Computational Investigations into Carbohydrate-Related Biochemistry

WenJuan Huang
University of Windsor

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

Part of the Chemistry Commons

Recommended Citation
Huang, WenJuan, 'Computational Investigations into Carbohydrate-Related Biochemistry' (2012). Electronic Theses and Dissertations. 4786.
https://scholar.uwindsor.ca/etd/4786

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 519-253-3000 ext. 3208.
Computational Investigations into
Carbohydrate-Related Biochemistry

by

WenJuan Huang

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

2012

© 2012 WenJuan Huang
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:
L’auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l’Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L’auteur conserve la propriété du droit d’auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n’y aura aucun contenu manquant.
Computational Investigations into Carbohydrate-Related Biochemistry

by

Wenjuan Huang

APPROVED BY:

__________________________________________________
S. Wetmore, External Examiner
University of Lethbridge

__________________________________________________
B. Zielinski
Department of Biology

__________________________________________________
S. Ananvoranich
Department of Chemistry and Biochemistry

__________________________________________________
J. Wang
Department of Chemistry and Biochemistry

__________________________________________________
J. Gauld, Advisor
Department of Chemistry and Biochemistry

__________________________________________________
A. Alfakih, Chair of Defense
Department of Mathematics & Statistics

June 19, 2012
Declaration of Co-Authorship/Previous Publication

I. Co-Authorship Declaration

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

Chapters 6 is in collaboration with my supervisor Professor James Gauld.

Chapter 3 and 4 are in collaboration with Professor Jorge Llano under the supervision of Professor James Gauld.

Chapter 5 is in collaboration with Mr. Rami Gherib under the supervision of Professor James Gauld.

Chapter 7 is in collaboration with Mr. Eric A. C. Bushnell and Professor Christopher S. Francklyn under the supervision of Professor James 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 thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.
II. Declaration of Previous Publication

This thesis includes 5 original papers that have been previously published/submitted for publication in peer reviewed journals, as follows:

<table>
<thead>
<tr>
<th>Thesis Chapter</th>
<th>Full citation</th>
<th>Publication status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5</td>
<td>Huang, W.-J.; Gherib, R.; Gauld, J. W J. Phys. Chem. B</td>
<td>Accepted pending minor revisions</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Huang, W.-J.; Gauld, J. W J. Am. Chem. Soc.</td>
<td>Under 1st revision</td>
</tr>
</tbody>
</table>

I certify that I have obtained written permission from the copyright owner(s) to include the above published materials in my thesis. 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 thesis 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 thesis, 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 written permission from the copyright owners to include such materials in my thesis.
I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University of Institution.
Abstract

As the most abundant class of biological molecules, carbohydrates are essential to all living organisms. Their derivatives play a variety of important roles in biological systems. In this thesis, carbohydrate-related biochemistry is investigated using various computational methods.

Chapter 1 presents an overview of the problems addressed in this thesis, and Chapter 2 discusses various theoretical methods.

Chapter 3 is an investigation on the NAD⁺-dependent oxidation mechanism of UDP–glucose dehydrogenase (UDPGlcDH), a target for the development of new antibacterial drugs. A large enzyme active-site modelling approach was used in the investigation of the mechanism.

Chapter 4 is a DFT investigation on how 6-Phospho-α-glucosidase (GlvA) exploits the inherent properties of modified sugar-rings that dramatically enhance the rate of glycosidic bond cleavage process. The driving-force of the mechanism and enzyme regioselectivity were explained using natural bond orbital theory (NBO) and second-order perturbation analyses.

Chapter 5 is a computational investigation, involving Docking and MD methods, to elucidate substrate binding within the active site of LuxS. In particular, we aim to determine the substrate binding conformations in the enzyme active site and the first stage of the enzyme mechanism.

Chapter 6 is a quantum mechanics/molecular mechanics (QM/MM) and DFT investigation on the catalytic mechanism of the flavoenzyme UDP-Galactopyranose Mutase (UGM). A complete understanding of its mechanism can potentially enable the development of new TB therapeutic drugs. We studied two enzyme active-site models with a protonated or neutral Histidine residue using DFT cluster approach. Then, a
QM/MM-based approach was used to include the steric and non-homogeneous polar environment of the enzyme's active site.

