Computational Insights into Nitrogen-Related Biocatalysis

Bogdan Florin Ion
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
https://scholar.uwindsor.ca/etd/5479

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.
Computational Insights into Nitrogen-Related Biocatalysis

By

Bogdan F. Ion

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Chemistry and Biochemistry
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
at the University of Windsor

Windsor, Ontario, Canada

© 2015 Bogdan F. Ion
Computational Insights into Nitrogen-Related Biocatalysis

by

Bogdan Ion

APPROVED BY:

______________________________
J. Pearson, External Examiner
University of Prince Edward Island

______________________________
M. Cioppa
Earth & Environmental Sciences

______________________________
S. Ananvoranich
Department of Chemistry & Biochemistry

______________________________
J. R. Green
Department of Chemistry & Biochemistry

______________________________
J. Gauld, Advisor
Department of Chemistry & Biochemistry

September 21, 2015
Declaration of Co-Authorship / Previous Publications

I. Co-Authorship Declaration

I hereby declare that this dissertation incorporates material that is result of joint research, as follows:

Chapter 3 was done in collaboration with Ms. Erum Kazim under the supervision of Prof. James W. Gauld.

Chapter 6 was done in collaboration with Mr. Sameer Jafar and Mr. Mohamed Aboelnga under the supervision of Prof. James W. Gauld.

Chapter 7 was done in collaboration with Mr. Phil De Luna and Dr. Eric Bushnell under the supervision of Prof. James W. Gauld.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation, and have obtained written permission from each of the co-author(s) to include the above materials in my dissertation.

I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publication

This dissertation includes six original papers that have been previously published in peer-reviewed journals or are to be submitted, as follows:
<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>Citation</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 4</td>
<td>Ion, B. F.; Gauld, J. W. <em>In preparation</em>.</td>
<td>To be submitted</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Ion, B. F.; Gauld, J. W. <em>In preparation</em>.</td>
<td>To be submitted</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>Ion, B. F.; Gauld, J. W. <em>In preparation</em>.</td>
<td>To be submitted</td>
</tr>
</tbody>
</table>

I certify that I have obtained a written permission from the copyright owners to include the above-published materials in my dissertation. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owners to include such materials in my dissertation.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.
Abstract

Nitrogen-dependent reactions are prevalent and essential in many biochemical systems. These chemical reactions are ensured to occur at physiological rates via the catalytic power of enzymes. Important to some reactions, their catalysis is also dependent on cofactors such as NAD\(^+\), metal ions, and active site water molecules. In this dissertation, several nitrogen-related biochemical systems are investigated using complementary computational methods such as docking, molecular dynamics simulations, quantum chemical clusters, and quantum mechanics/molecular mechanics.

The use of this multi-scale computational approach has been successfully applied to investigate the catalytic mechanisms, substrate binding, and roles of key active site residues of both metallo- (e.g., *Streptococcus pneumoniae* Nicotinamidase) and non-metalloenzymes (e.g., Ornithine Cyclodeaminase). Additionally, *in silico* mutations were done to examine the impact genetic mutations have on the catalytic site of physiologically important enzymes (e.g., \(\Delta^1\)-pyrroline-5-carboxylate dehydrogenase). The specificity of enzymes involved in protein synthesis (e.g., L-lysyl-tRNA synthetase) has also been studied along with their ability to discriminate with high-fidelity between chemically and structurally similar ligands.

The application of quantum chemical cluster methods to explore multiple X-ray crystal structures of an enzyme (e.g., pseudouridine-5'-monophosphate glycosidase) provided a greater understanding of its reaction mechanism. Moreover, the importance in carefully selecting a starting point from available crystal structures was shown when applying molecular modeling and simulation methods.
Dedication

I dedicate this work to my family, my best friend and loving Fiancée, Samantha Sabelli.
Acknowledgements

During the last five years as a graduate student in the Gauld research group, I have gained many rewarding and unforgettable experiences. Throughout my time, I met many amazing people, made friendships and connections, some of which I have had the opportunity to collaborate on various studies presented in this work. It would not have been an easy journey without their endless help and support, and for that I am extremely grateful.

In particular, I would like to properly recognize and thank my awesome supervisor, Professor James W. Gauld, for his valuable teachings, advice, and patience during this process. He has played an essential role in guiding me towards becoming a fellow scientist. Moreover, I truly value his outstanding support and the life lessons I received over the years; they prepared me for the professional realm as an independent researcher.

I would also like to thank my Ph.D. committee members, Professor Sirinart Ananvoranich, Professor Maria Cioppa, and Professor James Green. They have been very generous and patient with their time and fruitful discussions relating to my research over the course of my graduate studies. Furthermore, I acknowledge them for taking their time to read and assess my dissertation. In addition, I would like to recognize and thank Professor Jason Pearson for accepting the task as an external examiner.

The past and present members of the Gauld 'Unit' have truly been wonderful in creating a friendly and entertaining environment, especially during the many hours researching and writing. They have added their support and lightheartedness during my studies. Dr. Eric Bushnell has been very patient in educating me with useful advice, and familiarizing me with numerous computational chemistry concepts. As well, his involvement in the beginning of my academic journey is valued. I would also like to acknowledge Dr. Wenjuan Huang, Avtinder Basra, Joel Almasi, Grant Fortowsky,
Natalia Mroz, Rami Gherib, Anthony Deschamps, Peter Boateng, Omid Beiraghi, Travis DeWolfe, Philip Stamov, Arielle Desamito, Sarah Henshaw, Wanlei Wei, Jacqueline Gemus, and all who are mentioned below for all the interesting discussions and for making all academic and social events joyful. Moreover, I wish to thank the undergraduate students with whom I had the pleasure of working with, including Phil De Luna, Erum Kazim, Danielle Cooper, and Sameer Jafar. Most of which have been collaborating with me on some of the studies presented herein. Likewise, Dr. Hisham Dokainish, Gyllian Porteous, and Mohamed Aboelnga have partaken in my research as well, and their help is much appreciated.

Outside the research group circle, there are many important people that I owe extreme gratitude. That is, I am thankful for being blessed with an amazing family who has shown their constant and loving support provided during my long tedious years of studies. Furthermore, a heartfelt thank you to my amazing Fiancée, Samantha Sabelli, for always being so supportive and inspiring during these past few years. I would like to extend my acknowledgement to her family, for their encouragement and guidance, as well as the 'food packages' and care provided by Nonna Maria.

A special acknowledgement to my colleagues and friends throughout the years: Matthew and Irene Hurajt, Abhishek Jariwala, Dennis Ma, Akhil Vohra, Eric, Jacqueline, Wanlei, Mohamed, and the great Dan Simard for taking their time to proofread and revise my dissertation.

Consequently, all of the aforementioned people have played a tremendous part with their valuable time, guidance, and assistance during the completion of my studies. I am deeply thankful for having wonderful people in my life who have truly understood the sacrifices and work I have dedicated to complete my academic goals.

Lastly, I recognize and thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding, Ontario Graduate Scholarship (OGS) program for
financial support, SHARCNET, and Compute Calcul Canada for computational resources.
# Table of Contents

*Declaration of Co-Authorship / Previous Publications* ........................................ iii

*Abstract* .................................................................................................................. v

*Dedication* ............................................................................................................... vi

*Acknowledgements* ............................................................................................... vii

*List of Figures* ......................................................................................................... xv

*List of Schemes* ........................................................................................................ xxv

*List of Tables* ........................................................................................................... xxvii

*List of Abbreviations and Symbols* ........................................................................ xxix

**Chapter 1. Introduction** ......................................................................................... 1

  1.1. Introduction ........................................................................................................ 2
  1.2. References .......................................................................................................... 5

**Chapter 2. Computational Enzymology Methods** .................................................. 8

  2.1. Introduction ........................................................................................................ 9
  2.2. Molecular Modeling and Simulations of Enzymes .............................................. 9
  2.3. Choosing a Suitable Starting Structure ............................................................ 10
  2.4. Computational Methods for Modeling Enzyme-Catalyzed Reactions .......... 12

    2.4.1. *Molecular Dynamics (MD) Simulations* ................................................... 12
    2.4.1.1. Molecular Mechanics (MM) ................................................................. 14
    2.4.1.2. MD Protocols ...................................................................................... 16
    2.4.2. *Quantum Mechanics (QM) Cluster Approach* .................................... 16
    2.4.2.1. QM Methods ...................................................................................... 17
    2.4.2.2. Solvation ............................................................................................. 19
    2.4.2.3. Setup of QM Cluster Model ............................................................... 20
    2.4.3. *Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Approach* .... 21
    2.4.3.1. Setup of ONIOM(QM/MM) Model .................................................... 23
Chapter 3.  A Multi-Scale Computational Study on the Mechanism of Streptococcus pneumoniae Nicotinamidase (SpNic) ................................................................. 31
3.1. Introduction .............................................................................................................. 32
3.2. Computational Methods ......................................................................................... 36
  3.2.1. Molecular Dynamics (MD) Simulations ............................................................... 36
  3.2.2. ONIOM(QM/MM) Calculations .......................................................................... 36
3.3. Results and Discussion ........................................................................................... 38
  3.3.1. The Unbound Active Site .................................................................................... 38
  3.3.2. The Substrate-Bound Active Site ........................................................................ 40
  3.3.3. Catalytic Mechanism of SpNic .......................................................................... 42
    3.3.3.1. Stage 1: Formation of a Thioester Enzyme-Intermediate Complex with Loss of Ammonia .............................................................. 42
    3.3.3.2. Stage 2: Hydrolysis of the Thioester Enzyme-Intermediate and Product Formation .................................................................................. 48
  3.3.4. Inclusion of Dispersion Effects on the Catalytic Mechanism of SpNic .......... 51
3.4. Conclusions ............................................................................................................ 54
3.5. References .............................................................................................................. 55

Chapter 4.  A Multi-Scale Computational Study of the Catalytic Mechanism of Maleamate Amidohydrolase (NicF): A Non-Metallo Amidase and the Role of its Active site Threonine ................................................................. 59
4.1. Introduction ............................................................................................................ 60
4.2. Computational Methods ......................................................................................... 63
  4.2.1. Molecular Dynamics (MD) Simulations ............................................................... 63
  4.2.2. ONIOM(QM/MM) Calculations .......................................................................... 63
4.3. Results and Discussion ........................................................................................... 64
  4.3.1. The Maleamate-Bound Active Site .................................................................... 64
  4.3.2. The Oxyanion Hole ........................................................................................... 69
  4.3.3. Catalytic Mechanism of NicF ........................................................................... 71
Chapter 5.  Insights into the Substrate Active Site Binding and Mutagenesis of $\Delta^1$-Pyrroline-5-Carboxylate (P5C) Dehydrogenase using a Molecular Dynamics Approach ................................................. 85

5.1. Introduction ........................................................................................................ 86

5.2. Computational Methods ..................................................................................... 90

5.2.1. Chemical Models .............................................................................................. 90

5.2.2. Solvation and Energy Minimization .................................................................... 91

5.2.3. Thermal Relaxation ............................................................................................ 92

5.2.4. Subsequent Models and MD Simulations .......................................................... 92

5.3. Results and Discussion ......................................................................................... 94

5.3.1. Equilibrated Michaelis Complexes ................................................................... 94

5.3.2. Average Substrate-Bound Complexes ............................................................... 95

5.3.3. Activated- and Hydroxylated-Michaelis Complexes ........................................ 103

5.3.4. S352L Mutation ............................................................................................... 107

5.3.5. S352A Mutation ............................................................................................... 110

5.3.6. E314A Mutation ............................................................................................... 112

5.4. Conclusions .......................................................................................................... 114

5.5. References ............................................................................................................ 116

Chapter 6.  Computational Studies on the Pre-Transfer Editing of Lysyl-tRNASynthetase: Noncognate L-Ornityl Preference over Native L-Lysyl ....................... 120

6.1. Introduction .......................................................................................................... 121

6.2. Computational Methods ...................................................................................... 124

6.2.1. MD Models ...................................................................................................... 124

6.2.2. Energy Minimizations ..................................................................................... 124

6.2.3. MD Simulations ................................................................................................. 125

6.2.4. ONIOM(QM/MM) Calculations ....................................................................... 125
6.3. Results and Discussion ................................................................. 127
   6.3.1. Equilibrated Michaelis Complexes ............................................ 127
   6.3.2. Aminoacyl Binding ............................................................... 128
   6.3.3. Pre-transfer Editing of Orn-AMP versus Lys-AMP ....................... 130
6.4. Conclusions ................................................................................. 138
6.5. References .................................................................................. 139

Chapter 7. A Molecular Dynamics (MD) and Quantum Mechanics/Molecular
Mechanics (QM/MM) Study on Ornithine Cyclodeaminase (OCD): A Tale of Two
Iminiums .......................................................................................... 143
7.1. Introduction .................................................................................. 144
7.2. Computational Methods ............................................................... 147
   7.2.1. Molecular Dynamics (MD) Equilibration .................................... 147
   7.2.2. QM/MM Computations ........................................................... 149
7.3. Results and Discussion ................................................................. 151
   7.3.1. Structure of the Active site with the Substrate L-Ornithine and NAD\(^+\) Cofactor Bound ................................................................. 151
   7.3.2. Orn Oxidation and Formation of 2-Aminoproline (AP) .................. 152
   7.3.3. Deamination of 2-Aminoproline (AP) ........................................... 156
   7.3.4. Hydride Transfer from NADH onto P2C to Give the Final Product L-Proline ........................................................................ 160
7.4. Conclusions .................................................................................. 162
7.5. References .................................................................................. 163

Chapter 8. The Effects of Structural Variation on the Enzymatic Mechanism of
Pseudouridine-5'-Monophosphate (ΨMP) Glycosidase using Molecular Dynamics and
Quantum Chemical Cluster Approaches ................................................... 169
8.1. Introduction .................................................................................. 170
8.2. Computational Methods ............................................................... 173
   8.2.1. MD Simulations ....................................................................... 173
   8.2.2. Quantum Mechanical (QM) Cluster Calculations ......................... 174
8.3. Results and Discussion ................................................................. 176
   8.3.1. Equilibrated Starting Structures .................................................. 176
   8.3.2. Substrate Binding in the Three Reactant Complexes ..................... 179

xiii
8.3.3. Nucleobase Deprotonation ................................................................. 182
8.3.4. Covalent Linkage between Substrate and Lys166–NH$_2$ (C1'–$^{\text{K166}}$N$^+$H$_2^+$) .... 184
8.3.5. Concurrent Lys166 Deprotonation and Lys93 Protonation ...................... 185
8.3.5. $sp^3$ Hybridization of the C5 Pseudouracil Moiety ............................... 185
8.3.6. Schiff Base Formation ........................................................................ 187
8.3.7. Imine Hydrolysis .............................................................................. 187
8.3.8. R5P Cyclization ................................................................................ 188
8.3.9. Continuum Solvation Effects on Relative Energies ............................. 191
8.3.10. Variation in the Fixed Atom Positions ............................................ 191
8.3.11. Choosing the Right Model ............................................................... 193
8.4. Conclusions ......................................................................................... 194
8.5. References .......................................................................................... 196

Chapter 9. Conclusions .............................................................................. 200
9.1. Conclusions ......................................................................................... 201
9.2. References .......................................................................................... 204

Chapter 10. Future Work ............................................................................ 205

Appendices ................................................................................................. 207
Appendix A. Supporting Information ......................................................... 208
Appendix B. Copyright Permissions ........................................................... 211

Vita Auctoris ............................................................................................... 213
List of Figures

Figure 1.1. An example of Watson-Crick base pairing showing the hydrogen-bonding pattern between adenine (A) and thymine (T). ................................................................. 2

Figure 1.2. Nitrogen presence in selected biomolecules including (a) guanidine group of L-arginine (Arg), (b) choline moiety in platelet-activating factor (PAF), and (c) acetyl amine group of N-acetyl glucosamine (NAG). .................................................................................. 4

Figure 2.1. Graphic representation of the yearly rise in the number of protein structures (orange) in the Protein Data Bank and number of articles published per year (blue) which included "molecular model" and "enzyme" keywords in the Web of Science search engine. ................................................................................................................................. 10

Figure 2.2. Schematic illustration of the timescales associated with the conformational events in proteins/enzymes ranging from femtoseconds to seconds. ................................. 13

Figure 2.3. A peptide molecular model showing the bond stretching (str), valence angle bending (bend), torsional angle (tors), and non-covalent interactions such as van der Waals (vdW) or electrostatics. Note each R represents the side chain of an amino acid. ........................................................................................................................................... 15

Figure 2.4. Polarizable continuum model for methylamine, where the outer shell is represented by the dielectric constant, ε..................................................................................... 20

Figure 2.5. A 2D potential energy surface of a two-step reaction mechanism, where RC = reactant complex, TSx (x = 1, 2) = transition state, and PC = product complex. Note the rate-determining step is IC1 → PC since E’ is greater than E. ............................................. 24
Figure 3.1. Schematic representation of the substrate-bound active site model used in the QM/MM calculations. The outer circle represents the MM-layer, while those in the inner circle represent the QM-layer. Residues in red were mutated to glycyl (see text). ..........38

Figure 3.3. Optimized structure with (a) selected distances shown (Angstroms), (b) ESP charges, and (c) π-interactions in the substrate-bound active site, reactant complex (RC), obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow. .................................................................41

Figure 3.4. Potential energy surface (kJ mol⁻¹) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory for formation of the thioester enzyme-intermediate with concomitant loss of ammonia. .................................................................................................................................................................................................43

Figure 3.5. Optimized geometries obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory of the transition states (TS1 and TS2) and intermediates (IC1 and IC2) for the first stage of the overall mechanism of SpNic. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow. .................................................................45

Figure 3.6. ESP charges for select key species in the SpNic mechanism (H-atom inclusive). Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow. .....................46

Figure 3.7. Potential energy surface (kJ mol⁻¹) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory for the hydrolysis of the thioester intermediate and product complex formation. .............48
Figure 3.8. Optimized geometries obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory of the thioester intermediate + H₂O (IC₂') and other transition states (TSₓ, x = 3, 4, 5) and product complexes (PCₓ, x = a, b, c) for the second stage of the overall mechanism of SpNic. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 3.9. Potential energy surfaces for the overall catalytic mechanism of SpNic obtained at the ONIOM(DFTᵢ/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within the (a) ME and (b) EE formalism. DFTᵢ = B3LYP (black), B3LYP-D3 (red), or M06 (blue).

Figure 4.1. Optimized structure of the reactant complex (RC) with (A) selected distances shown (Angstroms), (B) π-interactions, obtained at ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Electrostatic potential surfaces mapped onto the electron density with key ESP charges of (C) maleamate and (D) maleamic acid substrates, as well as the oxyanion hole (E) without Thr146–OH (F) with Thr146–OH interaction, constructed using single points at B3LYP/6-311+G(2df,p). Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 4.2. Plots of (A) oxyanion hole RMSD and (B) O\textsubscript{carb}⋯HO–Thr146, (C) O\textsubscript{carb}⋯HN–Thr146, and (D) O\textsubscript{carb}⋯HN–Cys150 distances (Angstroms), obtained over a 10.1 ns MD simulation.

Figure 4.3. Plot of (A) the active site RMSD after 10.1 ns and a pictorial representation of (B) an overlay between the average structure (bold black) and the O\textsubscript{carb}⋯HO–Thr146 conformer with the shortest distance (orange). RMSD and distances are shown in Angstroms. Atom colour code: H, white; C, black/orange; N, blue; O, red; S, yellow.
Figure 4.4. PESs (kJ mol\(^{-1}\)) for the first stage of the overall mechanism of NicF, obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within ME (red) and EE (black) formalism. Note TS2* represents a constrained optimization...

Figure 4.5. Optimized geometries of the transition states (TS1 and TS2*) and intermediates (IC1 and IC2) for Stage 1, obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow...

Figure 4.6. Electrostatic potential surfaces mapped onto the electron density of select key species in the NicF mechanism for (A) neutral and (B) ionic carboxyl groups, obtained at B3LYP/6-311+G(2df,p) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow...

Figure 4.7. PESs (kJ mol\(^{-1}\)) for the second stage of the overall mechanism of NicF, obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within ME (red) and EE (black) formalism...

Figure 4.8. Optimized geometries of the transition states (TS3 and TS4) and product complexes (PC\(_a\) and PC\(_b\)) for Stage 2, obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow...

Figure 5.1. (A) Unbound active site of P5CDH [PDB ID: 3V9J] and its bound sites with (B) GLU product [PDB ID: 3V9K] and (C) NAD\(^+\) coenzyme [PDB ID: 3V9L]. Note Glu314* indicates that its R-group is missing...
Figure 5.2. Surrounding active site RMSDs of the P5CDH (A) reactant complex (RC; GSA- and NADH-bound) and (B) product complex (PC; GLU-bound) during the first 10.1 ns simulation.

Figure 5.3. RMSDs of GSA-bound RC based on the positions of active site residues such as the thiol of Cys348, carboxylic acid of Glu314, amide of Asn211, peptide backbone between Cys348 and Lys347, GSA substrate, in the (A) presence of GSA, absence of (B) GSA and (C) both GSA and amide of Asn211 positions. (D) RMSDs in the position of the GSA substrate alone. Note all of these calculations were based on the last 5 ns of the simulation.

Figure 5.4. Overlays of representative average structures for the GSA-bound RC, based on RMSD positions of active site residues in the (A) presence of GSA, absence of (B) GSA and (C) both GSA and Asn211 amide, (D) three plateau representatives. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.3. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 5.5. RMSDs of GLU-bound PC based on the positions of active site residues in the (A) presence of GLU, absence of (B) Cys348 thiol and (C) both thiol and carboxyl groups of Cys348 and Glu314, respectively. (D) RMSDs in the position of the GLU product alone. Note all of these calculations were based on the last 5 ns of the simulation.

Figure 5.6. Overlays of representative average structures of GLU-bound PC, based on RMSD positions of active site residues in the (A) presence of GLU, absence of (B) Cys348 thiol and (C) both Cys348 thiol and Glu314 carboxyl. Note in (D) the overall average structure is overlaid with the structures of the RMSD spike. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.5. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Figure 5.7. RMSDs of fully-bound RC based on the positions of active site residues and their respective ligands in the (A) active (Cys348–S'/Glu314–COOH) and (B)-(C) hydroxylated (4-OH-GSA) forms. In (C) the RMSD positions of the hydroxylated RC ligands are absent. Note all of these calculations were based on the last 5 ns of each simulation.

Figure 5.8. Overlays of representative average structures of the RC, based on RMSD positions of active site residues in the (A)-(B) active and (B)-(D) hydroxylated forms. Note in (D) the RMSD positions of the hydroxylated RC ligands are absent. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.7. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 5.9. Overlays of average structures for the fully-bound RC with (A) active and (B) hydroxylated RC active sites. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.7 and Appendix Figure A5.2A. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 5.10. Overlays of RC and PC average structures with the S352L mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.8 and A5.10. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure 5.11. Overlays of RC and PC average structures with the S352A mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.8 and A5.11. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Figure 5.12. Overlays of RC and PC average structures with the E314A mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.9 and A5.12. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow. 113

Figure 6.1. Surrounding active site RMSDs of the average LysU Michaelis complexes including (A) Orn-, (B) Lys-, (C) Cys-, and (D) Ser-AMP as substrates. 127

Figure 6.2. Overlaid average structures of aa- = Orn-, Lys-, Cys-, and Ser-AMP within the LysU active site. Atom colour code: H, white; C, gray/maroon/citrine/black; N, blue; O, red; S, yellow; P, pink. 128

Figure 6.3. Overlaid average structures of aa- = Orn- and Lys-AMP within the LysU active site. Atom colour code: H, white; C, gray/maroon; N, blue; O, red; P, pink. 130

Figure 6.4. Optimized geometries of the reactant complexes (aaRC) of (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Note aa-AMP* indicates that its ribose and nucleobase are not shown for simplicity. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange. 131

Figure 6.5. Optimized geometries of the aaTS1* transition states and aaIC1 intermediates for (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Note aa-AMP* indicates that its ribose and nucleobase are not shown for simplicity. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange. 132

Figure 6.6. Potential energy surface (kJ mol$^{-1}$) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d,p):AMBER96) for the editing mechanism of Orn-AMP by LysU. 133
Figure 6.7. Potential energy surface (kJ mol\(^{-1}\)) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d,p):AMBER96) for the editing mechanism of Lys-AMP by LysU.

Figure 6.8. Optimized geometries of the \(^{\text{aa}}\text{TS2}^*\) transition states and \(^{\text{aa}}\text{PC}\) product complexes for (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

Figure 6.9. Optimized geometries of the lactam products along with the AMP byproducts from \(^{\text{Orn}}\text{PC}\) and \(^{\text{Lys}}\text{PC}\) of the LysU editing mechanism. Atom colour code: H, white; C, gray/maroon; N, blue; O, red; P, pink.

Figure 7.1. Illustration of the ONIOM (QM/MM) fully bound active site model used: (a) the low or molecular mechanics (MM)-layer is shown in wire while the high or quantum mechanics (QM)-layer is shown in tube format; (b) those moieties within the inner circle were included within the QM-region, while those within the outer circle represent the residues contained in the MM-layer of the QM/MM model. Atom colour code: H, white; C, gray; N, blue; O, red.

Figure 7.2. (a) Optimized structure (see Computational Methods) of the initial fully bound active site complex \(\text{RC}\) with selected bond lengths shown (in Angstroms). (b) Solvated MD structure of the active site. Atom colour code: H, white; C, gray; N, blue; O, red.

Figure 7.3. Optimized structures (see Computational Methods) with selected bond lengths shown (in Angstroms) for the ornithine cyclodeaminase (OCD) catalytic mechanism prior to deamination as L-ornithine is converted to 2-aminoproline. Atom colour code: H, white; C, gray; N, blue; O, red.
Figure 7.4. Potential energy surface (PES) obtained (see Computational Methods) of the OCD catalytic mechanism prior to deamination as L-ornithine is converted to 2-aminoproline.  

Figure 7.5. Optimized structures (see Computational Methods) with selected bond lengths (in Angstroms) of the water-assisted intermediate complex (IC2′⋯H$_2$O), TS4, cyclic intermediate IC3, and Schiff base intermediate IC4. Atom colour code: H, white; C, gray; N, blue; O, red.  

Figure 7.6. PESs obtained (see Computational Methods) for (a) water-assisted deamination of 2-aminoproline to give Δ$^1$-pyrroline-2-carboxylate (P2C) and, (b) reduction of P2C to give L-proline.  

Figure 7.7. Optimized structures (see Computational Methods) with selected bond lengths shown (in Angstroms) of the Δ$^1$-pyrroline-2-carboxylate intermediate IC4'', TS7, and L-proline product (PC). Atom colour code: H, white; C, gray; N, blue; O, red.  

Figure 8.1. (A) Surrounding active site RMSDs of the ΨMP dock, ΨMP⋯A166K, ΨMP- and R5P-adducts during the course of the 10.1 ns MD simulation. Plots of distances with respect to time (ns) for the last 10 ns of the simulation of the four complexes are as follows: (B) Mn(II)⋯P, (C) K$^{166}_{\text{N}^\varepsilon}⋯\text{C1'}$, and (D) K$^{166}_{\text{N}^\varepsilon}–\text{C1'}$.  

Figure 8.2. Overlaid average structures of ΨMP⋯A166K with ΨMP- and R5P-adducts in the glycosidase active site. Atom colour code: H, white; C/Mn(II), green/yellow/gray; N, blue; O, red; P, pink.  

Figure 8.3. Optimized structures of the glycosidase reactant complexes (XRC, where X = DOCK, ΨMP, R5P), obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.
Figure 8.4. Optimized structures of the glycosidase intermediate complexes: (A) $\text{DOCK}^{\text{IC}x}$, (B) $\Psi^{\text{MP}}^{\text{IC}x}$, and (C) $\text{R5P}^{\text{IC}x}$, where $x = 1, 2, 3$, obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

Figure 8.5. Optimized structures of the glycosidase intermediate complexes: (A) $\text{DOCK}^{\text{IC}y}$, (B) $\Psi^{\text{MP}}^{\text{IC}y}$, and (C) $\text{R5P}^{\text{IC}y}$, where $y = 4, 5, 6$, obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

Figure 8.6. Optimized structures of the glycosidase product complexes ($X^{\text{PC}}$, where $X = \text{DOCK}, \Psi^{\text{MP}}, \text{R5P}$), obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.
List of Schemes

Scheme 1.1. Hydrolytic deamination of cytidine-5'-monophosphate (CMP) to form a uridine derivative (UMP) using CMP deaminase. ................................................................. 3

Scheme 1.2. Oxidative deamination of L-glutamate (Glu) to α-ketoglutarate (α -KG) by glutamate dehydrogenase (GDH)........................................................................................................ 3

Scheme 3.1. Overall reaction for conversion of nicotinamide to nicotinic acid as catalyzed by nicotinamidase (Nic). ........................................................................................................... 33

Scheme 3.2. Proposed mechanism(s) for conversion of nicotinamide to nicotinic acid as catalyzed by the nicotinamidase SpNic.23 .................................................................................................................. 35

Scheme 4.1. The overall deamination reaction of maleamate (1) to maleate (2) catalyzed by the amidohydrolase (NicF).9 ................................................................................................................. 61

Scheme 4.2. Proposed mechanism for conversion of maleamate to monoionic maleate catalyzed by the amidohydrolase NicF.9 ................................................................................................................. 62

Scheme 4.3. Optimized structures of two possible substrate ionization states in a dielectric constant (ε) of 78.39, obtained at IEFPCM-B3LYP/6-31G(d) level of theory. 68

Scheme 5.1. Overall reaction for the conversion of L-glutamate-γ-semialdehyde (GSA) to glutamate by Δ1-pyrroline-5-carboxylate dehydrogenase (P5CDH).23 ........................................... 87

Scheme 5.2. The three-stage mechanistic proposal for the conversion of glutamate-γ-semialdehyde to glutamate catalyzed by Δ1-pyrroline-5-carboxylate dehydrogenase.18 Note sequence numbering of P5CDH is based on mouse-model derived structures.23 .... 89
Scheme 6.1. Two-step reaction for the aminoacylation of tRNA^{Lys} catalyzed by the aminoacyl-tRNA synthetase, LysRS: (A) L-lysine activation via a displacement reaction followed by (B) L-lysyl transfer to tRNA^{Lys} biomolecule. ............................................... 122

Scheme 6.2. Proposed mechanism for the cyclization of mischarged Orn-AMP catalyzed by LysU. \(^{22}\) ........................................................................................................................................ 123

Scheme 6.3. QM-layer representation of the L-lysyl-bound active site model used in the ONIOM(QM/MM) calculations. ........................................................................................................ 126

Scheme 7.1. Overall reaction for conversion of L-ornithine to L-proline as catalyzed by ornithine cyclodeaminase (OCD). ........................................................................................................ 145

Scheme 7.2. Proposed (A) hydrolytic and (B) non-hydrolytic pathways for the conversion of L-ornithine to L-proline as catalyzed by the enzyme ornithine cyclodeaminase. ........................................................................................................ 146

Scheme 8.1. The conversion of ΨMP to uracil and ribose-5'-phosphate by ΨMP glycosidase. ........................................................................................................................................ 170

Scheme 8.2. Proposed mechanism catalyzed by ΨMP glycosidase for conversion of ΨMP to uracil and ribose-5'-phosphate. \(^{17}\) ........................................................................................................................................ 172

Scheme 8.3. Quantum mechanical cluster model of the ΨMP glycosidase active site with bound-ΨMP (fixed atoms are shown by asterisks). ........................................................................................................ 176

Scheme 8.4. Schematic summary of reactant, intermediates, and product complexes for ΨMP glycosidase using a QM cluster approach. Note that red arrows represent two other intermediate possibilities (not obtained). ........................................................................................................ 194
List of Tables

Table 5.1. Wild type reactant complexes considered in this study..........................93

Table 5.2. Mutated reactant complexes considered in this study............................93

Table 5.3. Product complexes considered in this study........................................94

Table 5.4. Average $r(C\text{carb} \cdots \cdot (H)S_{C348})$, $r(O\text{carb} \cdots \cdot HS_{C348})$, $r(O\text{carb} \cdots \cdot HN_{C348})$, $r(O\text{carb} \cdots \cdot HN_{N211})$, and $r(\partial C_{E314} \cdots (H)S_{C348})$ distances (Å) for model 5 (related to Figure 5.4).................................................................98

Table 5.5. Average $r(C\text{carb} \cdots \cdot (H)S_{C348})$, $r(O\text{carb} \cdots \cdot HS_{C348})$, $r(O\text{carb} \cdots \cdot HN_{C348})$, $r(O\text{carb} \cdots \cdot HN_{N211})$, and $r(\partial C_{E314} \cdots (H)S_{C348})$ distances (Å) for model 13 (related to Figure 5.6)........................................................................................................102

Table 5.6. Average $r(C\text{carb} \cdots \cdot (H)S_{C348})$, $r(O\text{carb} \cdots \cdot HS_{C348})$, $r(O\text{carb} \cdots \cdot HN_{C348})$, $r(O\text{carb} \cdots \cdot HN_{N211})$, $r(\partial C_{E314} \cdots (H)S_{C348})$, $r(C_{pyr} \cdots (H)S_{C348})$, and $r(C_{pyr} \cdots H_{carb})$ distances (Å) for models 1, 2, and 3 (related to Figures 5.8 and 5.9).................................................................107

Table 5.7. Average $r(C\text{carb} \cdots \cdot (H)S_{C348})$, $r(O\text{carb} \cdots \cdot HS_{C348})$, $r(O\text{carb} \cdots \cdot HN_{C348})$, $r(O\text{carb} \cdots \cdot HN_{N211})$, $r(\partial C_{E314} \cdots (H)S_{C348})$, $r(C_{pyr} \cdots (H)S_{C348})$, $r(C_{pyr} \cdots H_{carb})$, $r(\zeta C_{P449} \cdots \alpha C_{C348})$, and $r(\partial C_{E447} \cdots \alpha C_{C348})$ distances (Å) for models 7, 10, 15, and 18 (related to Figure 5.10)........................................................................................................110

Table 5.8. Average $r(C\text{carb} \cdots \cdot (H)S_{C348})$, $r(O\text{carb} \cdots \cdot HS_{C348})$, $r(O\text{carb} \cdots \cdot HN_{C348})$, $r(O\text{carb} \cdots \cdot HN_{N211})$, $r(\partial C_{E314} \cdots (H)S_{C348})$, $r(C_{pyr} \cdots (H)S_{C348})$, $r(C_{pyr} \cdots H_{carb})$, $r(\zeta C_{P449} \cdots \alpha C_{C348})$, and $r(\partial C_{E447} \cdots \alpha C_{C348})$ distances (Å) for models 8, 11, 16, and 19 (related to Figure 5.11)........................................................................................................112
Table 5.9. Average \( r(C_{\text{carb}}\cdots\cdot(H)S_{C348}) \), \( r(O_{\text{carb}}\cdots\cdot HS_{C348}) \), \( r(O_{\text{carb}}\cdots\cdot HN_{C348}) \), \( r(O_{\text{carb}}\cdots\cdot HN_{N211}) \), \( r(\delta C_{E314}\cdots\cdot(H)S_{C348}) \), \( r(C_{\text{pyr}}\cdots\cdot(H)S_{C348}) \), \( r(C_{\text{pyr}}\cdots\cdot H_{\text{carb}}) \), \( r(\zeta C_{P449}\cdots\cdot\alpha C_{C348}) \), and \( r(\partial C_{E447}\cdots\cdot\alpha C_{C348}) \) distances (Å) for models 1, 6, 9, 12, 13, 14, 17, and 20 (related to Figure 5.12). ^d

Table 6.1. Selected average distances (Angstroms) considered in the MD study when each noncognate Orn-, Ser-, Cys-, and cognate Lys-AMP are bound in the active site of LysU...

Table 6.2. Selected distances (Angstroms) considered in the QM/MM study when Orn- and Lys-AMP are curled within the active site of LysU...

Table 8.1. Summary of enzyme-substrate models considered in this study...

Table 8.2. Relative energies of QM-optimized complexes, obtained at B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory in gas phase (\( \epsilon = 1.0 \)).

Table 8.3. Relative energies of QM-optimized complexes, obtained at B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory with \( \Delta E_{\text{solv}}(\epsilon = 4.0) \) inclusion.

Table 8.4. Summary of key distances (Angstroms) between fixed atoms for the three models.
List of Abbreviations and Symbols

Abbreviations

aa-AMP  Aminoacyl-adenylate
aaRS    Aminoacyl-tRNA synthetase
aa-tRNA\textsuperscript{aa}  Aminoacyl-tRNA\textsuperscript{aa}
AMBER  Assisted Model Building with Energy Refinement (force field)
AP      2-Aminoproline
Arg     L-Arginine or arginy1
ATP     Adenosine triphosphate
B3      Becke's 3-parameter exchange functional
CHARMM  Chemistry at HArvard Macromolecular Mechanics (force field)
DFT     Density functional theory
EE      Electronic embedding
ESP     Electrostatic potential
IC      Intermediate complex
IEFPCM  Integral equation formalism-polarizable continuum model
LYP     Lee-Yang-Parr correlation functional
Lys     L-Lysine or lysyl
LysRS  L-Lysyl-tRNA synthetase
MD    Molecular dynamics
ME    Mechanical embedding
MM    Molecular mechanics
MOE   Molecular Operating Environment (program)
NAD⁺  Nicotinamide adenine dinucleotide
NieF  Maleamate amidase
OCD   Ornithine cyclodeaminase
ONIOM Our own n-layered integrated molecular orbital and molecular mechanics
Orn   L-Ornithine or ornityl
P2C   $\Delta¹$-Pyrroline-2-carboxylate
P5C   $\Delta¹$-Pyrroline-5-carboxylate
P5CDH $\Delta¹$-Pyrroline-5-carboxylate dehydrogenase
PA    Proton affinity
Parm10 Amber parameters for proteins and nucleic acids
Parm@Frosst Parameters for small molecules developed by Merck-Frosst
PC    Product complex
PES   Potential energy surface
PFROSST Force field consisting of parm10 and parm@Frosst parameters
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>Pro</td>
<td>L-Proline or prolyl</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum mechanics</td>
</tr>
<tr>
<td>QM/MM</td>
<td>Quantum mechanics/molecular mechanics</td>
</tr>
<tr>
<td>RC</td>
<td>Reactant complex</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>SP</td>
<td>Single-point calculation</td>
</tr>
<tr>
<td>SpNic</td>
<td><em>Streptococcus pneumoniae</em> nicotinamidase</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer ribonucleic acid</td>
</tr>
<tr>
<td>Trp</td>
<td>L-Tryptophan</td>
</tr>
<tr>
<td>TS</td>
<td>Transition state</td>
</tr>
<tr>
<td>vDW</td>
<td>van der Waals</td>
</tr>
</tbody>
</table>

**Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>ε</td>
<td>Dielectric constant</td>
</tr>
<tr>
<td>ψ</td>
<td>Molecular wavefunction</td>
</tr>
<tr>
<td>Ψ</td>
<td>Pseudouridine</td>
</tr>
<tr>
<td>ΨMP</td>
<td>Pseudouridine-5'-monophosphate</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction
1.1. Introduction

Nitrogen is an essential element for life; it is an integral component of all classes of biomolecules including nucleic acids, proteins, lipids, and carbohydrates.\textsuperscript{1-3} Consequently, the reactions of nitrogen lie at the heart of the chemistry of living organisms. For instance, nitrogen is important in nitrogenous building blocks that dictate the lifecycle of the cell.\textsuperscript{4} That is, its presence as an amine functional group contributes to reactions that deal with DNA damage\textsuperscript{5} while playing a base-excision DNA repair role as well to restore the original sequence.\textsuperscript{6}

In these chemical pathways, an amine has the ability to not only behave as a hydrogen donor but it can also provide an electron pair to act as a hydrogen acceptor. These capabilities make this moiety critical in base pair formation such as A⋯T, which are dominated by hydrogen bonding interactions as shown in Figure 1.1.\textsuperscript{7}

![Figure 1.1.](image)

Figure 1.1. An example of Watson-Crick base pairing showing the hydrogen-bonding pattern between adenine (A) and thymine (T).

