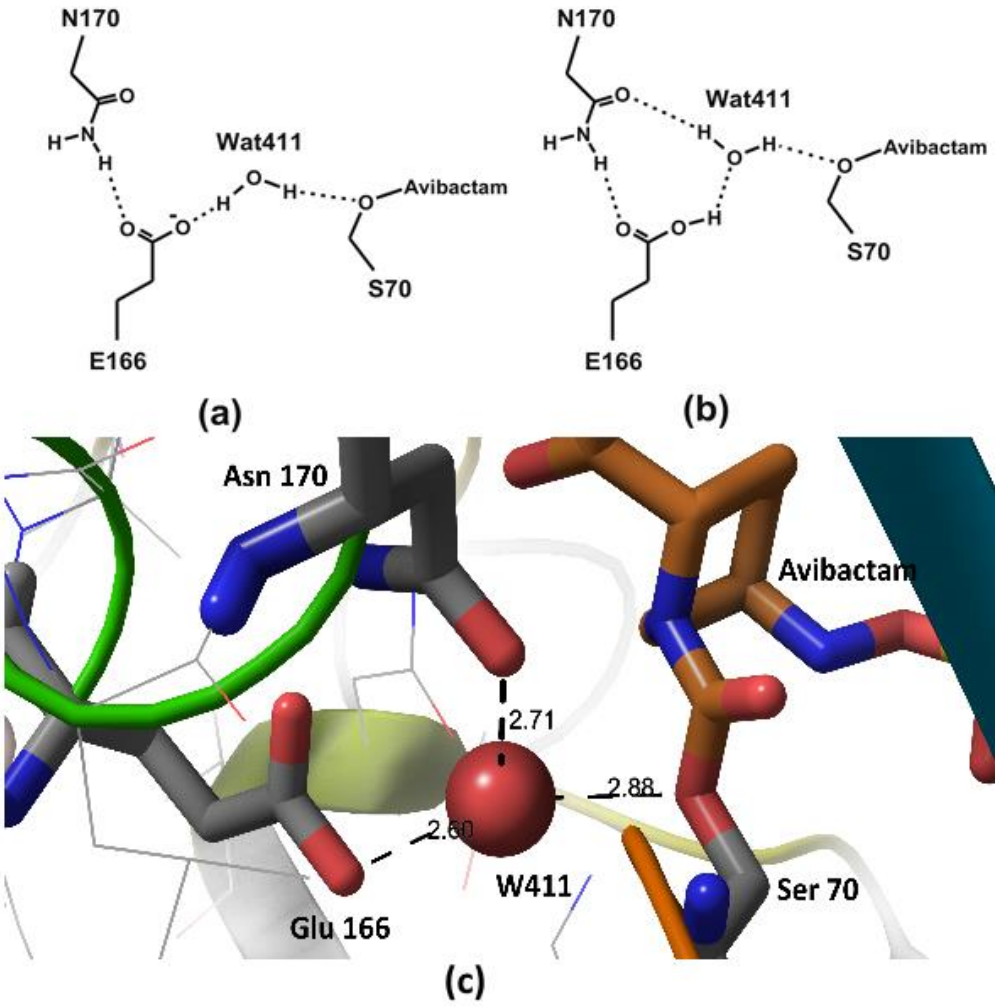


Figure 2.2. Possible protonation states of Glu166 based on crystallographic analysis of the CTX-M-15-avibactam complex represented by (a) ionized and (b) neutral forms. The dotted lines indicate potential hydrogen bonds. The structures were prepared by extracting the atomic coordinates from a CTX-M-15-avibactam complex X-ray structure. (c) Crystal structure of CTX-M-15-avibactam complex from PDB entry: 4HBU.

In simplified model systems for quantum chemical studies on the relative energetics of recyclisation versus hydrolysis of the avibactam derived acyl-enzyme propionate, acetamide, methanol, and butylamine were used to mimic the side chains of Glu166, Asn132, Asn170, Ser70, Ser130, Ser237 and Lys73 residues in the active site of CTX-M-15, respectively. Two structurally observed water molecules, Wat411 and Wat490, were included in the reaction models because they were found to interact with both

avibactam and the active-site residues in the original X-ray structure. The quantum chemical region of the system in which the chemical process takes place is treated, i.e. that surrounding the acylation position and hydrolytic water with relevant amino acids, should be considered. Overall, this model system comprised 98 atoms,

After the addition of 3 chloride ions to neutralise the total charge, the all-atom model for the CTX-M-15-avibactam complex was placed in a rectangular box of dimension 60.9 × 75.4 × 71.2 Å containing 8475 TIP3P[44] water molecules. After 2000 minimization cycles to remove poor steric contacts, the system was equilibrated beginning with 20 ps equilibration dynamics of solvent molecules at 300 K. The next step involved equilibration of the solutes with a fixed position of the solvent molecules at 10, 50, 100, 150, 200, 250, and 300 K for 10 ps at each temperature. Equilibration dynamics of the entire system were then performed at 300 K for 100 ps. Following equilibration, 10.8 ns production dynamics simulations were carried out with periodic boundary conditions in the NPT ensemble. The temperature and pressure were kept at 300 K and 1 atm using Berendsen temperature coupling[45] and isotropic molecule-based scaling (Berendsen thermostat-barostat with coupling constants of 0.1 ps and 0.5 ps), respectively. The SHAKE algorithm,[46] with a tolerance of 10^{-6} Å, was applied to fix all bond lengths involving hydrogen atoms. A time step of 2.0 fs and a nonbond-interaction cut-off radius of 12 Å was used; the trajectory was sampled every 0.4 ps (200-step intervals) for analysis. The criteria used to finalize the MD simulations followed a rigorous protocol to determine if the system is indeed in an equilibrium state. Thus, from 4 ns, I calculated the RMSD every 1 ns, verifying if the RMSD variation was lower than 1.5 and 1.0 Å for the protein backbone and side chains, respectively

2.2.2. Quantum mechanical calculations of CTX-M-15-avibactam complex

All structures corresponding to energy minima and transition states on the model enzymatic reaction pathways were fully optimized at the RHF/6-31G* level of theory using the GAMESS program.[47] Geometry optimizations were performed using analytically determined gradients and quasi-Newton-Raphson optimization algorithms.[48] The nature of each stationary-structure encountered on the reaction pathways was determined by the number of imaginary frequencies that were obtained following diagonalization of the analytical Hessian matrix. Each transition state structure was identified as having a single negative eigenvalue and the corresponding imaginary vibrational frequency was related to the motion that would connect the expected starting and final minima. The intrinsic reaction coordinate (IRC) connecting a transition state to the neighboring energy minima was determined using the Gonzalez-Schlegel second-order (GS2) method[49] at the same level of theory as used in the geometry optimizations.

To obtain a better prediction for the energetics, post-HF level calculations including the effect of electron correlation were conducted at the optimized stationary-state structures. These single point calculations were carried out with the 6-31G* basis set using Møller−Plesset second-order perturbation theory (MP2).[50] I also carried out the single point calculations using density functional theory at the B3LYP/6-31G** level.[58] The electronic energies computed in this way were used to calculate the relative free energies (ΔG) given by

$$\Delta G = \Delta E_{elec} + \Delta H' - T\Delta S \qquad (1)$$

Here, $\Delta H'$ denotes the enthalpy change due to thermal motions of the nuclei including the zero-point vibrational energies, and ΔS is the entropy change. Electronic energies (E_{elec}) were evaluated at MP2/6-31G* and B3LYP/6-31G** levels of theory and RHF/6-

31G* vibrational frequencies were used to calculate $\Delta H'$ and ΔS; thermal contributions were evaluated at 298K. Some water molecules are observed by crystallography in the active site;[34] since the calculation concerned inhibitor complexes, the effects of bulk solvent were not considered. The roles of Wat411 and Wat490 were investigated since the crystallographic analyses imply Wat411 may affect recyclization and/or hydrolysis. Wat411 is close to C1 (avibactam) as revealed by crystallography.[34]

In the explicit solvent approach, water molecules are treated as discrete entities necessitating a detailed description of interactions between solvent molecules at the atomic level. In contrast, implicit models dispense with the detail by considering the solvent to be a dielectric continuum, and are thus the more computationally efficient.

2.3. Results

2.3.1. MD simulation studies to characterise the protonation state of Glu166 of CTX-M-15 β-lactamase in complex with avibactam

To investigate the questions of reversible reaction of avibactam with its targets, I focused on the clinically important class A extended-spectrum β-lactamase CTX-M-15 for which crystal structures are available (PDB entry: 4HBU).[34] Crystallographic analyses suggest Wat411, which is positioned to hydrogen bond to Glu166, has a role in stabilizing the covalent CTX-M-15-avibactam acyl-enzyme complex (henceforth avibactam complex) (Figure. 2.2).[34] First, MD simulations to investigate the likelihood of the structure under simulation conditions and the Glu166 protonation state were carried out. The results indicate that CTX-M-15 enzyme is conformationally stable irrespective of the protonation state of Glu166 (Figure 2.3(a)).

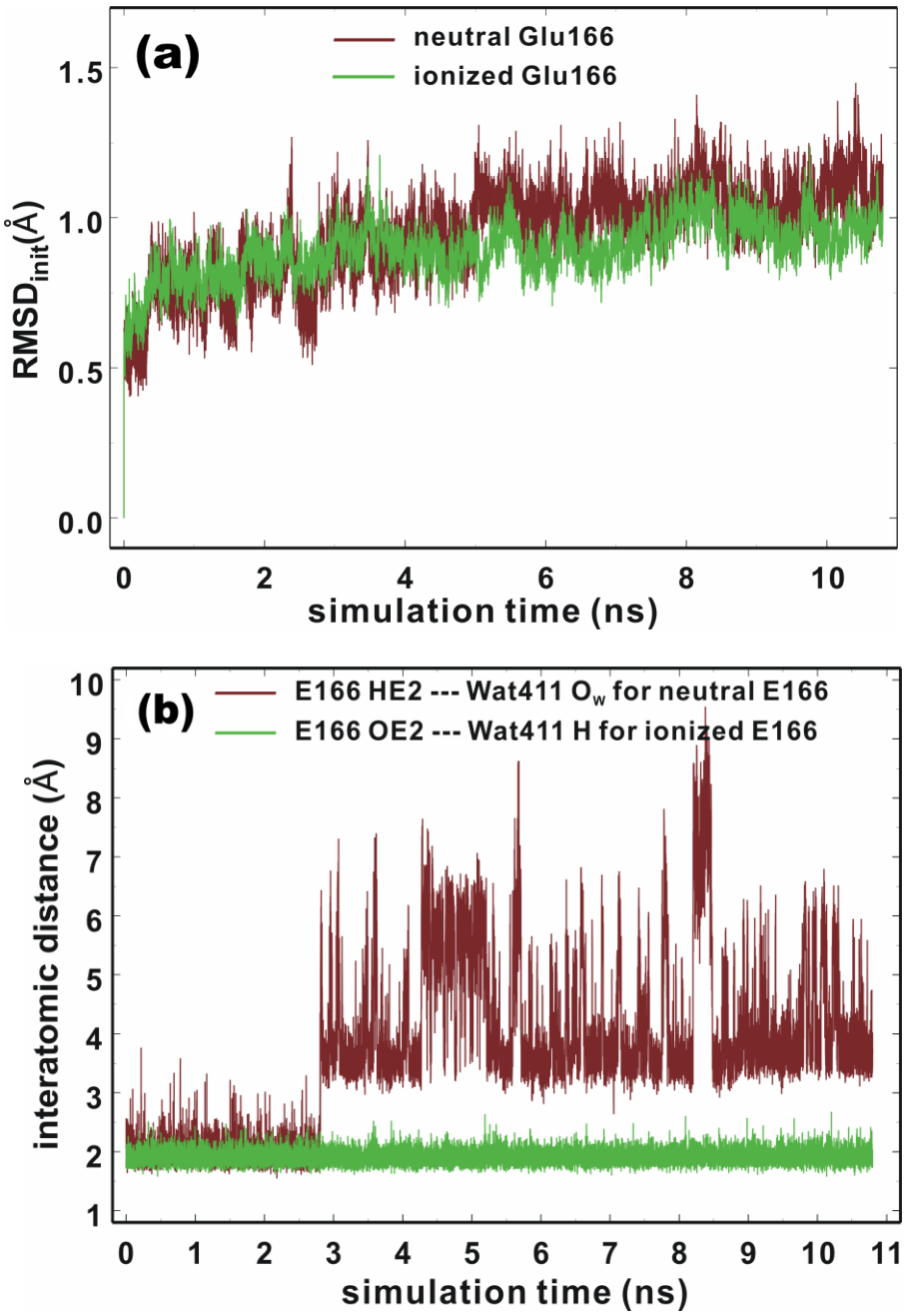


Figure 2.3. (a) Comparative view of the time evolutions of $RMSD_{init}$ values for protein backbone heavy atoms in CTX-M-15-avibactam complexes with protonated (brown) and deprotonated (green) forms of Glu166 (b) Hydrogen-bond distances between neutral Glu166 and Wat411 (brown) and between ionized Glu166 and Wat411 (green)as determined by molecular dynamic simulation of AMBER force field (ver. 12) (see Figure 2.2). Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4118. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

Studies on the time dependence of the distances associated with the hydrogen bonds of Wat411 reveal that when Glu166 is deprotonated, the hydrogen bond between Glu166 OE2 (oxygen of side chain carboxylic acid group) and Wat411 is sustained for 98% of

simulation time (Figure 2.3(b)) (when a cut-off for the O···H hydrogen bond is set at 2.2 Å).[42] The time-averaged distance between Glu166 OE2 and the Wat411 oxygen (O_W) is 2.68 Å; this is close to the reported refined distance of 2.60 Å in the CTX-M-15-avibactam structure.[34] In contrast, when Glu166 is in its neutral form, HE2 of Glu166 is positioned away from the Wat411 O_W atom, at a distance >3.0 Å for most of the simulation time after 2.81 ns. Similar dynamic behaviour is observed for the distances between Glu166 OE2 and Wat411 O_W with protonated Glu166. These results imply the crystallographically observed hydrogen bond between Glu166 and Wat411 is stable in the CTX-M-15 avibactam complex when Glu166 is deprotonated; note that this contrasts with the interpretation of the results of crystallographic studies on the CTX-M-15 avibactam acyl-enzyme complex.[34] The deprotonated form of Glu166 was thus used in subsequent investigations on recyclisation and hydrolysis of the CTX-M-15-avibactam complex.

2.3.2. MD simulations on the possibility of recyclisation of the β-lactamase-avibactam complex

Kinetic studies indicate that the avibactam complex preferentially reforms avibactam rather than undergoing hydrolysis.[27,34] To investigate the molecular basis of this, I compared the time evolution of the distance between O_W of Wat411 and the carbonyl carbon of ring-opened avibactam (C1) to that between the nitrogen of the avibactam derived hydroxylamine-O-sulfonate group (N_S) and C1 atom (henceforth N_S and C1) was compared (Figure 2.4(a)). The intramolecular N_S···C1 distances in avibactam were closer than the intermolecular O_W···C1 distances for >99% of simulation time, supporting the proposal that recyclisation is preferred over hydrolysis.

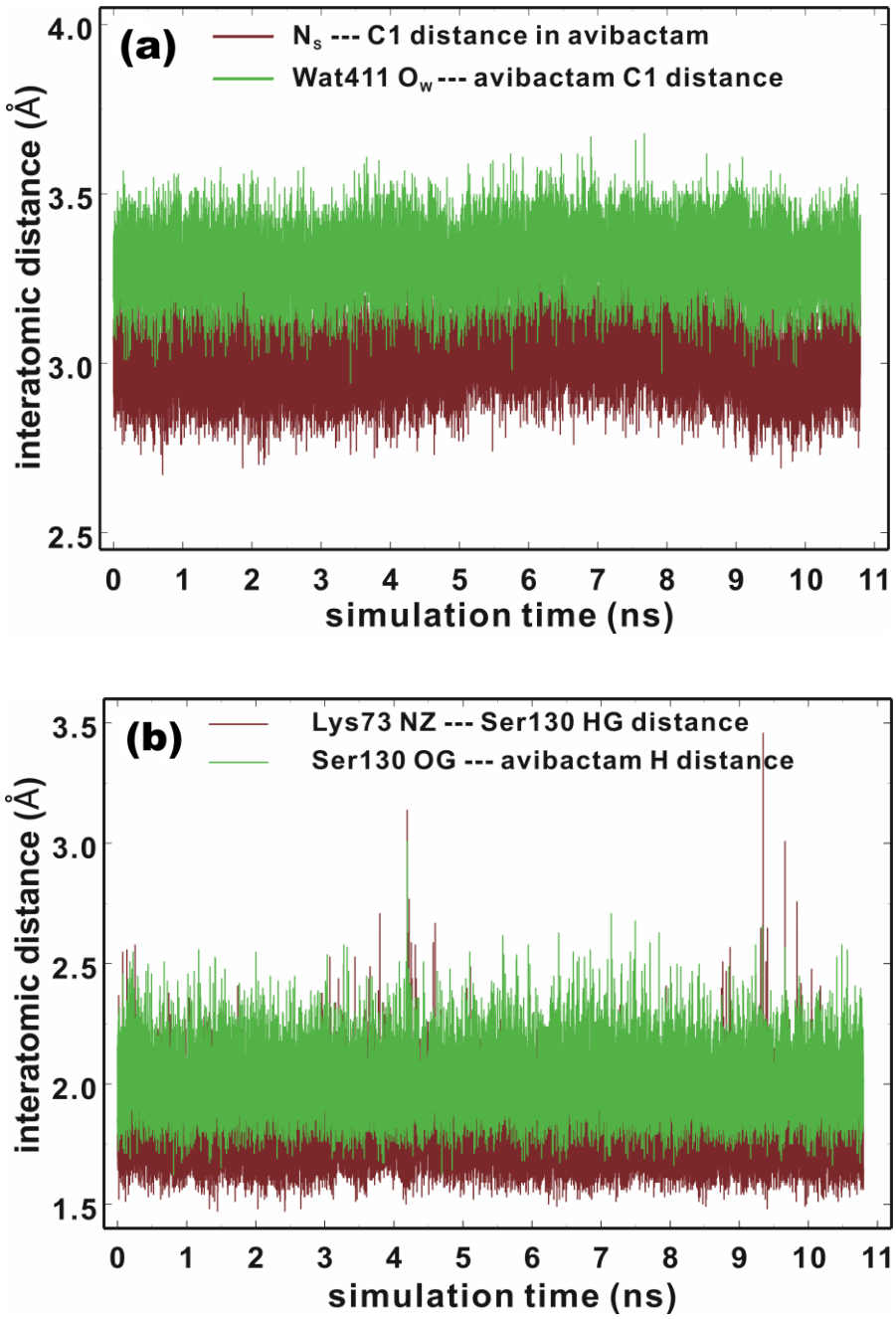


Figure 2.4. Time evolutions of selected interatomic distances during MD simulations of CTX-M-15-avibactam in explicit water model. See text and Figure 2.6 for the identification of atoms. (a) Interatomic distances between N_S and C1 atoms of avibactam (brown) and those between Wat411 O_W and C1 of avibactam (green). (b) Hydrogen bond distances between Lys73 and Ser130 (brown) and that between Ser130 and avibactam (green). Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4118. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

In order for deacylation of Ser70 in the acyl-enzyme complex to occur via an addition-elimination reaction, a stereoelectronically acceptable relationship between the nucleophile (N_S or O_W) and C1 carbonyl must be established [51] ('near attack conformer,

NAC' [53]). The N_S···C1 (i.e. the avibactam urea derived nitrogen and carbonyl) distance falls within the 3.2 Å (the proposed distance limit for NAC involving nucleophilic attack on a carbonyl)[53] for 97% of the MD simulation time, compared to only 14% for the O_W···C1 distance, again supporting recyclisation over hydrolysis. The angle formed between N_S and central carbonyl group (N_S···C1=O angle) ranges from 100 to 110° for 82% of residence time. These results indicate most of the MD trajectory snapshots of CTX-M-15-avibactam complex fall within criteria favourable for nucleophilic addition.[51,54] Note that, since MD simulations do not account for stereoelectronic effects, torsional and bonding interactions can likely generate arrangements favorable for ring closure.

According to X-ray analysis of the CTX-M-15 avibactam complex,[34] consecutive proton transfer processes (one from Ser130 to Lys73, and the other from Ns(avibactam) to Ser130) are required for recyclisation (similar processes occur during formation of the avibactam complex[38]).

The dynamic stabilities of the hydrogen bonds between Ser130 and Lys73 and between the avibactam –N_SH– group and Ser130 were thus investigated (Figure 2.4(b)). The Ser130 HG···Lys73 NZ and avibactam N_SH···Ser130 OG hydrogen bonds are present for 99% and 93% of simulation time, respectively, with respective time-averaged distances of 1.78 and 2.00 Å. Note that the hydrogen bond between Lys73 and Ser130 is unusually strong in some trajectory snapshots, with the associated N···H distance being close to 1.5 Å, supporting the possibility of proton transfer from Ser130 to Lys73, which would in turn will facilitate nucleophilic addition to C1 (carbonyl carb on of ring-opened avibactam) of the covalent complex by deprotonating N_S (hydroxyla mine-O-sulfonate group).

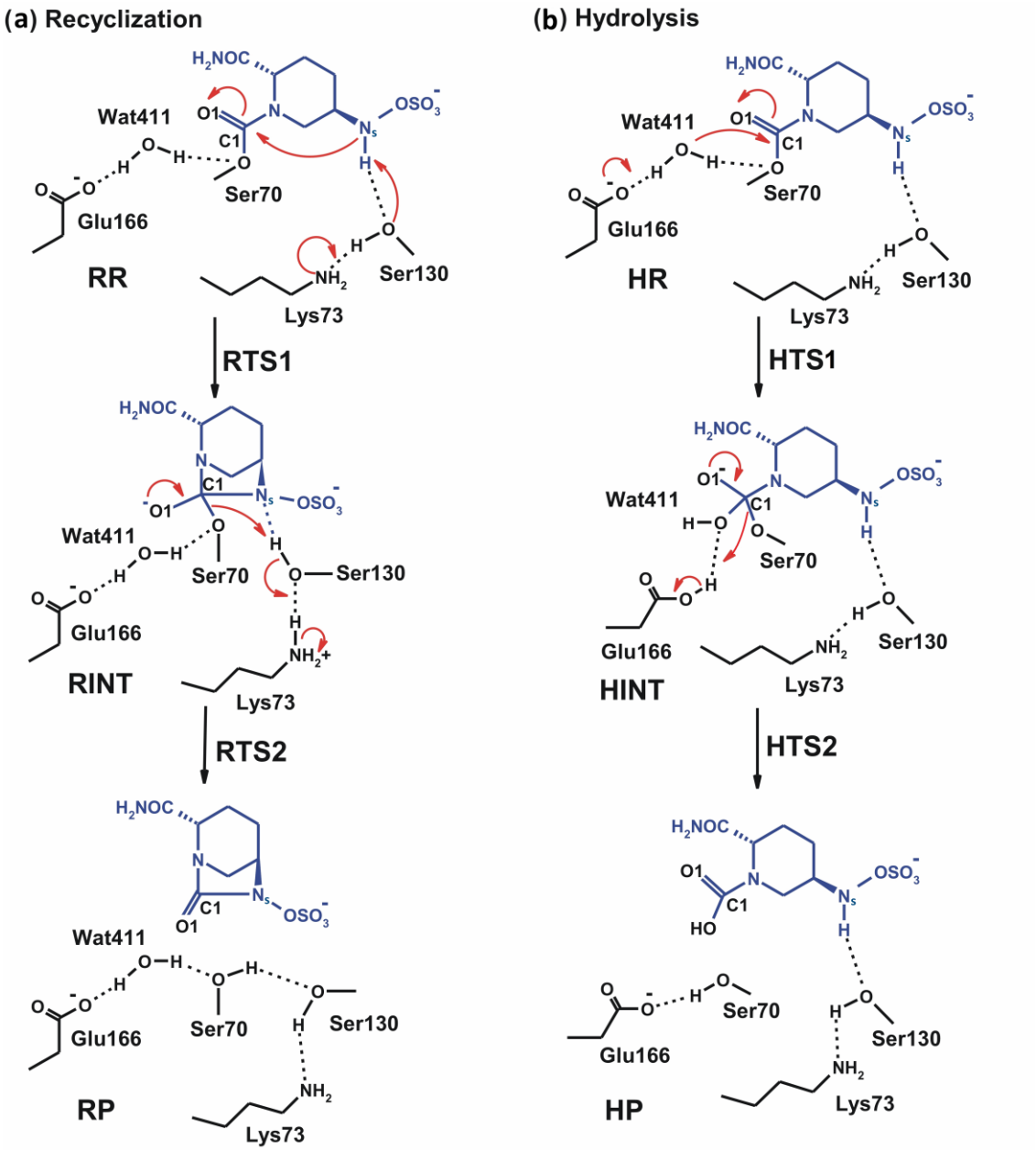


Figure 2.5. Possible fates for reaction of the CTX-M-15 avibactam covalent complex: (a) recyclisation, and (b) hydrolysis. **RR** and **HR**, **RTS1** and **HTS1**, **RINT** and **HINT**, **RTS2** and **HTS2**, and **RP** and **HP** denote the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during the recyclisation and hydrolysis reactions, respectively. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4119. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

The results again support recyclisation being favoured over carbamate hydrolysis (Figure 2.4(a)). However, hydrolysis of the avibactam complex is of interest with respect to evolution of resistance, especially given that Wat411 is close to C1 in some trajectory

snapshots (Figure 2.4(a)) in which the nucleophilic attack may be possible. Therefore, the two possibilities of reaction for the CTX-M-1 avibactam complex (Figure 2.5) using quantum chemical calculations at the MP2/6-31G*//RHF/6-31G* level were examined.

2.3.3. Avibactam recyclisation in the active site of CTX-M-15

Figure 2.6 displays the results of MP2/6-31G* calculations on the recyclisation reaction leading to avibactam. The first step involves intramolecular nucleophilic attack of N_S on C1, accompanied by proton abstraction from N_S by Ser130 (Figure 2.3).

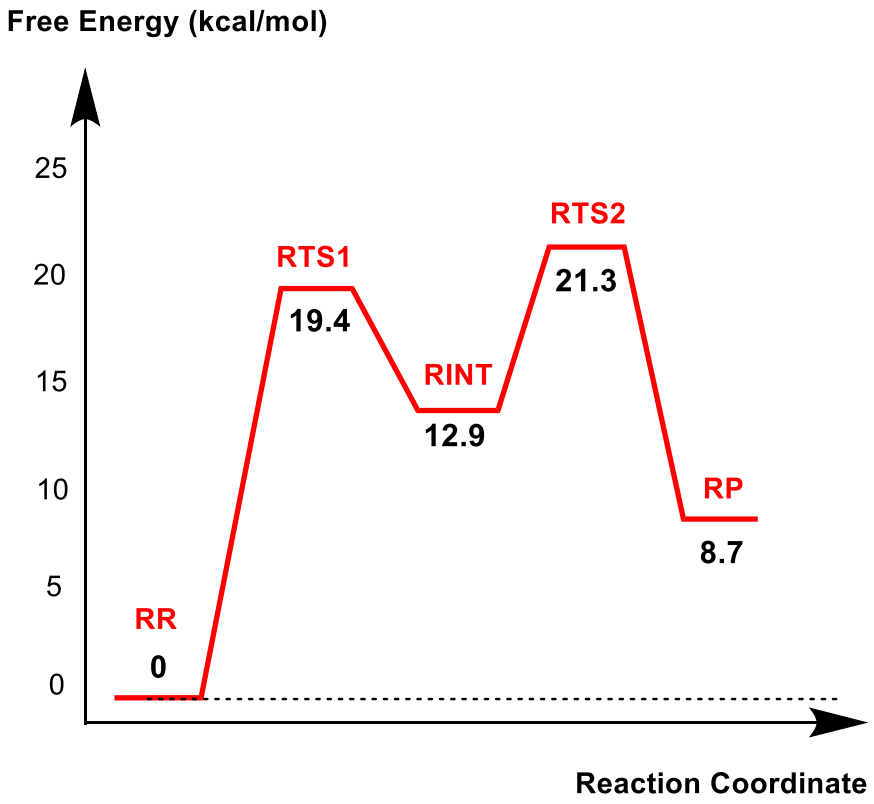


Figure 2.6. Free energy profile of the intrinsic reaction coordinates for the recyclisation to give avibactam from the covalent complex with CTX-M-15. Free energy is measured from reactant complex (**RR**) for each stationary-state structure using the MP2/6-31G* level of calculation. **RTS1**, **RINT**, **RTS2**, and **RP** represent the first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during the recyclisation reaction. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4120. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

The tetrahedral reaction intermediate (**RINT**) forms with a calculated free energy of 12.9 kcal/mol above the reactant complex (**RR**) via the first transition state (**RTS1**); the

calculated free energy of activation of which is 19.4 kcal/mol. The OG atom of Ser70 moves away from C1 in the second step to form a non-covalent avibactam complex (**RP**). The activation barrier for this reaction step is estimated to be 8.4 kcal/mol. The calculated overall free energy of activation for recyclisation amounts to 21.3 kcal/mol (from **RR** to **RTS2**).

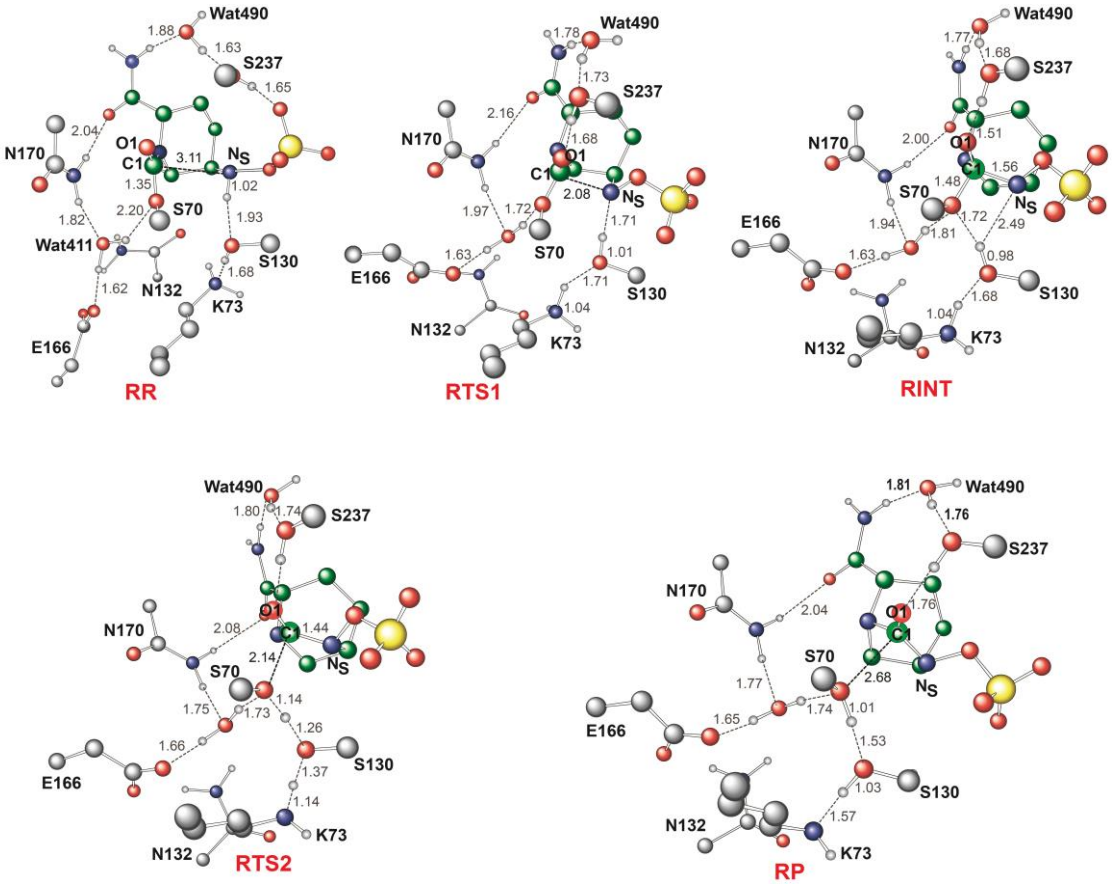


Figure 2.7. Energy minima structures and transition states identified in the reaction model for the recyclization of avibactam in the active site of CTX-M-15 using the RHF/6-31G* level of calculation. Selected interatomic distances are given in Å. **RR**, **RTS1**, **RINT**, **RTS2**, and **RP** represent the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during recylization (See Figure 2.5) Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4121. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

Although care should be taken in comparing the absolute values, due the different systems / methods involved, it is interesting to compare the calculated energetics for recyclisation to avibactam with those calculated for the avibactam ring-opening reaction

to form the covalent complex with the TEM-1 β-lactamase.[38] In the calculations at the MP2/6-31G*//RHF/6-31G* level, the activation barrier for the formation of tetrahedral intermediate (**RP** → **RINT**) is 6.1 kcal/mol higher than that for cleaving the C1-Ns bond (**RINT** → **RR**). This is consistent with the previous hybrid QM/MM investigations on the inhibition of TEM-1 enzyme by avibactam in which deprotonation of Ser70 to attack the C1 atom in order to form the tetrahedral intermediate was found to be the rate-limiting step.[38]

The structural changes occurring during recyclisation are now described (Figure 2.5 and Figure 2.7). At the starting point for recyclisation (**RR**), which was obtained with full geometry optimization, N_S of avibactam is close to C1 (3.11 Å) with a N_S···C1=O angle of 109°, i.e. within the Bürgi-Dunitz criteria.[53,54] In **RR**, however, Wat411 also remains in the vicinity of the carbamoyl group being positioned to make hydrogen bonds with the side chains of Glu166, Asn170, and Ser70. Overall, **RR** retains the structural features observed in the crystal structure of CTX-M-15-avibactam acyl-enzyme complex[34] and those in MD simulations in aqueous solution (data not shown because of similarity to the crystal structure).

The avibactam recyclisation reaction starts from **RR** with approach of N_S toward C1 (Figure 2.7). When the N_S···C1 interatomic distance falls to 2.08 Å, the reaction reaches **RTS1**; nucleophilic addition is facilitated by deprotonation of Ns by the Ser130 hydroxyl, which is in turn deprotonated by the neutral side chain of Lys73. The transition state **RTS1** is late in terms of both C1–N_S bond formation and proton transfer from N_S to Ser130, corresponding to ~65% and 95% towards formation of **RINT**, respectively. The single imaginary frequency of **RTS1** is dominated by the bond forming motions of C1–N_S bond in avibactam and OG–HG bond in Ser130, implying the activation free

energy of this step arises substantially from distortion of planar geometry around the acyl-enzyme complex carbonyl due to the nucleophilic attack of deprotonated N_S. Concomitant with approach of N_S toward C1, the Ser237 hydroxyl moves towards the carbonyl oxygen (O1) of the carbamate to avoid collision with the avibactam derived sulfate. This movement establishes a strong hydrogen bond between Ser237 and O1 in **RTS1** and **RINT**, which stabilises the developing negative charge on O1 due to the addition of N_S to C1. Simultaneously with reaction from **RR** to **RTS1**, the hydrogen bond between Wat411 and Ser70 is strengthened (the distance decreases from 2.20 to 1.72 Å). Based on the apparent maintenance of the Wat490···Ser237, Wat411···Glu166, and Asn170···Wat411 hydrogen bonds in going from **RR** to **RTS1**, it is likely that Wat490, Glu166 and Asn170 play roles in positioning Ser237 and Wat411, respectively.

Further approach of N_S to C1 to form the C1–N_S bond leads to formation of a 'tetrahedral' sp^3 hybridised intermediate (**RINT**) (Figure 2.7). Concomitant with this reaction, the Ser237···O1 hydrogen bond strengthens due to the accumulation of negative charge on O1. Interestingly, the Ser130 hydroxyl group also moves from N_S to Ser70 to serve as a hydrogen bond donor (as does Wat411). These two hydrogen bonds appear to promote cleavage of Ser70–C1 bond by stabilising the 'leaving' OG atom of Ser70.

The second step in recyclisation is initiated by elongation of the Ser70–C1 bond in **RINT**, accompanied by proton transfer from OG of Ser130 to Ser70 (Figure 2.7). This proton transfer is enabled by the protonation of Ser130 by the positively charged Lys73 side-chain. As a consequence of the two consecutive proton-transfer processes, the Ser70–C1 distance increases from 1.48 Å in **RINT** to 2.14 Å in **RTS2**. A significant feature associated with **RTS2** is that the two hydrogens attached to Ser130 and Lys73 are asymmetrically shared by the donor and acceptor atoms. This result differs from that

of **RTS1** in which the two protons are fully transferred to the acceptor atoms. Overall, the results predict that **RTS2** is intermediate and 'late' in terms of the rupture of the Ser70–C1 bond and the formation of O–H bond in Ser70, respectively, which exhibit 55% and 82% advancement to complete the second step of recyclisation.

From **RTS2**, a further increase in the Ser70···C1 distance carries the reaction system to the final minimum energy structure (**RP**), i.e. non-covalently bound avibactam in the active site (Figure 2.7). Planarity around C1 is restored in **RP** due to the cleavage of Ser70–C1 bond. Simultaneously, proton transfer processes involving protonated Ser70, Ser130, and Lys73 are completed to stabilize the leaving group. Wat411 maintains a hydrogen bond with Ser70 throughout the reaction, consistent with its role in enabling 'release' of Ser70 from the acyl-enzyme complex. The role of the Lys73-Ser130 dyad appears to be crucial in recyclisation (as in formation of the avibactam complex[38]). it acts as a general base to activate N_S as a nucleophile in the first step and as a general acid to stabilize the negative charge of the leaving Ser70 in the second step.

2.3.4. Mechanism of hydrolysis in the CTX-M-15 active site.

Figure 2.8 displays the calculated free energy profile for hydrolysis of the carbamoyl link between Ser70 and avibactam. The first activation barrier given by the free energy difference between the reactive complex (**HR**) and the first transition state (**HTS1**) is 6.3 kcal/mol higher than that for recyclisation. Similarly, the barrier for the formation of the tetrahedral reaction intermediate (**HINT**) increases by 5.6 kcal/mol when compared to that in recyclisation. Hence, consistent with experimental observations,[27] hydrolysis is calculated to be much less efficient than recyclisation. From **HINT**, an additional free energy barrier of 12.0 kcal/mol must be surmounted to reach the rate-limiting second transition state (**HTS2**) for the hydrolytic cleavage of Ser70–C1 in the acyl-enzyme complex.

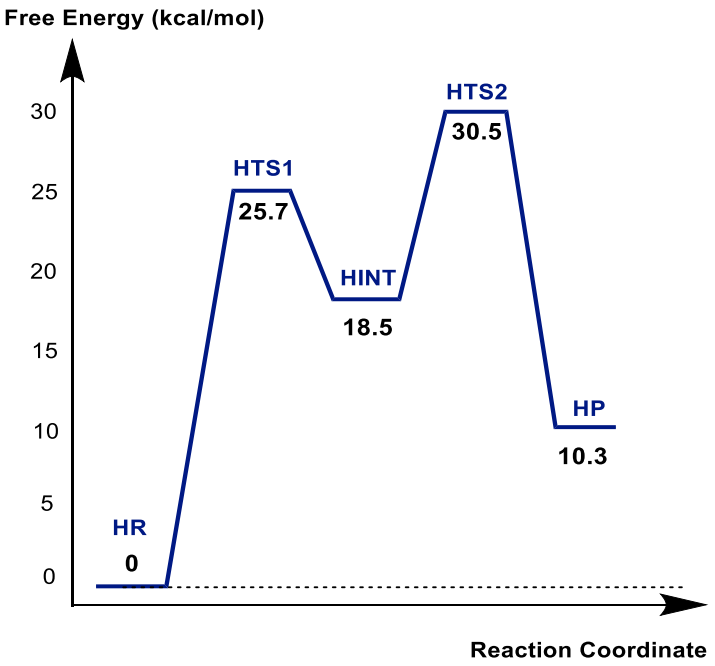


Figure 2.8. Free energy profile of the intrinsic reaction coordinate for hydrolysis of the Ser70-avibactam bond in the CTX-M-15 active site. Free energy is measured from reactant complex (**HR**) for each stationary-state structure using the MP2/6-31G* level of calculations. **HTS1**, **HINT**, **HTS2**, and **HP** represent the first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during the hydrolysis reaction. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4122. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

The calculated overall free energy of activation for hydrolysis is 30.5 kcal/mol. Importantly, this is substantially higher than for recyclisation (by 9.2 kcal/mol). These results clearly indicate a preference for recyclisation over hydrolysis, consistent with the reversibility observed in kinetic studies.[27] In contrast to the large difference in free energies of activation, the calculated free energy change for hydrolysis appears to be higher than that of the recyclisation by only 1.6 kcal/mol. Thus, kinetic factors are important in determining the fate of the CTX-M-15 avibactam complex.

Figure 2.9 displays the structures of the reacting system at the energy minima and transition states identified on the intrinsic reaction coordinate (IRC) for hydrolysis. In the reactive complex **HR** (identical to **RR** in Figure 2.7), Wat411 resides 3.83 Å distant from C1. The reaction proceeds from **HR** as the O_W atom of Wat411 approaches C1 in an approximately perpendicular manner to the molecular plane of the carbamoyl of the covalent complex. When the O_W···C1 distance reduces to 1.54 Å, the reaction system reaches **HTS1** in which C1 has undergone significant pyramidalization: the average of the four bond angles around C1 is 108° in **HTS1** as compared to 120 ° in **HR**. Nucleophilic addition is likely to be facilitated by the deprotonation of Wat411 by Glu166, as reflected by a decrease in the hydrogen bond distance from 1.62 Å in **HR** to 1.19 Å in **HTS1**. This partial deprotonation apparently causes the second hydrogen of Wat411 to 'flip' away from Ser70 towards the terminal aminocarbonyl oxygen of avibactam, leading to a new O–H···O hydrogen bond. This may have the effect of destabilizing **HTS1** due to the loss of a hydrogen bond around the reaction center. In **HTS1**, the role of general acid catalyst to stabilize the negatively charged O1 atom seems to be played by the side-chain amide moiety of Asn170 as compared to the side-chain hydroxyl group of Ser237 in **RTS1**.

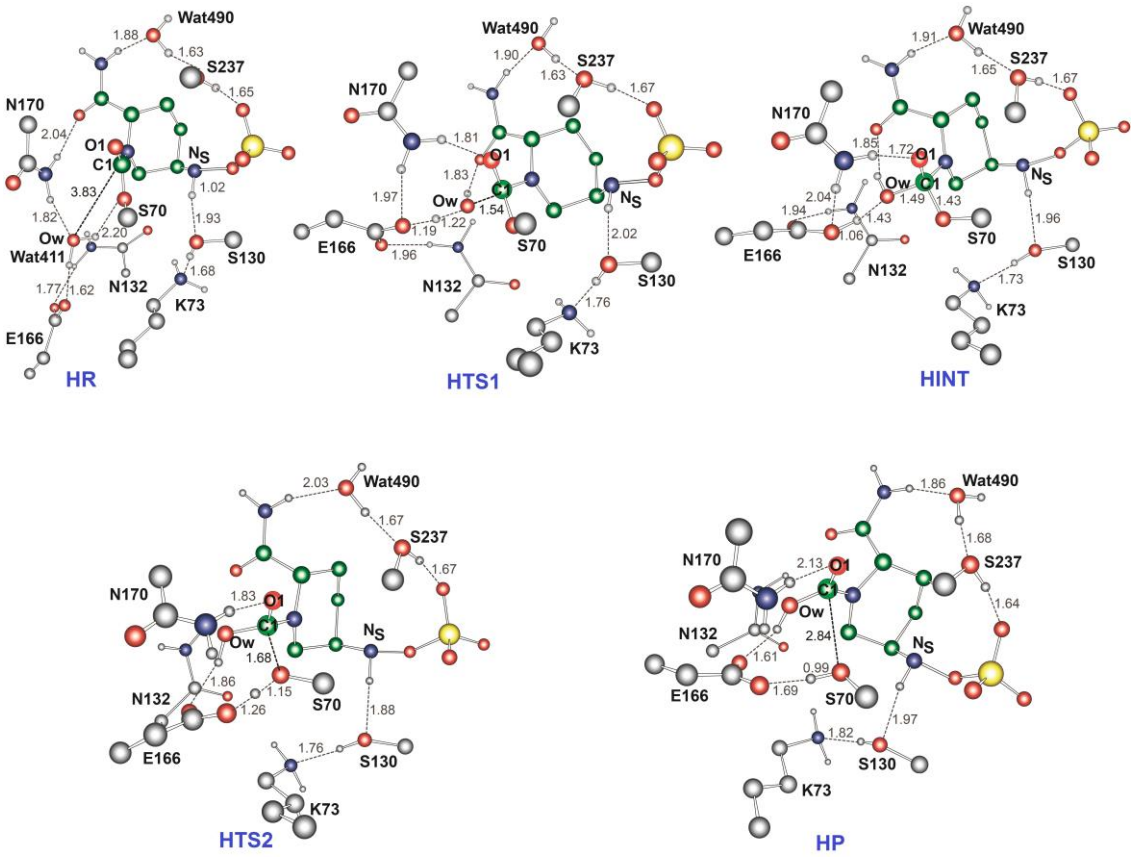


Figure 2.9. Energy minima structures and transition states identified in the enzymatic reaction model for the hydrolysis of avibactam in the active site of CTX-M-15 using the RHF/6-31G* level of calculation. Some selected interatomic distances are given in Å. **HR**, **HTS1**, **HINT**, **HTS2**, and **HP** represent the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during hydrolysis. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4123. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

However, the former seems to be less efficient than the latter in terms of the stabilization of O1 as reflected in the difference in the associated hydrogen bond distances: 1.81 Å in **HTS1** (Figure 2.9) as compared to 1.68 Å in **RTS1** (Figure 2.7). The higher activation barrier of the first step of hydrolysis mechanism than that of the recyclisation can thus be attributed to the combined effects of the loss of a hydrogen bond with Ser70 and the weakening of hydrogen-bond stabilization with respect to O1 in **HTS1**.

Further approach of O_W to C1 at **HTS1** carries the reaction system to the tetrahedral intermediate **HINT**, in which a proton from Wat411 is fully transferred to Glu166 with complete formation of the O_W–C1 bond (Figure 2.9). The second step involves the dissociation of Ser70–C1 bond and simultaneous proton transfer from the protonated Glu166 to the leaving OG atom of Ser70, which leads to the formation of **HTS2**. During these changes, the second hydrogen of Wat411 moves from the terminal aminocarbonyl oxygen of avibactam to Glu166 to make a new O–H···O hydrogen bond. It is noted that only Glu166 plays a substantial role in stabilizing the leaving OG atom of Ser70 in **HTS2** whereas Ser130 and Wat411 donate two hydrogen bonds to stabilize the leaving group in **RTS2** (Figure 2.7). The better stabilization of the leaving group (protonated Ser70) in **RTS2** than in **HTS2** rationalise the preference for recyclisation relative to hydrolysis with a large difference (9.2 kcal/mol) in the overall free energies of activation. **HTS2** then decays to form the final minimum energy structure (**HP**), i.e. the complex of the hydrolyzed reaction product (Figure 2.9). The covalent bond between Ser70 OG and C1 atoms is fully broken in **HP**, which leads to the restoration of sp^2 hybridization at the central C1 atom.

2.3.5. Putative recyclisation of oxacillin mechanism in the OXA-1 active site

The results described above imply that the recyclisation reaction of the avibactam derived acyl-enzyme complex is possible because it involves the formation of a five-membered ring rather than a four-membered one. To test this hypothesis, we investigated the free energy profile for recyclisation of the acyl-enzyme derived from the clinically important β-lactam antibiotic oxacillin. I worked with the OXA-1 β-lactamase for which a relevant high-resolution crystal structure with oxacillin is available (OXA-1 is a class D β-lactamase) (Figure 2.10).

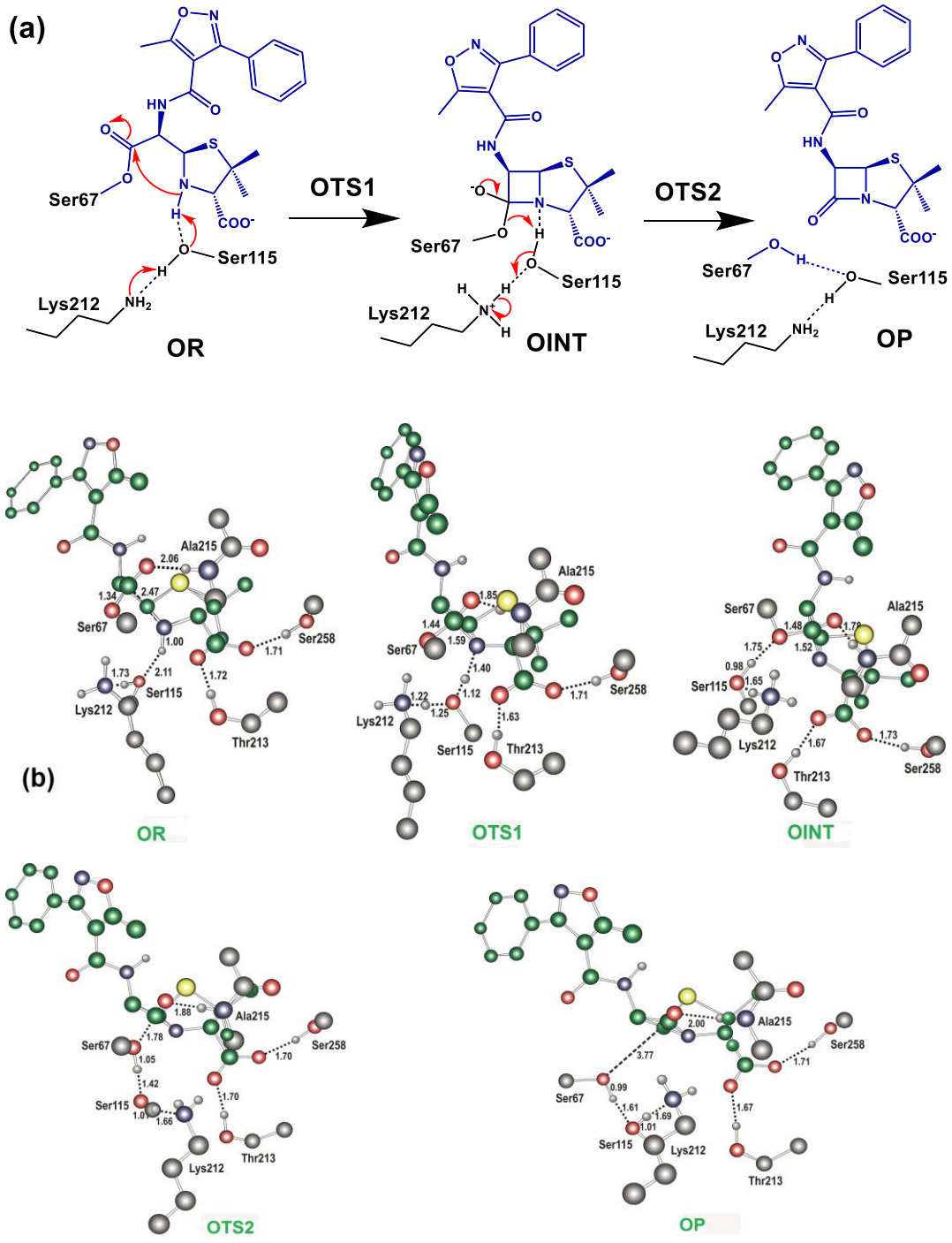


Figure 2.10. Reaction mechanism for the putative recyclisation of the oxacillin derived acyl-enzyme complex in the active site of **OXA-1**. (a) **OR**, **OTS1**, **OINT**, **OTS2**, and **OP** denote the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively. (b) Calculated structures of the energy minima and transition states of the recyclisation of oxacillin**.** The enzymatic reaction model for the recyclisation of oxacillin in the active site of OXA-1 using the RHF/6-31G* level of calculation. Some selected interatomic distances are given in Å. **OR**, **OTS1**, **OINT**, **OTS2**, and **OP** represent the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during recyclization Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4124. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

As in the case of avibactam, the MP2/6-31G*//RHF/6-31G* level of theory was employed to obtain the structures and free energies of the energy minima and transition states (**OR**, **OTS1**, **OINT**, **OTS2**, and **OP** in Figure 2.10(b)) using the X-ray crystal structure of OXA-1 in complex with the hydrolyzed oxacillin (PDB entry: 4MLL).[55]

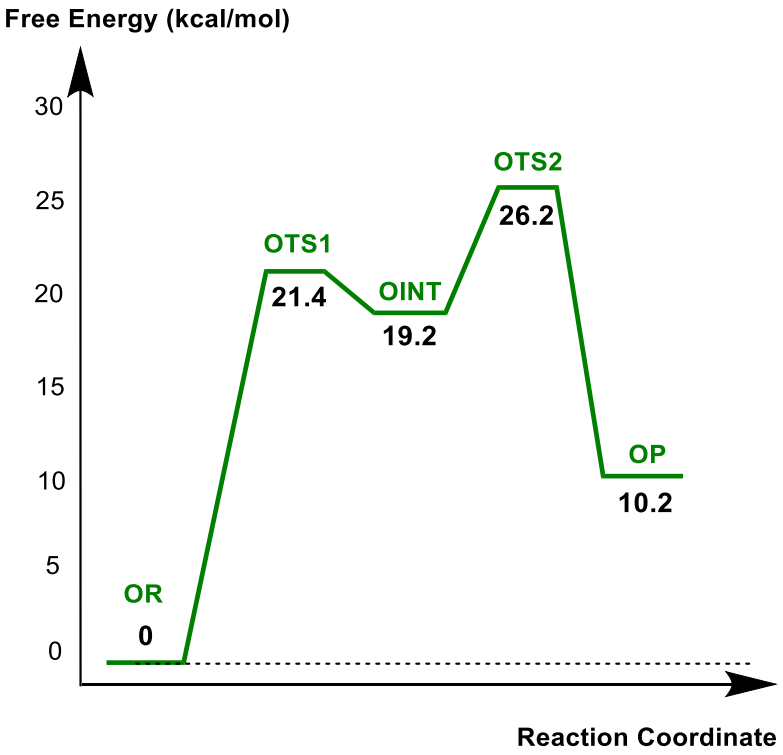


Figure 2.11. Free energy profile of the enzymatic reaction model for putative recyclisation of the oxacillin derived acyl-enzyme complex in the active site of OXA-1. Free energy is measured from the reactant complex (**OR**) for each stationary-state structure using the MP2/6-31G* level of calculation. **OTS1**, **OINT**, **OTS2**, and **OP** represent the first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during the recyclisation reaction. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4125. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

Figure 2.11 displays the free energy profile diagram along the reaction coordinate for recyclisation of the oxacillin derived acyl-enzyme complex. As for avibactam, the first step involves nucleophilic attack of the β-lactam derived nitrogen in the reactant complex (**OR**) onto the carbonyl of the ring-opened oxacillin to form a tetrahedral intermediate (**OINT**) with a free energy of 19.2 kcal/mol above **OR**. A calculated kinetic barrier of 21.4 kcal/mol must be overcome in this step to restore the four-membered ring. Cleavage

of C-O bond between oxacillin and Ser67 leads to the formation of intact oxacillin (**OP**) that is 10.2 kcal/mol higher in free energy than **OR**. The activation free energy in this second step is estimated to be 7.0 kcal/mol.