Chapter 7 presents a new type of substrate assisted-mechanism that was proposed for aminoacyl tRNA synthetase, an essential step in protein synthesis. Our DFT calculation results indicated that the neutral amine group on the substrate could act as the required general base in the mechanism. This is the first time that such a substrate assisted catalytic mechanism has been proposed for this presumably ancient class of enzymes.

Finally, Chapter 8 summarizes the main conclusions and possible extensions of the current work.
This work is dedicated to my family, the AWESOME Gauld Group, and the beautiful world.
Acknowledgments

First and foremost, I would like to express my sincere thanks to my supervisor, Dr. James W. Gauld. His guidance and continuous encouragement during the past five years helped me build up solid background knowledge and lab skills that will be invaluable for my future research. His support of my research throughout my graduate studies is greatly appreciated.

My sincere gratitude also goes to my Ph.D. committee members, Prof. Jichang Wang, Dr. Sirinart Ananvoranich, and Prof. Barbara Zielinski, for their time and patient instructions and discussions involving my research during the past few years. In addition, I also thank them for taking the time to read and evaluate this work.

I am also grateful to all the Gauld group members, over the years, for their support and assistance in my research projects as well as my colourful and varied foreign life in Canada. In particular, I would like to thank Eric A. C. Bushnell, Hisham Dokainsh, Bogdan Ion, Rami Gherib for proofreading this thesis and discussions on computational chemistry. I would also like to specifically thank to Prof. Jorge Llano and Dr. Haining Liu for enlightening me on computational biochemistry studies. Furthermore, my thanks are expressed to all the current and past group members: Erum Kazim, Grant Fortowsky, Natalia Mroz, Avtinder Basra, Joel Almasi, Jesse Robinet and Robert Ty. All of you have made my life here enjoyable. In addition, I would like to thank Prof. Bulent Mutus and his group members especially Rebecca Wang, and Prof. Jichang Wang and his group members as well as all other people in the department of chemistry and biochemistry for friendship and company in past few years.
Thanks are also extended to Mr. Joe Lichaa for technical support with computer problems and the department graduate secretary Mrs. Marlene Bezaire for all her assisting on my academic study and life at the University of Windsor.

The scholarships and funding from University of Windsor, SHARCNET, and NSERC are gratefully acknowledged.

Finally, I would like to thank my mother Shengju Hu and father Chongzhao Huang for their love and encouragement over the years. I also would like to thank Mr. and Mrs. Afolabi and their family for their care as my family in Windsor. Last, I would like to cherish the memory of my grandma in heaven.
# Table of Content

Declaration of Co-Authorship/Previous Publication ........................................ iii

Abstract ............................................................................................................. vi

Dedication .......................................................................................................... viii

Acknowledgments ............................................................................................ ix

List of Figures .................................................................................................... xvii

List of Schemes .................................................................................................. xxiv

List of Tables ...................................................................................................... xxvi

List of Appendices (on CD) ................................................................................ xxvii

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

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

  References ......................................................................................................... 6

Chapter 2 Theoretical Methods ......................................................................... 7