Aside from non-covalent interactions, amines can participate in nucleotide reactions such as in Scheme 1.1.\textsuperscript{8,9} In this particular example, a cytosine nucleobase can undergo hydrolytic deamination to form uracil with the release of an ammonia moiety (NH\textsubscript{3}) using CMP deaminase. This type of deamination has the potential to modify gene expression.\textsuperscript{9} In turn, the physiological role of encoded proteins is altered. If such base substitutions modify the amino acid sequence for mRNA coding regions, it could result in ineffective proteins.\textsuperscript{10,11} This commonly occurs in the coding regions for ion channel protein
subunits, resulting in altered membrane Ca\(^{2+}\) permeability, which has been shown to play a role in the central nervous system.\(^{11}\)

![Scheme 1](image1.png)

**Scheme 1.1.** Hydrolytic deamination of cytidine-5'-monophosphate (CMP) to form a uridine derivative (UMP) using CMP deaminase.

The biochemical significance of nitrogen goes beyond the nucleic acid class, where it continues to satisfy functions in amino acid reactions. As per the oxidative deamination in Scheme 1.2, a naturally encoded amino acid, \(\text{L}-\text{glutamate}\), is converted to \(\alpha\)-ketoglutarate (\(\alpha\)-KG) using glutamate dehydrogenase (GDH) enzyme.\(^{12,13}\) Interestingly, this intermediate is involved in nucleotide repair, specifically in oxidative alkylation.\(^{14}\) Furthermore, \(\alpha\)-KG plays an important role in the citric acid cycle as a biosynthetic intermediate.\(^{15}\) This critical compound allows for the production of adenosine triphosphate (ATP), the 'energy currency' of the cell.\(^{16}\)

![Scheme 2](image2.png)

**Scheme 1.2.** Oxidative deamination of \(\text{L}-\text{glutamate}\) (Glu) to \(\alpha\)-ketoglutarate (\(\alpha\)-KG) by glutamate dehydrogenase (GDH).
Chapter 1 — Introduction

The prevalence of nitrogen is also seen in the side chains of amino acids, lipids, and carbohydrates. For instance, an important R-group such as that of L-arginine is known to be involved in the urea cycle (Figure 1.2a). Its guanidine group is critical in the formation of urea, a safe carrier of toxic ammonia within our bodies. On the other hand, lipids, which are ether-based, are able to perform signaling functions apart from their structural role in the cellular membrane. One such signaling phospholipid is the platelet-activator factor depicted in Figure 1.2b, containing a tertiary amine. Some biological responses for this supramolecule encompass regulations of bronchoconstriction and aggregation/degranulation of platelets. When looking at carbohydrates such as chitin polymers, they are monomerically composed of N-acetylglucosamine subunits. The formation of this glucose derivative involves the dehydration between glucosamine and acetyl-CoA. One distinction between chitin and cellulose monomers is that chitin is found in the exoskeleton of arthropods composed of methyl amide subunits as opposed to the cell wall of higher plants whose monomers are based on hydroxyl functional groups.

![Figure 1.2](image-url)

**Figure 1.2.** Nitrogen presence in selected biomolecules including (a) guanidine group of L-arginine (Arg), (b) choline moiety in platelet-activating factor (PAF), and (c) acetyl amine group of N-acetyl glucosamine (NAG).
Chapter 1 — Introduction

The reactions that these biomolecules undergo can only be driven forward at necessary physiological rates in the presence of enzymes.\textsuperscript{27} For instance, the uncatalyzed polysaccharide hydrolysis (\textit{e.g.}, starch) under standard conditions in neutral solution is estimated to have a half-life of 5 to 8 million years. In the presence of an enzyme (\textit{e.g.}, amylase), the rate of this reaction is enhanced by a factor of more than $10^{17}$ or a catalytic rate of over a thousand times per second.\textsuperscript{28} It is also important to recognize that many enzymes rely on other molecules such as cofactors.\textsuperscript{29} These non-protein components are present in the active site and some common examples include nicotinamide adenine dinucleotide (NAD\textsuperscript{+}) coenzymes, transition metals (\textit{e.g.}, Zn(II), Mn(II)), and ubiquitous water moieties.\textsuperscript{30-32} Note that metal ions can be essential structurally in substrate orientation, catalytically by acting as a Lewis acid, or both.\textsuperscript{31} As a result, cellular homeostasis and life-sustaining conditions can be achieved.

Computational chemistry is an invaluable tool for investigating nitrogen-dependent enzymes and their associated biochemical processes. The application of computational modeling techniques is well accepted to complement experimental methods including site-directed mutagenesis, kinetic studies, and $pK_a$ determination amongst many others.\textsuperscript{33-36} In this dissertation several computational modeling tools are applied to understand and gain insights into the functional mechanisms of nitrogen-related chemistry involving biochemical systems.

1.2. References


Chapter 1 — Introduction


Chapter 1 — Introduction


Chapter 2. Computational Enzymology

Methods
2.1. Introduction

Almost four decades have passed since the groundbreaking study done by Warshel and Levitt\(^1\) on first applying combined quantum and molecular mechanics (QM/MM) to investigate the mechanistic pathway of a key glycosidase enzyme (\textit{i.e.}, egg-white lysozyme), revolutionizing the field of computational enzymology. Another method that has played a crucial role in driving this field forward is the application of molecular dynamics (MD) simulations, which can analyze enzymatic systems at a level of detail that has yet to be achieved experimentally.\(^2\) The introduction of potential energy functions by Karplus allowed simulations of these biomolecules to be made possible.\(^3\) It was not until recently that the incredible efforts of Warshel, Levitt, and Karplus were recognized such that they were awarded the prestigious Nobel Prize in Chemistry in 2013 for "developing multi-scale models for such complex chemical systems".\(^4\)

2.2. Molecular Modeling and Simulations of Enzymes

Computational modeling techniques (\textit{e.g.}, QM/MM) and MD simulations have seen tremendous growth since the 1990s, paralleling both the increase in computer technology and accuracy in method development.\(^5\) As depicted in Figure 2.1, the continuous growth of molecular modeling of enzymes is consistent with the rise in number of crystal structures developed.\(^6\) Also, it should be noted that only until the advent of the 21\textsuperscript{st} century, the number of protein structures developed has increased more than the number of publications reflecting enzymatic computational studies.\(^6\) Such advancement in solving protein structures could be due to the improvement in crystallographic techniques, making it easier to stabilize and thus develop, \textit{e.g.}, X-ray crystal structures of proteins, especially those of cellular membrane receptors.\(^7\) Nevertheless, the rise in the number of crystal structures provides an amazing opportunity for computational enzymology to mature even more and provide significant contributions to filling in the gap of experimental studies on biochemical reactions.\(^7,\,8\) For instance, molecular modeling and
simulations could elucidate directly the catalytic power of enzymes at an atomic level while also describing the effects of mutations, key active site residues important for catalysis, ligand binding modes, and structural dynamics.\textsuperscript{2,9}

**Figure 2.1.** Graphic representation of the yearly rise in the number of protein structures (orange) in the Protein Data Bank\textsuperscript{6} and number of articles published per year (blue) which included "molecular model**" and "enzyme" keywords in the Web of Science search engine.

### 2.3. Choosing a Suitable Starting Structure

Before attempting to describe the various computational chemistry methods within this dissertation, it is important to acknowledge the starting geometry for modeling the catalytic reaction of an enzyme. Two commonly used experimental methods, X-ray and NMR crystallography, can be used to generate the Cartesian coordinates for the chemical model.
X-ray crystallography can provide high-resolution protein structures either in an apo (i.e., unbound), holo form (i.e., cofactor-bound), or in complex with a substrate analogue, intermediate, product, and in some cases an inhibitor.\textsuperscript{10} These crystals give us non-dynamic protein structures and are in a non-solvated environment. Other limitations of this method include its inability to determine positions of hydrogen atoms.\textsuperscript{10} In addition, if the protein structure is too flexible (e.g., mobile N-termini), the method may not provide coordinate data for those segments due to disorder in the electron density. A crystallographic parameter that can measure the uncertainty in the atomic position flexibility is the B factor.\textsuperscript{11} A high B factor is proportional to coordinate error, thus one should proceed with caution especially if the X-ray crystal resolution is greater than 2.0 Å.\textsuperscript{12,13}

On the other hand, although NMR protein structures provide dynamic characteristics, they are only suitable for much smaller proteins.\textsuperscript{11,14} This is due to a higher propensity of identical amino acid residues observed in large proteins, which causes overlap and consequently a low crystal resolution.\textsuperscript{11}

It is not very uncommon for a protein to have multiple solved crystal structures.\textsuperscript{15} These can be used to generate different models to examine the effects of varying constraints in modeling and ultimately determine the enzymatic mechanism using different starting geometries.\textsuperscript{14}

In the event some protein structures are not available in the Protein Data Bank (PDB), one could apply homology modeling to create a starting structure. These models are built using the amino acid sequence of an unsolved protein in combination with a related protein whose structure is known.\textsuperscript{16} However, one should treat these molecular models with caution, as the actual protein may differ from the approximate structure.\textsuperscript{2}

Nonetheless, the use of crystallographic structures to create suitable starting models is essential, and in most cases complementary to computational chemistry tools, in investigating the properties and catalytic mechanisms of biochemical systems.
Consequently, in this chapter an overview of the molecular modeling and simulations applied in the subsequent chapters are briefly covered. It is noted that, with the exception of those structures obtained at the highest resolutions, all X-ray structures are supplemented with an MM energy function, which is empirically parameterized. Thus, most crystal structures are partially theoretical.\textsuperscript{17}

2.4. Computational Methods for Modeling Enzyme-Catalyzed Reactions

The science behind enzyme-catalyzed reactions is flourishing due to computational calculations that have been developed to answer inherent mechanistic questions. For instance, methods such as all-atom MD simulations, quantum mechanical (QM) clusters, and hybrid QM/MM have transformed the field of computational enzymology.\textsuperscript{1-3}

2.4.1. Molecular Dynamics (MD) Simulations

Once the 'static' crystal structure of an enzyme has been chosen, it is crucial to understand its dynamics. As a world leader in the study of enzyme catalysis, Jeremy Knowles, once said, "Taking a photograph of a horse does not necessarily tell you how fast it can run".\textsuperscript{8} Thus, using MD simulations we can capture the protein dynamics of an enzyme and determine the conformational changes it may undergo during its catalytic cycle.

Specifically, these computer simulations can be used to generate time-dependent conformers of an enzymatic system by integrating Newton's equations of motion.\textsuperscript{18} This can create a molecular trajectory of the starting reactant complex. Then it is statistically analyzed to select an average structure that is now thermally relaxed and represent better the reactant complex in a subsequent mechanism.
Chapter 2 — Computational Enzymology Methods

Figure 2.2. Schematic illustration of the timescales associated with the conformational events in proteins/enzymes ranging from femtoseconds to seconds.\(^{19}\)

Generally, the conformational changes of a protein range from femtoseconds to seconds and they are caused by motions such as those shown in the timescales of Figure 2.2.\(^{19}\) In particular, covalent bond vibrations of amino acids (e.g., Lys) represent some of the fastest motions occurring with time spans of femto- to picoseconds.\(^{19}\) For instance, a Trp residue on the outskirt of a protein can flip with a timescale ranging from pico- to nanoseconds.\(^{19}\) The movement of active site residues during ligand binding and/or enzyme catalysis can occur in milliseconds, depending on the magnitude of structural...
Finally, the rearrangement of protein chains which is necessary during (un)folding of β-barrel proteins, for example, can involve molecular motions on the second timescale and even hours in some other cases.\textsuperscript{20}

It is noted that the high computational costs of conventional MD can limit access to timescale events in the nanosecond range.\textsuperscript{21} Thus, one can see how this can be a problem especially for biochemical processes that are on much higher timescales of milliseconds and beyond. Fortunately, there are several types of solutions that can allow access to longer simulations such as enhanced sampling techniques and coarse-grain MD simulations.

2.4.1.1. Molecular Mechanics (MM)

Both conventional and enhanced sampling techniques are generally performed using empirical-based methods such as molecular mechanics (MM). With MM one can predict the total energy of the system as a function of conformation. A caveat of this method is that the electrons of the system are not taken into account when calculating the energy of the system.\textsuperscript{18} Only the system's nuclei are considered and as such bond breaking and forming processes cannot be described. Nevertheless, MM methods are still applicable to describe covalent and non-covalent interactions and thus they are essential for MD simulations used to study complex conformational events.\textsuperscript{2,22}

When considering covalent interactions, the energy is described by bonds, angles, and torsional angles, whereas non-covalent interactions are described by van der Waals (vdW) and electrostatic interactions. A pictorial representation is provided in Figure 2.2 below using a peptide model.
Chapter 2 — Computational Enzymology Methods

Figure 2.3. A peptide molecular model showing the bond stretching (str), valence angle bending (bend), torsional angle (tors), and non-covalent interactions such as van der Waals (vdW) or electrostatics. Note each R represents the side chain of an amino acid.

Mathematically, the total MM energy function ($E_{total}$) is represented as follows:

$$E_{total} = \sum E_{str} + \sum E_{bend} + \sum E_{tors} + \sum E_{vdW} + \sum E_{elect}$$  \hspace{1cm} (2.1)

In Equation 2.1, the first three terms on the right-hand side, $\sum E_{str}$, $\sum E_{bend}$, and $\sum E_{tors}$ represent the energy components with respect to the bond stretching, valence angle bending, and dihedral angle torsion. These functions are expressed as simple harmonic and periodic expressions as shown in Equation 2.2. The last two terms, $\sum E_{vdW}$ and $\sum E_{elect}$ represent the non-covalent contributions, specifically, vdW and electrostatic interaction energies, correspondingly. vdW interaction functions are expressed as Lennard-Jones potentials, whereas the electrostatic interactions are computed using Coulomb's law (Equation 2.2).

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2$$ \hspace{1cm} (2.2)

$$+ \sum_{dihedrals} \frac{V_n}{2} [1 - \cos (n\phi - \gamma)] + \sum_{i<j} \left( \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) + \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}}$$

As mentioned previously, molecular mechanics methods are empirical, meaning that experimentally or computationally derived parameters are incorporated into their energy functions. Together, the mathematical energy expression plus the empirical parameters
form the MM force field. Consequently, different force fields exist for various chemical systems including organic molecules, biomolecules, and many others. A commonly-used force field is Assisted Modeling Building with Energy Refinement (AMBER) which was developed to accurately represent conformations and intermolecular interactions for both proteins and nucleic acids.

2.4.1.2. MD Protocols

There are several ways of running an MD simulation for the purpose of obtaining an equilibrated protein structure and hence a suitable protein model in aqueous solution. One effective method is simulated annealing (applied in Chapter 3) which is analogous to the annealing of a semi-crystalline polymer, where the crystal is heated and maintained above its recrystallization temperature, and then gradually cooled until it becomes more crystalline. Specifically, an MD protocol would involve sequential heating and controlled cooling of the enzymatic system in order to reach a global conformational minimum without unnecessarily prolonging the simulation, thus minimizing computational costs. Another means of conformational convergence involves equilibration of the system between 150 and 300 K for a period of 100 ps followed by a production run of 10 ns (conformational sampling) as applied in most of the dissertation chapters. Consequently, the molecular trajectory is examined using a data-mining technique such as cluster analysis. A representative structure is then chosen to construct a chemical model for subsequent use in QM cluster or QM/MM studies. The latter two methods are covered in the next two sections, respectively.

2.4.2. Quantum Mechanics (QM) Cluster Approach

The next stage in computational enzymology involves investigations of the catalytic mechanism of an enzyme typically using a QM cluster approach. This method has proven particularly useful in modeling active sites of enzymes up to several hundred atoms, providing good accuracy. Unlike the MM approach discussed above, QM methods can
treat the behavior of electrons, which dictate chemical reactions. As a result, the formation and breaking of bonds can be described appropriately to determine the geometries and energies of intermediates and transition state structures.

2.4.2.1. QM Methods

There are several kinds of QM methods that are capable of calculating the electronic structure of a molecular system. These include *ab initio*, density functional theory, and semi-empirical calculations.

These methods provide a means of approximation to solving the time-independent Schrödinger equation shown below since it cannot be solved exactly for multi-electron systems. In Equation 2.3, $\hat{H}$ is the non-relativistic time-independent Hamiltonian operator, $\psi$ is the wavefunction of a particular system, depending on the particle positions, $r$, and $E$ is the energy of the system described by the wavefunction.

$$\hat{H}\psi(r) = E\psi(r) \quad (2.3)$$

For molecular systems, $\hat{H}$ can be written as a five-term summation:

$$\hat{H} = -\sum_{l=1}^{N_E} \frac{1}{2} v_l^2 - \sum_{\alpha=1}^{N_N} \frac{1}{2M_{\alpha}} v_{\alpha}^2 - \sum_{l=1}^{N_E} \sum_{\alpha=1}^{N_N} \frac{Z_{\alpha}}{r_{l\alpha}} + \sum_{l=1}^{N_E} \sum_{j>l} \frac{1}{r_{lj}} + \sum_{\alpha=1}^{N_N} \sum_{\beta>\alpha} \frac{Z_{\alpha}Z_{\beta}}{r_{\alpha\beta}} \quad (2.4)$$

In Equation 2.4, the first two summations represent the kinetic energies of $N_E$ electrons and $N_N$ nuclei, correspondingly. The subsequent double-summations are representative of the potential energies of electron–nuclei, electron–electron, and nuclei–nuclei electrostatic interactions, respectively. It is further noted that $M_{\alpha}$ is the mass of nucleus $\alpha$, $Z_{\alpha}$ is the atomic number of nucleus $\alpha$, and $r$ is the distance between two particles.

One way of simplifying the molecular Hamiltonian is to apply the Born-Oppenheimer (BO) approximation (Equation 2.4). This treats the electrons moving in a field where the
much more massive nuclei are assumed to be stationary. As a result, the kinetic energy of nucleus $\alpha$ is now zero and the nuclear–nuclear repulsion term is a constant, resulting in an electronic Hamiltonian.$^{34}$

The simplified Hamiltonian still remains unsolvable for multi-electron systems since the electron–electron repulsion term is dependent on two electrons, thus causing the electronic Schrödinger equation to be inseparable.$^{35}$ By ignoring these electron–electron interactions the N-electron wavefunction is separable into N-one electron wavefunctions. Thankfully, the orbital approximation is a reasonable assumption where the motion of an electron is assumed to be independent of the others. As a result, each can be assigned its own one-electron orbital. This implies that the potential energy term for the electron–electron interaction is effectively ignored. However, in reality electrons do interact with each other.

The conventional Hartree-Fock (HF) method has been established on the assumption that the electron–electron repulsion is treated in an average way where each electron moves through a smeared-out field produced by the other electrons.$^{33}$ Yet this method ignores electron correlation, underestimating the electron–electron repulsion, increasing the total electronic energy of a system.$^{34}$ There are extensions to the basic HF theory or otherwise known as post-HF methods that can describe the electron correlation. These methods however are not practical for studying large biochemical systems due to high computational costs incurred in performing these calculations.

Fortunately, an alternative QM method such as density functional theory (DFT) can allow larger systems to be studied in a feasible manner when compared to correlated ab initio methods.$^{36}$ In DFT, the energy of the N-electron system is computed from functionals of the electron density which is a function of only three variables, i.e., $x$, $y$, and $z$ electron coordinates (3N), regardless of the value of N.$^{36}$ This explains the reduced computational costs that are associated with the DFT approach in contrast to wavefunction-based methods whose $\psi$ depends on 3N variables for each electron.
Importantly, although exact DFT functionals are unknown, approximate forms have been developed.\textsuperscript{37} One functional is the local density approximation (LDA),\textsuperscript{38} which is dependent only on the local electron density. More complex DFT methods include generalized gradient approximation (GGA)\textsuperscript{39} and meta-GGA\textsuperscript{40} functionals that are dependent on the gradients and second derivatives of the electron density, respectively. Other DFT functionals that have proven to be mostly reliable in computational chemistry are hybrid functionals such as B3LYP\textsuperscript{41, 42} that contain a percentage of the exact exchange from HF theory. One drawback for this DFT functional however is that it cannot describe dispersion interactions, \textit{e.g.}, in ligand binding where the protein pocket is enriched in aromatic residues.\textsuperscript{27} In such cases, this can be solved by adding corrections such as those of Grimme's dispersion corrections\textsuperscript{43} or applying meta-hybrid GGA functionals such as M06\textsuperscript{44} that can describe such effects. Consequently, DFT methods are widely accepted in computational enzymology to model enzyme reactions.\textsuperscript{45}

2.4.2.2. Solvation

When applying a quantum chemical cluster approach to model the active site of enzymatic systems, it is critical to describe the protein environment. One method of describing the polarizing effect of the missing active site surroundings is to use an implicit solvation medium such as the polarizable continuum model (PCM) technique.\textsuperscript{46}

A common PCM-based method that is also used in this dissertation is Integral Equation Formalism-PCM (IEFPCM).\textsuperscript{47, 48} This approach creates a solute cavity such as that shown by the inner circle around methylamine in Figure 2.4, whereas its outer shell represents the polarized medium at a selected dielectric constant (\textit{e.g.}, $\varepsilon = 4$ for protein core and $\varepsilon = \sim 80$ for bulk water).\textsuperscript{49}
Figure 2.4. Polarizable continuum model for methylamine, where the outer shell is represented by the dielectric constant, $\varepsilon$.

Please note that the inner cavity is simplified and it is actually formed by interlocking spheres of their respective vdW radii. It is noted that with such a method, the interactions between solute and solvent are ignored. Thus, it may become important to include solvent molecules into the model in order to account for any key hydrogen bonds in the QM model. Nevertheless, the addition of a polarizable continuum to QM cluster models has been shown in many studies to accurately describe the solvation effects. Moreover, this method is proved most useful for systems that consist of less than ~150 atoms as the saturation of polarization effects is not converged in most cases.\textsuperscript{31}

Finally, for some investigations in this dissertation, the environment polarity of QM models has been described using a dielectric constant ($\varepsilon$) of 4.0 as this value has been found to be reliable for the active site surroundings of enzymes.\textsuperscript{31,50}

2.4.2.3. Setup of QM Cluster Model

Once we have equilibrated the enzyme complex using MD and chosen an average structure, the next step is to develop a QM cluster model (see Chapter 8). Usually this process involves minimizing the representative structure and truncating the complex to describe the active site region using a small to moderate-sized model (e.g., up to 200 atoms).\textsuperscript{9,31} It is important that this model includes atoms of key residues that are involved
in catalysis, substrate binding/orientation, and any other stabilizing interactions near the active site that could affect the catalytic mechanism. For example, the latter include H-bonding and electrostatic interactions. Generally, the whole ligand is included in the QM model. However, there are cases where only part of the ligand is involved in a chemical reaction and as a result only the atoms of interest are modeled to minimize unnecessary computational expense.\textsuperscript{26, 29} In addition, the side chain of amino acids can also be simplified to include only the key functional groups. A few examples include L-lysyl and L-arginyl whose R-groups are modeled as methylamine and guanidine molecules, respectively. It is noted that since there is no explicit protein environment, an approximation can be applied to mimic the solvation effects as described earlier using $\varepsilon = 4.0$.\textsuperscript{31, 50} One important technique that can be used to model the steric effects is to apply constraints by fixing key atoms located on the periphery of the QM cluster.\textsuperscript{35} Additionally, this coordinate-locking scheme will allow the integrity of the active site to be kept during subsequent geometry optimization.

2.4.3. Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Approach

As discussed in the Introduction of this chapter, the Nobel prize-winning QM/MM technique is able to combine high-level QM methods with cost-effective MM calculations in order to characterize the complex biochemical systems much studied in computational enzymology.

Specifically, this hybrid method describes the bond breaking and forming processes that occur within the reactive region of the enzyme (\textit{i.e.}, active site) using a high QM level of theory.\textsuperscript{51} Meanwhile, the remainder of the enzyme (\textit{i.e.}, surrounding active site environment) is treated at MM level of theory. The faster MM can thus provide a description of the steric and/or electrostatic effects of the protein environment (refer to Chapter 2 cover page for one schematic representation of the QM/MM enzyme model).
In contrast to the QM cluster approach where the size of the chemical model is limited to a few hundred atoms, QM/MM methods can be utilized to investigate enzymes using models composed of thousands of atoms. Furthermore, QM/MM is capable of modeling the non-homogeneous protein environment, whereas QM cluster is limited to implicit solvation.

There are various schemes of how the total QM/MM energy is calculated. For the purpose of this dissertation, only "Our own N-layered Integrated molecular Orbital molecular Mechanics" (ONIOM) scheme\(^\text{52}\) is discussed. The ONIOM energy \(E^{\text{ONIOM}}\) is obtained using a subtractive method as shown in Equation 2.5.

\[
E^{\text{ONIOM}} = E^\text{MM}_{\text{full}} + E^\text{QM}_{\text{inner}} - E^\text{MM}_{\text{inner}}
\]  

In the above mathematical expression, the total energy is obtained by adding the inner QM-layer energy and deducting the inner MM-layer energy from the MM energy of the full system.\(^\text{53}\)

Furthermore, the ONIOM(QM/MM) energy can be evaluated using different formalisms. In particular, mechanical or electronic embedding (ME or EE, respectively) can be used to describe the interaction between the atoms in the QM-layer and the ones in the outer MM-layer.\(^\text{53}\) In most cases where environmental electrostatic effects are not important, ME is sufficient and all interactions between the two layers are treated at the MM level of theory (see Equation 2.5).\(^\text{53}\) Consequently, the residues in the MM-layer only provide steric effects. On the other hand, EE is advantageous in accurately describing the electrostatic interactions between QM- and MM-layer. That is, the partial charges of the outer MM-region can polarize the electron density of the inner QM-region by creating field \(v\) as indicated in the expression below.\(^\text{53}\)

\[
E^{\text{ONIOM-EE}} = E^\text{MM}_{\text{full}} + E^v\text{QM}_{\text{inner}} - E^v\text{MM}_{\text{inner}}
\]  

\(^{22}\)
2.4.3.1. Setup of ONIOM(QM/MM) Model

When preparing a QM/MM study, there are several considerations that must be made to define the two layers.\textsuperscript{53} First and foremost one has to use chemical intuition to determine the atoms involved in the chemically active region of the full system. The lower MM level of theory cannot describe any bond breaking or formation processes and thus they cannot be included in this region. In the ONIOM formalism, univalent link atoms (e.g., hydrogen atoms) are typically added to uncapped bonds in the inner QM-layer in order to address the broken covalent interaction between the two partitions.\textsuperscript{53} Importantly, alkene or alkyne bonds cannot be included in separate regions of the model. For example, the aromatic rings of tryptophan should be placed in the same layer. Another aspect to consider when doing the ONIOM partitioning is to avoid cutting carbon–heteroatom bonds and only cut C–C bonds to minimize the error caused by the partitioning. Moreover, the reactive centers should be at least three bonds away from the MM-layer such that the MM parameters are kept the same within the two partitioned models during the course of the enzymatic reaction. As highlighted earlier, if the electrostatic effects of the protein environment are important, they can be described by extending the ONIOM formalism to electrostatic embedding.

2.5. Potential Energy Surfaces

Once an enzymatic model has been developed either via QM cluster or QM/MM approach, a potential energy surface (PES) can be generated to elucidate an enzyme's catalytic mechanism.\textsuperscript{27} Specifically, PESs portray multidimensional surfaces that describe the energy of a system in relation to its coordinates as a result of the BO approximation.\textsuperscript{46, 54} In computational enzymology, mechanistically relevant geometries include the reactant, transition state, intermediate, and product complexes.\textsuperscript{2} These stationary points represent the lowest energy pathway that an enzyme may take and they can be characterized as minima or maxima using frequency calculations.\textsuperscript{55} As a result, a
PES can describe structural and thermochemical information about a biochemical system such as the two-step mechanism example in Figure 2.5. Note that $E_a$ and $E'_a$ represent activation energies of the first and second step, respectively. It is evident that the activation barrier for $\text{IC1} \rightarrow \text{PC}$ is higher, causing this chemical step to be the rate-determining step (RDS). Sometimes there are much more complex catalytic mechanisms, making it difficult to establish the RDS. One technique involves dividing the complex energy pathway into simpler exothermic steps and choosing the one with the highest $E_a$ value.\(^{56}\)

Figure 2.5. A 2D potential energy surface of a two-step reaction mechanism, where $\text{RC} = \text{reactant complex}, \ \text{TS}_x (x = 1, 2) = \text{transition state}, \ \text{and PC} = \text{product complex}$. Note the rate-determining step is $\text{IC1} \rightarrow \text{PC}$ since $E'_a$ is greater than $E_a$.

### 2.6. Proton Affinities

Proton affinities (PAs) are calculated as the difference in energy between the parent species (AH) and its deprotonated derivative ($A^-$).\(^{57}\)

\[
E_{PA} = E_{A^\cdot} - E_{HA}
\]
2.7. Software Details

There are several computational software suites that have been used to perform the studies within this dissertation. In particular, the Molecular Operating Environment (MOE)\(^58\) and NAMD\(^59\) programs were used in developing the molecular models and running the molecular dynamics (MD) simulations, correspondingly. Gaussian 03\(^60\) and 09\(^61\) were used to optimize chemical structures as well to perform frequency calculations, single-point energy calculations, and electrostatic potential surfaces. Furthermore, UCSF Chimera package\(^62\) was also used to visualize protein structures and create the molecular graphic in the protein (un)folding event of Figure 2.2.

2.7. References


(33) Levine, I. N. *Quantum Chemistry*; Prentice Hall: New Jersey, **1991**.


(35) Atkins, P.; Julio, d. *Atkins' Physical Chemistry*; 8th ed.; Oxford University Press, **2006**.


Chapter 2 — Computational Enzymology Methods


(58) *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., Montréal, QC, Canada, 2015.
1802.

(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; 
Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; 
Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, 
O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 
Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; 
Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, 
Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; 
Raghavachari, K.; Scuseria, G. E.; Rachels, J. V.; Curtiss, R. C.; Cioslowski, J.; 
Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; 
Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, 
A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; 
Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision D.02*, Gaussian, Inc., Wallingford 
CT, 2004.

(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; 
Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; 
Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; 
Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, 
Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; 
Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.;
Chapter 2 — Computational Enzymology Methods


*Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT, USA, **2009**.

Chapter 3. A Multi-Scale Computational Study on the Mechanism of Streptococcus pneumoniae Nicotinamidase (SpNic)
3.1. Introduction

Metal ions often play central roles in protein biochemistry such as for their folding, stabilization, and biochemical function.\textsuperscript{1,2} For instance, approximately 40% of all known enzymes require at least one metal ion for their catalytic function.\textsuperscript{3-5} In such cases the metal may, for example, be central to substrate recognition and binding, \textit{e.g.}, \(\text{Mg}^{2+}\) in DNA polymerase,\textsuperscript{6} or redox active within the mechanism, \textit{e.g.}, the oxo-manganese cluster within photosystem II.\textsuperscript{7}

Zinc is one of the most biologically important metal ions,\textsuperscript{8} with peptide hydrolases\textsuperscript{9,10} being amongst some of the most well-known \(\text{Zn(II)}\)-containing enzymes. While the coordination of the \(\text{Zn(II)}\) may vary from 4–6,\textsuperscript{2} these enzymes widely share a general acid/base mechanism.\textsuperscript{11,12} More specifically, the \(\text{Zn(II)}\) facilitates formation of a suitable nucleophile \textit{via} \(\text{H}_2\text{O}\) activation. In addition, it binds to the carbonyl oxygen of the substrate's amide bond to be cleaved. This enhances the susceptibility of the bond to nucleophilic attack at its carbon centre, and helps stabilize the tetrahedral intermediate formed during the overall reaction. A general acid subsequently protonates the amide-nitrogen completing amide bond hydrolysis.\textsuperscript{10}

Nicotinamidases (Nics) are a family of peptide hydrolases that generally contain a \(\text{Zn(II)}\) ion.\textsuperscript{13,14} They catalyze hydrolysis of the amide bond in nicotinamide (Scheme 3.1),\textsuperscript{15,16} and as such play a key role in the metabolism of the ubiquitous and important enzyme cofactor \(\text{NAD}^+\). While Nics are widespread in nature they have not been found within mammals and thus present a potential drug target.\textsuperscript{14} Furthermore, they have recently been used in the activation of tuberculosis prodrugs such as pyrazinamide.\textsuperscript{15,17,18}

Experimentally, Du \textit{et al.}\textsuperscript{19} obtained X-ray crystal structures and activity measurements of a Nic from \textit{Pyrococcus horikoshii}. Based on their results and comparison with homologous enzymes they concluded that the active site contains a catalytic triad, comprised of a cysteiny1, aspartyl and lysyl, a \textit{cis}-amide bond as part of an oxyanion hole, and a catalytic \(\text{Zn(II)}\). They proposed that the Cys acts as a nucleophile while the
Zn(II) activates a coordinated H₂O to hydrolyze the resulting enzyme-intermediate thioester bond. Fyfe et al.²⁰ have obtained X-ray crystal structures of a Nic with either product or a product-analogue bound. Importantly, they concluded that the substrate binds to the Zn(II) via its pyridine ring nitrogen (N1). That is, in the family of nicotinamidases, the Zn(II) is atypical and does not bind to the substrate's amide bond.¹,⁴,²¹

\[
\begin{align*}
\text{nicotinamide} + \text{H}_2\text{O} \rightarrow & \text{nicotinic acid} + \text{NH}_3 \\
\text{Nic} & \\
\end{align*}
\]

Scheme 3.1. Overall reaction for conversion of nicotinamide to nicotinic acid as catalyzed by nicotinamidase (Nic).

Very recently, Sheng et al.²² performed a computational QM/MM study on a nicotinamidase from a yeast species (Pnc1). In Pnc1 the Zn(II) is coordinated to two H₂O molecules, and monodentately to two histidyl imidazoles, an aspartyl carboxylate as well as the substrate's N1 centre in an octahedral arrangement. In contrast to that proposed by Du et al.,¹⁹ they concluded that the hydrolyzing H₂O comes from the bulk solvent while the active site lysyl helps to stabilize some species along the mechanism.²² In addition, they suggested that the Zn(II)-binding site acts as a Lewis acid rather than only the Zn(II) ion. For formation of an enzyme-substrate thioester derivative and its subsequent hydrolysis they obtained rate-limiting barriers of approximately 107.5 and 117.6 kJ mol⁻¹, respectively.

Recently, French et al.¹⁵ examined the mechanism of *Streptococcus pneumoniae* Nic (SpNic) via X-ray crystal structures in combination with inhibition and mutagenic studies. In particular, structures were obtained of the apoenzyme, and both native and a C136S mutant with substrate, product, or inhibitor bound within the active site.¹⁵ They
concluded that in contrast to Pnc1, in SpNic the Zn(II) is coordinated by a single H$_2$O, and monodentately via two histidyl imidazoles (His55 and His71), two R-group carboxylates (Asp53 and Glu64), as well as the substrate's N1 centre. They further suggested that the Zn(II) helps bind and orient the substrate, as well as activating a water for hydrolysis of an intermediate's thioester bond. Almost simultaneously, French et al. performed experimental steady state kinetic and $^{18}$O isotope exchange studies and suggested that the active site lysyl or Zn(II)-bound H$_2$O may have duplicate roles. In particular, either may protonate the substrate's leaving –NH$_2$ group and subsequently activate the incoming water for hydrolysis of the enzyme-intermediate thioester bond.

Based in part on these findings they proposed the mechanism for SpNic shown in Scheme 3.2. More specifically, the thiol of the active site cysteiny1 (Cys136) is deprotonated by the R-group carboxylate of a nearby aspartyl (Asp9). The now activated cysteiny1 thiolate nucleophilically attacks the nicotinamide substrate's amide carbonyl carbon, forming an enzyme-substrate tetrahedral intermediate. The Asp9–COOH proton is then transferred to the leaving amine (–NH$_2$) of the substrate with concomitant collapse of the tetrahedral oxyanion intermediate and release of NH$_3$. It is noted that Sheng et al. suggested that for Pnc1 C–S bond formation occurs after proton transfer from the aspartyl to the substrate, thus also avoiding formation of a tetrahedral intermediate. Noting the potential duplicate roles of the active site lysyl (Lys103) and Zn(II)-bound H$_2$O, French et al. suggested that the mechanism may proceed via pathway A or B (see Scheme 3.2). In A, Lys103 protonates the leaving ammonia and then facilitates attack of an H$_2$O at the tetrahedral intermediate's thioester bond. In B, the Zn(II)-bound H$_2$O protonates the leaving ammonia and the resulting Zn(II)–OH then activates a bulk solvent water for attack at the intermediate's thioester bond. It is noted that in the computational study of Sheng et al. on Pnc1 neither pathway A or B were followed. Instead, in Pnc1 the active site lysyl helps stabilize mechanistic intermediates and transition states while Asp8 activates a water molecule for hydrolysis of the thioester bond. This results in formation
of a second tetrahedral intermediate that collapses to give nicotinate and a neutral Cys167.

**Scheme 3.2.** Proposed mechanism(s) for conversion of nicotinamide to nicotinic acid as catalyzed by the nicotinamidase SpNic.\(^{23}\)

In this present study, the catalytic mechanism of nicotinamidase from *Streptococcus pneumoniae* (SpNic) has been computationally investigated. More specifically, molecular dynamics (MD) and ONIOM(QM/MM)-based approaches have been complementarily applied to investigate substrate binding, as well as the catalytic mechanism and role of key active site residues. In addition, environmental and dispersion effects on the mechanism have been examined *via* the application of several DFT methods.
3.2. Computational Methods

3.2.1. Molecular Dynamics (MD) Simulations

MD simulations were performed using the Molecular Operating Environment (MOE)\textsuperscript{26} program. The X-ray crystal structure of the C136S mutant of homo-tetrameric SpNic complexed with nicotinamide was used as the template structure (PDB ID: 3O94).\textsuperscript{15} As the active site does not contain any monomer interface residues, an appropriate single monomer was selected for further calculations, Ser136 mutated to cysteine, while missing hydrogen atoms were added using the MOE default method. The resulting enzyme-substrate complex was solvated using a 7-Å spherical layer of water molecules. An ellipsoidal potential wall with a scaling constant of 2 was placed around the solvated enzyme-substrate complex, thus ensuring the system lies within the volume of space established by the ellipsoid. A damping functional factor was included to allow the electrostatic and van der Waals potentials to decay smoothly. The solvated complex was optimized using the AMBER99 force field until the root mean square gradient of the total energy was below 0.21 kJ mol\textsuperscript{-1} Å\textsuperscript{-2}. The optimized complex underwent thermal relaxation at constant pressure and temperature. The Nosé-Poincaré thermostat\textsuperscript{27} was coupled with the equations of motion and a 2 fs time step was set for numerical integration. The system was annealed by heating it from 150 to 300 K over a 25 ps period, then held at 300 K for a further 25 ps. Then it was heated to 400 K over a 25 ps period, then held at 400 K for a further 325 ps before being cooled down to 300 K over 50 ps. This annealing process was stopped after a further 50 ps interval held at 300 K. The final structure from the trajectory was then optimized using the AMBER99 force field. We have successfully used this MD protocol in studies on other enzymes.\textsuperscript{28,29}

3.2.2. ONIOM(QM/MM) Calculations

The QM/MM calculations were done using the ONIOM\textsuperscript{30–38} formalism in the Gaussian 09\textsuperscript{39} suite of programs. Optimized geometries and harmonic vibrational
frequencies, to characterize the nature of the stationary points, were obtained at the ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within the mechanical embedding (ME) formalism. That is, the reactive region (high-layer) was described at the B3LYP/6-31G(d)\textsuperscript{40-42} level of theory while the surrounding protein environment (low-layer) was described using the AMBER96\textsuperscript{43} force field. Relative energies were obtained via single-point calculations at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory, on the above optimized structures, within an electronic embedding (EE) formalism. The effects of high-layer method choice and dispersion were modeled via single-point calculations, on the above optimized structures, at the ONIOM(DFT\textsubscript{i}/6-311+G(2df,p):AMBER96) level of theory within both an ME and EE formalism, where DFT\textsubscript{i} = B3LYP-D3 or M06. B3LYP-D3 indicates that Grimme's D3 dispersion\textsuperscript{25} scheme as implemented in Gaussian 09 was applied to the B3LYP functional.