The second transition state (**OTS2**) has a higher calculated free energy than the first one (**OTS1**) by 4.9 kcal/mol. The second step in recyclisation is initiated by elongation of the Ser70–C1 bond in **OINT**, accompanied by proton transfer from OG of Ser115 to Ser67 (Figure 2.10). This proton transfer is enabled by the protonation of Ser115 by the positively charged Lys212 side-chain. As a consequence of the two consecutive proton-transfer processes, the Ser67–C1 distance increases from 1.48 Å in **OINT** to 1.78 Å in **OTS2**. Comparing this determining step of avibactam recyclization with oxacillin recyclization, avibactam is easily dissociated from covalent serine residue after ring closure (The distance of Ser70–C1 distance increases from 1.48 Å in **RINT** to 2.14 Å in **RTS2** in avibactam). A significant feature associated with **OTS2** is that the two hydrogens attached to Ser115 and Lys212 are asymmetrically shared by the donor and acceptor atoms. This result differs from that of **OTS1** in which the two protons are fully transferred to the acceptor atoms. The overall activation free energy for the conversion of **OR** to **OP** amounts to 26.2 kcal/mol, which is substantially higher (4.9 kcal/mol) than for recyclisation to form avibactam (Figure 2.6). Because the mechanisms of the two catalytic recyclisation reactions are similar, the increased ease of avibactam recyclisation compared to oxacillin is likely predominantly due to the lowering of kinetic barriers to form the five- rather than the four-membered ring. The reduced ring strain of avibactam relative to oxacillin is apparently overcome by the electronic activation provided by the bicyclic structure, thus providing lower $\Delta G^{\ddagger}$ values.

2.3.6. Hydrolysis mechanism of desulfated avibactam in the active site of KPC-2.

Recent kinetic studies demonstrate that the class A β-lactamase KPC-2 possesses the catalytic capability to hydrolyze avibactam with a different mechanism from that of CTX-M-15. [33] In the hydrolysis of avibactam in the active site of KPC-2 enzyme, an additional reaction step is required prior to the cleavage of Ser70–C1 bond, which makes the overall hydrolytic route very slow. This preliminary step involves the elimination of the sulfate group attached to the N_S atom in avibactam. This mechanism for the loss of sulfate is still not clear. However, the homology modelling and MD simulations suggest that Glu 276 can attack sulfate group of ring opened avibactam using a neighboring water. Hydrolysis at the sulfactam O-S bond would convert the sulfactam to the hydroxylamine, resulting in an acyl-enzyme complex (Figure 2.12). Dehydration of the hydroxylamine from acyl-enzyme complex would produce the imine. Because the sulfate group has been expected to contribute to the protection of the central carbamoyl linkage in β-lactamase-avibactam complexes, the desulfated avibactam seems to be more susceptible to the nucleophilic attack by water molecule than the intact avibactam.[33]

To examine the efficiency of the hydrolytic deacylation of Ser70 covalently bonded to the desulfated avibactam in the active site of KPC-2, I determined the free energy profile along the minimum energy path. In the simplified model system for KPC-2, ethanol was used to represent Thr237 instead of methanol which was employed to mimic Ser237 in the model system for CTX-M-15. The results of *ab initio* calculations for free energy changes along the KPC-2 hydrolysis reaction pathway are summarized in Figure 2.13. The activation free energies of the first and the second steps for the hydrolysis mechanism of the desulfated avibactam bound to KPC-2 are predicted to be 18.6 and 6.3 kcal/mol lower than those of the intact avibactam in the active site of CTX-M-15.

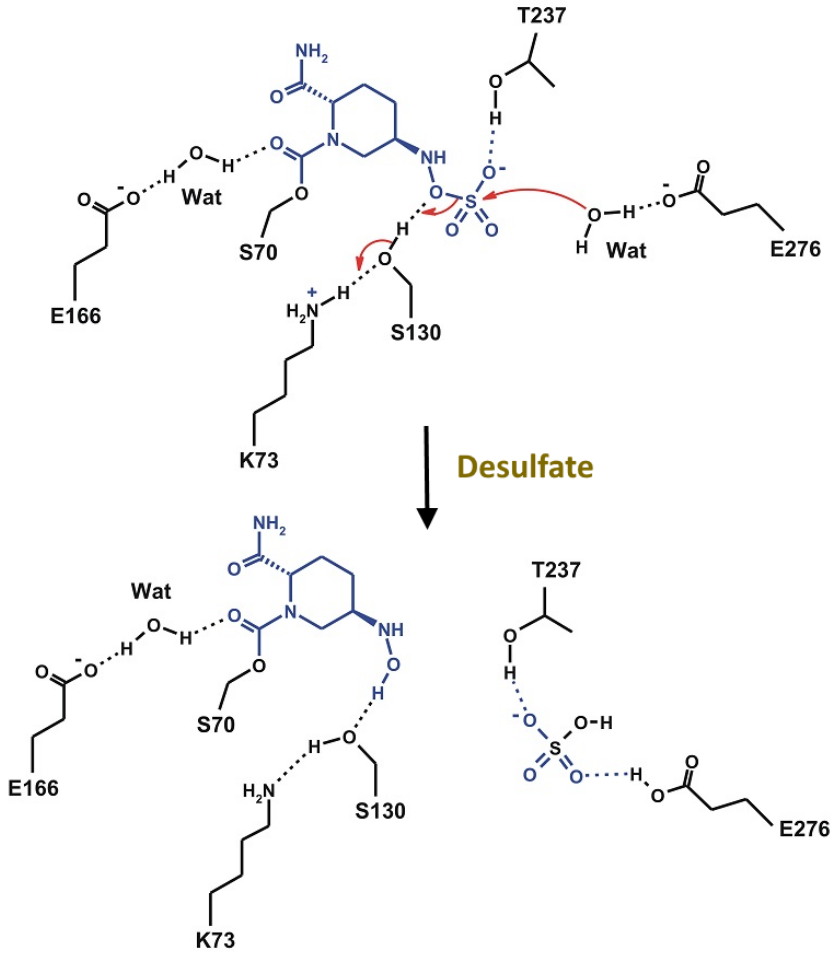


Figure 2.12. Proposed reaction mechanism for the desulfated fragmentation of ring opened avibactam by KPC-2.

The difference in overall activation free energies amounts to 21.1 kcal/mol, which indicates that the hydrolysis of desulfated avibactam by KPC-2 is catalytically more efficient than that of intact avibactam by CTX-M-15 by a factor of 10^{15}. One can infer that dissociation of sulfate from avibactam is a critical step in hydrolysis. Following loss of sulfate, the avibactam derived complex can be easily hydrolyzed. Furthermore, the thermodynamic potential for the formation of unstable reaction intermediate (**DINT**) appears to decrease from 18.5 kcal/mol in the hydrolysis mechanism of avibactam with CTX-M-15 to only 3.7 kcal/mol in that of desulfated avibactam with KPC-2. It should also be noted that the imine fragment generated as the result of the hydrolysis avibactam reaction is substantially stabilized upon complexation with the active-site residues of

KPC-2; this is reflected in the 17.4 kcal/mol decrease in free energy for the product complex (**DP**) compared to the initial reactive complex (**DR**). Thus, the hydrolytic cleavage of Ser70–C1 bond for the desulfated avibactam in complex with KPC-2 seems to be both kinetically and thermodynamically favored over that for the analogous avibactam derived complex in the active site of CTX-M-15. The energetic features of KPC-2-catalyzed hydrolysis of desulfated avibactam support the hypothesis that the sulfate group in avibactam would play an important role of protecting the covalent bond between Ser70 and the C1 atom of avibactam from the nucleophilic attack by the catalytic water molecule.

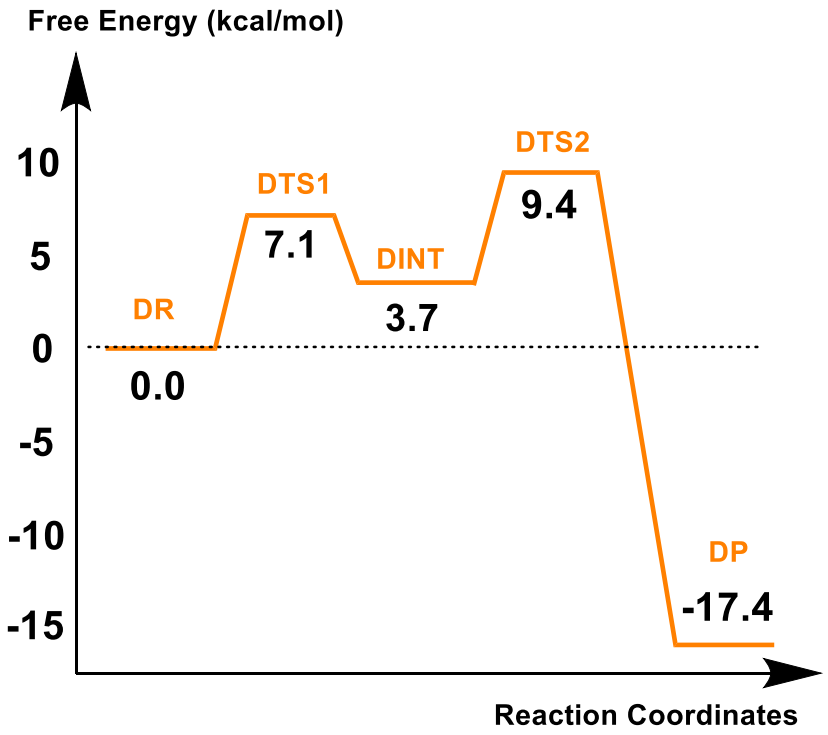


Figure 2.13. Free energy profile of the enzymatic reaction model for the hydrolysis of desulfated avibactam in the active site of KPC-2. For each stationary-state structure, the free energy measured from reactant complex (DR). DTS1, DINT, DTS2, and DP designate the first transition state, reaction intermediate, second transition state, and product complex, respectively.

I now discuss the geometrical features relevant to the hydrolysis reaction of the desulfated avibactam in the active site of KPC-2 enzyme. Figure 2.14 displays the structures of energy minima and transition states identified along IRC of the enzymatic reaction model.

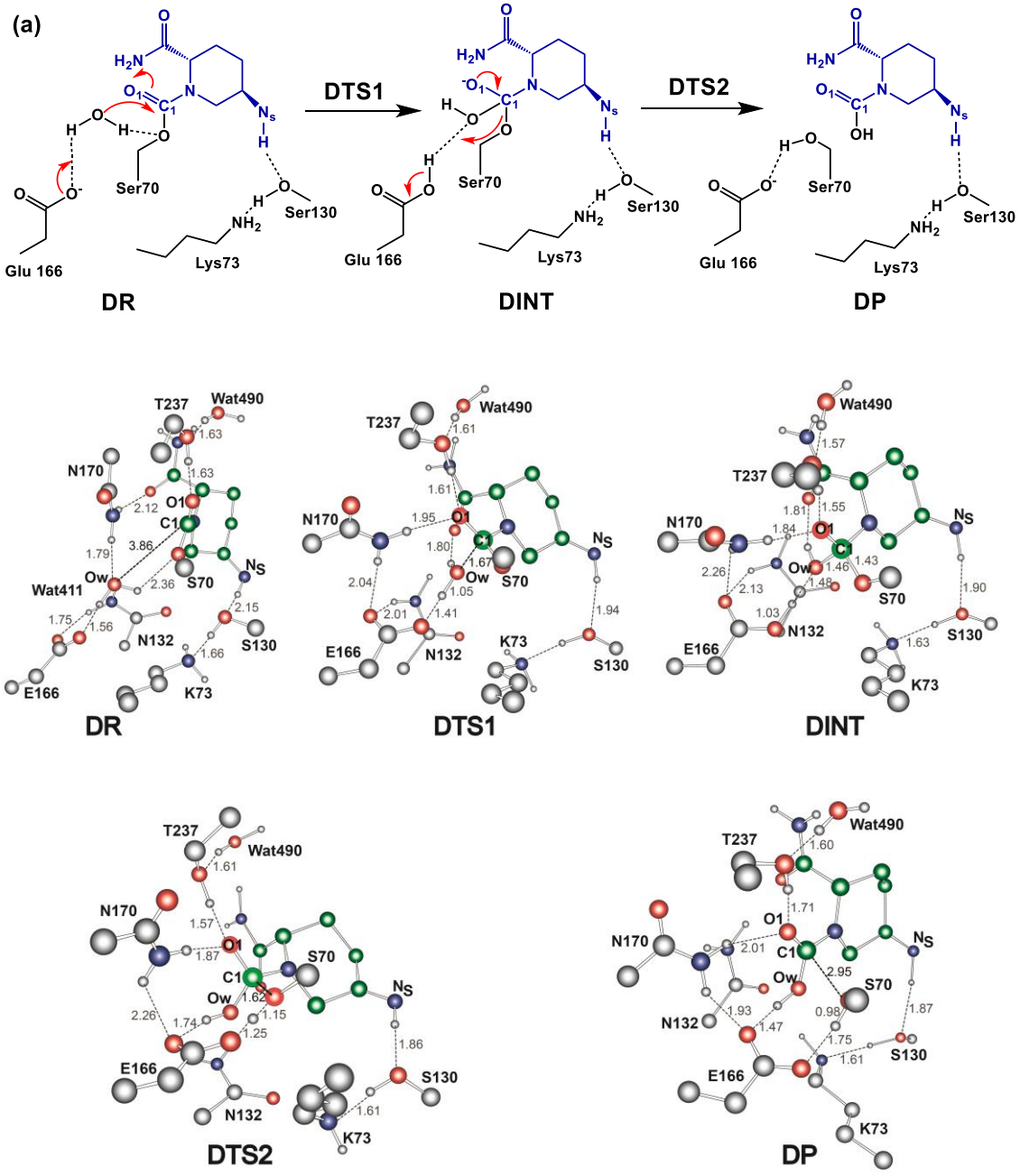


Figure 2.14. Free Structures of the energy minima and transition states identified in the enzymatic reaction model for the hydrolysis of desulfated avibactam in the active site of KPC-2. Some selected interatomic distances are given in Å.

The first minimum energy structure (**DR**) corresponds to the reactive complex in which the desulfated avibactam forms a covalent bond with OG atom of Ser70. The characteristic structural feature that discriminates **DR** from **HR** lies in that Thr237 donates a strong hydrogen bond to O1 with the associated O–H···O distance of 1.63 Å. The establishment of this hydrogen bond can apparently be made possible by the disappearance of the sulfate group in avibactam, which would have an effect of increasing the motional degrees of freedom for Thr237 that resides in close proximity to the sulfate group.

The formation of Thr237···O1 hydrogen bond induces the polarization of the carbonyl bond in the central carbamoyl group due to the electron transfer from non-bond orbital of O1 to anti-bonding O–H σ orbital of Thr237, which results in enhancement of the susceptibility of C1 to the nucleophilic attack by Wat411. Thus, the structural differences between **DR** and **HR** indicate that the removal of the sulfate group in avibactam will promote the hydrolytic cleavage of the Ser70–O1 bond in the active site of β-lactamases.

The hydrolysis reaction starts at **DR** with the proton transfer from Wat411 to the side-chain carboxylate group of Glu166 and the simultaneous approach of O_W to the central C1 atom of desulfated avibactam. The substantial contraction of O_W···C1 distance from 3.86 to 1.67 Å carries the reacting system to the first transition state (**DTS1**). In this structure, the side-chain amide group of Asn170 plays the role of electrophilic catalyst to stabilize the developing negative charge on O1 by donating a hydrogen bond together with Thr237. Like **HTS1**, the transition state **DTS1** is late in terms of the formation of O_W–C1 bond: the shortening of O_W···C1 distance by 2.19 Å in going from **DR** to **DTS1** corresponds to 91% advancement toward the formation of **DINT**. By contrast, the proton transfer from Wat411 to Glu166 makes only 28% progress to complete the first reaction

step, implying that proton abstraction lags behind formation of O_W–C1 bond. This 'transition state imbalance' may have a negative effect on the catalytic efficiency of KPC-2 due to the retardation in the charge transfer required for the hydrolysis reaction. However, such a barrier enhancing effect seems to be effectively negated by optimizing the electrophilic catalysis involving Thr237 and Asn170, which can be inferred from 18.6 kcal/mol decrease in the activation barrier for the first step as compared to that in the hydrolysis mechanism for CTX-M-15-avibactam complex.

The complete formation of the O_W–C1 bond makes the reacting system arrive at **DINT** in which a proton of Wat411 is fully transferred to Glu166. During the structural change from **DR** to **DINT** (Figure 2.14), the Thr237···O1 hydrogen bond matures into a stronger form with the contraction of H···O distance from 1.63 to 1.55 Å besides the establishment of an additional hydrogen bond between Asn170 and O1. These two strong hydrogen bonds seem to make a major contribution to the stabilization of the tetrahedral intermediate by neutralizing the negative charge on O1 developed with the nucleophilic addition of Wat411 to C1, which can be invoked to explain the low thermodynamic potential of **DINT** (3.7 kcal/mol).

A common mechanistic feature in the hydrolysis reactions of avibactam in the active sites of both CTX-M-15 and KPC-2 involves the role of Glu166 as a proton shuttle to deprotonate Wat411 in the first step and to protonate the leaving Ser70 in the second step (Figure 2.14). The action of protonated Glu166 as a general-acid catalyst with respect to the Ser70 side-chain facilitates of Ser70–O1 bond cleavage in **DINT** to form the second transition state (**DTS2**). In this structure, the proton appears to be symmetrically shared by Glu166 and Ser70. Because **DTS2** is an early transition state in terms of Ser70–O1

bond cleavage with only 13% accomplishment toward the formation of reaction products, the two hydrogen bonds involving O1 remain as strong as those in **DINT**.

Passing through **DTS2**, the reaction system falls into the final minimum energy structure (**DP**), which corresponds to the reaction product in complex with the active-site residues of KPC-2. Both hydrogen bonds involving O1 have undergone a significant weakening in going from **DTS2** to **DP** in response to the restoration of the planarity at C1. On the other hand, the carbamic acid moiety of the hydrolyzed avibactam appears to be stabilized effectively in **DP** by three hydrogen bonds with Glu166, Asn170, and Thr237 whereas the carbamoyl ester group in **DR** forms only one hydrogen bond with Thr237. Note that, in particular, that a strong hydrogen bond between the carboxylate groups Glu166 and the hydrolyzed product in **DP** is established, with an associated O···H distance of 1.47 Å. The effective stabilization of the carbamic acid group in **DP** by multiple and strong hydrogen bonds may be responsible for the highly negative reaction free energy (-17.4 kcal/mol) predicted in the hydrolysis of desulfated avibactam in the active-site of KPC-2.

2.3.7. Energetic and entropic contributions to the enzymatic reactions.

I then estimated the relative contributions of energetic and entropic contributions to the avibactam recyclisation and hydrolysis. Table 1 lists the calculated changes in electronic energy (ΔE_{elec}), thermal energy including the zero-point vibrational energy ($\Delta H^{\wedge}$), and entropic term ($-T\Delta S$) along the IRCs of the enzymatic reaction models under consideration. The thermodynamic parameters were calculated at 298.15 K and 1 atm using the usual statistical mechanical expressions and RHF/6-31G* vibrational frequencies.[56]

Energy minima and transition states	ΔE_{elec}	$\Delta H'$	$-T\Delta S$	ΔG_{MP2}
Recyclisation of avibactam by CTX-M-15				
RR	0 (0)	0	0	0 (0)
RTS1	18.7 (15.5)	-1.9	2.6	19.4 (16.2)
RINT	12.3 (14.2)	-1.5	2.2	12.9 (14.9)
RTS2	21.3 (18.7)	-1.2	1.2	21.3 (18.7)
RP	8.1 (6.6)	-0.2	0.7	8.7 (7.1)
Hydrolysis of avibactam by CTX-M-15				
HR	0 (0)	0	0	0 (0)
HTS1	26.1 (21.1)	-0.9	0.5	25.7 (20.7)
HINT	18.8 (15.9)	-1.1	0.8	18.5 (15.6)
HTS2	29.7 (25.2)	-1.3	2.2	30.5 (26.1)
HP	10.7 (7.4)	-0.1	-0.3	10.3 (7.0)
Recyclisation of oxacillin by OXA-1				
OR	0 (0)	0	0	0 (0)
OTS1	19.4 (16.1)	-4.3	6.3	21.4 (18.1)
OINT	15.2 (11.9)	-0.2	4.2	19.2 (15.9)
OTS2	27.0 (23.3)	-2.7	1.9	26.2 (22.5)
OP	13.8 (9.4)	-1.3	-2.3	10.2 (5.8)
Hydrolysis of desulfated avibactam by KPC-2				
DP	0(0)	0	0	0(0)
DTS1	6.5 (5.2)	-0.8	1.4	7.1(6.1)
DINT	2.8(1.9)	-0.8	1.7	3.7 (2.2)
DTS2	7.1(5.9)	-1.3	3.6	9.4 (7.8)
DP	-16.2 (-14.5)	0.2	-1.4	-17.4(-14.5)

Table 2.1. Relative contributions to the free energy changes of the recyclisation and hydrolysis reactions of avibactam (5-member ring) by CTX-M-15 and the recyclisation of oxacillin (4-member ring) catalyzed by OXA-1. Numbers inside and outside of parentheses are the energy values calculated at the MP2/6-31G* and wB97X/6-311++G** levels of theory, respectively. Energies are given in kcal/mol. The results for recyclisation of oxacillin by the OXA-1 β-lactamases are given for comparison. Table taken from Ch oi et al, *Org. Biomol. Chem. 14,* p.4125. Copyright 2016 Royal Society of Chemistry. Reprinted with pe rmission.

To address the dependences of the free energy profiles of the two enzymatic reaction mechanisms on the basis set and the treatment of electron correlation, I compared the ΔE_{elec} values for all the stationary-state structures (as defined in Table 1) calculated at both MP2/6-31G* and wB97X/6-311++G** levels of theory. The latter is a hybrid density functional with long-range correction for non-covalent interactions.[57] The free energy of activation ($\Delta G^{\ddagger}$) for recyclisation is calculated to be lower than that of hydrolysis based on the results of both the MP2/6-31G* and wB97X/6-311++G** calculations, i.e. 21.3 versus 30.5 and 18.7 versus 26.1 kcal/mol, respectively. The higher $\Delta G^{\ddagger}$ values in the MP2 results (9.2 kcal/mol) than in density functional calculations (7.4 kcal/mol) are unsurprising because of tendencies for MP2 to underestimate and for wB97X to overestimate electron correlation effects, respectively.[58,59] However, the differences in ΔE_{elec} values among the stationary-state structures appear similar to those in the corresponding $\Delta G^{\ddagger}$ values. Both the MP2/6-31G*//RHF/6-31G* (9.2 kcal/mol) and wB97X/6-311++G**//RHF/6-31G* (7.4 kcal/mol) levels of theory predict a large gap between the activation free energies of the recyclisation and hydrolysis reactions. The consistency between MP2 and DFT results has been observed in calculations on other protein catalysts (i.e. theozymes).[60] Thus, the overall results of MP2 and density functional calculations support a preference for avibactam recyclisation over hydrolysis.

It is notable that in both recyclisation and hydrolysis reaction of the avibactam complex, the major calculated contribution to the free energy of activation ($\Delta G^{\ddagger}$) comes from the activation enthalpy ($\Delta H^{\ddagger}$) (ΔE_{elec} plus $\Delta H^{\prime}$) rather than the entropy term ($-T\Delta S^{\ddagger}$). The entropy values amount to only 2.6 and 2.2 kcal/mol for recyclisation and hydrolysis, respectively, compared to the corresponding $\Delta H^{\ddagger}$ values of 20.1 and 28.4 kcal/mol calculated at the MP2/6-31G* level of theory. The tendency is retained when the method

to calculate ΔE_{elec} values is changed to the wB97X/6-311++G** level. Thus, recyclisation seems to be preferred over hydrolysis for CTX-M-15 enzyme because the former is more effective than the latter in lowering $\Delta H^{\ddagger}$ through the stabilization of the unstable transition states and reaction intermediates by 'direct' interactions.

Because the electronic energy term contributes predominantly to free energy changes along the reaction pathways (Table 2.1), the predicted difference in activation free energies for recyclisation and hydrolysis can likely be rationalized by proton transfers. It is notable in this regard that the required proton transfer reactions are more advanced in **RTS1** and **RTS2** (Figure 2.7) than in **HTS1** and **HTS2** (Figure 2.9), respectively, towards the formation of the intermediates and products. This proposal is consistent with the knowledge that, in general, an alkoxide is a stronger base than a carboxylate, and an alkyloxonium ion is a stronger acid than a carboxylic acid. It can thus be argued that in the recyclisation mechanism, Lys73-Ser130 dyad may serve not only as the better general base catalyst to activate the N-nucleophile in the first step, but also as the better general acid catalyst in the second step than the side chain of Glu166 that plays the similar roles in the hydrolysis mechanism. Hence, the preference for recyclisation over hydrolysis for the deacylation of Ser70 in the CTX-M-15-avibactam complex might in part be associated with better performance of the proton shuttle by Lys73-Ser130 dyad than by Glu166. The effectiveness of Lys-Ser dyad in general acid and general base catalyst has also been appreciated in catalytic reactions of various enzymes including VP4 protease,[61] penicillin-binding protein 5[62] and LexA protein[63] as well as other class A β-lactamases.[64,65]

2.4. Conclusions

Both efficient formation and the stability of the respective covalent complexes are important factors in determining the efficacy of antibiotics / β-lactamase inhibitors working via 'acylating' type mechanisms, including the β-lactam antibacterials and avibactam. The combined results presented here, and previously[37,38], support the proposal[22] that the bicyclic core of avibactam enables formation of a highly stable covalent complex with β-lactamases, in a manner comparable to that of β-lactam inhibitors. My work focused on potential reactions of this complex. Based on MD simulations of CTX-M-15 enzyme in complex with avibactam and quantum chemical calculations on relevant mechanistic reaction models, I examined the possibilities of the competing recyclisation and hydrolysis fates of the β-lactamase-avibactam complex, processes which are relevant to potency and, potentially, resistance. In most of the trajectory snapshots collected from MD simulations, the nitrogen (N_S) of the hydroxylamine-O-sulfonate group of avibactam is closer to the carbonyl carbon (C1) of the carbamoyl moiety than the active site water molecule (Wat411) bound to Glu166. The hydrogen bond between the N_S atom of avibactam and Lys73-Ser130 dyad is also observed to be in a stable form during almost the entire course of simulation. The MD simulation results thus support preferential recyclisation over hydrolysis for the avibactam complex. In contrast, analogous calculations imply that the recyclisation of an exemplary β-lactam, oxacillin, is disfavored, as anticipated based on work with small molecules.[66] The results of *ab initio* calculations at MP2/6-31G*//RHF/6-31G* and wB97X/6-311++G** levels of theory on the model systems provide further evidence that the avibactam recyclisation mechanism is kinetically favored over the hydrolysis mechanism with the difference of 9.2 kcal/mol in the activation free energies, supporting the previous experimental implications.[27,33] In both deacylation fates, the entropic

penalties for the formation of transition states and reaction intermediates were small compared to the variations of electronic energies. Consistent with the difference in activation free energies, recyclisation appears to be more efficient than hydrolysis in terms of stabilizing the negative charges developed on the carbonyl oxygen (O1) of the acyl-enzyme complex and the leaving OG atom of Ser70 of the transition states and the tetrahedral reaction intermediate. The differences in the structural features of the transition states involved in the two deacylation fates indicate that, when compared to the role of Glu166 in the hydrolysis, during recyclisation, the Lys73-Ser130 dyad can serve not only as the better general base to facilitate the nucleophile addition in the first step but also as the better electrophilic catalyst to stabilize the leaving group in the second step. The combination of the results of this study and those of previous work[27,33,34,38] implies that the avibactam template is not only optimized for binding to give a covalent complex, but that recyclisation to reform active avibactam is favoured over hydrolysis to give an inactive product. In this sense avibactam, and other reversibly and covalently binding inhibitors, may thus have an advantage over (at least bicylic) β-lactam inhibitors which recyclisation appears not be an option as shown by our calculations on the acyl-enzyme complex formed from oxacillin.

By any reasonable standards the β-lactam antibacterials exemplify the therapeutic power of covalent enzyme-inhibition via acylation of a nucleophilic residue. The special properties of β-lactams have been proposed in part to arise from their ability to acylate a nucleophilic active site residue which is unusually stable with respect to both recyclisation and, when appropriately functionalised, to hydrolysis.[18] In contrast, since the formation of larger lactam ring sizes can be chemically easier than β-lactams for stereoelectronic reasons,[66] they can undergo more facile recyclisation, a property that might normally be considered disfavourable in terms of potency.[16,18] However, when

considering resistance, reversible reaction of a covalently reacting inhibitor may be a useful property since recyclisation to reform an active inhibitor is referable to hydrolysis yielding an inactive one.[16,17] Expanded ring sizes may be favourable in this regard, and the inability of β-lactams to undergo recyclisation could be regarded an Achilles heel compared to analogous reversibly reacting inhibitors. With the expanding clinical use of the breakthrough non-β-lactam β-lactamase inhibitor avibactam and likely inspired successors, time will tell.

Chapter 2 References

1. Butler, M. S., Blaskovich, M. A., Cooper, M. A. Antibiotics in the clinical pipeline in 2013. *J. Antibiot.* **66**, 571-591 (2013).
2. Mattei, P. J., Neves, D., Dessen, A. Bridging cell wall biosynthesis and bacterial morphogenesis. *Curr. Opin. Struct. Biol.* **20**, 749-755 (2010).
3. Frère, J-M., Duez, C., Ghuysen, J-M., Vandekerkhove, J. Occurrence of a serine residue in the penicillin-binding site of the exocellular DD-carboxy-peptidase-transpeptidase from Streptomyces R61. *FEBS Lett.* **70**, 257-260 (1976).
4. Arya, N., Jagdale, A. Y., Patil, T. A., Yeramwar, S. S., Holikatti, S. S., Dwivedi, J., Shishoo, C. J., Jain, K. S. The chemistry and biological potential of azetidin-2-ones. *Eur. J. Med. Chem.* **74**, 619-56 (2014).
5. van Berkel, S. S., Nettleship, J. E., Leung, I. K. H., Brem, J., Choi, H., Stuart, D. I., Claridge, T. D. W., McDonough, M. A., Owens, R. J., Ren, J., Schofield, C. J. Binding of (5S)-penicilloic acid to penicillin binding protein 3. *ACS Chem. Biol.* **8**, 2112-2116 (2013).
6. Bebrone, C., Lassaux, P., Vercheval, L., Sohier, J. S., Jehaes, A., Sauvage, E., Galleni, M. Current challenges in antimicrobial chemotherapy: focus on ß-lactamase inhibition. *Drugs*, **70**, 651-679 (2010).
7. Padayatti, P. S., Helfand, M. S., Totir, M. A., Carey, M. P., Carey, P. R., Bonomo, R. A., van den Akker, F. High resolution crystal structures of the trans-enamine intermediates formed by sulbactam and clavulanic acid and E166A SHV-1 β-lactamase. *J. Biol. Chem.* **280**, 34900-34907 (2005).
8. Drawz, S. M., Bonomo, R. A. Three decades of beta-lactamase inhibitors. *Clin. Microbiol. Rev.* **23**, 160-201 (2010).
9. Bush, K., Jacoby, G. A. Updated functional classification of beta-lactamases. *Antimicrob. Agents Chemother.* **54**, 969-976 (2010).
10. Cornaglia, G., Giamarellou, H., Rossolini, G. M. Metallo-β-lactamases: a last frontier for β-lactams? *Lancet Infect. Dis.* **11**, 381-393 (2011).
11. Matagne, A., Dubus, A., Galleni, M., Frère, J. M. The beta-lactamase cycle: a tale of selective pressure and bacterial ingenuity. *Nat. Prod. Rep.* **16**, 1-19 (1999).
12. Hamilton-Miller, J. M. T. β-Lactams: variations on a chemical theme, with some surprising biological results. *J. Antimicrob. Chemother.* **44**, 729-734 (1999)
13. Clarke, H. T., Johnson, J. R., Robinson, R. *The Chemistry of Penicillin*, Princeton University Press, Princeton, New Jersey, (1949).
14. Baldwin, J. E., Lynch, G. P., Pitlik, J. Gamma-lactam analogues of beta-lactam antibiotics. *J. Antibiot.* **44**, 1-24 (1991).
15. Nozaki, Y., Katayama, N., Ono, N., Tsubotani, H., Harada, S., Okazaki, H., Nakao, Y. Binding of a non-β-lactam antibiotic to penicillin-binding proteins. *Nature* **325**, 179-180 (1987).
16. Wilmouth, R. C., Kassamally, S., Westwood, N. J., Sheppard, R. J., Claridge, T. D. W. Aplin, R. T., Wright, P. A., Pritchard, G. J., Schofield, C. J. Mechanistic insights into the inhibition of serine proteases by monocyclic lactams. *Biochemistry* **38**, 7989-7998 (1997).
17. Wilmouth, R. C., Westwood, N. J., Anderson, K., Brownlee, W., Claridge, T. D. W., Clifton, I. J., Pritchard, G. J., Aplin, R. T., Schofield, C. J. Inhibition of elastase by N-sulfonylaryl beta-lactams: anatomy of a stable acyl-enzyme complex. *Biochemistry* **37**, 17506-17513 (1998).
18. Westwood, N. J., Claridge, T. D. W., Edwards, P. N., Schofield, C. J. Reversible acylation of elastase by γ-lactam analogues of β-lactam inhibitors. *Bioorg. Med. Chem. Lett.* **7**, 2973-2978 (1997).
19. Wright, P. A. Wilmouth, R. C., Clifton, I. J., Schofield, C. J. 'pH-jump' crystallographic analyses of gamma-lactam-porcine pancreatic elastase complexes. *Biochem. J.* **351**, 335-340 (2000).
20. Macheboeuf, P., Fischer, D. S., Brown. Jr., T., Zervosen, A., Luxen, A., Joris, B., Dessen, A., Schofield, C. J. Structural and mechanistic basis of penicillin-binding protein inhibition by lactivicins. *Nat. Chem. Biol.* **3**, 565-569 (2007).
21. Migaud, M. E., Wilmouth, R. C., Mills, G. I., Wayne, G. J., Risley, C., Chambers, C., Macdonald, S. J., Schofield, C. J. 5,5-Fused thiophene gamma-lactams as templates for serine protease inhibition. *Chem. Commun*. **12**, 1274-1275 (2002).
22. Coleman, K. Diazabicyclooctanes (DBOs): a potent new class of non-β-lactam β-lactamase inhibitors. *Curr. Opin. Microbiol.* **14**, 550-555 (2011).
23. Vazquez, J. A., González-Patzán, L. D., Stricklin, D., Duttaroy, D. D, Kreidly, Z., Lipka, J., Sable, C. Efficacy and safety of ceftazidime-avibactam versus imipenem-cilastatin in the treatment of complicated urinary tract infections, including acute pyelonephritis, in hospitalized adults: results of a prospective, investigator-blinded, randomized study. *Curr Med Res Opin.* **28**, 1921-1931 (2012).

24. Lucasti, C., Popescu, I., Ramesh, M.K., Lipka, J., Sable, C. Comparative study of the efficacy and safety of ceftazidime/avibactam plus metronidazole versus meropenem in the treatment of complicated intra-abdominal infections in hospitalized adults: results of a randomized, double-blind, Phase II trial. *J Antimicrob Chemother*. **68**, 1183-1192 (2013).
25. Bush, K. Improving known classes of antibiotics: an optimistic approach for the future. *Curr. Opin. Pharmacol*. **12**, 527-534 (2012).
26. Aszodi, J., Rowlands, D. A., Mauvais, P., Collette, P., Bonnefoy, A., Lampilas, M. *Bioorg. Med. Chem. Lett.* Design and synthesis of bridged gamma-lactams as analogues of beta-lactam antibiotics. **14**, 2489-2492 (2004).
27. Ehmann, D. E., Jahic, H., Ross, P. L., Gu, R. F., Hu, J., Kern, G., Walkup, G. K., Fisher, S. L. Avibactam is a covalent, reversible, non-β-lactam β-lactamase inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11663-11668 (2012).
28. Stachyra, T., Levasseur, P., Pechereau, M. C., Girard, A. M., Claudon, M., Miossec, C., Black, M. T. In vitro activity of the β-lactamase inhibitor NXL104 against KPC-2 carbapenemase and Enterobacteriaceae expressing KPC carbapenemases. *J. Antimicrob. Chemother.* **64**, 326-329 (2009).
29. Stachyra, T., Pechereau, M. C., Bruneau, J. M., Claudon, M., Frere, J. M., Miossec, C., Black, M. T. Mechanistic studies of the inactivation of TEM-1 and P99 by NXL104, a novel non-beta-lactam beta-lactamase inhibitor. *Antimicrob. Agents Chemother.* **54**, 5132-5138 (2010).
30. Harada, S., Tsubotani, S., Hida, T., Ono, H., Okazaki, H. Structure of lactivicin, an antibiotic having a new nucleus and similar biological activities to β-lactam antibiotics. *Tetrahedron Lett.* **27**, 6229-6232 (1986).
31. Harada, S., Tsubotani, S., Hida, T., Koyama, K., Kondo, M., Ono, H. Chemistry of a new antibiotic: lactivicin. *Tetrahedron* **44**, 6589-6606 (1988).
32. Brown. Jr., T., Charlier, P., Herman, R. Schofield, C. J., Sauvage, E. Structural basis for the interaction of lactivicins with serine beta-lactamases. *J. Med. Chem.* **53**, 5890-5894 (2010).
33. Ehmann, D. E., Jahić, H., Ross, P. L., Gu, R.-F., Hu, J., Durand-Réville, T. F., Lahiri, S., Thresher, J., Livchak, S., Gao, N., Palmer, T., Walkup, G. K., Fisher, S. L. Kinetics of avibactam inhibition against Class A, C, and D β-lactamases. *J. Biol. Chem.* **288**, 27960-27971 (2013).
34. Lahiri, S. D., Mangani, S., Durand-Reville, T., Benvenuti, M., Luca, F. D., Sanyal, G., Docquier, J.-D. Structural insight into potent broad-spectrum inhibition with reversible recyclization mechanism: avibactam in complex with CTX-M-15 and Pseudomonas aeruginosa AmpC β-lactamases. *Antimicrob. Agents Chemother.* **57**, 2496-2505 (2013).
35. Lahiri, S. D., Mangani, S., Jahic, H., Benvenuti, M., Durand-Reville, T. F., Luca, F. D., Ehmann, D. E., Rossolini, G. M., Alm, R.A., Docquier, J.-D. Molecular basis of selective inhibition and slow reversibility of avibactam against class D carbapenemases: a structure-guided study of OXA-24 and OXA-48. *ACS. Chem. Biol.* **10**, 591-600 (2015).
36. King, D., King, A., Lal, S. M., Wright, G. D., Strynadka, N. C. J. Molecular Mechanism of Avibactam-Mediated β-Lactamase Inhibition. *ACS. Infect. Dis.***1**, 175-184 (2015).
37. Tripathi, R., Nair, N. N. Mechanism of Acyl–Enzyme Complex Formation from the Henry–Michaelis Complex of Class C β-Lactamases with β-Lactam Antibiotics. *J. Am. Chem. Soc.* **135**, 14679-14690 (2013).
38. Sgrignani, J., Grazioso, G., Amici, M. D., Colombo, G. Inactivation of TEM-1 by Avibactam (NXL-104): Insights from Quantum Mechanics/Molecular Mechanics Metadynamics Simulations. *Biochemistry* **53**, 5174-5185 (2014).
39. Del Rio, E., Lopez, R., Menendez, R. M. I., Sordo, T. L., Ruiz-Lopez, M. F. Theoretical study of ester enolate–imine condensation route to β-lactams. *J. Comput. Chem.* **19**, 1826-1833 (1998).
40. Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz. Jr., K. M., Onufriev, A., Simmerling, A. C., Wang, B., Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688 (2005).
41. Fox, T., Kollman, P. A. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B* **102**, 8070-8079 (1998).
42. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press, Oxford, (1997).
43. Ke, W., Bethel, C. R., Thomson, J. M., Bonomo, R. A., van den Akker, F. Crystal structure of KPC-2: insights into carbapenemase activity in class A β-lactamases. *Biochemistry* **46**, 5732-5740 (2007).
44. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.***79**, 926-935 (1983).
45. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).

46. Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).
47. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S. J., Windus, T. L., Dupuis, M., Montgomery, J. A. General atomic and molecular electronic structure system. *J. Comp. Chem.* **14**, 1347-1363 (1993).
48. Helgaker, T. Transition-state optimizations by trust-region image minimization. *Chem. Phys. Lett.* 182, 503-510 (1991).
49. Gonzalez, C., Schlegel, B. H. An improved algorithm for reaction path following. *J. Chem. Phys.* **90**, 2154-2161 (1989)
50. Møller, C., Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems *Phys. Rev*. **46**, 618-622 (1934).
51. Burgi, H. B., Dunitz, J. D., Shefter, E. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group.*J. Am. Chem. Soc.* **95**, 5065-5067 (1973).
52. Nozaki, Y., Katayama, N., Harada, S., Ono, H., Okazaki, H. Lactivicin, a naturally occurring non-beta-lactam antibiotic having beta-lactam-like action: biological activities and mode of action. *J. Antibiot*. 42, 84-93 (1989).
53. Bruice, T. C., Benkovic, S. J. Chemical Basis for Enzyme Catalysis. *Biochemistry* **39**, 6267-6274 (2000).
54. Wilmouth, R. C., Clifton, I. J., Robinson, C. V., Roach, P. L., Aplin, R. T., Westwood, N. J., Hajdu, J., Schofield, C. J. Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. *Nat. Struct. Biol.* **4**, 456-462 (1997).
55. June, C. M., Vallier, B. C., Bonomo, R. A., Leonard, D. A., Powers, R. A. Structural origins of oxacillinase specificity in class D β-lactamases *Antimicrob. Agents Chemother.* **58**, 333-341 (2014).
56. Merrick, J. P., Moran, D., Radom. L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **111**, 11683-11700 (2007).
57. Chai, J.-D., Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys*. **128**, 084106 (2008).
58. Cremer, D. Density functional theory: coverage of dynamic and non-dynamic electron correlation effects. *Mol. Phys.* **99**, 1899-1940 (2001).
59. Kim, K., Jordan. K. D. Comparison of Density Functional and MP2 Calculations on the Water Monomer and Dimer. *J. Phys. Chem.* **98**, 10089-10094 (1994).
60. DeChancie, J., Clemente, F. R., Smith, A. J. T., Gunaydin, H., Zhao, Y.-L., Zhang, X., Houk, K.N. How similar are enzyme active site geometries derived from quantum mechanical theozymes to crystal structures of enzyme-inhibitor complexes? Implications for enzyme design. *Protein Sci*. **16**, 1851-1866 (2007).
61. Chung, I. Y. W., Paetzel, M. Crystal structures of yellow tail ascites virus VP4 protease: trapping an internal cleavage site trans acyl-enzyme complex in a native Ser/Lys dyad active site. *J. Biol. Chem.* **288**, 13068-13081 (2013).
62. Shi, Q., Meroueh, S. O., Fisher, J. F., Mobashery, S. Investigation of the Mechanism of the Cell Wall dd-Carboxypeptidase Reaction of Penicillin-Binding Protein 5 of Escherichia coli by Quantum Mechanics/Molecular Mechanics Calculations. *J. Am. Chem. Soc.* **130**, 9293-9303 (2008).
63. Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M., Kim, B., Little, J. W., Strynadka, N. C. J. A Conformational Switch for Regulation of Self-Cleavage. *Cel*, **106**, 585-594 (2001).
64. Paetzel, M., Dalbey, R. E., Strynadka, N. C. J. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature* **396**, 186-190 (1998).
65. Strynadka, N. C. J., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., James, M. N. G. Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution. *Nature* **359**, 700-705 (1992).
66. Imming, P., Klar, B., Dix, D. Hydrolytic stability versus ring size in lactams: implications for the development of lactam antibiotics and other serine protease inhibitors. *J. Med. Chem*. **43**, 4328-4331 (2000).

Chapter 3

A Human Protein Hydroxylase that Accepts D-Residues

3.1. Introduction

Hydroxylations of biopolymers as catalysed by 2-oxoglutarate dependent oxygenases,[5-7] are important in a variety of biological processes including oxygen sensing,[1-4] chromatin regulation, biosynthesis, and RNA synthesis. Factor inhibiting hypoxia inducible factor (FIH) belongs to 2OG dependent dioxygenase superfamily and contains a non-heme ferrous iron at its acive site.[8] A characteristic feature that discriminates FIH from other studied oxygenases lies in its promiscuity in catalysing the hydroxylation of various amino acid residues.[9-11] FIH was first identified as an asparaginyl C-3 hydroxylase with catalytic activity on hypoxia inducible factor (HIF); such activity reduces the transcriptional activity of HIF to the extent to be physiologically relevant in the hypoxic response.[12] During the FIH catalyzed hydroxylation of HIF, FIH consumes molecular oxygen and 2OG produces CO_2 and succinate. FIH also catalyses the hydroxylation of multiple ankyrin repeat domain residues at the C-3 positions. Notably, in the case of these substrates, FIH can act on not only Asn-residues, but also an Asp- and His-residues.[13,14]

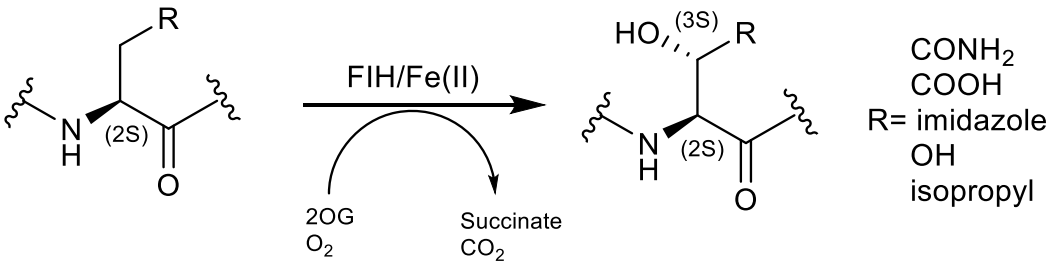


Figure 3.1. FIH is a 2OG oxygenase catalyzing the post-translational hydroxylation of proteins including the hypoxia-inducible factor (HIF; R=$CONH_2$) and ankyrin repeat domain proteins (R=$CONH_2$, COOH, OH, imidazole and isopropyl).

Furthermore, recent studies have shown that FIH can catalyse the C-3 hydroxylation of completely hydrophobic side chains, such as L-leucine to form products with the same stereochemical properties as in polar amino acids (Figure 3.1).[10] The results suggest that

FIH engineering to catalyze specific types of residue oxidation to enable targeted post-translational modification is of interest.