  2.1 Introduction ................................................................................................. 8

  2.2 The Schrödinger Equation .......................................................................... 9

  2.3 The Variational Theorem ........................................................................... 10

  2.4 Further Simplification of the Schrödinger Equation .................................. 11

      2.4.1 The Born-Oppenheimer (B-O) Approximation .................................... 11

      2.4.2 The Orbital Approximation ................................................................ 13

  2.5 The Hartree-Fock Self-Consistent Method ................................................ 14

      2.5.1 Self-Consistent-Field Process .............................................................. 15

      2.5.2 Electron Correlation ............................................................................ 17
2.6 Basis Sets .......................................................................................................................................................... 17
  2.6.1 Minimal Basis Set ......................................................................................................................................... 18
  2.6.2 Split-Valence Basis Sets ............................................................................................................................... 18
  2.6.3 Polarization Basis Sets ............................................................................................................................... 19
  2.6.4 Diffuse Functions ......................................................................................................................................... 19
2.7 Hierarchy of \textit{ab initio} Methods ...................................................................................................................... 20
2.8 Density Functional Theory ............................................................................................................................... 21
  2.8.1 Hohenberg-Kohn Theorem ......................................................................................................................... 22
  2.8.2 The Kohn-Sham Approach .......................................................................................................................... 23
2.9 Molecular Mechanics ......................................................................................................................................... 25
2.10 Molecular Dynamics ......................................................................................................................................... 28
2.11 Quantum Mechanics/Molecular Mechanics (QM/MM) Methods .................................................................... 29
2.12 Natural Bond Orbital (NBO) Method ................................................................................................................ 32
2.13 Atoms In Molecules (AIM) ............................................................................................................................ 34
  2.13.1 Topology of the Electron Density ............................................................................................................... 34
  2.13.2 Criteria for Defining Hydrogen Bond Interactions in AIM ......................................................................... 35
2.14 Potential Energy Surfaces ............................................................................................................................... 36
  2.14.1 Geometry Optimizations ............................................................................................................................ 39
  2.14.2 Frequency Calculations .............................................................................................................................. 39
  2.14.3 Single Point Energy Calculations ............................................................................................................... 40
2.15 Environmental Effects ....................................................................................................................................... 41
  2.15.1 Solvation Models ....................................................................................................................................... 41
  2.15.2 Electronic Embedding (EE) in QM/MM ...................................................................................................... 42
2.16 Technical Aspects and Units ............................................................................................................................ 43
References ............................................................................................................................................................... 44
Chapter 3 A DFT Study on the Catalytic Mechanism of UDP–Glucose Dehydrogenase .................................................................................................................................................................................. 48

3.1 Introduction ......................................................................................................................................................................................................................... 49
3.2 Computational Methods .............................................................................................................................................................................................. 51
3.3 Results ............................................................................................................................................................................................................................................. 53
  3.3.1 Protonation state of the active site of UDPGlcDH in the Michaelis complex .......... 53
  3.3.2 Substrate binding in the active-site of UDPGlcDH ................................................................................................................................. 56
  3.3.3 First catalytic step: Oxidation of glucose to form glucuronaldehyde ...................... 58
  3.3.4 Second catalytic step: Nucleophilic addition to form the 6-thiohemiacetal intermediate ........................................................................................................ 60
  3.3.5 Third catalytic step: Second oxidation to form the 6-thioester intermediate ...... 62
  3.3.6 Fourth catalytic step: Hydrolysis of the 6-thioester intermediate to form glucuronic acid ........................................................................................................ 66
3.4. Conclusions .................................................................................................................................................................................................................. 70
References ...................................................................................................................................................................................................................... 72

Chapter 4 Redox Mechanism of Glycosidic-Bond Hydrolysis Catalyzed by 6-Phospho-α-Glucosidase (GlvA): A DFT Study ................................................................. 75

4.1 Introduction .................................................................................................................................................................................................................. 76
4.2 Computational Methods .............................................................................................................................................................................................. 81
4.3 Results ............................................................................................................................................................................................................................................. 83
  4.3.1 Substrate Binding in the Active Site .................................................................................... 83
  4.3.2 3-Oxidation of Glucopyranose to Form the C2–C3=O Carbonyl Intermediate ...... 85
    4.3.2.1. Proton Transfer from the C3–OH Group to the Mn2+···OH− Moiety ......................... 86
    4.3.2.2. Hydride Transfer from the C3–H Group to NAD+ .................................................... 88
  4.3.3 2-Deprotonation of the C2–C3=O Carbonyl Intermediate to Form the C2=C3=O− Enolate .................................................................................................................................................................................................................. 89
4.3.4. Glycosidic bond cleavage ........................................................................................................ 91
4.3.5. Hydrolysis of the C1=C2–O\(^-\) Enolate and Subsequent 2-Protonation to Form the C2–C3=O Carbonyl Intermediate ........................................................................................................ 93
4.3.6. 3-Reduction of the C2–C3=O Carbonyl Intermediate and Subsequent Reprotonation to Form Glucopyranose ........................................................................................................ 95
4.3.7. Natural Population and Energy Analyses of the Electron Densities of Equilibrium Structures Along the Reaction Coordinate ........................................................................................................ 97
4.4. Conclusions .................................................................................................................................. 106
References ........................................................................................................................................... 108