The QM/MM chemical model was derived from the final optimized MD structure (see above) and is illustrated in Figure 3.1. As the steric and electrostatic effects of the environment surrounding the active site can have important affects on the mechanism,\textsuperscript{44} all residues and waters up to 15 Å away from the Zn(II) were included. The QM-region contained the Zn(II) ion, the R-groups of its first coordination sphere residues (His71, Glu64, Asp53, and His55), and its coordinated H\textsubscript{2}O. In addition, the nicotinamide substrate, R-groups of the catalytic triad residues (Lys103, Asp9, and Cys136), and peptide backbones between Val131–Leu132 (α-C\textsubscript{131}–CO–NH–α-C\textsubscript{132}) and Ile135–Cys136 (α-C\textsubscript{135}–CO–NH–α-C\textsubscript{136}) were included. The outer circle in Figure 3.1 indicates the residues and waters included in the MM-layer in their entirety except those partially in the high-layer, while those in red were mutated to glycyl, \textit{i.e.}, only their backbones were included. The latter was done due to their location at the periphery of the system and to reduce computational cost. The α-carbon of each residue in the MM-layer was held fixed at its final AMBER99 optimized position (see above). This computational approach has been successfully applied in studies on related enzymatic systems.\textsuperscript{28}
Figure 3.1. Schematic representation of the substrate-bound active site model used in the QM/MM calculations. The outer circle represents the MM-layer, while those in the inner circle represent the QM-layer. Residues in red were mutated to glycyl (see text).

For proton affinity (PA) calculations on isolated species in aqueous solution, optimized geometries were obtained at the B3LYP/6-31G(d) level of theory. Relative energies were obtained via single-point calculations on the above structures at the IEPFCM-B3LYP/6-311+G(2df,p) level of theory. IEPFCM indicates polar environment effects were included by use of the Integral Equation Formalism-PCM method with a dielectric constant (ε) of 78.3.\textsuperscript{45-48} Active site PAs were calculated at ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE level of theory on the above optimized structures.

3.3. Results and Discussion

3.3.1. The Unbound Active Site

As noted in the Introduction the catalytic activity of SpNic has been experimentally shown to depend on the triad comprising Lys103, Asp9, and Cys136.\textsuperscript{15,23} Thus, elucidating the initial protonation state of these residues is important to a more complete understanding of their possible roles in SpNic's mechanism. For instance, neutralization
of both the R-groups of Asp9 and Cys136 (i.e., Asp9–COOH and Cys136–SH) could inhibit the mechanism and in particular formation of the putative thioester intermediate.

In aqueous solution, the proton affinity (PA) of methyl-thiolate, a model for ionized cysteine, is calculated (see Computational Methods) to be 1249.7 kJ mol$^{-1}$. It is noted that this value is higher than that calculated for H$_2$O (1008.5 kJ mol$^{-1}$) at the same level of theory. However, within the SpNic active site a Cys136 thiolate has a decidedly higher PA of 1504.0 kJ mol$^{-1}$. This suggests that within the active site environment Cys136 is more likely to be neutral.

The role of Lys103 is somewhat ambiguous. It has been suggested by French et al.\textsuperscript{15,23} to play the role of an acid in the mechanism while Fyfe et al.\textsuperscript{20} have alternately proposed that it only acts as an electrostatic stabilizing factor. In aqueous solution, the R-group amine of lysine, modeled as methylamine, is calculated to have a PA of 1192.2 kJ mol$^{-1}$, which is higher than that calculated for H$_2$O in aqueous solution. Within the unbound enzyme active site, the PA of the R-group amine of Lys103 is calculated, at the ONIOM(B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d):AMBER96)-EE level of theory, to increase significantly to 1579.6 kJ mol$^{-1}$. This suggests that within the unbound active site the R-group of Lys103 likely exists in its protonated form (i.e., Lys103–NH$_3^+$). Furthermore, it appears unlikely to be able to act as a mechanistic acid as suggested by French et al.\textsuperscript{15,23} but instead may at least in part have a role as an electrostatic stabilizing factor as proposed by Fyfe et al.\textsuperscript{20} In the mechanism proposed by French et al.,\textsuperscript{15,23} the Zn(II)-coordinated H$_2$O hydrolyzes the thioester bond of a mechanistic intermediate. However, when the Zn(II)–OH$_2$ is modified to a Zn(II)–OH, the PA of Lys103–NH$_2$ increases even further to 1778.2 kJ mol$^{-1}$. Thus, it would again appear that in such a scenario Lys103–NH$_3^+$ is unlikely to be a suitable mechanistic acid.

The possible occurrence of a stable complex in which Cys136–SH has transferred its proton onto Asp9–COO$^-$ (i.e., Lys103–NH$_3^+$······Asp9–COOH······Cys136–S$^-$) was examined. However, no such complex was obtained at the present level of theory,
suggesting that nucleophilic attack of the sulfur of Cys136–SH at the carbonyl carbon (C_{carb}) of the substrate may occur with concomitant transfer of the thiol proton onto the carboxylate of Asp9.

3.3.2. The Substrate-Bound Active Site

The optimized structure of the preferred substrate-bound active site, the reactant complex (RC), is shown in Figure 3.2. It is noted that the RC was overlaid with the X-ray crystal structure (PDB ID: 3O94). This comparison shows that there were no significant differences between the two structures (e.g., Zn(II)⋯N1 distance is 2.27 Å in the PDB structure, whereas in the optimized structure it is 2.21 Å). In agreement with experiment, the Zn(II) ion adopts an essentially octahedral geometry. Specifically, it monodentately ligates to the R-group carboxylates of Asp53 and Glu64, which are coordinated trans to each other with similar Zn(II)⋯O_{carb} lengths of 2.01 and 2.03 Å, respectively. In addition, the Zn(II) ligates almost equidistantly to the R-group imidazoles of His55 (2.28 Å) and His71 (2.29 Å). It is noted that the former (His55) is trans to the nicotinamide substrate's Zn(II)-coordination site. The single H_{2}O bound to the Zn(II) ion has a Zn(II)⋯O_{W} distance of 2.12 Å, approximately 0.10 Å longer than the Zn(II)⋯O_{carb} distances involving Asp53 and Glu64 (see above). However, it should be noted that the water simultaneously forms very strong hydrogen bonds to the non-coordinated R-group carboxylate oxygens of both Asp53 and Glu64; r(O_{W}H_{2}⋯O_{Asp53/Glu64}) = 1.72 and 1.24 Å, respectively. In addition, the same carboxylate oxygen of Asp53 hydrogen bonds to the protonated R-group amine of Lys103 with r(Lys103–NH_{3}^{+}⋯O_{Asp53}) = 1.89 Å. Meanwhile, the nicotinamide substrate is bound via its pyridine N1 centre to the Zn(II) at a distance of 2.21 Å, which is slightly shorter than the Zn(II)⋯His55/71 coordination bonds.

As can be seen in Figure 3.2 the R-groups of the catalytic triad of Lys103, Asp9, and Cys136 form a strong hydrogen bond chain with r(Lys103–NH_{3}^{+}⋯O_{Asp9}) and
\( r(O_{\text{Asp9}} \cdots HS_{\text{Cys136}}) \) lengths of 1.59 and 1.86 Å, respectively. The backbone \( –NH– \) moieties of Leu132 (1.85 Å) and Cys136 (2.59 Å), which may play a role in stabilizing an oxyanion intermediate, form moderate and quite weak hydrogen bonds, respectively, with the substrate's carbonyl oxygen.

**Figure 3.2.** Optimized structure with (a) selected distances shown (Angstroms), (b) ESP charges, and (c) \( \pi \)-interactions in the substrate-bound active site, reactant complex (RC), obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

The structure of the active site-bound nicotinamide substrate was also compared to that obtained for the substrate in aqueous solution in order to provide insights into the effects of binding. Notably, upon binding the substrate's amide group C=O bond
Chapter 3 — A Multi-Scale Computational on the Mechanism of SpNic

lengthens slightly by 0.01 Å to 1.24 Å while the C–N bond shortens by 0.02 Å to 1.34 Å. Furthermore, the substrate's dihedral angle $\angle C2–C3–C(O)–N$ increases on binding by 15.1º to 35.6º.

There are several other interactions between substrate and active site residues that play a role in substrate binding and orientation. In particular, based on observed distances in X-ray crystal structures, Phe14, Phe68, and Tyr106 are thought to be involved in $\pi$-interactions with the nicotinamide ring. As shown in Figure 3.2c, these interactions were observed in the optimized structure obtained for RC. In particular, the $\pi-\pi$ interactions are positioned in an edge-to-face type of arrangement. Both Phe14 and Phe68 aromatic rings are stacked 3.67–3.80 Å away from the nicotinamide substrate's pyridyl. Meanwhile, the Tyr106 aromatic ring is 3.75–4.36 Å away.

It is also observed that upon binding within the active site, the positive charge on the substrate's C$_{\text{carb}}$ centre decreases slightly by 0.09 from that calculated for the isolated substrate in aqueous solution to 0.60 while the negative charge on O$_{\text{carb}}$ is essentially unchanged at -0.56. This would appear to suggest that at least in RC formation, the role of the Zn(II) is primarily to facilitate proper binding orientation of the substrate.

3.3.3. Catalytic Mechanism of SpNic

As described in the Introduction, it has been proposed that the mechanism occurs in two stages: (i) formation of a thioester covalently cross-linked enzyme-substrate complex with loss of the substrate's amine group, and (ii) hydrolysis of the enzyme-substrate's thioester bond and product formation.

3.3.3.1. Stage 1: Formation of a Thioester Enzyme-Intermediate Complex with Loss of Ammonia

The potential energy surface obtained, at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory, for Stage 1 of the overall mechanism of SpNic is shown in Figure 3.3. The
optimized structures of the corresponding intermediates and transition states are shown in Figure 3.4.

![Potential energy surface](image)

**Figure 3.3.** Potential energy surface (kJ mol\(^{-1}\)) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory for formation of the thioester enzyme-intermediate with concomitant loss of ammonia.

The first step is formation of a tetrahedral thioester enzyme-substrate intermediate (IC1). More specifically, the sulfur of Cys136 nucleophilically attacks the nicotinamidase substrate's carbonyl carbon (C\(_{\text{carb}}\)) while concomitantly the Cys136–SH thiol proton is transferred onto the substrate's amide group nitrogen. The proton transfer is facilitated directly by the carboxylate of Asp9. Indeed, in TS1 Asp9 more closely resembles a neutral aspartic acid (Asp9–COOH), indicating that transfer of the proton onto the substrate amide nitrogen occurs late in this step. The resulting tetrahedral intermediate IC1 lies only 21.2 kJ mol\(^{-1}\) higher in energy than RC and is able to reversibly rearrange back to the reactive complex essentially without a barrier. The lower relative energy of TS1
with respect to IC1 is a common artifact of the use of single-point calculations and/or empirical corrections on flat PESs and typically indicates that at the higher level of theory used to obtain relative energies, the reaction likely occurs without a barrier. Interestingly, the environment appears to play a significant role in this step by electrostatically stabilizing both TS1 and IC1. Single-point calculations at the same level of theory but without inclusion of the environment's electrostatic charge (i.e., ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-ME//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory) give a barrier for this step of 68.5 kJ mol$^{-1}$ while IC1 lies 49.2 kJ mol$^{-1}$ higher in energy than RC.
Figure 3.4. Optimized geometries obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory of the transition states (TS1 and TS2) and intermediates (IC1 and IC2) for the first stage of the overall mechanism of SpNic. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

In IC1 there now exists a weak C_{carb}–S_{Cys136} bond as indicated by its length of 1.95 Å while the substrate's –NH$_2$ group is now protonated (Figure 3.4). Concomitantly, the C_{carb}=O_{carb} bond has lengthened from that observed in RC by 0.05 Å to 1.29 Å, while the C_{carb}–N bond has lengthened significantly by 0.24 Å to 1.58 Å (Figure 3.4). There is a slight increase in oxyanionic character of the substrate's O_{carb} centre to −0.62 (Figure 3.5).
This in part causes both the Leu132/Cys136–NH–O\textsubscript{carb} hydrogen bonds to shorten significantly to 1.80 and 2.00 Å, respectively (Figure 3.4). It is noted that the Zn(II) charge (Figure 3.5) is calculated to have decreased slightly in IC\textsubscript{1} to 0.81 while Zn(II)–N1 distance has shortened slightly by 0.02 Å to 2.19 Å.

\textbf{Figure 3.5.} ESP charges for select key species in the SpNic mechanism (H-atom inclusive). Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

The second and final step of Stage 1 is collapse of the tetrahedral intermediate resulting in cleavage of the C\textsubscript{carb}–NH\textsubscript{3} bond. For this step we performed detailed scans of the PES. Importantly, an energy maximum (TS\textsubscript{2}) of 53.5 kJ mol\textsuperscript{−1} with respect to RC
(Figure 3.3) was obtained upon elongating the $C_{\text{carb}}\cdots \text{NH}_3$ bond to 1.98 Å at which distance it is effectively cleaved. This is in fact the rate-limiting step of the overall mechanism. The resulting enzyme-intermediate thioester complex IC2 lies higher in energy than RC by 18.1 kJ mol$^{-1}$ (Figure 3.3). Within IC2 the $C_{\text{carb}}\cdots \text{NH}_3$ distance has lengthened even further to 2.85 Å while the $C_{\text{carb}}\cdots \text{S}_{\text{Cys136}}$ bond has shortened significantly from that in IC1 (1.95 Å) to 1.77 Å and now resembles a typical C–S single bond (Figure 3.4). Furthermore, the $C_{\text{carb}}=\text{O}_{\text{carb}}$ bond has also shortened by 0.06 Å to 1.23 Å. In the calculated ESP charges for IC2 (Figure 3.5) the positive charge on $C_{\text{carb}}$ has decreased to 0.29 while that of $\text{O}_{\text{carb}}$ is now less negative (i.e., less oxyanion character) at $-0.44$.

With the ammonia moiety now effectively free in the active site, several possible scenarios exist where water may be made available for the subsequent hydrolysis stage. It has been suggested that the Zn(II)-bound H$_2$O may be the required water.$^{15}$ However, given its position relative to $C_{\text{carb}}$ ($r(C_{\text{carb}}\cdots \text{O}_{\text{water-Zn}}) = 5.10$ Å) and that it is hydrogen bonded to the nearby carboxylates of Asp53 and Glu64, this would seem unlikely. Alternatively, the cleaved NH$_3$ may leave the active site and be replaced by a solvent H$_2$O. It has been previously proposed by French et al.$^{15,23}$ that the cleaved NH$_3$ may be protonated by the R-group amine of Lys103. However, based on the optimized structure of IC2 this would appear unlikely to occur at least directly due to sterics and without some rearrangement of the active site's hydrogen bonding network. In particular, Lys103 and the leaving NH$_3$ are separated by 5.96 Å and with residues and the thioester intermediate between them, make an unlikely proton transfer. Furthermore, the Lys103–NH$_2$ group remains hydrogen bonded to both the carboxylate R-groups of Asp9 and Asp53 (Figure 3.4). Alternatively, however, the NH$_3$ may simply be eliminated from the active site and replaced by a H$_2$O. This latter option has been used to provide a suitable reactive complex for the second stage of the mechanism.
3.3.3.2. Stage 2: Hydrolysis of the Thioester Enzyme-Intermediate and Product Formation

Unfortunately, no present experimental X-ray crystal structures have been obtained with a water bound in an appropriate position for replacing the leaving NH$_3$. Hence, the NH$_3$ moiety of IC2 was appropriately substituted by a H$_2$O molecule, generating IC2'. The resulting potential energy surface obtained for hydrolysis of the thioester intermediate and product formation is given in Figure 3.6 while the optimized structures of the corresponding intermediates, transition states, and product complexes are given in Figure 3.7.

![Potential Energy Surface](image)

**Figure 3.6.** Potential energy surface (kJ mol$^{-1}$) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory for the hydrolysis of the thioester intermediate and product complex formation.

In the optimized structure of the thioester intermediate IC2', the H$_2$O sits such that its oxygen O$_w$ is 2.59 Å from C$_{carb}$ with one of its lone pairs directed towards C$_{carb}$ (Figure 3.7). The most significant changes observed on introducing the water are due to the strong hydrogen bond (1.87 Å) it forms with the R-group carboxylate of Asp9. This induces a change in the hydrogen bond network of the catalytic triad. Specifically, the R-
group carboxylate of Asp9 transfers a proton back to the R-group amine of Lys103, which in turn affects the latter hydrogen bond interaction with the Zn(II)-coordinated Asp53. This induces a slight increase of 0.08 in the charge on the Zn(II) centre (Figure 3.5).

The first step in the second stage is Asp9-facilitated nucleophilic attack of the water's oxygen (Ow) at the thioester's C_carb centre with concomitant cleavage of the C_carb–S_{Cys136} bond and transfer of a water proton onto the R-group carboxylate of Asp9. This exothermic reaction step proceeds with a quite low barrier of just 26.0 kJ mol\(^{-1}\) via TS3 (Figure 3.6). The product complex formed, PC_a, lies lower in energy than IC2' by 42.4 kJ mol\(^{-1}\). In PC_a, neutral nicotinic acid has been formed and is bound within the active site. It can also be seen that Lys103 plays a stabilizing role by forming a relatively strong hydrogen bond (2.09 Å) with the thiolate of Cys136 (Figure 3.7). In the computational study of Sheng et al.\(^{22}\) on Pnc1 it was similarly concluded that in that case the active site lysyl (Lys122) aids in stabilizing stationary points along the mechanism, although for SpNic it appears more directly involved in stabilizing active site residues in such species. The formation of PC_a in essence marks successful completion of the catalytic mechanism of SpNic. The nicotinic acid can now be released from the active site while the catalytic triad residues are likely to easily return to their initial starting states.

However, an alternate product complex (PC_c) lies marginally lower in energy than PC_a by 2.6 kJ mol\(^{-1}\) (Figure 3.6) at the present level of theory (see Computational Methods). Like PC_a the R-group of Asp9 is non-ionized (i.e., Asp9–COOH) while that of Lys103 is protonated (i.e., Lys103–NH\(_3^+\)) as it can be seen in Figure 3.7. The differences are that in PC_c the thiol of Cys136 is now neutral (i.e., Cys136–SH) while the product is now nicotinate. This rearrangement of the hydrogen bond network can occur in two steps via the intermediate product complex PC_b at a cost of 48.8 kJ mol\(^{-1}\) with respect to PC_a. In PC_b, Cys136 is neutral while Asp9 is ionized, and the there has been a rotation about the C_carb–OH bond in the nicotinic acid product. Both transition structures TS4 and TS5 are
calculated to lie lower in energy than \( \text{PC}_b \) at the level of theory used to obtain the PES in Figure 3.6, indicating that \( \text{PC}_b \) likely rearranges without a barrier to either \( \text{PC}_a \) or \( \text{PC}_c \).

**Figure 3.7.** Optimized geometries obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory of the thioester intermediate + \( \text{H}_2\text{O} \) (IC\( ^2 \)) and other transition states (TS\(_x\), \( x = 3, 4, 5 \)) and product complexes (PC\(_x\), \( x = a, b, c \)) for the second stage of the overall mechanism of SpNic. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
3.3.4. Inclusion of Dispersion Effects on the Catalytic Mechanism of SpNic

Similar to the computational work of Sheng et al.\textsuperscript{22} on the Pnc1 nicotinamidase, we have used the B3LYP functional to describe the high-layer or reactive region of the QM/MM model. This functional has been and currently continues to be widely used in such studies. However, it is unable to describe dispersion interactions which may be important in enzymatic reactions.\textsuperscript{24} Indeed, as noted above for RC, in the case of SpNic several non-polar or hydrophobic groups are present in and around the active site and substrate. Fortunately, there are now corrections that can be applied such as those of Grimme\textsuperscript{25} and newer functionals such as M06 that better account for such effects. The above PESs obtained for the mechanism of SpNic were recalculated using such corrections and functionals within both an ME (Figure 3.8a) and EE (Figure 3.8b) formalism (\textit{i.e.}, without and with inclusion of the effects of the environment's polarity, respectively).

First it should be noted that comparison of the corresponding coloured lines in Figure 3.8a \textit{versus} 3.8b indicates the effects of including the effects of the electrostatic environment on the entire mechanism of SpNic. The same trends are observed regardless of the functional used. Namely, the electrostatic environment lowers all reaction barriers, \textit{i.e.}, preferentially stabilizes transition states (TSs), some even to the extent that they (\textit{i.e.}, TS\textsubscript{1} and TS\textsubscript{5}) essentially do not exist (see above). As a result, Stage 1 of the mechanism becomes a one-step reaction. The smallest decreases of 11.0–11.1 kJ mol\textsuperscript{−1} are observed for TS\textsubscript{4}, while the largest changes are 60.5–61.8 kJ mol\textsuperscript{−1} for TS\textsubscript{1}. While the polar environment also lowers the relative energies of almost all intermediate and product complexes, the observed decreases are significantly less than observed for the TSs. For instance, the smallest decreases are observed for IC\textsubscript{2} and are now just 0.4–1.2 kJ mol\textsuperscript{−1}, while the largest decreases are 44.3–44.6 kJ mol\textsuperscript{−1} and occur for PC\textsubscript{e}. The only exception is PC\textsubscript{b} whose energy increases by 3.1–4.1 kJ mol\textsuperscript{−1}. It is noted that Sheng \textit{et al.}\textsuperscript{22}
concluded that in Pnc1 the Zn(II) binding site acts as a Lewis acid to influence the reaction.

![Figure 3.8](image_url)

**Figure 3.8.** Potential energy surfaces for the overall catalytic mechanism of SpNic obtained at the ONIOM(DFT$_i$6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within the (a) ME and (b) EE formalism. DFT$_i$ = B3LYP (black), B3LYP-D3 (red), or M06 (blue).

The inclusion of dispersion interaction effects via use of Grimme's correction (i.e., DFT$_i$ = B3LYP-D3) generally has negligible or only minor effects on the calculated PES, within either the ME or EE formalism (i.e., without or with inclusion of the polar environment effects, respectively). Furthermore, its effects can be to either increase (e.g., IC2) or
decrease (e.g., \( \text{PC}_b \)) the relative energy. In particular, the largest increases are 16.3–16.4 kJ mol\(^{-1}\) and are observed for \( \text{IC2} \). Meanwhile, the most significant decreases at just –6.4–7.0 kJ mol\(^{-1}\) are obtained for \( \text{PC}_b \). Thus, compared to the impact of including the environment's electrostatic effects (i.e., ME \( \rightarrow \) EE), the inclusion of dispersion effects via Grimme's corrections (i.e., \( \text{DFT}_i = \text{B3LYP} \rightarrow \text{B3LYP-D3} \)) generally has less significant effects. Importantly, however, the overall mechanism and identity of the rate-limiting barrier (TS2) remains unchanged.

For the case in which \( \text{DFT}_i = \text{M06} \) for the high-layer in the single-point calculation (see above) slightly different behaviours are observed between when used within an ME or EE formalism (i.e., without or with inclusion of the environment's electrostatic effects, respectively). Within the former (Figure 3.8a), the M06 functional gives the same general overall mechanism for SpNic as obtained for \( \text{DFT}_i = \text{B3LYP} \). However, now TS1, IC1 and TS2 have lower energies relative to RC by 5.9, 8.5 and 2.2 kJ mol\(^{-1}\), respectively. More significant changes are observed for Stage 2. In particular, while \( \text{DFT}_i = \text{M06} \) generally gives relative energies in agreement with those obtained using B3LYP-D3, it gives a markedly higher barrier for TS4 of 14.5 kJ mol\(^{-1}\) (c.f., Figure 3.5). Indeed, for \( \text{DFT}_i = \text{M06} \) this reaction step essentially has a higher barrier than for \( \text{C}_{\text{carb}} – \text{N} \) cleavage via TS2 though by only 7.5 kJ mol\(^{-1}\), and is thus now the rate-limiting step. In addition, it also predicts notably lower relative energies for \( \text{PC}_b \) and TC5, which are now in fact in good agreement with those obtained using \( \text{DFT}_i = \text{B3LYP-D3} \) (Figure 3.8a). When the polar environment is included via use of the EE formalism, some notable changes in the PES upon changing \( \text{DFT}_i \) to M06 are observed. In particular, as for \( \text{DFT}_i = \text{B3LYP} \) both TS1 and IC1 are greatly stabilized by the polar environment but now they are almost thermoneutral with RC having relative energies of 0.8 and 12.8 kJ mol\(^{-1}\), respectively (c.f., Figure 3.2). For the second stage, the largest changes are observed for TS4 which now has a notably higher energy relative to IC2' by 8.7 kJ mol\(^{-1}\), while both \( \text{PC}_b \) and TS5 are stabilized by 4.3 and 2.0 kJ mol\(^{-1}\), respectively.
**3.4. Conclusions**

Several computational approaches including DFT-small chemical models, molecular dynamics (MD) and ONIOM quantum mechanics/molecular mechanics (QM/MM), have been complementarily applied to the study of a nicotinamidase from *Streptococcus pneumoniae*, referred to as SpNic. Specifically, its catalytic mechanism as well as substrate binding and the role of key active site residues has been investigated.

Initial studies examined the proton affinities of the catalytic triad residues Cys136 and Lys103. It has been suggested that Lys103 may help facilitate proton transfer from the Zn(II)-bound water to give a Zn(II)-bound hydroxyl *(i.e., Zn(II)…OH₂ → Zn(II)…OH⁻)*. However, the proton affinity of the amine R-group of Lys103 increases in the presence of Zn(II)…OH⁻ suggesting that it is unlikely to be able to participate in such a proton transfer. Rather, Lys103 may play a stabilizing role in the mechanism, in particular for the thiol/thiolate of Cys136.

Using ONIOM(QM/MM) approach within both a mechanical embedding (ME) and electronic embedding (EE) formalism, *(i.e., without and with inclusion of the environment's electrostatic effects, the overall two-stage catalytic mechanism of SpNic was elucidated.*

It is shown that the electrostatic environment has a significant impact on the overall mechanism. In particular, within an ME formalism, Stage 1, formation of an enzyme-substrate thioester intermediate with loss of NH₃, occurs *via* a two-step mechanism. At the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory the first step, formation of a tetrahedral enzyme-substrate intermediate *(IC1)* is rate-determining with a barrier of 68.5 kJ mol⁻¹. However, inclusion of the effects of the polar environment results in stabilization of all transition states and, to a lesser extent, intermediates, along the mechanism with only a few exceptions. Indeed, using the same level of theory but within an EE formalism, Stage 1 becomes a one-step reaction: nucleophilic attack of the sulfur of Cys136 at the substrate's C_carb centre occurs.
with concomitant cleavage of the $C_{\text{carb}}$–NH$_2$ bond to give the enzyme-substrate thioester intermediate (IC2) with loss of NH$_3$. Furthermore, the calculated barrier is now only 53.5 kJ mol$^{-1}$. In contrast, Stage 2 of the overall mechanism, hydrolysis of the $C_{\text{carb}}$–S bond in IC2' with formation of the nicotinic acid product essentially occurs in one-step within both the ME and EE formalism with barriers of 62.0 and 26.0 kJ mol$^{-1}$, respectively.

Dispersion interaction effects were modeled via application of Grimme's dispersion corrections$^{25}$ to the B3LYP method, i.e., B3LYP-D3. The largest effects were observed for IC2 which was destabilized by 16.4 kJ mol$^{-1}$ and the alternate product complex PC$_b$ and TS5 which were stabilized by 7.0 and 5.0 kJ mol$^{-1}$, respectively. Notably, their effects are much less in general than those due to polarity of the environment.

Use of the M06 functional, i.e., ONIOM(M06/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96), within both the ME and EE formalism gave results in reasonable agreement within those obtained using B3LYP and B3LYP-D3. Importantly, the same overall catalytic mechanism for SpNic was obtained.

### 3.5. References

Chapter 3 — A Multi-Scale Computational on the Mechanism of SpNic


(26) Molecular Operating Environment (MOE), 2012.10; Chemical Computing Group Inc.: Montréal, QC, Canada, 2014.


Chapter 3 — A Multi-Scale Computational on the Mechanism of SpNic

Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.;
Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.;
Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo,
C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.;
Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.;
Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.

Gaussian 09, Revision D.01; Gaussian, Inc.: Wallingford, CT, USA, 2009.

Chapter 4. A Multi-Scale Computational Study of the Catalytic Mechanism of Maleamate Amidohydrolase (NicF): A Non-Metallo Amidase and the Role of its Active site Threonine
4.1. Introduction

Vitamins play a number of important metabolic roles including in catalysis and receptor binding. For example, vitamin B₆ (pyridoxal-5'-phosphate) is important in Schiff-base enzymatic catalysis, while vitamin D₃ binds to receptors involved in bone and calcium metabolism. Due to their insufficient production within the cell, vitamins must be obtained via our diets. However, they must often be modified after ingestion in order to become biologically active. For instance, vitamin B₃ (nicotinate), critical for NAD⁺ metabolism, must be obtained from nicotinamide by the catalytic action of amide hydrolases such as nicotinamidases. This catabolism is also essential for many physiologically important bacterial enzymes that require N-heteroaromatic derivatives as substrates. Many of these bacteria are present in soil and/or are human pathogens, such as Bordetella bronchiseptica, Streptococcus pneumoniae, and Mycobacterium tuberculosis. As a result, there is also tremendous interest from both agricultural and human health perspectives in understanding the catalytic chemistry of amide hydrolases.

The majority of amide hydrolases contain a metal ion within their active site. These ions are important for both substrate recognition and binding and furthermore, often act as Lewis acids and thus have a key catalytic role. Some well-known examples are the Zn(II)-dependent nicotinamidase and thermolysin. In both cases the metal helps bind and orient the amide substrate, as well as facilitate cleavage and hydrolysis of its amide bond. Notably, the substrates bind quite differently to the Zn(II) ion in these enzymes. In thermolysin, the Zn(II) binds directly to the amide bond oxygen while in nicotinamidase it binds to a pyridyl ring nitrogen some distance from the amide bond being cleaved. Nevertheless, in both systems the Zn(II) acts as a Lewis acid.

Maleamate amidohydrolase (NicF) is a related member of the same enzyme family, and has a key role in the nicotinate catabolic pathway of aerobic bacteria. In particular, it catalyzes the conversion of maleamate to the fumarate-precursor, maleate (Scheme 4.1).
Based on gene studies in *Pseudomonas putida*, Jimenez *et al.*\(^8\) proposed that, as in homologous enzymes such as nicotinamidas,\(^16\) NicF may possess a catalytic triad composed of an aspartyl (Asp31), lysyl (Lys121), and cysteinyl (Cys154).

More recently, Kincaid *et al.*\(^9\) obtained an X-ray crystal structure of an unbound maleamate amidohydrolase from *B. bronchiseptica* (PDB ID: 3UAO) in which Cys150 had been mutated to a cysteine sulfenate (Cso). Based on the structure obtained, they also suggested that Asp29, Lys117, Cys/Cso150 (Cso = cysteine–sulfenate) likely form a catalytic triad.

\[
\begin{align*}
\text{H}_2\text{O} + \text{NicF} & \rightarrow \text{O} \\
\text{1} & \rightarrow \text{2} + \text{NH}_4 \\
\end{align*}
\]

**Scheme 4.1.** The overall deamination reaction of maleamate (1) to maleate (2) catalyzed by the amidohydrolase (NicF).\(^9\)

However, they also proposed as no Zn(II) metal ion was observed to be present, that it did not play a role in catalysis. In addition, they performed kinetic studies and showed that the enzyme functioned efficiently without a metal ion. Thus, unlike the majority of amide hydrolases,\(^18\) maleamate amidohydrolase (NicF) appears to be a metal-independent (non-metallo) cysteine amidase.

As a result of their studies they proposed the catalytic mechanism shown in Scheme 4.2. More specifically, the thiol of the active site cysteinyl (Cys150) is deprotonated by the R-group carboxylate (R–COO\(^-\)) of the nearby aspartyl (Asp29) residue that was observed to be within hydrogen bonding distance.\(^9\) The resulting cysteinyl thiolate is now a markedly stronger nucleophile and attacks the carbonyl carbon (C\(_{\text{carb}}\)) of the maleamate substrate's amide bond. This gives rise to a tetrahedral oxyanion with a covalent C\(_{\text{carb}}\)--S\(_{\text{Cys150}}\) enzyme-intermediate cross-link. The now neutral Asp29 R-group carboxylic acid
then donates its proton, derived from the Cys150 thiol, to the substrate's amine group. This induces cleavage of the \( C_{\text{carb}}-\text{NH}_2 \) bond to give a thioester cross-linked enzyme-intermediate complex and \( \text{NH}_3 \). The \( \text{NH}_3 \) is then replaced by a \( \text{H}_2\text{O} \) molecule. It has been suggested that the Asp29 is able to activate the water for hydrolysis of the thioester bond in part due to the presence of the nearby \( R \)-group of Lys117.\(^9\) The incoming \( \text{H}_2\text{O} \) nucleophilically attacks the \( C_{\text{carb}} \) centre while transferring a proton to Asp29. The resulting tetrahedral oxyanion intermediate ultimately collapses to proton with cleavage of the \( C_{\text{carb}}-\text{S}_{\text{Cys150}} \) bond. Meanwhile, Arg175 and Lys190' (not shown for simplicity) are thought to help stabilize the substrate's anionic charge.\(^9\) However, a number of key questions still remain about the catalytic mechanism of NicF and the exact roles of the active site residues. Furthermore, insights into its non-metallo mechanism can provide an opportunity to compare and contrast with the more common metalloamidases.

In this study a complementary multi-scale computational approach has been used to elucidate the catalytic mechanism of maleamate amidohydrolase from *Bordetella bronchiseptica* (NicF). In particular, molecular dynamics (MD) simulations and an ONIOM(QM/MM)-based method have been used to investigate substrate binding, the catalytic mechanism, and the role of the protein environment and key active site residues.

\[ \text{Scheme 4.2. Proposed mechanism for conversion of maleamate to monoionic maleate catalyzed by the amidohydrolase NicF.} \]
4.2. Computational Methods

4.2.1. Molecular Dynamics (MD) Simulations

Both MD simulation and model preparation were done using the Molecular Operating Environment (MOE)\textsuperscript{21} program. The X-ray crystal structure of the NicF complexed with an acetate ion was used to obtain the chemical model (PDB ID: 3UAO).\textsuperscript{9} Specifically, each active site in the enzyme complex contained an acetate ion that was modified to the maleamate substrate. The cysteine–sulfinate (Cso150) was mutated to cysteine while hydrogen atoms were added using the default MOE method. The enzyme-substrate complex was solvated by a 4-Å spherical layer of water molecules. The resulting solvated complex was restrained within the volume established by an ellipsoidal potential wall with a scaling constant of 2. The electrostatic and van der Waals potentials decayed smoothly by a damping functional factor. The structure was then minimized using the AMBER99 force field until the root mean square gradient of the total energy fell below 0.42 kJ mol\(^{-1}\) Å\(^{-2}\).

The complex then underwent thermal relaxation at constant pressure and temperature. The Nosé-Poincaré thermostat\textsuperscript{22} was coupled with the equations of motion, where a 2-fs time step was set for numerical integration. The system was equilibrated at 150 K for 0.1 ns, after which the system was set to 300 K for an interval of 10 ns. Based on root mean square deviation (RMSD) and cluster analyses, an average structure from the trajectory was then optimized using the AMBER99 force field. This MD protocol has been successfully applied in other enzymatic studies.\textsuperscript{23, 24}

4.2.2. ONIOM(QM/MM) Calculations

Hybrid QM/MM methods were performed using the ONIOM\textsuperscript{25-33} formalism in the Gaussian 09\textsuperscript{34} program. More specifically, optimized geometries and harmonic vibrational frequencies were obtained at the ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within a mechanical embedding (ME) formalism. The reactive region (QM-
layer) was described at the B3LYP/6-31G(d)\textsuperscript{35-37} level of theory while the surrounding protein environment (MM-layer) was described using the AMBER96\textsuperscript{38} force field. Single-point calculations at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory within both an ME and electronic embedding (EE) formalism, on the above optimized structures, were done to obtain the relative energies as well as the effects of the protein environment. Electrostatic potential surfaces were derived from the optimized QM/MM geometries and generated using single points at the B3LYP/6-311+G(2df,p) level of theory. For the isolated species in aqueous solution, structures were optimized at the IEFPCM-B3LYP/6-31G(d) level of theory with a dielectric constant (\(\varepsilon\)) of 78.34.

The QM/MM chemical model was derived from the final optimized AMBER99 structure (see MD section above). All residues and waters up to 26 Å from the amide carbon of the maleamate substrate were included as the steric and electrostatic effects of the environment surrounding the active site can have important effects on the mechanism.\textsuperscript{39} The reactive region included the maleamate substrate, R-groups of the proposed catalytic triad residues (Lys117, Asp29, and Cys150), partial R-groups of the substrate–COO\(^-\) stabilizing residues (Lys190\(^\prime\) and Arg175), and peptide backbones between Ala145–Thr146 (\(\alpha\)-C\(_{145}\)–CO–NH–\(\alpha\)-C\(_{146}\) including \(\beta\)-OH moiety of Thr146) and Gly149–Cys150 (\(\alpha\)-C\(_{149}\)–CO–NH–\(\alpha\)-C\(_{150}\)). All surrounding residues and waters were included in the MM-layer. The \(\alpha\)-carbon of each residue in the MM-layer was kept fixed at its final MD optimized position. This computational approach has been successfully applied on related enzymes.\textsuperscript{18}

4.3. Results and Discussion

4.3.1. The Maleamate-Bound Active Site

As described in the Introduction, in the experimentally proposed mechanism the carboxylate group of maleamate (-C_4O_1O_2^-) is suggested to bind with Arg175 and Lys190\(^\prime\).\textsuperscript{9} In the average MD structure of the enzyme-substrate complex, the substrate's
carboxylate does indeed form salt bridges with the R-groups of Arg175 and Lys190' with \(C_{\zeta} \text{Arg175} \cdots \text{O}_1'\) and \(N_{\text{Lys190}'} \cdots \text{O}_2'\) distances of 3.38 and 2.74 Å, respectively (Table A4.1). Additionally, however, the β-hydroxyl group of Thr146 also strongly hydrogen bonds to the substrate carboxylate with a \(\text{Thr146OH} \cdots \text{O}_1'\) distance of 1.77 Å.

In the resulting QM/MM optimized structure of the reactant complex (RC), shown in Figure 4.1A, the R-group amine of Lys190' has transferred a proton onto the substrate's carboxylate (i.e., both Lys190' and the substrate are neutral). Furthermore, the Thr146 R-group β-hydroxyl is now hydrogen bonded with the maleamate's amide carbonyl oxygen. Despite these differences, the R-groups of Arg175 and Lys190' remain hydrogen bonded with the substrate's carboxylic group with \(C_{\zeta} \text{Arg175} \cdots \text{O}_1'\) and \(N_{\text{Lys190}'} \cdots \text{O}_2'\) distances of 2.66 and 3.22 Å, respectively. It is noted that the R-group of Arg175 also forms a moderately strong hydrogen bond (1.98 Å) to the β-hydroxyl oxygen of Thr146, \(r(N_{\text{Lys190}'H} \cdots \text{O}_2') = 1.98\) Å.