Mechanistic investigations into the 2OG oxygenases and closely related enzymes are briefly summarized in Figure 3.2. In essence, catalysis by 2OG oxygenases involves the oxidative decarboxylation of 2OG by reaction with dioxygen to generate carbon dioxide, succinate and a reactive oxidizing species that mediates substrate oxidation. One of the oxygen atoms from dioxygen is incorporated into succinate and, in the case of hydroxylation reactions, the other is incorporated, at least partially, into the hydroxyl group of the product.[15,16]

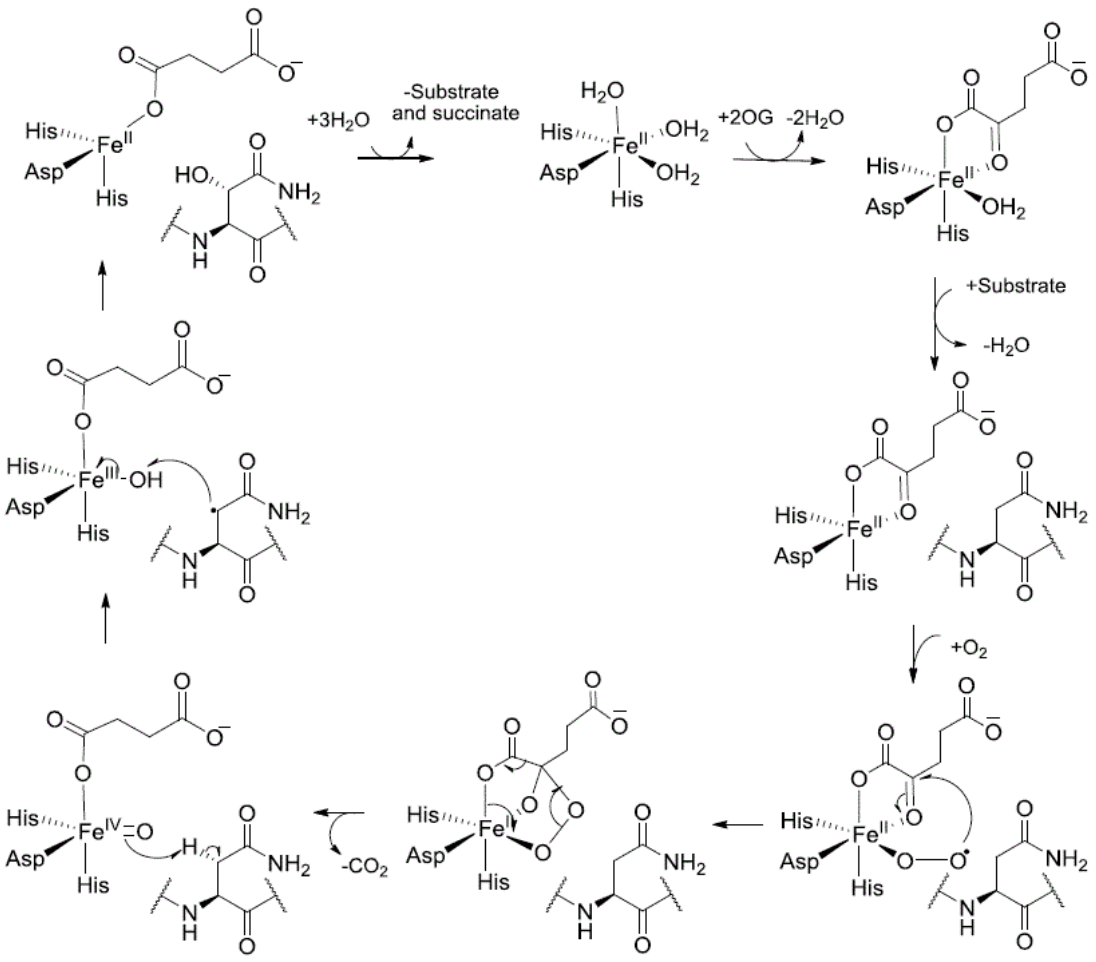


Figure 3.2. Consensus outline mechanism proposed for a hydroxylation reaction as catalyzed by the 2OG oxygenases.

Oxidative decarboxylation of 2OG yields CO_2, an Fe(IV)-oxo species (or ferryl intermediate) with bound succinate. The ferryl species abstracts a hydrogen atom from the primary substrate to generate an Fe(III)-OH species and a substrate radical. In the conventional mechanism, hydroxyl radical rebound yields the hydroxylated product, which with succinate, dissociates to complete the catalytic cycle.[16]

In this chapter I describe investigations on the structural and energetic features of the hydroxylation reactions of leucine in the active site of FIH by means of *ab initio* density functional theory (DFT) calculations. Extensive DFT calculations were carried out to determine the stationary-state structures on the calculated minimum energy paths of enzymatic reaction models of the hydroxylations of L- and D-leucine. The simplified model system for the FIH-catalysed reaction involves unmodified and hydroxylated leucine, 2OG, and the reactive oxygen bound to the central iron ion. Methylimidazole and acetate ion were employed in the structural model of the enzyme-substrate complex to respectively represent the two histidine residues (His199 and His279) and Asp201 that coordinate the active site iron ion. Interest was focused on addressing the stereoselectivity and the possibility of double hydroxylations in the catalytic action of FIH on L- and D-leucine (an experimental observation - see below). Catalytic mechanisms for L- and D-leucines were proposed by examining their complexation patterns in the active site of FIH and the structures of products generated in the hydroxylation reactions. Finally, the computational findings for the catalytic mechanism of the stereoselective hydroxylation by FIH are supported by experimental evidence obtained from mass spectrometric analysis by Helen Thorbjornsrud and X-ray crystal structures of FIH in complex with various peptide substrates and products by Drs. Michael McDonough and John Scotti.

This work described in this chapter concerns the possible hydroxylation of D-amino acid residue as catalysed by FIH. This work was inspired by the discovery that FIH can catalyse hydroxylation of hydrophobic residues. The potential hydroxylation of D-residues is of interest because of their increased prevalence in aged cell[30,31]and by the results with allylglycine containing substrates as described below.

3.2. Experimental methods

3.2.1. Computational Methods

3.2.1.1. Quantum chemical investigations of catalytic hydroxylations of D- and L-leucine in the active site of FIH.

The starting coordinates for *ab initio* DFT calculations were extracted from the crystal structure of FIH in complex with N-oxalylglycine (NOG) and an oligopeptide including a D-leucine residue (PDB entry: 4JAA and Schofield group unpublished data); NOG was used as the surrogate for the cosubstrate 2OG. To construct an appropriate enzymatic reaction model for quantum chemical calculations, the oligopeptide substrate was simplified to D-leucine (Figure 3.3) and N-oxalylglycine was modified to 2OG. In this structural model for the enzyme-substrate complex, methylimidazole and an acetate ion were used to mimic the side chains of His and Asp residues coordinated to the iron ion. One vacant coordination site was then occupied by the reactive oxygen.

As the starting point to study the catalytic hydroxylation of the substrate leucine residue, the catalytically active form of FIH including FIH, 2OG, and dioxygen (O_2) were prepared as shown in Figure 3.3. This structural model was established based on the catalytic mechanism of FIH as proposed by Chowdhury et al.,[15] which involves the attack of dioxygen on the carbonyl carbon of 2OG to form the reactive oxygen to activate the C-H bond of substrate leucine residue. Although the central none-heme iron resides as

the ferrous ion in the resting form of FIH, it is assumed to be oxidized to the ferric ion in the catalytically active form as suggested in the proposed catalytic mechanism. All geometries corresponding to the energy minima on the reaction path were fully optimized using the Gaussian09 program.[32] These geometry optimizations were carried out with the def2-TZVP basis set for all atoms. To calculate the electronic energies of the reactants and products in complex with the catalytic center of FIH, we employed the B3LYP hybrid density functional. This consists of Becke's hybrid gradient-corrected exchange functional[17] and the gradient-corrected correlation functional of Lee, Yang, and Parr.[18] The spin multiplicity of the central ferric ion was assumed to sextet because it coordinates six weak ligands. Based on the optimized structures on the reaction pathways, I also investigate the thermodynamic aspects of the FIH catalysed hydroxylation of D- and L-leucine residues by FIH.

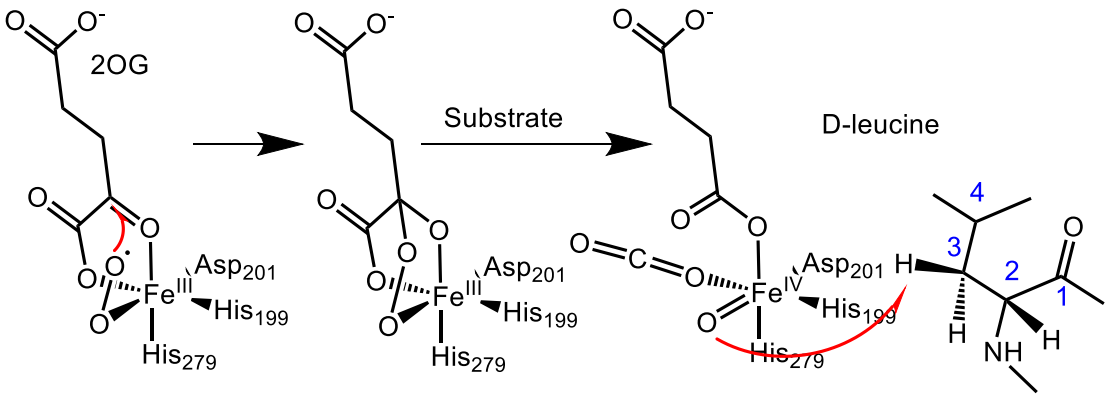


Figure 3.3. Schematic diagram for the formation of the D-leucine substrate oxidizing form of FIH. The number of the carbon atoms in D-leucine are indicated.

3.3. Results and Discussion

3.3.1. Initial L-/D-allylglycine hydroxylation

The promiscuity of FIH catalysed protein hydroxlyations was precedented by earlier work on 2OG oxygenases acting on small molecules, in particular those involved in β-lactam antibiotic biosynthesis.[16] In some cases, unsaturated substrate analogues were found to undergo rearrangements, likely (in some examples) via radical intermediates.[19-21] Inspired by these studies, [22] Dr. Wei Ge tested whether FIH can catalyse hydroxylation of an allylglycine residue, anticipating that hydroxylation may be more efficient than for leucine due to the weaker allylic C-H bond, or that a 1,3-allylic re-arrangement, or epoxide formation may occur, as precedented for 2OG oxygenases catalysing small molecule oxidation.[16] A potential peptide substrate for FIH with allylglycine at the potential hydroxylation position was therefore synthesized. For reasons of expediency, the peptide was initially made with an epimeric (D/L) mixture of allylglycine at the hydroxylation position. When incubated with FIH under standard conditions, it was found that the epimeric mixture reproducibly gave an approximate 50% conversion to a hydroxylated product; as evidenced by a +16 Da mass increase using matrix assisted laser desorption ionisation time of flight (MALDI-ToF) mass spectral analysis of the incubation product mixture. To investigate this further the corresponding peptide with L-allylglycine at the hydroxylation position was then made. Surprisingly, given the known hydroxylation substrate for FIH, this peptide did not undergo hydroxylation (<5%). In contrast, the subsequently synthesized D-allylglycine peptide underwent relatively efficient hydroxylation demonstrating that FIH can catalyse hydroxylation of D-residues and that its side chain selectivity for D- and L-residues differs substantially. To investigate further peptides with D-residues of common proteinogenic residues with a set of side chains including those of 'L-residue' FIH substrates, the results demonstrated that

D-Leu, D-His and D-allylglycine underwent hydroxylation with the following order of efficiency: D-Leu > D-His > D-allylglycine.

3.3.2. Unexpected dihydroxylation of D-leucine substrate by FIH

3.3.2.1. Experimental confirmation of dihydroxylation of D-leucine by FIH.

To examine the promiscuity of FIH in catalyzing the hydroxylation of D-residues, several D-form amino acid residues were substituted at the hydroxylation position of the known consensus ankyrin substrate (CA: sequence: HLEVVKLLLEHGADV**N**AQDK) peptide substrates in which the asparagine residue (bold) subject to FIH-catalysed modification was substituted with all 20 D-amino acids. Figure 3.4 shows the results of enzymatic hydroxylation reaction of D-leucine in CA peptide substrate at varying concentrations of FIH. Surprisingly, in the case of D- (but not L-) leucine not only a +16 Da but also a +32 Da mass shift was observed. Note that the peak at 2432.5 Da is reduced as the concentration of FIH increases from 0 to 10 and 50 μM due to its consumption. The second (+16 Da) and the third peaks (+32 Da) appear at 2448.4 and 2464.4 Da, respectively, only in the presence of FIH. The uniform spacing of 16 Da among the three peaks indicates the formation of both singly and doubly hydroxylated products. Driven by the fact that the availability of the second hydroxylation reaction depends on the stereochemical properties of the substrate, I performed DFT calculations and X-ray crystallographic analysis with Drs. Michael McDonough and John Scotti to investigate the detailed energetic and structural features of FIH-substrate complexes responsible for the hydroxylation state of the products.

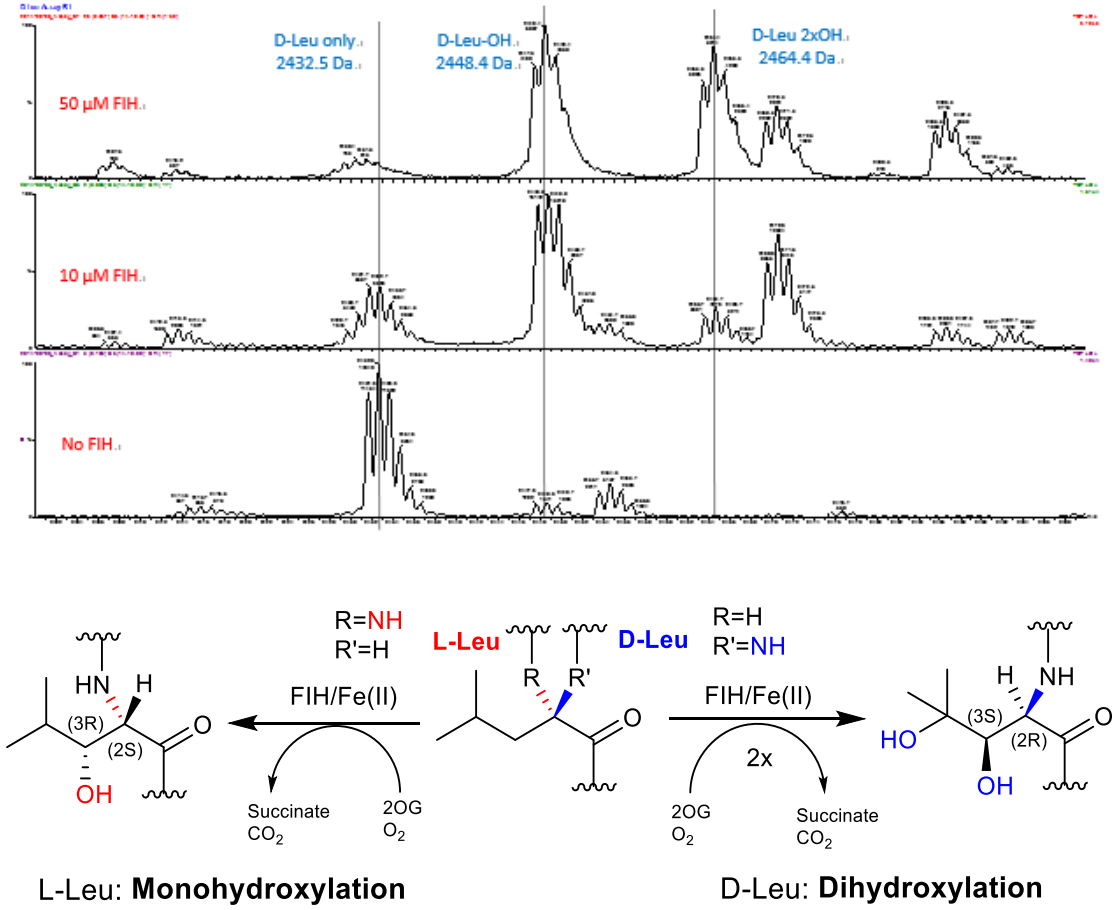


Figure 3.4. The single and double hydroxylations of L- and D-leucine, respectively, in the FIH-catalyzed hydroxylation of an ankyrin substrate. Mass spectrometric analysis of FIH catalyzed hydroxylation of D-leucine in CA (HLEVVKLLLEHGADV**X**AQDK) peptide substrate where **X** is the hydroxylated residue. Time course analysis implies sequential monohydroxlyation followed by dihydroxylation. The peaks were assigned through MS fragmentation and NMR analyses.

3.3.2.2. DFT investigations of the single hydroxylation reaction by the catalytic unit of FIH.

To address the energetic and structural features in the catalytic hydroxylation of D- and L-leucine in the active site of FIH, the reaction pathways were explored with density functional calculations at the B3LYP/def2-TZVP level of theory using the simplified model systems described in the previous section. Figure 3.5(b) shows the optimized structures of the catalytic unit in complex with D-leucine (**SDR**), the hydroxylated product bound in the active site (**SDP**), and the transition state (**SDTS**) encountered along

the reaction path. In the reactive complex **SDR**, the reactive oxygen (O1) coordinated to the ferric ion resides in proximity to the C_β atom of D-leucine (C3 in Figure 3.5) at 2.19 Å distance to the *pro-S* hydrogen, the substitution of which would generate the *S* enantiomer. In comparison, *pro-R* hydrogen in **SDR** points outward from the active site at the distance of 3.49 Å from O1. Due to the closer proximity of *pro-S* hydrogen to O1 than the *pro-R* counterpart, the (*S*)-3-hydroxy product seems to be produced in preference to the *R* enantiomer.

The catalytic hydroxylation reaction started from **SDR** with the approach of O1 to C3, which would be accompanied by the proton transfer from C3 by O1. When the O1···C3 distance contracts from 3.13 Å in **SDR** to 2.50 Å, the reacting system reaches the transition state (**SDTS**) in which the *pro-S* hydrogen is partially transferred to O1. During the reaction, the formation of O1–C3 bond lags behind the proton transfer in the context that the associated O···H distance decrease significantly from 2.19 Å in **SDR** to 1.15 Å in **SDTS**, which corresponds to 94% advancement toward the formation of the reaction product (**SDP**). The single imaginary frequency for **SDTS** is dominated not only by the O1–C3 bond forming motion but also by the C3-H bond breaking motion, indicating that the proton transfer from C3 to O1 is necessary to cross the activation barrier for the hydroxylation reaction. Despite the extensive search for the potential energy surface, neither carbanionic nor radical intermediate was found along the reaction path. This disproves the stepwise mechanism involving the proton abstraction and the subsequent rebound steps observed in the C–H hydroxylation by the other mononuclear non-heme enzymes.[23] Thus, the FIH-catalysed hydroxylation step for D-leucine seems to proceed via a concerted mechanism.[24,25] The concurrent occurrence of proton transfer and O–C bond formation has also been proposed on the basis of experimental studies on the catalytic mechanism of various 2OG dependent oxygenases.[26-29]

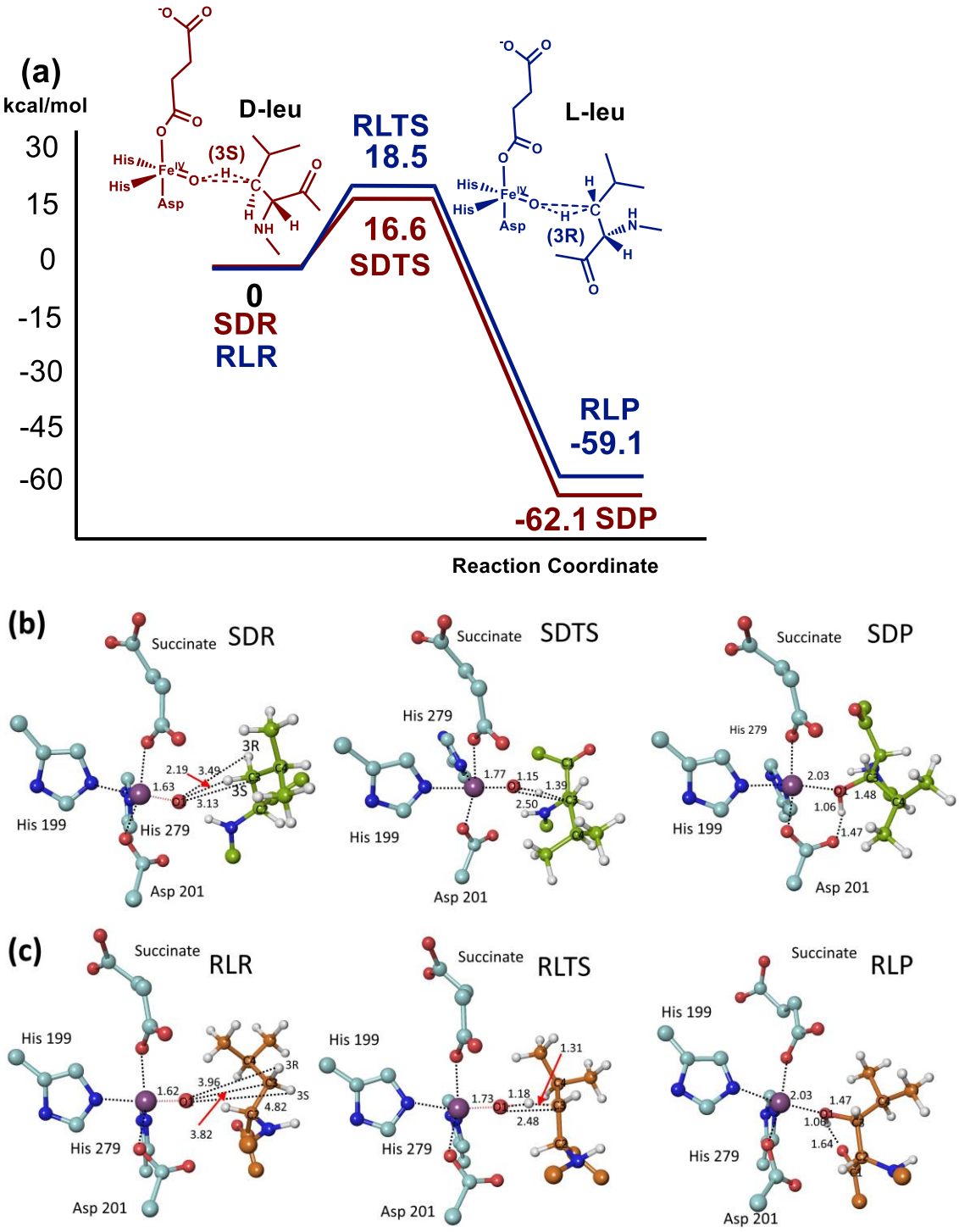


Figure 3.5. (a) Potential energy profile diagrams along the reaction coordinates of the FIH-catalysed hydroxylation of D-leucine (brown) and L-leucine (blue). The potential energies of stationary-state structures for the hydroxylation of D- and L-leucine are measured from **SDR** and **RLR**, respectively. (b) Optimized structures of energy minima and transition state for the single hydroxylation reaction at C-3 position of D-leucine in the active site of FIH. (c) Optimized structures of the energy minima and the transition state for the single hydroxylation reaction at C-3 position of L-leucine in the active site of FIH. Some selected interatomic distances are indicated in Å. Structures are colored as follows: Fe (blue purple), N(blue), O(red), C(turquoise(protein), substrate (D-leucine: yellowgreen, L-leucine: orange), H(white).

The complete formation of O1–C3 bond carries the reacting system to the other energy minimum (**SDP** in Figure 3.5(b)) that corresponds to the singly hydroxylated product in complex with the active site of FIH. As implicated by the different interatomic distances between O1 and the hydrogens attached to C3 in **SDR**, the *pro-S* hydrogen is fully transferred to O1 to form the *S* enantiomer of the hydroxylated D-leucine. In response to the protonation, the coordination distance from O1 to the central ferric ion increases from 1.63 Å in **SDR** to 2.03 Å in **SDP**. At **SDP**, the newly formed hydroxyl moiety points toward the side-chain carboxylate group of Asp201 to form a strong hydrogen bond at the distance of 1.47 Å. The establishment of such a strong hydrogen bond would serve as a driving force to facilitate the hydroxylation reaction.

Figure 3.5(c) shows the structures of the energy minima and the transition state encountered on the reaction coordinate of the hydroxylation of L-leucine in the active site of FIH. As the substrate changes from D- to L-leucine in the reactive complex involving the catalytic unit of FIH (**RLR**), a positional exchange can be seen between the hydrogen and the side-chain 2,2-dimethylpropyl group attached to the C_α atom (C2 in Figures 3.5). The characteristic feature that discriminates **RLR** from **SDR** lies in that O1 atom stays closer to *pro-R* hydrogen (3.96 Å) than *pro-S* hydrogen (4.82 Å). Judging from this binding configuration, the enantiomer of β-hydroxy-L-leucine is expected to be produced in the FIH-catalysed hydroxylation reaction rather than the *S* counterpart. It is also noteworthy that the distance between O1 and C3 atoms increases from 3.13 Å in **SDR** to 3.82 Å in **RLR**, which implies the higher activation barrier in the hydroxylation of L-leucine than in that of D-leucine.

At the reactive complex (**RLR**), the hydroxylation reaction proceeds with the approach of O1 toward C3. When the interatomic distance falls to 2.48 Å, the reacting system develops into the transition state (**RLTS** in Figure 3.5) in which the *pro-R*

hydrogen attached to C3 is partially transferred to O1. As can be inferred from the 96% advancement toward the formation of O–H bond in the (*R*)-β-hydroxy product complexed in the active site of FIH (**RLP** in Figure 3.5(c)), the transition state **RLTS** is late in terms of the proton transfer from C3 to O1. This is in contrast to only 57% completion of the O1–C3 bond formation. The main structural features of **RLTS** are thus similar to **SDTS**, indicating the difference in the activation barriers would stem from the difference in binding patterns of D- and L-leucine in the active site of FIH. The role of hydrogen bond acceptor with respect to the product hydroxyl group changes from Asp201 in **SDP** (Figure 3.5(b)) to the backbone carbonyl oxygen in **RLP** (Figure 3.5(c)) in accordance with the configurational change at C2.

Figure 3.5(a) displays the potential energy profiles along the reaction paths depicted in Figures 3.5(b) and 3.5(c). The stereoselective hydroxylation reactions of D- and L-leucine seem to be thermodynamically favorable in the active site of FIH because **SDP** and **RLP** are approximately 60 kcal/mol lower in potential energy than **SDR** and **RLR**, respectively. These large stabilizations can be attributed to the formation of the two new covalent bonds involving O1 atom at the expense of one C–H bond. The activation barrier for the hydroxylation of D-leucine appears to be lower than that of L-leucine by 1.9 kcal/mol. The increase in activation energy with the substrate change from D- to L-leucine is actually not surprising because O1 resides more distant in **RLR** (Figure 3.5(c)) than in **SDR**. Judging the small difference in the activation barriers, however, both D- and L-leucine are expected to be oxidized to the respective (*S*)-β-hydroxy-D-leucine and (*R*)-β-hydroxy-L-leucine in the active site of FIH, respectively.

3.3.2.3. DFT calculations on the second hydroxylation reactions in the active site of FIH.

I next turned to addressing the possibility of the successive hydroxylation of the singly hydroxylated D- and L-leucine in the active site of FIH. To prepare the starting structures of these second hydroxylation reactions, the geometries of the hydroxylated D- and L-leucine in complex with the catalytic unit of FIH were fully optimized at the B3LYP/def2-TZVP level of theory. The structural model for the reactive FIH-D-leucine complex (**SDR2**) is shown in Figure 3.6(b), together with the doubly hydroxylated product bound to the active site (**SDP2**) and the transition state (**SDTS2**) connecting the two energy minima.

In **SDR2**, the hydrogen attached to C_γ atom (C4 in Figure 3.6) points toward to the reaction center, which would have the effect of facilitating the attack of O1 to C4. Although O1 stays a little closer to C3 (3.09 Å) than C4 (3.17 Å) in **SDR2**, the possibility of double hydroxylation at C3 position was excluded in this study due to the thermodynamic instability of the resulting germinal diol product (a ketone was not observed experimentally (Schofield group, unpublished results)). Furthermore, the hydrogen atom attached to C3 resides in the 'opposite' direction to O1 in **SDR2**, which would be inappropriate for the formation of new O1–C3 bond that requires hydrogen abstraction from C3. A vicinal diol is thus expected to be produced in the second hydroxylation reaction to introduce a hydroxyl moiety at C4 atom of (*S*)-β-hydroxy-D-leucine.

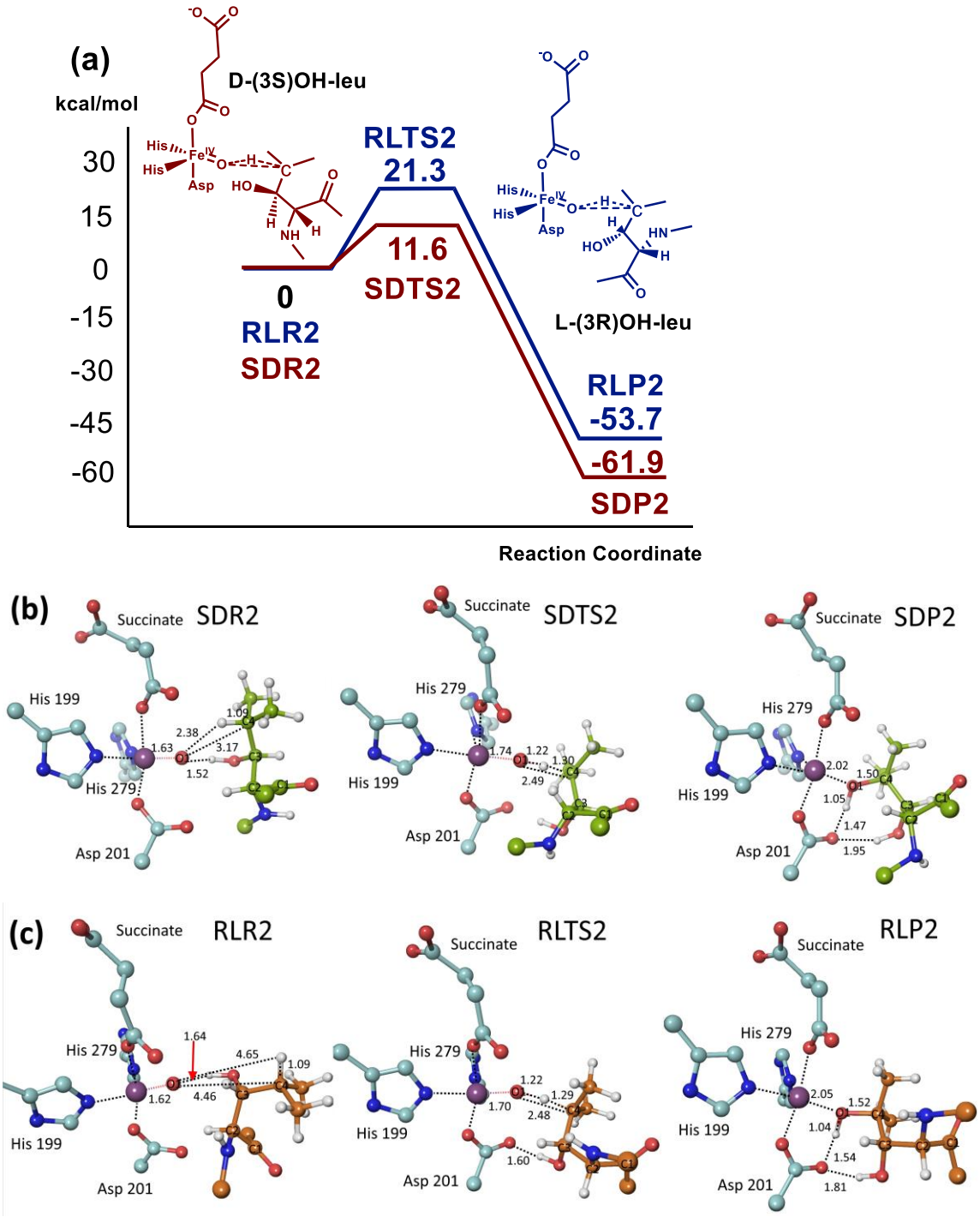


Figure 3.6. (a) Potential energy profile diagrams along the reaction coordinates of the FIH-catalysed second hydroxylation of (*3S*)-β-hydroxy-D-leucine (brown) and (*3R*)-β-hydroxy-L-leucine (blue). The potential energies of stationary-state structures on the reaction coordinates are measured from the respective reactant complexes (**SDR2** and **RLR2**). (b) Optimized structures of energy minima and the transition state for the second hydroxylation reaction of (*3S*)-β-hydroxy-D-leucine in the active site of FIH. (c) Optimized structures of the energy minima and the transition state for the second hydroxylation reaction of (*3R*)-β-hydroxy-L-leucine in the active site of FIH. Structures are colored as follows: Fe (blue purple), N(blue), O(red), C(turquoise(protein), substrate (D-leucine: yellowgreen, L-leucine: orange), H(white). Some selected interatomic distances are indicated in Å.

When the C4–H bond increases from 1.09 Å in **SDR2** to 1.30 Å, the enzymatic reaction model is optimized to the transition state (**SDTS2**) that is late in terms of the O1–H bond formation. During this structural change, the hydrogen-bond acceptor with respect to the substrate hydroxyl group changes from O1 to a side-chain carboxylate oxygen of Asp201. The contributions to the single imaginary frequency inherent in **SDTS2** come from the proton transfer from C4 to O1 as well as from the O1–C4 bond forming motion. This implies that the proton abstraction and the O1–C4 bond formation occur in the concerted fashion as in the single hydroxylations of D- and L-leucine. The O1···H distance contracts from 2.38 to 1.22 Å in going from **SDR2** to **SDTS2**, which corresponds to 87% completion of the O1–H bond toward the formation of **SDP2**. In contrast to this significant advancement, O1 approaches C4 only by 0.69 Å during the formation of the transition state to make only 41% progress to generate the β,γ-dihydroxy-D-leucine product. **SDTS2** is thus similar to **SDTS** (Figure 3.5(b)) and **RLTS** (Figure 3.5(c)) in that the O–C bond formation lags behind the proton transfer.

As a consequence of complete proton transfer from C4 to O1, the reacting system converges to the second energy minimum (**SDR2**) in which the β,γ-dihydroxy-D-leucine product is complexed at the catalytic center of FIH. In this complex, both hydroxyl moieties donate hydrogen bond with the side-chain carboxylate ion of Asp201 in the bifurcated form. Apparently, these two hydrogen bonds would have the effect of making the formation of the β,γ-dihydroxy-D-leucine product thermodynamically more favorable.

As the starting point for the second hydroxylation of L-leucine, we also optimized the structure of the reactive complex (**RLR2**) established between (*3R*)-β-hydroxy-L-leucine and the simplified model for the catalytic unit of FIH involving the reactive

oxygen (O1). As shown in Figure 3.4, **RLR2** differs from **SDR2** in that the C4 atom resides distant from the catalytic center of FIH. The C4···O1 distance amounts to 4.46 Å in **RLR2** as compared to 3.17 Å in **SDR2** (Figure 3.4). This stems from the increase in the steric repulsion between 2-OG and the methyl groups attached to C4 due to the change of the substrate from (*3S*)-β-hydroxy-D-leucine to (*3R*)-β-hydroxy-L-leucine. The appearance of such a catalytically unfavorable configuration in **RLR2** indicates that the second hydroxylation of (*3R*)-β-hydroxy-L-leucine in the active site of FIH would be less efficient than that of (*3S*)-β-hydroxy-D-leucine.

From comparing the structures of **SDR2** and **RLR2** (Figure 3.6), it clearly follows that the substrate hydroxyl group introduced at C3 plays the role of hydrogen-bond donor with respect to O1 in both reactive complexes. The O–H···O hydrogen bond in **SDR2** contributes to sustaining the reactive conformation of (*3S*)-β-hydroxy-D-leucine in which the hydrogen atom attached to C4 resides in close proximity (2.38 Å) to O1. On the other hand, the corresponding hydrogen bond in **RLR2** apparently serves as an obstacle to achieving the reactive configuration in which the proximity of O1 to the C4–H bond is necessary for hydrogen abstraction and the simultaneous O1–C4 bond formation. The conformational restraint caused by establishing a strong hydrogen bond with O1 is thus likely to make it difficult for FIH to catalyse the second hydroxylation reaction of (*3R*)-β-hydroxy-L-leucine

At **RLR2**, the second hydroxylation reaction proceeds with the approach of O1 to C4. The shortening of O1···C4 distance to 2.48 Å leads to the formation of the transition state (**RLTS2** in Figure 3.6(c)) in which the hydrogen attached to C4 is partially transferred to O1. In contrast to the conformational difference between (*3R*)-β-hydroxy-L-leucine in **RLR2** and (*3S*)-β-hydroxy-D-leucine in **SDR2**, **RLTS2** appears to be very

similar to **SDTS2** in the imbalance between the proton transfer and the O1–C4 bond formation as well as in the establishment of a hydrogen bond between the substrate hydroxyl moiety and Asp201. The complete formation of O1–C4 bond carries the reacting system to the final energy minimum (**RLP2** in Figure 3.6(c)) that corresponds to the β,γ-dihydroxy-L-leucine product in complex with the active site of FIH. **RLP2** is also similar to **SDP2** in that the doubly hydroxylated product is stabilized with the two hydrogen bonds with the side-chain carboxylate group of Asp201 in the bifurcated form.

Figure 3.6(a) compares the potential energy profile diagrams along the reaction coordinates of the FIH-catalysed second hydroxylation of (*3S*)-β-hydroxy-D-leucine and (*3R*)-β-hydroxy-L-leucine. The β,γ-dihydroxy-D-leucine appears to be produced in complex with the catalytic center of FIH (**SDP2**) with the potential energy of 61.9 kcal/mol above the reactant complex (**SDR2**), which indicates that the second hydroxylation reaction of (*3S*)-β-hydroxy-D-leucine is relatively thermodynamically favourable. The activation barrier of 11.6 kcal/mol can be overcome to form the product via the transition state (**SDTS2**). It is interesting to compare the calculated energetics for the second hydroxylation of (*3S*)-β-hydroxy-D-leucine with those of (*3R*)-β-hydroxy-L-leucine. The activation barrier for the formation of the vicinal diol product from (*3S*)-β-hydroxy-D-leucine is 9.7 kcal/mol lower than that from (*3R*)-β-hydroxy-L-leucine. As can be inferred by comparing the energy difference between **SDR2** and **SDP2** to that between **RLR2** and **RLP2**, furthermore, the formation of β,γ-dihydroxy-D-leucine seems to be thermodynamically more favored than β,γ-dihydroxy-L-leucine by 8.2 kcal/mol. On the basis of kinetic and thermodynamic aspects deduced from *ab initio* DFT calculations, it can be argued that the second hydroxylation reaction is available only for D-leucine in the active site of FIH. Despite the large difference in the activation barriers and the reaction energies, the structural features around the reacting centres of **SDTS2**

and **SDP2** are very similar to those of **RLTS2** and **RLP2**, respectively. For example, the differences between the corresponding interatomic distances involving O1 fall within 0.04 Å (Figures 3.5 and 3.6). On the other hand, O1···C4 distance increases significantly from 3.17 to 4.46 Å as the substrate changes from (*3S*)-β-hydroxy-D-leucine to (*3R*)-β-hydroxy-L-leucine. As indicated in the structures of **SDR2** and **RLR2**, the substituted hydroxyl group plays the role of hydrogen-bond donor with respect to O1 in both reactive complexes (Figures 3.6). The O–H···O hydrogen bond in **SDR2** contributes to sustaining the reactive conformation of (*3S*)-β-hydroxy-D-leucine in which the vicinal hydrogen atom points toward the catalytic center of FIH. Although a similar hydrogen bond is also established in **RLR2**, the substrate cannot be placed in such a catalytically favorable conformation because the positions of the hydrogen atom and the isopropyl group attached to C3 have to be exchanged to retain the *R* configuration (Figures 3.6(c)). The difference in activation barriers and reaction energies in the second hydroxylations of (*3S*)-β-hydroxy-D-leucine and (*3R*)-β-hydroxy-L-leucine can thus be attributed to exchanged *pro-R* and *pro-S* configuration and the spin-off effect of their different hydroxylation positions in the active site of FIH.

3.3.2.4. Amino acid analyses of stereochemistry of the hydroxylated D-leucine residue.

The results of described computational studies on the simplified enzymatic reaction model for FIH indicate the production of *(3S)*-stereoisomers with a preference for *(3R)* counterpart in the single hydroxylation of D-leucine. To validate this hypothesis, the two possible enantiomers of the singly hydroxylated D-leucine were prepared by Dr. Marios Markoulides and amino acid analyses were performed by Dr. Adam Hardy for comparison. A mixture of racemic: (*R*)-β-hydroxy-D-leucine (*2R,3R*) with (*S*)-β-

hydroxy-L-leucine (*2S,3S*) and (*S*)-β-hydroxy-D-leucine (*2R,3S*) with (*R*)-β-hydroxy-L-leucine (*2S,3R*) were synthesized, respectively.

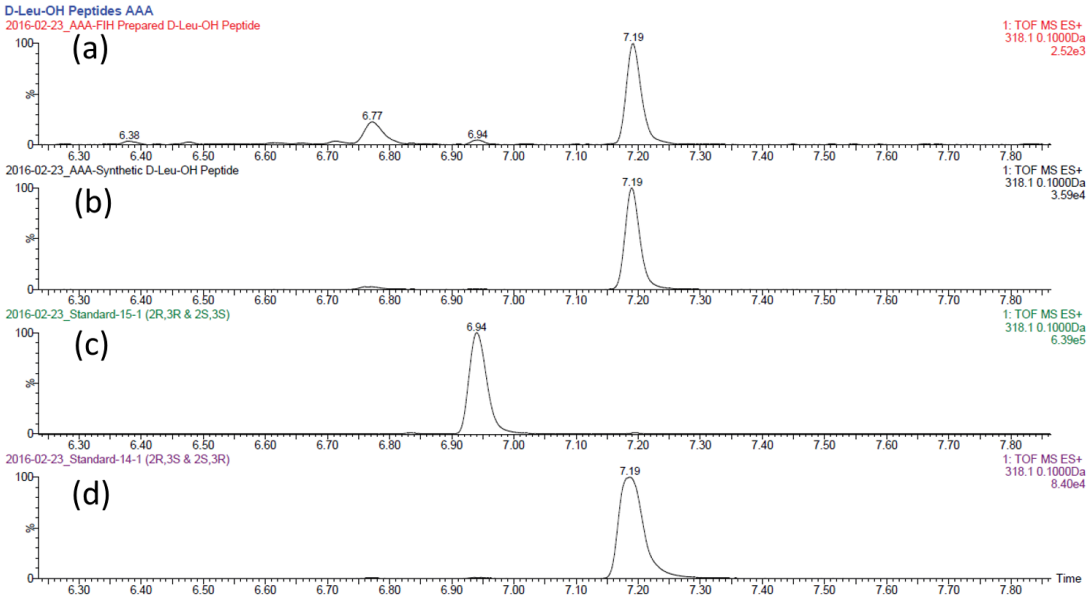
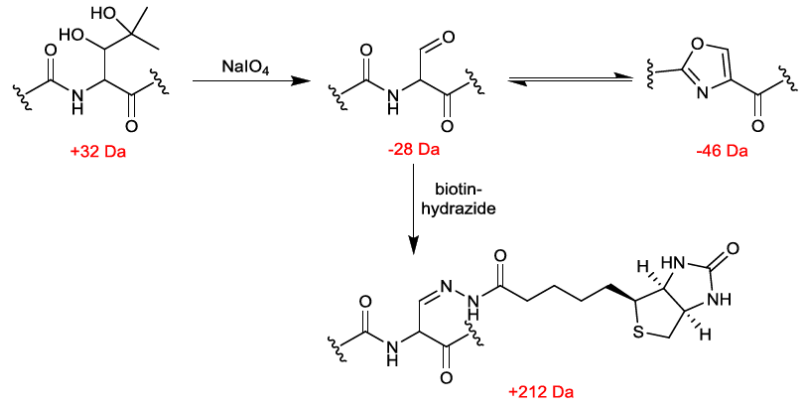


Figure 3.7. Amino acid analysis of the CA_D-leucine peptide after incubation with FIH demonstrates that (*S*)-β-hydroxy-D-leucine (*2R,3S*) is formed. Traces correspond to: (a) β-hydroxy-D-leucine peptide incubated with FIH and co-factors (retention time 7.19 min) and (b) β-hydroxy-D-leucine synthetic peptide (retention time 7.19 min). (c) Pure *racemic* (*R*)-β-hydroxy-D-leucine (*2R,3R*)/(*S*)-β-hydroxy-L-leucine (*2S,3S*) mixture (retention time 6.94 min); (d) Pure *racemic* (*S*)-β-hydroxy-D-leucine (*2R,3S*)/(*R*)-β-hydroxy-L-leucine (*2S,3R*) mixture (retention time 7.19 min).

As shown in Figure 3.7, only the mixture containing (*3S*)-β-hydroxy-D-leucine has the same peak at a retention time 7.19 for the FIH-produced and synthetic peptides with a singly hydroxylated D-leucine. In contrast, (*3R*)-β-hydroxy-D-leucine has a different peak at 6.94 retention time. The results provide experimental evidence for the enantioselective formation of (*S*)-β-hydroxy-D-leucine in the catalytic hydroxylation of D-leucine by FIH. Amino acid analysis of the D-leucine peptide after incubation with FIH demonstrates that (*S*)-β-hydroxy-D-leucine (*2R,3S*) is formed. The traces in Figure 3.7 correspond to the mass chromatograms (m/z 318.1) of the hydroxylated peptides produced by FIH catalysis and the synthetic peptides with the singly hydroxylated D-

leucine clearly show the consensus of the stereochemical preference at the hydroxylated C_β atom of D-leucine.

3.3.2.5. Chemical determination of diol



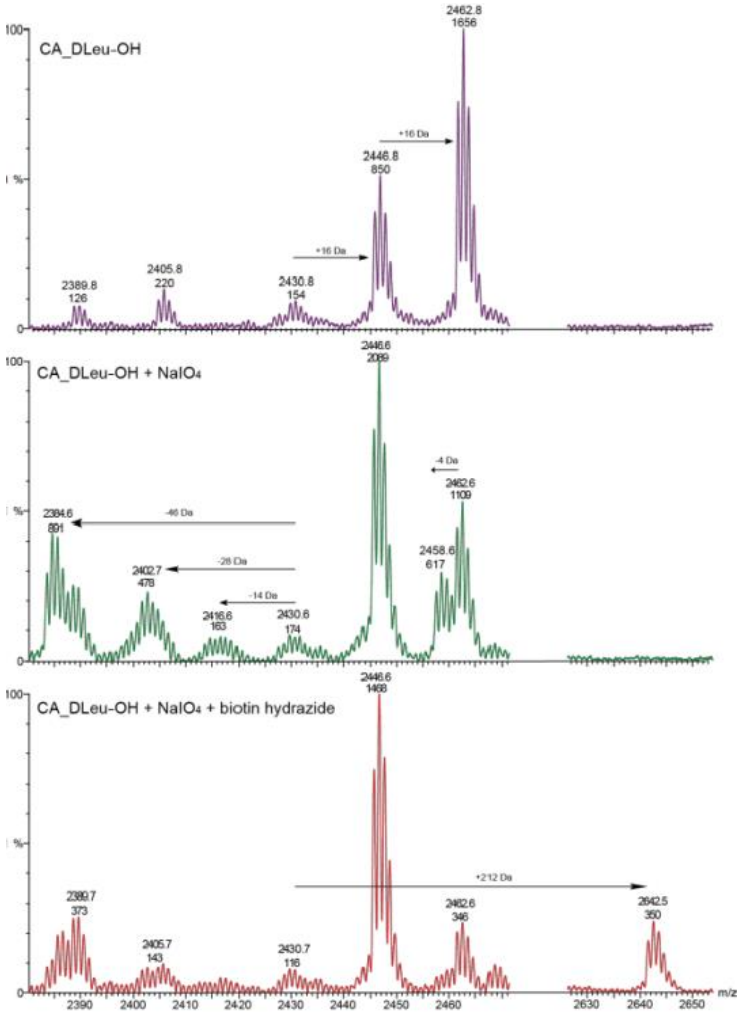


Figure 3.8. Treatment of dihydroxylated CA_D-Leu with excess $NaIO_4$ likely results in oxidative cleavage, followed by oxazole formation. The intermediate aldehyde may be trapped as a hydra-zone with biotin hydrazide. MALDI-TOF MS data for reaction of NaIO4 with CA_D-Leu-OH, followed by addition of biotin hydrazide.

The results of the incubation of purified D-Leu-diol peptide with $NaIO_4$ suggested that a 1,2 diol is produced. $NaIO_4$ treatment should cleave a vicinal diol and produce a peak which is -28Da (2402 Da) relative to the D-Leu peptide peak. Reaction of the intermediate aldehyde produced a -46 Da (2384 Da) peak (Figure 3.8). The result of biotin hydrazide treatment supports production of an aldehyde as a +212 Da peak was observed (2642 Da). This work was performed by Philip Gerken. The experimental detection of aldehyde by $NaIO_4$ treatment provides additional evidence for the computational prediction for the production of β,γ-dihydroxy-D-leucine (Figure 3.3) due to the double hydroxylation of D-leucine in the active site of FIH.

3.3.3. Crystallography of FIH with substrates

3.3.3.1. Overall structures of FIH in complex with D-leucine, D-β-(2R,3S)-hydroxyleucine, and D-allylglycine CAP

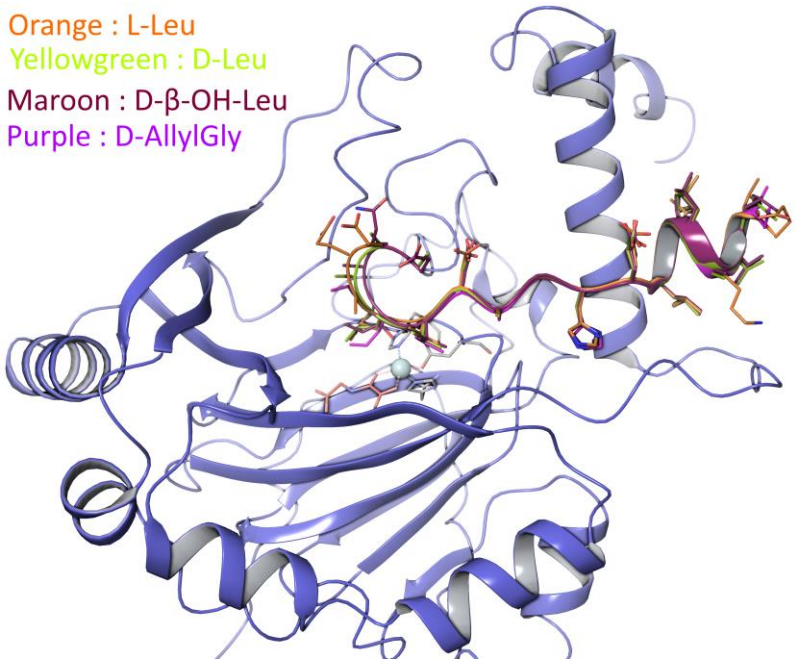


Figure 3.9. FIH binds peptides with different hydroxylated residues in a similar manner. Ribbon representation of superimposed peptide substrates when bound to FIH (violet); orange, L-leucine (PDB: 4B7E); yellowgreen, D-leucine (PDB: 4JAA); maroon, D-β-(2R,3S)-hydroxyleucine (unpublished PDB); and turquoise sphere, Fe.

To investigate the structural basis on the differential selectivity of FIH for D- over L-residues, we attempted X-ray crystallographic analysis of FIH complexed with D-residue containing substrates (collaboration with Dr. Michael McDonough and John Scotti). Views form structures of FIH in complex with L-leucine, D-leucine and D-β-(2*R*,3*S*)-hydroxyleucine and D-allylglycine CA peptide which were obtained are shown in Figure 3.9.