Chapter 5 An Active Site Water Broadens Substrate Specificity in S-ribosylhomocysteinase (LuxS): A Docking, MD and QM/MM Study ..................113
5.1. Introduction .................................................................................................................................. 114
5.2. Computational Methods .................................................................................................................. 118
5.3. Results and Discussion .................................................................................................................... 122
5.3.1 Docking and MD Simulations ........................................................................................................ 122
5.3.1.1 The active site-bound 2-keto-SRH intermediate complex ....................................................... 122
5.3.1.2 The active site-bound β- and α-SRH complexes ...................................................................... 126
5.3.1.3 The active site-bound linear aldose-SRH complex ................................................................. 129
5.3.2 QM/MM investigation on the catalytic mechanism ...................................................................... 131
5.3.2.1 Interconversion of active site-bound substrate complexes ....................................................... 135
5.3.2.2 The proposed mechanism for formation of the 2-keto-SRH intermediate .............................. 137
5.3.2.3 Alternative mechanisms for formation of the 2-keto-SRH intermediate ............................... 140
5.4. Conclusions .................................................................................................................................. 145
References ........................................................................................................................................... 148

Chapter 6 Tautomerization in the UDP-Galactopyranose Mutase Mechanism: A DFT-Cluster and QM/MM Investigation .............................................151
6.1 Introduction ......................................................................................................................... 152
6.2 Computational methods .......................................................................................................... 156
6.3 Results and Discussion .......................................................................................................... 159
  6.3.1 Deprotonation of the initial flavin cofactor ................................................................. 159
  6.3.2 Catalytic mechanism obtained using the DFT-cluster model................................. 161
    6.3.2.1 Formation of a flavin-substrate (N5PAD—C1gal) covalent bond .................. 162
    6.3.2.3 Conformational rearrangement of the linear sugar moiety .......................... 167
    6.3.2.4 Formation of the galactofuranose ring ......................................................... 168
  6.3.3 Catalytic mechanism obtained using the QM/MM model ........................................ 172
6.4 Conclusions ......................................................................................................................... 177

References ..................................................................................................................................... 181

Chapter 7 The α-Amino Group of the Threonine Substrate As The General Base During tRNA Aminoaacylation: A New Version of Substrate Assisted Catalysis Predicted by Hybrid DFT ................................................................................................................................. 185

7.1 Introduction ......................................................................................................................... 186
7.2. Computational Methods ....................................................................................................... 190
7.3. Results .................................................................................................................................. 193
  7.3.1. Substrate-bound active site ......................................................................................... 193
    7.3.1.1. 'Pre-reactive' complexes (PRC's) .................................................................. 194
    7.3.1.2. 'Reactive' complexes (RC's) ........................................................................ 195
  7.3.2. The protonated His309-assisted mechanism ............................................................ 200
    7.3.2.1. Formation of the tetrahedral intermediate complex .................................... 200
    7.3.2.2. Rearrangement of the tetrahedral intermediate complex .......................... 204
    7.3.2.3. Formation of the product complex ........................................................... 205
  7.3.3. The neutral His309-assisted mechanism ................................................................. 206
7.4. Conclusions ......................................................................................................................... 212
References ........................................................................................................................................... 215

Chapter 8 Conclusions .................................................................................................................... 219

Vita Auctoris ..................................................................................................................................... 225
## List of Figures

**Figure 1.1** Examples of carbohydrate derivatives with the sugar component highlighted in red: (a) cellulose, (b) uridine diphosphate glucose (UDP), and (c) a fragment of deoxyribonucleic acid (DNA). ................................................................. 2

**Figure 1.2** Glucose conformational interconversion reactions......................................................... 3

**Figure 2.1** A Pople diagram illustrating the relationship between basis set size, extent of inclusion of electron correlation, and the exact solution of the Schrödinger equation. ............................................................................................................ 20

**Figure 2.2** Illustration of a two-layer QM/MM model: the atoms shown in ball and stick format are in the high-layer (QM region) while those shown in wire format are in the low-layer (MM region). ............................................................................................................. 30

**Figure 2.3** Schematic illustration of the energies of the perturbative donor-acceptor interaction involving the occupied and unoccupied orbitals in the NBO method...... 33