The effects of binding on the maleamate substrate were considered by comparing its bound structure (Figure 4.1A) with both its unbound ionized and neutral forms in an aqueous environment (Scheme 4.3). Upon binding maleamate's amide carbonyl, it slightly lengthens by 0.01 Å to 1.25 Å while the \(C_2=C_3\) bond shortens by 0.02 Å to 1.33 Å. Neutralization of the substrate's (maleamic acid) carboxylate via protonation, mimicking what was observed upon active site binding, shortens the alkenyl bond by 0.01 Å to 1.33 Å. Meanwhile, the amide carbonyl decreases by 0.02 Å to 1.25 Å. Perhaps most importantly, for both the ionized and neutral forms of the substrate in aqueous solution, the carbon backbones are almost planar with the amide and carboxylate groups perpendicular to each other. In contrast, in the active site-bound conformation, the C backbone of maleamic acid is approximately 10.0° from planar.

Also in Figure 4.1A, the catalytic triad R-groups of Lys117, Asp29, and Cys150 form moderate hydrogen bonds with \(r(N_{\text{Lys117}'} \cdots \text{HOOC}–\text{Asp29})\) and \(r(O_{\text{Asp29}} \cdots \text{HS}–\text{Cys150})\) lengths of 1.63 and 2.26 Å, respectively. A putative oxyanion hole, composed of the
backbone –NH– groups of Thr146 and Cys150, has been suggested to play a role in stabilizing oxyanion intermediates formed during the reaction. Specifically, O_{carb}⋯HN–Thr146 and O_{carb}⋯HN–Cys150 form moderate hydrogen bonds of 2.26 and 2.20 Å, respectively with the substrate’s carbonyl oxygen, accounting for the lengthening of the carbonyl bond upon binding. In addition, as previously elucidated by the MD results, the same carbonyl oxygen of maleamate hydrogen bonds to the β-OH of Thr146 with r(O_{carb}⋯HO–Thr146) = 1.84 Å.

Similarly to SpNic, there are several other interactions between substrate and active site residues that play a role in binding and orientation of the substrate. In NicF, these include π-interactions between aromatic side chains of Phe13, Phe34, Phe83 with the maleamate's alkene group as shown in Figure 4.1B. Specifically, the π–π interactions are arranged in an edge-to-side type of arrangement. The aromatic rings of Phe13, Phe34, and Phe83 are aligned 3.65–4.51 Å, 4.23–4.65 Å, and 4.73–4.50 Å away from the maleamate substrate's alkenyl, respectively.

It has been shown experimentally that the NicF activity depends on the catalytic triad consisting of Lys117, Asp29, and Cys150. In a previous computational study done by us on SpNic, a metal-dependent amidohydrolase, we showed the initial protonation state of these residues using proton affinity calculations. For instance, in the reactant complex (RC) of SpNic, the active site Cys136 is likely to be neutral (i.e., Cys136–SH), Asp9 occurs in the ionic state, whereas Lys103 exists in the protonated form (i.e., Lys103–NH₃⁺), which we concluded it likely plays a stabilizing role in the thiol/thiolate of Cys136 and carboxylate/carboxylic group of Asp9 during the mechanism.

However, in the optimized RC structure of NicF, Lys117's amine R-group provides its proton to the R-group carboxylate of Asp29, resulting in Lys117–NH₂ and Asp29–COOH (Figure 4.1A). This proton transfer could be an artifact of the optimization calculation being done in gas phase. Also, the absence of a metal to position the substrate's amide
group closer to the triad in the NicF active site (i.e., in contrast to Zn(II)-bound SpNic site) may be responsible for the increase in acidity of Lys117 amine.

Figure 4.1. Optimized structure of the reactant complex (RC) with (A) selected distances shown (Angstroms), (B) π-interactions, obtained at ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Electrostatic potential surfaces mapped onto the electron density with key ESP charges of (C) maleamate and (D) maleamic acid substrates, as well as the oxyanion hole (E) without Thr146–OH (F) with Thr146–OH interaction, constructed using single points at B3LYP/6-311+G(2df,p). Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Nevertheless, the neutral Asp29 side chain does not inhibit the mechanism since Lys117 shows the ability to act as a general base, particularly, in the formation of the IC1 tetrahedral intermediate (Figure 4.5).

Scheme 4.3. Optimized structures of two possible substrate ionization states in a dielectric constant ($\varepsilon$) of 78.39, obtained at IEFPCM-B3LYP/6-31G(d) level of theory.

From Figure 4.1C-D, it can be seen that the potential mapped onto the electron density emphasizes a pronounced negative charge around the alkenyl and carbonyl oxygen in the maleamate substrate. However, if the carboxylate group is neutral as in the RC (i.e., maleamic acid), the negative effect is much less pronounced on the surface. The ESP charges support this, where the C$_3$ charge increases by 0.21 to $-0.20$ in maleamic acid, while C$_2$ decreases to $-0.23$. In addition, the neutralization of the substrate's carboxylate group results in slight stabilization of the amide's carbonyl oxygen, decreasing its charge by 0.01. Also, a 0.06 increase in the electrophilic character of the C$_{\text{carb}}$ occurs, making it a better candidate for the cysteinyl's nucleophilic attack at $+0.66$. Together these results suggest there is electron delocalization when maleamate is protonated.

In parts E and F of Figure 4.1, we examined the effect of Thr146–OH addition to the oxyanion hole. First, it is noted that in the absence of Thr146–OH, the oxyanion hole does not affect the O$_{\text{carb}}$ charge, while the C$_{\text{carb}}$ does show a much higher electrophilic character at $+0.66$. In Figure 4.1F, the $\beta$-OH of Thr146 helped decrease the negative
charge on $O_{\text{carb}}$ by 0.07 to $-0.45$, showing the potential to stabilize oxyanion intermediates in the NicF mechanism.

### 4.3.2. The Oxyanion Hole

Previous studies of amidohydrolases have suggested that the active site is composed of two backbone amine ($-\text{NH}$) groups that are involved in stabilizing the oxyanion intermediates formed involving the substrate's carbonyl oxygen ($O_{\text{carb}}$) during the reaction.\(^9\)\(^{40}\) In particular, a putative oxyanion hole in the NicF active site involves hydrogen bond interactions between $O_{\text{carb}}$⋅⋅⋅$\text{HN}$–Thr146 and $O_{\text{carb}}$⋅⋅⋅$\text{HN}$–Cys150. Based on our MD results, we observed that an additional H-bond donor exists, where the R-group of threonyl ($O_{\text{carb}}$⋅⋅⋅$\text{HO}$–Thr146) interacts moderately via H-bonding with the carbonyl oxygen of the maleamate substrate at several points during the simulation (see Figures 4.2B and 4.3B).

**Figure 4.2.** Plots of (A) oxyanion hole RMSD and (B) $O_{\text{carb}}$⋅⋅⋅$\text{HO}$–Thr146, (C) $O_{\text{carb}}$⋅⋅⋅$\text{HN}$–Thr146, and (D) $O_{\text{carb}}$⋅⋅⋅$\text{HN}$–Cys150 distances (Angstroms), obtained over a 10.1 ns MD simulation.
Consequently, we performed an RMSD analysis on the oxyanion hole atoms including the $O_{\text{carb}} \cdots HO$–Thr146 interactive atoms too. As shown in Figure 4.2A, we note that there is $0.54 \pm 0.06$ Å fluctuation when plotted over time. In addition we graphed each individual interaction to establish the cause for the large conformational changes that occur during the MD simulation. It was determined that $O_{\text{carb}} \cdots HO$–Thr146 distance fluctuates the most, ranging from 1.96 to 3.43 Å (Figure 4.2B). In contrast, the plots in Figure 4.1C-D show both $O_{\text{carb}} \cdots HN$–Thr146 and $O_{\text{carb}} \cdots HN$–Cys150 average distances of approximately 1.88 Å, a rather moderate hydrogen bond.

With an extension of the MD analysis to the active site residues, we were able to determine the average structure by plotting the RMSD over the 10.1 ns course of the simulation followed by cluster analysis (Figure 4.3). Once the representative structure was obtained, it was overlaid with the conformer that showed the shortest hydrogen bond distance (see Figures 4.2B and 4.3B). Although the average structure does show a longer distance of 2.72 Å for $O_{\text{carb}} \cdots HO$–Thr146 than the other conformer (1.96 Å), the interaction is still present. These differences do have an influence on key active site residues as shown in Figure 4.3B. For instance, in the average structure, Cys150 and Asp29 R-groups interact via H-bonding, facilitating the proton transfer for nucleophilic attack ($i.e.$, Cys150–SH$\cdots\cdots$OOC–Asp29). Though the carboxylate of Asp29 is now rotated, the amine group of Lys117 adjusts accordingly to maintain the electrostatic interaction with the Asp29's carboxylate.
4.3.3. Catalytic Mechanism of NicF

As mentioned in the Introduction, the mechanistic proposal proceeds via a nucleophilic addition-elimination sequence. For instance, in Stage 1, the formation of a thioester enzyme-intermediate complex occurs with loss of the substrate's amine group in the form of ammonia. This is followed by the hydrolysis of the enzyme-substrate's thioester bond and formation of product.

4.3.3.1. Stage 1: Thioester Enzyme-Intermediate Formation with Concomitant Loss of NH₃

The NicF potential energy surfaces (PESs) for Stage 1 of the overall mechanism are shown in Figure 4.4, while the corresponding optimized structures are shown in Figure
4.5. These were obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within ME (red) and EE (black) formalism.

It should be noted that the comparison of the coloured lines illustrated in Figure 4.4 represent the influence of including the environmental polarization effects on the relative energies of the first stage of NicF mechanism.

**Figure 4.4.** PESs (kJ mol\(^{-1}\)) for the first stage of the overall mechanism of NicF, obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory within ME (red) and EE (black) formalism. Note \(\text{TS2}^*\) represents a constrained optimization.

In the first chemical step of the Stage 1 mechanism, the sulfur of Cys150 performs a nucleophilic attack at the carbonyl carbon (\(C_{\text{carb}}\)) of the maleamic acid substrate. Concurrently, the carboxylic acid proton of Asp29 is transferred onto the Lys117's amine
group, now allowing for the cysteiny1 proton transfer onto the available Asp29 carboxylate (see Figures 4.4 and 4.5). In the absence of polar environment effects, the tetrahedral oxyanion intermediate IC1 lies very high in energy relative to RC at 140.8 kJ mol\(^{-1}\). Also IC1 is able to reversibly rearrange back to the reactive complex without a barrier. However, once the environment's electrostatic charge is accounted for, both TS1 and IC1 are significantly stabilized. For instance, this reaction step now has a barrier of only 34.4 kJ mol\(^{-1}\) while IC1 lies 35.1 kJ mol\(^{-1}\) higher in energy than RC, indicating that the environment plays a major role on stabilizing the anionic O\(_\text{carb}\). Note the lower energy of TS1 relative to IC1 is a commonly observed artifact when using higher level of theory single-point calculations on flat PESs.

In Figure 4.5, it is observed that IC1 now has a weak C\(_\text{carb}\)–S\(_\text{Cys150}\) bond as indicated by its length of 2.07 Å with concomitant lengthening of the C\(_\text{carb}\)=O\(_\text{carb}\) bond from that seen in RC by 0.06 Å to 1.31 Å. In addition, the C\(_\text{carb}\)–N bond has lengthened by 0.17 Å to 1.51 Å. There is a great increase from –0.52 to –0.76 in the oxyanionic character of the O\(_\text{carb}\) atom when forming the tetrahedral intermediate in IC1 (see Figures 4.5 and 4.6A). As a result, the Thr146/Cys150–NH···O\(_\text{carb}\) hydrogen bonds shorten greatly to 1.96 and 1.98 Å, respectively (Figure 4.5). It is seen that the Thr146–OH···O\(_\text{carb}\) distance has lengthened slightly by 0.17 Å to 2.01 Å but still maintaining a moderate hydrogen bond with O\(_\text{carb}\). Furthermore, in the presence of an ionic carboxyl group in the IC1-based structure, the negative charge is enhanced even more (~0.79), indicating that the R-groups of Lys190' and Arg175 may play a part as well in stabilizing the oxyanionic intermediate (see Figures 4.5 and 4.6B).

In the next and final step of Stage 1, Asp29–COOH is able to protonate the amine leaving group, resulting in the collapse of the tetrahedral intermediate. Specifically, the proton transfer occurs concurrently with the C\(_\text{carb}\)···NH\(_3\) bond cleavage, forming a thioester intermediate, IC2. This step was performed using a constrained TS optimization (TS2\(^*\)), similarly to SpNic,\(^{18}\) where we elongated the C\(_\text{carb}\)···NH\(_3\) bond to 1.98 Å at
which distance it is essentially cleaved. In this TS, the Asp29 carboxyl proton has fully transferred to the amine leaving group as indicated by \( r(H_{\text{Asp29}}-\text{NH}_2) = 1.02 \, \text{Å} \). Meanwhile, the R-group of Asp29 concurrently received a proton from the Lys117 amine. These suggest that the two proton transfers occur early in this reaction step. An adequate estimate of the \( \text{TS2}^* \) energy barrier was obtained at 95.2 kJ mol\(^{-1}\) within the ME formalism with respect to RC (Figure 4.4). When including polar environment effects, the relative energy of IC2 is lowered by 16.2 kJ mol\(^{-1}\) to 79.0 kJ mol\(^{-1}\). This is actually the rate-limiting barrier of the overall mechanism within the EE formalism.

**Figure 4.5.** Optimized geometries of the transition states (TS1 and TS2*) and intermediates (IC1 and IC2) for Stage 1, obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
The subsequent thioester complex IC2 lies higher in energy than RC by 55.0 kJ mol\(^{-1}\) within ME formalism (Figure 4.4). However, with inclusion of electrostatic effects, IC2 is marginally higher at 59.5 kJ mol\(^{-1}\) with respect to the RC. This emphasizes that the stabilization of the O\(_{\text{carb}}\) atom in TS1 and IC1 is highly dependent on the polar environment. Thus, the use of EE formalism is necessary to describe these environmental interactions. In addition, the C\(_{\text{carb}}\)···NH\(_3\) distance in IC2 has lengthened even further to 2.85 Å while the C\(_{\text{carb}}\)–S\(_{\text{Cys150}}\) bond has shortened greatly from that in IC1 (2.07 Å) to that of a typical C–S single bond of 1.77 Å (Figure 4.5). Moreover, the C\(_{\text{carb}}\)=O\(_{\text{carb}}\) bond has also shortened by 0.07 Å to 1.23 Å.

Interestingly, the comparison of the Stage 1 PES within the EE formalism between the non-metallo NicF enzyme with that of the zinc-dependent SpNic\(^{18}\) shows an important aspect of catalysis in cysteine amidases. In particular, both enzymes involve similar reaction steps yet the relative energies of all stationary points on the surface are higher in NicF with values ranging between 13.6–36.9 kJ mol\(^{-1}\). Additionally, the rate-limiting barrier is 53.5 kJ mol\(^{-1}\) in SpNic, whereas that of NicF is moderately higher by 25.5 kJ mol\(^{-1}\) at an energy of 79.0 kJ mol\(^{-1}\) with respect to RC. These results could suggest that the effect of the Zn(II) metal is much more significant in stabilizing the oxyanionic character of O\(_{\text{carb}}\) as well as positioning the substrate with higher precision near the catalytic triad and oxyanion hole within the active site (i.e., in contrast to weaker basic residues such as Lys190' and Arg175 in NicF which only anchor the maleamate substrate at the carboxyl end, causing more flexibility).

More importantly, without the inclusion of the environmental polar effects (i.e., done within ME formalism), the relative energies of NicF on the calculated Stage 1 PES are even higher than those of SpNic with differences of 28.2–91.1 kJ mol\(^{-1}\). This comparison suggests that the zinc-dependent enzyme, SpNic, is able to compensate much more effectively for the lack of electrostatic environment description in contrast to NicF, where the metal is absent. Nevertheless, the formation of a thioester enzyme-intermediate with
concomitant loss of NH$_3$ still proceeds catalytically without depending on a metal. This is in agreement with experimental work that has indicated NicF has catalytic activity in the absence of Zn(II).

In Figure 4.6A-B, both electrostatic potential maps for the IC2-based intermediates show there is less electron density around the C$_{\text{carb}}$ centre, especially in the neutral carboxyl group case. Likewise, calculated ESP charges indicate that the positive charge on C$_{\text{carb}}$ has decreased while that of O$_{\text{carb}}$ is now less negative.

**Figure 4.6.** Electrostatic potential surfaces mapped onto the electron density of select key species in the NicF mechanism for (A) neutral and (B) ionic carboxyl groups, obtained at B3LYP/6-311+G(2df,p) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
4.3.3.2. Stage 2: Maleic Acid Formation via Hydrolysis of the Thioester Enzyme-Intermediate

As previously proposed in our work on SpNic, the free NH$_3$ may exit the active site and be exchanged with a solvent H$_2$O. Thus, the NH$_3$ was simply removed from the active site and replaced by a H$_2$O. This provided a suitable reactive complex for the second stage of the NicF mechanism. Thus, the NH$_3$ byproduct of IC2 was appropriately exchanged with a H$_2$O molecule, denoted by IC2$'$ in Figure 4.7 and 4.8.

The subsequent potential energy surfaces for the maleic acid formation via the hydrolysis of the thioester intermediate are given in Figure 4.7 while the corresponding optimized structures are provided in Figure 4.8.

![Relative Energy (kJ mol$^{-1}$)](image)

**Figure 4.7.** PESs (kJ mol$^{-1}$) for the second stage of the overall mechanism of NicF, obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)/ONIOM(B3LYP/6-311G(d):AMBER96) level of theory within ME (red) and EE (black) formalism.
In the first step of the Stage 2 mechanism, the catalytic water within IC2' is positioned such that its oxygen O\textsubscript{w} is 2.69 Å from C\textsubscript{carb} (Figure 4.8). O\textsubscript{w} is now capable of performing an S\textsubscript{N}2 reaction with the C\textsubscript{carb} atom where the C\textsubscript{carb}–S\textsubscript{Cys150} bond is cleaved concurrently. It is noted, that the Asp29 carboxyl proton is transferred early in this chemical step onto the Lys117’s amine group, allowing for the transfer of the O\textsubscript{w} proton onto the available Asp29 carboxylate (see Figures 4.7 and 4.8).

**Figure 4.8.** Optimized geometries of the transition states (TS\textsubscript{3} and TS\textsubscript{4}) and product complexes (PC\textsubscript{a} and PC\textsubscript{b}) for Stage 2, obtained at the ONIOM(B3LYP/6-31G(d):AMBER96)-ME level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Within the ME formalism, this chemical step proceeds with a barrier of 102.2 kJ mol\(^{-1}\) as shown in Figure 4.7 by TS\textsubscript{3}. The product complex formed, PC\textsubscript{a}, lies slightly higher in energy than IC2' by 8.1 kJ mol\(^{-1}\), where a neutral maleic acid and an ionic Cys150 are obtained. However, under the environment’s polarity, this step becomes barrierless where TS\textsubscript{3} lies lower in energy than IC2' by \(-5.4\) kJ mol\(^{-1}\), representing a 107.6 kJ mol\(^{-1}\)
energy drop in comparison to the calculations done within ME formalism. Consequently, the product formation step is exothermic within the EE formalism, where $PC_a$ lies at $-42.8$ kJ mol$^{-1}$ relative to $IC2'$.

During product formation, the role of Thr146–OH as a member of the oxyanion hole is strengthened. In particular, the hydrogen bond distance between $O_{\text{carb}}$⋯$HO$–Thr146 is slightly shortened in $TS3$ with respect to $IC2'$ by 0.02 Å to 2.10 Å (Figure 4.8). More significantly, a stronger H-bond forms in $PC_a$ at 1.87 Å in comparison to $IC2'$. As Figure 4.6A-B shows, this is important also because the electron density around the carbonyl group becomes richer upon product formation. Also, the ESP charges show that maleic acid has an $O_{\text{carb}}$ with a more negative charge (0.10 decrease), while $C_{\text{carb}}$ is more positive (0.16 increase).

It is concluded that the active site threonyl side chain (Thr146–OH) aids in stabilizing both intermediates and transition states in the catalytic mechanism, and hence may play an integral part in the oxyanion hole of the non-metallo NicF amidase.

In addition, as shown in Figure 4.8, the product is in the same conformation and ionic state ($i.e.$, maleic acid) in both product complexes. Meanwhile, in $PC_a$, both R-groups of Cys150 and Lys117 are ionic and the Thr146 and Cys150 groups participating in the oxyanion hole are slightly stronger. However, in $PC_b$ the thiol of Cys150 and the amine of Lys117 are now neutral. The oxyanion hole does show weaker H-bond interactions by 0.04 and 0.02 Å between $O_{\text{carb}}$⋯$HN$–Thr146 and $O_{\text{carb}}$⋯$HO$–Thr146, respectively. These may suggest the maleic acid product is ready to be released from the active site.

### 4.4. Conclusions

In the present study, we used multi-scale computational methods to gain insights into the catalytic mechanism of *Bordetella bronchiseptica* NicF amidase as well as its substrate binding, oxyanion hole, and role of key active site residues. In particular, DFT,
molecular dynamics (MD), and ONIOM quantum mechanics/molecular mechanics (QM/MM), have been applied in a complementary approach.

The active site residues involved in the oxyanion hole were examined using both MD and ONIOM(QM/MM). It has been suggested that this region helps facilitate the stabilization of the anionic character of O_{carb} substrate via O_{carb}···HN–Thr146 and O_{carb}···HN–Cys150 interactions. However, our results suggest that Thr146–OH may also play a stabilizing role in the oxyanion hole by H-bonding with the O_{carb} during the mechanism.

The overall catalytic mechanism of NicF was elucidated over two stages using an ONIOM(QM/MM) approach within both a mechanical embedding (ME) and electronic embedding (EE) formalism (i.e., without and with inclusion of the polar environment effects, respectively).

This study also suggests that the electrostatic environment has a significant impact on the overall mechanism. For instance, within an ME formalism, the first step of Stage 1 is the formation of a tetrahedral intermediate (IC1) and it was concluded to be rate-determining at a cost of 140.8 kJ mol\(^{-1}\). On the other hand, the inclusion of the environment's electrostatic charges results in stabilization of all reaction barriers and almost all intermediates on the PESs. Subsequently, within an EE formalism, the calculated barrier is now only 79.0 kJ mol\(^{-1}\).

In Stage 2 of the catalytic mechanism, the cleavage of the C_{carb}–S_{Cys150} bond in IC2' occurs in one step within the ME formalism, resulting in the maleic acid product (PC\(_a\)) with an energy barrier of 102.2 kJ mol\(^{-1}\). When applying EE formalism to the calculations, this chemical step occurs without a barrier at –5.4 kJ mol\(^{-1}\). Here, the rate-determining step is the proton transfer to the R-group of Cys150 with a reaction barrier of only 7.0 kJ mol\(^{-1}\).

Consequently, even though the NicF enzyme lacks a metal for precise substrate placement and Lewis acid catalysis, the non-metallo cysteine amidase active site shows
catalytic activity and proceeds via an enzymatically feasible mechanism. In addition, the identity of the catalytic water has been clarified.

4.5. References


(21) *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., Montréal, QC, Canada, **2015**.


Chapter 5. Insights into the Substrate

Active Site Binding and Mutagenesis of

$\Delta^1$-Pyrroline-5-Carboxylate (P5C)

Dehydrogenase using a Molecular Dynamics Approach
5.1. Introduction

Pyridine nucleotides are coenzymes that play important roles in protein biochemistry including metabolic and regulatory signaling pathways. More specific functions of these nucleotides include redox reaction and ion channel regulation, apoptosis and survival and cell signaling; these occur under both physiological and pathological conditions.

The NAD$^+$ dinucleotide and its phosphorylated form, NADP$^+$, commonly serve as coenzymes in cellular metabolism. In such cases they not only behave as electron carriers in oxidative phosphorylation, but also act as substrates for ADP-ribosylation reactions, as well as precursors of the calcium-mobilizing cyclic ADP-ribose. More importantly, they are involved in proline metabolism with functions such as osmolytic control and regulation of cellular stress responses.

Dehydrogenases are an important group of NAD$^+$-dependent enzymes that play significant roles in metabolism. They are responsible for oxidation-driving mechanisms and consequently involved in disease-causing disorders. In particular, type II hyperprolinemia is an autosomal genetic disorder that causes neurodegeneration problems in humans. The pathology is linked to a deficiency in Δ$^1$-pyrroline-5-carboxylate dehydrogenase (P5CDH) activity, where its substrate, L-glutamate-γ-semialdehyde (GSA) is converted to L-glutamate (Scheme 5.1). GSA is intimately related to P5C via a non-enzymatic hydrolysis. Given the reversibility of this reaction, a deficiency in P5CDH results in overproduction of the P5C substrate. This accumulation causes an elevation in cellular plasma L-proline, an important amino acid believed to have a central nervous system-related role in mammals. P5CDH is also involved in the hydroxyproline catabolism, making it a dual-substrate specific enzyme. Similarly to GSA, P5CDH can bind its hydroxylated form, 4-erythro-hydroxy-L-glutamate-γ-semialdehyde (4-OH-GSA) and form the respective L-glutamate product (4-OH-GLU). The nature of this dual substrate specificity is not yet known.
Chapter 5 — Insights into Substrate Binding and Mutagenesis of P5CDH

Early site-directed mutagenesis studies by Wang et al.\textsuperscript{28} have suggested that mitochondrial aldehyde dehydrogenase (ALDH) relies on a glutamate residue (Glu268) for initial activation of the essential cysteine amino acid, Cys302. Point mutation studies by Farrés et al.\textsuperscript{29} confirmed that Cys302 is indeed an active site residue and its role is to behave as a nucleophile.

\[ \text{GLU} \rightarrow \text{GLU} \]

\[ \text{GSA} \rightarrow \text{GSA} \]

\[ \text{P5C} \]

\[ \text{P5CDH} \]

Scheme 5.1. Overall reaction for the conversion of L-glutamate-\( \gamma \)-semialdehyde (GSA) to glutamate by \( \Delta^1 \)-pyrroline-5-carboxylate dehydrogenase (P5CDH).\textsuperscript{23}

Unlike other NAD\(^+\)-binding proteins, aldehyde dehydrogenases have unique structural features that allow them to make fewer contacts with the coenzyme after binding.\textsuperscript{30-32} Crystallographic studies including X-ray, NMR, and fluorescence techniques by Hammen et al.\textsuperscript{33} have shown that the nicotinamide ring of NAD\(^+\) can have multiple conformations upon binding to ALDH. In particular, one NAD\(^+\) conformation prevents the general base, Glu268, from activating the nucleophilic Cys302. Another conformation shows that the nicotinamide moiety is not close enough to the aldehyde substrate for an efficient hydride transfer. Perez-Miller et al.\textsuperscript{34} used X-ray crystallography and mutagenic studies to determine the structures of the wild type and C302S mutant, complexed with both NAD\(^+\) and NADH. Their results suggested the oxidized cofactor prefers the extended conformation necessary for hydride transfer, whereas NADH binds in a contracted conformation, allowing the deacylation reaction to occur.

More recently, Inagaki et al.\textsuperscript{18} proposed a mechanism for \textit{Thermus thermophilus} P5CDH using X-ray crystallography. In particular, crystal structures of the apoenzyme,
NAD\textsuperscript{+}, NADH, and product (GLU) bound within the active site were obtained. However, they declared that no activator was found for the nucleophilic attack of Cys302, but instead suggested that the oxyanion hole could increase the electrophilic character of the aldehyde carbon (δC) and hence activate the Cys302 residue for attack.\textsuperscript{18} On the other hand, Tsybovsky \textit{et al.}\textsuperscript{32} have proposed that the active site glutamate in aldehyde dehydrogenases may not only be involved in the thioester hydrolysis but also in the activation of the catalytic cysteinyl. Also, it is unclear as to how nicotinamide derivative is displaced from the active site due to uninterpretable electron density in the NADH-bound enzyme.\textsuperscript{18}

Consequently, the steps of the catalytic mechanism of P5CDH have been proposed to occur in three stages as per Scheme 5.2.\textsuperscript{18} In Stage 1, the NAD\textsuperscript{+} coenzyme and GSA substrate bind to the active site, resulting in conformational changes of several active site residues. In particular, the R-group of Glu314 (equivalent of Glu268) rotates, pointing away from the active site, whereas the peptide backbone between Cys315 and Gly316 is shifted. The latter shift allows the peptide carbonyl to interact \textit{via} hydrogen bonding with the nicotinamide --NH. After the binding stage, Cys348 thiol (equivalent of Cys302) is believed to attack δC. However, the nature of activating this nucleophile is unclear. After the Cys348 attack, the formation of a hemithioacetal intermediate occurs, which is believed to be stabilized by hydrogen bonding from Asn211 amide group and the --NH--backbone of Cys348. Similarly to several amidase studies,\textsuperscript{35-37} these two participating residues are known to form the oxyanion hole. During Stage 2, the collapse of this tetrahedral intermediate, results in the formation of a thioacyl enzyme with the concurrent production of NADH (acylation stage). Once the coenzyme leaves the active site, both the R-group of Glu314 and peptide backbone between Cys315 and Gly316 reform their initial conformations, allowing Stage 3 to occur (\textit{i.e.}, deacylation). At this point, a water molecule enters the site and allows the hydrolysis to occur, resulting in another tetrahedral intermediate. Once the product, glutamate (GLU) is formed, Cys348 R-group
is positioned outside the active site, permitting the release of GLU and then the catalytic cycle can continue.

Scheme 5.2. The three-stage mechanistic proposal for the conversion of glutamate-\(\gamma\)-semialdehyde to glutamate catalyzed by \(\Delta^1\)-pyrroline-5-carboxylate dehydrogenase.\(^{18}\) Note sequence numbering of P5CDH is based on mouse-model derived structures.\(^{23}\)

Using available X-ray crystal structures,\(^{18}\) Hempel et al.\(^{38}\) tried to understand the impact the S352L mutation has on the P5CDH structure. Specifically, they performed molecular dynamics simulations using a related aldehyde dehydrogenase and they determined that this hyperprolinemia-associated mutation causes the water network between Ser326 (equivalent of S352) and the catalytic Cys322 (equivalent of Cys348) to collapse. However, X-ray crystallographic and mutagenic studies done by Srivastava et al.\(^{23}\) suggest that the S352L mutation disrupts the substrate recognition, NAD\(^+\) binding,
and Cys348 orientation. These can in turn potentially affect the cysteinylnucleophilic attack and thus P5CDH catalysis.

Here, in this study, Stage 1, the substrate-binding mechanism of the P5CDH active site has been studied using both reactant and product complexes. Specifically, molecular dynamics (MD) simulations have been performed to elucidate the roles active site residues play on the binding of substrate, coenzyme, and product. In addition, *in silico* mutations have been done to explore the effect important genetic mutants have on the binding site.

### 5.2. Computational Methods

The molecular dynamics (MD) calculations were done using the Molecular Operating Environment (MOE)\(^{39}\) and NAMD\(^{40}\) programs.

#### 5.2.1. Chemical Models

The reactant and product complexes were modeled based on several X-ray crystallographic structures of a P5CDH homo-dimer as shown in Figure 5.1. Specifically, the reactant complex (RC) was prepared from a holoenzyme, bound with the NAD\(^+\) coenzyme (PDB ID: 3V9L).\(^{23}\) This structure was then docked with the L-glutamate-γ-semialdehyde (GSA) substrate, where the binding interactions were based on the GLU-bound P5CDH (PDB ID: 3V9K)\(^{23}\) due to the similarity of the ligands. The template structure for the product complex (PC) is the unbound P5CDH active site (PDB ID: 3V9J),\(^{23}\) where the glutamate product was then added and positioned in accordance with the X-ray crystal structure of the GLU-bound complex. Note that the 3V9K product structure was not chosen to model the PC since the R-group of Glu314, a critical amino acid, was not fully crystallized in this complex. In addition, the thiol group of Cys348 was modeled as neutral in both RC and PC complexes. Note that even though the protein is homo-dimeric, we have chosen to keep both monomers due to presence of interface residues.
Figure 5.1. (A) Unbound active site of P5CDH [PDB ID: 3V9J] and its bound sites with (B) GLU product [PDB ID: 3V9K] and (C) NAD\(^+\) coenzyme [PDB ID: 3V9L]. Note Glu314* indicates that its R-group is missing.

5.2.2. Solvation and Energy Minimization

In both RC and PC models, the missing hydrogen atoms were added using the MOE default method. After removing the solvent from each X-ray crystal structure, both enzyme-ligand complexes were then solvated with a 4-Å spherical layer of waters. In order to force the system to lie within the volume of space defined by the surrounding
shell of waters, an ellipsoidal potential wall with a scaling constant of 2 was placed around the solvated enzyme-ligand complexes. To allow the electrostatic and van der Waals potentials to decay smoothly beyond 8–10 Å, a damping functional factor was included. The geometries of the solvated complexes were then optimized using the AMBER12:EHT force field until the root mean square gradient of the total energy fell below 0.21 kJ mol$^{-1}$ Å$^{-1}$. Additionally, only those residues and water molecules within the second environmental shell neighboring the P5CDH active site were free to move in order to reduce computational expense.

5.2.3. Thermal Relaxation

The MD simulations were then performed under constrained pressure and temperature. The equations of motion were coupled with the Nosé-Poincaré thermostat$^{41}$ and the time step for numerical integration was set to 2 fs. Initially, the systems were heated from 150 to 300 K for a period of 100 ps, followed by a production run of 10 ns at 300 K and pressure of 1 atm. Based on active site RMSD and cluster analyses, a representative structure was chosen from each RC and PC trajectory and then optimized using AMBER12:EHT.

5.2.4. Subsequent Models and MD Simulations

From each optimized representative structure of the above 10.1 ns thermal relaxation process, RC and PC systems were run for an additional 5.1 ns (100 ps for equilibration, followed by a 5 ns production run), totaling a 15.2 ns simulation length for each complex. Furthermore, several other models were generated and energy minimized using the AMBER12:EHT force field until the root mean square gradient of the total energy fell below 0.21 kJ mol$^{-1}$ Å$^{-1}$. In total, 20 models were obtained and they are summarized in Tables 5.1, 5.2, and 5.3. In particular, Table 5.1 lists wild type RCs both as Michaelis complexes (MCs; fully-bound) and holo/apoenzymatic forms. The MCs were prepared with either both NAD$^+$/GSA ligands or NAD$^+$/4-OH-GSA bound to the active site to
investigate differences in substrate binding to the P5CDH active site. In addition, an MC model (GSA- and NAD\(^+\)-bound) was prepared with an ionic Cys348 (R–S\(^−\)) and a neutral Glu314 (R–COOH) to determine whether Glu314 plays major role in acylation. The holo/apoenzyme models were obtained with either NAD\(^+\) coenzyme or GSA substrate as well as without any ligands in order to study the unbound or partially bound active site differences.

**Table 5.1.** Wild type reactant complexes considered in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Michaelis Complex</th>
<th>Holo/Apoenzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fully Bound</td>
<td>Activated Cys348</td>
</tr>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

In addition, we performed *in silico* mutations such as S352L, S352A, and E314A for the RC complexes in both fully bound and unbound active sites (Table 5.2).

**Table 5.2.** Mutated reactant complexes considered in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>RC: P5CDH Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fully Bound</td>
</tr>
<tr>
<td></td>
<td>S352L</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 5.3 below, both wild type and mutated PC models are summarized with and without the GLU product in order to study the unbound or bound active site differences along with the substrate orientation.
Table 5.3. Product complexes considered in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Wild Type</th>
<th>Mutations</th>
<th>PC: P5CDH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GLU Bound</td>
<td>Unbound</td>
<td>GLU Bound</td>
</tr>
<tr>
<td>13</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As described in the Thermal Relaxation section, the same simulation procedure was applied for all models with the exception where a 5 ns production run was used instead of the previously applied 10 ns run. All of these calculations were subjected to several RMSD and cluster analyses, resulting in average structures. Each model representative was then used for geometric and comparison analysis to other complexes. We have successfully applied this MD protocol in the study of other enzymatic systems.42, 43

5.3. Results and Discussion

5.3.1. Equilibrated Michaelis Complexes

In our initial studies of the reactant and product complexes of P5CDH, we equilibrated their active sites using 10.1 ns MD simulations (see Computational Methods). As shown in Figure 5.2 below, the RMSDs in the positions of nearby active site residues, waters, and ligand(s) were calculated relative to the energy minimized starting structure to ensure conformational convergence. In particular, the reactant complex (RC), bound to both the GSA substrate and the NAD$^+$ coenzyme, reached equilibrium with surrounding active site RMSDs lying within a narrow range of 0.57 ± 0.10 Å (Figure 5.2A). For the GLU-bound product complex (PC), the RMSDs were calculated similarly, obtaining a much narrower range of 0.39 ± 0.08 Å, possibly due to the absence of the flexible NAD$^+$ coenzyme.
(Figure 5.2B). Note that the greatest deviation in both active sites occurs during the first 100 ps of the MD simulations as the temperature is increasing.

![Figure 5.2](image)

**Figure 5.2.** Surrounding active site RMSDs of the P5CDH (A) reactant complex (RC; GSA- and NAD\(^+\)-bound) and (B) product complex (PC; GLU-bound) during the first 10.1 ns simulation.

5.3.2. *Average Substrate-Bound Complexes*

As mentioned in the Computational Methods, the averages of the 10.1 ns equilibrated RC and PC were used to prepare substrate-bound models to examine the active site consistency. These were subsequently run for 5.1 ns to ensure conformational equilibrium is achieved. For the GSA-bound RC or model 5 (Table 5.1), the active site RMSDs include the thiol of Cys348, carboxyl of Glu314, amide of Asn211, peptide backbone between Cys348 and Lys347, and GSA substrate. A plot of the calculated active site RMSDs with a range of 0.87 ± 0.22 Å is shown in Figure 5.3A. Here, midway through the simulation, the RMSD plateau rises significantly from approximately 0.5 to 0.9 Å, but remains consistent. This was investigated by sampling the GSA-bound RC trajectory structures using cluster analysis, grouping the RMSDs into five clusters. Consequently, five representative average structures were obtained and their GSA-bound active sites were overlaid as seen in Figure 5.4A. These overlaid structures show the GSA substrate to vary in the aldehyde group position.
Figure 5.3. RMSDs of GSA-bound RC based on the positions of active site residues such as the thiol of Cys348, carboxylic acid of Glu314, amide of Asn211, peptide backbone between Cys348 and Lys347, GSA substrate, in the (A) presence of GSA, absence of (B) GSA and (C) both GSA and amide of Asn211 positions. (D) RMSDs in the position of the GSA substrate alone. Note all of these calculations were based on the last 5 ns of the simulation.