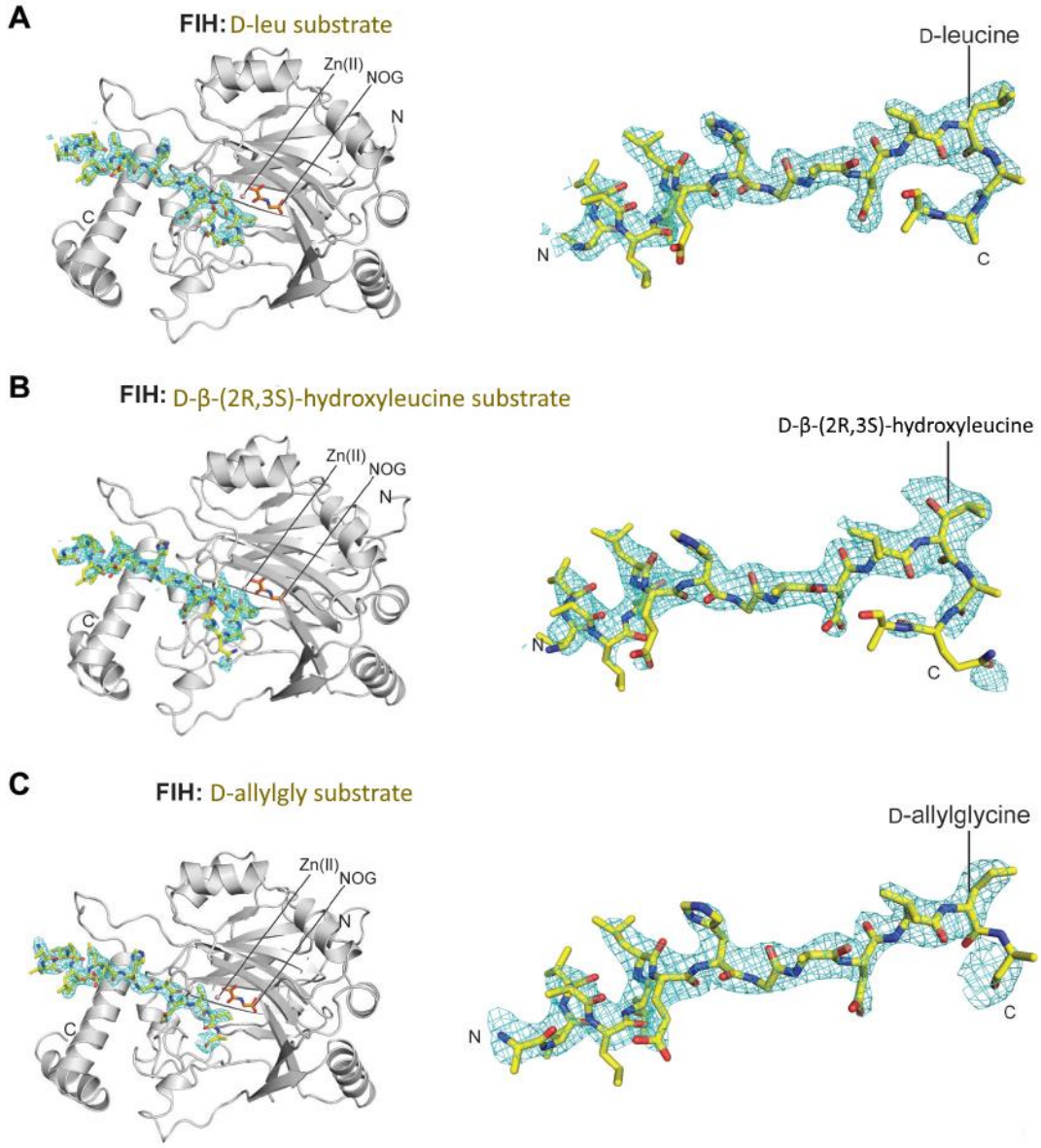


Figure 3.10. Views from crystal structures of FIH in complex with a CA peptide containing an unnatural D-amino acid. The overall structure of FIH in complex with (**A**) CA (D-leucine), (**B**) CA (D-β-(2*R* ,3*S*)-hydroxyleucine), and (**C**) CA (D-allyglycine); the electron density OMIT $|mF_o - DF_c|$ (shown in cyan mesh) are contoured to 3.0 σ.

The overall structures of FIH in complex with D-leucine, D-β-(2*R*,3*S*)-hydroxyleucine, and D-allylglycine CAP revealed that the peptide main chains bind to FIH in a similar manner (Figure 3.10.A). Importantly, in the case of the FIH with D-β-(2*R*,3*S*)-hydroxyleucine substrate, the 2.2 Å resolution dataset was sufficient to clearly view the β-hydroxyl group of the active site D-leucine within the electron density map (Figure 3.10.*B*).

3.3.3.2. Crystallographic analysis of bifurcating monohydroxylation vs. dihydroxylation by stereochemical D-/L-leucine

To rationalise the different availabilities of double hydroxylation between L- and D-leucine, I compared the conformations of L- and D-leucine in the active site of FIH as revealed by the X-ray crystal structures (Figure 3.10). As shown in Figure 3.11(a), the crystallographic analyses show that the backbone amide carbonyl groups of L- and D-leucine are directed in opposite directions (L-/D- main chain direction of substrate) due to the different stereochemistry at the C_α atom, which also leads to the different positioning of the side chains and to the change of loop structure of CA peptide substrates.

The torsional motion of the isopropyl group of L-leucine appears restricted because it establishes a close van der Waals contacts with the side chains of Tyr102, His199, and Gln239 at the top of active site in Figure 3.11(b). On the other hand, the side-chain isopropyl moiety of D-leucine points toward bulk solvent without significant interactions with nonpolar residues of FIH. Because the backbone amide carbonyl oxygen of D-leucine and that of the neighboring Ala804 of CA peptide substrate appear to repel the side chain of Tyr102 away from the active site due to steric clash, the isopropyl group of D-leucine should be able to rotate freely around in active site, such that the hydrogen atom bonded to C_γ atom can point toward the iron. These observations are consistent with

the computational predictions. Crystallography shows that the side chain of L-leucine is confined by side chain rings of His199 and Tyr102 (Figure 3.11(b)), However, the wide loop turn of main chain of D-leucine substrate can open 'the Tyr102 ring gate' to release the side chain of D-leucine. (Figure 3.11(a)). Thus, the side chain (2,2-dimethylpropyl) of L-/D-leucine preferentially adopt different conformations (sterically confined vs. freely rotatable due to sharp turn vs. wide turn) to open the route for different possibilities of FIH catalysed hydroxylation.

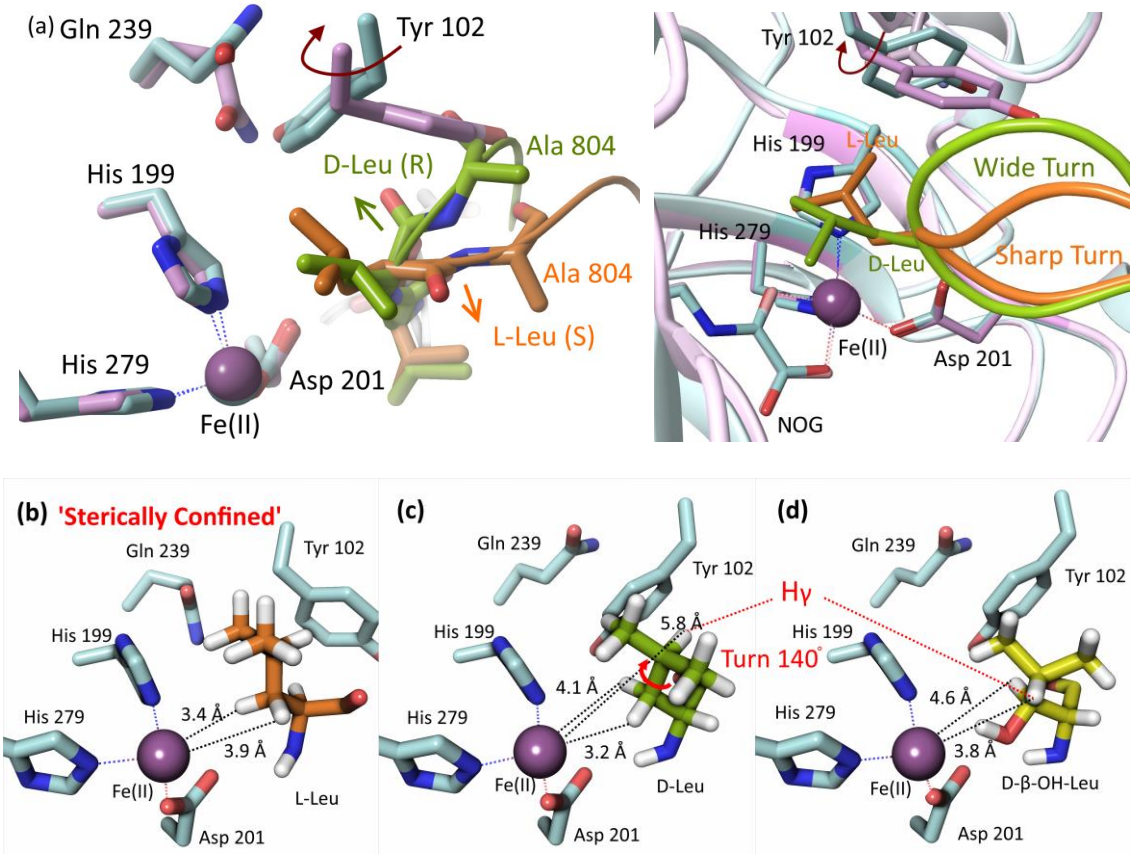


Figure 3.11. Superposition of X-ray crystal structures of FIH in complex with CA peptides containing (a) L-leucine (orange) and D-Leucine (yellowgreen). Carbon atoms of FIH in complex with L- and D-leucine are indicated in pink and cyan, respectively; (b) L-leucine, (c) D-leucine, and (d) D-β-(2*R*,3*S*)-hydroxyleucine in CA peptides.

As shown in Figure 3.11(b), the product of the first hydroxylation of L-leucine is predicted to be L-β-(2*S*,3*R*)-hydroxyleucine, because the *pro*-3*R* hydrogen resides closer to the central ferrous ion (3.4 Å) than pro-3*S* hydrogen (3.9 Å).[10] However, the positions

of the *pro*-3*R* and *pro*-3*S* hydrogens are "exchanged" in D-leucine (relative to L-leucine) with the distances of 4.1 and 3.2 Å from the ferrous ion (Figure 3.11(c)), respectively, due to the change in the stereochemistry of the C_α atom. This structural feature supports the possibility of producing D-β-(2*R*,3*S*)-hydroxyleucine in the first hydroxylation of D-leucine in the active site of FIH. Consistent with the structure of **SDP** in Figure 3.4, the substituted hydroxyl group establishes a hydrogen bond with the side-chain carboxylate ion of Asp199 in the structure of FIH in complex with D-β-(2*R*,3*S*)-hydroxyleucine (Figure 3.11(d)). The hydrogen bond of D-β-(2*R*,3*S*)-hydroxyleucine with the carboxylate of Asp199 would have the effect of sustaining the conformation of D-β-(2*R*,3*S*)-hydroxyleucine in the active site for the hydrogen attached to C_γ atom to point toward the iron at the distance of 3.8 Å. The proximity between C_γ hydrogen and the iron implicates the possibility of the additional hydroxylation of D-β-(2*R*,3*S*)-hydroxyleucine to form a 1,2-diol product as presented by hydroxylation of the ferryl-oxo oxygen as indicated in the section of 3.3.2.3. DFT calculations on the second hydroxylation reactions in the active site of FIH.

3.4. Conclusions

It is evident from X-ray crystallographic analyses of the structures of FIH in complex with various CA peptides that the C_α stereochemistry of leucine has a deterministic influence on the first and the second hydroxylation reactions due to the changes in the binding modes around the hydroxylation positions. The different hydroxylation outcome observed with D-leucine compared with L-leucine suggest that the reactivity and selectivity of FIH-catalyzed hydroxylation can be altered by site-directed mutagenesis. The biological relevance, if any, of the results on FIH-catalyzed D-residue hydroxylation is unknown; however, they imply HIF-α epimerisation is a plausible mechanism for reversible oxygen sensing. Epimerisation could be catalysed reversibly by oxygenases including FIH in such a way to regulate the activity of protein substrates. For example, a transcription factor protein could be synthesised to include an L-residue that is subjected to oxidative epimerisaion to a D-epimer with different transcriptional activity. The extent of epimerisation could determine the level of activity. This process could enable reversible oxygen / oxygenase dependent regulation of expression (note such as with the established hypoxia sensing mechanism of HIF-CAD hydroxylation.[1,2])

Although FIH can accept various amino acids as the substrate, it catalyses the hydroxylation reaction of leucines in a stereoselective fashion. By means of *ab initio* density functional calculations at the B3LYP/def2-TZVP level of theory, I addressed the structural and energetic features of the catalytic hydroxylations of L- and D-leucine using a simplified enzymatic reaction model for FIH. The results imply the formation of (*R*)-β-hydroxy-L-leucine and (*S*)-β-hydroxy-D-leucine from L- and D-leucine, respectively, in an exclusive manner. These predictions are validated by experimental observations. Only D-leucine is oxidized to β,γ-dihydroxy-D-leucine in the active site of FIH due to

freely rotatable side chain and favourable transition state energy barriers. The different complexation modes of L- and D-leucine in the active FIH site caused by the change in stereochemical property are responsible for such a large difference in the results of hydroxylation reactions. The computational findings for the enantioselective catalytic mechanisms of FIH were validated by the experimental evidences from mass spectrometric analyses and X-ray crystal structures of FIH in complex with various peptide substrates.

Chapter 3 References

1. Schofield, C.J.; Ratcliffe, P.J. Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* **5**, 343-354 (2004).
2. Semenza, G. L. HIF-1 and mechanisms of hypoxia sensing. *Curr. Opin. Cell Biol.* **13**, 167-171 (2001).
3. Kaelin, W.G. Jr., Ratcliffe, P.J. Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol Cell* **30**, 393–402 (2008).
4. Loenarz, C. et al. The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO Rep.* **12**, 63-70 (2011).
5. Schofield, C.J., Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722-731 (1999).
6. Kershaw, N.J., Caines,M.E.C., Sleeman, M.C., Schofield, C.J. The enzymology of clavam and carbapenem biosynthesis. *Chem. Commun.* **34**, 4251-4263 (2005).
7. Borowski, T., de Marothy, S., Broclawik, E., Schofield, C.J., Siegbahn, P.E.M. Mechanism for Cyclization Reaction by Clavaminic Acid Synthase. Insights from Modeling Studies. *Biochemistry* **46**, 3682–369 (2007).
8. Loenarz, C.; Schofield, C. J. Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. chem. Boil.* **4**, 152-156 (2008).
9. Cockman, M.E., Lancaster, D.E., Stolze, I.P., Hewitson, K.S., McDonough, M.A., Coleman, M.L., Coles, C.H., Yu, X., Hay, R.T., Ley, S.C., Pugh, C.W., Oldham, N. J., Masson, N., Schofield, C.J., Ratcliffe, P.J. Posttranslational hydroxylation of ankyrin repeats in IκB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc. Natl. Acad. Sci. USA* **103**, 14767-14772 (2006).
10. Yang, M., Hardy, A.P., Chowdhury, R., Loik, N.D., Scotti, J.S., McCullagh, J.S., Claridge, T.D., McDonough, M.A., Ge, W., Schofield, C.J. Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl.* **52**, 1700-1704 (2013).
11. Linke, S., Stojkoski, C., Kewley, R.J., Booker, G.W., Whitelaw, M.L., Peet, D.J. Substrate Requirements of the Oxygen-sensing Asparaginyl Hydroxylase Factor-inhibiting Hypoxia-inducible Factor. *J. Biol. Chem.* **279**, 14391-14397 (2004).
12. McDonough, M.A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Liénard, B.M.R., Zondlo J, Oldham NJ, Clifton IJ, Lewis J, *et al*. Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. USA* **103,** 9814 -9819 (2006).
13. Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinović, J., Edelmann, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., *et al*. Asparaginyl Hydroxylation of the Notch Ankyrin Repeat Domain by Factor Inhibiting Hypoxia-inducible Factor. *J. Biol. Chem.* **282**, 24027 - 24038 (2007).
14. Hardy, A.P., Prokes, I., Kelly, L., Campbell, I.D., Schofield, C.J. Asparaginyl β-Hydroxylation of Proteins Containing Ankyrin Repeat Domains Influences Their Stability and Function. *J. Mol. Biol.* **392**, 994-1006 (2009).
15. R. Chowdhury, A. Hardy, C. J. Schofield. The human oxygen sensing machinery and its manipulation. *Chem. Soc. Rev.* **37**, 1308–1319 (2008).
16. Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J. Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. inorg. Biochem.* **100**, 644-669 (2006).
17. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. chem. Phys.* **98**, 5648-5652 (1993).
18. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **37**, 785-789 (1988).
19. Clifton, I. J.; Doan, L. X.; Sleeman, M. C.; Topf, M.; Suzuki, H.; Wilmouth, R. C.; Schofield, C. J. Crystal structure of carbapenem synthase (CarC). *J. Biol. Chem.* **278**, 20843-20850 (2003).
20. Stapon, A., Li, R., & Townsend, C. A. (2003). Synthesis of (3 S, 5 R)-Carbapenam-3-carboxylic Acid and Its Role in Carbapenem Biosynthesis and the Stereoinversion Problem. *J. Am. Chem. Soc.***125**, 15746-15747 (2003).
21. Baldwin, J. E., & Bradley, M. Isopenicillin N synthase: mechanistic studies. *Chem. Rev.* **90**, 1079-1088 (1990).
22. Ge, W.; Wolf, A.; Feng, T.; Ho, C. H.; Sekirnik, R.; Zayer, A.; Granatino, N.; Cockman, M.E.; Loenarz, C.; Loik, N. D.; Hardy, A. P.; Claridge, T. D. Hamed, R. B.; Chowdhury, R.; Gong, L.; Robinson, C. V.; Preston, G. M.; Coleman, M. L.; Schofield, C. J. (2012). Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat. chem. Biol.*, **8**, 960-962 (2012).

23. Cho, K.-B.; Hirao, H.; Shaik, S.; Nam, W. To rebound or dissociate? This is the mechanistic question in C–H hydroxylation by heme and noneheme metal-oxo complexes. *Chem. Soc. Rev.* **45**, 1197-1210 (2016).
24. Newcomb, M., Le Tadic-Biadatti, M. H., Chestney, D. L., Roberts, E. S., Hollenberg, P. F. A nonsynchronous concerted mechanism for cytochrome P-450 catalyzed hydroxylation. *J. Am. Chem. Soc.* **117**, 12085-12091 (1995).
25. Shaik, S., Hirao, H., & Kumar, D. Reactivity of high-valent iron–oxo species in enzymes and synthetic reagents: a tale of many states. *Acc. Chem. Res.* **40**, 532-542 (2007).
26. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Jr. Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:r-Ketoglutarate Dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008-13009 (2003).
27. Hoffart, L. M.; Barr, E. W.; Guyer, R. B.; Bollinger, J. M., Jr.; Krebs, C. Direct Spectroscopic Detection of a C-H-Cleaving High-Spin Fe(IV) Complex in a Prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. USA* **103**, 14738–14743 (2006).
28. Grzyska, P. K.; Appelman, E. H.; Hausinger, R. P.; Proshlyakov, D. A. Insight into the Mechanism of an Iron Dioxygenase by Resolution of Steps Following the Fe^{IV}═O Species. *Proc. Natl. Acad. Sci. USA* **107**, 3982–3987 (2010).
29. Martinez, S.; Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702-20711 (2015).
30. Fisher, G. H. Appearance of D-amino acids during aging: D-amino acids in tumor proteins. *Exs*, **85**, 109-118 (1997).
31. Fujii, N. D-amino acid in elderly tissues. *Biol. Pharm.Bull.* **28**, 1585-1589 (2005).
32. Gaussian 09, Revision D.01, Frisch, M. J. *et al.* Gaussian, Inc., Wallingford CT (2009).

Chapter 4

Computational Prediction of Molecular Hydration Entropy with Hybrid Scaled Particle Theory and the Free Energy Perturbation Method

4.1 Introduction

The hydration free energy (ΔA_{hyd}) quantifies the free energy change in the transfer of a solute from the gas phase to aqueous solution. Although ΔA_{hyd} can serve as a fundamental thermodynamic quantity for the behavior of a small molecule in a variety of chemical and biological systems, the experimental measurement of ΔA_{hyd} has lagged behind the need for its determination. To the best of my knowledge, experimental ΔA_{hyd} values are available only for a few thousands of the millions of organic compounds reported to date. Accordingly, a number of efficient computational methods to predict ΔA_{hyd} have been proposed on the basis of various theoretical frameworks, and assessed in Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) challenges.[1]

Although ΔA_{hyd} can be obtained straightforwardly from the solubility and vapor pressure data of a solute,[2] the latter is difficult to measure precisely with standard apparatus whereas the former can be measured with accuracy by determining the concentration of saturated solution.[11] The scant ΔA_{hyd} data can thus be attributed to the difficulty in measuring the vapor pressure of a solute. Because the vapor pressure can be computed from the hydration entropy (ΔS_{hyd}) and vice versa,[3] the development of reliable computational methods to estimate ΔS_{hyd} seems to give different approaches to calculate the ΔA_{hyd} data for various solute molecules.

Knowledge of ΔS_{hyd} would also be informative for elucidating the thermodynamic aspects of protein-ligand interactions in which the major contribution to the binding free energy comes from the entropic term without significant enthalpy-entropy compensation during the conformational changes for binding.[4] In contrast to the general notion that ligand binding to a protein is accompanied by the entropy loss, it has been demonstrated

that the binding entropy (ΔS_{bind}) can be positive during the formation of some protein-ligand complexes,[5,6] furthermore, in some case the $-T\Delta S_{bind}$ values can be more negative than ΔH_{bind}.[7-10] These observations indicate that the increase of entropy can serve as a driving force for binding of a ligand to the receptor protein. The optimal activity of many drug molecules can therefore be achieved only by a compromise between the maximization of the protein-ligand interactions and the minimization of ΔS_{hyd} to maximize ΔS_{bind}, supporting the development of reliable computational methods to estimate ΔS_{hyd} values.

Despite the lack experimental data for validation, the development of computational methods for estimating ΔS_{hyd} has been pursued by several groups following the pioneering work of Ben-Naim and Marcus on solvation thermodynamics.[11] The results of ideal solution theory-based calculations indicate that the entropic term associated with the first hydration shell will dominate the solvation free energies of alkanes.[12] The Langevin dipoles solvation model has proved to be efficient in predicting the ΔS_{hyd} values of some small neutral molecules and monoatomic ions.[13] Free energy perturbation (FEP) and thermodynamic integration (TI) methods have been applied most frequently in estimating the molecular ΔS_{hyd} values through the thermodynamic decomposition of ΔA_{hyd}.[14-17] Recently, Gerogiokas et al. proposed "the grid cell theory" to calculate ΔS_{hyd} by discretizing the cell theory methodology on a three dimensional grid to resolve the density, enthalpy, and entropy of water molecules around the solute.[18] Through the graphical analyses of the spatially resolved components of ΔA_{hyd} obtained from only a single simulation, they were able to predict ΔS_{hyd} with comparable accuracy to FEP and TI-based methods.[18] However, these methods seem to be insufficient to be useful in

practical applications because large deviations from experimental data were observed even in validation tests with less than 30 solute molecules.

The present study aimed to develop a new computational method for estimating the molecular ΔS_{hyd} values with high accuracy based on the modified thermodynamic decomposition approach. A key feature of this method lies in that the solute-water interactions are decomposed into hydrophobic and electrostatic parts to calculate their respective contributions to ΔS_{hyd} within different theoretical frameworks. This decomposition is required because full FEP calculations have often been unsuccessful in reproducing the ΔS_{hyd} values of small solute molecules.[19,20] The extended thermodynamic decomposition approach was thus anticipated to enhance the accuracy in ΔS_{hyd} prediction by complementing the conventional FEP methods.

4.2. Theory and Computational methods

4.2.1. Hybrid Scaled Particle Theory and Free Energy Perturbation

A solute molecule dissolves in water when the resulting solute-water interactions are strong enough to surmount the entropic penalty for establishing the optimal configuration of hydration.[11] The loss of entropy during hydration stems from the reordering of solute and water molecules to form a complex with the optimal intermolecular interactions. In this context, ΔS_{hyd} may be decomposed into van der Waals (ΔS_{vdW}) and electrostatic (ΔS_{elec}) contributions.

$$\Delta S_{hyd} = \Delta S_{vdW} + \Delta S_{elec} \quad (1)$$

In the usual FEP and TI methods, the ΔS_{vdW} and ΔS_{elec} terms are calculated simultaneously to obtain ΔS_{hyd} using Leonard-Jones and Coulomb potentials, respectively.[37] However, it is often difficult to successfully complete molecular dynamics

(MD) simulations for FEP and TI procedures due to the appearance of singularity in the Leonard-Jones potential as the potential function is scaled from the initial to final state.[22] Although this singularity problem can be overcome by scaling the electrostatic and van der Waals interaction terms in the sequential manner, it is difficult to completely avoid the numerical instability in FEP calculations due to the possibility of the appearance of unwanted minima in the non-bonded potential energy at intermediate λ values. This problem may arise especially when the attractive electrostatic interactions are counterbalanced insufficiently by van der Waals repulsions.[21] Because of such technical difficulties in coping with the van der Waals interaction term, it seems to be a reasonable choice to compute the ΔS_{vdW} term separately with a simpler method than used in full FEP calculations. Recently, the introduction of the term $\alpha(1-\lambda_{vdw})^2$ of soft-core potential can eliminate the singularity problem of the van der Waals interaction during simulations.[82] However, in this soft core potential method, they still cannot reduce uncertainty of energy (ΔE) and entropy ($-T\Delta S$) since they are amplified by the factor $T/\Delta T$.[37]

The sign of ΔS_{vdW} is negative for most organic solutes due to the immiscibility between their hydrophobic parts and water. ΔS_{vdW} can therefore be approximated as the loss of entropy associated with the reordering of water molecules to form a cavity (ΔS_{cav}) in which a solute molecule can be accommodated. Then, ΔS_{hyd} can be expressed in the following form as eq (2):

$$\Delta S_{hyd} = \Delta S_{cav} + \Delta S_{elec} \qquad (2)$$

Some implicit solvent models such as solvent-accessible surface area (SASA) method seems to be appropriate for the calculation of molecular hydration free energy.[22] It would be possible to calculate the S_{cav} term within the SASA framework upon the availability

of a proper method for calculating the first derivative of hydration free energy with respect to temperature. In the S_{cav} term can be obtained straightforwardly from the relation between the hydration free energy and temperature, two theoretical frameworks can be applied to obtain the molecular ΔS_{cav} values. I first employed the scaled-particle theory (SPT) developed for describing the radial distribution of hard sphere mixtures in the liquid phase.[23,24] Because SPT proved to be appropriate for describing the solute-solvent interactions for hydrophobic solutes,[25,26] it is anticipated to be useful for estimating the ΔS_{vdW} values of organic solute molecules.

4.2.2. Scaled Particle Theory

SPT aims to calculate the chemical potential ($\Delta\mu$) for the insertion of a cavity into a hard sphere fluid.[27] Because the formation of a cavity for a hydrophobic solute in water may be approximated as an entropic process due to the weak solute-water interactions, ΔS_{cav} can be expressed as follows.

$$\frac{\Delta S_{cav}}{k_B} = -\ln(1-\varphi) + 3\left(\frac{\varphi}{1-\varphi}\right)R + \left[\left(\frac{3\varphi}{1-\varphi}\right) + 4.5\left(\frac{\varphi}{1-\varphi}\right)\right]R^2 + \frac{\varphi P}{\rho k_B T}R^3 \qquad (3)$$

Here, φ and ρ represent the volume fraction of water molecules with respect to the total volume (0.371) and the number density of water (0.333×10^{24} m^{-3}), respectively. R denotes the ratio of the radius of solute to that of water molecule (r_t/r_w) while P stands for the pressure of water (2.306 kPa). Because r_w is known to be 1.4 Å,[28] the knowledge of r_t is the only requisite for calculating the ΔS_{cav} value of a solute molecule. The r_t parameter of each solute was set equal to the radius of the water-accessible volume (WAV) that was assumed to have spherical symmetry. Within this approximation, r_t was calculated straightforwardly from the WAV value that could be set equal to $(4/3)\pi r_t^3$.

The accuracy of the present hybrid SPT/FEP method is limited because a hard-sphere model had to be used for solutes in SPT calculations. Prior to calculating the ΔS_{cav} values within the SPT framework, therefore, I carried out MD simulations with the full molecular structures of solutes and explicit water molecules to calculate the WAV values of all solute molecules as accurately as possible. As it was anticipated that MD-based calculations would have the effect of reflecting the conformational diversity of solute molecules, which was shown to be a significant contributor to hydration entropy.[29] The disadvantage of conventional SPT can thus be complemented in part by statistical sampling of various conformations of solute molecules in aqueous solution. The WAV value of each solute molecule was then calculated with the analytical method proposed by Petitjean[30] in which it was given by the union of the van der Waals volumes of all the solute atoms.

4.2.3. Information Theory

As an alternative to SPT, information theory (IT) can also be applied to calculate ΔS_{cav} as suggested by Hummer et al.[31] In this theoretical framework, the probability (p_n) that n water molecules are found in a cavity is defined as follows;

$$p_n = \exp(\omega_0 + \omega_1 n + \omega_2 n^2) \qquad (4)$$

Here, ω_0, ω_1, and ω_2 parameters can be determined by solving Lagrange equations with the pair correlation function ($g(r)$) between a pair of oxygen atoms in water. MD simulations were performed with the NAMD program of version 2.8[32] in the calculation of $g(r)$. To select the number of water molecules to describe the bulk solvent, we compared the *g(r)* profiles obtained with 233 and 1981 TIP3P[33] waters used as the solvent model for 2-propanol. As shown in Figure 4.1, *g(r)* remains almost intact with the increase in the number of water molecules from 233 to 1981. To save the computational

burden, therefore, we carried out MD simulation with 233 TIP3P waters to calculate the ΔS_{cav} values of the solute molecules.

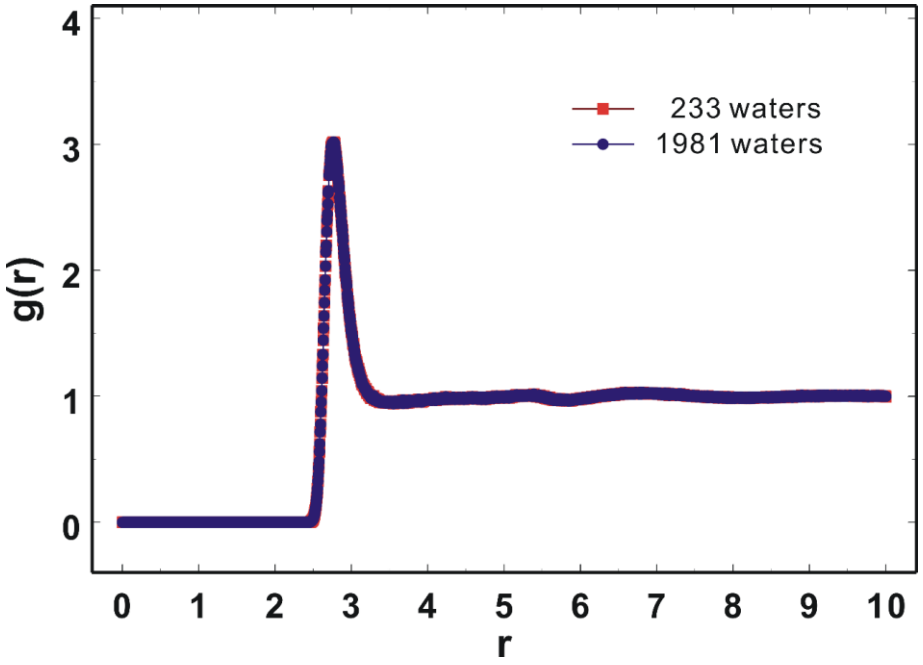


Figure 4.1. Comparison of the pair correlation functions calculated via 1 nanosecond MD simulations with 233 and 1981 water molecules. Figure taken from Choi et al, *J. Chem. Theory Comput*. *11*, p.4935. Copyright 2015 American Chemical Society. Reprinted with permission.

After the equilibration dynamics for 1 ns, water configurations were sampled for 1 ns production dynamics in NVT ensemble to obtain $g(r)$. ω_0, ω_1, and ω_2 parameters were then determined by numerically solving the Lagrange equations with a multi-dimensional root finding algorithm.[34] Finally, ΔS_{cav} value of each solute molecule was calculated with the negative logarithm of p_0. the probability (p_0) means that *zero* water molecule is found in a cavity.

$$\Delta S_{cav} = -k_B ln p_0 \qquad (5)$$

4.2.4. Implementation of electrostatic terms using Free Energy Perturbation

As mentioned above, it would be reasonable to separately calculate the entropic cost for van der Waals interactions between solute and water with SPT or IT. Then, FEP

simulations can be performed for the purpose of calculating the ΔS_{elec} term only by scaling the Coulomb potential with the unstable Leonard-Jones potential being kept intact throughout the simulations. This was accomplished using a scaling parameter (λ_{elec}) that played the role of creating and annihilating the partial charges of solute and water atoms in the FEP simulations of solute hydration. The electrostatic part of ΔA_{hyd} calculated with FEP was then used to obtain ΔS_{elec}, which can be expressed as the temperature derivative of ΔA_{hyd}. As widely adopted in the literature,[35-39] we used the finite difference method to calculate the ΔS_{elec} value of each solute molecule.

Because entropy can be given by the derivative of Helmholtz free energy (A) with respect to temperature at constant volume, the entropy change (ΔS) associated with the solute-water interactions is written as follows using the finite difference method.

$$\Delta S = -\frac{\Delta A(T+\Delta T)-\Delta A(T-\Delta T)}{2\Delta T} \quad (6)$$

If one neglects the van der Waals term in the potential energy function, ΔA at a given temperature is the difference of Helmholtz free energies between the charged and uncharged states of the system. In this case, ΔS_{elec} can be given by eq 7.

$$\Delta S_{elec} = -\frac{\{\Delta A_q(T+\Delta T)-\Delta A_0(T+\Delta T)\}-\{\Delta A_q(T-\Delta T)-\Delta A_0(T-\Delta T)\}}{2\Delta T} \quad (7)$$

Here, A_q and A_0 are the Helmholtz free energies of charged and uncharged systems, respectively. It is thus necessary to compute the differences between A_q and A_0 at temperatures $T + \Delta T$ and $T - \Delta T$ to obtain ΔS_{elec}, which were accomplished successfully with the constrained FEP simulations.

To select a proper ΔT value for obtaining the first derivative of ΔA by finite difference method, we calculated the ΔS_{elec} values with varying ΔT (10, 20, and 30 K) using 2-

propanol as the probe molecule. In this test case, the statistical errors of the calculated ΔS_{elec} values appeared to become larger as ΔT decreased from 30 to 20 and 10 K.[17] This validation result is consistent with the previous computational finding that the finite difference method can be valid near the room temperature with $\Delta T = 30\ K$.[17] Therefore, ΔT was set equal to 30 K in this study.

Because ΔA_{hyd} refers to the change in free energy for a solute molecule to leave the gas phase and to enter the liquid water, a thermodynamic cycle may be useful to calculate the ΔA_{hyd} values at varying temperatures, which is illustrated in Figure 4.2. Because A is thermodynamically a state function, the sum of all free energy changes in the cycle must be zero. ΔA_{hyd} can therefore be expressed as follows.

$$\Delta A_{hyd} = \Delta A_{gas} - \Delta A_{aq} \tag{8}$$

ΔA_{aq} and ΔA_{gas} correspond to the difference between A_q and A_0 values of the solute in aqueous solution and that in the gas phase, respectively. I obtained the ΔA_{hyd} values of all the solute molecules through the FEP calculations of ΔA_{aq} and ΔA_{gas} based on the independent MD simulations in water and in the gas phase, respectively.

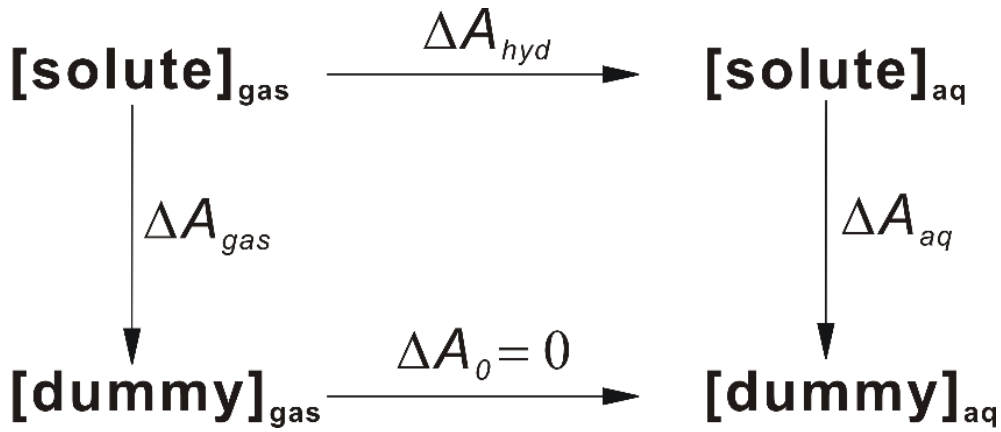


Figure 4.2. Thermodynamic cycle for calculating the hydration free energies of solute molecules. Note that $\Delta A_0 = 0$ because solute is not involved in the process. Figure taken from Choi et al, *J. Chem. Theory Comput. 11*, p.4935. Copyright 2015 American Chemical Society. Reprinted with permission.

To determine the ΔA_{aq} and ΔA_{gas} values, the solute molecule with atomic charges (state 0) was changed gradually to the dummy molecule in which all the atoms were uncharged (state 1) by perturbing the Hamiltonian of the system. This transformation could be accomplished by scaling the electrostatic interaction potential with the change of a state variable (λ_{elec}) that maps the initial and the final states onto 0 and 1, respectively. ΔA_{aq} and ΔA_{gas} values for the change from the initial to final state were then calculated by summing the incremental free energy changes over the several windows of λ_{elec} that was varied from 0 to 1 with the discrete and uniformly spaced intervals.[40]

$$\Delta A(0 \rightarrow 1) = -RT \sum_i ln \langle exp\left(-\frac{V_{\lambda(i+1)} - V_{\lambda(i)}}{RT}\right)\rangle_{\lambda(i)} \qquad (9)$$

Here V_i is the potential energy function for the representative state i, and $\langle \cdots \rangle_i$ denotes the ensemble average over the samples of the enclosed quantity in state i.

FEP calculations to compute ΔA_{aq} and ΔA_{gas} values were performed through MD simulations of the solutes in aqueous solution and in the gas phase, respectively. In the former case, each solute molecule was immersed in a periodic box with the dimension of 29.1×29.1×29.1 Å, which was followed by the addition of 824 TIP3P water molecules in the simulation box to be consistent with the density of water at 298 K. For all the solute molecules under investigation, the entire water-solute system exhibited a good convergent behavior at 268 *K* as well as 328 *K* after around 500 ps of simulation time because it included only a small molecule and waters.

After the equilibration dynamics for 1 ns, I carried out 18 ns of perturbation using the NAMD program of version 2.8[32] at 268 and 328 K with varying λ_{elec} values. Each perturbation consisted of nine windows with 1 ns of equilibration and 1 ns of data collection. CGenFF[41] parameter sets were used for all the solute molecules and

CHARMM[42] parameter sets for TIP3P water. Atomic charges were assigned using the extended bond-charge increment scheme to be able to capture short- and medium-range inductive effects. All MD-based FEP calculations were conducted with periodic boundary conditions in the NVT ensemble and with the nonbond-interaction cutoff radius of 12 Å. We used the time step of 1 fs to integrate the Langevin equations of motion with the friction coefficient of 5 ps^{-1}. MD simulations started from $\lambda_{elec} = 0$, which corresponds to the fully charged system. Both intramolecular and intermolecular electrostatic interactions were gradually made weaker by increasing λ_{elec} until it reaches 1 at which all atomic charges of the solute molecule vanish. Because only the electrostatic part of ΔA_{hyd} (ΔA_{elec}) needed to be calculated in FEP simulations, the Coulomb potential was scaled by parameter $1-\lambda_{elec}$, whereas the Leonard-Jones term remained unchanged during the entire course of simulations. The configurations were sampled at every 10 time steps to calculate the variation of electrostatic interaction energies with the increase of λ_{elec} value by 0.125. At each λ_{elec} value, we took the ensemble average of total energies to obtain the incremental hydration free energy ($\Delta A_{elec}(\lambda_{elec})$) as implemented in the NAMD program. All these $\Delta A_{elec}(\lambda_{elec})$ values were then summed over λ_{elec}'s to obtain the ΔA_{elec} values at 268 K and 328 K, which were in turn combined to calculate ΔS_{elec} using eq 7. A doublewide sampling was performed for all the charge transformations, and the final results were obtained by taking the averages from the backward and forward simulations.

4.2.5. Comparison of hybrid SPT/FEP with Full FEP

To compare the performances of our hybrid SPT (or IT) and FEP method with the conventional full FEP simulations, I also calculated the ΔS_{hyd} values of 18 representative solute molecules shown in Figure 4.3. This test set included methane, propane, butane,

2-methylpropane, toluene, methanol, ethanol, *p*-cresol, acetamide, and N-methylacetamide, which had been widely studied as the simplified structural models for side-chain and backbone groups of proteins. Some large hydrophobic molecules such as acenaphthene, phenanthrene, pyrene, 2,6-ditert-butylpyridine and 2-adamantanone were included in the comparative analysis because SPT and IT may become less effective,[26,31] in estimating the ΔS_{cav} values with the increase of van der Waals volume. Despite the extensive search of literature and publicly accessible chemical databases, I found that the experimental hydration entropy data were unavailable for the solutes with molecular weight larger than 300 amu as compared to the abundance of hydration free energy data. Therefore, the validation set was further enriched with piperidine, 1,4-dioxane, and morpholine because many drug molecules include the heterocyclic moieties. Although some of the calculated ΔS_{hyd} values were available in the literature, we recalculated all of them because of the strong dependence of FEP results for ΔS_{hyd} on force fields and water models.[43]

Full FEP calculations were carried out by scaling both van der Waals and electrostatic interaction terms with the two state variables λ_{vdW} and λ_{elec}, respectively. More specifically, ΔA_{hyd} values of the test molecules were obtained by switching off the electrostatic and van der Waals interaction potentials in the sequential way to warrant the stability of the entire system including the solute and water molecules. We first defined 21 windows to perform the FEP simulation of each solute with the λ values ranging from 0 to 1 ($\Delta\lambda = 0.05$). To further reduce the potential statistical errors in FEP calculations, the two λ states were added near the physical end states ($\lambda = 0.01$ and $\lambda = 0.99$) as recommended in the literature.[44] Following the suggestion by Pohorille et al.,[45] we

combined the results of forward and backward simulations to obtain the ΔA_{hyd} values of all solute molecules rather than using the single unidirectional simulation results.

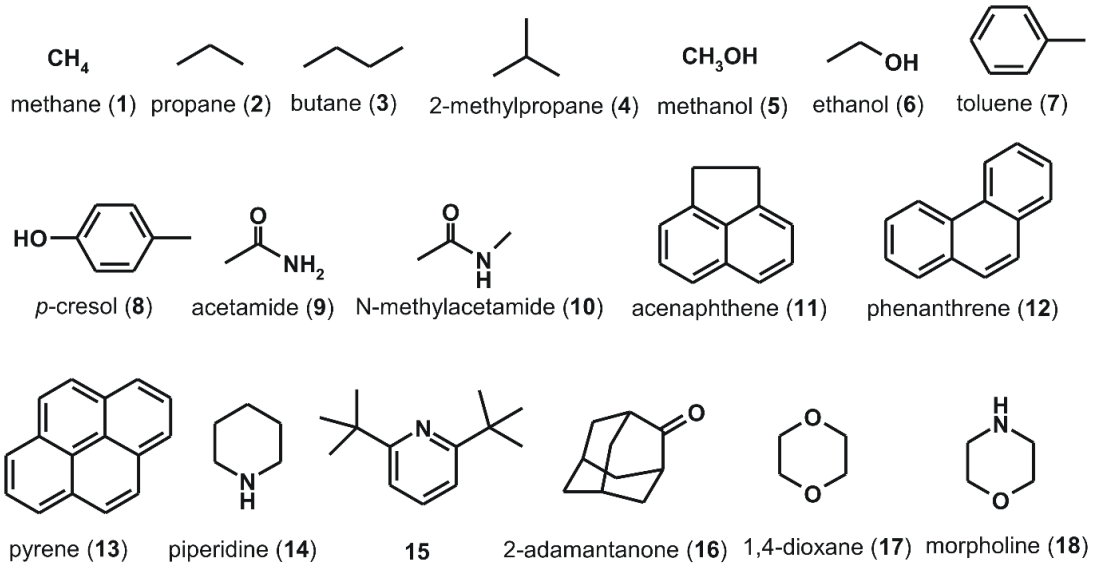


Figure 4.3. Structures of the molecules used to compare the performances of hybrid SPT/FEP method with those of the conventional full FEP simulations. IUPAC name of **15** is 2,6-ditert-butylpyridine. Figure taken from Choi et al, *J. Chem. Theory Comput*. *11*, p.4936. Copyright 2015 American Chemical Society. Reprinted with permission.

To improve the poor convergent behavior of the van der Waals interaction term in the full FEP calculations, we used the soft-core potential (U_{vdW}^{SC}) shown in eq 10[46,47] instead of the original Leonard-Jones potential.

$$U_{vdW}^{SC}(r,\lambda_{vdw}) = 4\lambda_{vdw}\epsilon\left[\left(\frac{\sigma^6}{\alpha(1-\lambda_{vdw})^2\sigma^6+r^6}\right)^2 - \left(\frac{\sigma^6}{\alpha(1-\lambda_{vdw})^2\sigma^6+r^6}\right)\right] \qquad (10)$$

Here, α is set equal to 0.5 to control the magnitude of soft-core term. This soft-core potential has a functional dependence on λ_{vdW} and reduces to the original Leonard-Jones potential when λ_{vdW} is 1. Due to the presence of the soft-core term in both repulsive and attractive parts, the problematic singularity can be avoided in a large part in FEP simulations.

To establish a method for estimating ΔS_{hyd}, it was required to prepare a reference dataset comprising various organic molecules and their experimental ΔS_{hyd} values. Therefore, we built a chemical library of 90 organic molecules (21 hydrocarbons and 69 polar molecules) for which the experimental hydration enthalpies were available in Minnesota Solvation Database and Physical/Chemical Property Database,[48] and the corresponding experimental heat capacity data could be extracted from *CRC Handbook of Chemistry and Physics*.[49] The enthalpy and heat capacity data were then combined to generate the experimental ΔS_{hyd} values that served as the yardsticks to validate the corresponding computational results. The standard states of the reference data set were normalized to be consistent with the definition of hydration entropy by Ben-Naim and Marcus.[11]

4.3. Results

4.3.1 Correlation for cavity formation of small molecules with SPT and IT

With respect to the calculation of ΔS_{cav}, the result indicates SPT and IT can be applied because both proved to be effective in estimating the entropic cost for creating a void volume in which a solute molecule can be accommodated. To select the more appropriate theoretical tool, we compared the results for the ΔS_{cav} values of 90 solute molecules calculated with SPT and IT. As shown in Figure 4.4, the ΔS_{cav} calculation results of SPT are in nearly complete agreement with those of IT with the associated squared linear correlation coefficient (R^2) of 0.999. Furthermore, the slope and intercept of the regression line amount to 1.005 and 0.111, respectively. This excellent agreement indicates almost the same performance of SPT as IT in estimating the ΔS_{cav} value of a solute. Considering the small difference between SPT and IT results, I will hereafter refer to the average of the two ΔS_{cav} values for each solute molecule calculated with SPT and IT as its ΔS_{cav} value.

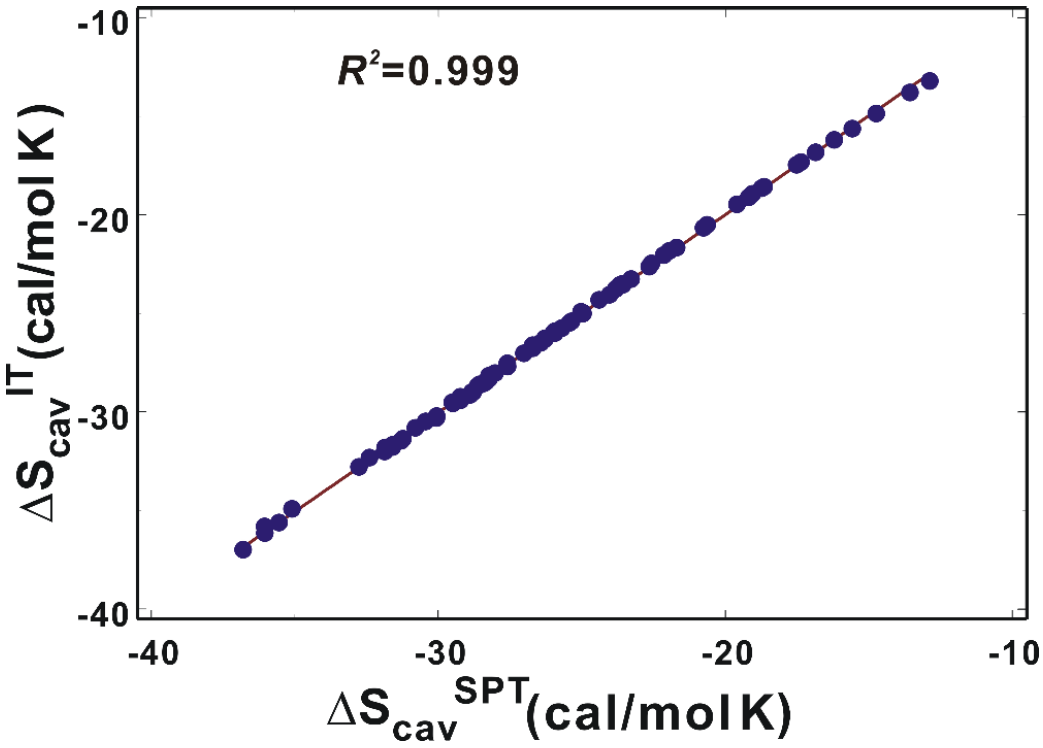


Figure 4.4. Correlation diagram for the cavity formation entropies (ΔS_{cav}) of 90 organic molecules calculated with scaled particle theory (SPT) and information theory (IT). The slope and intercept of the fitting are 1.005 and 0.111, respectively. Figure taken from Choi et al, *J. Chem. Theory Comput. 11*, p.4937. Copyright 2015 American Chemical Society. Reprinted with permission.

4.3.2 Calculated hydration entropies vs. Experimental hydration entropies

With respect to the high correlation between the ΔS_{cav} values calculated with SPT and IT, [26] we note that all 90 solute molecules are small with molecular weight lower than 200 amu. For this reason, their van der Waals radii fall within 3.5 Å, which has the effect of limiting the number of water molecules in the cavity formed by a solute molecule to one or two. In this case, the probability of finding no water molecule in the cavity (p_0) proved to be very similar in SPT and IT.[50] Because ΔS_{cav} is given by the negative logarithm of p_0 (eq 5), the small cavity volume of solutes can be invoked to explain the coincidence of the ΔS_{cav} results obtained with SPT and IT.