**Figure 2.4** Schematic illustration of a 2-D PES for a chemical reaction generated by plotting energy versus reaction coordinate: RC = reactant complex, TS = transition state, and PC = product complex.............................................................. 37

**Figure 2.5** Schematic illustration of a 3-D PES for a chemical reaction generated by plotting energy versus two structural coordinates. ........................................................... 38

**Figure 2.6** Schematic illustration of multiple reaction pathways interconnecting the reactant (RC) and product (PC) complexes via different possible intermediates (ICs). ......................................................................................................................... 38

**Figure 2.7** Graphic description of a polarized continuum model.......................... 42
**Figure 3.1** Structural models of the active site: (a) Arrangement of the catalytically active residues of UDP-glucose dehydrogenase according to the crystal structure (PDB accession code: 1DLJ). (b) Catalytic groups replacing the residues and glucose replacing UDP-glucose with protonated Lys204 residue. (c) Catalytic groups replacing the residues, and glucose replacing UDP-glucose with neutral Lys204 residue. Atoms marked with an asterisk were held fixed during geometry optimizations. ........................................................................................................................................53

**Figure 3.2** Optimized structure of the reactant complex (RC): substrate and cofactor bound to the active-site model of UDPGlcDH........................................................................................................................................57

**Figure 3.3** Potential energy surface for the oxidation of the 6-hydroxymethyl group of glucopyranose (RC) to form glucuronaldehyde (IC2) through the 6-oxymethyl anion intermediate (IC1)........................................................................................................................................59

**Figure 3.4** Optimized geometries of the 6-oxyanion intermediate (IC1) and the 6-formyl intermediate (IC2) in the active-site model........................................................................................................................................60

**Figure 3.5** Potential energy surface for the nucleophilic addition of the Cys260 thiol group to the 6-formyl end of glucuronaldehyde in order to form a 6-thiohemiacetal intermediate (IC4)........................................................................................................................................61

**Figure 3.6** Optimized structures of the thiohemiacetal anion (IC3) and neutral (IC4) intermediates. ........................................................................................................................................................................................................62

**Figure 3.7** Optimized structures of the 6-thiohemiacetal complex with both NADH and NAD⁺........................................................................................................................................................................................................63

**Figure 3.8** Potential energy surface for the oxidation of the 6-thiohemiacetal intermediate (IC4) to the 6-thioester intermediate (IC6) through a 6-thiohemiacetal oxyanion intermediate (IC5). ........................................................................................................................................................................................................65
**Figure 3.9** Optimized structures of the 6-thiohemiacetal oxyanion intermediate (IC5) and the 6-thioester intermediate (IC6)...........................................................................................................65

**Figure 3.10** Potential energy surface for hydrolysis of the 6-thioester intermediate (IC6) assisted by Glu141 and two water molecules, and subsequent formation of a product complex containing GlcUA (PC)...........................................................................................................67

**Figure 3.11** Optimized structures of the 6-thioester intermediate incorporating glutamate and two water molecules (IC6···Glu141•2H2O), the 6-thioether-6,6-diol anion intermediate (IC7) and the product complex with GlcUA (PC)...........................................68

**Figure 4.1** Structural models of the active site: (a) Arrangement of the catalytically active residues of GlvA and the 6-phospho-β-glucopyranose analog substrate in the crystal structure (PDB accession code: 1U8X); (b) Chemical model for the mechanistically important species, with 1-methyl α-glucopyranoside replacing the 6-phospho-β-glucopyranose analog substrate. ..................................................................................................................83

**Figure 4.2** Optimized structure of the active-site complex with bound αGlc1Me substrate (RC). ........................................................................................................................................................................84

**Figure 4.3** Potential energy surface for the first-half reaction (3-oxidation and deglycosylation) in the catalytic mechanism of redox hydrolysis of αGlc1Me by GlvA. ................................................................................................................................................................................86

**Figure 4.4** Optimized structures for the transition states TS1 and TS2, the 3-oxyanion intermediate IC1 and the 3-keto intermediate IC2.................................................................87

**Figure 4.5** Optimized structures of the transition state TS3 and the C2=C3–O-enolate intermediate IC3. ...........................................................................................................................................90

**Figure 4.6** Optimized structures for the transition state TS4 and the 1-en-3-keto-2-oxyanion intermediate complex IC4···MeOH.................................................................92
Figure 4.7 Potential energy surface for the second-half reaction (1-hydroxylation and 3-reduction) in the catalytic mechanism of redox hydrolysis of αGlc1Me by GlvA.