In the absence of GSA RMSD position, a plot is obtained as shown in Figure 5.3B along with the overlaid structures in Figure 5.4B. The RMSD plateau remains consistent with a few short-lived deviations where the Asn211 amide side chain becomes distal in position with respect to the substrate's carbonyl oxygen (O$_{\text{carb}}$) at a lengthened distance of 8.17 Å (1.88 Å away from the average B of O$_{\text{carb}}$···HN$_{\text{N211}}$ in Table 5.4). Even at the average B distance of 6.29 Å, there is a lack of a hydrogen bond between the O$_{\text{carb}}$ and amide Asn211, an important interaction in the oxyanion hole and hence catalysis. This is likely due to the apoenzymatic nature of model 5, being in the absence of the NAD$^+$ coenzyme.
Figure 5.4. Overlays of representative average structures for the GSA-bound RC, based on RMSD positions of active site residues in the (A) presence of GSA, absence of (B) GSA and (C) both GSA and Asn211 amide, (D) three plateau representatives. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.3. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

When the RMSDs were obtained in absence of both GSA and Asn211 positions, the range was quite narrow with a value of only 0.48 ± 0.07 Å (Figure 5.3C). As seen in Figure 5.4C, the overlay shows that the remaining active site residues maintain a relatively consistent conformation with the minor exception where the thiol proton of Cys348 rotates in and out of the active site. This is rather reasonable due to the oxyanion hole being unoccupied by the GSA's O\textsubscript{carb}, and thus this enables a rotation of the Cys348 thiol proton.
Finally, looking at a calculation of the GSA RMSD position (Figures 5.3D and 5.4D), it is noted that during the course of the MD simulation the substrate undergoes three main conformational changes from the starting point (see Appendix Figure A5.1). Note the RMSD range in each plateau is negligible with the highest range being 0.30 ± 0.05 Å. Based on the average distances from Table 5.4 for the 1st, 2nd, and 3rd plateau, the GSA's carbonyl carbon (C\textsubscript{carb}) is gradually moving further away from the mechanistic nucleophile, (H)S\textsubscript{C348} sulfur (0.45 Å between the 1st and the last plateau). Note in the average D structure, the C\textsubscript{carb}⋯(H)S\textsubscript{C348} distance is even longer at 5.14 Å. Also, O\textsubscript{carb}⋯HN\textsubscript{C348} and O\textsubscript{carb}⋯HS\textsubscript{C348} distances are lengthened and shortened, respectively, during the simulation (Table 5.4). Together, these results show that although the GSA is bound to the P5CDH active site and anchored to the Glu165, the substrate is not yet positioned in a favourable mechanistic position while in the absence of the NAD\textsuperscript{+} coenzyme.

The bindings of GSA/NAD\textsuperscript{+} ligands (1), NAD\textsuperscript{+} coenzyme (4), and unbound RC (6), all reach conformational equilibrium as shown in Figure A5.2 of the Appendix. Once again, cluster analysis was used to obtain five representative average structures, where each active site overlay shows the ligand interactions to be consistent (see Appendix Figure A5.3). It is noted that upon full binding of the P5CDH active site, the average...
$r(C_{\text{carb}} \cdots (H)S_{C348}) = 4.13 \text{ Å (Table 5.6)}, 0.25 \text{ Å closer than in the GSA-bound RC (c.f. Avg. B in Table 5.4).}$ In addition, the oxyanion hole comprises of $r(O_{\text{carb}} \cdots HN_{C348}) = 3.29 \text{ Å}$ and $r(O_{\text{carb}} \cdots HN_{N211}) = 5.39 \text{ Å}, 0.39$ and $0.90 \text{ Å}$ closer, respectively. Nevertheless, these distances with respect to the aldehyde carbonyl are too long, suggesting that the substrate is not ready for nucleophilic attack as mentioned in the Introduction (also see Activated- and Hydroxylated-Michaelis Complexes in the next section).

When superimposing the average structures of the NAD$^+$-bound (4) and complex 1, the active site interactions are fairly consistent (see Appendix Figure A5.4A). However, once the GSA-bound RC is overlaid against the fully-bound model, several key differences are observed between each other. For instance, in the presence of the coenzyme, Glu447 carboxyl interacts with the ribose moiety of the NAD$^+$ nicotinamid. This interaction is also in good agreement with that observed in the X-ray crystal structure (PDB ID: 3V9L). In the NAD$^+$ absence (5), this glutamyl side chain is now hydrogen bonded to the hydroxyl of Ser287 (see Appendix Figure A5.4B), suggesting Glu447 has an important role in the coenzyme binding as previously proposed by Srivastava and coworkers.23

For the GLU-bound PC or model 13 (Table 5.3), the active site RMSDs are composed of Cys348 thiol, Glu314 carboxylic acid, Asn211 amide, peptide backbone between Cys348 and Lys347, and GLU product positions. The calculated active site RMSDs are plotted against time with a range of $0.56 \pm 0.14 \text{ Å}$ (Figure 5.5A). Although the range is narrow, the fluctuations in this graph are quite frequent.

Looking at the five-structure overlay of Figure 5.6A as well as Table 5.5, we see that most of the proposed mechanistic interactions (i.e., oxyanion hole) are reasonably consistent with $O_{\text{carb}} \cdots HN_{C348}$ and $O_{\text{carb}} \cdots HN_{N211}$ distances varying up to 0.18 and 0.29 Å, respectively. However, the variability is not as fairly consistent for the cysteinyl R-
group (Cys348–SH). In particular, the $C_{\text{carb}}\cdots (H)S_{348}$ and $O_{\text{carb}}\cdots HS_{348}$ distances show fluctuations up to 1.55 and 3.19 Å, correspondingly.

Figure 5.5. RMSDs of GLU-bound PC based on the positions of active site residues in the (A) presence of GLU, absence of (B) Cys348 thiol and (C) both thiol and carboxyl groups of Cys348 and Glu314, respectively. (D) RMSDs in the position of the GLU product alone. Note all of these calculations were based on the last 5 ns of the simulation.

As mentioned in the Introduction, in order to release the GLU product, the Cys348 thiol group is repositioned away from the catalytic site. Thus, steady plot fluctuations in Figure 5.5A illustrate this process as an in- and out- 'switch' between inside (~0.3 Å RMSD line) and outside (~0.6 Å RMSD line) the active site over the course of the 5 ns production run. In addition, the $\angle N\alpha C\gamma C\delta$ dihedral angle of Cys348 varies between $-27.9$ and $-153.7^\circ$, representing a wide range of 125.8°. Moreover, the 'switch' shows that the PC is ready for release of its product and hence, the beginning of the next catalytic cycle.
**Figure 5.6.** Overlays of representative average structures of GLU-bound PC, based on RMSD positions of active site residues in the (A) presence of GLU, absence of (B) Cys348 thiol and (C) both Cys348 thiol and Glu314 carboxyl. Note in (D) the overall average structure is overlaid with the structures of the RMSD spike. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.5. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

In the absence of Cys348 and/or Glu314 R-group(s), the active site RMSDs along with the representative structures are much more consistent at a range of 0.52 ± 0.06 and/or 0.37 ± 0.06 Å, respectively (Figures 5.5B-C and 5.6B-C). A short-lived spike occurs at 1.67 ns (Figure 5.5B) and when overlaid with the average structure in Figure 5.6D, it is clearly seen that the Glu314 carboxylic acid proton is positioned out. This is a rather less favourable orientation due to absence of intramolecular H-bonding within the carboxyl group. Here, the $\partial C_{E314} \cdots (H)S_{C348}$ interaction is 4.84 Å apart. It should be noted that the average $D \partial C_{E314} \cdots (H)S_{C348}$ distance in PC is 4.08 Å (Table 5.5), indicating that Glu314
carboxyl is well placed near the ligand. This proximity suggests Glu314 can act as a base catalyst in the deacylation stage of the mechanism in contrast to the RC model where the average $D \overline{\partial}_{E314} \cdots (H)_{S_{C348}} = 10.26$ Å (Table 5.4). Also, $O_{\text{carb}} \cdots H_{N_{C348}}$ and $O_{\text{carb}} \cdots H_{N_{N211}}$ (Avg. D) are within H-bonding distances of 1.88 and 1.87 Å, respectively. This indicates that the product is properly bound from a mechanistic point of view. Furthermore, the GLU conformation appears to have reached equilibrium, showing a very narrow range at $0.28 \pm 0.05$ Å (Figure 5.5D).

Table 5.5. Average $r(C_{\text{carb}} \cdots (H)_{S_{C348}})$, $r(O_{\text{carb}} \cdots H_{S_{C348}})$, $r(O_{\text{carb}} \cdots H_{N_{C348}})$, $r(O_{\text{carb}} \cdots H_{N_{N211}})$, and $r(\overline{\partial}_{E314} \cdots (H)_{S_{C348}})$ distances (Å) for model 13$^a$ (related to Figure 5.6).

<table>
<thead>
<tr>
<th>Structure</th>
<th>$C_{\text{carb}} \cdots (H)<em>{S</em>{C348}}$</th>
<th>$O_{\text{carb}} \cdots H_{S_{C348}}$</th>
<th>$O_{\text{carb}} \cdots H_{N_{C348}}$</th>
<th>$O_{\text{carb}} \cdots H_{N_{N211}}$</th>
<th>$\overline{\partial}<em>{E314} \cdots (H)</em>{S_{C348}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. A1</td>
<td>3.68</td>
<td>2.51</td>
<td>1.89</td>
<td>2.01</td>
<td>4.65</td>
</tr>
<tr>
<td>Avg. A2</td>
<td>4.39</td>
<td>4.23</td>
<td>1.75</td>
<td>1.81</td>
<td>4.40</td>
</tr>
<tr>
<td>Avg. A3</td>
<td>4.19</td>
<td>2.98</td>
<td>1.91</td>
<td>2.01</td>
<td>4.85</td>
</tr>
<tr>
<td>Avg. A4</td>
<td>4.84</td>
<td>5.34</td>
<td>1.93</td>
<td>1.78</td>
<td>4.00</td>
</tr>
<tr>
<td>Avg. A5</td>
<td>5.23</td>
<td>5.70</td>
<td>1.82</td>
<td>1.72</td>
<td>4.82</td>
</tr>
<tr>
<td>Avg. D</td>
<td>5.30</td>
<td>5.72</td>
<td>1.88</td>
<td>1.87</td>
<td>4.08</td>
</tr>
<tr>
<td>Spike$^*$</td>
<td>5.20</td>
<td>5.57</td>
<td>2.03</td>
<td>1.81</td>
<td>4.84</td>
</tr>
</tbody>
</table>

$^a$ – indicates measurement for one structural point from the spike.

In the unbound PC model (14), the active site RMSDs reach convergence within the first 0.50 ns (see Appendix Figure A5.5). Notably, the starting structure of the 5 ns production run shows that upon ligand (GLU) removal, the oxyanion hole collapses, positioning Asn211 amide and Cys348 –NH– backbone away from the active site with $\angle \Delta (N-\alpha C-\beta C-\delta C) = 94.0^\circ$ and $\angle \Delta (N-\alpha C-\beta C-S) = 103.7^\circ$, respectively (see Appendix Figure A5.6A). Thus, this active site adjustment results in more space allocation for the entrance of GSA and NAD$^+$ in the next round of the catalytic cycle (see Appendix Figure A5.6B).
Chapter 5 — Insights into Substrate Binding and Mutagenesis of P5CDH

5.3.3. Activated- and Hydroxylated-Michaelis Complexes

As stated in the Introduction, it has been suggested that Glu314 may be suitable to act as a base and abstract the Cys348 thiol (–SH), thereby activating P5CDH for catalysis.\(^{28,32}\) Thus, a fully-bound RC model in the 'active' form was prepared with Cys348–S\(^-\)/Glu314–COOH (complex 2). A plot of the active site RMSDs is seen in Figure 5.7A with a spike occurring at 0.28 ns. This large RMSD deviation is due to the repositioning of NAD\(^+\) coenzyme within the catalytic site. In particular, the nicotinamide ring almost bisects the average structure from the top as shown in Figure 5.8B (see Table 5.6 for RMSD Spike distances).

Furthermore, the residue positions in the five-structure overlay of the active RC are consistent, indicating that the conformational equilibrium has been reached (Figures 5.7A and 5.8A). Comparing the average structures of the fully-bound RC (Cys348–SH/Glu314–COO\(^-\)) with its active form (Figure 5.9A and Table 5.6), several key observations can be made. For instance, the nucleophilic distance from the GSA's C\(_{\text{carb}}\) is \(r(C_{\text{carb}}\cdots\text{(H)S}_{\text{C}348}) = 3.55\,\text{Å},\) 0.58 Å closer than its neutral form (i.e., Cys348–SH). In addition, the members of oxyanion hole are closer as well, where the O\(_{\text{carb}}\cdots\text{HN}_{\partial\text{N}211}\) is now within H-bonding distance at 2.21 Å (Table 5.6). Also, it is seen that the C\(_{\text{pyr}}\cdots\text{H}_{\text{carb}}\) is 0.81 Å closer at 6.39 Å, facilitating the proposed aldehyde hydride transfer to the NAD\(^+\) coenzyme. Collectively, these calculations suggest that the presence of an ionic Cys348, aids in binding of GSA to the oxyanion hole, forming an active RC.

According to one crystallographic study,\(^{32}\) two apparent conformations exist in another member of the aldehyde dehydrogenase family, one where the cysteinylnucleophile and the C4 of the nicotinamide ring are covalently attached and one where the sulfur atom is 2.60 Å away from the C4 carbon. In our MD results for the P5CDH system, the average structure of the active complex shows the C\(_{\text{pyr}}\cdots\text{S}_{\text{C}348}\) to be 3.65 Å (Table 5.6) with a planar nicotinamide ring. It is also observed that the average \(r(S_{\text{C}348}\cdots\text{HN}_{\partial\text{N}211}) = 2.19\,\text{Å}\) and the amide moiety of the nicotinamide ring flips in
contrast to model 1, resulting in \( r(S_{C348} \cdots H_{	ext{pyr}}) = 2.09 \text{ Å} \) (Figure 5.9A). These interact moderately to stabilize the negative charge of the Cys348 sulfur atom. As a result, this conformation appears to be dominant at equilibrium, especially since it is critical for the Cys348–S⁻ to be nucleophilically available such that the catalytic mechanism can occur.

**Figure 5.7.** RMSDs of fully-bound RC based on the positions of active site residues and their respective ligands in the (A) active (Cys348–S⁻/Glu314–COOH) and (B)-(C) hydroxylated (4-OH-GSA) forms. In (C) the RMSD positions of the hydroxylated RC ligands are absent. Note all of these calculations were based on the last 5 ns of each simulation.
As stated in the Introduction, P5CDH is able to catalyze the oxidation for not only GSA but its hydroxylated form as well, 4-OH-GSA. The binding of the hydroxylated aldehyde to the holoenzymatic active site (model 3) reached conformational equilibrium after 1.2 ns as shown in Figure 5.7B (0.48 ± 0.09 Å range).

Figure 5.8. Overlays of representative average structures of the RC, based on RMSD positions of active site residues in the (A)-(B) active and (B)-(D) hydroxylated forms. Note in (D) the RMSD positions of the hydroxylated RC ligands are absent. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.7. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
In the overlaid structures of Figure 5.8C-D, obtained from clustering (Figure 5.7B-C), it was determined that the γ-hydroxyl group of GSA maintains a stable interaction via H-bonding with the Glu165 carboxyl at an average \( r(\text{HO}_4 \cdots \text{OOC}_{E165}) = 1.63 \) Å. Thus, this interaction positions the 4-OH-GSA substrate similarly to complex 1 (Figure 5.9B), making it a compelling model for catalysis. For example, the average \( r(\text{C}_{\text{carb}} \cdots (\text{H})\text{S}_{C348}) = 4.17 \) Å (slightly longer by 0.04 Å than model 1) and the \( O_{\text{carb}} \) distances from the oxyanion hole interactive atoms of Cys348 –NH– and Asn211 amide, are 3.75 and 4.58 Å away (+0.46 and –0.81 Å than model 1), respectively (Table 5.6).

![Figure 5.9](image.png)

**Figure 5.9.** Overlays of average structures for the fully-bound RC with (A) active and (B) hydroxylated RC active sites. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.7 and Appendix Figure A5.2A. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

On the other hand, the coenzyme's nicotinamide ring appears to shift away from the active site during which the amide moiety is flipped, now H-bonded to the Cys348 sulfur at 2.65 Å (see Figure 5.8C-D and Appendix Figure A5.7). Due to the coenzyme's partial repositioning, the \( C_{\text{pyr}} \cdots (\text{H})S_{C348} \) and \( C_{\text{pyr}} \cdots H_{\text{carb}} \) distances are longer by 0.27 and 0.63 Å.
with average distances of 3.81 and 7.83 Å, correspondingly (Figure 5.8C and Table 5.6). Because of these longer interactions, the NAD$^+$ hydride transfer appears to be less facile in the presence of neutral Cys348. Thus, this further supports the idea that the Cys348 should be in its active form (i.e., ionic) to facilitate and attract the C4 positive charge of the NAD$^+$ closer to the catalytic region of P5CDH.

Table 5.6. Average $r(C_{\text{carb}}\cdots(H)S_{C348})$, $r(O_{\text{carb}}\cdots HS_{C348})$, $r(O_{\text{carb}}\cdots HN_{C348})$, $r(O_{\text{carb}}\cdots HN_{N211})$, $r(\partial C_{E314}\cdots(H)S_{C348})$, $r(C_{\text{pyr}}\cdots(H)S_{C348})$, and $r(C_{\text{pyr}}\cdots H_{\text{carb}})$ distances (Å) for models 1, 2, and 3 (related to Figures 5.8 and 5.9).

<table>
<thead>
<tr>
<th>#</th>
<th>Structure</th>
<th>$C_{\text{carb}}\cdots (H)S_{C348}$</th>
<th>$O_{\text{carb}}\cdots HS_{C348}$</th>
<th>$O_{\text{carb}}\cdots HN_{C348}$</th>
<th>$O_{\text{carb}}\cdots HN_{N211}$</th>
<th>$\partial C_{E314}\cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}}\cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}}\cdots H_{\text{carb}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Avg.</td>
<td>4.13</td>
<td>2.81</td>
<td>3.29</td>
<td>5.39</td>
<td>9.77</td>
<td>3.54</td>
<td>7.20</td>
</tr>
<tr>
<td>2</td>
<td>Avg. B</td>
<td>3.55</td>
<td>N/A</td>
<td>2.21</td>
<td>4.52</td>
<td>8.80</td>
<td>3.65</td>
<td>6.39</td>
</tr>
<tr>
<td>3</td>
<td>Spike</td>
<td>4.71</td>
<td>N/A</td>
<td>3.45</td>
<td>5.69</td>
<td>8.75</td>
<td>3.48</td>
<td>7.26</td>
</tr>
<tr>
<td>3</td>
<td>Avg. C</td>
<td>4.17</td>
<td>2.04</td>
<td>3.75</td>
<td>4.58</td>
<td>9.60</td>
<td>3.81</td>
<td>7.83</td>
</tr>
</tbody>
</table>

5.3.4. S352L Mutation

As previously mentioned in the Introduction, it has been experimentally shown that type II hyperprolinemia is associated with P5CDH genetic mutations. For example, one such mutation is converting Ser352 to Leu (S352L), causing a deficiency in enzymatic activity and hence an increase in P5C and Pro levels in the cellular plasma. It is noted that Ser352 lies at the end of a catalytic loop. Thus, S352L mutant models were prepared for both reactant and product complexes 7, 10, 15, and 18 in both fully-bound and unbound states and compared with their respective average structures for the wild type (WT) enzyme active sites, 1, 6, 13, and 14.

For binding of GSA in the S352L-holoenzyme mutant (7), it was seen that in the absence of the seryl side chain (–OH), the NAD$^+$ coenzyme is shifted away from the active site causing several conformational changes. For instance, in the overlay composed of the average WT and S352L structures of Figure 5.10A, the hydrophobic/bulky leucyl
(Leu352) R-group pushes against the phenyl ring of Phe449, thus breaking the H-bond interaction between the Glu447 R-group and 3'-OH moiety of NAD⁺'s nicotinamide ribose ring. The lack of this interaction in the mutant results in a change of the coenzyme's position and thus possible catalytic inactivity. In Table 5.7, average \( r(\zeta_{C_{P449}}\cdots\alpha_{C_{C348}}) = 8.03 \) Å and \( r(\partial_{E447}\cdots\alpha_{C_{C348}}) = 6.76 \) Å show that Phe449 and Glu447 as a result, migrated significantly by 4.11 and 4.35 Å, respectively, upon S352L mutation (see Appendix Table A5.2). This supports crystallographic/mutagenesis work done by Srivastava et al.,\(^\text{23}\) wherein the role of Glu447 in the product complex of P5CDH was proposed to aid NAD⁺ binding.

Also, the catalytic thiol of Cys348 is now H-bonded to the oxygen of the nicotinamide carbonyl, placing the aldehyde \( C_{\text{carb}} \) 0.22 Å further away from the cysteinyl sulfur atom in the S352L mutant (Table 5.7). Thus this makes \( C_{\text{carb}} \) less available for the GSA oxidation mechanism. In the unbound RC mutant (10), both Phe449 and Glu447 remain essentially unchanged at \( r(\zeta_{C_{P449}}\cdots\alpha_{C_{C348}}) = 3.98 \) Å and \( r(\partial_{E447}\cdots\alpha_{C_{C348}}) = 11.29 \) Å, correspondingly. However, the aromatic phenyl ring of Phe449 is tilted horizontally by approximately 90° in contrast to models 1 and 6.

For the PC models, the S352L mutation seems to have no effect on the catalytic site of P5CDH. As detailed in the Introduction, stage 3 of the mechanism involves ligand deacylation where the NADH coenzyme is absent (\textit{i.e.}, once reduced, NADH is believed to leave the active site). Consequently, the proposed mechanistic interactions remain virtually unchanged in the presence of L352 while in the bound PC (Figure 5.10C, Table 5.7, and Appendix Table A5.2). However, in the unbound product complex, the oxyanion hole of the dehydrogenase enzyme collapses. Notably this occurs in both WT and mutant forms of unbound RCs as observed in Figure 5.10B.
**Figure 5.10.** Overlays of RC and PC average structures with the S352L mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.8 and A5.10. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Table 5.7. Average $r(C_{\text{carb}} \cdots (H)S_{C348})$, $r(O_{\text{carb}} \cdots HS_{C348})$, $r(O_{\text{carb}} \cdots HN_{C348})$, $r(O_{\text{carb}} \cdots HN_{N211})$, $r(\partial C_{E314} \cdots (H)S_{C348})$, $r(C_{\text{pyr}} \cdots (H)S_{C348})$, $r(C_{\text{pyr}} \cdots H_{\text{carb}})$, $r(\zeta C_{P449} \cdots \alpha C_{C348})$, and $r(\partial C_{E447} \cdots \alpha C_{C348})$ distances (Å) for models 7, 10, 15, and 18 (related to Figure 5.10).

<table>
<thead>
<tr>
<th>#</th>
<th>$C_{\text{carb}} \cdots (H)S_{C348}$</th>
<th>$O_{\text{carb}} \cdots HS_{C348}$</th>
<th>$O_{\text{carb}} \cdots HN_{C348}$</th>
<th>$O_{\text{carb}} \cdots HN_{N211}$</th>
<th>$\zeta C_{C348} \cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}} \cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}} \cdots H_{\text{carb}}$</th>
<th>$\zeta C_{P449} \cdots \alpha C_{C348}$</th>
<th>$\zeta C_{E447} \cdots \alpha C_{C348}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>4.49</td>
<td>4.74</td>
<td>3.89</td>
<td>4.86</td>
<td>10.04</td>
<td>3.36</td>
<td>7.56</td>
<td>8.03</td>
<td>6.76</td>
</tr>
<tr>
<td>10</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>9.54</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3.98</td>
</tr>
<tr>
<td>15</td>
<td>5.12</td>
<td>5.23</td>
<td>1.87</td>
<td>1.81</td>
<td>4.53</td>
<td>N/A</td>
<td>N/A</td>
<td>4.44</td>
<td>7.49</td>
</tr>
<tr>
<td>18</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>6.37</td>
<td>N/A</td>
<td>N/A</td>
<td>4.81</td>
<td>9.19</td>
</tr>
</tbody>
</table>

5.3.5. S352A Mutation

According to a S352A mutagenesis study done previously,\textsuperscript{23} it was proposed that the –OH removal in mutating Ser352 to Leu plays no role in the loss of P5CDH catalytic activity. Similarly, in order to examine the effect of the hydroxyl moiety removal of Ser352, we first performed a S352A in silico mutation on fully-bound RC and PC, resulting in MD-equilibrated models 8 and 16, respectively.

In the comparison of model 8 with its native structure (1), it was observed that the absence of the seryl side chain (–OH) does not appear to cause any major conformational changes. The average structures of WT and S352A in Figure 5.11A show that nicotinamide ribose of NAD\textsuperscript{+} maintains its H-bonding interaction with the Glu447 carboxyl (c.f. Figure 5.10A). It is also noted that the Cys348 thiol orientation remains consistent between the two models (i.e., 1 and 8). In Table 5.8, average $r(\zeta C_{P449} \cdots \alpha C_{C348}) = 5.19$ Å and $r(\partial C_{E447} \cdots \alpha C_{C348}) = 10.58$ Å show that Phe449 and Glu447 migrated negligible distances of 1.27 and 0.53 Å, respectively, in the S352A mutant (c.f. Appendix Table A5.2). This suggests that the active site is not compromised by the S352A mutation. The Phe449 side chain in the unbound RC S352A mutant (Figure 5.11B) keeps a similar position to that seen in the S352L model (c.f., Figure 5.10B).
Moreover, all four S352L and S352A PC models 15, 16, 18, and 19 show similar active site interactions and thus exhibit the potential for catalytic activity in the absence of NAD$^+(H)$ (Figures 5.10BC-11BC and Tables 5.7-8).

**Figure 5.11.** Overlays of RC and PC average structures with the S352A mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.8 and A5.11. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Table 5.8.  Average \( r(C_{\text{carb}}\cdots(H)S_{\text{C348}}) \), \( r(O_{\text{carb}}\cdots(H)S_{\text{C348}}) \), \( r(O_{\text{carb}}\cdots\text{HN}_{\text{N211}}) \), \( r(\partial C_{E314}\cdots(H)S_{\text{C348}}) \), \( r(C_{\text{pyr}}\cdots(H)S_{\text{C348}}) \), \( r(C_{\text{pyr}}\cdots\text{H}_{\text{carb}}) \), \( r(\zeta C_{P449}\cdots\alpha C_{\text{C348}}) \), and \( r(\partial C_{E447}\cdots\alpha C_{\text{C348}}) \) distances (Å) for models 8, 11, 16, and 19 (related to Figure 5.11).

<table>
<thead>
<tr>
<th>#</th>
<th>( C_{\text{carb}}\cdots(H)S_{\text{C348}} )</th>
<th>( O_{\text{carb}}\cdots(H)S_{\text{C348}} )</th>
<th>( O_{\text{carb}}\cdots\text{HN}_{\text{N211}} )</th>
<th>( \partial C_{E314}\cdots(H)S_{\text{C348}} )</th>
<th>( C_{\text{pyr}}\cdots(H)S_{\text{C348}} )</th>
<th>( C_{\text{pyr}}\cdots\text{H}_{\text{carb}} )</th>
<th>( \zeta C_{P449}\cdots\alpha C_{\text{C348}} )</th>
<th>( \partial C_{E447}\cdots\alpha C_{\text{C348}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4.19</td>
<td>1.95</td>
<td>4.38</td>
<td>6.81</td>
<td>9.88</td>
<td>3.75</td>
<td>7.64</td>
<td>5.19</td>
</tr>
<tr>
<td>11</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>10.00</td>
<td>N/A</td>
<td>N/A</td>
<td>3.86</td>
</tr>
<tr>
<td>16</td>
<td>4.98</td>
<td>5.21</td>
<td>1.84</td>
<td>1.87</td>
<td>4.15</td>
<td>N/A</td>
<td>N/A</td>
<td>4.39</td>
</tr>
<tr>
<td>19</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>7.25</td>
<td>N/A</td>
<td>N/A</td>
<td>4.41</td>
</tr>
</tbody>
</table>

5.3.6. E314A Mutation

Based on previous work done on 10-formyltetrahydrofolate dehydrogenase, a mutation of Glu673 (equivalent of Glu314 in P5CDH) to an alanine residue resulted in an inactive enzyme. So it was proposed that this active site glutamyl is involved in ligand acylation. As such we prepared several models such as 9, 12, 17, and 20, where the Glu314 was changed to alanine (Figure 5.12), in order examine its effect on the active site residues.

Similarly to the S352L model, the glutamyl mutation (E314A) shows breaking of the H-bond interaction between the Glu447 R-group and 3'-OH moiety of NAD⁺'s nicotinamide ribose ring. Consequently, the lack of this interaction in both E314A and S352L mutants results in NAD⁺ disorientation and as such a possible inactive enzyme. Although the outcome of these mutations is the same, there are several differences as shown in both Figure 5.12 and Table 5.9 (c.f. Figure 5.10 and Table 5.7). For instance, the average \( r(\zeta C_{P449}\cdots\alpha C_{\text{C348}}) = 3.77 \text{ Å} \) (i.e., model 9), meaning that Phe449 has not shifted much at all and thus is not the cause of the Glu447···NAD⁺ interaction loss in the E314A case. More specifically, we see that upon the E314A mutation, there is a significant change in the backbone between –NH– backbone of Cys315, pushing the NAD⁺ partially outside the active site (Figure 5.9A). In addition, the Cys348 thiol is now
H-bonded to the carbonyl oxygen of the coenzyme's nicotinamide ring, consistent with that seen in the S352L mutation. Furthermore, the average $r(\hat{\epsilon}C_{E447} \cdots \alpha C_{C348}) = 7.14 \, \text{Å}$ shows that Glu447 has migrated significantly by 4.03 Å in the E314A model (Table 5.9).

**Figure 5.12.** Overlays of RC and PC average structures with the E314A mutation. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Appendix Figures A5.9 and A5.12. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

As proposed in the mechanism within the Introduction, upon the hydride transfer to NAD$^+$, the reduced coenzyme is then able to leave the active site. Importantly, any
mutation (i.e., S352L, S352A, or E314A) done on the RC model in the unbound state or in the PC, either bound or unbound, has no impact on the conformation of the active site residues (Figure 5.12B-D), thus remaining catalytically active.

Table 5.9. Average $r(C_{\text{carb}} \cdots (H)S_{C348})$, $r(O_{\text{carb}} \cdots HS_{C348})$, $r(O_{\text{carb}} \cdots HN_{C348})$, $r(O_{\text{carb}} \cdots HN_{N211})$, $r(\partial C_{E314} \cdots (H)S_{C348})$, $r(C_{\text{pyr}} \cdots (H)S_{C348})$, $r(C_{\text{pyr}} \cdots H_{\text{carb}})$, $r(\zeta C_{P449} \cdots \alpha C_{C348})$, and $r(\partial C_{E447} \cdots \alpha C_{C348})$ distances (Å) for models 1, 6, 9, 12, 13, 14, 17, and 20 (related to Figure 5.12).\(^d\)

<table>
<thead>
<tr>
<th>#</th>
<th>$C_{\text{carb}} \cdots (H)S_{C348}$</th>
<th>$O_{\text{carb}} \cdots HS_{C348}$</th>
<th>$O_{\text{carb}} \cdots HN_{C348}$</th>
<th>$O_{\text{carb}} \cdots HN_{N211}$</th>
<th>$\partial C_{E314} \cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}} \cdots (H)S_{C348}$</th>
<th>$C_{\text{pyr}} \cdots H_{\text{carb}}$</th>
<th>$\zeta C_{P449} \cdots \alpha C_{C348}$</th>
<th>$\partial C_{E447} \cdots \alpha C_{C348}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.19</td>
<td>2.23</td>
<td>3.58</td>
<td>4.81</td>
<td>9.97</td>
<td>3.91</td>
<td>8.57</td>
<td>3.95</td>
<td>11.17</td>
</tr>
<tr>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>10.09</td>
<td>N/A</td>
<td>N/A</td>
<td>4.46</td>
<td>11.01</td>
</tr>
<tr>
<td>9</td>
<td>3.93</td>
<td>4.62</td>
<td>3.41</td>
<td>4.49</td>
<td>3.91</td>
<td>7.78</td>
<td>3.77</td>
<td>7.14</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>5.30</td>
<td>N/A</td>
<td>11.09</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4.30</td>
<td>3.09</td>
<td>1.86</td>
<td>1.93</td>
<td>4.71</td>
<td>N/A</td>
<td>N/A</td>
<td>4.09</td>
<td>7.43</td>
</tr>
<tr>
<td>14</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>6.04</td>
<td>N/A</td>
<td>4.44</td>
<td>7.95</td>
</tr>
<tr>
<td>17</td>
<td>4.92</td>
<td>6.64</td>
<td>3.08</td>
<td>3.87</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3.84</td>
<td>7.93</td>
</tr>
<tr>
<td>20</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3.94</td>
<td>7.83</td>
</tr>
</tbody>
</table>

\(^d\) – note each average of models 1, 6, 13, and 14 is based on E314 active site RMSD extension.

5.4. Conclusions

A series of molecular dynamics simulations were performed to gain insights into the substrate (i.e., GSA, 4-OH-GSA, NAD\(^\text{+}\), GLU) active site binding of P5CDH in both reactant and product complexes. Moreover, the WT models were compared to several mutant enzymes including S352L, S352A, and E314A.

In the holoenzyme RC, the GSA substrate was observed to anchor to Glu165 carboxyl via its $\alpha$-NH\(^3\), yet it was not positioned in the oxyanion hole. When compared to the Michaelis complex of the RC, it appears that the aldehyde carbonyl $O_{\text{carb}}$ is positioned closer though still not within H-bonding distance. It is only when the active site Cys348 is deprotonated (with Glu314 modeled as a base) that the oxyanion hole member, $-\text{NH}$–
backbone of Cys348, is interacting moderately at an average of 2.21 Å for O_{\text{carb}}•••HN_{\text{Cys348}}.

Our MD results also indicate that the average structure of the active complex (i.e., Cys348–S{ }^-) shows the C_{\text{pyr}}•••S_{\text{C348}} to be 3.65 Å with a planar nicotinamide ring. Along with several other interactions such as the average \( r(S_{\text{C348}}•••H\tilde{N}_{\text{N211}}) = 2.19 \) Å and \( r(S_{\text{C348}}•••HN_{\text{pyr}}) = 2.09 \) Å, these stabilize the negative charge of the Cys348 sulfur atom. Consequently, this conformation appears to be dominant at equilibrium, particularly since Cys348–S{ }^- is key for the catalysis to occur.

It was also shown that upon GLU removal from the PC, the oxyanion hole collapses, placing Asn211 amide and Cys348 –NH– backbone away from the catalytic site. Thus, this active site change could possibly be the enzyme's way of dealing with the entrance of GSA and NAD{ }^+ in the next round of the catalytic cycle.

In the active site overlay between the fully-bound RC with GSA and 4-OH-GSA, it was determined that the \( \gamma \)-hydroxyl moiety of GSA maintains a stable interaction via H-bonding with the Glu165 carboxyl at an average \( r(\text{HO}_{4\text{-OH-GSA}}•••\text{OOC}_{\text{E165}}) = 1.63 \) Å. This interaction seems to have no effect on its enzymatic binding and positions 4-OH-GSA analogously to the GSA-bound enzyme, making it a compelling model for catalysis.

When looking at the overlay of the average WT and S352L structures, the bulky leucyl (Leu352) side chain sterically pushes against Phe449 phenyl and breaks the H-bond interaction between the Glu447 R-group and 3'-OH moiety of the nicotinamide sugar. The absence of this interaction in the mutant results in a change of the NAD{ }^+ position and consequently the Cys348–SH position, possibly hindering catalytic activity of P5CDH.

All four S352L and S352A PC models 15, 16, 18, and 19 show comparable active site interactions and in the absence of NAD{ }^+(H) appears to remain unaffected catalytically.

Likewise to the S352L mutant, the glutamyl mutation (E314A) disrupts the H-bond interaction between the Glu447 side chain and 3'-OH moiety of nicotinamide ribose.
Subsequently, the lack of this interaction in both enzyme mutants (i.e., E314A and S352L) results in NAD$^+$ misplacement and an inactive enzyme.

5.5. References


Chapter 5 — Insights into Substrate Binding and Mutagenesis of P5CDH


(39) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., Montréal, QC, Canada, 2015.


Chapter 6. Computational Studies on the Pre-Transfer Editing of Lysyl-tRNA Synthetase: Noncognate L-Ornityl Preference over Native L-Lysyl
6.1. Introduction

While enzymes are known to enable life, they generally have a single specific catalytic function.\textsuperscript{1,2} For instance, proteolytic enzymes such as trypsin are equipped with an active site, optimized for highly specific peptidase activity.\textsuperscript{3} Another example is methionine sulfoxide reductase, an antioxidant enzyme critical for the reduction of methionine sulfoxide to methionine.\textsuperscript{4} However, there are other systems that are capable of multi-tasking.\textsuperscript{5,6} One such group that has been gaining recognition over the years is the family of aminoacyl-tRNA synthetases (aaRSs).\textsuperscript{7,8} In particular, aaRSs are involved in several physiologically important processes such as protein synthesis,\textsuperscript{9} inflammation,\textsuperscript{10} apoptosis,\textsuperscript{11} and viral assembly.\textsuperscript{12} During protein translation two functions are carried out within the aminoacylation active site of these enzymes (Scheme 6.1).\textsuperscript{13-15} First, aaRSs match each of the 20 naturally-encoded amino acids with their cognate tRNA (tRNA\textsuperscript{aa}). This is followed by chemical activation, a two-step reaction as shown in Scheme 6.1A. Here, an amino acid, L-lysine, bound to lysyl-tRNA synthetase (LysRS) is first activated \textit{via} a displacement reaction, producing the corresponding aminoacyl-adenylate monophosphate (L-lysyl-AMP) and pyrophosphate (PP\textsubscript{i}). The second half-reaction occurs within the same active site, where the L-lysyl-AMP can be transferred and linked to tRNA\textsuperscript{Lys} (Scheme 6.1B). Specifically, in all eukaryotic LysRS, the transfer of the L-lysine occurs at the 3'-OH of the ribose of the 3'-terminal of tRNA adenylate 76 (characteristic of most class II aaRSs).\textsuperscript{16,17} However, in some bacteria and most Archaea, aminoacylation by LysRS happens at the 2'-OH of the ribose, a common feature of class I enzymes.\textsuperscript{16,18} Notably, LysRS is the single example of crossover between the two classes of aaRS enzymes.\textsuperscript{19}

aaRSs are highly conserved amongst eukaryotes and prokaryotes and they are known as ancient enzymes.\textsuperscript{20} Despite the early origin, their fidelity in protein synthesis is remarkable. For instance, LysRS has evolved to have outstanding discriminatory power between its cognate L-lysine (Lys) and the structurally similar, noncognate L-ornithine
(Orn) by a factor of more than $10^6$. Specifically, for every molecule of misacylated Orn-tRNA$^{\text{Lys}}$, more than one million Lys-tRNA$^{\text{Lys}}$ are correctly charged.

Scheme 6.1. Two-step reaction for the aminoacylation of tRNA$^{\text{Lys}}$ catalyzed by the aminoacyl-tRNA synthetase, LysRS: (A) L-lysine activation via a displacement reaction followed by (B) L-lysyl transfer to tRNA$^{\text{Lys}}$ biomolecule.

As such, these enzymes are involved in specific aminoacyl tRNA charging and thus they have developed several proofreading mechanisms to prevent its misacylation by noncognate amino acids. Specifically, aaRSs are equipped with either one or both pre- and post-transfer editing mechanisms. The pre-transfer editing approach involves removal of the noncognate aminoacyl (aa) prior to tRNA$^{\text{aa}}$ charging. On the other hand, in post-transfer editing, aa is removed from the mischarged tRNA$^{\text{aa}}$ complex. In the case of LysRS, it lacks a post-transfer editing mechanism. Thus high-accuracy in protein translation is solely dependent on the pre-transfer correcting pathway.
The editing mechanism for lysyl-tRNA synthetases against L-ornithine is proposed as per Scheme 6.2. In particular, L-ornityl is converted to L-ornityl-1,5-lactam via a self-cyclization. It is noted that LysU represents the heat-inducible isoform of *Escherichia coli* LysRS.

Similarly, within our group, several catalytic mechanisms done by Fortowsky *et al.* were proposed computationally based on the pre-transfer editing of MetRS against the highly toxic L-homocysteinyl-AMP. In particular, these results indicated that aside from the substrate-assisted aminoacyl mechanism, it is possible that editing may occur via a more favourable pathway where Asp259 acts as a base in the phosphoester cleavage, resulting in a lactam derivative likewise to that obtained in the scheme below.