Table 4.1 lists the calculated ΔS_{cav}, ΔS_{elec}, and ΔS_{hyd} values in comparison with the corresponding experimental hydration entropy (ΔS_{hyd}^{X}) values for all 90 solute molecules under investigation.

compound	ΔS_{cav}	ΔS_{elec}	ΔS_{hyd}	ΔS_{hyd}^{X}	compound	ΔS_{cav}	ΔS_{elec}	ΔS_{hyd}	ΔS_{hyd}^{X}
methane	-13.67	0.21	-13.46	-16.33	3-methylpyridine	-26.29	-2.28	-28.56	-27.76
ethane	-17.34	0.39	-16.94	-20.47	4-methylpyridine	-26.30	-2.39	-28.69	-27.95
propane	-20.57	0.57	-20.00	-22.91	3-ethylpyridine	-28.91	-2.91	-31.82	-27.44
butane	-23.55	0.42	-23.13	-26.85	4-ethylpyridine	-28.87	-1.95	-30.81	-28.38
2-methylpropane	-23.53	0.59	-22.93	-25.17	3,4-dimethylpyridine	-28.88	-2.77	-31.64	-27.86
pentane	-26.27	0.41	-25.86	-30.51	3,5-dimethylpyridine	-28.89	-2.74	-31.63	-32.26
2,2-dimethylpropane	-26.27	0.70	-25.57	-27.14	4-tert-butylpyridine	-33.37	-2.60	-35.97	-31.41
hexane	-28.91	0.31	-28.60	-33.67	acetonitrile	-17.48	2.01	-15.47	-14.94
heptane	-31.29	0.26	-31.03	-35.95	propanenitrile	-20.70	1.17	-19.53	-18.91
ethene	-19.14	0.49	-18.65	-14.86	2-butanone	-24.04	-3.81	-27.84	-24.60
1-propene	-16.20	1.64	-14.56	-15.25	2-pentanone	-26.66	-3.37	-30.03	-27.98
cyclopropane	-19.53	1.75	-17.78	-21.09	3-pentanone	-26.73	-3.95	-30.68	-28.31
benzene	-24.35	0.27	-24.08	-22.57	3-methyl-2-butanone	-26.66	-5.36	-32.01	-35.21
methylbenzene	-27.02	1.17	-25.85	-26.09	2-hexanone	-29.31	-3.81	-33.11	-30.49
ethylbenzene	-29.51	0.42	-29.08	-29.59	4-methyl-2-pentanone	-29.22	-4.87	-34.09	-31.98
m-xylene	-29.52	0.75	-28.77	-28.14	2-heptanone	-31.68	-4.68	-36.37	-35.42
p-xylene	-29.52	0.75	-28.77	-30.13	4-heptanone	-31.63	-5.37	-37.00	-36.76
propylbenzene	-31.94	0.68	-31.26	-33.40	2-nonanone	-36.08	-4.33	-40.41	-41.63
2-phenylpropane	-31.82	-0.16	-31.98	-32.22	5-nonanone	-35.92	-4.76	-40.68	-41.38
9H-fluorene	-35.57	0.69	-34.88	-32.78	propanoic acid	-22.51	-2.50	-25.01	-23.58
anthracene	-36.89	1.32	-35.57	-33.72	butanoic acid	-25.38	-2.55	-27.92	-26.38
methanol	-15.59	-2.70	-18.29	-18.53	diethylether	-24.96	-4.16	-29.12	-31.29
ethanol	-18.99	-3.97	-22.96	-23.02	1,2-dimethoxyethane	-26.46	-3.72	-30.18	-31.35
1-propanol	-22.09	-4.50	-26.59	-27.01	tetrahydrofuran	-23.27	-2.44	-25.71	-26.25
2-propanol	-22.09	-5.47	-27.56	-29.12	tetrahydropyran	-25.96	-5.86	-31.82	-28.72
1-butanol	-24.96	-5.18	-30.14	-33.55	2-ethyltetrahydrofuran	-26.01	-3.34	-29.35	-30.11
2-methyl-1-propanol	-24.98	-5.44	-30.42	-32.67	2,5-Dimethyloxolane	-28.57	-4.19	-32.76	-35.34
2-butanol	-24.96	-5.86	-30.83	-34.58	1,3-dioxolane	-21.89	-0.91	-22.80	-18.26
2-methyl-2-propanol	-24.97	-5.76	-30.74	-33.90	methylacetate	-22.63	-1.46	-24.09	-22.95
1-pentanol	-27.63	-3.74	-31.37	-34.43	ethylacetate	-25.44	-3.15	-28.60	-26.23
3-pentanol	-27.55	-6.68	-34.24	-33.17	methylbutanoate	-28.02	-2.93	-30.95	-29.06
2-methyl-2-butanol	-27.64	-6.68	-34.31	-37.76	N-butylacetate	-30.46	-4.55	-35.01	-33.69
1-hexanol	-30.18	-4.75	-34.93	-38.45	ethylpentanoate	-32.77	-4.51	-37.28	-36.98
3-hexanol	-30.12	-5.81	-35.93	-42.11	2-methoxyethanol	-23.61	-4.93	-28.54	-28.30
4-heptanol	-32.35	-5.72	-38.07	-46.13	2-propoxyethanol	-29.01	-6.75	-35.76	-34.56
1,2-ethanediol	-20.57	-5.59	-26.15	-26.77	2-butoxyethanol	-31.37	-6.57	-37.94	-38.01
cyclopentanol	-25.92	-5.76	-31.68	-35.18	3-hydroxybenzaldehyde	-28.67	0.27	-28.40	-24.87
cyclohexanol	-28.48	-5.78	-34.26	-37.87	3-hydroxybenzonitrile	-28.39	0.18	-28.21	-24.27
cycloheptanol	-30.80	-4.05	-34.85	-41.37	fluoromethane	-14.79	2.03	-12.76	-12.17
phenol	-25.72	-0.13	-25.84	-24.06	2,2,2-trifluoroethanol	-21.67	-0.36	-22.03	-24.51
o-cresol	-28.27	-0.66	-28.93	-30.54	formic acid	-16.13	-3.82	-19.96	-19.49
p-cresol	-28.20	-0.48	-28.68	-28.83	dimethylformamide	-18.68	-3.51	-22.19	-24.23
4-tert-butylphenol	-35.00	-2.87	-37.87	-36.16	acetamide	-20.09	-1.53	-21.62	-22.17
pyrroidine	-23.79	-5.95	-29.74	-32.57	N-methylacetamide	-23.19	-5.38	-28.57	-31.06
pyridine	-23.57	-3.21	-26.77	-24.65	tetrahydro-2-nmethanol	-27.32	-1.92	-29.25	-29.47

Table 4.1. ΔS_{cav}, ΔS_{elec}, and ΔS_{hyd} values of 90 solute molecules calculated at 298 K (in cal·mol^{-1}·K^{-1}) in comparison with the corresponding experimental hydration entropy (ΔS_{hyd}^{X}). grey background means 21 hydrocarbons. Table taken from Choi et al, *J. Chem. Theory Comput. 11*, p.4938. Copyright 2015 American Chemical Society. Reprinted with permission.

ΔS_{cav} tends to become more negative with the increase in the molecular weight of solutes. This is unsurprising because the larger a solute molecule is, the more it will make the water molecules lose the degree of freedoms to construct a void volume. More interestingly, ΔS_{hyd} values appear to be dominated in magnitude by ΔS_{cav} rather than ΔS_{elec}. For example, the magnitude of the latter is only 11.3% of that of the former on average for all the solute molecules. Such a minor contribution of electrostatic solute-water interactions to hydration entropy was also observed in thermodynamic decomposition analysis of hydration free energies of alkanes and amines.[17,18] The dominance of ΔS_{cav} term in ΔS_{hyd} is also consistent with the previous computational finding that the loss of entropy during the hydration of a solute could be attributed in a large part to the formation of a solvent-excluded volume with only a minor contribution of the electrostatic solute-water interactions.[51,52] Thus, the computational results confirm that the entropic penalty for hydration stems from the repulsive hydrophobic solute-water interactions rather than the electrostatic ones. It should be noted that only the repulsive van der Waals interactions were considered in this study to calculate ΔS_{cav} with the attractive part being neglected for simplicity. To address the validity of this approximation, we compared the calculated ΔS_{cav} values of 21 hydrocarbons for which the electrostatic interactions with water would be negligibly small with the corresponding experimental ΔS_{hyd} values. As shown in Figure 4.5, the ΔS_{cav} values compare reasonably well with experimental ΔS_{hyd} data with the associated R^2 and root mean square error (RMSE) of 0.842 and 2.65 cal·mol^{-1}·K^{-1}, respectively. The slope and intercept of the linear correlation diagram amount to 0.914 and –1.501, respectively. These results support the correct assumption that the attractive van der Waals interactions between the hydrophobic solute and water would have little effect on ΔS_{cav}.

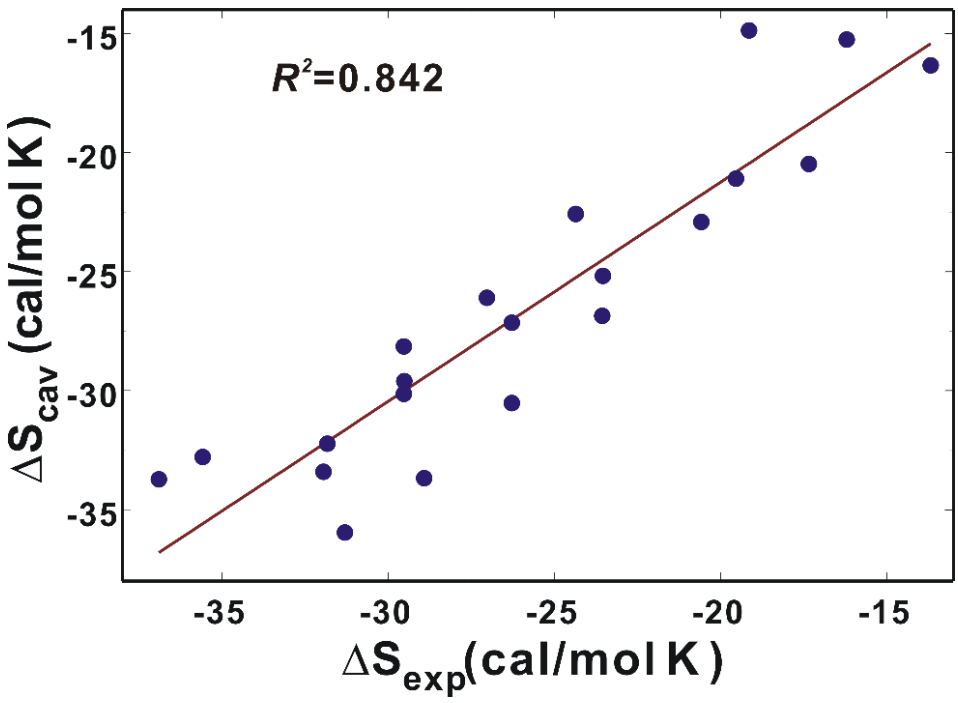


Figure 4.5. Linear correlation diagram between the experimental hydration entropies and the calculated ΔS_{cav} values of 21 hydrocarbons. The slope and intercept of the fitting are 0.914 and –1.501, respectively Figure taken from Choi et al, *J. Chem. Theory Comput*. *11*, p.4939. Copyright 2015 American Chemical Society. Reprinted with permission.

4.3.2.1. Calculated hydration entropies of hydrophobic molecules

The largest differences between ΔS_{cav} and ΔS_{elec} are observed in hydrocarbons: the average absolute value of the latter amount to only 2.6% of that of the former for 21 hydrocarbons from methane to anthracene (grey background in Table 4.1). Such a low absolute ΔS_{elec} value for hydrocarbons is actually not surprising because their individual atoms have insignificant partial charges due to the similarity of electronegativities between carbon and hydrogen. It is also noteworthy that in contrast to the solutes with polar moieties, most hydrocarbons have positive ΔS_{elec} values, which implies the gain of entropy through the solute-water electrostatic interactions. This may be understood in the context that the configurational degree of freedoms of hydrocarbons would increase in water due to the lack of a stable conformation. Judging from the highly negative ΔS_{cav} values (Table 4.1), however, such a gain of entropy seems to be negligible when

compared to the loss of entropy in the formation of a cavity. Thus, the difficulty in the formation of a void volume may be responsible for the immiscibility of hydrocarbons in water.

4.3.2.2. Calculated hydration entropies of hydrophilic molecules

Most solute molecules with polar group(s) appear to have negative ΔS_{elec} values, except for acetonitrile, propanenitrile, 3-hydroxybenzaldehyde, and 3-hydroxybenzonitrile (Table 4.1), which indicate the loss of configurational entropy in the electrostatic complexation between the solute and water molecules. Because the hydration of a hydrophilic solute is thermodynamically favorable, the entropic penalties both for the formation of a cavity and for the establishment of electrostatic interactions seem to be compensated with extra energetic stabilizations of the polar groups in water. ΔS_{elec} values of 3-hydroxybenzaldehyde and 3-hydroxybenzonitrile are likely to be overestimated in FEP calculations because they are positive (albeit quite small) despite the possession of two polar groups. More serious overestimations of ΔS_{elec} are observed in acetonitrile and propanenitrile (Table 4.1). This was unexpected because the polar –C≡N moiety can establish a hydrogen bond with water in the ordered form. Therefore, the good agreement between the computational and experimental ΔS_{hyd} values of both nitrile compounds is most likely to be achieved by error cancellation in the ΔS_{cav} and ΔS_{elec} results.

In Table 1, it is also worth noting that the polar solutes with a phenolic –OH moiety have very small negative ΔS_{elec} values. The average of ΔS_{elec} values for phenol, *o*-cresol, and *p*-cresol is -0.40 cal·mol^{-1}·K^{-1} while that of the other polar solutes with at least one hydrogen-bonding group amounts to -3.64 cal·mol^{-1}·K^{-1}. This may be understood in terms of the asymmetric distribution of negative and positive charges in water molecule, which makes the polar groups hydrated differently according to the sign and magnitude of atomic charge. In this regard, the solute atoms with large negative charges were found

to be hydrated in preference to those with positive charges.[53] Because it is peculiar to the phenolic oxygen that one of its lone electron pairs tends to delocalize into the adjacent phenyl ring, its low negative ΔS_{elec} value is consistent with the weakening of electrostatic interactions with water molecules caused by the decrease of the negative charge on the phenolic oxygen.

4.3.2.3. Correlation of calculated hydration entropies vs. experimental hydration entropies

We compared the ΔS_{hyd} values of all 90 solute molecules given by ΔS_{cav} plus ΔS_{elec} values calculated with SPT (or IT) and MD-based FEP simulations, respectively, to the corresponding experimental ones. The linear correlation diagram between the experimental and calculated results is shown in Figure 4.6. Overall, the estimated hydration entropies compare reasonably well with the experimental ones with the associated RMSE of 2.68 cal·mol^{-1}·K^{-1}, which corresponds to 9.2% of the average of the absolute experimental values. In this fitting, the R^2 value amounts to 0.843 with the slope and the intercept values of 1.011 and -0.148, respectively. The present method for estimating ΔS_{hyd} is found to outperform the Langevin dipoles solvation model in terms of both R^2 and RMSE values.[13] The ΔS_{hyd} results in this study are also in better agreement with the experimental ones than those obtained by thermodynamic decomposition analysis of the ΔA_{hyd} values calculated from full FEP simulations.[14,17] Furthermore, our hybrid SPT/FEP hydration model shows higher performance in estimating ΔS_{hyd} than the grid cell theory in which R^2 and RMSE values for 12 solutes amounted to 5.089 cal·mol^{-1}·K^{-1} and 0.675, respectively.[18] These comparisons imply that the combination of SPT (or IT) and electrostatic perturbation may serve as a new efficient computational tool for estimating the hydration entropies of organic molecules.

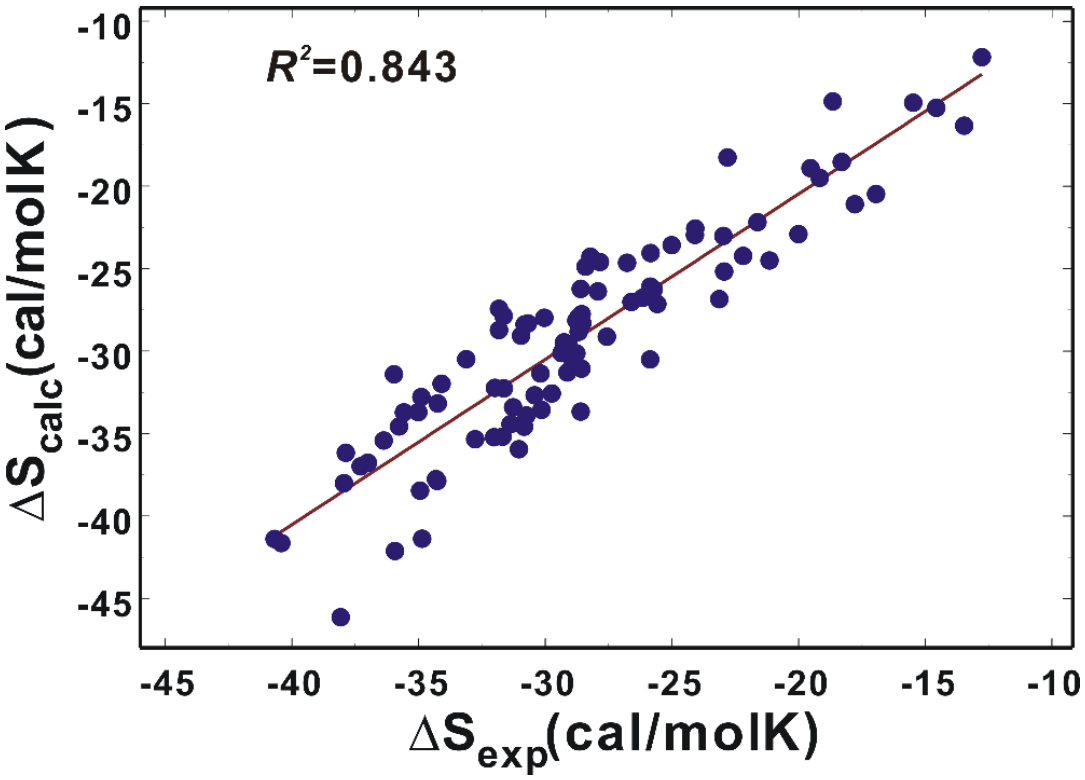


Figure 4.6. Linear correlation diagram between the experimental and calculated hydration entropies of 90 compounds at 298 K. The slope and intercept of the fitting are 1.011 and –0.148, respectively. Figure taken from Choi et al, *J. Chem. Theory Comput*. *11*, p.4939. Copyright 2015 American Chemical Society. Reprinted with permission.

The largest differences between the experimental and calculated ΔS_{hyd} values are observed in a few polar solutes including 4-heptanol, cycloheptanol, and 3-hexanol: the deviations amount to larger than 6.0 cal·mol^{-1}·K^{-1}. The present computational method also shows a poor performance in estimating the ΔS_{hyd} values of alkane compounds with five to seven carbons (pentane, hexane, and heptane) with the difference between experimental and calculated results larger than 4.5 cal·mol^{-1}·K^{-1}. For the better prediction of ΔS_{hyd}, therefore, some modifications seem to be required both in SPT/IT approach to compute ΔS_{cav} and in FEP calculation of ΔS_{elec}. In case of N-methylacetamide that has been actively studied as a simplified structural model for protein backbone,[54] our method yields only 8.0% error in the estimation of ΔS_{hyd} as compared to 30.4% in full FEP simulation results of Wan et al.[37]

The overall accuracy of the calculated ΔS_{hyd} values of hydrocarbons appears to be similar to that of polar solutes with the associated RMSE values of 2.80 and 2.64 cal·mol^{-1}·K^{-1}, respectively. In case of hydrocarbons, significant errors may be accumulated in ΔS_{cav} due to the imperfect conformational sampling in MD simulations as well as to the ambiguity in choosing a representative structure. Such difficulties in the structural characterization of hydrocarbons basically stems from a large number of conformational degree of freedoms afforded by the immiscibility with water molecules. On the other hand, the inaccuracy of ΔS_{hyd} values of some polar molecules may be attributed to the difficulty in unraveling the effects of solute-water interactions on the configurational change of water molecules. Because polar solutes interact directly with the water molecules in the first hydration shell, they induce the larger configurational change than hydrophobic solutes. Indeed, the ordered first hydration shell waters were found to fractionate into not only a more highly ordered but also into a more disordered component in the presence of polar solutes.[55]

4.3.3. Comparison of hybrid SPT/FEP with full FEP

Despite the inherence of uncertainty in ΔS_{cav} and ΔS_{elec} values, the hybrid SPT/FEP method seems to significantly outperform the full FEP method due to the better description of the solute-water van der Waals interaction term that makes much more contribution to ΔS_{hyd} than the electrostatic counterpart. To address this hypothesis, I compared the deviations of the ΔS_{hyd} values of 18 solute molecules calculated with the hybrid SPT/FEP and full FEP methods with respect to the experimental ones. In the full FEP calculations of ΔS_{hyd}, we used the identical method as adopted in the hybrid SPT/FEP approach to warrant a fair comparison. The ΔS_{hyd} values calculated both in forward and in backward FEP simulations are presented in Appendices. The standard deviations for ΔS_{hyd} values range from 0.18 to 5.02 cal·mol^{-1}·K^{-1}, which correspond to

0.3-8.7% with respect to the actual ΔS_{hyd} values. As shown in Table 4.2, the hybrid SPT/FEP method appears to predict the ΔS_{hyd} values of all the solute molecules with even higher accuracy than the full FEP method. The average of deviations amounts to only 2.01 in the former as compared to 20.42 cal·mol^{-1}·K^{-1} in the latter. These comparisons confirm the effectiveness of the separate calculations of van der Waals and electrostatic interaction terms with SPT (or IT) and FEP methods, respectively, in the estimation of molecular hydration entropy.

solute	deviation from experimental data		solute	deviation from experimental data	
	full FEP	hybrid		full FEP	hybrid
1	5.59	2.98	**10**	18.70	2.40
2	10.13	2.86	**11**	32.41	1.23
3	15.05	3.71	**12**	30.76	4.44
4	30.11	2.18	**13**	32.37	1.74
5	14.21	0.45	**14**	7.50	3.21
6	13.99	0.05	**15**	29.04	0.06
7	37.21	0.26	**16**	26.13	4.37
8	26.71	0.01	**17**	16.39	5.37
9	19.30	0.51	**18**	1.95	0.42

Table 4.2. Comparison of the discrepancies of the ΔS_{hyd} values (in cal·mol^{-1}·K^{-1}) obtained with hybrid SPT/FEP and those with full FEP method with respect to the experimental ones. See Figure 4.2 for the identification of solute molecules. Table taken from Choi et al, *J. Chem. Theory Comput. 11*, p.4940. Copyright 2015 American Chemical Society. Reprinted with permission.

Because the numerical instability termed end-point catastrophe is not observed explicitly for 18 representative solute molecules when λ approaches 0 or 1, the large errors in the ΔS_{hyd} values calculated with the full FEP method may be attributed to the fact that the current NAMD force fields are not optimized adequately for ΔS_{hyd} calculations. At these end point, interaction of the reference or the target topology with its environment is

extremely weak, yet still nonzero, which in turn allows the surrounding atoms to clash against the appearing or vanishing chemical moieties. The resulting numerical instabilities cause large fluctuations in the estimated $\langle \Delta U \rangle$. Therefore, the development of the new potential parameters to enhance the computational accuracy would be prerequisite for the full FEP calculations to be useful for ΔS_{hyd} prediction. Because the imperfection of van der Waals interaction term proves to be responsible for the large error of ΔS_{hyd} estimation in this study, the reoptimization of force fields should be focused on the depths of the potential energy wells and the equilibrium separations between the solute and solvent atoms.

4.3.4. Hydration free energy of thermodynamic decomposition

Finally, I turned to the comparison of the ΔG_{hyd} data obtained by combining the calculated ΔS_{hyd} and ΔH_{hyd} values with the corresponding experimental ΔG_{hyd} data. This comparative analysis was performed to validate the usefulness of the ΔS_{hyd} values calculated with the hybrid SPT/FEP method for the enrichment of ΔG_{hyd} data. To compute the ΔH_{hyd} values of 90 molecules listed in Table 4.1, we employed the solvent-contact model suggested by Stouten et al.[56] Within the framework of solvent-contact model, ΔH_{hyd} for a solute molecule can be expressed as follows.

$$\Delta H_{hyd} = \sum_{i}^{atoms} S_i (O_i^{max} - \sum_{j \neq i}^{atoms} V_j e^{\frac{r_{ij}^2}{2\sigma^2}}) \quad (11)$$

Here, the gaussian-type envelope function with respect to the interatomic distance between solute atoms (r_{ij}) and a constant (σ = 3.5 Å) is employed to define the occupied volume to which the approach of solvent molecules is restricted. S_i, O_i^{max}, and V_j represent the atomic hydration energy per unit volume, the maximum atomic occupancy, and the atomic fragmental volume, respectively. These three atomic parameters for all the atom types available in the dataset were optimized with a standard genetic algorithm

using the training set comprising 45 molecules. All details for the optimized atomic parameters and the calculated ΔH_{hyd} values of 90 solute molecules under investigation were provided in Supporting Information (Table A.4.7. and A.4.8.).

The calculated ΔH_{hyd} values were then combined with the ΔS_{hyd} data obtained with the hybrid SPT/FEP method to estimate the molecular ΔG_{hyd} values, which were compared with the corresponding experimental ΔG_{hyd} results. As shown in Figure 4.7, the calculated ΔG_{hyd} values appear to be in good agreement with the experimental ones with the associated R^2 and RMSE values of 0.855 and 1.15 kcal/mol, respectively. The slope and intercept of the linear correlation diagram amount to 0.978 and 0.173, respectively. Judging from such a reasonably good fit, the combination of ΔH_{hyd} and ΔS_{hyd} values calculated respectively with the solvent-contact model and the hybrid SPT/FEP method seems to be useful for estimating the ΔG_{hyd} values of small solute molecules.

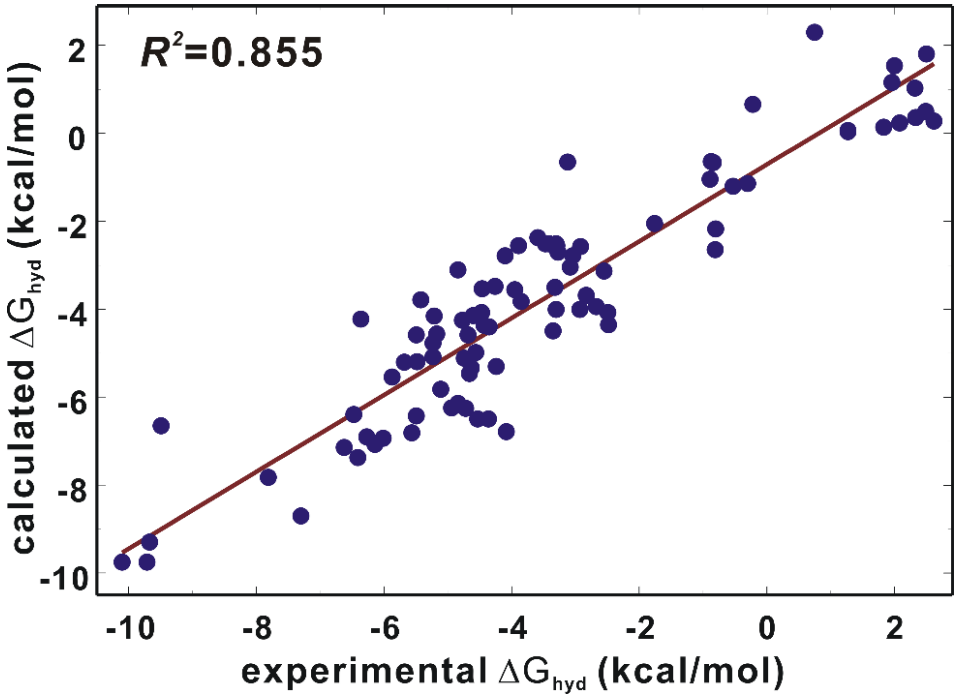


Figure 4.7. Linear correlation diagram between the experimental ΔG_{hyd} values and those obtained by combining the calculated ΔS_{hyd} and ΔH_{hyd} values. Figure taken from Choi et al, *J. Chem. Theory Comput. 11*, p.4940. Copyright 2015 American Chemical Society. Reprinted with permission.

Despite the significant improvement of accuracy in predicting the hydration entropies of organic molecules, we expect that the present hybrid SPT/FEP method can be further modified in several aspects. First, the decomposition of the hydration entropy into a cavity term and an electrostatic term may be a rough approximation due to the complexity in solute-water interactions. This is likely to be the case for large solute molecules whose structures in water are determined by a compromise between van der Waals and electrostatic interactions. The change of the theoretical framework may therefore be required for the accurate estimation of hydration entropies of large solute molecules.

The use of Leonard-Jones potential in SPT calculations instead of the simple hard-sphere model is most likely to lead to the better estimation of the molecular ΔS_{cav} values due to the better description of solute-water interactions. In this case, however, the calculation of ΔS_{cav} seems to be more difficult than in the hard-sphere model because of the lack of the analytical expression for ΔS_{cav}. Extensive atomistic simulations are therefore needed to calculate the various contributions to ΔS_{cav} in the mixtures of Leonard-Jones particles.

It should also be taken into consideration that only the non-bonded electrostatic interactions are evaluated in FEP simulations despite the inherence of covalent character in strong hydrogen bonds. Therefore, the ΔS_{elec} values of the solutes that can form strong hydrogen bond(s) with water are likely to be less negative than the actual values. This problem is expected to be resolved in a large part by the use of hybrid QM/MM method instead of classical MD simulation in FEP calculations because the former can model the electron-transfer interactions between hydrogen-bond donor and acceptor groups. Finally, the use of high-quality atomic charges would also enhance the ΔS_{elec} calculation results because only the electrostatic solute-water interactions are considered in the FEP simulations. Therefore, the atomic charge calculations need to be performed using the sophisticated quantum chemical methods at post-Hartree-Fock level such as

configuration interaction, Møller-Plesset perturbation, and coupled cluster theories. In this case, it should be kept in mind that that the accuracy of our hybrid SPT/FEP method for the estimation of S_{hyd} may be limited by decomposing the potential function whose parameters were optimized simultaneously. To enhance the accuracy of S_{hyd} prediction, therefore, all the potential parameters in SPT and FEP terms need to be reoptimized in such a way to minimize the difference between experimental and calculated S_{hyd} values. Future studies on the hydration entropy would focus on the further improvement of the accuracy under consideration of the above-mentioned points.

4.4. Inclusion of self-solvation terms in hydration free energy

4.4.1 Derivation of self-solvation free energies based on solvent-contact model

The solvent-contact model for molecular solvation is based on several assumptions. First, the solvation free energy (ΔG_{sol}) of a molecule can be approximated by the sum of individual atomic contributions as follows.

$$\Delta G_{sol} = \sum_{i}^{atoms} \Delta G_{sol}^{i} \tag{12}$$

Second, the solvation free energy of an atom i can be given by the product of the atomic solvation parameter (S_i) and its volume exposed to bulk solvent (F_i).

$$\Delta G_{sol}^{i} = S_i F_i \tag{13}$$

Third, the atomic volume exposed to solvent is assumed to be equal to the unoccupied volume around the atom of interest. The occupied volume around the atom i (O_i) indicates the region to which the approach of solvent molecule is forbidden due to the intramolecular interactions with the rest of solute atoms. O_i can be determined by summing over the product of an atomic volume parameter (V_j) representing the fragmental volume of the other atoms and a suitable envelope function, $E(r_{ij})$, with respect to the distance between the centers of atoms i and j.

$$O_i = \sum_{j \neq i}^{atoms} V_j E(r_{ij}) \tag{14}$$

Here, the Gaussian envelope function is used with the variable r_{ij} representing the interatomic distance. Because F_i is the difference between the maximum occupancy of atom i (O_i^{max}) and O_i,[57] the solvation free energy of a molecule can be expressed in the following form.

$$\Delta G_{sol} = \sum_{i}^{atoms} S_i \left(O_i^{\max} - \sum_{j \neq i}^{atoms} V_j e^{-\frac{r_{ij}^2}{2\sigma^2}} \right) \tag{15}$$

Although the solvation free energies of small organic molecules calculated with eq. 15 compared well with experimental results, this solvation model needs to be modified in order to be useful in practical applications because of the neglect of the self-solvation effect. Indeed, it has been demonstrated that the effects of the intramolecular non-bond interactions between solute groups should be reflected in the solvation free energy function to describe the solute-solvent interactions in a quantitative fashion.[58-60] For example, the intramolecular hydrogen bond and van der Waals interactions established in the occupied volume of the solute can affect the strength of the solute-solvent interactions through the change in electron distribution in the outer region exposed to bulk solvent. The presence of this self-solvation effect can be attributed to the intramolecular stabilization/destabilization of the atoms in the solvent-exposed region by the atoms in the occupied volume. The pattern for solute-solvent interactions can thus be affected significantly by the intramolecular interactions that serve as the reason for the charge redistribution at the solute-solvent interface. Therefore, the solvation energy function in eq 15 should be insufficient to fully describe the stabilization of a solute molecule in solution because it contains the solute-solvent interaction term only and lacks the self-solvation term. To define the self-solvation energy of an organic molecule in

solution, we assume that it can be obtained by the summation of all individual atomic contributions. The self-solvation energy of an atom *i* (ΔG^{i}_{self}) can then be approximated by the product of the atomic self-solvation parameter (P_i) and the occupied volume around the atom *i* as follows.

$$\Delta G^{i}_{self} = P_i O_i \tag{16}$$

Here, P_i describes the stabilization energy of solute atom *i* per unit volume due to the intramolecular interactions with the other atoms of the solute. This self-solvation term can be appended to the solvation term in eq 15 to obtain the improved solvation free energy function for organic molecules, which can be expressed in the following form.

$$\Delta G_{sol} = \sum_{i}^{atoms} \left\{ S_i \left(O_i^{\max} - \sum_{j \neq i}^{atoms} V_j e^{-\frac{r_{ij}^2}{2\sigma^2}} \right) + P_i \sum_{j \neq i}^{atoms} V_j e^{-\frac{r_{ij}^2}{2\sigma^2}} \right\} \tag{17}$$

The first and second terms in eq 17 correspond to the contribution from solute-solvent interactions and that from the intramolecular interactions between solute atoms to the stabilization of the solute molecule in solution, respectively. This new solvation free energy function places an emphasis on the fact that organic solute molecules can be stabilized in solution as a consequence of the coordination between the solute-solvent interactions and the structural stabilization of the solute in itself. The four key atomic parameters in the present solvation model include the maximum atomic occupancy (O_i^{max}), the atomic fragmental volume (V_i), and the atomic solvation (S_i) and self-solvation (P_i) energies per unit volume. The extent of contribution from the intramolecular interactions to molecular solvation free energy is assumed to be determined by P_i parameters in the present solvation model. Thus, four different atomic parameters should be optimized for all possible atom types to complete the solvation free energy function.

4.4.2. Optimization of atomic volume parameters with genetic algorithm.

Three kinds of atomic parameters need to be optimized in order to calculate the solvation free energy of a molecule based on the solvent contact model. Among them, the atomic volume parameter V_j represents the fragmental volume of atom *j* in a molecule. Because the V_j values exhibited a bad convergent behavior in the simultaneous optimization of the three kinds of parameters, they were optimized with the operation of an independent genetic algorithm as detailed below.

The total volume of a molecule should be determined prior to the parametrization of V_j values. For this purpose, each molecule in the training and the test sets was placed in a 3-D box whose length, width, and height correspond to the maximum distances along the three axes defining the coordinate system of the van der Waals volume of the molecule. Monte Carlo simulations involving random selections of a point in the predefined 3-D box were then carried out to calculate the total volume of the molecule (V_{mol}) embedded in the box. In this simulation, V_{mol} could be obtained by the volume of the box (V_{box}) multiplied by the ratio of the number of trials to select a point in the molecular van der Waals volume (N_{hits}) to the total number of trials (N_{trials}). Thus, we have

$$V_{mol} = V_{box} \times \frac{N_{hits}}{N_{trials}}. \tag{18}$$

With the calculated V_{mol} values in hand, the atomic volume parameters were optimized with the standard genetic algorithm. A generation was defined with 100 vectors comprising the V_j parameters, followed by the removal of 50 with a bias toward preserving the most fit with the lowest error. The empty 50 vectors were then filled with point mutations to alter the value of one of the parameters with probability 0.01, and with cross breeds with probability 0.6 to select some parameters from one vector to replace the elements of another vector of the top 50. The 50 newly created vectors were then

evaluated together with the top 50. This cycle was repeated as many times as desired. In the evaluation of the 100 vectors, we used a gradient-based minimization method on the error hypersurface (F_V). This hypersurface is defined by the sum of the absolute values of the differences between the calculated V_{mol} value of a molecule and the sum of V_j values in the molecule.

$$F_V = \sum_{k}^{molecules} \left| V_{mol}^{k} - \sum_{j}^{atoms} V_j \right| . \qquad (19)$$

4.4.3. Correlation for calculated self-solvation free energies vs. experimental solvation free energies

Figure 4.8 illustrates the correlations between the experimental solvation free energies and those calculated using eq 17. The parameters, molecular structures, experimental and calculated solvation free energies of all 404 molecules are shown in Appenidix Section (Table A.4.9., A.4.10. and A.4.11.). We note that the R^2 values for training and test sets show 0.88 and 0.85 in the present solvation model, respectively. Apparently, these improvements should be attributed to the inclusion of the self-solvation term in the solvation free energy function and to the extension of atomic parameter space to cope with various chemical environments in large molecules. The accuracy of the present solvation model is actually comparable to those of the quantum chemical dielectric continuum solvation model (COSMO)[61] and molecular dynamics free energy perturbation method.[62] These comparisons indicate that our GA-based solvent-contact model would be more efficient in estimating the molecular solvation free energies than the sophisticated quantum chemical method and statistical simulations with all-atom models because the former can produce the energy values from a simplified potential function. In comparison with the results for the previous solvation model that neglected the self-solvation effect, the largest improvements in solvation free energies are observed

for the molecules in which the strong intramolecular hydrogen bonds or intramolecular van der Waals contacts can be established. This substantial improvement further exemplifies the importance of intramolecular interactions in the stabilization of organic molecules in solution, which was also proposed for the structural stability of proteins in solution.[63] The earlier solvation models such as solvent accessible surface area, hydrophobicity scales, and group additivity models proved to be insufficient to explain the solvation properties of organic molecules due to the assumption that the stability of solutes in solution should be determined by the extent of the solvent-exposed regions.[64] This linearity criterion for solute-solvent interactions is actually inapplicable to large organic molecules because their buried regions can also contribute to the structural stability in solution. Our modified solvent-contact model confirmed that the solvation free energies of large organic molecules could be estimated with reasonable accuracy by combining the contributions from the solvent-exposed and self-solvation regions, the relative importance of which should be dependent on the molecular conformations in solution. The significant contribution of the self-solvation term to solvation free energies of organic molecules is thus consistent with the nonlinearity in solute-solvent interactions that stems from the intramolecular interactions between solute atoms in solution. Despite the improved accuracy in estimating the solvation free energies of organic molecules, some problems still remain to be solved for the present solvation model to be employed extensively in practical applications. First, the atomic parameters of some atom types such as the cationic carbon and the hydrogen attached to sulfur atom could not be determined in this study due to the lack of corresponding experimental data. The atomic parameter space needs to be extended to cover a more variety of atom types in molecules when the more experimental data for molecular solvation free energies will be available in the future. Second, conformational diversity of organic molecules should be

considered in the parameterization because the volumes of solvent-exposed and buried regions can vary with the conformational changes. For this purpose, molecular dynamics or Monte Carlo simulations can be applied prior to the parameterization to collect various local structural minima of the solutes.

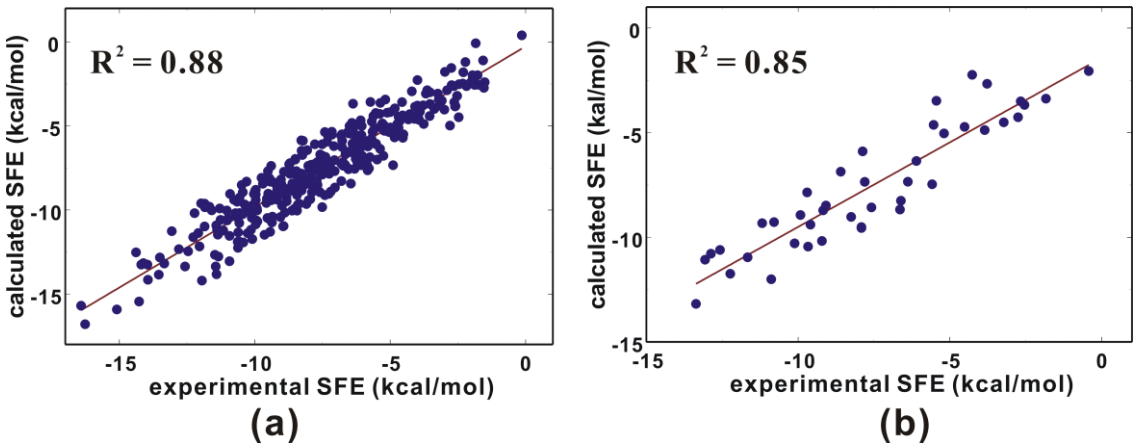


Figure 4.8. Correlation diagrams for experimental solvation free energies (SFEs) versus those obtained with the solvation free energy function with self-solvation term for (a) 362 molecules in the training set and (b) 42 molecules in the test set. Figure taken from Choi et al, *J. Cheminfomatics 5*, p.17 Copyright 2013 Springer. Reprinted with permission.

Finally, the solvation free energy function needs to be decomposed into enthalpy and entropy terms. Because both thermodynamic quantities are experimentally accessible, the potential parameters in the enthalpic and entropic terms can be optimized independently using their respective corresponding experimental data. Apparently, this dual parameterization warrants the better correlation between the experimental and computational solvation free energies than the single parameterization because more diverse experimental data can be included in reference dataset. Because the sign of solvation free energy is determined by the combination of enthalpic and entropic contributions, the decomposition analysis of solvation free energy can also provide thermodynamic insight into the solvation mechanism. Future studies for solvation can

focus on further improvement in the accuracy of solvation free energy function with the three above-mentioned points kept in mind.

4.5. Application of hydration free energies to virtual screening

4.5.1. BRAF kinase inhibitors

VRAF murine sarcoma viral oncogene homologue B1 (BRAF) is a serine-threonine kinase and a component of the Ras/Raf/MEK/ERK signaling pathway. This pathway normally regulates cell growth, division and differentiation, and has long been associated with human cancers due to the frequent oncogenic mutations identified in rapidly growing fibrosarcoma (RAF) family members. An activating mutation in the gene encoding BRAF is known to be responsible for 40–60% of melanomas.[65] Most of the reported BRAF mutants involve the substitution of glutamate for valine at amino acid residue number 600. This V600E mutation constitutively activates BRAF and downstream signal transduction in the MAPK (mitogen-activated protein kinase) signaling pathway.[66] This mutation-induced overactivation of BRAF has also been identified in a variety of other human cancers such as ovarian,[67] colorectal,[68] and thyroid carcinomas.[69] Therefore, the inhibition of BRAF kinase activity has been considered to be a promising therapeutic strategy for the treatment of cancer.[70]

I applied the methods to identify novel classes of BRAF inhibitors by means of a structure-based drug design protocol involving the virtual screening with docking simulations and in vitro enzyme assay. Virtual screening has not always been successful due to the inaccuracy in the scoring function, which leads to a weak correlation between the enrichment in virtual screening and binding mode prediction.[71] The characteristic feature that discriminates our virtual screening approach from the others lies in the implementation of an accurate solvation model in calculating the binding free energy

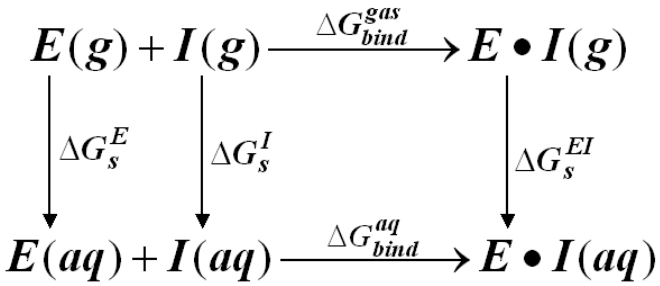
between BRAF and the putative ligands, which has been known to cause the increase in the accuracy of the scoring function of docking programs.[72] The results show that the docking simulations with the improved binding free energy function can be a useful tool for elucidating the activities of the identified inhibitors, as well as for enriching the chemical library with molecules that are likely to have desired biological activities.

4.5.2. High throughput Docking Simulations

Three dimensional (3D) coordinates of the X-ray crystal structure of BRAF in complex with a potent inhibitor (PDB code: 2FB8)[73] were selected as the receptor model in the virtual screening with docking simulations. After removing the solvent molecules, hydrogen atoms were added to each protein atom. Special attention was paid to assign the protonation states of the ionizable Asp, Glu, His, and Lys residues in the X-ray structure of BRAF. The side chains of Asp and Glu residues were assumed to be neutral if one of their carboxylate oxygens pointed toward a hydrogen-bond accepting group including the backbone aminocarbonyl oxygen at a distance within 3.5 Å, a generally accepted distance limit for a hydrogen bond of moderate strength.[74] Similarly, the lysine side chains were assumed to be protonated unless the NZ atom was in proximity of a hydrogen-bond donating group. The same procedure was also applied to determine the protonation states of ND and NE atoms in His residues.

The docking library for BRAF comprising about 240,000 compounds was constructed from the latest version of the chemical database distributed by Interbioscreen (http://www.ibscreen.com) containing approximately 477,000 synthetic and natural compounds. Prior to the virtual screening with docking simulations, they were filtrated on the basis of Lipinski's "Rule of Five" to adopt only the compounds with the physicochemical properties of potential drug candidates[75] and without reactive functional group(s). All of the compounds included in the docking library were then subjected to

the Corina program to generate their 3D atomic coordinates, followed by the assignment of Gasteiger-Marsilli atomic charges.[76] We used the AutoDock program[77] in the virtual screening of BRAF inhibitors because the outperformance of its scoring function over those of the others had been shown in several target proteins.[78] The AMBER force field parameters were assigned for calculating the van der Waals interactions and the internal energy of a ligand as implemented in the original AutoDock program. Docking simulations with AutoDock were then carried out in the ATP-binding site of BRAF to score and rank the compounds in the docking library according to their calculated binding affinities.



$$\Delta G_{bind}^{aq} = \Delta G_{bind}^{gas} + \Delta G_{s}^{EI} - \Delta G_{s}^{E} - \Delta G_{s}^{I}$$

$$\Delta G_{bind}^{aq} \cong \Delta G_{bind}^{gas} - \Delta G_{s}^{I}$$

Figure 4.9. Ligand solvation energy (ΔG_s^I) of Thermodynamic cycle of protein-ligand. Enzyme-Ligand binding energy in aqueous state (ΔG_{bind}^{aq}), Enzyme-Ligand binding energy in vacuum state (ΔG_{bind}^{gas}), Enzyme solvation energy (ΔG_s^E) and Enzyme-Ligand complex solvation energy (ΔG_s^{EI}), respectively.

Most of biological processes occur in aqueous solution, ligand-solvent interactions contribute deeply to binding energies. The estimation of solvation free energy with reasonable accuracy should thus be prerequisite for exact binding energies. As shown in Figure 4.9, ligands bind to protein pocket from bulk solvent to nearly vacuum state of protein pocket, this desolvation cost should be considered. Most protein-ligand free energies generally include about 5~20kcal/mol during protein-ligand complexation. [79]

In the actual docking simulation of the compounds in the docking library, I used the empirical AutoDock scoring function improved by the implementation of a new solvation model for a compound. The modified scoring function has the following form:

$$\Delta G_{bind}^{aq} = W_{vdW}\sum_{i=1}\sum_{j>i}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+W_{hbond}\sum_{i=1}\sum_{j>i}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)$$
$$+W_{elec}\sum_{i=1}\sum_{j>i}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+W_{tor}N_{tor}+W_{sol}\sum_{i=1}S_i\left(Occ_i^{\max}-\sum_{j>i}V_j e^{-\frac{r_{ij}^2}{2\sigma^2}}\right), \qquad (20)$$

where W_{vdW}, W_{hbond}, W_{elec}, W_{tor}, and W_{sol} are the weighting factors of van der Waals, hydrogen bond, electrostatic interactions, torsional term, and desolvation energy of inhibitors, respectively. r_{ij} represents the interatomic distance, and A_{ij}, B_{ij}, C_{ij}, and D_{ij} are related to the depths of the potential energy well and the equilibrium separations between the two atoms. The hydrogen bond term has an additional weighting factor, $E(t)$, representing the angle-dependent directionality. The cubic equation approach was applied to obtain the dielectric constant required in computing the interatomic electrostatic interactions between BRAF and a ligand molecule.[80] In the entropic term, N_{tor} is the number of all rotatable bonds in the ligand. In the desolvation term, S_i and V_i are the solvation parameter and the fragmental volume of atom i,[56] respectively, while $Occ_i^{\max}$ stands for the maximum atomic occupancy. In the calculation of molecular solvation free energy term in Eq. (20), we used the atomic parameters developed by Kang *et al.*[57] because those of the atoms other than carbon were unavailable in the current version of the AutoDock program. This modification of the solvation free energy term is expected to increase the accuracy in virtual screening because the underestimation of ligand solvation often leads to the overestimation of the binding affinity of a ligand with many polar atoms.[72]

4.5.3. Docking Simulation results with experimental confirmation.

The docking simulation of a compound in the docking library started with 3D grid calculations of the interaction energy for all of the possible atom types present in chemical database. These uniquely defined potential grids for the receptor protein were then used in common for docking simulations of all compounds in the docking library. As the center of the common grids, we used the center of mass coordinates of the inhibitor SB-590885 bound in the ATP-binding site of BRAF.[73] The calculated grid maps were of dimension 61 × 61 × 61 points with the spacing of 0.375 Å, yielding a receptor model that includes atoms within 22.9 Å of the grid center. For each compound in the docking library, 10 docking runs were performed with the initial population of 50 individuals. The maximum number of generations and energy evaluation were set to 27,000 and 2.5 × 10^5, respectively. Docking simulations with the AutoDock program were then carried out in the ATP-binding site of BRAF to score and rank the compounds in the docking library.

Of the 240,000 compounds subject to the virtual screening with docking simulations, the 200 top-scored compounds were selected as virtual hits. 184 of them were available from the compound supplier and were tested over BRAF at 50 μM concentration in a high-throughput binding assay (KINOMEscan, Ambit Biosciences).[81] Six compounds were discovered as hits with percent of control (POC) values less then 60, and were selected for the determination of K_d values. The chemical structures and the inhibitory activities of the newly identified inhibitors are shown in Figure 4.10 and Table 4.3, respectively. It is noted that compounds **1** and **2** possess 9H-purin and phthalazine groups, respectively. These two-membered rings seem to serve as an effective mimic for the adenine group of ATP because **1** and **2** are found to be potent inhibitors with the associated K_d values of 0.7 and 2.1 μM, respectively. To the best of my knowledge, compounds **1-6** have not

been reported as BRAF inhibitors so far. As can be seen in Table 4.3, six inhibitors reveal a high potency against BRAF with the K_d values ranging from 0.7 to 20 μM. Because these inhibitors are structurally diverse with good physicochemical properties as a drug candidate, they deserve consideration for further development by structure-activity relationship (SAR) studies to optimize the inhibitory activities to develop a new anticancer medicine.

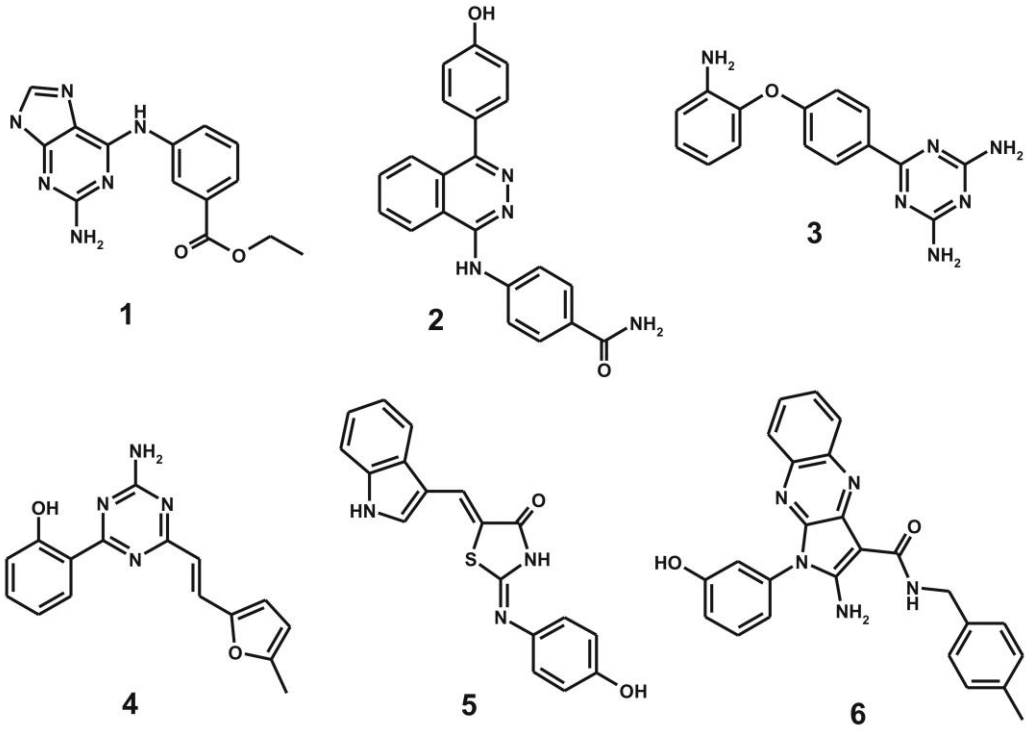


Figure 4.10. Chemical structures of the newly identified BRAF inhibitors**.**

compounds	POC (50 μM)	K_d (μM)
1	15	0.7
2	12	2.1
3	52	11.8
4	60	20.0
5	60	18.5
6	49	12.6

Table 4.3. The inhibitory activities of compounds 1-6 against BRAF. Only the compounds with POC < 60 are shown

To obtain structural insights into the inhibitory mechanisms of the identified BRAF inhibitors, their binding modes in the ATP-binding site were investigated. The results of these docking simulations are self-consistent because the functional groups of similar chemical character are found to be placed in similar ways with comparable interactions with the protein groups. For example, the hydrogen-bond donating groups of the inhibitors point toward the backbone groups at the ATP-binding site of BRAF with their terminal hydrophobic groups being situated between the Gly loop and another loop at the top of C-terminal domain. In order to examine the possibility of the allosteric inhibition of BRAF by the identified inhibitors, docking simulations were carried out with the grid maps for the receptor model so as to include the entire part of BRAF. However, the binding configuration in which an inhibitor resides outside the ATP-binding site was not observed for any of the new inhibitors. These results support the possibility that the inhibitors would impair the activity of BRAF through the specific binding in the ATP-binding site.