Figure 4.8 Optimized structures for the 2-en-3-oxyanion intermediate IC5, the 3-keto intermediate IC6, the 3-oxyanion intermediate IC7 and the product complex consisting of glucose in the regenerated active site (PC).

Figure 5.1 Schematic illustration of the ONIOM QM/MM active site-containing chemical model with 2-keto-SRH bound: the inner circle indicates those moieties included in the QM region while the outer circle indicates those residues and H2O’s included in the MM layer.

Figure 5.2 Plots of RMSD's calculated during the MD simulations for the enzyme-2-keto-SRH intermediate complex: (a) RMSD-QM versus time (ps); (b) RMSD-active site residues versus time (ps); (c) RMSD-substrate versus time (ps).

Figure 5.3 Overlay of select residues and active site moieties of the five (5) average structures obtained from cluster analysis of RMSD-QM's of the active site bound-2-keto-SRH complex. For clarity, only selected hydrogen atoms are shown. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

Figure 5.4 For the active site-bound b-SRH substrate complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

Figure 5.5 For the active site-bound α-SRH substrate complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).
**Figure 5.6** For the active site-bound linear aldose SRH substrate (Linear-SRH) complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua)............................. 131

**Figure 5.7** Optimized structures with selected bond distances (angstroms) of the active site-bound SRH substrate complexes in which the substrates ribose sugar is in its α- (α-RC) or β-furanose (β-RC) anomer or corresponding linear aldose configurations (α-RC' and β-RC'), and transition structures for ring opening α-TS and β-TS (see text). Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (purple).................................................................. 135

**Figure 5.8** Schematic PESs for unassisted inter-conversion of (a) α-RC and α-RC', and (b) β-RC and β-RC'. ................................................................. 136

**Figure 5.9** Calculated PES for reaction of the aldehyde-containing active site-bound substrate complex α-RC' to give the 2-keto containing derivative IC2 (see text).... 138

**Figure 5.10** Optimized structures with selected bond distances (angstroms) of α-TS1, α-IC1, α-TS2, IC2; see text. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (purple).................................................................. 139

**Figure 5.11** Calculated PESs for the alternate 'water assisted' pathway for reaction of (a) α-RC to give the 2-keto derivative IC2 and (b) β-RC to give the 2-keto derivative IC2. .................................................................................. 142

**Figure 5.12** Optimized structures with selected bond distances (angstroms) of a-TS1', b-TS1', b-IC1' and b-TS2' (see text). Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (purple).................................................................................. 144

**Figure 5.13** Comparison of the PESs for reaction of a-RC and b-RC via a 'water assisted' ring opening process alternative mechanisms................................................. 145
Figure 6.1 The large fully-bound active site model used in the DFT-cluster calculations. Atoms marked by an asterisk (*) were fixed at their X-ray crystal structure (PDB ID: 3INT) position

Figure 6.2 The QM/MM substrate-bound active site model: the inner circle indicates those moieties placed in the QM layer while the outer circle indicates those placed in the MM layer. Residue color code: included in their entirety (red); backbone was included with R-group modeled by a methyl (blue); only R-groups included (black).

Figure 6.3 PES for the catalytic mechanism of UGM (with a neutral His60) obtained at the IEFPCM(ε=4.0) B3LYP/6-311+G(2d,p)//B3LYP/6-31G(d) level of theory... 160

Figure 6.4 Optimized structures with selected bond lengths (Angstroms) of \( b^1 \)RC, \( b^1 \)TS1 and \( b^1 \)C1 (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).

Figure 6.5 Optimized structures with selected bond lengths (Angstroms) of \( b^1 \)TS2, \( b^1 \)C2, \( b^1 \)TS3 and \( b^1 \)C3 (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).