**Scheme 6.2.** Proposed mechanism for the cyclization of mischarged Orn-AMP catalyzed by LysU.

Recently, Desogus *et al.* obtained multiple X-ray crystal structures of the *E. coli* lysyl-tRNA synthetase (LysU) in the presence of Lys, Lys-AMP intermediate, and ATP analogue (AMP-PCP). Based on their structural studies, they observed that the LysU active site is designed to facilitate the nucleophilic attack of the L-lysyl α-COO⁻ towards the ATP α-phosphate without the direct involvement of active site residues. Consequently, this leads to the formation of the chemically active L-lysyl (*i.e.*, Lys-AMP) and PPᵢ as previously shown in Scheme 6.1A.

In this study, the pre-transfer editing mechanism of L-lysyl-tRNA synthetase from *E. coli* (LysU) has been computationally investigated using a Lys-AMP-bound complex.
Several approaches such as docking, molecular dynamics (MD), and ONIOM(QM/MM) have been used to study substrate specificity of LysU against L-ornithine as well as its editing pathway. Additionally, MD was applied to also study the binding of Ser- and Cys-AMP, noncognate amino acids that are misacylated onto tRNA$^{Lys}$.

6.2. Computational Methods

The model preparation, energy minimizations, and average structure assessments were done using the Molecular Operating Environment (MOE) software. The NAMD program was used to carry out the molecular dynamics (MD) simulations.

6.2.1. MD Models

The X-ray crystal structure of LysU complexed with L-lysyl-adenylate intermediate (Lys-AMP) was used as a template to obtain four models (PDB ID: 1E1T). Specifically, all models were based on the same active site where the L-lysyl moiety was modified to another aminoacyl (aa) such as L-ornityl, L-cysteinyl, and L-seryl, forming a LysU⋯aa-AMP enzyme complex. The missing hydrogen atoms were added using the MOE default method and any uncry stallized residues were added on the basis of the amino acid sequence. The solvation of each resulting enzyme-substrate complex was done using a 6-Å layer of water molecules, surrounding each solute atom.

6.2.2. Energy Minimizations

The electrostatic and van der Waals potentials decayed smoothly using a damping functional factor. The AMBER12:EHT force field was used to optimize each solvated complex until the root mean square gradient fell below 0.42 kJ mol$^{-1}$ Å$^{-2}$. Prior to MD simulations, all residues, waters, and aa-AMP substrate within 36 Å of the AMP phosphorus atom were kept free, leaving all other atoms frozen at the positions that they ended up after the energy minimization to reduce computational costs.
6.2.3. MD Simulations

All MD simulations underwent thermal relaxation at constant pressure and temperature. The Nosé-Poincaré thermostat\textsuperscript{26} was coupled with the Hamiltonian equations of motion, where they were discretized using a 2-fs time step. For each model, a 0.10 ns equilibration was done at 150 K followed by a sampling phase until all the structures reached conformational convergence. In the LysU\textcdots Lys-AMP system this was particularly followed by a production run kept at 300 K for an interval of 13.35 ns. The other LysU simulations in complex with Orn-, Cys-, and Ser-AMP were run over 12.85, 13.85, and 13.60 ns, respectively. It should be noted that the LysU\textcdots Orn-AMP simulation did not reach a plateau until the final structure of the earlier trajectory was taken and equilibrated for another 0.10 ns at 150 K, followed by a production run of 10 ns. Based on RMSD and cluster analyses, an average structure from each aa-AMP trajectory was then optimized using the AMBER12:EHT force field. This MD protocol was successfully applied in other enzymes.\textsuperscript{27,28}

6.2.4. ONIOM(QM/MM) Calculations

Hybrid QM/MM calculations were applied using the ONIOM\textsuperscript{29-37} formalism in the Gaussian 09\textsuperscript{38} program. In particular, ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory was used to optimize geometries of Lys-AMP and Orn-AMP models and obtain harmonic vibrational frequencies to characterize the stationary points for each potential energy surface. The reactive region (QM-layer) for both models was described at the B3LYP/6-31G(d,p)\textsuperscript{39-41} level of theory while AMBER96\textsuperscript{42} force field was used to describe the surrounding protein environment (MM-layer). Single-point calculations were done at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory on the above optimized structures, in order to obtain the relative energies.

The two QM/MM models were derived from the final optimized AMBER12:EHT LysU\textcdots Lys-AMP structure (see MD Simulations above). All residues and waters up to
25 Å away from the AMP-phosphorus atom were included because the steric and electrostatic interactions of the protein environment surrounding the active site can affect the mechanism.\textsuperscript{43}

The QM-region of each model included the substrate (Lys- or Orn-AMP), partial R-groups of important residues surrounding the aminoacyl (Glu240, Lys244, Glu278, Glu428), and peptide backbones between Leu474–Gly475–Ile476 (C\textsubscript{474}–O–NH–\(\alpha\)-C\textsubscript{475}–CO–NH–\(\alpha\)-C\textsubscript{476}). Also, two active site water molecules were included along with the partial R-groups of Arg262 and Asn424 positioned within hydrogen bonding distance of the phosphate moiety (see Scheme 6.3 for QM-selected atoms in Lys-AMP model). The surrounding protein environment is composed of residues and waters included in the MM-layer. If three bonds away from the QM-layer, the \(\alpha\)-carbon of each residue in the MM-layer was kept fixed at its final MD optimized position. This computational method was applied to maintain the overall structure of the active site as shown in other enzymatic studies.\textsuperscript{44,45}

\textbf{Scheme 6.3.} QM-layer representation of the L-lysyl-bound active site model used in the ONIOM(QM/MM) calculations.
6.3. Results and Discussion

6.3.1. Equilibrated Michaelis Complexes

Initially, all four aa-AMP models (Orn-, Lys-, Cys-, and Ser-AMP) of LysU were equilibrated using MD simulations long enough to ensure conformational convergence as discussed in the Computational Methods section. Each simulation was described using RMSDs in the positions of nearby active site residues, waters, and aa-AMP ligand. For example, LysU···Orn-AMP complex reached conformational equilibrium in the second MD run where surrounding active site RMSDs had a range of just $0.38 \pm 0.06\ \text{Å}$ (Figure 6.1A). Similarly, the remaining aminoacyl-bound enzyme complexes – Lys-, Cys-, and Ser-AMP – showed active site RMSDs with narrow ranges of $0.66 \pm 0.12$, $0.93 \pm 0.14$, and $0.79 \pm 0.13\ \text{Å}$, respectively (Figure 6.1B-D).

![Graphs showing RMSDs](image)

**Figure 6.1.** Surrounding active site RMSDs of the average LysU Michaelis complexes including (A) Orn-, (B) Lys-, (C) Cys-, and (D) Ser-AMP as substrates.
6.3.2. Aminoacyl Binding

As mentioned in the Introduction, LysU is able to discriminate between its cognate Lys-AMP substrate and noncognate Orn-AMP. However, other aminoacyls such as cysteinyl and seryl evade this editing process and are mischarged onto tRNA$^{\text{Lys}}$. As a result, using MD, we examined the active site binding of all these noncognate amino acids against the native L-lysyl. In an overlay as shown in Figure 6.2, the average structures of all aminoacyl substrates were superimposed. It is noted that each ribose ring and adenine nucleobase exhibit similar positions within the LysU active site. However, the side chains of the aminoacyl substrates (i.e., R–NH$_3^+$, R–SH, R–OH) do show great conformational variation. Specifically, the phosphoester bond of the L-cysteinyl ligand is a complete mirror image of the others, possibly a factor in the inability of this aaRS active site to proofread against it as experimentally observed by Jakubowski.\textsuperscript{22} Almost similarly, L-seryl bends down away from the ester bonds of Orn- and Lys-AMP. Another interesting characteristic of binding is the linear positioning of both Orn and Lys moieties. This indicates that the L-ornityl self-cyclization is dependent on the active site in order to be 'bent' or oriented for proper nucleophilic attack of the aminoacyl's carbonyl carbon ($C_{\text{carb}}$).

**Figure 6.2.** Overlaid average structures of aa- = Orn-, Lys-, Cys-, and Ser-AMP within the LysU active site. Atom colour code: H, white; C, gray/maroon/citrine/black; N, blue; O, red; S, yellow; P, pink.
Another aspect of this MD study was to determine whether an active site carboxylate is indeed available to act as a catalytic base in the editing mechanism. As such Glu421 is the equivalent residue to Asp259 in MetRS, however its position indicates Glu421 is not suitable for catalysis as a base in the self-cyclization mechanism (Table 6.1). In the case of Orn-AMP, Orn-N⋯O\textsubscript{phos} lies closer by 4.59 Å at 7.21 Å in contrast to Orn-N⋯O\textsubscript{E421} which is 11.80 Å apart. Also, O\textsubscript{E421}⋯C\textsubscript{carb} distance is not close as it is 1.58 Å longer than Orn-N⋯C\textsubscript{carb}. Analogously, the other aminoacyl models show similar results, suggesting Glu421 may not play a role in the editing process of LysU. This would not be uncommon since the aaRS ancestral origin has led many of these enzymes to diversify their active site and hence possibly portray different proofreading mechanisms.

**Table 6.1.** Selected average distances (Angstroms) considered in the MD study when each noncognate Orn-, Ser-, Cys-, and cognate Lys-AMP are bound in the active site of LysU.

<table>
<thead>
<tr>
<th>Model</th>
<th>Average Distances (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aaX*⋯O\textsubscript{phos}</td>
</tr>
<tr>
<td>Orn-AMP</td>
<td>7.21</td>
</tr>
<tr>
<td>Ser-AMP</td>
<td>4.85</td>
</tr>
<tr>
<td>Cys-AMP</td>
<td>6.17</td>
</tr>
<tr>
<td>Lys-AMP</td>
<td>7.54</td>
</tr>
</tbody>
</table>

X* = N, O, or S heteroatom in the R-group of the respective aminoacyl.

In Figure 6.3, several active site residues, responsible for substrate binding, are overlaid for both Lys- and Orn-AMP. It is indeed observed that their positioning within the active site is very similar with an RMSD of 0.82 Å. For instance, several electrostatic interactions are consistent in each aminoacyl complex, where e.g., R–COO\textsuperscript{−} of Glu278 and Glu428 form salt bridges with the aminoacyl's α-NH\textsubscript{3}\textsuperscript{+} and R–NH\textsubscript{3}\textsuperscript{+} groups, respectively. These observations further emphasize that not only these aminoacyls are structurally similar but also their interactive behavior within the active site is
outstandingly alike. Consequently, the catalytic power of LysU to mechanistically select and edit the misactivated L-ornityl against its native L-lysyl remains to be established by further studies using QM/MM models as discussed later on in this section.

**Figure 6.3.** Overlaid average structures of aa- = Orn- and Lys-AMP within the LysU active site. Atom colour code: H, white; C, gray/maroon; N, blue; O, red; P, pink.

### 6.3.3. *Pre-transfer Editing of Orn-AMP versus Lys-AMP*

As stated in the Introduction, pre-transfer editing mechanisms of aaRSs are generally believed to work by converting the incorrect aminoacyl to a lactam derivative, thereby preventing tRNA mischarge (*i.e.*, Orn-tRNA$^{\text{Lys}}$). Several other aspects can be seen with respect to the binding interactions of the aa-AMP substrates in the reactant complexes of Figure 6.4. One set of interactions responsible for substrate positioning includes the hydrogen bonding of the α-NH$_2$ of L-ornityl with Glu278 and Glu240 carboxyl groups at 1.66 and 1.92 Å, correspondingly (Figure 6.4A). The amine tail (R–NH$_3^+$) of the aminoacyl in Orn$^{\text{RC}}$ has been deprotonated by Glu428–COO$^-$, forming a nucleophilic R–NH$_2$ side chain with a H-bonding distance of 1.64 Å. In addition, the non-bridging oxygen atoms of α-phosphate are stabilized *via* multiple H-bonds involving Asn424 and
Arg262 side chains, as well as two water molecules (see Figure 6.4A). Similarly, these H-bonding interactions involving the phosphate are observed in $^{\text{Lys}}\text{RC}$ of Figure 6.4B. However, the L-lysyl's amine tail remains protonated unlike that of L-ornityl. This indicates that the Orn-AMP interactions with the active site residues (i.e., $\text{E428}^{\text{COOH}} \cdots \text{N}^{\text{orn}}$, $\text{K244}^{\text{N}} \cdots \text{HOOC}_{\text{E240}}$) are not neutralized simply due to the gas phase optimization. For instance, Lys-AMP $\text{R}^{\text{–NH}_3^+}$ electrostatically interacts with both Glu428 and Glu240 side chains due to closer and more stable H-bonds at 1.55 and 2.14 Å, respectively. Together these results suggest that the editing of Orn-AMP is favoured and facilitated by the active site residues, thus possibly being a factor in the discrimination between the two aminoacyl substrates.

![Figure 6.4](image)

**Figure 6.4.** Optimized geometries of the reactant complexes ($^{\text{aa}}\text{RC}$) of (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Note $^{\text{aa}}\text{-AMP}^*$ indicates that its ribose and nucleobase are not shown for simplicity. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

However, it is noted that both $^{\text{aa}}\text{-AMP}$ substrates ($\text{aa} = \text{Orn}, \text{Lys}$) from both MD (Figure 6.3) and optimized QM/MM reactant complexes ($^{\text{aa}}\text{RC}$) of LysU (Figure 6.4) are not properly oriented for nucleophilic attack at the aminoacyl's sp$^2$ carbon carbonyl, $\text{C}_{\text{carb}}$. For instance, both $^{\text{Orn}}\text{N} \cdots \text{C}_{\text{carb}}$ and $^{\text{Lys}}\text{N} \cdots \text{C}_{\text{carb}}$ in the $^{\text{aa}}\text{RC}$ are positioned at distances of 6.37 and 6.72 Å, respectively, making the intramolecular attack unlikely (Figure 6.4).
Consequently, a conformational change should occur within the active site to permit the substrate to 'bend' and facilitate the attack. In particular, the linear aa R-group curls within the active site as shown in Figure 6.5.

**Figure 6.5.** Optimized geometries of the \(^{aa}TS1^*\) transition states and \(^{aa}IC1\) intermediates for (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Note aa-AMP* indicates that its ribose and nucleobase are not shown for simplicity. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

Using a number of detailed systematic scans of the potential energy surface (PES), the TS curling (\(^{Orn}TS1^*\)) occurs with an upper limit of 82.0 kJ mol\(^{-1}\) (Figure 6.6). This approximate approach was successfully applied in our previous enzymatic work done on OCD,\(^{46}\) where L-ornithine is the substrate and a similar TS value was obtained for its curling step. It is noted that in \(^{Orn}RC\), the \(^{Orn}N\cdots C_{\text{carb}}\) distance has decreased from 6.37 Å to 2.54 Å in \(^{Orn}IC1\) (Figure 6.5B). This indicates the desired \(^{Orn}N\cdots C_{\text{carb}}\) bond has not yet formed, but it is now in a favourable nucleophilic position for the \(C_{\text{carb}}\) attack.
Also it is observed in both \textsuperscript{Orn}TS1\textsuperscript{*} and \textsuperscript{Orn}IC1, one of the non-bridging phosphate oxygen atoms has lost its H-bonding contact with the Arg262 guanidine group. Moreover, it is now in contact with the \(\alpha\)-NH\(_2\) group of Orn-AMP at distances of 2.82 and 2.46 Å for \textsuperscript{Orn}TS1\textsuperscript{*} and \textsuperscript{Orn}IC1, respectively. In Figure 6.5B, we can see the Lys-AMP amine tail, now neutral, curls in towards the C\textsubscript{carb} as shown in both \textsuperscript{Lys}TS1\textsuperscript{*} and \textsuperscript{Lys}IC1, where \(r(\textsuperscript{Lys}\text{N}…\text{C}\text{carb})\) is 3.29 and 2.72 Å, correspondingly.

![Figure 6.6. Potential energy surface (kJ mol\(^{-1}\)) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)/ONIOM(B3LYP/6-31G(d,p):AMBER96) for the editing mechanism of Orn-AMP by LysU.](image)

The approximate barrier for the curling step \((\textsuperscript{Lys}TS1\textsuperscript{*})\) is 91.5 kJ mol\(^{-1}\), meaning it is 9.5 kJ mol\(^{-1}\) higher in energy than the 'bending' of the L-ornityl R-group (see Figures 6.6 and 6.7). On the other hand, \textsuperscript{Lys}IC1 lies 19.5 kJ mol\(^{-1}\) lower in energy at 45.5 kJ mol\(^{-1}\) (Figure 6.7). The slightly more favourable \textsuperscript{Lys}IC1 can be partially attributed to the
additional stabilizing H-bond towards O\textsubscript{phos} at a distance of 2.26 Å, an interaction that is absent in the case of Orn-AMP.

**Figure 6.7.** Potential energy surface (kJ mol\(^{-1}\)) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)/ONIOM(B3LYP/6-31G(d,p):AMBER96) for the editing mechanism of Lys-AMP by LysU.

When looking at the \(^{\text{aa}}\text{RC}\) distances between the amine nitrogen of the aa R-group and the phosphate oxygen atom (\(^{\text{aa}}\text{N} \cdots \text{O} \text{phos}\)) from the MD study (c.f. Table 6.1), it can be seen that both decreased significantly to 4.18 and 2.93 Å in \(^{\text{Lys}}\text{IC1}\) and \(^{\text{Orn}}\text{IC1}\), respectively (Table 6.2). Consequently, both \(^{\text{aa}}\text{O} \text{phos}\) atoms present themselves as possible candidates in aiding the intramolecular transfer of the R-group amine proton transfer. Meanwhile, when considering whether Glu421 plays any catalytic role in the aminoacyl editing by LysU, it becomes clear that its positioning in the active site is not valuable towards catalysis unlike that seen in the case of MetRS\(^{23}\) (Table 6.2).
Next, in the final step of the editing process, the amine tail of the aminoacyl (\(\text{aa}^N\)) can readily attack the sp\(^2\) center C\(_{\text{carb}}\) in an S\(_N\)2 fashion as shown in Figure 6.8. Specifically, in \(\text{Orn}^\text{TS2*}\) the \(\text{Orn}^N\cdots\text{C}_{\text{carb}}\) distance has shortened significantly from 2.54 Å to 1.79 Å while the bond between the bridging phosphate oxygen (O'\(_{\text{phos}}\)) and C\(_{\text{carb}}\) has elongated by 0.18 Å to 1.58 Å (Figures 6.5A and 6.8A).

Table 6.2. Selected distances (Angstroms) considered in the QM/MM study when Orn- and Lys-AMP are curled within the active site of LysU.

<table>
<thead>
<tr>
<th>Model</th>
<th>Distances (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\text{aa}^N\cdots\text{O}_{\text{phos}})</td>
</tr>
<tr>
<td>(\text{Orn}) IC1</td>
<td>4.18</td>
</tr>
<tr>
<td>(\text{Lys}) IC1</td>
<td>2.93</td>
</tr>
</tbody>
</table>

This cyclization step occurs with a reaction barrier of 36.2 kJ mol\(^{-1}\) (\(\text{Orn}^\text{TS2*}\)), thus making it the rate-determining step of the editing mechanism (Figure 6.6). It is noted that although this scan is a fixed optimization, the frequency calculation confirmed indeed that it is a transition structure. For instance, it was indicated that O'\(_{\text{phos}}\), 2.40 Å away from \(\text{Orn}^\text{NH}\), is likely to act as the base to abstract one of the \(\text{Orn}^N\) protons. It would appear that this proton transfer occurs late in the transition state. Early changes that occur within the active site also indicate the cleavage of the O'\(_{\text{phos}}\)–C\(_{\text{carb}}\) is favoured; a stronger H-bonding interaction (0.41 Å closer) exists between the aminoacyl's \(\alpha\)-NH\(_2\) and one of the phosphate oxygen atoms.

In light of the suggested catalytic base residue present in MetRS,\(^{23}\) we also considered Asn424 amide group which appears to be in an appropriate catalytic position with \(\text{Orn}^N\cdots\text{N}_{\text{N424}},\ \text{N}_{\text{N424}}\cdots\text{O}'_{\text{phos}}\) distances of 2.18 and 2.28 Å, respectively. However, despite our efforts to include Arg412 in the QM-layer to stabilize the Asn424 amide and lower the pK\(_a\) to promote its use as a proton shuttle,\(^{47}\) no convincing results were determined. Consequently, this reinforces the idea that although aaRSs are part of the
same family, their active sites are designed and have adapted differently, operating via different editing mechanisms.\(^{19}\)

**Figure 6.8.** Optimized geometries of the \(^{\text{ aa }}\)TS\(^{2*}\) transition states and \(^{\text{ aa }}\)PC product complexes for (A) Orn- and (B) Lys-AMP in the LysU editing mechanism with selected distances shown (Angstroms). Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

Nonetheless, this reaction step results in the favourable production of the L-ornithine lactam derivative and AMP byproduct (\(^{\text{ Orn}}\)PC), lying at an energy of \(-92.9 \text{ kJ mol}^{-1}\) relatively to \(^{\text{ Orn}}\)RC (Figure 6.6). As shown in Figure 6.8A, the amine proton of Orn-aminoacyl is transferred onto a phosphate oxygen atom, where HO'\(_{\text{phos}}\) forms an H-bond of 1.97 Å with the bridging amine (\(^{\text{ OmN}}\)). In addition, \(^{\text{ OmN}}\)–C\(_{\text{carb}}\) bond is now formed (1.41 Å), while the aminoacyl's O'\(_{\text{phos}}\)–C\(_{\text{carb}}\) bond is cleaved.
Looking at the Lys-AMP QM/MM model, Lys-aa cyclization occurs via $\text{Lys}^{\text{TS2*}}$ in Figure 6.7 with a large barrier of 94.7 kJ mol$^{-1}$. Similarly to the Orn-aa pathway, $\text{Lys}^{\text{TS2*}}$ is the rate-determining step, however this leads to a non-favourable product complex, $\text{Lys}^{\text{PC}}$, lying endothermically at a much higher relative energy of 21.6 kJ mol$^{-1}$ with respect to $\text{Lys}^{\text{RC}}$ or 114.5 kJ mol$^{-1}$ higher in energy than $\text{Orn}^{\text{PC}}$ (see Figures 6.6 and 6.7). In Figure 6.8B, analogous to the Orn-lactam, the amine proton of Lys-aminoacyl is transferred onto the phosphate oxygen, now forming an H-bond of 2.17 Å, longer by 0.2 Å in contrast to that seen in $\text{Orn}^{\text{PC}}$. The $\text{Lys}^{\text{N}}$–$\text{C}_{\text{carb}}$ bond is now 1.38 Å, slightly shorter by 0.03 Å than the cyclic Orn derivative. Another structural difference in the Orn-AMP product complex is that the active site Arg262 guanidine group is participating in H-bonding with both of its amino moieties towards stabilizing the negative charge on the phosphate oxygen. This active site feature is absent in $\text{Lys}^{\text{PC}}$, thus possibly contributing to its disfavour.

Moreover, the overlay of the two product complexes, $\text{Orn}^{\text{PC}}$ and $\text{Lys}^{\text{PC}}$ in Figure 6.9, clearly shows another perspective as to why the energy difference between the two structures is so high. The L-ornithine-1,5-lactam forms the favourable chair conformer (six-membered cyclic structure), whereas the L-lysine-1,6-lactam is in the much less favourable boat conformation (seven-membered cyclic structure), where the steric and angle strains are greater. Together these results show that the selectivity power of LysU for the L-ornityl aminoacyl trumps that of L-lysyl in the pre-transfer editing mechanism.
6.4. Conclusions

In this study we have considered the pre-transfer editing mechanism of lysyl-tRNA synthetase, LysU, using a multi-scale approach. Specifically, we have applied docking, molecular dynamics (MD), and hybrid quantum mechanics/molecular mechanics (QM/MM) methods to examine the self-cyclization of a noncognate aminoacyl, L-ornityl, against the LysU-native aminoacyl, L-lysyl. Also, we have gained insight into the binding of noncognate Cys- and Ser-AMP substrates within the LysU active site, possibly revealing why they evade cyclization and hence the reason why these noncognate aminoacyls are mischarged onto tRNA\textsuperscript{Lys}.

The self-cyclization mechanism of LysU, where Orn-AMP is a substrate, resulted in the thermodynamically favourable formation of L-ornithine lactam product. The \textsuperscript{Orn}PC \textsuperscript{PC} lies at $-92.9$ kJ mol\textsuperscript{-1} relative to \textsuperscript{Orn}RC. The reaction barrier for this step (\textsuperscript{Orn}TS\textsuperscript{2*}) is rate-determining at 36.2 kJ mol\textsuperscript{-1}, while also making it an enzymatically feasible editing process. On the other hand, the substrate-assisted editing of Lys-AMP resulted in an endothermic product complex with a relative energy value of 21.6 kJ mol\textsuperscript{-1} with respect to \textsuperscript{Lys}RC. Moreover, although \textsuperscript{Lys}TS\textsuperscript{2*} is similarly rate-determining, its energy barrier is 94.7 kJ mol\textsuperscript{-1} with respect to \textsuperscript{Lys}IC\textsuperscript{1}. Consequently, these results indicate that the pre-
transfer editing of Orn-AMP is much more favourable than the self-cyclization of the
cognate Lys-AMP, suggesting misaminoacylation by LysU of its tRNA$^{\text{Lys}}$ with Orn is
highly unlikely.

6.5. References

(24) *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., Montréal, QC, Canada, **2015**.


Chapter 6 — Computational Studies on the Pre-Transfer Editing of LysU


Chapter 7. A Molecular Dynamics (MD) and Quantum Mechanics/Molecular Mechanics (QM/MM) Study on Ornithine Cyclodeaminase (OCD): A Tale of Two Iminiums
Chapter 7 — An MD and QM/MM Study on OCD

7.1. Introduction

L-proline (Pro) is one of the naturally occurring 20 genetically encoded amino acids and is unique amongst them in being the only one with a secondary α-amine. The effect of its structure and functionality upon the proteins in which it is found has long been studied. In addition, however, it is also known to have a number of key physiologically important roles. For example, it has been shown to be an important antioxidant needed by microorganisms, plants, and animals.\(^1\)\(^-\)\(^8\) Furthermore, in certain pathogenic bacteria, it has been suggested that proline metabolism plays a role in enabling the pathogen to survive under harsh conditions.\(^9\)\(^-\)\(^11\) Plants also depend on proline production for cell wall biosynthesis, while mediating abiotic and biotic cell stresses.\(^12\)\(^,\)\(^13\) In mammals, the interconversion between proline and Δ\(^1\)-pyrroline-5-carboxylate (P5C), an intermediate in its biosynthesis and catabolism, is believed to be involved in cell apoptosis.\(^1,\)\(^14\)\(^-\)\(^18\) Importantly, this cycle provides a redox shuttle between the cytosol and mitochondria, controlling the formation of reactive oxygen species.\(^14\)\(^,\)\(^19\)\(^,\)\(^20\) Consequently, it is important to understand the functions of the enzymes that metabolize proline.

In general, Pro is synthesized within cells and organisms via multi-enzymatic pathways from either glutamate (Glu) or arginine (Arg).\(^9\)\(^,\)\(^14\)\(^,\)\(^21\)\(^-\)\(^24\) In the former, the enzymes involved require the use of γ-glutamyl kinase and glutamate-γ-semialdehyde dehydrogenase, which require the cofactors ATP and NADH or NADPH, respectively. In contrast, the latter pathway utilizes the metalloenzyme arginase and the pyridoxal-5'-phosphate (PLP)-dependent enzyme, ornithine δ-aminotransferase. Both pathways, however, lead to the formation of a common intermediate, glutamic-semialdehyde. This then undergoes a non-enzymatic cyclization via an intramolecular condensation reaction to give the imine, P5C.\(^14\)\(^,\)\(^25\) Then, the common enzyme P5C reductase (P5CR) reduces the latter cyclic intermediate to form Pro.\(^9\)\(^,\)\(^26\) In mammals, the former pathway in which glutamate is converted to proline is believed to be the major metabolic route.
More recently, however, several bacteria have been shown to be able to utilize an unusual enzymatic route for proline biosynthesis. More specifically, they use the NAD$^+$-dependent non-metalloenzyme ornithine cyclodeaminase (OCD) to directly convert the amino acid L-ornithine (Orn), itself also an intermediate along the above 'Arg-pathway', to proline. Thus, in contrast to the two general pathways discussed above in which the actual cyclization step is done without enzymatic participation, the mechanism of OCD includes the cyclization of a linear intermediate to a cyclic product (Scheme 7.1). Furthermore, it not only produces L-proline stereospecifically and without additional enzymes being involved, but does so in an irreversible fashion.

\[ \text{OCD} \quad \text{L-ornithine} \quad \text{NAD}^+ \quad \text{L-proline} + \text{NH}_4^+ \]

**Scheme 7.1.** Overall reaction for conversion of L-ornithine to L-proline as catalyzed by ornithine cyclodeaminase (OCD).

It is generally believed that the overall mechanism of OCD begins with an initial hydride transfer from the Orn substrate's C$_\alpha$–H moiety onto the C$_4$ center of the nicotinamide ring of the NAD$^+$ cofactor. This results in the formation of an iminium (C$_\alpha$=NH$_2^+$ containing) intermediate complex. Unfortunately, however, it is then unclear how the mechanism proceeds. Based on the results of experimental studies including X-ray crystallographic structures and mass spectroscopic data, two possible pathways have been proposed (Scheme 7.2). While they share some common features such as involving formation of a Schiff base, a chemically and biochemically important reaction process that has been previously studied in detail both experimentally and computationally, they also have some very important differences. In particular, in the hydrolytic pathway, an active site water nucleophilically attacks the C$_\alpha$ center of the
iminium intermediate, resulting in loss of ammonium with formation of the keto acid, 5-amino-2-oxopentanoate (Scheme 7.2A).\textsuperscript{28,34} This then undergoes a cyclization reaction in which the δ-amine nucleophilically attacks the carbonyl carbon to form a cyclic $\Delta^1$-pyrroline-2-carboxylate (P2C) species. This latter intermediate has previously been shown to be involved in other biochemical processes including, for example, lysine catabolism. It is further noted that this ring formation step can be considered analogous to the cyclization of glutamic γ-semialdehyde in which an amine (α-amine) reacts with a carbonyl carbon center. This step is then followed by a hydride transfer from the NADH moiety onto the C$_\alpha$ position (C2) of P2C to give L-proline.

Scheme 7.2. Proposed (A) hydrolytic and (B) non-hydrolytic pathways for the conversion of L-ornithine to L-proline as catalyzed by the enzyme ornithine cyclodeaminase.
In contrast, in the alternate proposed mechanism known as the non-hydrolytic pathway, and shown above in Scheme 7.2B, the δ-amino tail of the iminium intermediate is neutralized by a suitable active site residue. Consequently, it is now able to nucleophilically attack the $C_\alpha$ center of the iminium intermediate to form a cyclic 2-aminoproline species. This then undergoes loss of the $\alpha$-amino group as ammonia or ammonium to give P2C. The final step is as in the hydrolytic pathway; hydride transfer from the NADH moiety onto the $C_\alpha$ position (C2) of P2C to give the final product.

As noted above, the enzyme OCD itself has only been found in a few select bacteria. However, it has been noted that it shows close phylogenetic resemblance to some crystallin enzymes, in particular those found in mammals. These proteins and enzymes are primarily known for their role in lens and cornea structure and transparency. However, their malfunctioning has been linked to a number of diseases including cataract formation and cancer. Indeed, OCD is a member of the $\mu$-crystallin family of enzymes which have been found to be abundant in the eye lens of marsupials and are believed to have a similar chemical mechanism to that of OCD. Thus, a clearer elucidation of the mechanism of OCD can provide invaluable insights not only into its catalytic abilities and proline biosynthesis, but also into related physiologically important enzymes.

Computational chemistry is a proven invaluable tool for the study of enzymatic mechanisms. Thus, we have complementarily applied molecular dynamics (MD) simulations and an ONIOM(QM/MM) approach to investigate the catalytic mechanism of OCD. In particular, we have examined the feasibility of both the proposed hydrolytic and non-hydrolytic pathways for conversion of L-ornithine to L-proline.

### 7.2. Computational Methods

#### 7.2.1. Molecular Dynamics (MD) Equilibration

The Molecular Operating Environment (MOE) program was used for model preparation and the molecular dynamics (MD) simulations. For the model, the structure
of OCD used was taken from an X-ray crystallographic structure of an OCD homo-
dimer; each active site was complexed with L-ornithine (Orn) and NADH (PDB ID: 1X7D). While the protein is homo-dimeric, we have chosen to use a single monomer for the MD simulations. This model simplification is reasonable since the catalytic site does not include interface residues. Thus, a single protomeric enzyme-substrate-cofactor complex was selected while the other was removed. In addition, all selenomethionine (Se-Met) residues were mutated to the native methionines. The bound NADH cofactor was oxidized to give the catalytically required NAD$^+$ (i.e., a hydride was removed from the C$_4$ center of NADH). The coordinates of missing hydrogens were added using the MOE default method. A 7-Å spherical layer of water molecules was then added to solvate the enzyme-substrate complex. In order to force the system to lie within the volume of space defined by the surrounding shell of waters, an ellipsoidal potential wall with a scaling constant of 2 was placed around the solvated enzyme-substrate complex. To allow the electrostatic and van der Waals potentials to decay smoothly, a damping functional factor was included. The geometry of each solvated complex was then optimized using the CHARMM22 force field until the root mean square gradient of the total energy fell below 0.21 kJ mol$^{-1}$ Å$^{-2}$. The MD simulations were performed under constrained pressure and temperature. The equations of motion were coupled with the Nosé-Poincaré thermostat and the time step for numerical integration was set to 2 fs. Initially, the system was heated from 150 to 300 K for a period of 50 ps, followed by an equilibration period of 100 ps at the constant temperature of 300 K and pressure of 1 atm. Based on RMSD and cluster analyses, a representative structure was chosen from the trajectory. This structure was then optimized with the CHARMM22 force field and used to obtain a suitable enzyme complex for further investigation (see below). We have successfully applied this MD protocol in the study of other enzymatic systems.
Chapter 7 — An MD and QM/MM Study on OCD

7.2.2. QM/MM Computations

The Gaussian 09 suite of programs\textsuperscript{56} was used for all QM/MM calculations using the ONIOM formalism.\textsuperscript{57–65} Optimized structures were obtained at the ONIOM (B3LYP/6-31G(d):AMBER96) level of theory.\textsuperscript{66–69} Specifically, the AMBER96 molecular mechanics force field was used to describe the low-layer (\textit{i.e.}, protein environment surrounding the active center) while the high-layer (\textit{i.e.}, the active center) was described using the density functional theory-based method B3LYP in conjunction with the 6-31G(d) basis set. Harmonic vibrational frequencies were calculated at this same level of theory in order to obtain the corresponding Gibbs free energy corrections at SATP (\(\Delta E_{\text{Gibbs}}\)) in addition to characterize the nature of the stationary points on the potential energy surface (PES) (\textit{i.e.}, as energy minima or first-order transition states (TSs)). Single-point energy calculations on the above optimized structures were performed at the ONIOM (B3LYP/6-311+G(2df,p):AMBER96//B3LYP/6-31G(d):AMBER96) + \(\Delta E_{\text{Gibbs}}\) level of theory within a mechanical embedding (ME) formalism.

The final CHARMM22 optimized MD structure (see above) was used to obtain a suitable chemical model for the fully bound active site for use in the QM/MM-based mechanistic studies. More specifically, all residues and waters up to 15 Å from the L-ornithine substrate were extracted to be used as the enzyme-substrate-cofactor model. This distance was chosen as it has been previously shown that the steric and electrostatic effects arising from the protein environment surrounding the active site within this distance can have important effects on the mechanism and its intermediates and transition states.\textsuperscript{70}

The QM-region included the L-ornithine substrate and the active site groups with which it directly interacts or are experimentally known or proposed to be involved in the mechanism. Namely, the R-groups of Arg112, Lys69, Glu56 and Asp228 were included along with two active site waters observed within the MD simulations. The NAD\textsuperscript{+} cofactor was also included in part, the asterisk (*) indicating that only its nicotinamide
and ribose ring was within the QM-region. The rest was replaced and modeled by a methyl group as shown in Figure 7.1. All those residues or waters shown in the outer-circle in Figure 7.1b were included within the MM-layer in their entirety. The only exceptions being those in which the R-group was included within the QM region. To ensure the integrity of the active site model during calculations, the α-carbons of each residue as well as that of the capping methyl in NAD⁺ were held fixed at their final CHARMM22 optimized positions (see above). This computational approach has been widely used previously, and successfully applied in studies on related enzymatic systems.  

**Figure 7.1.** Illustration of the ONIOM (QM/MM) fully bound active site model used: (a) the low or molecular mechanics (MM)-layer is shown in wire while the high or quantum mechanics (QM)-layer is shown in tube format; (b) those moieties within the inner circle were included within the QM-region, while those within the outer circle represent the residues contained in the MM-layer of the QM/MM model. Atom colour code: H, white; C, gray; N, blue; O, red.
7.3. Results and Discussion

7.3.1. Structure of the Active Site with the Substrate L-Ornithine and NAD\(^+\) Cofactor Bound

We began this study by obtaining and examining the optimized structure of the fully bound active site complex (RC). That is, the complex in which both the NAD\(^+\) cofactor and L-ornithine (Orn) substrate are bound and which is shown in Figure 7.2a. As it can be seen, the Orn is bound electrostatically via interactions of each of its charged groups with active site residues. In particular, the $\alpha$-COO\(^-\) forms short, strong hydrogen bonds with the protonated R-groups of both Arg112 and Lys69 with distances of 1.73 and 1.53 Å, respectively. Meanwhile, its $\alpha$-NH\(_3\)+ group is strongly hydrogen bonded with the nearby R-group carboxylate of the active site aspartyl (Asp228) with an $r(\alpha$-NH$\cdots$OOC–Asp228) distance of 1.62 Å. Notably, this results in a modest lengthening in the participating N$_{\alpha}$–H bond to 1.08 Å. For the R-group, $\delta$-NH$_3^+$, however, it was found that it transferred a proton to the R-group carboxylate of the active site glutamyl residue (Glu56) with which it hydrogen bonds. Although they now form a neutral N$_{\delta}$⋯HOOC–Glu56 hydrogen bond, its length is still decidedly short with a length of only 1.68 Å.

Importantly, in this bound conformation, the distance between the substrate's C$_{\alpha}$ and the NAD\(^+\) cofactor's C$_4$ center to which a hydride would be transferred, \textit{i.e.}, $r(C_{\alpha}\cdots C_4(\text{NAD}^+))$, is 3.98 Å. It is noted that while in the X-ray crystal structure (PDB ID: 1X7D), the enzyme was co-crystallized with the substrate and NADH, the above calculated distance is in reasonable agreement with the experimentally measured distance of 3.8 Å. It is noted that the mechanistically important C$_4\cdots$H–C$_{\alpha}$ distance is 3.51 Å and that the C$_{\alpha}$–H bond has lengthened marginally to 1.10 Å.

In addition, it is noted that in the MD structure of the initial reactant complex RC and shown in Figure 7.2b, three waters (W1, W2, and W3) were observed to lie in the active site and interact with the substrate. More specifically, one (W1) forms a hydrogen bonding bridge between the $\delta$-NH$_3^+$ group and a sugar hydroxyl of the NAD\(^+\) cofactor. A
second (W2) is simply hydrogen bonded to the substrate's δ-NH$_3^+$ group while a third (W3) forms a hydrogen bonding bridge between the anionic R-group carboxylate of Glu56 and the substrate's protonated α-NH$_3^+$ moiety. It should be noted that the oxygen of this latter water, O$_{W3}$, is 3.50 Å away from the C$_\alpha$ center. These same interactions were also in agreement with the subsequent QM/MM optimized structure (not shown).