4.5.3. Analysis of binding modes

The calculated binding mode of **1** in the ATP-binding site of BRAF is shown in Figure 4.11. It should be noted that the amino group attached to the inhibitor purine ring forms the bifurcated hydrogen bonds with the two backbone aminocarbonyl oxygens of Ala481 and Ile527. A stable hydrogen bond is also established between the carbonyl oxygen of the inhibitor ester group and the backbone amidic nitrogen of Cys532. These three hydrogen bonds seem apparently to play a role of anchor for binding of **1** in the middle of the ATP-binding site. The inhibitor **1** can be further stabilized in the ATP-binding site by hydrophobic interactions of its nonpolar groups with the side chains of Ala481, Leu514, Trp531, and Phe583. Judging from the overall structural features derived from docking simulations, the submicromolar inhibitory activity of **1** is likely to stem from the

multiple hydrogen bonds and hydrophobic interactions established simultaneously in the ATP-binding site. Considering the low molecular weight of ~298, **1** could serve as a new inhibitor scaffold from which much more potent inhibitors can be developed.

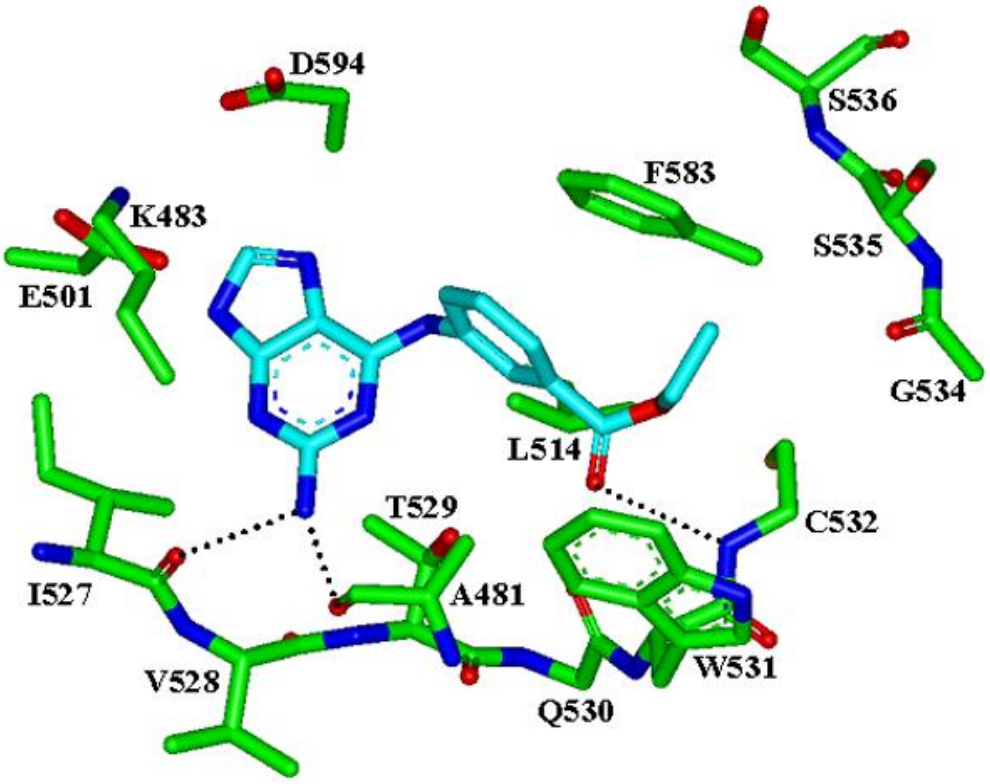


Figure 4.11. Calculated binding mode of 1 in the ATP-binding site of BRAF. Carbon atoms of the protein and the ligand are indicated in green and cyan, respectively. Each dotted line indicates a hydrogen bond.

Figure 4.12 shows the lowest-energy binding mode of **2** in the ATP-binding site of BRAF. It is seen that the terminal amidic nitrogen donates a hydrogen bond to the backbone aminocarbonyl oxygen of Ala481. The formation of a hydrogen bond with backbone amide group of Ala481 appears to be a common feature in the binding modes of all six inhibitors found in this study. The terminal phenolic group of **2** also forms a hydrogen bond with the backbone aminocarbonyl oxygen of Gly534, which should also be a significant binding force that stabilizes **2** in the ATP-binding site of BRAF. It is also noteworthy that the number of hydrogen bonds decreases from three in the BRAF-**1** to two in the BRAF-**2** complex, which would have an effect of lowering the binding affinity.

Hydrophobic interactions in the BRAF-**2** complex are established in a similar way to those in the BRAF-**1** complex: its phenyl and phthalazine rings are stabilized with the interactions with the side chains of Ala481, Leu514, Trp531, and Phe583. Two additional residues on Gly loops, Ile463 and Val471, are also involved in the stabilization of **2** at the ATP-binding site. This indicates that the hydrophobic interactions in BRAF-**2** complex should be stronger than those in BRAF-**1** complex. Thus, the loss of one hydrogen bond seems not to be compensated for effectively by establishing the stronger hydrophobic interactions, which can be an explanation for the three-fold lower inhibitory activity of **2** than **1** (Table 4.3). To gain some insight into improving the inhibitory activities of **1** and **2**, I compared their calculated the binding modes in the ATP-binding site of BRAF with that of SB-590885 in the X-ray crystal structure. This comparative analysis would be informative for making a strategy for lead optimization because SB-590885 is known as a potent and selective BRAF inhibitor with the associated K_d value of 0.15 nM.[73] As can be seen in Figure 4.13, **1** and **2** seem to be stabilized in the ATP-binding site in a similar way to the substructure of SB-590885 that includes three aromatic rings.

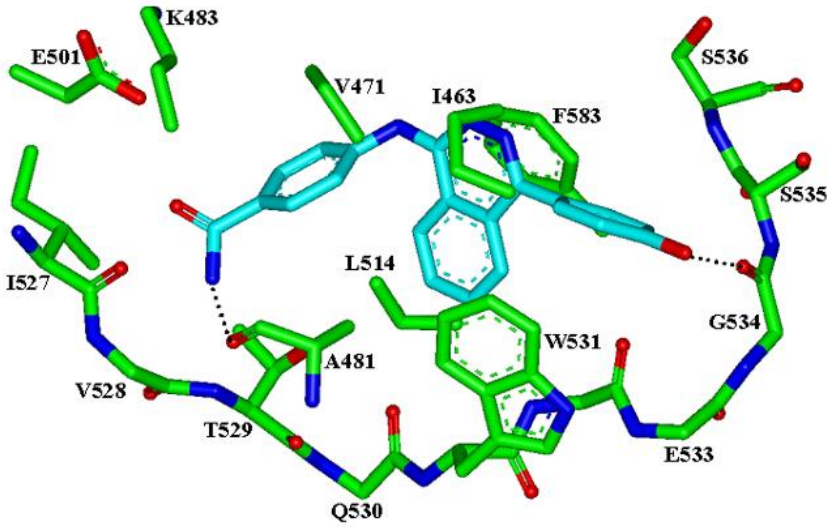


Figure 4.12. Calculated binding mode of **2** in the ATP-binding site of BRAF. Carbon atoms of the protein and the ligand are indicated in green and cyan, respectively. Each dotted line indicates a hydrogen bond.

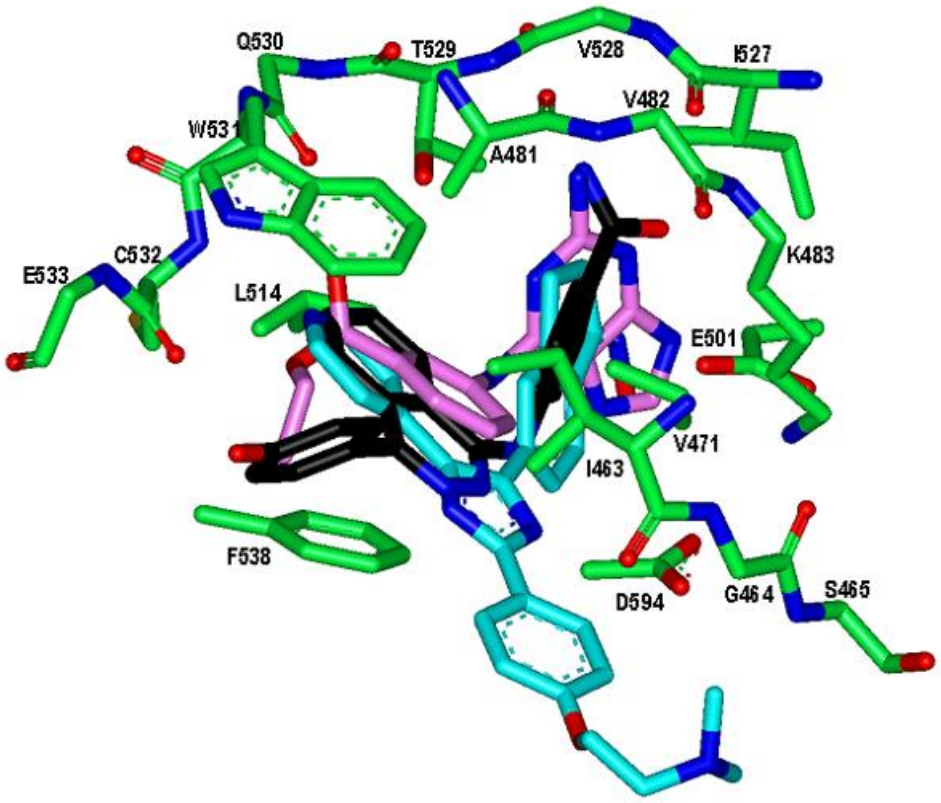


Figure 4.13. Comparative view of the calculated binding modes of **1** and **2** in the ATP-binding site of BRAF and that of SB-590885 in the X-ray crystal. Carbon atoms of SB-590885, **1**, and **2** are indicated in cyan, pink, and black, respectively.

In summary, the results identify six novel inhibitors of BRAF by applying a computer-aided drug design protocol involving the structure-based virtual screening with docking simulations under consideration of the effects of ligand solvation in the scoring function. These inhibitors have desirable physicochemical properties as candidates to develop new potent inhibitors and manifest a high potency with K_d values ranging from 0.7 to 20 μM. Therefore, each of the newly discovered inhibitors deserves consideration for further development by SAR studies to optimize the inhibitory activities. Detailed binding mode analyses with docking simulations show that the inhibitors can be stabilized in ATP-binding site by the simultaneous establishment of multiple hydrogen bonds and van der Waals contacts.

Due to a significant increase in molecular size, however, SB-590885 interacts with more amino acid residues than **1** and **2** including Gly464 and Ser465, which appear to stabilize

the terminal tertiary amine group of SB-590885. It is noted that these residues are located in the Gly loop that is also known as activation loop in the case of BRAF.[73] Judging from the increase in the inhibitory activity by a factor of $10^3 – 10^4$ in going from **1** and **2** to SB-590885, derivatization for increasing the potency should be designed in such a way to maximize the interactions with the residues in the activation loop.

4.6. Conclusions

Molecular hydration entropies could be predicted with reasonable accuracy by means of the hybrid SPT/FEP method to separately calculate the cavity formation entropy (ΔS_{cav}) as the hydrophobic contribution and electrostatic solute-water interaction term (ΔS_{elec}). Due to the effective partitioning of the two entropic components, the prediction accuracy of ΔS_{hyd} is enhanced to a substantial extent when compared to the conventional full FEP calculations. This improvement was made possible by successfully overcoming the difficulty in calculating ΔS_{vdw} in full FEP calculations. SPT and IT exhibited almost the same performance in the estimation of ΔS_{cav}. ΔS_{cav} was found to dominate over ΔS_{elec} in magnitude not only for hydrocarbons but also for polar solutes. This implies that most of the entropic cost for the hydration of a solute is spent for the formation of solvent-excluded volume rather than for the direct solute-water interactions. The entropic features of hydration found in this study could serve as useful information in drug discovery. In order to optimize the binding affinity of a ligand for the target protein, for example, the ligand structure should be modified in such a way to maximize the strength of attractive protein-ligand interactions and simultaneously to maximize the entropic penalty for cavity formation in water to minimize the desolvation cost for protein-ligand binding.

Chapter 4 References

. Mobley, D. L., Wymer, K. L., Lim, N. M., Guthrie, J. P. Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput. Aided Mol. Des.* **28**, 135–150 (2014).

2. Thompson, J. D., Cramer, C. J., Truhlar, D. G. Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances. *J. Chem. Phys.* **119**, 1661–1670 (2003).
3. Hagen, D. E., Salk, S. H. S. Entropy and enthalpy change in hydrated ion clusters. *Chem. Phys.* **102**, 459–466 (1986).
4. Ahmad, M., Helms, V., Lengauer, T., Kalinina, O. V. Enthalpy–Entropy Compensation upon Molecular Conformational Changes. *J. Chem. Theory Comput.* **11**, 1410−1418 (2015).
5. Reynolds, C. H., Holloway, M. K. Thermodynamics of ligand binding and efficiency. *ACS Med. Chem. Lett.* **2**, 433–437(2011).
6. Olsson, T. S., Williams, M. A., Pitt, W. R., Ladbury, J. E. The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. *J. Mol. Biol.* **384**, 1002–1017 (2008).
7. Kavanagh, K. L., Guo, K., Dunford, J. E., Wu, X., Knapp, S., Ebetino, F. H., Rogers, M. J., Russell, R. G. G., Oppermann, U. The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *Proc. Natl. Acad. Sci. USA* **103**, 7829–7834 (2006).
8. Kaul, M., Barbieri, C. M., Pilch, D. S. Defining the basis for the specificity of aminoglycoside-rRNA recognition: a comparative study of drug binding to the A sites of Escherichia coli and human rRNA. *J. Mol. Biol.* **346**, 119–134 (2005).
9. Carbonell, T., Freire, E. Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry* **44**, 11741–11748 (2005).
10. Yanchunas, J., Jr., Langley, D. R., Tao, L., Rose, R. E., Friborg, J., Colonno, R. J., Doyle, M. L. Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrob. Agents Chemother.* **49**, 3825–3832 (2005).
11. Ben-Naim, A., Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **81**, 2016–2027 (1984).
12. Giesen, D. J., Cramer, C. J., Truhlar, D. G. Entropic contributions to free energies of solvation. *J. Phys. Chem.* **98**, 4141–4147 (1994).
13. Florian, J., Warshel, A. Calculations of hydration entropies of hydrophobic, polar, and ionic solutes in the framework of the Langevin dipoles solvation model. *J. Phys. Chem. B* **103**, 10282–10288 (1999).
14. Wyczalkowski, M. A., Vitalis, A., Pappu, R. V. New estimators for calculating solvation entropy and enthalpy and comparative assessments of their accuracy and precision. *J. Phys. Chem. B* **114**, 8166–8180 (2010).
15. Jedlovszky, P., Idrissi, A. Hydration free energy difference of acetone, acetamide, and urea. *J. Chem. Phys.* **129**, 164501 (2008).
16. Gallicchio, E., Kubo, M. M., Levy, R. M. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B* **104**, 6271–6285 (2000).
17. Kubo, M. M., Gallicchio, E., Levy, R. M. Thermodynamic decomposition of hydration free energies by computer simulation: application to amines, oxides, and sulfides. *J. Phys. Chem. B* **101**, 10527–10534 (1997).
18. Gerogiokas, G., Calabro, G., Henchman, R. H., Southey, M. W. Y., Law, R. J., Michel, J. Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *J. Chem. Theory Comput.* **10**, 35–38 (2014).
19. Pitera, J. W., van Gunsteren, W. F. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simul.* **28**, 45–65 (2002).
20. Blondel, A. Ensemble variance in free energy calculations by thermodynamic integration: theory, optimal “alchemical” path, and practical solutions. *J. Comput. Chem.* **25**, 985–993 (2004).
21. Genheden, S., Kongsted, J., Söderhjelm, P., Ryde, U. Nonpolar Solvation Free Energies of Protein− Ligand Complexes. *J. Chem. Theory Comput.* **6**, 3558–3568 (2010).
22. Gapsys, V., Seeliger, D., de Groot, B. L. New soft-core potential function for molecular dynamics based alchemical free energy calculations. *J. Chem. Theory Comput.* **8**, 2373−2382 (2012).
23. Reiss, H., Frisch, H. L., Lebowitz, J. L. Statistical mechanics of rigid spheres. *J. Chem. Phys.* **31**, 369–380 (1959).
24. Pierotti, R. A. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.* **76**, 717–726 (1976).

25. Ashbaugh, H. S., Pratt, L. R. Colloquium: Scaled particle theory and the length scales of hydrophobicity. R*ev. Mod. Phys.* **78**, 159–178 (2006).
26. Floris, F. M., Selmi, M., Tani, A., Tomasi, J. Free energy and entropy for inserting cavities in water: comparison of Monte Carlo simulation and scaled particle theory results. *J. Chem. Phys.* **107**, 6353–6365 (1997).
27. Lebowitz, J. L., Helfand, E., Praestgaard, E. Scaled particle theory of fluid mixtures. *J. Chem. Phys.* **43**, 774–779 (1965).
28. Richards, F. M. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
29. Mobley, D. L., Dill, K. A., Chodera, J. D. Treating entropy and conformational changes in implicit solvent simulations of small molecules. *J. Phys. Chem. B* **112**, 938–946 (2008).
30. Petitjean, M. On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects. *J. Comput. Chem*. **15**, 507–523 (1994).
31. Hummer, G., Garde, S., Garcia, A. E., Pohorille, A., Pratt, L. R. An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci. USA* **93**, 8951–8955 (1996).
32. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
33. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
34. Powell, M. J. D. A Hybrid Method for Nonlinear Equations. In *Numerical Methods for Nonlinear Algebraic Equations*; Rabinowitz, P., Ed.; Gordon and Breach, 1970; pp 87–114.
35. Yu, H. A., Karplus, M. A thermodynamic analysis of solvation. *J. Chem. Phys*. **89**, 2366–2379 (1988).
36. Levy, R. M., Gallicchio, E. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* **49**, 531–567 (1998).
37. Wan, S., Stote, R. H., Karplus, M. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* **121**, 9539–9548 (2004).
38. Smith, D. E., Haymet, A. D. J. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *J. Chem. Phys*. **98**, 6445–6454 (1993).
39. Peter, C., Oostenbrink, C., van Dorp, A., van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. *J. Chem. Phys*. **120**, 2652–2661 (2004).
40. Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* **22**, 1420-1426 (1954).
41. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E.; Guvench, O., Lopes, P., Vorobyov, I., Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
42. Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem*. **30**, 1545–1615 (2009).
43. Hess, B., van der Vegt, N. F. A. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B* **110**, 17616–17626 (2006).
44. Bruckner, S., Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *J. Comput. Chem.* **32**, 1303–1319 (2011).
45. Pohorille, A., Jarzynski, A., Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **114**, 10235–10253 (2010).
46. Zacharias, M., Straatsma, T. P., McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **100**, 9025–9031 (1994).
47. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R., van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **222**, 529–539 (1994)

48. Cabani, S., Gianni, P., Vincenzo, M., Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solution Chem*. **10**, 563–595 (1981).
49. Haynes, W. M. *CRC Handbook of Chemistry and Physics*; CRC Press, Boca Raton, FL, 2012.
50. Berne, B. J. Inferring the hydrophobic interaction from the properties of neat water. *Proc. Natl. Acad. Sci. USA* **93**, 8800–8803 (1996).
51. Madan, B., Lee, B. Role of hydrogen bonds in hydrophobicity: the free energy of cavity formation in water models with and without the hydrogen bonds. *Biophys. Chem.* **51**, 279–289 (1994).
52. Durell, S. R., Wallqvist, A. Atomic-scale analysis of the solvation thermodynamics of hydrophobic hydration. *Biophys. J.* **71**, 1695–1706 (1996).
53. Mobley, D. L., Baker, J. R., Barber, A. E., II, Fennell, C. J., Dill, K. A. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B* **112**, 2405–2414 (2008).
54. Graziano, G. Hydration thermodynamics of N-methylacetamide. *J. Phys. Soc. Jpn.* **69**, 3720–3725 (2000).
55. Godec, A., Smith, J. C., Merzel, F. Increase of both order and disorder in the first hydration shell with increasing solute polarity. *Phys. Rev. Lett.* **107**, 267801 (2011).
56. Stouten, P. F. W., Frömmel, C., Nakamura, H., Sander, C. An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.* **10**, 97–120 (1993).
57. Kang, H., Choi, H., Park, H. Prediction of molecular solvation free energy based on the optimization of atomic solvation parameters with genetic algorithm. *J. Chem. Inf. Model.* **47**, 509–514 (2007).
58. König, G.; Boresch, S. Hydration free energies of amino acids: Why side chain analog data are not enough. *J. Phys. Chem. B.* **113**, 8967-8974 (2009).
59. Chang, J.; Lenhoff, A. M.; Sandler, S. I. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B.* **111**, 2098–2106 (2007).
60. Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5024 (1996).
61. Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the free energy of hydration of a challenging set of pesticide-like compounds. *J. Phys. Chem. B* **113**, 4508–4510 (2009).
62. Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).
63. Baldwin, R. L. Desolvation penalty for burying hydrogen-bonded peptide groups in protein folding. *J. Phys. Chem. B* **114**, 16223–16227 (2010).
64. Gromiha, M. M.; Selvaraj, S. Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol.* **86**, 235–277 (2004).
65. Flaherty, K.T.; Puzanov, I.; Kim, K.B.; Ribas, A.; McArthur, G.A.; Sosman, J.A.; O'Dwyer, P.J.; Lee, R.J.; Grippo, J.F.; Nolop, K.; Chapman, P.B. Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med*, **363**, 809-819 (2010).
66. Niault, T. S.; Baccarini, M. Targets of Raf in tumorigenesis. *Carcinogenesis*, **31**, 1165-1174 (2010).
67. Singer, G.; Oldt, R.; Cohen, Y.; Wang, B. G.; Sidransky, D.; Kurman, R. J.; Shih, I. M. Mutations in BRAF and KRAS characterize the development of low-grade ovarian serous carcinoma. *J. Natl. Cancer Inst.* **95**, 484-486 (2003).
68. Tol, J.; Nagtegaal, I. D.; Punt, C. J. BRAF mutation in metastatic colorectal cancer. *N. Engl. J. Med.*, **361**, 98-99 (2009).
69. Cohen, Y.; Xing, M.; Mambo, E.; Guo, Z.; Wu, G.; Trink, B.; Sidransky, D. BRAF mutation in papillary thyroid carcinoma. *J. Natl. Cancer Inst.* **95**, 625-627 (2003).
70. Dienstmann, R., & Tabernero, J. BRAF as a target for cancer therapy. *Anti-Cancer Agents Med. Chem.* **11**, 285-295 (2011).
71. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **49**, 5912-5931 (2006).
72. Park, H.; Jeon, J.-Y.; Kim, S. Y.; Jeong, D. G.; Ryu, S. E. Identification of novel inhibitors of mitogen-activated protein kinase phosphatase-1 with structure-based virtual screening. *J. Comput. Aided Mol. Des.* **25**, 469-475 (2011).
73. King, A. J.; Patrick, D. R.; Batorsky, R. S.; Ho, M. L.; Do, H. T.; Zhang, S. Y.; Kumar, R.; Rusnak, D. W.; Takle, A. K.; Wilson, D. M.; Hugger, E.; Wang, L.; Karreth, F.; Lougheed, J. C.; Lee, J.; Chau, D.; Stout, T. J.; May, E. W.; Rominger, C. M.; Schaber, M. D.; Luo, L.; Lakdawala, A. S.; Adams, J. L.; Contractor, R. G.; Smalley, K. S. M.; Herlyn, M.; Morrissey, M. M.; Tuveson, D. A.; Huang, P. S.

Cancer Res. Demonstration of a genetic therapeutic index for tumors expressing oncogenic BRAF by the kinase inhibitor SB-590885.Cancer research, **66**, 11100-11105. (2006).
74. Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, 1997.
75. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **23**, 3-17 (1997).
76. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **36**, 3219-3228 (1980).
77. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639-1662 (1998).
78. Park, H.; Lee, J.; Lee, S. Critical assessment of the automated AutoDock as a new docking tool for virtual screening. *Proteins* **65**, 549-554 (2006).
79. Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins* **34**, 4-16 (1999).
80. Park, H.; Jeon, J. H. Cubic equation governing the outer-region dielectric constant of globular proteins. *Phys. Rev. E* **75**, 021916 (2007).
81. Fabian, M. A.; Biggs III, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Bendetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J.-M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule–kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **23**, 329-336 (2005).
82. Chipot, C., Pohorille, A. *Free energy calculations*. Springer-Verlag Berlin Heidelberg, **2007**.

Chapter 5

Summary of work described in Chapter 2-4.

5.1. Reversibility of avibactam mediated serine β-Lactamase inhibition

5.2. Hydroxylation of D-amino acids by FIH

5.3. Computational prediction of molecular hydration entropies

5.4. Overall conclusions and future directions

5.1. Reversibility of avibactam mediated serine β-Lactamase inhibition

The development of new antibiotics is currently a topic of intense global interest. [1] Avibactam is arguably the most important molecule to be introduced for clinical use in the antibiotic field for more than 20 years. Although avibactam inhibits the same targets as the β-lactam based inhibitors of the β-lactamases it does so via a mechanism in which it reacts differently. The key mechanistic question of potential recyclization in the enzymatic active site via addition-elimination mechanisms has not previously been addressed even for β-lactam antibiotics.[2,3] Chapter 2 describes MD simulations and quantum calculations on the recyclization and hydrolysis of the avibactam derived acyl-enzyme complex with a clinically important SBL and compare the results with those for a β-lactam (oxacillin). The computational results studies reveal that recyclisation of the Avibactam derived carbamoyl (acyl-enzyme) complex via a stereoelectronically favoured reaction is preferred over hydrolysis. In contrast, analogous the recyclization in β-lactam mediated inhibition is disfavored. The results reveal avibactam recyclization is promoted by a proton shuttle, a 'structural' water protonating the nucleophilic serine, and stabilization of negative charge on the avibactam derived carbonyl (Figure 5.1). In most of the trajectory snapshots collected from the MD simulations, the nitrogen (N_S) of the hydroxylamine-O-sulfonate group of avibactam was observed to be closer to the carbonyl carbon (C1) of the carbamoyl moiety than the active site water molecule (Wat411) bound to Glu166. The hydrogen bond between the N_S atom of avibactam and Lys73-Ser130 dyad was also observed to be stable during the simulations. The results thus support preferential recyclisation over hydrolysis for the avibactam complex.[4]

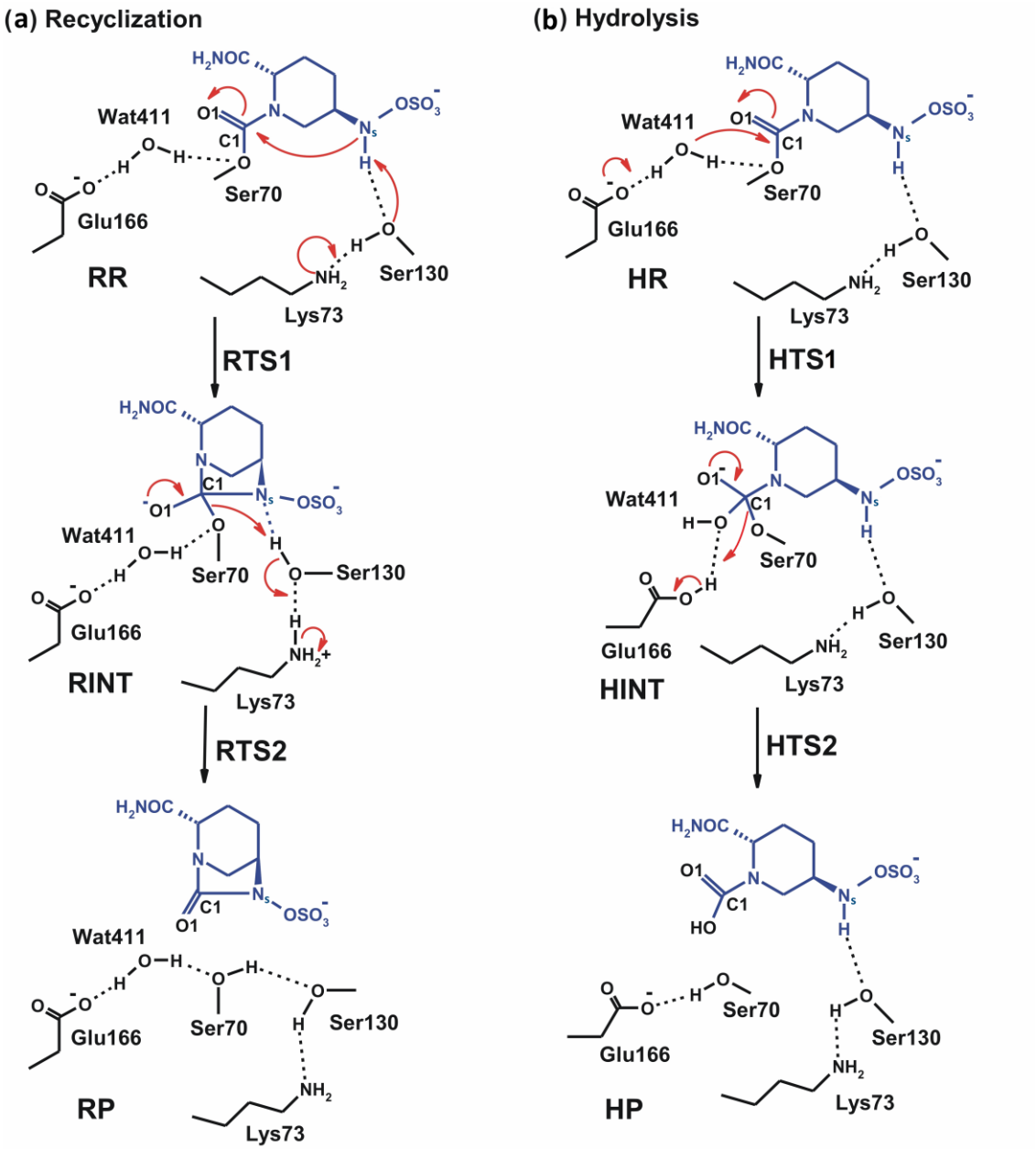


Figure 5.1. Possible fates for reaction of the CTX-M-15 avibactam covalent complex: (a) recyclisation, and (b) hydrolysis. **RR** and **HR**, **RTS1** and **HTS1**, **RINT** and **HINT**, **RTS2** and **HTS2**, and **RP** and **HP** denote the reactive complex, first transition state, reaction intermediate, second transition state, and final energy minimum, respectively, during the recyclisation and hydrolysis reactions, respectively. Figure taken from Choi et al, *Org. Biomol. Chem. 14*, p.4119. Copyright 2016 Royal Society of Chemistry. Reprinted with permission.

Consistent with the difference in activation free energies, recyclisation appears to be more efficient than hydrolysis in terms of stabilizing the negative charges developed on the carbonyl oxygen (O1) of the acyl-enzyme complex and the leaving OG atom of Ser70 of the transition states and the tetrahedral reaction intermediate. The differences in the

structural features of the transition states involved in the two deacylation fates indicate that, when compared to the role of Glu166 in the hydrolysis, during recyclisation, the Lys73-Ser130 dyad can serve not only as the better general base to facilitate the nucleophile addition in the first step but also as the better electrophilic catalyst to stabilize the leaving group in the second step.

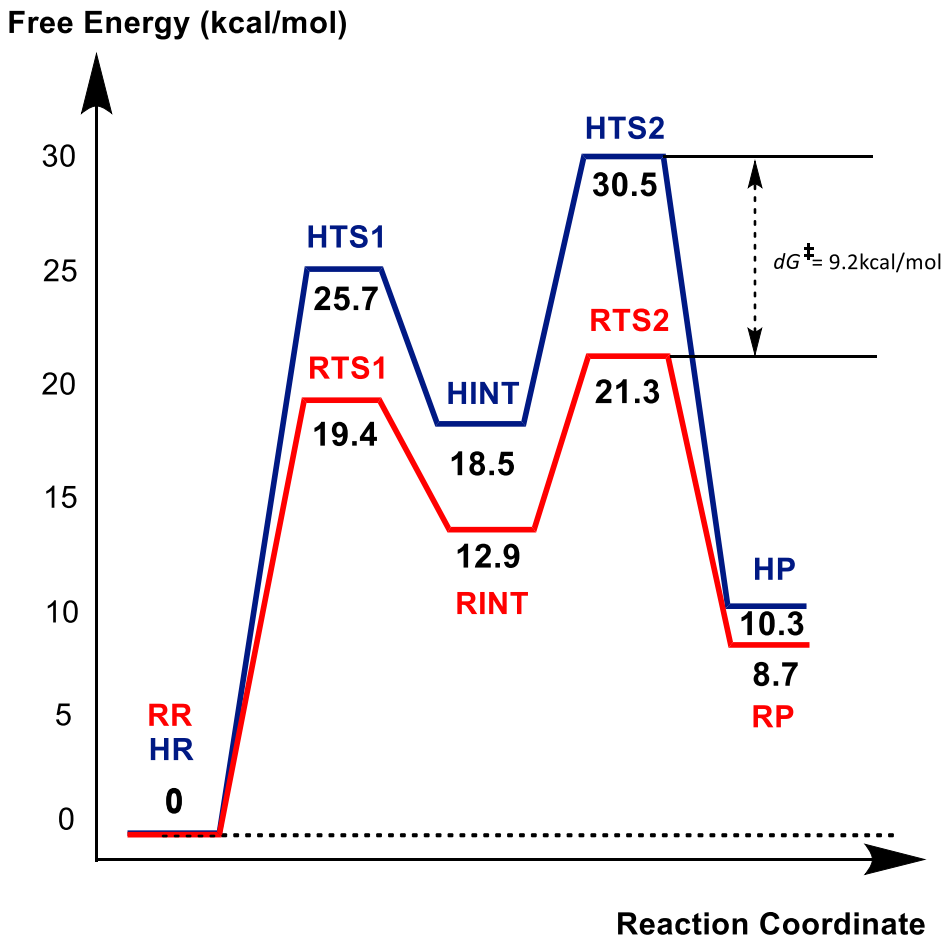


Figure 5.2. Free energy profile of the intrinsic reaction coordinate for recyclisation and hydrolysis of the Ser70-avibactam bond in the CTX-M-15 active site. Free energy is measured from reactant complex (**RR, HR**) for each stationary-state structure using the MP2/6-31G* level of calculations. **RTS1**, **RINT**, **RTS2**, **RP** and **HTS1**, **HINT**, **HTS2**, **HP** represent the first transition state, reaction intermediate, second transition state, and final energy minimum, during the recyclisation reaction and the hydrolysis reaction, respectively.

The results of *ab initio* calculations at MP2/6-31G*//RHF/6-31G* and wB97X/6-311++G** levels of theory on the model systems support the proposal that recyclisation

is kinetically favored over hydrolysis with a calculated difference of 9.2 kcal/mol in the two activation free energies in Figure. 5.2. The combination of the results of this study and those of previous studies implies that the avibactam template is optimized for binding to give a covalent complex, but that recyclisation albeit slowly to reform active avibactam is favoured over hydrolysis to give an inactive product. The special properties of β-lactams have been proposed in part to arise from their ability to acylate a nucleophilic active site residue which is unusually stable with respect to both recyclisation and, when appropriately functionalised, to hydrolysis.[5-7] However, avibactam, and other appropriately optimized reversibly and covalently binding inhibitors, may have an advantage over (at least bicylic) β-lactam inhibitors which recyclisation appears not be an option as shown by our calculations on the acyl-enzyme complex formed from oxacillin.

5.2. Hydroxylation of D-amino acids by FIH

One characteristic feature distinguishing FIH from most other 2OG oxygenases lies in its promiscuity in catalyzing the hydroxylation of amino acid residues.[8-10] FIH was identified as an asparaginyl C-3 hydroxylase with catalytic activity on hypoxia inducible factor (HIF); such activity reduces the transcriptional activity of HIF to the extent to be physiologically relevant to the hypoxic response. FIH also catalyses the hydroxylation of multiple ankyrin repeat domains at the C-3 positions; not only of Asn-, but also of Asp-, and His-residues.[11,12] The work described in Chapter 3 concerned the possible hydroxylation of D-amino acid residues by FIH. This work was inspired by the discovery that FIH can catalyse hydroxylation of hydrophobic residues; the potential hydroxylation of D-residues is of interest because of their increased prevalence in aged cells.[13,14]

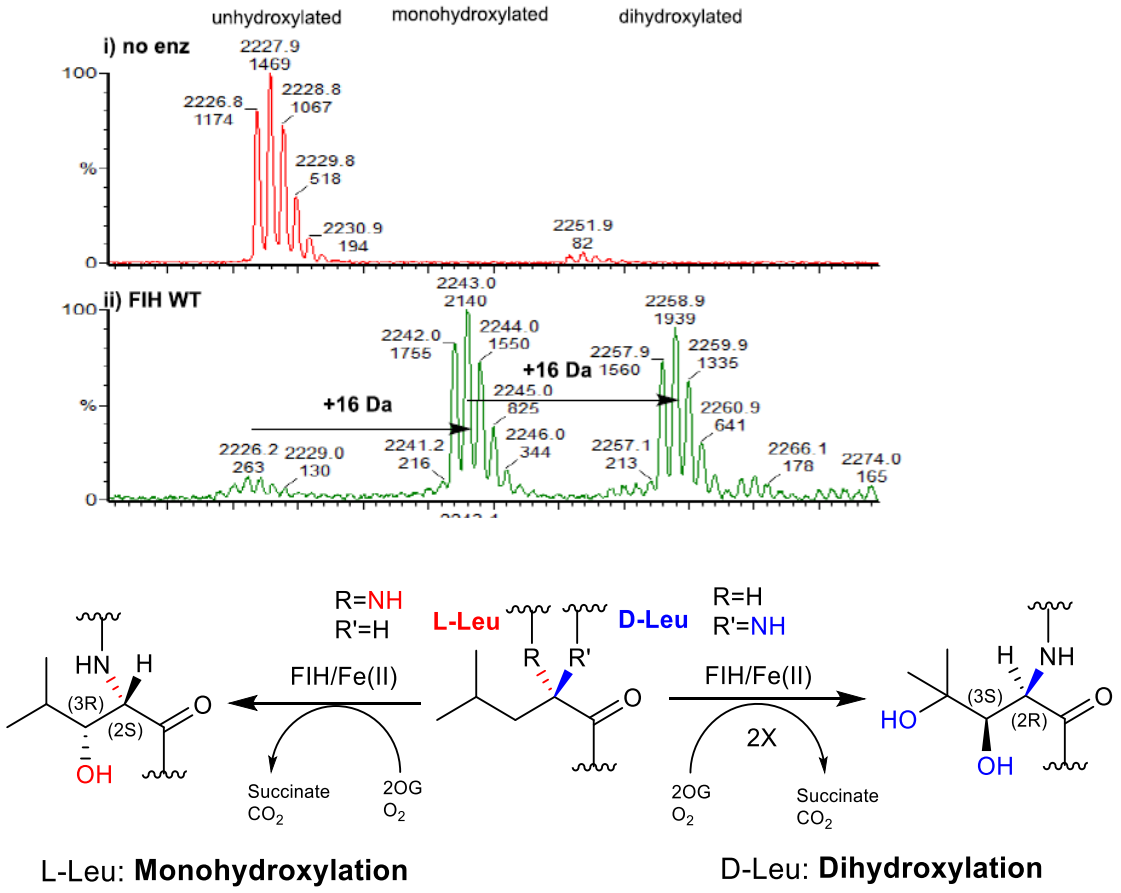


Figure 5.3. The single and double hydroxylations of L- and D-leucine, respectively, in the FIH-catalyzed hydroxylation of an ankyrin substrate. Mass spectrometric analysis of FIH catalyzed hydroxylation of D-leucine in CA peptide substrate. Time course analysis implies sequential monohydroxlyation followed by dihydroxylation. Peaks were assigned through MS fragmentation and NMR analyses.

To examine the promiscuity of FIH in catalyzing the hydroxylation of unusual amino acid residues, several D-amino acid residues were substituted at the hydroxylation position of the consensus ankyrin (CA) peptide substrates. Figure 5.3 shows the results of enzymatic hydroxylation reaction of D-leucine in CA peptide substrate at varying concentrations of FIH. Surprisingly, in the case of D- (but not L-) leucine not only a +16 Da but also a +32 Da mass shift was observed. Driven by the knowledge that the availability of the second hydroxylation reaction depends on the stereochemical properties of the substrate, I performed DFT calculation and collaborative X-ray crystallographic analyses to investigate the detailed energetic and structural features of FIH-substrate complexes responsible for the hydroxylation state of the products.

The FIH-catalyzed hydroxylation of D-leucine seems to proceed via a concerted mechanism. The concurrent occurrence of proton transfer and O–C bond formation has also been observed in experimental studies on the catalytic mechanism of various of the 2OG dependent oxygenases.[15-18] Figure 5.4(a) displays the potential energy profiles along the reaction paths depicted. The stereoselective hydroxylation reactions of D- and L-leucine seem to be thermodynamically favorable in the active site of FIH because **SDP** and **RLP** are approximately 60 kcal/mol lower in potential energy than **SDR** and **RLR**, respectively. These large stabilizations can be attributed to the formation of the two new covalent bonds involving O1 atom at the expense of one C–H bond. The activation barrier for the hydroxylation of D-leucine appears to be lower than that of L-leucine by 1.9 kcal/mol.

Figure 5.4(b) compares the potential energy profile diagrams along the reaction coordinates of the FIH-catalyzed second hydroxylation of (*3S*)-β-hydroxy-D-leucine and (*3R*)-β-hydroxy-L-leucine. The β,γ-dihydroxy-D-leucine appears to be produced as an FIH bound complex (**SDP2**) with a potential energy of 61.9 kcal/mol above that of the reactant complex (**SDR2**); this indicates that the second hydroxylation reaction of (*3S*)-β-hydroxy-D-leucine is thermodynamically favourable. An activation barrier of 11.6 kcal/mol must be overcome to form the product via the transition state (**SDTS2**). It is interesting to compare the calculated energetics for the second hydroxylation of (*3S*)-β-hydroxy-D-leucine with those of (*3R*)-β-hydroxy-L-leucine. The activation barrier for the formation of the vicinal diol product from (*3S*)-β-hydroxy-D-leucine is 9.7 kcal/mol lower than that from (*3R*)-β-hydroxy-L-leucine. As can be inferred by comparing the energy difference between **SDR2** and **SDP2** to that between **RLR2** and **RLP2**, the formation of β,γ-dihydroxy-D-leucine seems to be thermodynamically more favored than

β,γ-dihydroxy-L-leucine by 8.2 kcal/mol. On the basis of kinetic and thermodynamic aspects deduced from *ab initio* DFT calculations, it can be argued that the second hydroxylation reaction is available only for D-leucine in the active site of FIH.

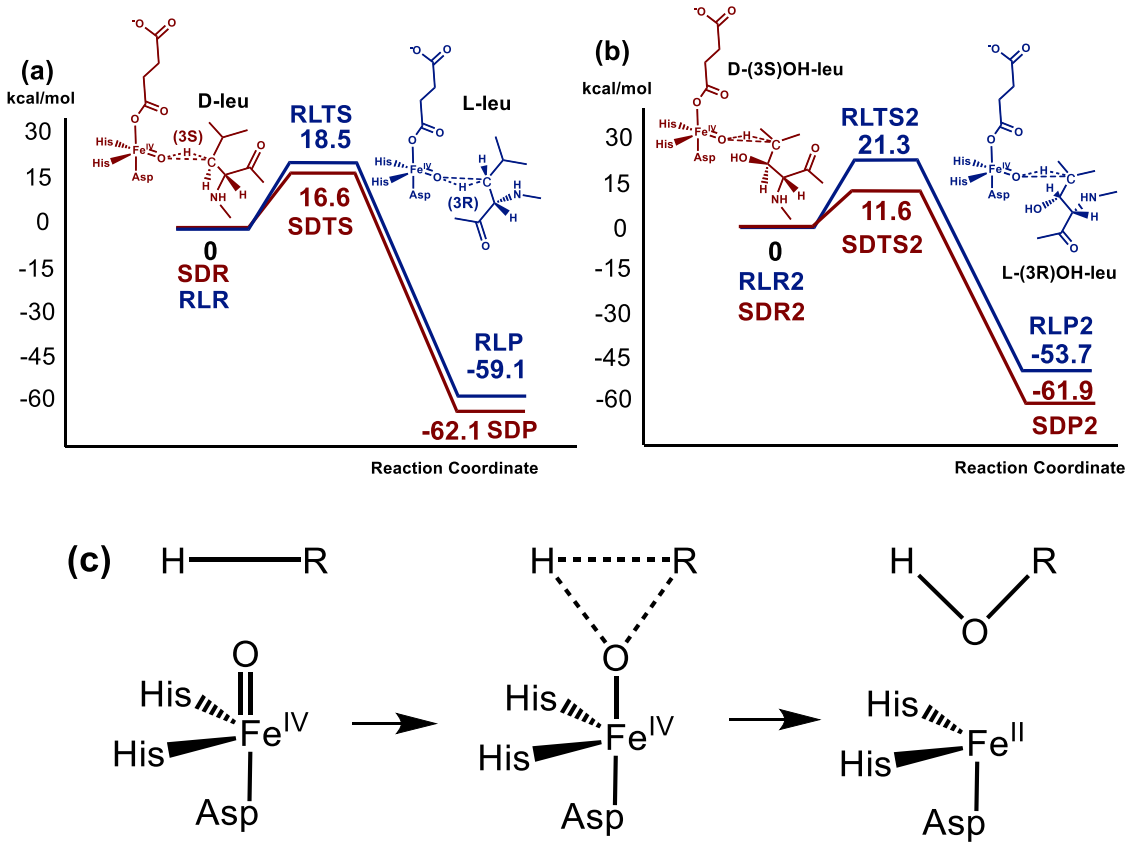


Figure 5.4. (a) Potential energy profile diagrams along the reaction coordinates of the FIH-catalyzed hydroxylation of D-leucine (brown) and L-leucine (blue). **SDR**, **SDTS**, **SDP**, **RLR**, **RLTS** and **RLP** represent the reactive complex, first hydroxylation transition state, hydroxylated product bound in the active site, D-leucine and L-leucine, respectively (b) Potential energy profile diagrams along the reaction coordinates of the FIH-catalyzed second hydroxylation of (3*S*)-β-hydroxy-D-leucine (brown) and (3*R*)-β-hydroxy-L-leucine (blue). **SDR2**, **SDTS2**, **SDP2**, **RLR2**, **RLTS2** and **RLP2** represent the reactive complex, second hydroxylation transition state, second hydroxylated product bound in the active site, (3*S*)-β-hydroxy-D-leucine (brown) and (3*R*)-β-hydroxy-L-leucine, respectively (c) Schematic nuclear motion of FIH-catalysed hydroxylation.

The difference in activation barriers and reaction energies for the observed and putative second hydroxylations of (*3S*)-β-hydroxy-D-leucine and (*3R*)-β-hydroxy-L-leucine can thus be attributed to the effect of the C-α stereochemistry on active site orientations to change the loop structure of CA peptide substrate and different hydrogen

bond pattern originated from D-/L-first hydroxylation (favourable hydrogen bond from first (*3S*)-β-hydroxy-D-leucine vs. obstacle hydrogen bond from first (*3R*)-β-hydroxy-L-leucine).

To rationalise the different availabilities of double hydroxylation between L- and D-leucine, I compared the conformations of L- and D-leucine in the active site of FIH as revealed in X-ray crystal structures. As shown in Figure 5.5(a), the crystallography shows that the backbone aminocarbonyl groups of the former and the latter are directed in opposite directions (L-/D- main chain direction of substrate) due to the different stereochemical properties at C_α atom, which leads to differential positioning of the side chains and to the change of loop structure of CA peptide substrates. The torsional motion of the isopropyl group of L-leucine appear restricted because it establishes a close van der Waals contact with the side chains of Tyr102, His199, and Gln239 in the active site (Figure 5.5(b)). On the other hand, the side-chain isopropyl group of D-leucine points toward bulk solvent without significant interactions with nonpolar residues in FIH. Because the backbone aminocarbonyl oxygen of D-leucine and that of the neighbouring Ala804 of CA peptide substrate repel the side chain of Tyr102 away from the active site, the isopropyl group of D-leucine should be allowed to rotate freely, so enabling the hydrogen atom bonded to C_γ atom to the reaction oxidised species of the ferrous ion. Crystallographic analysis shows that the side chain of L-leucine is confined by two rings of His199 and Tyr102 (Figure 5.5(b)), However, the relatively blunt curvature of the main chain of D-leucine substrate can open 'the Tyr102 ring gate' to release the side chain of D-leucine to undergo oxidation (Figure 5.5(a)). Thus, the stereoselective double hydroxylation leucine by FIH stems from the attainability of a reactive conformation for D-, but not L-leucine. To obtain structural insight into the stereochemical dependences in the catalytic hydroxylation of leucine by FIH, crystal structures of FIH in complex

with L-leucine, D-leucine, and D-β-(2*R*,3*S*)-hydroxyleucine were determined. As implied by analysis of the structures (Figure 5.5(b)), the product of the first hydroxylation of L-leucine is expected to be L-β-(2*S*,3*R*)-hydroxyleucine because the *pro-3R* hydrogen resides closer to the central ferrous ion (3.4 Å) than *pro-3S* hydrogen (3.9 Å). However, the positions of the *pro-3R* and *pro-3S* hydrogens are "exchanged" in D-leucine with the distances of 4.1 and 3.2 Å from the ferrous ion (Figure 5.5(c)), respectively. This is due to the change in the stereochemistry of the C_α atom. The results imply that computationally based prediction of oxygenase activity may be useful.

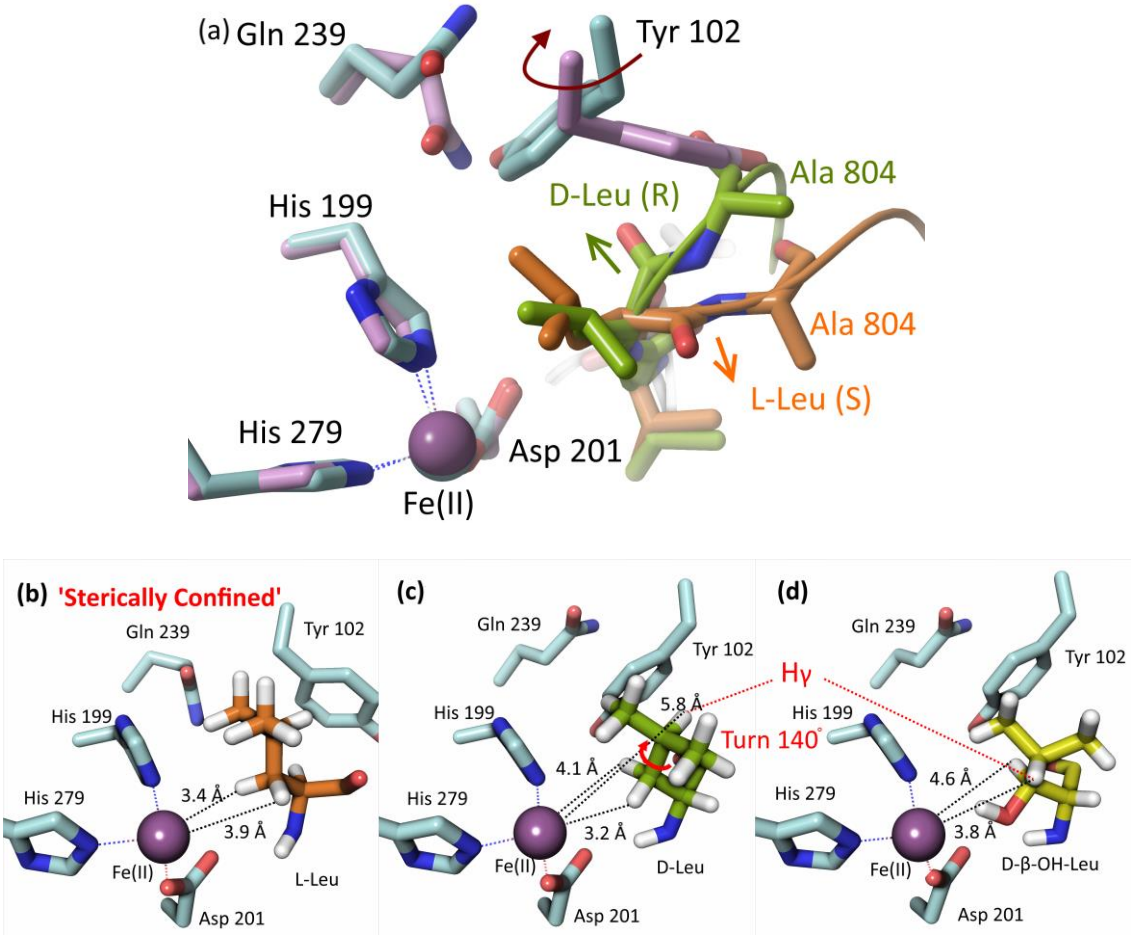


Figure 5.5. Superposition of X-ray crystal structures of FIH in complex with CA peptides containing (a) L-leucine (orange) and D-Leucine (yellowgreen) at residue 803. Carbon atoms of FIH in complex with L- and D-leucine are indicated in pink and cyan, respectively, (b) L-leucine, (c) D-leucine, and (d) D-β-(2*R*,3*S*)-hydroxyleucine at residue 803.

5.3. Computational prediction of molecular hydration entropies

Despite the importance of the knowledge of molecular hydration entropy (ΔS_{hyd}) in chemical and biological processes, the exact calculation of ΔS_{hyd} is difficult due to the complexity in solute-water interactions. My studies aimed at a novel computational method for estimating the molecular hydration entropy (ΔS_{hyd}) values with high accuracy based on a modified thermodynamic decomposition approach. Knowledge of ΔS_{hyd} would also be informative for elucidating the thermodynamic aspects of protein-ligand interactions in which the major contribution to the binding free energy comes from the entropic term without significant enthalpy-entropy compensation during the conformational changes for binding.[19] In contrast to the general notion that ligand binding to proteins is accompanied by the entropy loss, it has been demonstrated that the binding entropy (ΔS_{bind}) can be positive during the formation of some protein-ligand complexes,[20,21] furthermore, in some case the $-T\Delta S_{bind}$ values can be more negative than ΔH_{bind} in some cases.[22-25] A solute molecule dissolves in water when the resulting solute-water interactions are strong enough to surmount the entropic penalty for establishing the optimal configuration of hydration. The loss of entropy during hydration stems from the reordering of solute and water molecules to form a complex with the optimal intermolecular interactions. In this context, ΔS_{hyd} may be decomposed into van der Waals (ΔS_{vdW}) and electrostatic (ΔS_{elec}) contributions.

$$\Delta S_{hyd} = \Delta S_{vdW} + \Delta S_{elec} \qquad (1)$$

The sign of ΔS_{vdW} is negative for most organic solutes due to the immiscibility between their hydrophobic parts and water. ΔS_{vdW} can therefore be approximated as the loss of entropy associated with the reordering of water molecules to form a cavity (ΔS_{cav}) in

which a solute molecule can be accommodated. Then, ΔS_{hyd} can be expressed in the following form as eq (2);

$$\Delta S_{hyd} = \Delta S_{cav} + \Delta S_{elec} \qquad (2)$$

SPT aims to calculate the chemical potential (Δμ) for the insertion of a cavity into a hard sphere fluid. Because the formation of a cavity for a hydrophobic solute in water may be approximated as an entropic process due to the weak solute-water interactions, ΔS_{cav} can be expressed as follows.[26]

$$\frac{\Delta S_{cav}}{k_B} = -\ln(1-\varphi) + 3\left(\frac{\varphi}{1-\varphi}\right)R + \left[\left(\frac{3\varphi}{1-\varphi}\right) + 4.5\left(\frac{\varphi}{1-\varphi}\right)\right]R^2 + \frac{\varphi P}{\rho k_B T}R^3 \qquad (3)$$

Here, φ and ρ represent the volume fraction of water molecules with respect to the total volume (0.371) and the number density of water (0.333×10^{24} m^{-3}), respectively. R denotes the ratio of the radius of solute to that of water molecule (r_t/r_w) while P stands for the pressure of water (2.306 kPa). Because r_w is known to be 1.4 Å, the knowledge of r_t is the only requisite for calculating the ΔS_{cav} value of a solute molecule.

ΔA at a given temperature is the difference of Helmholtz free energies between the charged and uncharged states of the system.[27-31] In this case, ΔS_{elec} can be given by eq 4.

$$\Delta S_{elec} = -\frac{\{\Delta A_q(T+\Delta T) - \Delta A_0(T+\Delta T)\} - \{\Delta A_q(T-\Delta T) - \Delta A_0(T-\Delta T)\}}{2\Delta T} \qquad (4)$$

Here, A_q and A_0 are the Helmholtz free energies of charged and uncharged systems, respectively. It is thus necessary to compute the differences between A_q and A_0 at temperatures $T + \Delta T$ and $T - \Delta T$ to obtain ΔS_{elec}, which were accomplished successfully with the constrained FEP simulations.

Because ΔA_{hyd} refers to the change in free energy for a solute molecule to leave the gas phase and to enter the liquid water, a thermodynamic cycle may be useful to calculate the ΔA_{hyd} values at varying temperatures, which is illustrated in Figure 5.6. Because A is thermodynamically a state function, the sum of all free energy changes in the cycle must be zero. ΔA_{hyd} can therefore be expressed as follows.

$$\Delta A_{hyd} = \Delta A_{gas} - \Delta A_{aq} \quad (5)$$

ΔA_{aq} and ΔA_{gas} correspond to the difference between A_q and A_0 values of the solute in aqueous solution and that in the gas phase, respectively. I obtained the ΔA_{hyd} values of a test set of solute molecules through the FEP calculations of ΔA_{aq} and ΔA_{gas} based on the independent MD simulations in water and in the gas phase, respectively.

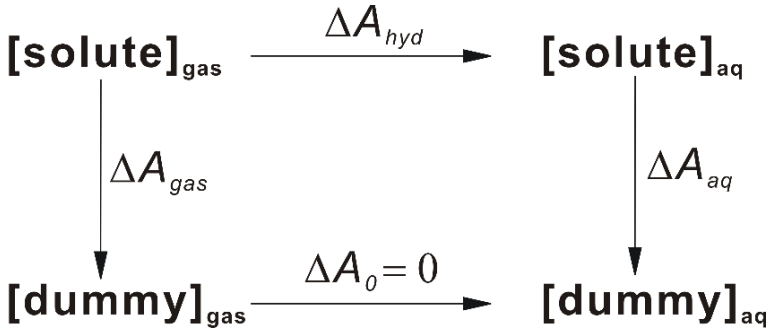


Figure 5.6. Thermodynamic cycle for calculating the hydration free energies of solute molecules. Note that $\Delta A_0 = 0$ because solute is not involved in the process.

I compared the ΔS_{hyd} values of all 90 solute molecules given by ΔS_{cav} plus ΔS_{elec} values calculated with SPT (or IT) and MD-based FEP simulations, respectively, to the corresponding experimental ones. The linear correlation diagram between the experimental and calculated results is shown in Figure 5.7.

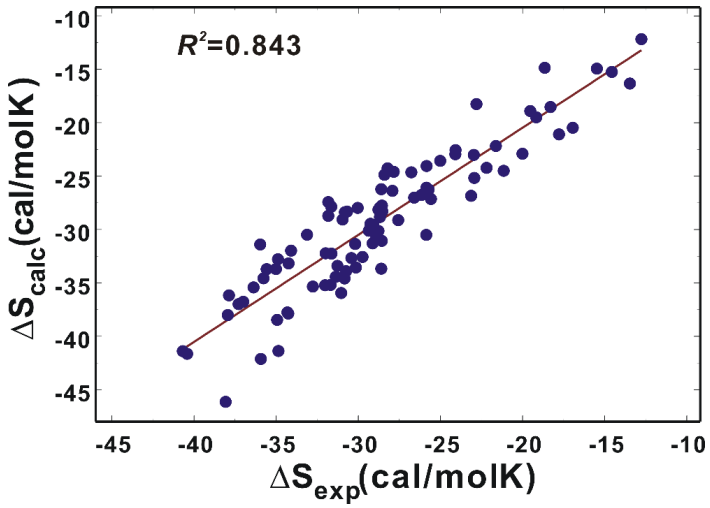


Figure 5.7. Linear correlation diagram between the experimental and calculated hydration entropies of 90 compounds at 298 K. The slope and intercept of the fitting are 1.011 and –0.148, respectively. Figure taken from Choi et al, *J. Chem. Theory Comput*. *11*, p.4939. Copyright 2015 American Chemical Society. Reprinted with permission.

Overall, the estimated hydration entropies compare reasonably well with the experimental ones with the associated RMSE of 2.68 cal·mol^{-1}·K^{-1}, which corresponds to 9.2% of the average of the absolute experimental values. In this fitting, the R^2 value amounts to 0.843 with the slope and the intercept values of 1.011 and -0.148, respectively.[32]

Molecular hydration entropies can be predicted with reasonable accuracy by means of the hybrid SPT/FEP method to separately calculate the cavity formation entropy (ΔS_{cav}) as the hydrophobic contribution and electrostatic solute-water interaction term (ΔS_{elec}). Due to the effective partitioning of the two entropic components, the prediction accuracy of ΔS_{hyd} was enhanced to a substantial extent when compared to the conventional full FEP calculations. This improvement was made possible by successfully overcoming the difficulty in calculating ΔS_{vdw} in full FEP calculations. SPT and IT exhibited almost the same performance in the estimation of ΔS_{cav}.

5.4. Overall conclusions and future directions

Overall the work of this thesis exemplifies how computational studies can give detailed insights into protein chemistry. The work on β-lactamases reveals that computational calculations are of use in characterizing inhibition mechanisms, in particular with respect to formation of acyl-enzyme complexes and their reactivity, including with respect to reformation of the inhibitor and hydrolysis of the acyl-enzyme complex. The work with oxygenases shows that computational methods have potential for predicting the hydroxylation positon catalyzed by uncharacterized enzymes. Finally, the work on hydration free energies serves as a fundamental thermodynamic quantity for the behavior of a small molecule in a variety of chemical and biological systems. The development of more sophisticated methods and increased computing power will allow us to expand applications to biological systems.

Future studies could focus on the physical and chemical understanding of Interactions of biological phenomena, protein-ligand interactions (drug discovery), and protein-substrate, protein-protein interactions from the perspective of the thermodynamic decomposition. Enthalpic and entropic decompositions of protein-ligand interactions could give valuable insights into new development of new drugs, especially by understanding entropy driven protein-ligand binding. In addition, a detailed examination of the thermodynamics of binding, including the free energy changes resulting from displacing water molecules in the active site may help in structure-based drug design. Calculations on the organometallic chemistry of 2-oxoglutarate (2OG) and ferrous iron dependent oxygenase superfamily may reveal new mechanism for uncharacterized enzymes. For example, in work submitted for publication calculations by me have informed on the stereo electronic preference of Grob type fragmentation has been proposed for a 2OG dependent bacterial ethylene-forming enzyme (*Ps*EFE).[33]

Furthermore, the prediction of how oxygenase active sites can modify substrates should enable the modification of 2OG oxygenase for use in biocatalysis.

Chapter 5 References

1. Butler, M. S., Blaskovich, M. A., Cooper, M. A. Antibiotics in the clinical pipeline in 2013. *J. Antibiot.* **66**, 571-591 (2013).
2. Burgi, H. B., Dunitz, J. D., Shefter, E. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group.*J. Am. Chem. Soc.* **95**, 5065-5067 (1973).
3. Bruice, T. C., Benkovic, S. J. Chemical Basis for Enzyme Catalysis. *Biochemistry* **39**, 6267-6274 (2000).
4. Wilmouth, R. C., Clifton, I. J., Robinson, C. V., Roach, P. L., Aplin, R. T., Westwood, N. J., Hajdu, J., Schofield, C. J. Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. *Nat. Struct. Biol.* **4**, 456-462 (1997).
5. Ehmann, D. E., Jahić, H., Ross, P. L., Gu, R.-F., Hu, J., Durand-Réville, T. F., Lahiri, S., Thresher, J., Livchak, S., Gao, N., Palmer, T., Walkup, G. K., Fisher, S. L. Kinetics of avibactam inhibition against Class A, C, and D β-lactamases. *J. Biol. Chem.* **288**, 27960-27971 (2013).
6. Lahiri, S. D., Mangani, S., Durand-Reville, T., Benvenuti, M., Luca, F. D., Sanyal, G., Docquier, J.-D. Structural insight into potent broad-spectrum inhibition with reversible recyclization mechanism: avibactam in complex with CTX-M-15 and Pseudomonas aeruginosa AmpC β-lactamases. *Antimicrob. Agents Chemother.* **57**, 2496-2505 (2013).
7. Sgrignani, J., Grazioso, G., Amici, M. D., Colombo, G. Inactivation of TEM-1 by Avibactam (NXL-104): Insights from Quantum Mechanics/Molecular Mechanics Metadynamics Simulations. *Biochemistry* **53**, 5174-5185 (2014).
8. Cockman, M.E., Lancaster, D.E., Stolze, I.P., Hewitson, K.S., McDonough, M.A., Coleman, M.L., Coles, C.H., Yu, X., Hay, R.T., Ley, S.C., Pugh, C.W., Oldham, N. J., Masson, N., Schofield, C.J., Ratcliffe, P.J. Posttranslational hydroxylation of ankyrin repeats in IκB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc. Natl. Acad. Sci. USA* **103**, 14767-14772 (2006).
9. Yang, M., Hardy, A.P., Chowdhury, R., Loik, N.D., Scotti, J.S., McCullagh, J.S., Claridge, T.D., McDonough, M.A., Ge, W., Schofield, C.J. Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl.* **52**, 1700-1704 (2013).
10. Linke, S., Stojkoski, C., Kewley, R.J., Booker, G.W., Whitelaw, M.L., Peet, D.J. Substrate Requirements of the Oxygen-sensing Asparaginyl Hydroxylase Factor-inhibiting Hypoxia-inducible Factor. *J. Biol. Chem.* **279**, 14391-14397 (2004).
11. Coleman, M.L., McDonough, M.A., Hewitson, K.S., Coles, C., Mecinović, J., Edelmann, M., Cook, K.M., Cockman, M.E., Lancaster, D.E., Kessler, B.M., *et al*. Asparaginyl Hydroxylation of the Notch Ankyrin Repeat Domain by Factor Inhibiting Hypoxia-inducible Factor. *J. Biol. Chem.* **282**, 24027 - 24038 (2007).
12. Hardy, A.P., Prokes, I., Kelly, L., Campbell, I.D., Schofield, C.J. Asparaginyl β-Hydroxylation of Proteins Containing Ankyrin Repeat Domains Influences Their Stability and Function. *J. Mol. Biol.* **392**, 994-1006 (2009).
13. Fisher, G. H. Appearance of D-amino acids during aging: D-amino acids in tumor proteins. *Exs*, **85**, 109-118 (1997).
14. Fujii, N. D-amino acid in elderly tissues. *Biol. Pharm.Bull.* **28**, 1585-1589 (2005).
15. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Jr. Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:r-Ketoglutarate Dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008-13009 (2003).
16. Hoffart, L. M.; Barr, E. W.; Guyer, R. B.; Bollinger, J. M., Jr.; Krebs, C. Direct Spectroscopic Detection of a C-H-Cleaving High-Spin Fe(IV) Complex in a Prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. USA* **103**, 14738–14743 (2006).
17. Grzyska, P. K.; Appelman, E. H.; Hausinger, R. P.; Proshlyakov, D. A. Insight into the Mechanism of an Iron Dioxygenase by Resolution of Steps Following the Fe^{IV}═O Species. *Proc. Natl. Acad. Sci. USA* **107**, 3982–3987 (2010).
18. Martinez, S.; Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702-20711 (2015).
19. Ahmad, M., Helms, V., Lengauer, T., Kalinina, O. V. Enthalpy–Entropy Compensation upon Molecular Conformational Changes. *J. Chem. Theory Comput.* **11**, 1410−1418 (2015).
20. Reynolds, C. H., Holloway, M. K. Thermodynamics of ligand binding and efficiency. *ACS Med. Chem. Lett.* **2**, 433–437(2011).

21. Olsson, T. S., Williams, M. A., Pitt, W. R., Ladbury, J. E. The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. *J. Mol. Biol*. **384**, 1002–1017 (2008).
22. Kavanagh, K. L., Guo, K., Dunford, J. E., Wu, X., Knapp, S., Ebetino, F. H., Rogers, M. J., Russell, R. G. G., Oppermann, U. The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *Proc. Natl. Acad. Sci. USA* **103**, 7829–7834 (2006).
23. Kaul, M., Barbieri, C. M., Pilch, D. S. Defining the basis for the specificity of aminoglycoside-rRNA recognition: a comparative study of drug binding to the A sites of Escherichia coli and human rRNA. *J. Mol. Biol*. **346**, 119–134 (2005).
24. Carbonell, T., Freire, E. Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry* **44**, 11741–11748 (2005).
25. Yanchunas, J., Jr., Langley, D. R., Tao, L., Rose, R. E., Friborg, J., Colonno, R. J., Doyle, M. L. Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrob. Agents Chemother.* **49**, 3825–3832 (2005).
26. Lebowitz, J. L., Helfand, E., Praestgaard, E. Scaled particle theory of fluid mixtures. *J. Chem. Phys.* **43**, 774–779 (1965).
27. Yu, H. A., Karplus, M. A thermodynamic analysis of solvation. *J. Chem. Phys.* **89**, 2366–2379 (1988).
28. Levy, R. M., Gallicchio, E. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* **49**, 531–567 (1998).
29. Wan, S., Stote, R. H., Karplus, M. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* **121**, 9539–9548 (2004).
30. Smith, D. E., Haymet, A. D. J. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *J. Chem. Phys*. **98**, 6445–6454 (1993).
31. Peter, C., Oostenbrink, C., van Dorp, A., van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. *J. Chem. Phys*. **120**, 2652–2661 (2004).
32. Kubo, M. M., Gallicchio, E., Levy, R. M. Thermodynamic decomposition of hydration free energies by computer simulation: application to amines, oxides, and sulfides. *J. Phys. Chem. B* **101**, 10527–10534 (1997).
33. Zhang, Z., Smart, T. J, Choi, H., Hardy, F., Lohans, C. T., Richardson, M. S. W., Paton, R. S., McDonough, M. A., Schofield, C. J. Structural and Stereoelectronic Insights into Oxygenase Catalysed Formation of Ethylene from 2-Oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *submitted*

Appendices

Chapter 3 Appendices

Table A.3.1. Crystallographic data and refinement statistics

Chapter 4 Appendices

Table A.4.1. ΔS_{cav}^{SPT}, ΔS_{cav}^{IT}, and ΔS_{hyd} values of 90 solute molecules calculated at 298 K

Table A.4.2. ΔA_{elec} values of 90 solute molecules calculated at 268 and 328 K

Table A.4.3. ΔS_{hyd} values (in $cal \cdot mol^{-1} \cdot K^{-1}$) of 18 reference molecules from full FEP simulations

Table A.4.4. ΔS_{cav} values of 18 reference molecules calculated from SPT and IT, ΔS_{elec} values from partial FEP simulations

Table A.4.5. ΔA_{elec}, ΔA_{vdw}, ΔA_{aq}, ΔA_{gas}, and ΔA_{hyd} values of 18 reference molecules calculated from full FEP simulations at 268 K

Table A.4.6. ΔA_{elec}, ΔA_{vdw}, ΔA_{aq}, ΔA_{gas}, and ΔA_{hyd} values of 18 reference molecules calculated from full FEP simulations at 328 K

Table A.4.7. The optimized atomic fragmental volume (V_j), maximum atomic occupancy (O_i^{max}), and atomic solvation (S_i) parameters for various atom types.

Table A.4.8. Molecular ΔH_{hyd} values of 90 solute molecules calculated with solvent contact model in comparison with the corresponding experimental results.

Table A.4.9. The optimized atomic fragmental volume (V_j), maximum atomic occupancy (O_i^{max}), and atomic solvation (S_i) parameters and atomic self-solvation parameters (P_i) for various atom types.

Table A.4.10. Experimental (ΔG^s_{exp}) and calculated (ΔG^s_{calc}) solvation free energies (in kcal/mol) of the 362 compounds in the training set.

Table A.4.11. Experimental (ΔG^s_{exp}) and calculated (ΔG^s_{calc}) solvation free energies (in kcal/mol) of the 42 compounds in the test set.

Protein: Substrates	**FIH:CAP(D-leucine)**	**FIH:CAP(D-β-hydroxyleucine)**	**FIH:CAP(D-allylglycine)**
X-ray source	Diamond Light Source beamline I04-1	Diamond Light Source beamline I04	Diamond Light Source beamline I04-1
Wavelength (Å)	0.9173	0.97949	0.9173
PDB Acquisition Code	4JAA	-	P4NR1
Resolution (Å)	2.39 (2.55–2.39)[§]	2.20 (2.28-2.20)[§]	2.68 (2.75–2.68)[§]
Space group	$P\,4_1\;\;2_1\;\;2$	$P\,4_1\;\;2_1\;\;2$	$P\,4_1\;\;2_1\;\;2$
Unit Cell Dimensions (a Å, b Å, c Å)	86.08, 86.08, 147.03	86.42, 86.42, 146.70	86.37, 86.37, 148.72
Molecules per a.u.	1	1	1
Total Number of Reflections Observed	112791	640987	164722
Number of Unique Reflections	22451 (3937)[§]	29287 (2878)[§]	16485 (1178)[§]
Redundancy	5.0 (4.6)[§]	21.9 (21.1)[§]	10.0 (10.8)[§]
Completeness (%)	99.3 (98.7)[§]	100.0 (100.0)[§]	99.9 (99.9)[§]
Wilson B	44.6	53.9	67
I/σ(I)	9.3 (2.6)[§]	17.9 (2.1)[§]	17.3 (2.9)[§]
[§§]R_{merge}	0.089	0.184	0.1
*R_{cryst}	0.184	0.213 (unrefined)	0.164
[†]R_{free}	0.219	0.247 (unrefined)	0.21
¶RMS deviation	0.01 (1.1°)	(unrefined)	0.01 (1.2°)
Average B factors (Å^2)	41.7	(unrefined)	63.4
Number of Water Molecules	120	(unrefined)	54

Table A.3.1. Crystallographic data and refinement statistics

[§]Parentheses indicate high resolution shell

[§§]$R_{\text{merge}} = \sum_j \sum_h \left| I_{hj} - < I_h > \right| / \sum_j \sum_h \left| < I_h > \right| \times 100$

*$R_{\text{cryst}} = \sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \left| F_{obs} \right| \times 100$

[†]R_{free}, based on 2-5% of the total reflections

¶RMS deviation from ideality for bonds (followed by the value for angle)

Crystals of FIH in complex with Zn(II), NOG and CAP peptides were grown in sitting drops using the vapour diffusion method (drop size: 200-300 nL) at 293 K in 96-well Intelliplates (Art Robbins). Crystals were cryo-protected by transfer to 25% (v/v) glycerol in well solution and then harvested in nylon loops (Hampton Research) and cryo-cooled by plunging in liquid nitrogen. Data were collected at 100 K using single crystals at Diamond Light Source beamline I04-1 [FIH:CAP(D-leucine) and FIH:CAP(D-allylglycine)] with a MarMosaic 300 mm CCD detector, and Diamond Light Source beamline I04 [FIH:CAP(D-β-hydroxyleucine)] with a Pilatus 6M-F detector. Data were then indexed,

integrated, and scaled using SCALA [FIH:CAP(D-leucine) and FIH:CAP(D-allylglycine)][1] and HKL3000 [FIH:CAP(D-β-hydroxyleucine)].[2] All structures were determined by molecular replacement (MR) using the MR-PHASER[3] subroutine of PHENIX[4] using *H. sapiens* FIH (PDB ID: 1H2K)[5] as the search model. Model building and refinement were performed iteratively using COOT[6] and PHENIX until converging R and R_{free} no longer decreased. Mn(II), NOG, and CAP were modelled in the final stages of refinement based on the $F_{obs} - F_{calc}$ electron density map.

Chapter 3 Appendices References

1. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K.S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr*. **67**, 235-242 (2011).
2. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* **276**, 307-326 (1997).
3. McCoy, A. J.; Storoni, L. C.; Bunkoczi, G.; Oeffner, R. D.; Read, R. J.; Grosse-Kunstleve, R. W.; Adams, P.D.; Winn, M. D. Phaser (CCP4: Supported Program) NAME phaser-2.5. 0-Maximum Likelihood Analysis and Phasing. SYNOPSIS phaser. *J. Appl. Cryst*, **40**, 658-674 (2007).
4. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J., Hung, L.W.; Kapral, G. J.; Grosse-Kunstleve, R. W. and McCoy, A. J. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr*. **66**, 213-221 (2010).
5. Elkins, J. M.; Hewitson, K. S.; McNeill, L. A.; Seibel, J. F.; Schlemminger, I.; Pugh, C. W.; Ratcliffe, P. J., Schofield, C. J. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *J. Biol. Chem.* **278**, 1802-1806 (2003).
6. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132 (2004).

compound	ΔS_{cav}^{SPT}	ΔS_{cav}^{IT}	ΔS_{hyd}	ΔS_{hyd}^{X}	compound	ΔS_{cav}^{SPT}	ΔS_{cav}^{IT}	ΔS_{hyd}	ΔS_{hyd}^{X}
methane	-13.56	-13.79	-13.46	-16.33	3-methylpyridine	-26.29	-26.28	-28.56	-27.76
ethane	-17.35	-17.32	-16.94	-20.47	4-methylpyridine	-26.29	-26.30	-28.69	-27.95
propane	-20.63	-20.50	-20.00	-22.91	3-ethylpyridine	-28.82	-29.00	-31.82	-27.44
butane	-23.57	-23.53	-23.13	-26.85	4-ethylpyridine	-28.79	-28.95	-30.81	-28.38
2-methylpropane	-23.57	-23.49	-22.93	-25.17	3,4-dimethylpyridine	-28.78	-28.98	-31.64	-27.86
pentane	-26.29	-26.26	-25.86	-30.51	3,5-dimethylpyridine	-28.79	-28.98	-31.63	-32.26
2,2-dimethylpropane	-26.27	-26.28	-25.57	-27.14	4-tert-butylpyridine	-33.41	-33.33	-35.97	-31.41
hexane	-28.82	-28.99	-28.60	-33.67	acetonitrile	-17.50	-17.45	-15.47	-14.94
heptane	-31.23	-31.35	-31.03	-35.95	propanenitrile	-20.75	-20.65	-19.53	-18.91
ethene	-19.18	-19.10	-18.65	-14.86	2-butanone	-24.03	-24.05	-27.84	-24.60
1-propene	-16.20	-16.19	-14.56	-15.25	2-pentanone	-26.71	-26.61	-30.03	-27.98
cyclopropane	-19.59	-19.46	-17.78	-21.09	3-pentanone	-26.70	-26.75	-30.68	-28.31
benzene	-24.39	-24.30	-24.08	-22.57	3-methyl-2-butanone	-26.71	-26.61	-32.01	-35.21
methylbenzene	-27.02	-27.01	-25.85	-26.09	2-hexanone	-29.22	-29.39	-33.11	-30.49
ethylbenzene	-29.51	-29.50	-29.08	-29.59	4-methyl-2-pentanone	-29.23	-29.22	-34.09	-31.98
m-xylene	-29.48	-29.56	-28.77	-28.14	2-heptanone	-31.60	-31.77	-36.37	-35.42
p-xylene	-29.48	-29.55	-28.77	-30.13	4-heptanone	-31.59	-31.66	-37.00	-36.76
propylbenzene	-31.88	-32.00	-31.26	-33.40	2-nonanone	-36.04	-36.13	-40.41	-41.63
(1-methylethyl)benzene	-31.85	-31.80	-31.98	-32.22	5-nonanone	-36.05	-35.79	-40.68	-41.38
9H-fluorene	-35.54	-35.60	-34.88	-32.78	propanoic acid	-22.56	-22.46	-25.01	-23.58
anthracene	-36.80	-36.98	-35.57	-33.72	butanoic acid	-25.35	-25.40	-27.92	-26.38
methanol	-15.56	-15.62	-18.29	-18.53	diethylether	-25.02	-24.91	-29.12	-31.29
ethanol	-19.05	-18.93	-22.96	-23.02	1,2-dimethoxyethane	-26.43	-26.49	-30.18	-31.35
1-propanol	-22.14	-22.03	-26.59	-27.01	tetrahydrofuran	-23.28	-23.26	-25.71	-26.25
2-propanol	-22.14	-22.04	-27.56	-29.12	tetrahydropyran	-25.93	-25.99	-31.82	-28.72
1-butanol	-24.96	-24.96	-30.14	-33.55	2-methyltetrahydrofuran	-26.01	-26.02	-29.35	-30.11
2-methyl-1-propanol	-24.97	-25.00	-30.42	-32.67	2,5-dimethyltetrahydrofuran	-28.55	-28.58	-32.76	-35.34
2-butanol	-24.96	-24.96	-30.83	-34.58	1,3-dioxolane	-21.95	-21.82	-22.80	-18.26
2-methyl-2-propanol	-24.95	-24.99	-30.74	-33.90	methylacetate	-22.65	-22.61	-24.09	-22.95
1-pentanol	-27.58	-27.68	-31.37	-34.43	ethylacetate	-25.42	-25.47	-28.60	-26.23
3-pentanol	-27.59	-27.52	-34.24	-33.17	methylbutanoate	-28.01	-28.02	-30.95	-29.06
2-methyl-2-butanol	-27.58	-27.69	-34.31	-37.76	N-butylacetate	-30.44	-30.48	-35.01	-33.69
1-hexanol	-30.06	-30.29	-34.93	-38.45	ethylpentanoate	-32.77	-32.77	-37.28	-36.98
3-hexanol	-30.05	-30.20	-35.93	-42.11	2-methoxyethanol	-23.67	-23.55	-28.54	-28.30
4-heptanol	-32.39	-32.30	-38.07	-46.13	2-propoxyethanol	-28.89	-29.13	-35.76	-34.56
1,2-ethanediol	-20.64	-20.50	-26.15	-26.77	2-butoxyethanol	-31.29	-31.45	-37.94	-38.01
cyclopentanol	-25.94	-25.90	-31.68	-35.18	3-hydroxybenzaldehyde	-28.64	-28.69	-28.40	-24.87
cyclohexanol	-28.41	-28.54	-34.26	-37.87	3-hydroxybenzonitrile	-28.33	-28.45	-28.21	-24.27
cycloheptanol	-30.80	-30.80	-34.85	-41.37	fluoromethane	-14.73	-14.85	-12.76	-12.17
phenol	-25.69	-25.74	-25.84	-24.06	2,2,2-trifluoroethanol	-21.69	-21.65	-22.03	-24.51
o-cresol	-28.23	-28.31	-28.93	-30.54	formic acid	-16.06	-16.20	-19.96	-19.49
p-cresol	-28.24	-28.17	-28.68	-28.83	N,N-dimethylformamide	-23.27	-23.21	-22.19	-24.23
4-tert-butylphenol	-35.10	-34.91	-37.87	-36.16	acetamide	-20.13	-20.04	-21.62	-22.17
pyrrolidine	-23.82	-23.76	-29.74	-32.57	N-methylacetamide	-23.19	-23.19	-28.57	-31.06
pyridine	-23.61	-23.52	-26.77	-24.65	tetrahydro-2-furanmethanol	-27.31	-27.34	-29.25	-29.47

Table A.4.1. ΔS_{cav}^{SPT}, ΔS_{cav}^{IT}, and ΔS_{hyd} values of 90 solute molecules calculated at 298 K (in cal·mol^{-1}·K^{-1}) in comparison with the corresponding experimental hydration entropy (ΔS_{hyd}^{X}).

compound	ΔA_{elec}^{268}	ΔA_{elec}^{328}	compound	ΔA_{elec}^{268}	ΔA_{elec}^{328}
methane	0.02	0.03	3-methylpyridine	-4.72	-4.86
ethane	-0.96	-0.94	4-methylpyridine	-6.10	-6.24
propane	-0.07	-0.04	3-ethylpyridine	-4.81	-4.98
butane	-0.43	-0.41	4-ethylpyridine	-4.76	-4.88
2-methylpropane	0.90	0.94	3,4-dimethylpyridine	-5.11	-5.28
pentane	-0.53	-0.50	3,5-dimethylpyridine	-5.49	-5.65
2,2-dimethylpropane	0.78	0.82	4-tert-butylpyridine	-5.19	-5.35
hexane	-0.65	-0.63	acetonitrile	-3.87	-3.75
heptane	-0.76	-0.75	propanenitrile	-3.58	-3.51
ethene	-1.94	-1.91	2-butanone	-4.29	-4.52
1-propene	-1.38	-1.28	2-pentanone	-4.47	-4.67
cyclopropane	-0.65	-0.54	3-pentanone	-4.68	-4.92
benzene	-2.06	-2.04	3-methyl-2-butanone	-4.26	-4.58
methylbenzene	-1.89	-1.82	2-hexanone	-4.80	-5.03
ethylbenzene	-1.80	-1.77	4-methyl-2-pentanone	-4.56	-4.85
m-xylene	-1.73	-1.68	2-heptanone	-4.67	-4.95
p-xylene	-1.54	-1.49	4-heptanone	-4.28	-4.60
propylbenzene	-1.73	-1.69	2-nonanone	-4.43	-4.69
(1-methylethyl)benzene	-1.96	-1.97	5-nonanone	-4.74	-5.02
9H-fluorene	-3.73	-3.69	propanoic acid	-8.13	-8.28
anthracene	-3.99	-3.91	butanoic acid	-8.14	-8.29
methanol	-5.81	-5.97	diethylether	-1.67	-1.92
ethanol	-5.99	-6.23	1,2-dimethoxyethane	-1.13	-1.35
1-propanol	-5.58	-5.85	tetrahydrofuran	-1.55	-1.70
2-propanol	-5.19	-5.52	tetrahydropyran	-1.68	-2.03
1-butanol	-5.40	-5.71	2-methyltetrahydrofuran	-1.45	-1.65
2-methyl-1-propanol	-5.57	-5.90	2,5-dimethyltetrahydrofuran	-1.28	-1.53
2-butanol	-5.72	-6.07	1,3-dioxolane	-3.38	-3.43
2-methyl-2-propanol	-6.09	-6.43	methylacetate	-4.53	-4.62
1-pentanol	-5.29	-5.51	ethylacetate	-4.43	-4.62
3-pentanol	-5.34	-5.74	methylbutanoate	-4.06	-4.24
2-methyl-2-butanol	-5.23	-5.63	N-butylacetate	-4.50	-4.77
1-hexanol	-5.17	-5.46	ethylpentanoate	-4.80	-5.07
3-hexanol	-5.74	-6.09	2-methoxyethanol	-8.74	-9.04
4-heptanol	-5.90	-6.24	2-propoxyethanol	-8.64	-9.04
1,2-ethanediol	-12.22	-12.56	2-butoxyethanol	-8.84	-9.23
cyclopentanol	-5.41	-5.76	3-hydroxybenzaldehyde	-7.71	-7.69
cyclohexanol	-5.14	-5.49	3-hydroxybenzonitrile	-9.98	-9.97
cycloheptanol	-5.70	-5.94	fluoromethane	-1.12	-1.00
phenol	-7.51	-7.52	2,2,2-trifluoroethanol	-8.49	-8.51
o-cresol	-7.66	-7.70	formic acid	-8.17	-8.40
p-cresol	-7.81	-7.84	N,N-dimethylformamide	-6.18	-6.39
4-tert-butylphenol	-7.75	-7.92	acetamide	-9.32	-9.41
pyrrolidine	-5.43	-5.79	N-methylacetamide	-9.39	-9.71
pyridine	-4.84	-5.03	tetrahydro-2-furanmethanol	-8.39	-8.51

Table A.4.2. ΔA_{elec} values of 90 solute molecules calculated at 268 and 328 K (in kcal·mol^{-1}).

compound	ΔS_{hyd} *(forward)*	ΔS_{hyd} *(backward)*	ΔS_{hyd}	ΔS_{hyd}^{X}	*estimation error*
methane	-22.33	-21.51	-21.92 (±0.58)	-16.33	5.59
propane	-32.84	-33.24	-33.04 (±0.28)	-22.91	10.13
butane	-40.42	-43.39	-41.91 (±2.10)	-26.85	15.05
2-methylpropane	-55.42	-55.16	-55.29 (±0.18)	-25.17	30.11
methanol	-33.73	-31.75	-32.74 (±1.40)	-18.53	14.21
ethanol	-35.98	-38.05	-37.01 (±1.46)	-23.02	13.99
toluene	-63.43	-63.17	-63.30 (±0.18)	-26.09	37.21
p-cresol	-53.22	-57.85	-55.53 (±3.27)	-28.83	26.71
acetamide	-41.73	-41.20	-41.47 (±0.37)	-22.17	19.30
N-methylacetamide	-52.13	-47.39	-49.76 (±3.35)	-31.06	18.70
acenaphthene	-58.79	-64.47	-61.63 (±4.02)	-29.22	32.41
phenanthrene	-60.09	-62.12	-61.10 (±1.43)	-30.35	30.76
pyrene	-64.91	-72.01	-68.46 (±5.02)	-36.09	32.37
piperidine	-41.15	-44.45	-42.80 (±2.33)	-35.30	7.50
2,6-ditert-butylpyridine	-71.33	-75.46	-73.39 (±2.92)	-44.35	29.04
2-adamantanone	-59.17	-55.07	-57.12 (±2.90)	-30.99	26.13
1,4-dioxane	-35.93	-40.65	-38.29 (±3.34)	-21.90	16.39
morpholine	-34.36	-32.76	-33.56 (±1.13)	-31.61	1.95

Table A.4.3. ΔS_{hyd} values (in cal·mol^{-1}·K^{-1}) of 18 reference molecules calculated from forward and backward full FEP simulations in comparison with experimental results (ΔS_{hyd}^{X}).

compound	ΔS_{cav}^{SPT}	ΔS_{cav}^{IT}	ΔS_{elec}	$\Delta S_{hyd}^{SPT/FP}$	ΔS_{hyd}^{X}	*error*
methane	-13.56	-13.79	0.20	-13.35	-16.33	2.98
propane	-20.63	-20.51	0.57	-20.05	-22.91	2.86
butane	-23.57	-23.54	0.42	-23.14	-26.85	3.71
2-methylpropane	-23.57	-23.49	0.58	-22.99	-25.17	2.18
methanol	-15.57	-15.62	-2.52	-18.08	-18.53	0.45
ethanol	-19.05	-18.93	-3.92	-22.97	-23.02	0.05
toluene	-27.01	-27.01	1.18	-25.83	-26.09	0.26
p-cresol	-28.23	-28.16	-0.58	-28.81	-28.83	0.01
acetamide	-20.13	-20.04	-1.53	-21.66	-22.17	0.51
N-methylacetamide	-23.19	-23.18	-5.47	-28.66	-31.06	2.40
acenaphthene	-31.43	-31.53	0.99	-30.45	-29.22	1.23
phenanthrene	-35.85	-35.97	1.06	-34.79	-30.35	4.44
pyrene	-39.55	-39.64	1.72	-37.83	-36.09	1.74
piperidine	-26.27	-26.14	-5.82	-32.09	-35.30	3.21
2,6-ditert-butylpyridine	-42.84	-42.70	-1.57	-44.41	-44.35	0.06
2-adamantanone	-33.88	-33.97	-1.48	-35.36	-30.99	4.37
1,4-dioxane	-24.68	-24.70	-2.59	-27.27	-21.90	5.37
morpholine	-25.53	-25.56	-6.50	-32.03	-31.61	0.42

Table A.4.4. ΔS_{cav} values of 18 reference molecules calculated from SPT and IT, ΔS_{elec} values from partial FEP simulations, and comparison of ΔS_{hyd} values obtained with hybrid SPT/FEP with those of experimental data. The unit of all entropy values is cal·mol^{-1}·K^{-1}.

compound	ΔA_{elec}	ΔA_{vdw}	ΔA_{aq}	ΔA_{gas}	ΔA_{hyd}
methane	-0.02	1.42	-1.44	0.00	1.44(±0.12)
propane	-0.05	1.07	-4.02	-3.00	1.01(±0.09)
butane	-0.41	1.78	-5.60	-4.30	1.30(±0.21)
2-methylpropane	0.93	1.99	14.51	17.33	2.81(±0.15)
methanol	-6.01	0.09	-9.59	-15.32	-5.73(±0.16)
ethanol	-5.82	1.11	10.84	5.96	-4.88(±0.21)
toluene	-1.91	0.25	-4.88	-6.52	-1.64(±0.23)
p-cresol	-7.68	4.53	10.58	7.24	-3.35(±0.18)
acetamide	-9.17	0.65	87.65	79.01	-8.64(±0.48)
N-methylacetamide	-9.24	0.55	40.36	31.12	-9.24(±0.56)
acenaphthene	-1.17	-2.41	-29.62	-33.35	-3.73(±0.23)
phenanthrene	-0.82	-4.02	-38.60	-43.35	-4.76(±0.18)
pyrene	-0.96	-5.01	-44.49	-50.07	-5.58(±0.47)
piperidine	-6.81	0.11	36.87	30.55	-6.33(±0.22)
2,6-ditert-butylpyridine	-4.33	-0.95	-24.00	-29.10	-5.10(±0.45)
2-adamantanone	-5.87	-0.54	1.56	-5.18	-6.74(±0.47)
1,4-dioxane	-6.28	0.97	24.69	18.91	-5.77(±0.36)
morpholine	-8.44	0.35	26.99	19.14	-7.85(±0.57)

Table A.4.5. ΔA_{elec}, ΔA_{vdw}, ΔA_{aq}, ΔA_{gas}, and ΔA_{hyd} values of 18 reference molecules calculated from full FEP simulations at 268 K. All energy values are given in kcal/mol.

compound	ΔA_{elec}	ΔA_{vdw}	ΔA_{aq}	ΔA_{gas}	ΔA_{hyd}
methane	0.01	2.68	-2.76	0.00	2.76(±0.09)
propane	-0.03	3.06	-6.22	-3.23	2.99(±0.11)
butane	-0.45	4.47	-8.54	-4.73	3.81(±0.13)
2-methylpropane	1.02	5.17	11.28	17.41	6.13(±0.34)
methanol	-6.12	2.22	-11.46	-15.22	-3.76(±0.21)
ethanol	-6.13	3.57	8.91	6.25	-2.66(±0.13)
toluene	-1.83	3.96	-8.64	-6.48	2.15(±0.16)
p-cresol	-7.30	7.29	7.44	7.42	-0.02(±0.23)
acetamide	-9.64	3.45	85.53	79.38	-6.15(±0.18)
N-methylacetamide	-9.33	3.48	37.09	30.84	-6.25(±0.34)
acenaphthene	-1.16	1.14	-34.04	-34.07	-0.03(±0.43)
phenanthrene	-1.35	0.23	-44.58	-45.68	-1.10(±0.12)
pyrene	-0.24	-1.37	-49.37	-50.85	-1.48(±0.23)
piperidine	-6.50	2.51	32.88	29.12	-3.76(±0.26)
2,6-ditert-butylpyridine	-5.26	4.53	-30.33	-31.02	-0.69(±0.08)
2-adamantanone	-5.17	1.77	-2.01	-5.32	-3.31(±0.32)
1,4-dioxane	-6.93	3.64	21.58	18.11	-3.47(±0.24)
morpholine	-8.36	2.34	23.96	18.12	-5.84(±0.56)

Table A.4.6. ΔA_{elec}, ΔA_{vdw}, ΔA_{aq}, ΔA_{gas}, and ΔA_{hyd} values of 18 reference molecules calculated from full FEP simulations at 328 K. All energy values are given in kcal/mol.

atom type	description	V_j (Å^3)	O_i^{max} (Å^3)	S_i (kcal/molÅ^3)
C.3_1	sp^3 carbon with 1 substituent	13.417	373.0	-0.952
C.3_2	sp^3 carbon with 2 substituents	10.308	306.3	-1.810
C.3_3	sp^3 carbon with 3 substituents	14.130	395.2	-1.238
C.3_4	sp^3 carbon with 4 substituents	8.904	387.3	0.952
C.2	sp^2 carbon	13.661	374.6	-1.914
C.ar_2	aromatic carbon with 2 substituents	9.237	314.3	-3.238
C.ar_3	aromatic carbon with 3 substituents	16.349	376.2	-2.349
C.CO_1	carbonyl carbon with 1 substituent	10.354	377.8	-5.238
C.CO_2	carbonyl carbon with 2 substituents	9.915	379.4	-5.762
C.1	sp carbon	8.656	385.7	-3.619
N.3_1	sp^3 nitrogen	13.718	373.4	-19.514
N.am	amidic nitrogen	9.236	336.9	-17.619
N.ar	aromatic nitrogen	18.684	356.0	-17.937
N.1	sp nitrogen	8.303	344.8	-12.540
O.3_1	sp^3 oxygen with 1substituent	6.186	350.0	-22.286
O.3_2	sp^3 oxygen with 2 substituents	12.772	310.3	-10.333
O.2	sp^2 oxygen	7.064	305.6	-16.332
O.pl	planar oxygen	11.016	270.6	-23.048
O.es_1	sp^3 oxygen in carboxylic acids	14.831	307.1	-16.619
O.es_2	sp^3 oxygen in esters	18.328	261.6	-4.254
F	fluorine	6.075	225.0	-3.810
H.C	hydrogen bonded to carbon	2.627	177.0	-2.000
H.N3	hydrogen bonded to sp^3 nitrogen	5.442	245.2	-7.334
H.Na	hydrogen bonded to amidic nitrogen	1.262	197.6	-7.676
H.O3	hydrogen bonded to sp^3 oxygen	1.302	216.7	-9.619
H.Op	hydrogen bonded to planar nitrogen	4.082	154.8	-10.730
H.Oa	hydrogen in carboxylic acids	4.630	184.9	-2.000

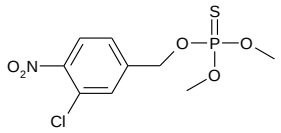
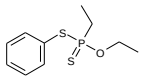
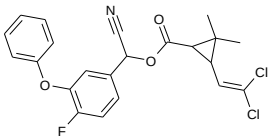
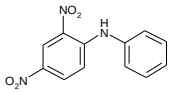
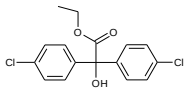
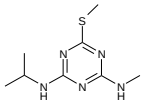
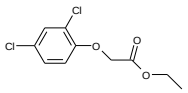
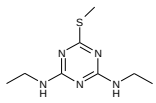
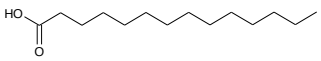
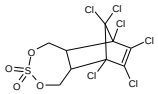
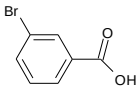
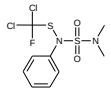
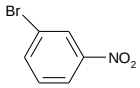
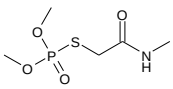
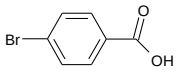
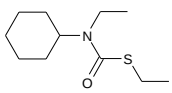
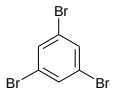
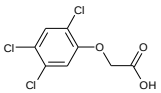
Table A.4.7. The optimized atomic fragmental volume (V_j), maximum atomic occupancy (O_i^{max}), and atomic solvation (S_i) parameters for various atom types.