**Figure 7.2.** (a) Optimized structure (see Computational Methods) of the initial fully bound active site complex RC with selected bond lengths shown (in Angstroms). (b) Solvated MD structure of the active site. Atom colour code: H, white; C, gray; N, blue; O, red.

### 7.3.2. Orn Oxidation and Formation of 2-Aminoproline (AP)

The first catalytic step of OCD involves a hydride transfer from Orn to NAD$^+$ via TS1 (Figure 7.3). Specifically, the hydride is transferred from the C$_\alpha$–H moiety of the Orn substrate onto the C$_4$ center of the nicotinamide ring of NAD$^+$. This process occurs via TS1 with a relatively high barrier of 90.6 kJ mol$^{-1}$ with respect to RC (Figure 7.4). This may reflect in part the lack of any stabilizing interactions between the active site and the hydride entity as well as the large structural changes predicted. For example, as it can be seen in Figure 7.3, in TS1 a significant reduction in the C$_\alpha$⋯C$_4$(NAD$^+$) distance is seen. Specifically, it has shortened by 1.24 Å to 2.74 Å. As a result, the hydride is almost equidistant between the C$_\alpha$ and C$_4$ centers with C$_\alpha$⋯H$^-$ and C$_4$⋯H$^-$ distances of 1.48 and
1.28 Å, respectively. However, concomitant with the H⁻ transfer, a proton from α-NH₃⁺ group has transferred to the R-group carboxylate of Asp228 (Figure 7.3). Indeed, in TS1, the proton has fully transferred to the carboxylate as indicated by an r(H–O₂Asp228) distance of 0.98 Å while the Cα–N bond length has already significantly shortened by 0.16 Å to 1.35 Å. This suggests that the proton transfer occurs early in this reaction step.

Figure 7.3. Optimized structures (see Computational Methods) with selected bond lengths shown (in Angstroms) for the ornithine cyclodeaminase (OCD) catalytic mechanism prior to deamination as L-ornithine is converted to 2-aminoproline. Atom colour code: H, white; C, gray; N, blue; O, red.

Complete H⁻ transfer onto the NAD⁺ cofactor results in formation of the iminium ion intermediate (IC1) which lies lower in energy than RC by −15.3 kJ mol⁻¹ (Figure 7.4). In IC1, the Cα⋯C₄(NADH) distance has lengthened significantly to 4.51 Å and is in fact now even greater than that seen in RC (Figure 7.3). Furthermore, the Cα–N bond has shortened even further to just 1.29 Å indicating the formation of a Schiff base. In
addition, the hydrogen bond interaction between the α-NH$_2$ and now neutral Asp228–COOH group has been weakened with an r(Asp228–COOH⋯N$_\alpha$) distance of 4.14 Å.

**Figure 7.4.** Potential energy surface (PES) obtained (see Computational Methods) of the OCD catalytic mechanism prior to deamination as L-ornithine is converted to 2-aminoproline.

As detailed in the Introduction section, it has been proposed that subsequent to hydride transfer and formation of the Schiff base intermediate, cyclization of Orn may occur *via* one of two possible pathways. Namely, it may either involve a first reaction with an active water to give the keto acid, 5-amino-2-oxopentanoate (A2O), or direct nucleophilic attack of the δ-NH$_2$ nitrogen at the C$_\alpha$ center. Indeed, the former A2O intermediate is similar to the glutamate γ-semialdehyde, common in both Arg- and Glu-pathways for proline biosynthesis. In particular, the ketone functionality is exposed to nucleophilic attack by either the α-amine in the non-enzymatic reaction or by δ-amine in OCD. As
noted previously for \textbf{RC}, the nearest active site water that may nucleophilically attack the \( C_\alpha \) center lies approximately 3.5 Å from the \( C_\alpha \) center. We did attempt to elucidate a possible reaction pathway for reaction of \textbf{IC1} with a water molecule but were unable to do so. Analysis of the structure of \textbf{IC1}, however, suggests that in agreement with Goodman \textit{et al.},\textsuperscript{31} the steric presence of Arg45 packing against the substrate in \textbf{IC1} prevents the water from approaching much closer. Indeed, the nearest distance between Arg45 and the \( C_\alpha/N_\alpha \) centers is only 3.40 Å. As a result, the hydrolytic pathway would seem unlikely to be physically feasible and hence is not being discussed further herein. Thus, the remainder of the discussion solely concerns the non-hydrolytic pathway unless otherwise mentioned. However, it is interesting to note that an iminium ion simulates a Lewis acid-activated carbonyl functional group.\textsuperscript{71} In such Lewis acid activation, a lowering of the LUMO energy of the C=O \( \pi \)-system occurs promoting nucleophilic attack. Thus, it appears that OCD has optimized proline synthesis by maintaining the existence of a far more reactive intermediate such as an iminium ion rather than a ketone derivative.

With the iminium intermediate \textbf{IC1} formed, it is now susceptible to nucleophilic attack by \( N_\delta \) of Orn. However, presumably prior to attack, the 'tail' is required to curl within the active site. It is noted that the \( \delta\text{-NH}_2 \) group in \textbf{IC1}, unlike say the \( \alpha\text{-COO}^- \) group, forms just one hydrogen bond interaction with an active site residue and that being with the now neutral R-group of Glu56. This would be expected to facilitate the required cleavage of this interaction and subsequent curling of the 'tail' of \textbf{IC1}. Unfortunately, at the present level of theory we were unable to exactly optimize a TS for this process. Instead, a number of detailed systematic scans of the PES for such a curling was performed in order to determine an upper limit for the energy required for this step. In particular, it was found that a decrease in the \( \angle C_\alpha-C_\beta-C_\gamma-C_\delta \) by 55.4° to 128.5° gave a structure (\textbf{TS2}) that corresponded to an energy maximum of 55.2 kJ mol\(^{-1}\) with respect to \textbf{RC}, which is 70.5 kJ mol\(^{-1}\) with respect to \textbf{IC1}. It is noted that in \textbf{TS2}, the \( N_\delta\cdots C_\alpha \)}
distance has decreased from 4.51 Å in IC1 to 3.92 Å. Thus, clearly the desired Cα–Nδ bond has not yet formed. Interestingly, it should be noted that all attempts to optimize energy minima complexes with ∠Cα–Cβ–Cγ–Cδ less than 128.5° (without constraining the Cα–Cβ–Cγ–Cδ dihedral angle) led directly to the formation of the very low energy 2-aminoproline intermediate complex IC2 lying 84.2 kJ mol⁻¹ lower in energy than RC. Importantly, in IC2, a Cα–Nδ bond has been formed with a length of 1.45 Å, typical for a single C–N bond. This suggests that in OCD, the active site structure allows for a reasonably low energy curling of the 'tail' of the Schiff base intermediate and that once suitably positioned, the Nδ center appears able to nucleophilically attack at the Cα center readily. Thus, as stated above, given that OCD has protected the iminium intermediate from reacting with water, it has allowed for an apparently barrierless C–N bond formation without the need for activation by a Lewis acid. This is far different than what has been seen in various carbonyl analogues where significant barriers for initial C–N bond formation exist.³⁸–⁴⁷

7.3.3. Deamination of 2-Aminoproline (AP)

In order to lose the α-amine of IC2 (i.e., deamination), it must first be protonated. This is analogous to the required protonation of the hydroxide to form water prior to Schiff base formation in the aldehyde analogues.³⁸,⁴⁰,⁴¹ One possibility is that an intramolecular proton transfer may occur from the ring –NH₂⁺– group directly onto the leaving amine. However, this would necessarily involve a four-membered ring transition structure which is well-known to be higher in energy due to the inherent strain involved.⁷² Alternatively, an active site residue may be able to act as a proton donor. Within the active site of OCD, however, the nearest acidic residue is the neutralized R-group carboxylic acid of Asp228. Within the present computational model in the formation of IC2 (i.e., during the preceding cyclization step), it was found that the Asp228–COOH group was no longer in a position to transfer a proton to the leaving α-amine (Figure 7). Thus, the Asp228–
COOH⋯α-NH₂ distance must first shorten allowing, for the formation of a hydrogen bond to the leaving α-amine. However, for this to occur IC₂ must undergo a conformational or hydrogen bonding network change. In particular, the proton on Glu56 must rotate allowing for a rearrangement of the active site waters allowing for the Asp228–COOH⋯α-NH₂ distance to be reduced.

The process occurs via TS₃ with a barrier of 50.6 kJ mol⁻¹ with respect to IC₂ to give the alternate complex IC₂' (Figure 7.4). The latter complex, while lying 25.4 kJ mol⁻¹ higher in energy than IC₂, still lies markedly lower in energy than RC by 58.8 kJ mol⁻¹. As expected, this rotation of the proton on Glu56–COOH group caused Asp228–COOH to hydrogen bond with the α-amino group of the substrate. Furthermore, the hydrogen bonding network between the Glu56–COOH group and Asp228–COOH involving the two waters also changed. Specifically, the Glu56–COOH moiety now acts as a hydrogen bond donor via the waters to the carboxylic group of Asp228.

However, within the present computational model we were unable to locate a transition structure for proton transfer from Asp228–COOH onto the leaving amine. This is likely due to the fact that this would result in several charged groups (Asp228–COO⁻, α-NH₃⁺, and −NH₂⁻), all being in close proximity and thus highly sensitive to the degree of inclusion of the electrostatics of the protein environment. While it is plausible that under experimental conditions Asp228–COOH may transfer its proton onto the α-NH₂ group, previous computational investigations on Schiff base formation reactions have shown that such proton transfers generally occur via intramolecular proton transfer from the bridging amine; furthermore such proton transfers are facilitated by a water or other suitable group to prevent the formation of a four-membered TS.³⁸,⁴⁰,⁴¹

Thus, we examined the possibility of a water molecule aiding proton transfer from either Asp228–COOH or the bridging −NH₂⁻ to the α-NH₂ moiety. Indeed, once the cyclic pyrrolidine species is formed in IC₂' it is noted that there is now solvent-accessible space in the vicinity of the Asp228–COOH group. The resulting IC₂' complex
in which a water was added is hereafter denoted as $\text{IC}2'\cdots\text{H}_2\text{O}$ (Figure 7.5). The resulting PES obtained for the subsequent proton transfers is shown in Figure 7.6a. Importantly, in $\text{IC}2'\cdots\text{H}_2\text{O}$, a water simultaneously acts as a hydrogen bond acceptor via its oxygen with the Asp228–COOH group and pyrrolidine's $–\text{NH}_2^+$ moiety, while also acting as a hydrogen bond donor to the $\alpha$-NH$_2$ group of the pyrrolidine and the 2'-OH group of the NAD$^+$ cofactor's ribose. Indeed, in the optimized structure of $\text{IC}2'\cdots\text{H}_2\text{O}$, $r(\text{Asp228–COOH}\cdots\text{OH}_2)$ and $r(\alpha\text{N}^\cdots\text{H}_2\text{O})$ were found to be 1.74 and 1.77 Å, respectively, while for the ring amine $r(\text{NH}_2^+\cdots\text{OH}_2)$ is slightly longer at 1.95 Å (Figure 7.5).

![Optimized structures](image)

**Figure 7.5.** Optimized structures (see Computational Methods) with selected bond lengths (in Angstroms) of the water-assisted intermediate complex ($\text{IC}2'\cdots\text{H}_2\text{O}$), TS4, cyclic intermediate IC3, and Schiff base intermediate IC4. Atom colour code: H, white; C, gray; N, blue; O, red.
Interestingly, rather than a proton transfer from Asp228–COOH, it was found that the 2-aminoproline itself underwent an intramolecular proton transfer from its ring –NH₂⁺– moiety onto the α-NH₂ group via the bridging water molecule. This process occurred via the six-membered ring transition structure, TS₄, at a very low cost of only 26.0 kJ mol⁻¹ with respect to IC₂'⋯H₂O (Figure 7.6a). The resulting intermediate IC₃ formed lies only 13.3 kJ mol⁻¹ higher in energy that IC₂'⋯H₂O (Figure 7.6a). Importantly, the protonation of the α-amine results in a significant lengthening in the Cα–Nα bond in IC₃ by 0.09 Å to 1.54 Å and a shortening of the Cα–Nδ bond (0.10 Å) within the pyrrolidine. Such bond changes are likely to aid in deamination and Schiff base formation.

In previous computational studies on Schiff base formation involving a ketone/aldehyde, the loss of the leaving group –OH₂ has been shown to occur in two steps.³⁸ First the adjacent bridging amine undergoes an inversion thus allowing for favourable overlap of its lone pair and the anti-bonding orbital of C–OH₂ bond to be cleaved. The second step is then cleavage of the C–OH₂ bond itself. For the 2-amino-2-carboxy-pyrrolidine, no stable intermediate corresponding to inversion of the ring –NH– was obtained nor was a concerted TS involving cleavage of the Cα–N bond with inversion of the bridging amine. However, it is noted that, while in previous studies, the barrier for loss of the water for the analogous carbonyl systems was found to be generally rate-limiting,³⁸,³⁹,⁴¹,⁴² cleavage of the Cα–NH₃⁺ bond in OCD was found to be exothermic. In particular, with IC₄ lying 34.9 kJ mol⁻¹ lower in energy than IC₃ or –21.6 kJ mol⁻¹ relative to IC₂'⋯H₂O (Figure 7.6a). Thus, there is clearly a driving force for formation of this Schiff base intermediate. C–N bond cleavage with concomitant Schiff base formation results in the Δ¹-pyrroline-2-carboxylate (P2C) containing complex IC₄. It should also be noted that in IC₄, the Asp228–COOH proton has also transferred onto the leaving NH₃ to give an ammonium ion, NH₄⁺ via H₂O. The latter is then free to leave the active site.
7.3.4. Hydride Transfer from NADH onto P2C to Give the Final Product L-Proline

For the final stage, we considered formation of the L-proline product from P2C after loss of the cleaved \( \text{NH}_4^+ \) from the active site. Only minor structural changes were noticed in the resulting complex \( \text{IC4}'–\text{NH}_4^+ \). In principle, L-proline can be formed via hydride transfer from the NADH moiety on the C\(_2\) (what will become C\(_\alpha\)) of P2C. However, in \( \text{IC4}'–\text{NH}_4^+ \), the key \( \text{NADH}C_4\cdots C_2 \) is quite long at 4.00 Å. Furthermore, the P2C itself is not ideally positioned for the transfer. However, it is able to undergo an intramolecular rotation about its C\(_2\)–COO\(^-\) bond, i.e., a change in its \( \angle\text{O}_1–\text{C}_1–\text{C}_\alpha–\text{N}_\alpha \) dihedral angle, where the oxygen involved is the one hydrogen bonded to the Lys69. This process occurs via \( \text{TS6} \) at a markedly low cost of only 14.4 kJ mol\(^{-1}\) with respect to \( \text{IC4}'–\text{NH}_4^+ \) suggesting that the rotation is likely reasonably unhindered (Figure 7.6b). It is noted that in \( \text{TS6} \) the \( \angle\text{O}_1–\text{C}_1–\text{C}_\alpha–\text{N}_\alpha \) has decreased by 43.0° to 93.8°. This rotation is also exergonic with the resulting alternate conformer complex \( \text{IC4}'' \) being lower in energy than \( \text{IC4}'–\text{NH}_4^+ \) by 19.1 kJ mol\(^{-1}\). Importantly, as a result of this reorientation of P2C within the active site, its C\(_2\) center is now more suitably positioned for hydride transfer and the \( \text{NADH}C_4\cdots C_2 \) distance has decreased significantly to 3.14 Å (Figure 7.7).
Figure 7.6. PESs obtained (see Computational Methods) for (a) water-assisted deamination of 2-aminoproline to give $\Delta^1$-pyrroline-2-carboxylate (P2C) and, (b) reduction of P2C to give L-proline.

Reduction of P2C via a hydride transfer from $\text{NADH}C_4-H$ onto its C2 center can then occur via TS7 with a barrier of 65.2 kJ mol$^{-1}$ with respect to IC4'' (Figure 7.6b). This is notably lower than the cost of the initial hydride transfer from the Orn substrate to NAD$^+$. This final step is also exergonic with the final product complex (PC) in which the L-proline is bound within the active site being a further 8.0 kJ mol$^{-1}$ lower in energy than IC4'', and with a relative free energy of $-27.1$ kJ mol$^{-1}$ with regards to IC4$'–\text{NH}_4^+$ (Figure 7.6b).
Figure 7.7. Optimized structures (see Computational Methods) with selected bond lengths shown (in Angstroms) of the Δ₁-pyrroline-2-carboxylate intermediate IC₄'', TS₇, and L-proline product (PC). Atom colour code: H, white; C, gray; N, blue; O, red.

7.4. Conclusions

Using a combination of MD and QM/MM methods, the catalytic mechanism of OCD has been investigated. From the results it was found that the initial hydride transfer from the C₄α–H group of the L-ornithine substrate to the C₄ center of the NAD⁺ cofactor with concomitant formation of a Schiff base, is the rate-limiting step. In particular, this process occurred with a relative free energy barrier of 90.6 kJ mol⁻¹. Experimentally, this hydride transfer step might be examined by deuterating H–(C₄α) of Orn and conducting a kinetic isotope effect study.

For the remaining steps of the mechanism, while two pathways have been proposed, it appears that the enzyme most likely operates via a non-hydrolytic pathway. In particular, the MD and QM/MM results suggest that water is sterically hindered from attacking C₄α.
after the initial Schiff base formation. Given that iminium ions are generally more reactive it seems that the active site of OCD evolved to guarantee such an intermediate exists by preventing its reaction with water. Indeed, following a conformational change of the substrate within the active site, a barrierless C–N bond formation occurred. This is considerably different than seen in previous investigations of various carbonyl analogues where significant barriers to C–N bond formation exists. While a transition state for deamination could not be found, the overall process was found to be thermodynamically favourable. Importantly, with deamination, a second Schiff base was formed. Like the initial C–N bond formation, this Schiff base would likely become more reactive. Indeed, the final step in the reaction was a H⁻ transfer with a low barrier of 65.2 kJ mol⁻¹. The resulting product, L-proline, being thermodynamically more favoured than the preceding intermediate.

Thus, while the typical biosynthetic pathway for L-proline from arginine requires two enzymes and non-enzymatic cyclization of a glutamate γ-semialdehyde intermediate, this cyclic intermediate being common to the pathway involving glutamate, is later reduced to form Pro. In contrast, ornithine cyclodeaminase appears to first exploit the formation of a highly reactive C₆=NH₂⁺–containing iminium ion in order to enable cyclization and ultimately formation of a second iminium ion, Δ¹-pyrroline-2-carboxylate, formed after loss of the α-NH₃ group. The latter ion is readily reduced via a hydride transfer from the NADH cofactor onto its 'C₆-center', resulting in the formation of L-proline.

7.5. References

Chapter 7 — An MD and QM/MM Study on OCD

Chapter 7 — An MD and QM/MM Study on OCD


Chapter 7 — An MD and QM/MM Study on OCD


(53) Molecular Operating Environment (MOE), 2010.10; Chemical Computing Group Inc.: Montréal, QC, Canada, 2012.


Chapter 8. The Effects of Structural Variation on the Enzymatic Mechanism of Pseudouridine-5’-Monophosphate (ΨMP) Glycosidase using Molecular Dynamics and Quantum Chemical Cluster Approaches
Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase

8.1. Introduction

Carbohydrates are an important class of biomolecules, whose metabolic actions are critical for cell survival when stressful conditions arise.\(^1\) For instance, excess glucose can be converted to polysaccharide forms for storage of energy.\(^4\) In mammals, glucose can be stored in the liver and muscle as glycogen,\(^5\) whereas plants are able to convert it to starch.\(^6\) Aside from energy storage capability, carbohydrates can serve as structural elements, some of which include chitin in the exoskeleton of arthropods\(^7\) and cellulose in the cell wall of fungi.\(^8\) More importantly, 2'-deoxyribose and ribose are the only sugar components of nucleic acids, forming the backbones of DNA and RNA, respectively.\(^9\)

DNA-cleaving enzymes such as N-glycosidases\(^10,11\) are capable of hydrolyzing sugar–nucleobase bonds (C–N glycosidic bonds). This group of enzymes plays an active role in DNA repair, excising nucleobases that have been damaged by oxidation, alkylation, or deamination.\(^12\) On the other hand, some nucleic acid-cleaving enzymes cut C–C glycosidic bonds and they are known as C-glycosidases.\(^15,16\) An example of such enzyme is pseudouridine-5'-monophosphate (ΨMP) glycosidase,\(^17\) where the C–C glycosyl isomer of uridine-5'-monophosphate, ΨMP, is hydrolyzed into uracil and ribose-5'-phosphate (R5P) as shown in Scheme 8.1.

![Scheme 8.1](image_url)

**Scheme 8.1.** The conversion of ΨMP to uracil and ribose-5'-phosphate by ΨMP glycosidase.
ΨMP glycosidase was reported by Huang et al.\textsuperscript{17} to be the first C-glycosidase that was mechanistically characterized using structural, mutagenic, and kinetic studies. They obtained several X-ray crystal structures of this glycosidase from \textit{Escherichia coli} including the unbound enzyme, ΨMP-bound mutant (K166A), ΨMP-, and R5P-adduct intermediate complexes. Note that both the enzyme mutant and ΨMP-adduct were derived from the incubation of R5P and uracil with the glycosidase, whereas the R5P-adduct, covalently attached to Lys166 side chain, was the result of single addition of R5P to the enzyme. Together, these four snapshots of the reaction pathway along with the mass spectrometry studies, confirmed the ring-opening of the ribose moiety and the presence of a lysine covalently bound to the substrate. Thus, it was suggested ΨMP glycosidase catalysis occurs \textit{via} an unforeseen C-glycosyl bond hydrolysis by a retroaldol-type mechanism unlike most glycosidases that typically rely on a dissociative mechanism for C–N and C–O glycosidic bond cleavage.\textsuperscript{17-19}

As shown in Scheme 8.2, the novel mechanism proposed by Huang et al.\textsuperscript{17} for the ΨMP glycosidase, is a five-step process including ring-opening, conjugate addition, C–C glycosidic bond cleavage, imine hydrolysis, and cyclization. In the first chemical step, the ring-opening of ΨMP ribose leads to formation of an alkene bond between the nucleobase and the sugar moiety (\textit{i.e.}, C1'=C5). Here, the ring oxygen atom (O4') is protonated \textit{via} acid/base catalysis possibly conducted by water-mediated protonation/deprotonation. The next reaction could involve the nucleophilic conjugate addition where the neutral Lys166–NH\textsubscript{2} attacks the unsaturated C1'=C5 bond at the C1' centre with the C5 becoming protonated, possibly by a Glu31 carboxyl group. The C1'–C5 glycosidic bond can now be cleaved in a retroaldol-type reaction, forming a Schiff base intermediate and a uracil anion. Subsequently, the imine hydrolysis and cyclization leads to enzyme regeneration and the release of R5P sugar.

This present study examines the effects that starting structures have on the mechanism of ΨMP glycosidase in \textit{Escherichia coli} using several computational approaches. For
example, molecular dynamics (MD) along with a QM cluster approach have been used to investigate each reaction coordinate developed via X-ray crystallography to determine whether each starting point makes a viable model for catalysis and leads to the same enzymatic pathway. In addition, substrate binding in the C-glycosidase and the function of key active site residues such as Asp31, Asp149, and Lys93 were elucidated.

Scheme 8.2. Proposed mechanism catalyzed by ΨMP glycosidase for conversion of ΨMP to uracil and ribose-5′-phosphate.17
8.2. Computational Methods

The Molecular Operating Environment (MOE)\textsuperscript{20} and NAMD\textsuperscript{21} programs were used for developing the molecular models and performing the molecular dynamics (MD) simulations, respectively.

8.2.1. MD Simulations

All four models were prepared from the X-ray crystal structures of \textit{E. coli} ΨMP glycosidase done by Huang \textit{et al.}\textsuperscript{17} including the apoenzyme-····SO$_4^{2-}$ (PDB ID: 4GIJ), ΨMP····K166A mutant (PDB ID: 4GIM), ring-opened ribose ΨMP covalent adduct (PDB ID: 4GIL), and ring-opened R5P adduct (PDB ID: 4GIK).

The overall structure of ΨMP glycosidase is a homotrimer where the Mn(II) metal binding site includes interface residues. Thus, instead of using a monomer, each crystal structure was simplified by selecting the substrate and expanding it 25 Å away, assuring the presence of residues from all interacting chains. In all models, the missing hydrogen atoms were added using the MOE default method.

In total, four models were generated, one for each available crystal structure of the glycosidase enzyme as shown in Table 8.1. For instance, model 1 was generated based on the sulfate-bound apoenzyme, where the anion was replaced with the ΨMP ligand. Its binding interactions were based on ΨMP-bound K166A complex (PDB ID: 4GIM). In model 2, the 'K166A' structure was mutated back to the native form, resulting in A166K reactant complex. For models 3 and 4, no active site structural modifications were done as they possibly represent pathway intermediates. However, note that any missing residues due to uncrystallization were added based on the amino acid sequence of the \textit{E. coli} glycosidase enzyme.

All crystallographic waters were removed, followed by the solvation of the resulting enzyme-substrate complexes using a 4-Å spherical layer of water molecules. These solvated complexes were forced to lie within the volume created by an ellipsoidal
potential wall with a scaling constant of 2. Using a damping functional factor, the electrostatic and van der Waals potentials decayed smoothly. The PFROSST force field was used to minimize the geometry of each solvated complex such that the root mean square gradient of the total energy fell below 0.42 kJ mol\(^{-1}\) Å\(^{-2}\). Once optimized, only those residues and water molecules that lie beyond the second environmental sphere were fixed in order to reduce computational costs, whereas all other residues were set free.

Table 8.1. Summary of enzyme-substrate models considered in this study.

<table>
<thead>
<tr>
<th>MD Model</th>
<th>QM Model</th>
<th>PDB Origin</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>'DOCK'</td>
<td>4GIJ</td>
<td>ΨMP Docking</td>
</tr>
<tr>
<td>2</td>
<td>N/A</td>
<td>4GIM</td>
<td>A166K</td>
</tr>
<tr>
<td>3</td>
<td>'ΨMP'</td>
<td>4GIL</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>'R5P'</td>
<td>4GIK</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The subsequent structures underwent thermal relaxation at constant pressure and temperature using molecular dynamics. The Nosé-Poincaré thermostat\(^{22}\) was coupled with the equations of motion and the numerical integration was set up with a 2-fs time step. Each system was equilibrated at 150 K over a 0.1 ns period and then set to 300 K for an interval of 10 ns. Using RMSD and cluster analyses, a representative structure was taken from each trajectory and optimized using the PFROSST force field. As shown herein, this MD protocol was applied successfully in other enzymatic studies.\(^{23-25}\)

8.2.2. Quantum Mechanical (QM) Cluster Calculations

The Gaussian 09\(^{26}\) program suite was used for all QM cluster calculations. More specifically, the hybrid exchange-correlation B3LYP functional\(^{27-29}\) was applied to the native QM-large active site clusters. The optimized geometries and their harmonic vibrational frequencies used to characterize the intermediates of the catalytic pathway, were computed \textit{via} 6-31G(d) basis set. On the above optimized geometries, single-point calculations at the B3LYP/6-311+G(2df,p) level of theory were done to obtain the
Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase

relative energies. Siegbahn and Himo\textsuperscript{30} have proposed that the effect of the protein environment on the enzymatic active site becomes negligible if the QM cluster model reaches a size of approximately 150–200 atoms. For each of our models, we simulated the protein solvation effects by applying the IEFPCM (integral equation formalism polarizable continuum model) method,\textsuperscript{31-34} with a dielectric constant (\(\varepsilon\)) of 4.0.\textsuperscript{30, 35}

QM cluster models were truncated from the final optimized PFROSST structures of the ΨMP-docked reactant complex (model 1), ring-opened ribose ΨMP covalent adduct (model 3), and ring-opened R5P adduct (model 4) (see MD Simulations above). In particular, the quantum chemical cluster model for the reactant complex is shown in Scheme 8.3. Each model included a ligand corresponding to a conformer along the catalytic pathway (i.e., ΨMP substrate (1), ring-opened ribose ΨMP (3), and ring-opened R5P (4)). Note that the uracil nucleobase was added in model 4 to maintain atom consistency with the other native models. A manganese ion (Mn(II)) along with its first-shell residues Asp145, Glu176', Glu179' (latter two obtained from neighboring monomer), R-group of catalytically important Lys166 and other R-groups such as Glu31, Lys93, Thr130, Ser147, Asp149, Thr270, Asn289 were also included. Amino acid backbones that make up the substrate-binding pocket include Thr112, Val113, Ala114 (\(\alpha\)-\(\text{C}_{112}\)–\(\text{CO}\)–\(\text{NH}\)–\(\alpha\)-\(\text{C}_{113}\)–\(\text{CO}\)–\(\text{NH}\)–\(\alpha\)-\(\text{C}_{114}\)) and Thr130, Gly131, Gly132, Ile133 (\(\alpha\)-\(\text{C}_{130}\)–\(\text{CO}\)–\(\text{NH}\)–\(\alpha\)-\(\text{C}_{131}\)–\(\text{CO}\)–\(\text{NH}\)–\(\alpha\)-\(\text{C}_{132}\)–\(\text{CO}\)–\(\text{NH}\)–\(\alpha\)-\(\text{C}_{133}\)). The surrounding substrate environment is comprised of 13 water molecules. The amino acids were truncated either at the \(\alpha\), \(\beta\), \(\gamma\), or \(\epsilon\)-carbon and the hydrogen atom addition was done manually. Several carbon atoms were kept fixed at their final MD optimized position to preserve the integrity of the active site (see asterisks in Scheme 8.3). This computational method was successfully applied in studies on other enzymes.\textsuperscript{36, 37}
8.3. Results and Discussion

8.3.1. Equilibrated Starting Structures

Based on the available crystal structures of *E. coli* ΨMP glycosidase, all four models reached conformational convergence as shown in Figure 8.1A. In particular, it can be seen that ΨMP dock, ΨMP···A166K, ΨMP-, and R5P-adducts have RMSDs with narrow ranges of 0.37 ± 0.07, 0.55 ± 0.06, 0.43 ± 0.06, and 0.38 ± 0.07 Å, respectively. As expected, the mutant-based structure shows the highest RMSD range as Ala166 residue was modeled back to its native lysine (Lys166) before being submitted for the MD simulation. Moreover, in Figure 8.1A, it appears that the ΨMP···A166K simulation (green) shows the greatest fluctuation within the first 2 ns. This may be indicative that the long side chain of the lysyl residue is attempting to reach a stable conformation. Another interesting note is that the ring-opened ribose ΨMP-adduct shows the next highest fluctuation (red). This is not surprising since the X-ray crystal structure for this model
was solved at a lower resolution of 2.54 Å, where the average B factor for both protein and ligand are indeed the highest at 53.5 and 49.5 Å², correspondingly. These high values indicate that not only the resolution for the ΨMP-adduct (PDB ID: 4GIL) is poor but also the residue flexibility in this glycosidase complex is 12.1 Å² higher, relative to the other intermediate, *i.e.*, R5P-adduct (PDB ID: 4GIK). This shows the former to have a higher coordinate error, and thus one has to use this model with caution.

![Diagrams](image)

**Figure 8.1.** (A) Surrounding active site RMSDs of the ΨMP dock, ΨMP…A166K, ΨMP- and R5P-adducts during the course of the 10.1 ns MD simulation. Plots of distances with respect to time (ns) for the last 10 ns of the simulation of the four complexes are as follows: (B) Mn(II)…P, (C) K₁₆₆N⁶…C¹’, and (D) K₁₆₆N⁶–C¹’.

It has been suggested that the manganese metal plays a role in anchoring the ΨMP phosphate moiety during catalysis. As such we measured the distance between the phosphorus atom and Mn(II) ion to assess whether it remains consistent in each starting geometry (Figure 8.1B). It was noted that all models with the exception of 2 (ΨMP…A166K) lie within 0.17 Å of each other, ranging from 5.07 ± 0.18 to 5.24 ± 0.20...
Å apart. It was determined that \( r(\text{Mn(II)}\cdots\text{P}) = 3.16 \pm 0.13 \, \text{Å} \) in \( \Psi\text{MP}\cdots\text{A166K} \) model, suggesting that the original alanine mutation could have compromised the active site of the glycosidase enzyme. In addition, the \( K^{166}\text{N}^\varepsilon\cdots\text{Cl}' \) distance was used to compare the A166K structure (green) to the docked reactant complex (purple). Specifically, the measurements in Figure 8.1C show \( K^{166}\text{N}^\varepsilon\cdots\text{Cl}' \) to be \( 5.22 \pm 0.63 \) and \( 6.56 \pm 0.36 \, \text{Å} \) for models 1 and 2, respectively. In other words, the \( K^{166}\text{N}^\varepsilon\cdots\text{Cl}' \) distance in the \( \Psi\text{MP}\cdots\text{A166K} \) model is \( 1.34 \, \text{Å} \) longer, in contrast to the \( \Psi\text{MP} \) dock. This makes it unlikely for the amine tail of the Lys166 (R–NH\(_2\)) to perform a nucleophilic attack at the anomeric carbon of the ribose moiety (Cl'). Also note that the distance between the two atoms drops as low as \( 3.44 \, \text{Å} \) in the docked model, shortly after 2 ns in the simulation, where then it rises after 5 ns.

When looking at the intermediate structures of models 3 and 4, the average bond lengths of \( K^{166}\text{N}^\varepsilon\cdots\text{Cl}' \) during the MD simulation are \( 1.50 \pm 0.03 \) and \( 1.34 \pm 0.03 \, \text{Å} \), respectively (Figure 8.1D). The length of the latter bond indicates the formation of an iminium ion (Cl'\( =K^{166}\text{N}^\varepsilon\text{H}^+ \)) between the sugar moiety and the catalytic lysyl side chain (Lys166). This was also confirmed experimentally via mass spectrometry studies done by Huang et al.\(^{17}\) which proved that Lys166 is indeed the R5P-modified residue.

In the overlay between the average structure of \( \Psi\text{MP}\cdots\text{A166K} \) and the intermediate complexes (\( \Psi\text{MP}- \) and R5P-adducts), it is clearly seen that the binding site of the manganese metal ion is disordered in the mutant-derived model (Figure 8.2). In particular, Mn(II) is now nearly 2 Å closer to the phosphate group of \( \Psi\text{MP} \), losing its monodentate coordination to both carboxyl side chains of Asp145 and Glu176' ligands. Therefore, \( \Psi\text{MP}\cdots\text{A166K} \) starting structure was not considered further for QM-modeling.
**Figure 8.2.** Overlaid average structures of ΨMP···A166K with ΨMP- and R5P-adducts in the glycosidase active site. Atom colour code: H, white; C/Mn(II), green/yellow/gray; N, blue; O, red; P, pink.

### 8.3.2. Substrate Binding in the Three Reactant Complexes

As mentioned in the Computational Methods, three average MD structures were used and truncated to build the respective QM cluster models. In particular, Figure 8.3 shows the optimized reactant complexes of the ΨMP-bound active sites, $^{\text{DOCK}}\text{RC}$, $^{\text{ΨMP}}\text{RC}$, and $^{\text{R5P}}\text{RC}$.

Overall, most ΨMP-bound reactant complexes fall in good agreement with the experimentally developed X-ray crystal structures (*i.e.*, PDB IDs: 4GIL, 4GIK). For instance, Mn(II) metal maintains its octahedral coordination within all of our active site models with the exception of the ΨMP-adduct derived structure ("ΨMP"). In the latter, Mn(II) is bound to three ligands (*i.e.*, R-groups of Asp145, Glu176', and Glu179') and two water molecules, forming a trigonal bipyramidal-like geometry. In the other two $^{\text{X}}\text{RC}$ structures, the metal ion is bound to three water molecules with distances between 2.13 and 2.47 Å, as well as the R-group carboxyl of Asp145, Glu176', and Glu179' in a monodentate fashion (coordination distances ranging between 2.06–2.36 Å in all $^{\text{X}}\text{RC}$). It is noted that Glu176' and Glu179' side chains were originally in the second sphere of
Mn(II), where two other waters were ligated instead. Nevertheless, the two conserved residues from the neighboring chain are proximal and thus have the potential to displace two water molecules from the metal's first coordination sphere.

Figure 8.3. Optimized structures of the glycosidase reactant complexes ($^X$RC, where $X = \text{DOCK, } \Psi\text{MP, } \text{R5P}$), obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.
In addition, the octahedral geometry of Mn(II) is maintained in two of the three of our models. Note that although the coordination geometry of Mn(II) in the ΨMP-adduct may have changed by one ligand, it still makes a suitable study model since Mn(II)⋯P is 6.36 Å or 0.39 Å shorter than in the R5P-adduct model whose metal geometry is octahedral. Thus all models can be relied upon, especially since the role of the metal is likely structural and is not likely to affect the catalytic mechanism of C–C glycosyl bond hydrolysis.

Also, in Figure 8.3A-C it can be seen that within all reactant complexes, the R-group carboxyl of Asp31 forms a hydrogen bond to a water molecule which is in direct contact with the N3 atom of the pseudouracil moiety. This interaction is important as the nucleobase deprotonation by the Asp31 base would promote the ribose ring-opening of the ΨMP substrate and thus initiate catalysis. Another crucial interaction for the ring-opening process we observed in each \(^{X}RC\), is the presence of an acid to protonate the ribose ring oxygen (O4'). In the 'DOCK' reactant complex, the proton transfer could occur directly from Lys93 amine side chain as it resides only 2.20 Å away from the sugar ring oxygen. However, the H-bond between \(^{K93}N^\epsilon H\) and O4' is much weaker at 4.22 Å in \(^{ΨMP}RC\).

On the other hand, in the \(^{R5P}RC\) structure, a water molecule acts as a proton shuttle between the Lys93–NH\(^3\) and O6 with \(r(^{K93}N^\epsilon H⋯O_W) = 2.91\) Å and \(r(H_W⋯O6) = 2.13\) Å. These observations are in line with the K93A mutation done by Huang et al.,\(^{17}\) where its \(k_{cat}\) was 17-fold lower than its native enzyme, suggesting Lys93 is relatively replaceable and may not be very crucial. Nonetheless, it seems that Lys93 residue could act as an acid/base catalyst towards the glycosidic hydrolysis of this enzyme based on the arrangement of the active site.

It is worth mentioning that Lys166–NH\(^2\) was chosen to be neutral in our models since it appears that no base is present to activate the lysyl nucleophile. Moreover, this also falls in accordance with the mechanistic proposal of Huang and coworkers.\(^{17}\) As such the
K166N···C1' distances in Figure 8.3 indicate Lys166 amine N^e atom is in a reasonable position at 5.26, 4.99, and 4.73 Å for DOCKRC, ΨMPRC, and RSPRC, respectively.

8.3.3. Nucleobase Deprotonation

As stated earlier, Lys166 R-group is involved in the nucleophilic attack at the anomeric C1' carbon. Before this process can occur, the pseudouracil nucleobase must lose a proton to promote ring-opening of the ribose sugar as shown by DOCKIC1 in Figure 8.4A. In all of our active site models, we observe that Asp31 carboxyl group is involved in proton abstraction via the water bridged to H_N3, resulting in the anionic nucleobase. The negative charge on N3 from DOCKIC1 is stabilized by both a strong hydrogen bond interaction of r(H_W···N3) = 1.69 Å and electron delocalization across the pseudouracil ring (e.g., 0.02 Å elongation in the C2=O relative to DOCKRC). Consequently, this results in the formation of an endothermic intermediate, 34.6 kJ mol⁻¹ higher than DOCKRC (Table 8.2).