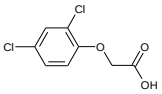
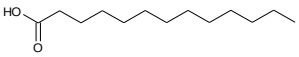
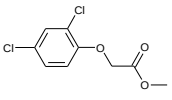
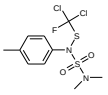
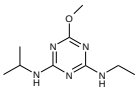
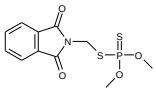
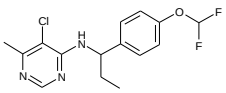
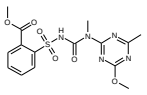
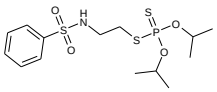
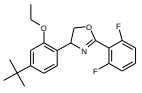
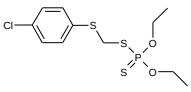
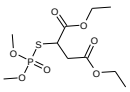
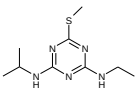
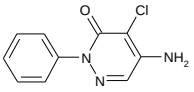
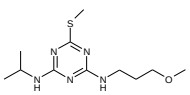
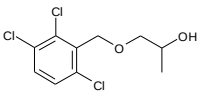
compound	ΔH_{hyd}^{calc}	ΔH_{hyd}^{exp}	compound	ΔH_{hyd}^{calc}	ΔH_{hyd}^{exp}
methane*	-2.44	-2.87	3-methylpyridine*	-12.76	-13.05
ethane	-4.92	-4.27	4-methylpyridine	-14.79	-13.27
propane*	-4.82	-4.88	3-ethylpyridine*	-13.60	-12.78
butane	-6.66	-5.93	4-ethylpyridine	-14.28	-13.20
2-methylpropane*	-5.82	-5.19	3,4-dimethylpyridine*	-13.55	-13.52
pentane	-7.35	-6.76	3,5-dimethylpyridine	-15.54	-14.46
2,2-dimethylpropane*	-5.82	-5.59	4-tert-butylpyridine*	-14.26	-13.83
hexane	-8.01	-7.55	acetonitrile*	-7.17	-8.34
heptane*	-8.96	-8.10	propanenitrile	-9.66	-9.48
ethane*	-3.27	-3.68	2-butanone*	-10.67	-10.92
1-propene	-4.30	-3.27	2-pentanone	-12.96	-11.64
cyclopropane	-5.25	-5.02	3-pentanone*	-11.65	-11.85
benzene*	-7.83	-7.59	3-methyl-2-butanone	-12.26	-13.77
methylbenzene*	-8.74	-8.67	2-hexanone*	-12.40	-12.38
ethylbenzene	-10.84	-9.62	4-methyl-2-pentanone	-13.64	-13.79
m-xylene*	-9.23	-9.23	2-heptanone*	-13.60	-13.60
p-xylene	-11.21	-9.79	4-heptanone	-15.03	-13.89
propylbenzene*	-10.50	-10.49	2-nonanone*	-16.10	-14.90
(1-ethylethyl)benzene	-10.68	-9.91	5-nonanone	-16.10	-15.01
9H-fluorene	-14.88	-13.12	propanoic acid*	-13.86	-13.51
anthracene*	-14.13	-14.00	butanoic acid	-12.54	-14.22
methanol*	-11.26	-10.64	diethylether	-10.75	-11.09
ethanol	-11.94	-12.09	1,2-dimethoxyethane*	-12.09	-14.20
1-propanol*	-13.15	-13.73	tetrahydrofuran	-10.18	-11.30
2-propanol	-13.01	-13.91	tetrahydropyran*	-10.13	-11.68
1-butanol	-15.24	-14.72	2-methyltetrahydrofuran	-11.26	-12.28
2-methyl-1-propanol	-14.39	-14.38	2,5-dimethyltetrahydrofuran	-12.33	-13.46
2-butanol*	-13.77	-14.99	1,3-dioxolane*	-9.59	-9.54
2-methyl-2-propanol*	-13.72	-15.28	methylacetate	-10.69	-10.16
1-pentanol	-15.83	-14.79	ethylacetate*	-11.56	-10.90
3-pentanol*	-14.62	-14.24	methylbutanoate	-12.91	-11.50
2-methyl-2-butanol	-14.58	-15.69	N-butylacetate*	-13.56	-12.60
1-hexanol*	-16.87	-15.82	ethylpentanoate	-15.46	-13.50
3-hexanol*	-17.47	-16.63	2-methoxyethanol*	-15.46	-14.45
4-heptanol*	-16.66	-18.00	2-propoxyethanol	-17.99	-16.72
1,2-ethanediol	-18.62	-17.28	2-butoxyethanol*	-18.19	-17.60
cyclopentanol*	-14.03	-15.98	3-hydroxybenzaldehyde	-15.11	-16.91
cyclohexanol	-17.00	-16.85	3-hydroxybenzonitrile*	-17.68	-16.75
cycloheptanol*	-16.81	-17.82	fluoromethane	-3.13	-3.85
phenol	-14.84	-13.79	2,2,2-trifluoroethanol*	-12.03	-11.97
o-cresol*	-14.15	-14.98	formic acid	-9.71	-11.23
p-cresol	-15.63	-14.73	N,N-dimethylformamide*	-14.98	-15.03
4-tert-butylphenol*	-15.39	-15.25	Acetamide*	-16.21	-16.32
pyrrolidine*	-14.07	-15.20	N-methylacetamide	-18.27	-19.36
pyridine	-12.97	-11.91	tetrahydro-2-furanmethanol	-17.42	-16.09

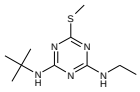
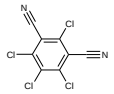
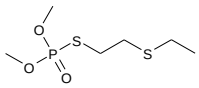
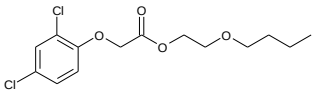
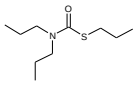
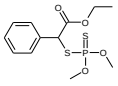
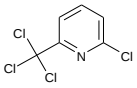
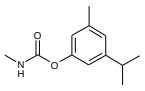
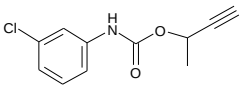
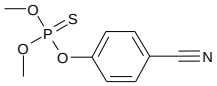
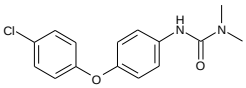
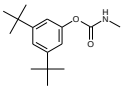
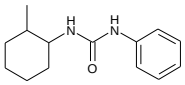
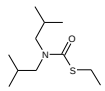
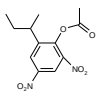
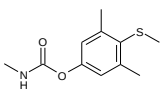
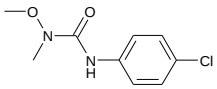
Table A.4.8. Molecular ΔH_{hyd} values of 90 solute molecules calculated with solvent contact model in comparison with the corresponding experimental results. All energy values are given in kcal/mol.*Indicated in asterisk are the molecules included in the training set.

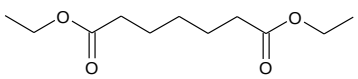
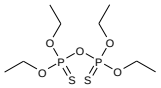
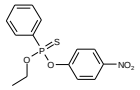
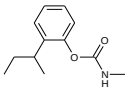
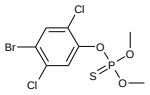
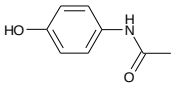
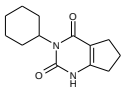
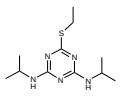
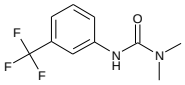
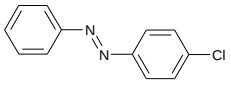
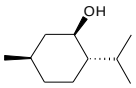
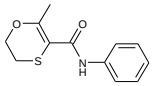
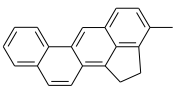
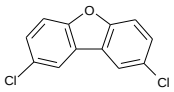
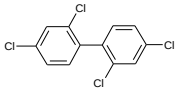
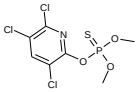
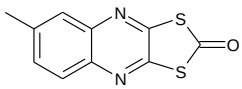
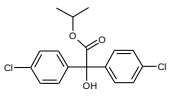
atom type	description	V_j (Å^3)	O_i^{max} (Å^3)	S_i (kcal/molÅ^3)	P_i (kcal/molÅ^3)
C.3_4	sp^3 carbon with 4 substituents	6.342	316.5	0.310	-0.905
C.3_3	sp^3 carbon with 3 substituents	7.977	321.4	0.302	-1.432
C.3_2	sp^3 carbon with 2 substituents	9.365	334.2	0.270	-2.055
C.3_1	sp^3 carbon with 1 substituent	11.341	343.0	1.937	-3.095
C.2_3	sp^2 carbon with 3 substituents	8.571	322.3	-0.429	-1.397
C.2_2	sp^2 carbon with 2 substituents	9.365	338.2	1.914	-2.476
C.2_1	sp^2 carbon with 1 substituent	10.952	352.5	-1.381	-3.846
C.1_2	sp carbon with 2 substituents	8.968	325.3	-0.730	-3.201
C.1_1	sp carbon with 1 substituent	17.698	342.8	1.857	-3.286
C.ar_3	aromatic carbon with 3 substituents	7.342	346.9	1.324	-3.095
C.ar_2	aromatic carbon with 2 substituents	9.762	354.2	0.032	-3.787
N.3_3	sp^3 nitrogen with 3 substituents	6.587	314.5	-9.680	20.746
N.3_2	sp^3 nitrogen with 2 substituents	7.853	327.3	-15.984	19.305
N.3_1	sp^3 nitrogen with 1 substituent	10.952	341.5	-16.776	10.095
N.2	sp^2 nitrogen	8.275	328.7	-8.079	5.643
N.1	sp nitrogen	10.952	364.3	-8.540	9.968
N.am_3	amidic nitrogen with 3 substituents	6.164	348.2	-10.619	7.286
N.am_2	amidic nitrogen with 2 substituents	8.653	356.2	-12.571	16.220
N.am_1	amidic nitrogen with 1 substituent	11.349	368.3	-20.952	20.048
N.ar	aromatic nitrogen	8.243	338.3	-7.937	8.397
N.pl3_3	planar nitrogen with 3 substituents	6.984	321.8	-1.587	5.873
N.pl3_2	planar nitrogen with 2 substituents	8.245	330.6	-14.021	9.143
N.pl3_1	planar nitrogen with 1 substituent	10.556	338.4	-14.571	9.422
O.3_2	sp^3 oxygen with 2 substituents	6.532	363.8	-7.831	9.571
O.3_1	sp^3 oxygen with 1 substituent	5.397	371.6	-15.089	12.857
O.2	sp^2 oxygen	7.833	344.3	-5.032	4.873
S.3	sp^3 sulfur	16.896	340.6	-1.905	-2.307
S.2	sp^2 sulfur	17.143	348.3	1.190	9.397
S.O	sulfoxide sulfur	13.810	345.2	-5.857	1.381
S.O2	sulfone sulfur	11.905	338.2	4.315	3.984
P	phosphorine	10.029	330.2	5.286	-17.857
F	fluorine	8.454	326.0	1.365	-2.303
Cl	chlorine	16.905	326.3	0.435	-2.937
Br	bromine	22.460	330.4	0.185	-5.073
H.C	hydrogen bonded to carbon	2.714	366.4	-0.476	-6.048
H.N	hydrogen bonded to nitrogen	1.786	364.8	4.730	-17.863
H.O	hydrogen bonded to oxygen	5.571	362.1	6.524	-22.032

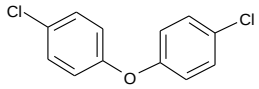
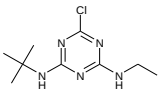
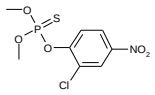
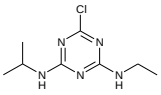
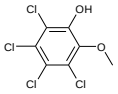
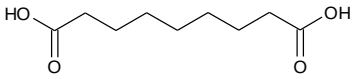
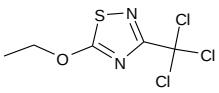
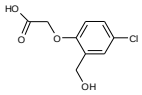
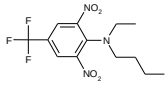
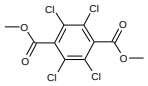
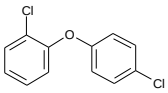
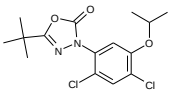
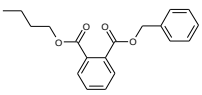
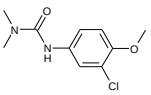
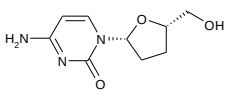
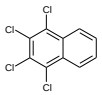
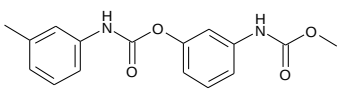
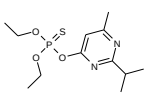
Table A.4.9. The optimized atomic fragmental volume (V_j), maximum atomic occupancy (O_i^{max}), and atomic solvation parameters (S_i) in the solvation model including self-solvation effects.

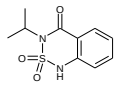
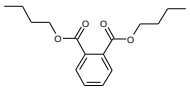
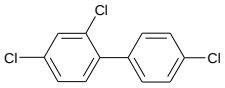
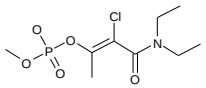
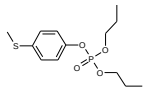
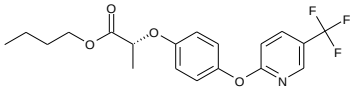
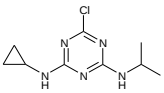
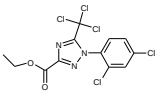
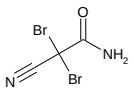
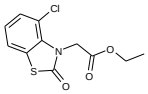
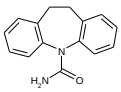
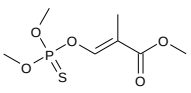
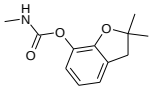
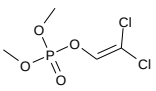
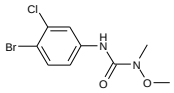
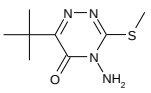
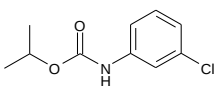
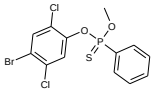
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.77	-7.24		-4.83	-3.95
	-7.95	-9.35		-12.22	-10.18
	-7.41	-9.11		-10.33	-9.74
	-5.29	-4.84		-10.56	-10.17
	-6.54	-7.57		-6.65	-6.67
	-7.65	-7.59		-7.92	-7.30
	-5.62	-4.40		-12.8	-12.32
	-7.84	-7.58		-4.99	-5.08
	-2.04	-2.42		-7.69	-8.47

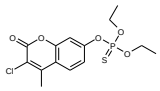
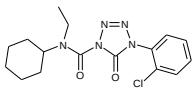
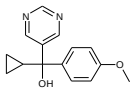
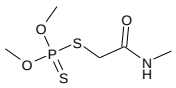
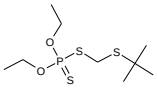
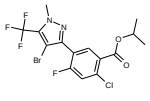
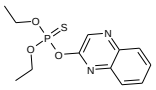
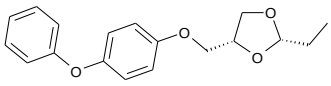
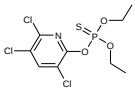
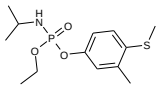
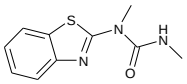
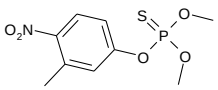
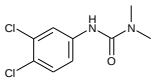
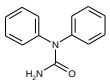
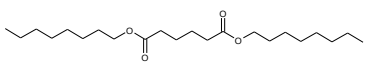
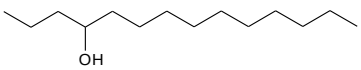
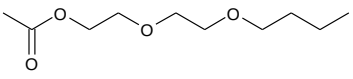
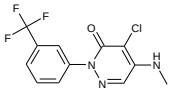
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-3.09	-3.48		-7.83	-8.61
	-7.3	-7.36		-5.38	-5.06
	-6.04	-6.76		-10.37	-9.66
	-8.82	-9.47		-8.03	-8.32
	-4.87	-5.48		-13.34	-13.18
	-8.41	-8.52		-7.33	-5.93
	-6.78	-6.24		-10.26	-11.37
	-9.27	-9.30		-10.56	-11.88
	-10.57	-10.76		-6.41	-6.44

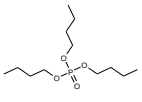
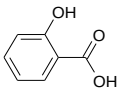
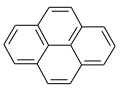
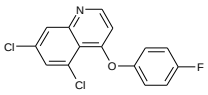
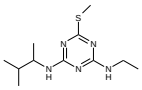
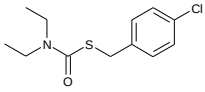
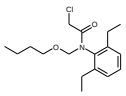
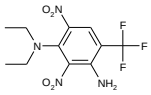
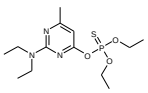
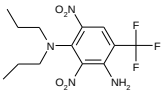
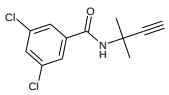
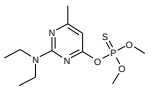
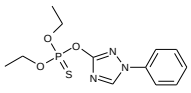
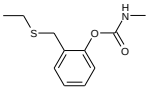
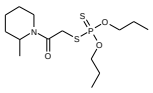
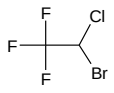
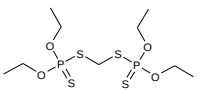
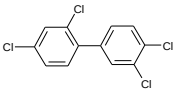
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-8.26	-8.71		-6.81	-5.57
	-7.98	-7.58		-7.08	-6.19
	-3.95	-4.28		-7.35	-7.53
	-4.2	-2.93		-7.42	-7.52
	-8.65	-8.84		-4.89	-7.33
	-10.61	-11.32		-7.68	-8.50
	-11.67	-9.80		-3.24	-3.07
	-3.37	-3.98		-7.22	-7.38
	-9.43	-8.30		-9.41	-10.60

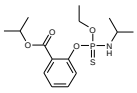
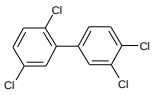
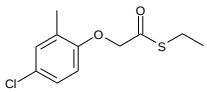
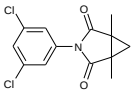
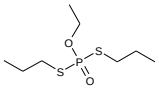
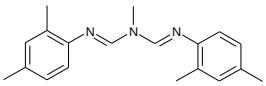
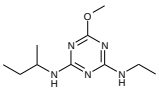
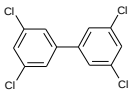
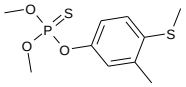
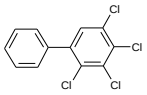
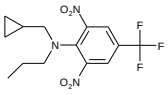
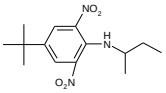
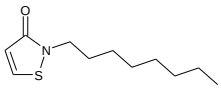
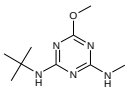
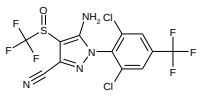
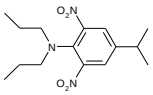
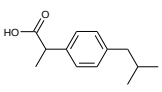
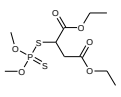
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-6.46	-5.45		-5.65	-5.34
	-7.19	-8.89		-7.4	-7.74
	-2.79	-4.97		-10.14	-11.17
	-11.59	-10.01		-7.99	-8.44
	-9.51	-7.68		-5.82	-5.89
	-4	-3.86		-10.76	-9.39
	-5	-4.17		-3.53	-2.99
	-2.18	-2.62		-5.2	-5.28
	-7.51	-6.76		-7.59	-8.93

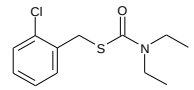
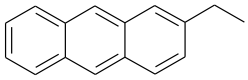
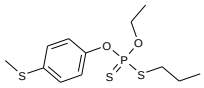
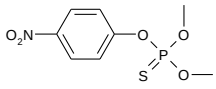
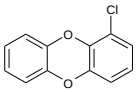
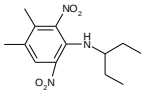
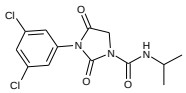
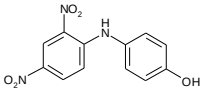
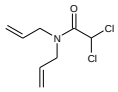
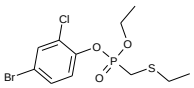
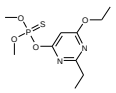
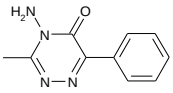
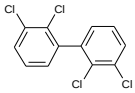
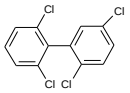
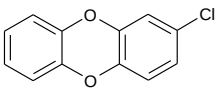
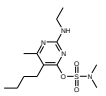
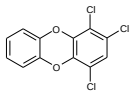
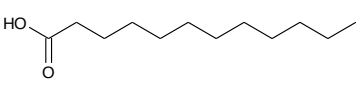
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-3.78	-3.76		-7.89	-8.94
	-7.25	-7.22		-9.57	-9.57
	-4.84	-5.44		-14.11	-13.15
	-6.74	-7.14		-11.47	-12.66
	-2.63	-3.58		-5.52	-6.05
	-2.6	-3.83		-7.46	-7.50
	-5.85	-7.30		-8.36	-9.61
	-15.09	-15.91		-2.74	-1.54
	-14.27	-15.44		-6.05	-4.99

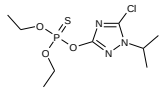
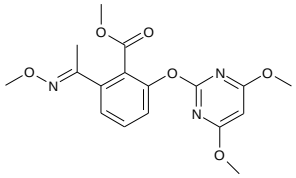
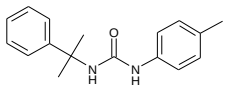
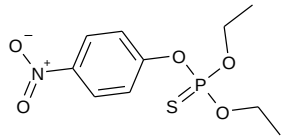
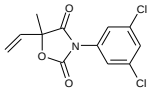
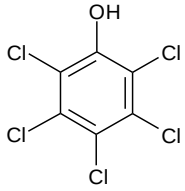
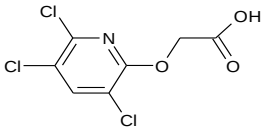
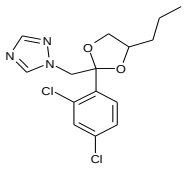
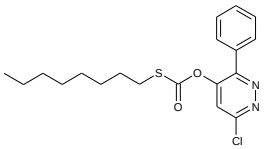
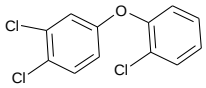
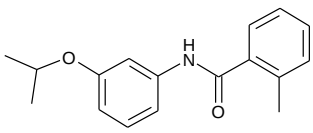
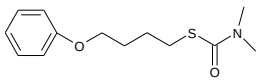
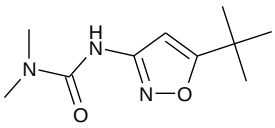
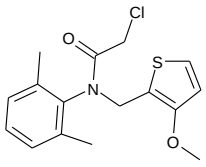
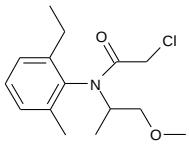
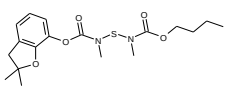
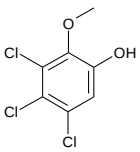
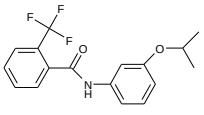
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-9.45	-8.12		-6.23	-5.55
	-2.73	-2.56		-13.06	-11.25
	-9.45	-9.27		-7.5	-7.22
	-9.92	-9.89		-9.11	-8.36
	-8.33	-8.46		-8.13	-8.31
	-9.39	-9.41		-5.91	-6.17
	-9.01	-8.47		-6.31	-7.54
	-9.05	-10.41		-11.16	-11.20
	-6.21	-7.14		-6.39	-6.04

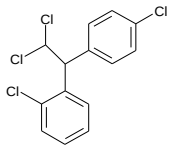
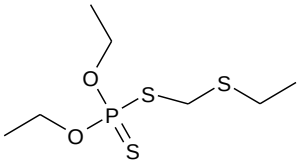
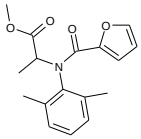
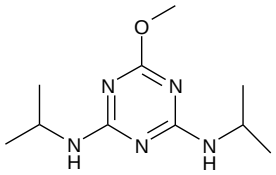
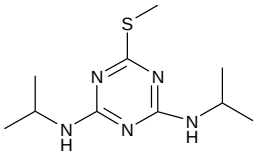
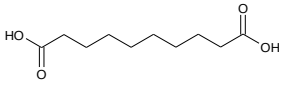
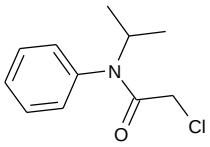
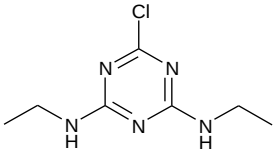
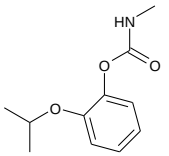
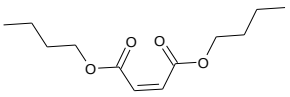
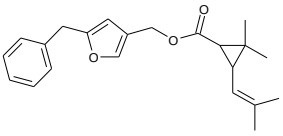
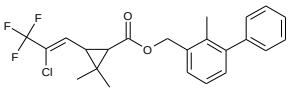
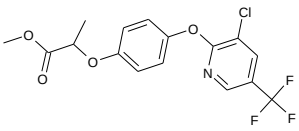
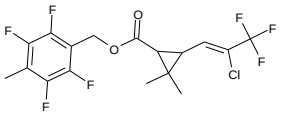
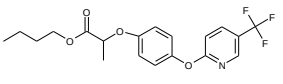
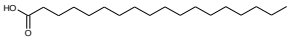
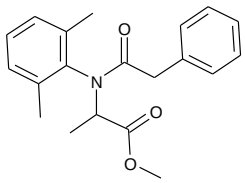
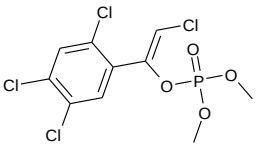
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.91	-7.90		-11.41	-13.82
	-11.42	-10.91		-10.91	-9.04
	-4.1	-4.57		-6.03	-5.88
	-7.67	-7.29		-7.59	-7.75
	-4.73	-5.15		-9.96	-9.40
	-10.42	-11.92		-6.27	-6.63
	-10.48	-8.50		-8.51	-7.85
	-6.25	-6.73		-4.66	-4.60
	-6.5	-4.99		-10.71	-10.18

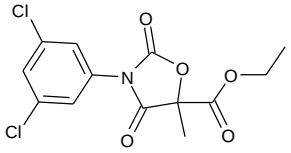
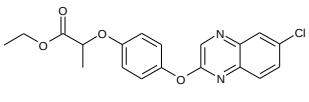
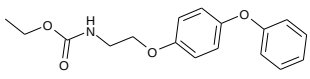
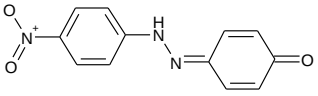
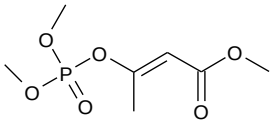
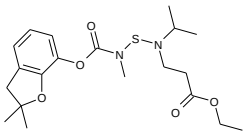
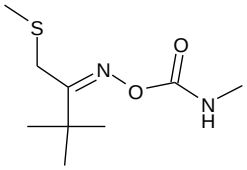
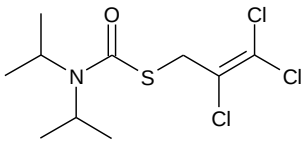
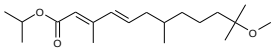
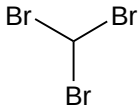
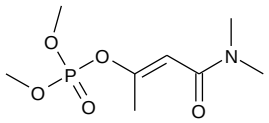
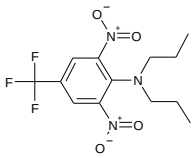
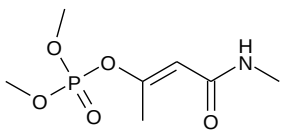
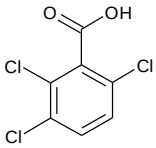
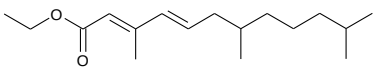
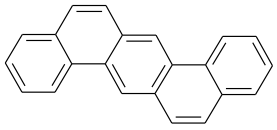
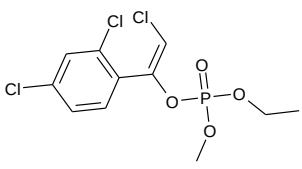
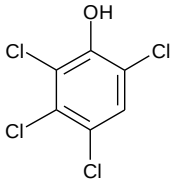
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-5.78	-7.52		-8.95	-10.96
	-4.69	-5.11		-6.37	-5.18
	-8.6	-8.63		-6.77	-5.87
	-7.66	-7.22		-5.79	-5.99
	-3.93	-4.63		-6.06	-5.54
	-8.73	-7.27		-6.09	-4.66
	-8.18	-8.51		-9.99	-8.87
	-9.03	-8.17		-0.14	0.40
	-6.56	-7.50		-2.32	-2.47

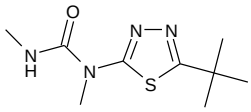
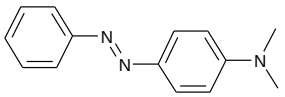
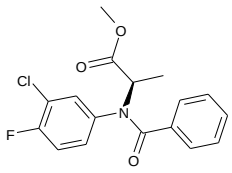
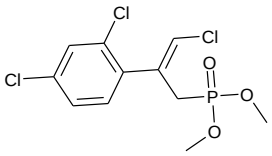
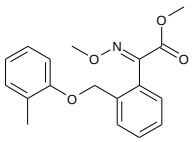
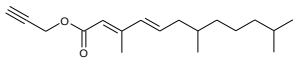
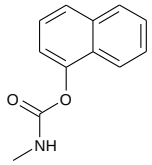
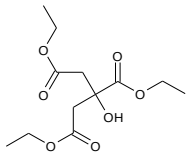
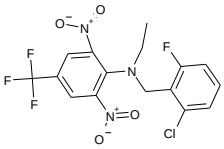
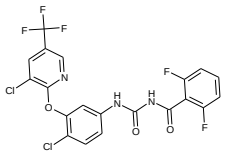
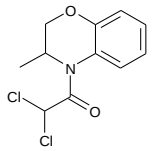
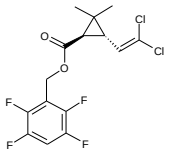
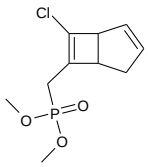
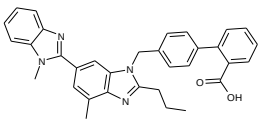
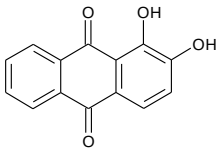
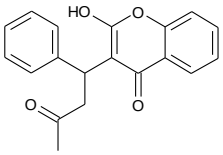
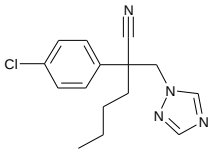
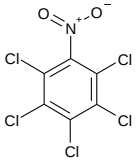
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.58	-6.69		-2.46	-2.51
	-4.08	-3.84		-4.53	-5.70
	-7.06	-7.02		-6.14	-4.84
	-9.18	-9.37		-1.51	-2.39
	-6.38	-4.88		-1.88	-2.55
	-2.59	-3.81		-5.04	-5.35
	-8.28	-5.85		-9.18	-8.92
	-10.18	-10.15		-3.19	-3.77
	-6.27	-6.94		-8.7	-7.84

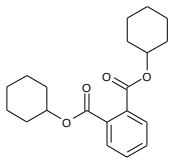
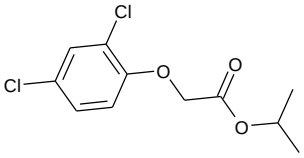
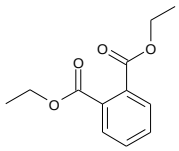
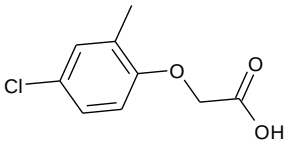
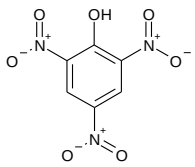
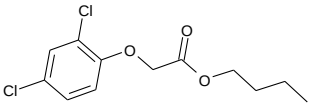
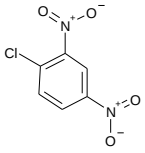
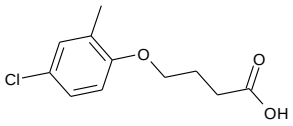
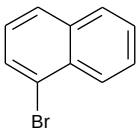
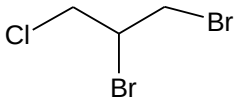
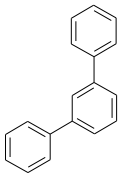
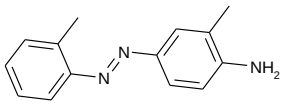
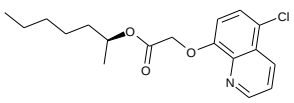
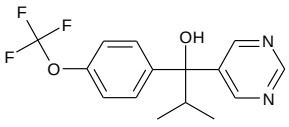
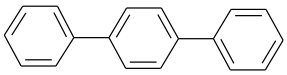
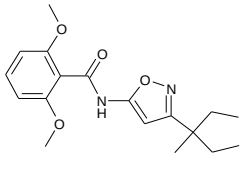
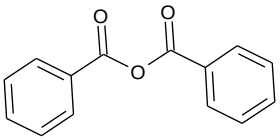
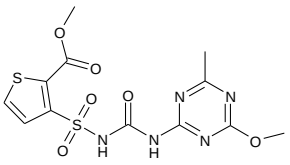
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-5.72	-6.02		-3.63	-2.87
	-5.69	-5.75		-8.02	-7.37
	-3.54	-4.40		-4.33	-4.60
	-11.34	-12.76		-13.94	-14.14
	-6.64	-6.52		-8.44	-9.10
	-6.14	-6.38		-13.96	-13.24
	-1.54	-2.72		-2.9	-2.87
	-4.54	-4.36		-8.5	-6.99
	-3.9	-4.13		-6.06	-7.14

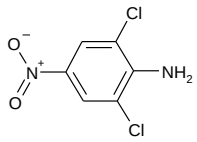
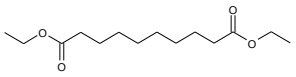
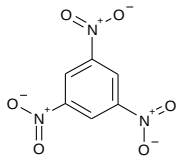
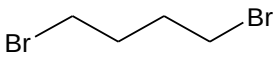
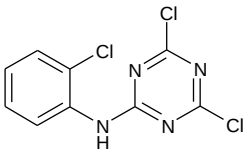
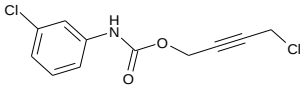
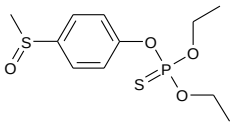
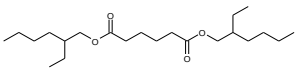
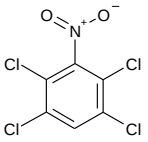
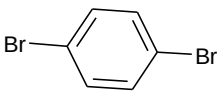
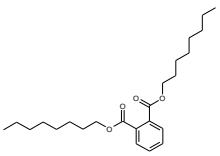
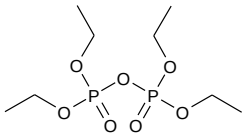
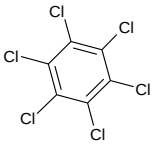
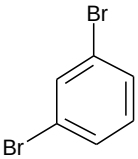
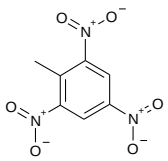
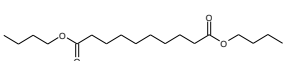
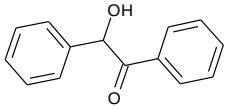
chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-6.37	-6.27		-9.96	-10.84
	-10.08	-10.16		-6.73	-7.36
	-8.24	-7.99		-5.38	-4.27
	-10.1	-9.73		-9.75	-8.08
	-10.56	-9.25		-3.99	-3.78
	-8.47	-7.74		-8.46	-7.68
	-11.26	-10.58		-8.83	-9.24
	-8.24	-6.72		-9.9	-10.67
	-6.06	-5.65		-9.39	-7.37

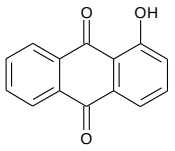
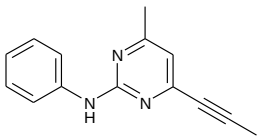
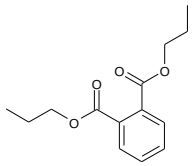
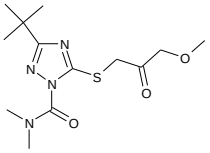
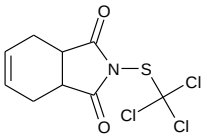
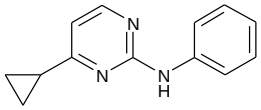
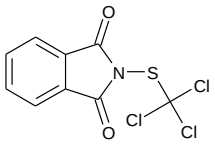
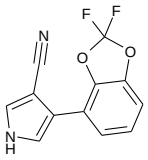
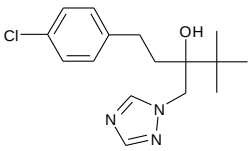
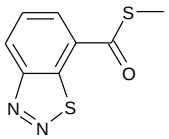
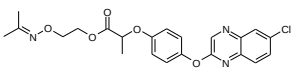
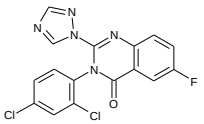
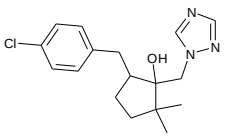
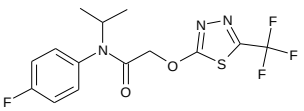
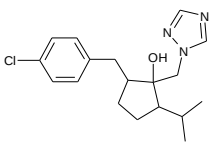
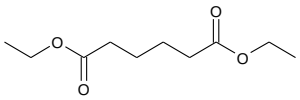
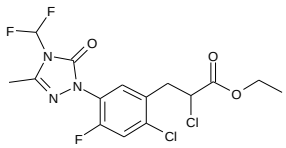
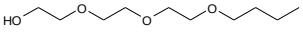
chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-4.74	-4.88		-5.11	-4.20
	-9.96	-9.07		-9.96	-8.79
	-8.61	-8.41		-12.57	-13.35
	-7.29	-5.59		-10.11	-10.45
	-9.7	-8.43		-6.57	-5.37
	-7.3	-7.15		-5.98	-5.98
	-6.66	-7.30		-1.57	-1.09
	-6.8	-7.23		-6.45	-8.42
	-7.26	-9.03		-9.55	-9.32

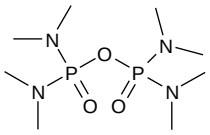
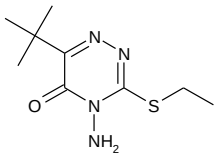
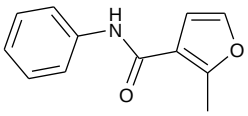
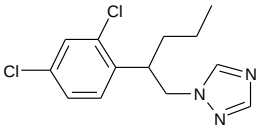
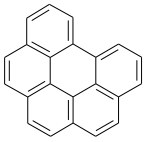
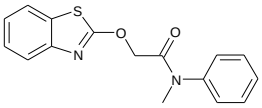
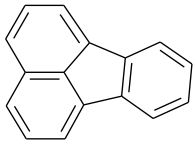
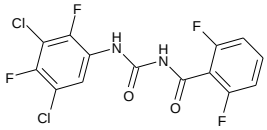
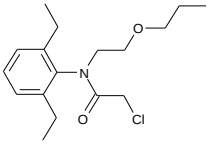
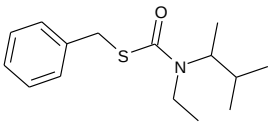
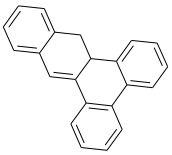
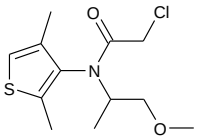
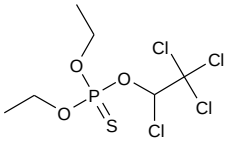
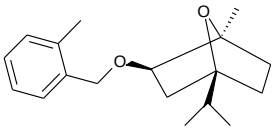
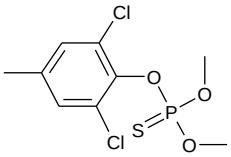
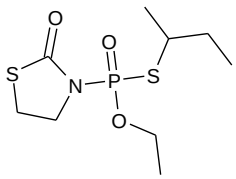
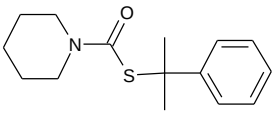
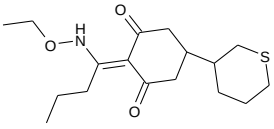
chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-8.27	-8.83		-8.53	-9.92
	-10.62	-12.24		-12.27	-11.61
	-11.84	9.83		-8.39	-8.23
	-8.75	-8.74		-4.51	-5.49
	-4.84	-5.67		-2.22	-1.19
	-11.85	-10.97		-3.2	-3.42
	-14.19	-13.24		-8.27	-6.46
	-3.76	-4.87		-7.15	-5.51
	-8.09	-9.64		-4.69	-4.44

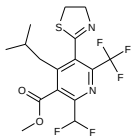
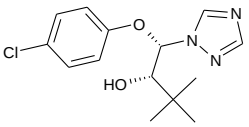
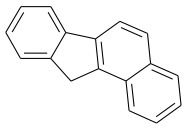
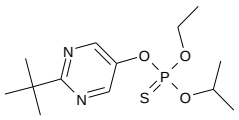
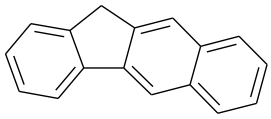
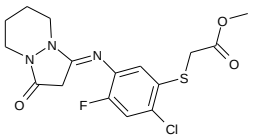
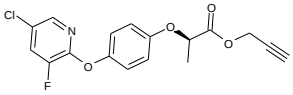
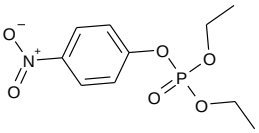
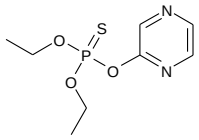
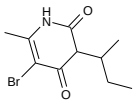
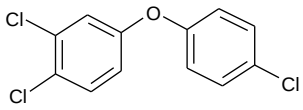
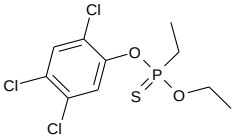
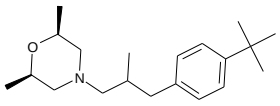
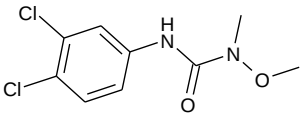
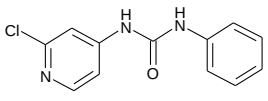
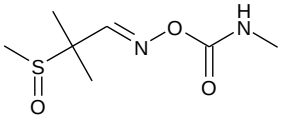
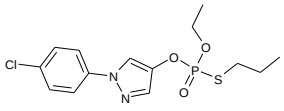
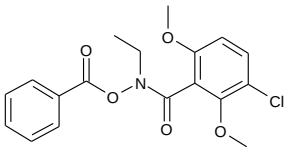
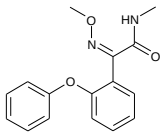
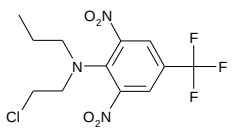
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-12.04	-12.16		-7.41	-5.35
	-6.8	-8.27		-8.01	-9.41
	-9.17	-8.72		-6.29	-6.29
	-9.37	-9.21		-9.28	-9.09
	-5.6	-5.21		-11.95	-14.19
	-7.39	-6.83		-3.99	-2.26
	-7.04	-8.65		-16.41	-15.69
	-14.38	-12.51		-9.36	-10.73
	-9.21	-9.04		-3.68	-3.07

chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-7.38	-6.96		-5.53	-4.22
	-6.32	-5.54		-9.54	-8.01
	-12.46	-12.45		-6.41	-4.81
	-5.37	-6.35		-9.32	-8.34
	-2.64	-2.62		-2.98	-1.77
	-5.24	-4.49		-8.03	-8.16
	-8.06	-6.43		-9.91	-10.01
	-5.68	-4.73		-9.93	-10.62
	-5.25	-6.77		-16.26	-16.79

chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.67	-7.13		-5.53	-6.07
	-8.97	-8.97		-2.3	-1.80
	-10.65	-9.50		-1.95	-2.20
	-8.62	-10.03		-8.7	-7.21
	-6.25	-6.01		-2.53	-3.20
	-1.96	-1.97		-5.48	-6.99
	-11.44	-13.35		-1.84	-0.07
	-1.77	-1.97		-8.28	-7.82
	-7.78	-6.08		-8.62	-8.65

chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-8.9	-8.71		-8.27	-7.92
	-6.52	-5.43		-11.36	-11.46
	-8.92	-7.72		-7.45	-6.96
	-7.51	-7.89		-10.45	-10.24
	-11.03	-10.02		-7.25	-7.84
	-13.5	-12.82		-12.08	-11.36
	-9.35	-10.39		-8.89	-10.15
	-9.53	-10.47		-5.22	-5.27
	-9.43	-11.29		-10.3	-8.30

chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-10.65	-10.42		-8.93	-10.95
	-10.44	-8.93		-8.5	-6.95
	-7.15	-5.34		-10.52	-10.22
	-4.65	-4.71		-10.98	-10.49
	-8.68	-5.87		-6.35	-6.00
	-8.22	-7.62		-7.46	-7.05
	-2.47	-4.47		-6.37	-3.67
	-3	-4.02		-11.16	-11.22
	-7.69	-7.15		-10.03	-9.04

chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-6.43	-6.55		-11.85	-9.61
	-5.73	-3.56		-5.27	-5.13
	-5.18	-3.42		-9.64	-11.47
	-9.47	-9.12		-11.39	-11.16
	-6.07	-7.33		-11.29	-9.90
	-3.53	-3.68		-4.56	-3.57
	-5.36	-3.62		-8.99	-9.73
	-13.54	-13.84		-10.09	-11.72
	-10.93	-11.53		-7.71	-8.26
	-10.63	-11.88		-4.28	-4.29

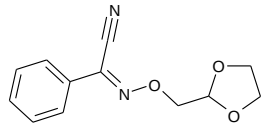
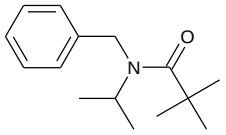
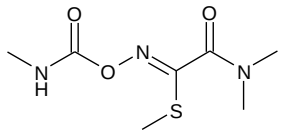
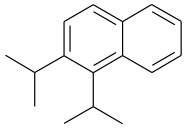
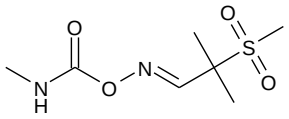
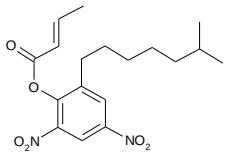
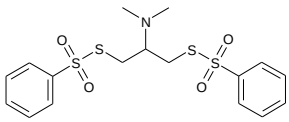
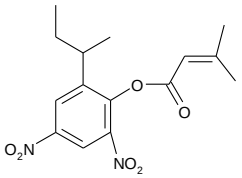
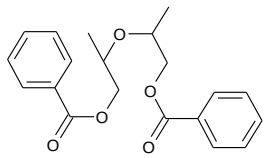
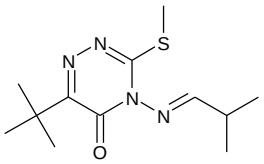
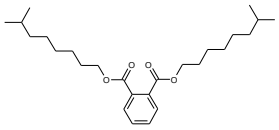
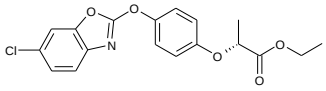
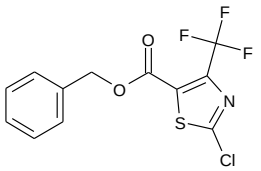
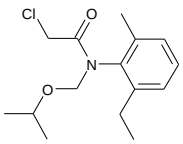
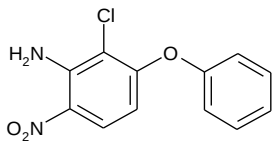
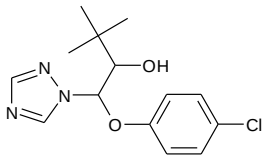
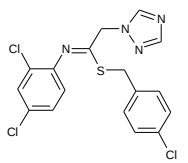
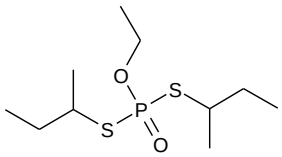
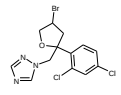
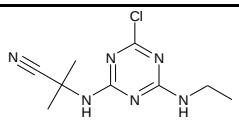
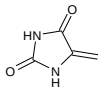
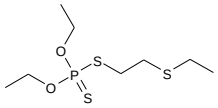
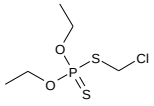
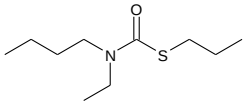
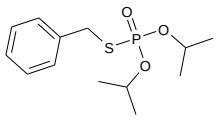
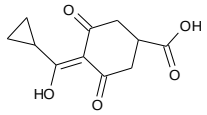
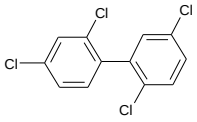
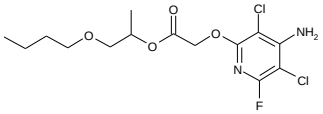
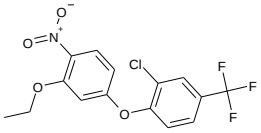
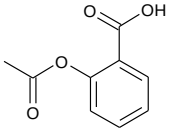
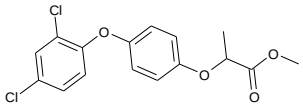
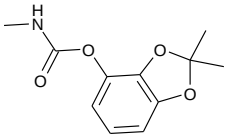
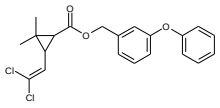
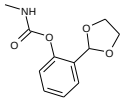
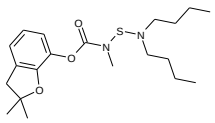
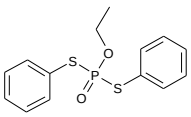
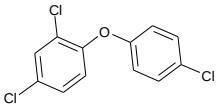
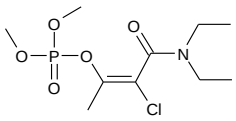
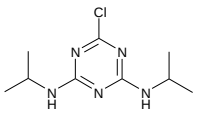
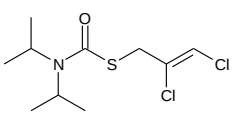
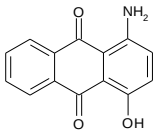
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.51	-9.82		-6.78	-5.49
	-10.93	-13.04		-1.75	-2.54
	-9.5	-9.02		-9.15	-8.87
	-5.82	-6.21		-7.05	-7.68
	-8.52	-8.50		-8.44	-8.65
	-5.75	-6.85		-8.44	-10.03
	-6.81	-7.21		-7.51	-6.08
	-8.41	-9.68		-11.98	-9.60
	-11.02	-10.61		-5.83	-7.63

Table A.4.10. Experimental (ΔG^s_{exp}) and calculated (ΔG^s_{calc}) solvation free energies (in kcal/mol) of the 362 compounds in the training set.

chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}	chemical structure	ΔG^{s}_{exp}	ΔG^{s}_{calc}
	-10.79	-9.27		-10.88	-12.00
	-11.18	-9.32		-5.53	-4.62
	-6.64	-8.67		-2.66	-3.50
	-2.75	-4.26		-7.91	-9.55
	-13.36	-13.18		-3.77	-2.66
	-9.59	-9.39		-6.1	-6.35
	-9.92	-8.93		-5.58	-7.46
	-7.91	-9.51		-9.7	-7.85
	-12.57	-10.60		-6.38	-7.34
	-9.67	-10.44		-2.54	-3.67
	-13.06	-11.07		-9.17	-8.71
	-5.19	-5.03		-12.23	-11.74

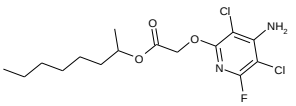
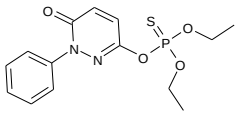
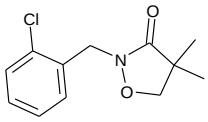
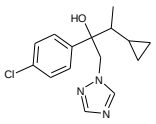
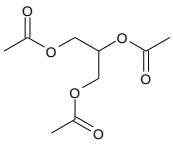
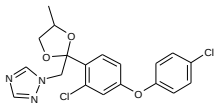
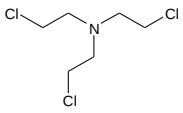
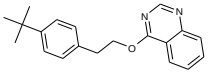
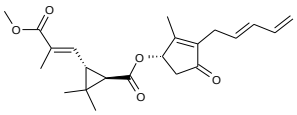
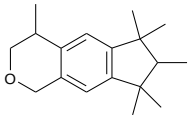
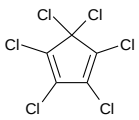
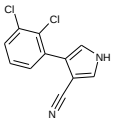
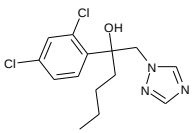
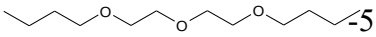
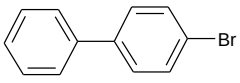
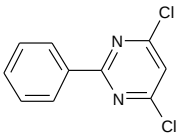
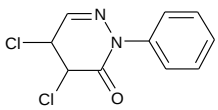
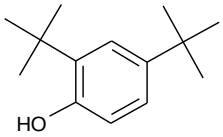
chemical structure	ΔG^s_{exp}	ΔG^s_{calc}	chemical structure	ΔG^s_{exp}	ΔG^s_{calc}
	-7.58	-8.57		-11.66	-10.95
	-7.87	-5.89		-10.11	-10.29
	-8.59	-6.86		-12.87	-10.78
	-4.26	-2.23		-7.8	-7.35
	-8.25	-9.02		-3.22	-4.50
	-0.42	-2.05		-9.08	-8.48
	-9.21	-10.17		-5.44	-3.47
	-1.83	-3.37		-4.51	-4.72
	-6.61	-8.25		-3.85	-4.87

Table A.4.11. Experimental (ΔG^s_{exp}) and calculated (ΔG^s_{calc}) solvation free energies (in kcal/mol) of the 42 compounds in the test set.