Similarly, in ΨMPIC1, a strong H-bond exists between the ionic N3 pseudouracil atom and water moiety that is bridged to the Asp31–COOH with a slightly elongated r(H_W···N3) = 1.80 Å (Figure 8.4B). In RSPIC1, this distance is marginally weaker at 1.83 Å (Figure 8.4C). However, in the latter two models, Ser147–OH is involved in the stabilization of the ionic ΨMP phosphate in contrast to 'DOCK' model, which lacks this favourable interaction.
Figure 8.4. Optimized structures of the glycosidase intermediate complexes: (A) DOCKICx, (B) \( \Psi \text{MP} \) ICx, and (C) \( \text{R5P} \) ICx, where \( x = 1, 2, 3 \), obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

As a result, both DOCKIC1 and \( \Psi \text{MP} \) IC1 have lower energy, \(-13.0\) and \(3.4\) kJ mol\(^{-1}\), correspondingly. Note that the ring-opened ribose \( \Psi \text{MP} \) derived model shows an exothermic \( \Psi \text{MP} \) RC \( \rightarrow \) \( \Psi \text{MP} \) IC1 step, possibly due to the slightly stronger H-bond interaction with the pseudouracil anion and C1’ being further away from the Lys166-NH\(_2\) group than in \( \text{R5P} \) IC1 by 0.5 Å with \( r(K166^N\cdots C1') = 5.21 \) Å, slightly relieving steric hindrance.
Table 8.2. Relative energies of QM-optimized complexes, obtained at B3LYP/6-31+G(2df,p)//B3LYP/6-31G(d) level of theory in gas phase (ε = 1.0).

<table>
<thead>
<tr>
<th>Model</th>
<th>RC</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>'DOCK'</td>
<td>0.0</td>
<td>34.6</td>
<td>107.4</td>
<td>100.8</td>
<td>163.6</td>
<td>51.3</td>
<td>31.4</td>
<td>−5.0</td>
</tr>
<tr>
<td>'ΨMP'</td>
<td>0.0</td>
<td>−13.0</td>
<td>181.8</td>
<td>127.3</td>
<td>160.8</td>
<td>145.1</td>
<td>32.5</td>
<td>36.8</td>
</tr>
<tr>
<td>'R5P'</td>
<td>0.0</td>
<td>3.4</td>
<td>62.5</td>
<td>−22.7</td>
<td>47.2</td>
<td>−46.2</td>
<td>−33.4</td>
<td>−72.2</td>
</tr>
</tbody>
</table>

8.3.4. Covalent Linkage between Substrate and Lys166–NH₂ (C1'–K166₄N₆H₂⁺)

The sugar ring oxygen is within 2.27 Å of Lys93–NH₃⁺ in DOCKIC1, thus possibly allowing for a direct proton transfer at O4' once the electron delocalization from the anionic N3 has resulted in the formation of an addition product. This step is essential as it involves the ring-opening of the substrate's ribose moiety. Subsequently, sp²-C1' centre can now undergo a nucleophilic attack by the Lys166–NH₂ (r(K₁₆₆₄N₆…C1') = 5.16 Å in DOCKIC1), forming the covalent adduct DOCKIC2 at an energy value of 107.4 kJ mol⁻¹ relative to DOCKRC. The K₁₆₆₄N₆–C1' bond in the newly formed covalent DOCKIC2 adduct is 1.51 Å, typical of a single bond, where Lys166–NH₂ in DOCKIC1 migrated 3.65 Å to perform the nucleophilic attack.

It is interesting to note that while both ΨMPIC1 and DOCKIC1 are set with the Lys93 side chain to interact with C6=O carbonyl oxygen directly, R5PIC1 shows a rather less direct H-bond interaction. That is, a water is seen to bridge between the O6 nucleobase atom and the amine R-group of Lys93 with r(K₉₃₄N₆…O₆) = 2.96 Å and r(H₆N₆…O₆) = 2.06 Å. Likewise to DOCKIC2, the subsequent intermediates of the docked and Schiff base derived structures show endothermic relative energies at 181.8 and 62.5 kJ mol⁻¹, respectively to each reactant complex. The high-energy values for the two former models are most likely due to the poor stabilization of the anionic charge in the phosphate and pseudouracil moieties. For instance, as depicted in Figure 8.4A-B, only one water is H-bonded to the ionic N3 atom in DOCKIC2 while in the ΨMPIC2 no water is seen to be in contact. On the other hand, the 'R5P' model not only makes two strong hydrogen bonds
to stabilize the anionic intermediate but its active site allows the cationic charge to be distributed across an H-bond chain towards a phosphate oxygen (see Figure 8.4C). This active site arrangement partially explains why the 44.9–119.3 kJ mol\(^{-1}\) raise in energies are observed in DOCKIC2 and \(^\Psi\)MPIC2 models relative to RSPIC2.

8.3.5. Concurrent Lys166 Deprotonation and Lys93 Protonation

The positive charge on the Lys166 amine is stabilized via an H-bonding network connected to the negatively charged non-bridging phosphate (O\(_{\text{phos}}\)). Specifically, in DOCKIC2 this chain is composed of K\(_{166}\)N\(_{\varepsilon}\)H…O4', O4'H…K\(_{93}\)N\(_{\varepsilon}\), and K\(_{93}\)N\(_{\varepsilon}\)…O\(_{\text{phos}}\) with distances of 2.34, 1.67, and 2.41 Å, respectively. As such, the former two proton donor-acceptors participate in a proton transfer reaction to stabilize the covalently-bound Lys166 amine, resulting in DOCKIC3. Here, Lys93 R-group is now ionic, capable of interacting once again via a salt bridge with the phosphate moiety at 2.97 Å (Figure 8.4A). In addition, with the Lys166 adduct now being neutral, K\(_{166}\)N\(_{\varepsilon}\)–C1' bond has shortened by 0.05 Å to 1.46 Å, slightly shorter than a typical C–N single bond. Consequently, this intermediate complex now lies slightly lower in energy at 100.8 kJ mol\(^{-1}\) relative to DOCKRC. The greatest energy decrease of 54.5 kJ mol\(^{-1}\) is noted in the \(^\Psi\)MPIC2 \(\rightarrow\) \(^\Psi\)MPIC3 chemical step, indicating that the loss of K\(_{166}\)N\(_{\varepsilon}\) ionic charge is indeed favourable. In the case of RSPIC3, its relative energy lies exothermically at −22.7 kJ mol\(^{-1}\) with respect to its reactant complex.

8.3.5. \(sp^3\) Hybridization of the C5 Pseudouracil Moiety

In DOCKIC3, a water moiety is positioned 3.05 Å away from the alkenyl C5=C6 of the nucleobase, being the likely source in the \(sp^2\)-to-\(sp^3\) hybridization change of the C5 centre. Note that this water acts as a proton shuttle between the acidic Asp31 carboxyl group and C5, where D\(_{31}\)COOH…O\(_{W}\) is within H-bonding distance at 1.68 Å. As shown in DOCKIC4 of Figure 8.5A, C5 is now \(sp^3\) hybridized, neutralizing the charge on the nucleobase ring N1 atom. Several bond changes occur showing that this reaction process
Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase

occurred within the active site of the docked glycosidase. One example is the lengthening of C5=C6 bond from 1.37 Å to 1.49 Å, indicative of a single bond, while a shortening in C6–N1 occurs, forming an imine bond with $r$(C6=N1) = 1.29 Å in the nucleobase.

![Optimized structures of the glycosidase intermediate complexes](image)

**Figure 8.5.** Optimized structures of the glycosidase intermediate complexes: (A) $\text{DOCKICy}$, (B) $\Psi\text{MPICy}$, and (C) $\text{RSPICy}$, where $y = 4, 5, 6$, obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.

As it appears in Figure 8.5A-B and Table 8.2, both the docked and ΨMP-intermediate models show analogous interactions leading to relative energies of 163.6 and 160.8 kJ mol$^{-1}$ respectively, to each reactant complex. Meanwhile, in Figure 8.5C, $\text{RSPIC4}$ lies only 47.2 kJ mol$^{-1}$ above $\text{RSPIC}$RC. It becomes clear that the orientation of several residues in
Chapter 8 — The Effects of Structural Variation on the Mechanism of ψMP Glycosidase

this model promote more stabilizing effects. One example includes the stronger hydrogen bond between $K93N^\varepsilon H$ and $O_{\text{phos}}$ at 1.61 Å, whereas $r(K93N^\varepsilon H\cdots O_{\text{phos}}) = 2.97$ Å and $r(K93N^\varepsilon H\cdots O_{\text{phos}}) = 2.63$ Å in $\text{DOCK IC4}$ and $\text{ΨMP IC4}$, correspondingly. Another crucial interaction that is absent in both of the higher-energy models is the $K166N^\varepsilon H\cdots O4'H\cdots O_{\text{phos}}$ H-bond network.

8.3.6. Schiff Base Formation

As it is seen in $\text{DOCK IC4}$, the active site is now prepared to undergo C1'-C5 glycosidic cleavage and form an iminium ion between the Lys166 R-group and the R5P ligand. Along with the released anionic uracil, this intermediate complex is described by the $\text{DOCK IC5}$ structure in Figure 8.5A. Specifically, this step involves a retroaldol-type reaction where C1'-C5 bond is cleaved and the electron lone pair of $K166N^\varepsilon$ is used to form the Schiff base with bond length of 1.29 Å, indicating a shortening of 0.16 Å in the $K166N^\varepsilon$-C1' bond relative to the previous intermediate complex, $\text{DOCK IC4}$. It is important to recognize that both $\text{DOCK IC5}$ and $\text{RSP IC5}$ imine bonds are formed with the E-configuration at much lower energies of 51.3 and $-46.2$ kJ mol$^{-1}$ with respect to their reactant complex. However, in the 'ΨMP' model, the less-stable Z-isomer is obtained and it is shown to lie at a higher energy value of 93.8 kJ mol$^{-1}$ when compared to $\text{DOCK IC5}$ (see Table 8.2).

The anionic uracil in $\text{DOCK IC5}$ is stabilized by a water molecule 2.01 Å away. Also several bond changes in the nucleobase occur via electron delocalization, suggesting the stabilization of its excess negative charge (e.g., shortening of C5=C6 by 0.09 Å to 1.43 Å and lengthening of the C6=O carbonyl bond by 0.02 Å to 1.24 Å).

8.3.7. Imine Hydrolysis

The next chemical step in the mechanism is the nucleophilic attack of the Schiff base forming a hydroxylated C1' centre whose structure is shown in $\text{DOCK IC6}$. As mentioned in the previous step, in $\text{DOCK IC5}$, the nucleophilic water is H-bonded to the anionic
nucleobase at 2.01 Å and it is 4.48 Å away from the C1' electrophilic carbon. This suggests uracil ion could act as a catalytic base and enhance the nucleophilic character of the water cofactor. Also in DOCKIC6, we note that the K166Nε–C1' is now a single bond at 1.44 Å and C1'–OH is 1.43 Å, resulting in the production of an imine hydrate intermediate.

It is noted that the formation of the imine hydrate ligand (DOCKIC6) is also seen as an intermediate in the Schiff base synthesis work done by Erdtman et al. ⁷ which has implications in the catalytic mechanism of phorphobilinogen synthase. Similarly, in their computational work, they suggested that it is possible for the enzyme-substrate to be covalently-bound to the imine hydrate but instead forms a Schiff base with the loss of water. In our current study, the glycosidase uses the same mechanism but in reverse, where the iminium ion is hydrolyzed instead.

Once the imine bond of the R5P-adduct is hydrolyzed and the imine hydrate is created, the E/Z-configuration is lost and so is the high-energy difference between the DOCKIC6 and ΨMPIC6 structures. For instance, their energies have dropped significantly, reaching similar values at 31.4 and 32.5 kJ mol⁻¹, correspondingly, to DOCKRC and ΨMPRC. While the relative energies of these models have fallen in ΧIC6, the Schiff base-derived model has marginally increased by 12.7 kJ mol⁻¹ to −33.4 kJ mol⁻¹, yet it still remains thermodynamically favourable.

8.3.8. R5P Cyclization

Now that the Schiff base hydrolysis has occurred and the imine hydrate is formed, it is important to enhance the electrophilic character of the anomeric C1' centre in order to promote ribose ring closure.

Specifically, K166Nε–C1' bond of the imine hydrate needs to be cleaved before the aldehyde intermediate can be formed. In particular, K166Nε likely receives a proton from the ionic Lys93 amine tail, now forming two neutral, unattached active site lysyl residues.
and a curled protonated O4' aldehyde substrate (see DOCK IC7 Cartesian coordinates in Appendix Table A8.1).

In DOCK IC6, the hydroxyl group of the imine hydrate makes a strong H-bond to C3'OH oxygen \( r(C1'OH \cdots OC3') = 1.70 \, \text{Å} \) which in turn interacts with the carboxylate side chain of Asp149 \( r(C3'OH \cdots OOC^{D149}) = 1.73 \, \text{Å} \). This strongly suggests that the Asp149 carboxyl R-group is capable to act as an acid/base catalyst and thus can facilitate the formation of R5P cyclization. More importantly, this observation is consistent with the lack of catalytic activity for D149A glycosidase mutant, which was previously believed to be due to a structural misfold.\(^{17}\)

The aldehyde C1' carbon is within nucleophilic distance of O4' at 2.54 Å and \( r(K93N^{ε} \cdots O4') = 1.65 \, \text{Å} \), suggesting the ring-closing reaction is feasible. During this process, protonation of C1'=O oxygen by Asp141–COOH via C3'OH hydroxyl occurs, resulting in the α-form of the phosphorylated D-ribose sugar, R5P, as shown in DOCK PC of Figure 8.6A. The formation of this phosphorylated sugar is thermodynamically stable at a rather modest value of –5.0 kJ mol\(^{-1}\) with respect to its reactant complex.

It should be noted that while Asp149 carboxyl group may act as a catalytic base in the 'DOCK' model, RSP IC6 shows Asp31–COO\(^{-}\) to be in a more suitable position to abstract the C1'–OH proton, thus producing the RSP IC7 (see Figure 8.5C and RSP IC7 Cartesian coordinates in Appendix Table A8.1). However, the protonation of the Asp31 carboxylate leads to a linear aldehyde intermediate, placing the O4' nucleophile 4.08 Å away from the electrophilic C1' centre. This observation further reinforces the proposal that Asp149 is indeed the acid/base catalyst in the formation of the R5P product.

Nonetheless, RSP PC is obtained from an exothermic reaction at –72.2 kJ mol\(^{-1}\) with respect to RSP RC. Unfortunately, when considering the 'ΨMP' model, ΨMP IC7 was not successfully optimized to form the aldehyde intermediate, however its product was formed with an energy of 36.8 kJ mol\(^{-1}\) relative to ΨMP RC.
Figure 8.6. Optimized structures of the glycosidase product complexes (\(X^{\text{PC}}\), where \(X = \text{DOCK, } \Psi\text{MP, } \text{R5P}\)), obtained at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory. Atom colour code: H, white; C, gray; N, blue; O, red; P, orange.
8.3.9. **Continuum Solvation Effects on Relative Energies**

As mentioned in the Computational Methods, the three active site models were treated using a QM cluster approach, where each is composed of 204 atoms. Based on the size of these models, the solvation effects should be saturated, thus providing accurate description of the long-range polarization from the surrounding substrate environment.\(^3\)

However as shown in Table 8.3 (c.f. Table 8.2), after applying a homogenous polarizable medium to the models using a standard dielectric constant of 4.0, we observed relative energy differences as high as 27.5 kJ mol\(^{-1}\) in the IC5 of 'DOCK' model and as low as –59.5 kJ mol\(^{-1}\) in the IC2 of 'ΨMP' model. These results suggest that although the two systems reach a significant model size, their active site is still not considered suitable to fully represent the stabilizing electrostatic effects.

**Table 8.3.** Relative energies of QM-optimized complexes, obtained at B3LYP/6-311+G(2df,p)/B3LYP/6-31G(d) level of theory with \(\Delta E_{\text{solv}}(\epsilon = 4.0)\) inclusion.

<table>
<thead>
<tr>
<th>Model</th>
<th>RC</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
<th>IC6</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>'DOCK'</td>
<td>0.0</td>
<td>23.4</td>
<td>85.3</td>
<td>86.8</td>
<td>144.3</td>
<td>78.8</td>
<td>54.3</td>
<td>14.4</td>
</tr>
<tr>
<td>'ΨMP'</td>
<td>0.0</td>
<td>–20.6</td>
<td>122.3</td>
<td>112.8</td>
<td>161.1</td>
<td>129.4</td>
<td>45.1</td>
<td>49.9</td>
</tr>
<tr>
<td>'R5P'</td>
<td>0.0</td>
<td>2.7</td>
<td>41.9</td>
<td>–10.2</td>
<td>41.6</td>
<td>–31.6</td>
<td>–39.2</td>
<td>–56.2</td>
</tr>
</tbody>
</table>

When considering the 'R5P' model and accounting for the error margin in B3LYP functional of up to ~13 kJ mol\(^{-1}\), we note that these solvation effect differences almost become irrelevant with an energy rise of ca. 3 kJ mol\(^{-1}\) in PC and IC2 stabilization of ca. –8 kJ mol\(^{-1}\) (see Table 8.3).\(^3\)

8.3.10. **Variation in the Fixed Atom Positions**

Another way in which the exclusion of protein environment can affect the QM cluster models is the modeling of steric effects imposed on the active site. For instance, a coordinate-locking scheme can be applied where key atoms at the boundary of the quantum chemical model are fixed to their crystal-derived structural position (see
Computational Methods), thus keeping the integrity of the model. In particular, 14 carbon atoms located at α-, β-, γ-, or ε-position on the periphery of each 204-atom quantum chemical model, were fixed. The origin of each system is derived from a different X-ray crystal snapshot, indicating that the corresponding chemical model will show displacement variation between the fixed atoms.

As shown in Table 8.4, three residues whose key carbon is fixed, were selected to assess the effect of steric constraints, both around the ligand (i.e., Asp31, Lys93, Lys166) and Mn(II) metal ion (i.e., Asp145, Glu176', Glu179'). The atom distances for $K_{166}C_{\alpha} \cdots D_{31}C_{\beta}$, $K_{166}C_{\alpha} \cdots K_{93}C_{\varepsilon}$, and $K_{93}C_{\varepsilon} \cdots D_{31}C_{\beta}$ in the 'ΨMP' model are 11.60, 11.97, and 9.89 Å, respectively. When compared to the other two models, 'ΨMP' measurements around the ligand are much shorter and they can be up to 1.42 Å closer together, thus suggesting this model undergoes more rigid constraints and is less flexible.

When looking at the fixed carbons of the metal coordination residues, it appears that $E_{176'}C_{\gamma}$ and $E_{179'}C_{\gamma}$ are only 4.06 Å apart or 0.61 and 0.23 Å closer than in 'R5P' and 'DOCK' models, respectively. These results could indicate that the collapse of the Mn(II) octahedral coordination to trigonal bipyramidal in 'ΨMP' may be due to the carboxyl groups of Glu176' and Glu179' having higher steric constraints.

**Table 8.4.** Summary of key distances (Angstroms) between fixed atoms for the three models.

<table>
<thead>
<tr>
<th>Fixed Distance (Å)</th>
<th>'DOCK'</th>
<th>'ΨMP'</th>
<th>'R5P'</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{166}C_{\alpha} \cdots D_{31}C_{\beta}$</td>
<td>13.02</td>
<td>11.60</td>
<td>12.64</td>
</tr>
<tr>
<td>$K_{166}C_{\alpha} \cdots K_{93}C_{\varepsilon}$</td>
<td>12.04</td>
<td>11.97</td>
<td>13.15</td>
</tr>
<tr>
<td>$K_{93}C_{\varepsilon} \cdots D_{31}C_{\beta}$</td>
<td>11.14</td>
<td>9.89</td>
<td>9.58</td>
</tr>
<tr>
<td>$D_{145}C_{\beta} \cdots E_{176'}C_{\gamma}$</td>
<td>6.72</td>
<td>6.85</td>
<td>6.47</td>
</tr>
<tr>
<td>$D_{145}C_{\beta} \cdots E_{179'}C_{\gamma}$</td>
<td>8.10</td>
<td>8.36</td>
<td>8.62</td>
</tr>
<tr>
<td>$E_{176'}C_{\gamma} \cdots E_{179'}C_{\gamma}$</td>
<td>4.29</td>
<td>4.06</td>
<td>4.67</td>
</tr>
</tbody>
</table>
8.3.11. Choosing the Right Model

Initially, we started out using the four available X-ray crystal structures as MD models to depict the different conformers of the glycosidase catalytic pathway. It was determined that model 2 (A166K modification) is not reasonable as a starting geometry to explore the enzymatic activity using the subsequent QM cluster approach. Thus, we created three quantum chemical models where we studied the relative energies of the intermediates both under gas and solvation media (c.f. Tables 8.2 and 8.3). It was concluded that the solvation effects in the 'R5P' model became irrelevant indicating that this system was best described as being functional and appropriate to study the mechanistic reactant, intermediates, and product as summarized in Scheme 4. Moreover, the residue flexibility offered by the 'R5P' model (see Table 8.4) appears to be suitable for modeling the steric effects and likely provides the correct energetics as obtained in Tables 8.2 and 8.3.
8.4. Conclusions

The complementary use of MD and QM cluster methods has been applied to provide significant insights into the modeling of chemical reactions of the $\Psi$MP glycosidase mechanistic pathway. The possible roles of key active site residues such as Asp149, Asp31, and Lys93 have been examined along with the effects that polarizable continuum have on the relative energies. Most importantly, the effect of the starting geometry has
been explored by using all four available X-ray crystal structures of \textit{E. coli} \(\Psi\)MP glycosidase.

Consequently, four MD models were developed including \(\Psi\)MP-docked reactant complex (model 1), \(\Psi\)MP-bound A166K (model 2), ring-opened ribose \(\Psi\)MP covalent adduct (model 3), and ring-opened R5P adduct (model 4). It was concluded that model 2 is not a reasonable starting structure due to several reasons. One reason is that its active site Lys166 amine tail is considerably far away from the anomeric C1' carbon of \(\Psi\)MP at 6.56 ± 0.36 Å, a significant 1.34 Å increase when compared to \(\text{K}^{166}\text{N}^{\varepsilon} \cdots \text{C}1'\) distance in model 1. Another reason why \(\Psi\)MP\(\cdots\)A166K is not a viable option to study the glycosidase mechanism is because its Mn(II) binding site is disordered, losing the metal's monodentate coordination to both carboxyl side chains of Asp145 and Glu176' ligands. Therefore, this starting structure was not considered further.

The remaining three models, `'DOCK'`, `'\(\Psi\)MP'`, and `'R5P'`, were described using a QM cluster approach. It was interesting to see that while these starting geometries possess distinct ligand conformations in some intermediates, all three models 'speak the same language', meaning they express the same mechanistic steps. In particular, possible chemical reactions are as follows:

1. \(\Psi\)MP nucleobase deprotonation at ring nitrogen N3 \textit{via} indirect Asp31–COO\(^-\) proton abstraction;
2. Covalent linkage between the substrate Lys166–NH\(_2\), forming C1'–K\(^{166}\text{N}^{\varepsilon}\text{H}^+\);
3. Hybridization change of the C5 pseudouracil moiety (sp\(^2\) → sp\(^3\)) served using Asp31–COOH;
4. Schiff base formation with concomitant production of uracil anion;
5. C1'–K\(^{166}\text{N}^{\varepsilon}\text{H}^+\) hydrolysis using uracil ion as a catalytic base to activate H\(_2\)O cofactor, forming an imine hydrate;
6. Open-form aldehyde is produced \textit{via} proton abstraction by Asp149 carboxyl;
(7) Ring-closing reaction resulting in the $\alpha$-form D-ribose product using Lys93–NH$_2$ as a catalytic base.

However, when considering the relative energies, the three models show considerable differences. For example, 'DOCK', 'ΨMP', and 'R5P' models show relative energies as high as 163.6, 181.8, and 62.5 kJ mol$^{-1}$, respectively to $^X$RC. Accounting for solvation effects using $\varepsilon = 4.0$, we discover the higher relative energies are lowered, especially in the 'ΨMP' model where IC2 is stabilized by $-59.5$ kJ mol$^{-1}$ relative to vacuum, but still not sufficiently in some cases. The inclusion of continuum solvation to the relative energies in 'R5P' are irrelevant, confirming that the 204-atom QM cluster model is indeed saturated in electrostatic effects. Furthermore, each model showed a variation of steric effects. The coordinate-locking scheme when applied to the same key atoms, which differ in position, resulted in different active site constraints. Together, these results suggest the mass-spectrometry confirmed 'R5P' model represents the most suitable structure to study the ΨMP glycosidase mechanism.

8.5. References

Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase


(20) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., Montréal, QC, Canada, 2015.


Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase


Chapter 8 — The Effects of Structural Variation on the Mechanism of ΨMP Glycosidase


Chapter 9. Conclusions
Chapter 9 — Conclusions

9.1. Conclusions

The complementary use of several computational enzymology methods enabled us to achieve greater insights into the fundamental chemistry underpinning the catalytic power of nitrogen-dependent enzymes and their biochemical role.

In Chapter 3, MD and ONIOM(QM/MM) methods have been applied to study the catalytic mechanism of SpNic. In particular, the overall two-stage nicotinamidase pathway was investigated using an ONIOM(QM/MM) approach within both ME and EE formalism, that is, without and with inclusion of the environment's electrostatic effects, respectively. In Stage 1, within the ME formalism, formation of an enzyme-bound thioester intermediate with loss of ammonia, occurs via a two-step mechanism. The first step, \( RC \rightarrow IC_1 \), is rate-limiting with a barrier of 68.5 kJ mol\(^{-1}\) at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)//ONIOM(B3LYP/6-31G(d):AMBER96) level of theory. On the other hand, at the same level of theory but within an EE formalism, Stage 1 becomes a single chemical step, where nucleophilic attack of the Cys136 sulfur at the carb centre of nicotinamide occurs with concurrent bond breaking of \( C_{\text{carb}}-NH_2 \) to produce \( IC_2 \) with loss of ammonia. The energy barrier is now only 53.5 kJ mol\(^{-1}\). Nevertheless, the hydrolysis of the \( C_{\text{carb}}-S \) bond in \( IC_2' \) resulting in the formation of the nicotinic acid, occurs in one-step reaction within both the ME and EE formalism. The energy barriers are 62.0 and 26.0 kJ mol\(^{-1}\), correspondingly. The inclusion of dispersion effects were modeled using Grimme's D3 dispersion corrections to the B3LYP method and M06 functional within both the ME and EE formalism. The results were all in reasonable agreement, giving rise to the same overall catalytic mechanism for SpNic.

In Chapter 4, DFT, MD, and ONIOM(QM/MM), have been applied to study the catalytic mechanism of NicF amidase. For instance, in addition to the previously suggested oxyanion hole interactions, \( O_{\text{carb}}\cdots\text{HN–Thr146} \) and \( O_{\text{carb}}\cdots\text{HN–Cys150} \), our results indicate that Thr146–OH may also play a stabilizing role by hydrogen bonding with the \( O_{\text{carb}} \) during the mechanism. The catalytic mechanism of NicF was examined
using an ONIOM(QM/MM) approach within both a ME and EE formalism as per the SpNic study. Within the ME formalism, the formation of a tetrahedral intermediate (IC1) was concluded to be rate-limiting at 140.8 kJ mol\(^{-1}\). However, after accounting for electrostatic effects, i.e., within an EE formalism, the calculated barrier is only 79.0 kJ mol\(^{-1}\). In the second stage of the mechanism, the cleavage of C\(_{\text{carb}}\)-S\(_{\text{Cys150}}\) in IC2' occurs in one chemical step within the ME formalism, whereas extending to EE formalism, this chemical step occurs without a barrier at \(-5.4\) kJ mol\(^{-1}\). Here, the rate-limiting step is the H\(^+\) transfer to Cys150 side chain with a reaction barrier of 7.0 kJ mol\(^{-1}\).

In Chapter 5, MD methods allowed us to gain insights into GSA, 4-OH-GSA, NAD\(^+\), and GLU binding to the P5CDH active site in both reactant and product complexes. In addition, several mutant enzymes were modeled including S352L, S352A, and E314A. When comparing the holoenzyme RC, Michaelis complex of the RC, and activated-Cys348 RC, it is concluded that in model 2 the oxyanion hole member, \(-\text{NH– backbone of Cys348}, is interacting at an average of 2.21 Å for O\(_{\text{carb}}\)–HN\(_{\text{Cys348}}\), thus allowing GSA to be catalytically positioned. Also, our results indicate that the average structure of Cys348–S\(^-\) model shows the C\(_{\text{pyr}}\)–S\(_{\text{C348}}\) is 3.65 Å with a planar nicotinamide ring. In addition, several other interactions such as the average \(r(S_{\text{C348}}\cdots\text{HN}_{\text{N211}}) = 2.19\) Å and \(r(S_{\text{C348}}\cdots\text{HN}_{\text{pyr}}) = 2.09\) Å, help stabilize the sulfur negative charge of Cys348. As a result, this conformation appears to be dominant at equilibrium, especially since the ionic Cys348 is key for catalysis to occur. It was also seen that upon removing the GLU product from PC, the oxyanion hole collapses, placing Asn211 amide and Cys348 –NH– backbone away from the active site. When considering the fully-bound RC with GSA and 4-OH-GSA, it was observed that the \(\gamma\)-hydroxyl moiety of GSA keeps a stable interaction via H-bonding with the Glu165 carboxyl at an average \(r(\text{HO}_{4\text{-OH-GSA}}\cdots\text{OOC}_{\text{E165}}) = 1.63\) Å. It appears that this interaction does not have an effect on the enzymatic binding of 4-OH-GSA, making it a convincing model for catalysis.
In the comparison between the average WT and S352L structures, the bulky Leu352 side chain pushes against Phe449 side chain, causing steric hindrance, thus breaking the H-bond interaction between the Glu447 carboxylate and 3'-OH moiety of the nicotinamide ribose. When looking at E314A, the H-bond interaction between the Glu447 side chain and 3'-OH is also disrupted. Thus, the lack of this interaction in both enzyme mutants results in NAD\(^+\) disorientation and likely an inactive enzyme.

In Chapter 6, we applied MD and QM/MM methods to examine the self-cyclization of the noncognate L-ornityl against the native L-lysyl in LysU. When Orn-AMP is used as a substrate, the \(\text{OrnPC}\) lies at \(-92.9 \text{ kJ mol}^{-1}\) relative to \(\text{OrnRC}\). Moreover, the reaction barrier for this step is rate-determining at 36.2 kJ mol\(^{-1}\), indicating the LysU editing process is feasible. However, the substrate-assisted editing of Lys-AMP forms an endothermic product complex with 21.6 kJ mol\(^{-1}\) relative to \(\text{LySRC}\). Also, \(\text{LySTS2}^*\) is similarly rate-determining, although its energy barrier lies much higher at 94.7 kJ mol\(^{-1}\). Together these results suggest that the pre-transfer editing of Orn-AMP is more favourable than the self-cyclization of the cognate Lys-AMP. This further suggests that mischarged tRNA\(^{\text{Lys}}\) with Orn is highly unlikely.

In Chapter 7, the catalytic mechanism of OCD has been investigated using a combination of MD and QM/MM methods. It was found that the initial hydride transfer from the \(C_\alpha-H\) group of the L-ornithine substrate to the \(C_4\) center of the NAD\(^+\) cofactor with concurrent formation of a Schiff base, is the rate-limiting step with an energy barrier of 90.6 kJ mol\(^{-1}\). Notably, with deamination, a second Schiff base was formed. Similarly to the initial \(C-N\) bond formation, this Schiff base would likely become more reactive. As such, the final chemical step in the reaction was a hydride transfer with a low barrier of 65.2 kJ mol\(^{-1}\), resulting in the feasible formation of L-proline.

In Chapter 8, the modeling of chemical reactions of the \(\Psi\)MP glycosidase mechanistic pathway were done via MD, QM cluster methods, and four X-ray crystal structures. Subsequently, from the four MD models, it was concluded that 2 is not a
reasonable starting structure due to multiple reasons. For instance, active site Lys166 amine tail is considerably far away from the anomic C1' carbon of ΨMP at 6.56 ± 0.36 Å, a significant 1.34 Å increase when compared to $^{K166N^e}\cdots C1'$ distance in model 1. Also, the Mn(II) binding site is disordered in ΨMP···A166K, making it an unreliable model to study the glycosidase mechanism. The other three models, 'DOCK', 'ΨMP', and 'R5P', were described using a QM cluster approach. It was determined that they all lead to the same mechanistic steps. Nevertheless, when considering the relative energies, the three models indicate considerable deviations. 'DOCK', 'ΨMP', and 'R5P' show relative energies as high as 163.6, 181.8, and 62.5 kJ mol$^{-1}$, correspondingly to $^XRC$. When we account for solvation effects using a dielectric constant of 4.0, we see the higher relative energies are reduced. The largest stabilizing effect is seen in the 'ΨMP' model, where $IC2$ is stabilized by –59.5 kJ mol$^{-1}$ relative to gas phase. However, in some cases this is still not sufficient. Using IEFPCM, the relative energy differences of 'R5P' become irrelevant, confirming that the solvation effects in the 204-atom QM cluster model are indeed saturated. Furthermore, each model showed a variation of steric effects as they differed in atom position. Finally, these results suggest the 'R5P' model is best suited to study the catalytic mechanism of ΨMP glycosidase.

9.2. References

Chapter 10.    Future Work

"Science [...] is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth."
- Jules Verne
Chapter 10 — Future Work

In this dissertation, computational modeling tools such as fully atomistic molecular dynamics (MD) simulations, \textit{in silico} mutagenesis, quantum mechanical (QM) clusters, and quantum mechanics/molecular mechanics (QM/MM) have proven to be invaluable in the mechanistic characterization of several types of nitrogen-dependent enzymes. It is hoped that the results obtained from the investigations presented herein will provide a way towards the design of new peptidomimetic catalysts.

Computational approaches are becoming even more robust and accurate as the technology performance is constantly growing to support efficient and well-behaved theoretical methods. For instance, in a biological system, the reactive center and its surrounding environment can then be expanded in model size to proficiently describe even the most of complicated mechanistic problems. Another example is the development of enhanced sampling methods with reduced bias that are much needed to access millisecond time scale events effectively such as protein folding and life-signaling processes.

The increasing availability of high-resolution X-ray crystal structures of proteins provide the opportunity to go beyond the static structural view and explore the protein dynamics using the ever-increasing power of computational chemistry. Most importantly, \textit{in silico} methods such as homology modeling are getting better at guessing the three-dimensional structures of unknown or hard-to-crystallize proteins such as membrane receptors, thus paving the way for future development of unknown crystals. It is believed that scientific collaborations between computational chemists and experimentalists are becoming more prevalent, enabling growth in many chemical and pharmaceutical industries. Consequently, treatments for a number of diseases can be developed by the insights gained into the mechanisms of proteins and enzymes \textit{via} the marriage of computer-aided drug design and experimental studies.
Appendices
Appendix A. Supporting Information

For access to Appendix A please refer to the University of Windsor online database of Chemistry and Biochemistry Theses and Dissertations.

**Table A3.1.** Cartesian coordinates and partial charges from the QM/MM optimized structures of each stationary point along the SpNic pathway.

**Table A4.1.** Optimized Cartesian coordinates and partial charges from the QM/MM structures of the NicF catalytic pathway.

**Figure A5.1.** RMSDs in the substrate position for the GSA-bound RC. (A) First plateau. (B) Second plateau. (C) Third plateau. Note all of these calculations were based on the last 5 ns of the simulation.

**Figure A5.2.** RMSDs in the positions of RC active site residues along with (A) both NAD\(^+\) and GSA ligands and (B) NAD\(^+\) coenzyme. In addition, the unbound site is represented in (C). Note all of these calculations were based on the last 5 ns of each simulation.

**Figure A5.3.** Overlays of representative average structures of the (A) fully-bound, (B) NAD\(^+\)-bound, and (C) unbound RC active sites. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure A5.2. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

**Figure A5.4.** Overlays of representative average structures for the fully-bound RC with (A) NAD\(^+\)-bound, (B) GSA-bound, and (C) unbound RC active sites. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figures A5.1 and 5.3A. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.
Figure A5.5. RMSDs in the positions of active site residues for the unbound PC. Note this calculation was based on the last 5 ns of the simulation.

Figure A5.6. Overlays of representative PC average structures for the (A) unbound, (B) GLU-bound and unbound active sites. The bolder atoms represent those used in the RMSDs from Figures A5.5 and 5.5A. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure A5.7. Overlays of representative average structures for the 4-OH-GSA- and NAD$^+$-bound RC based on RMSD active site residues and (A) 4-OH-GSA, NAD$^+$, (B) absence of both ligand positions. For clarity, not all hydrogen atoms are shown. The bolder atoms represent those used in the RMSDs from Figure 5.7B-C. Atom colour code: H, white; C, gray; N, blue; O, red; S, yellow.

Figure A5.8. RMSDs of RC and PC based on the positions of Ser352 with surrounding residues and respective ligands. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. Note all of these calculations were based on the last 5 ns of each simulation.

Figure A5.9. RMSDs of RC and PC based on the positions of Glu314 with surrounding residues and respective ligands. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. Note all of these calculations were based on the last 5 ns of each simulation.

Figure A5.10. RMSDs of RC and PC with the S352L mutation are based on the positions of Leu352 with surrounding residues and ligands. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. Note all of these calculations were based on the last 5 ns of each simulation.
**Figure A5.11.** RMSDs of RC and PC with the S352A mutation are based on the positions of Ala352 with surrounding residues and ligands. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. Note all of these calculations were based on the last 5 ns of each simulation.

**Figure A5.12.** RMSDs of RC and PC with the E314A mutation are based on the positions of Ala314 with surrounding residues and ligands. (A) Fully-bound and (B) unbound RCs. (C) GLU-bound and (D) unbound PCs. Note all of these calculations were based on the last 5 ns of each simulation.

**Table A5.1.** Average $r(C_{\text{carb}}\cdots(H)SC_{348})$, $r(O_{\text{carb}}\cdots HS_{C348})$, $r(O_{\text{carb}}\cdots HN_{C348})$, $r(O_{\text{carb}}\cdots HN_{N211})$, $r(\partial C_{E314}\cdots(H)SC_{348})$, $r(C_{\text{pyr}}\cdots(H)SC_{348})$, and $r(C_{\text{pyr}}\cdots H_{\text{carb}})$ distances (Å) for model 3b (related to Figures 5.8 and 5.9).

**Table A5.2.** Average $r(C_{\text{carb}}\cdots(H)SC_{348})$, $r(O_{\text{carb}}\cdots HS_{C348})$, $r(O_{\text{carb}}\cdots HN_{C348})$, $r(O_{\text{carb}}\cdots HN_{N211})$, $r(\partial C_{E314}\cdots(H)SC_{348})$, $r(C_{\text{pyr}}\cdots(H)SC_{348})$, $r(C_{\text{pyr}}\cdots H_{\text{carb}})$, $r(\zeta C_{P449}\cdots \alpha C_{C348})$, and $r(\partial C_{E447}\cdots \alpha C_{C348})$ distances (Å) for models 1, 6, 13, and 14 (related to Figures 5.10 and 5.11).

**Figure A6.1.** RMSDs for the Orn-AMP bound LysU structure. Note a second run was performed since the active site RMSDs didn't reach convergence (refer to manuscript).

**Table A6.1.** Optimized Cartesian coordinates and partial charges from the QM/MM structures of both Orn- and Lys-AMP bound LysU catalytic pathways.

**Table A7.1.** Optimized Cartesian coordinates and partial charges from the QM/MM structures of the OCD catalytic pathway.

**Table A8.1.** Optimized Cartesian coordinates from the QM-cluster structures of ΨMP glycosidase based on three X-ray crystal structures as starting points.
Appendix B. Copyright Permissions

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).
This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).
Vita Auctoris

Name: Bogdan Florin Ion

Place of Birth: Bucharest, Romania

Year of birth: 1988

Education:

Ph.D. Chemistry
University of Windsor
October 2015
Windsor, Ontario, Canada

B.Sc. Chemistry [H]
Minor: Mathematics
University of Windsor
June 2010
Windsor, Ontario, Canada