Figure 6.6 Optimized structures with selected bond lengths (Angstroms) of \( b^1 \)TS4, \( b^1 \)C4, \( b^1 \)TS5 and \( b^1 \)C5 (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).

Figure 6.7 Optimized structures with selected bond lengths (Angstroms) of \( b^1 \)TS6, \( b^1 \)C6, \( b^1 \)TS7, \( b^1 \)C7, \( b^1 \)TS8 and \( b^1 \)PC (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).... 171

Figure 6.8 PES for the catalytic mechanism of UGM (with a neutral His60) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d):Amber96)-ME level of theory.
Figure 7.1 Optimized structures with selected bond distances (Angstroms) of the fully bound active site pre-reactive and reactive complexes in which His309 is protonated (aPRC and aRC respectively) or neutral (bPRC and bRC respectively).... 194

Figure 7.2 Calculated PES for the protonated His309-assisted mechanism of ThrRS. ................................................................................................................................................................. 200

Figure 7.3 Optimized structures with selected bond lengths (Angstroms) of the intermediates (IC), transition structures (TS) and product complex (PC) of the protonated His309-assisted mechanism of ThrRS. For clarity, only some residues are shown in ball and stick format. Color key: C (gray); O (red); N (blue); P (orange); S (yellow), Zn (lavender) and H (white). ........................................................................................................................................ 203

Figure 7.4 Calculated PES for the neutral His309-assisted mechanism of ThrRS. . 206

Figure 7.5 Optimized structures with selected bond lengths (Angstroms) of the intermediates (IC), transition structures (TS) and product complex (PC) of the neutral His309-assisted mechanism of ThrRS. For clarity, only key residues are shown in ball and stick format. Color key: C (gray); O (red); N (blue); P (orange); S (yellow), Zn (lavender) and H (white). ........................................................................................................................................ 209

Figure 7.6 Calculated PES for the neutral His309-assisted mechanism of ThrRS in which the lysyl residue (Lys465) is included in the chemical model (see text) ....... 211
**List of Schemes**

**Scheme 3.1** Uridine 5'-diphosphate glucuronic acid, abbreviated UDPGlcUA ........49

**Scheme 3.2** The proposed catalytic mechanism of UDP-glucose dehydrogenase (UDPGlcDH).\(^\text{10}\) ..........................................................................................................................51

**Scheme 4.1** The NAD\(^+\)-Dependent Redox Pathway of Glycosidic-Bond Hydrolysis Catalyzed by GlvA.\(^\text{16,17,19-22,30}\) ..........................................................................................................................79

**Scheme 5.1** Illustration of the enzymatic pathway of AI-2 biosynthesis..............115

**Scheme 5.2** The proposed catalytic mechanism of LuxS.\(^\text{19}\) .................................................118

**Scheme 5.3** Illustration of the structures of the 2-keto-ribose intermediate and different substrate and substrate-analogues (see text). .................................................................120

**Scheme 6.1** The overall mutase reaction catalyzed by UDP-galactopyranose mutase (UGM). .......................................................................................................................152

**Scheme 6.2** The catalytic mechanism of UGM as proposed by Soltero-Higgin \textit{et al.}\(^\text{10}\) and Gruber \textit{et al.}\(^\text{6}\) involving formation of a covalent bond between the N\(_{5}\text{FAD}\) centre of FAD and the anomic carbon (C\(_1\text{Gal}\)) of UDP-galactopyranose..............................154

**Scheme 6.3** Schematic summary of the reactant, intermediates and product complexes obtained for the overall mechanism of UGM using the DFT-cluster approach (see text). ..........................................................................................................................170

**Scheme 7.1** General catalytic mechanism of aaRS's in which the substrates pro-S non-bridging oxygen acts as the mechanistic base.\(^\text{16}\) .....................................................................................187

**Scheme 7.2** Proposed mechanism of ThrRS.\(^\text{21}\) ...............................................................189
**Scheme 7.3** Schematic illustration of the two ThrRS active site models used in this study in which His309 (a) is neutral and acts as a base and (b) is protonated and acts as an acid.
List of Tables

Table 2.1 Force field examples.\textsuperscript{5}.................................................................27

Table 2.2 The characterization and the nature of four types of critical points........35

Table 3.1 Proton affinities and basicities (in kJ mol\textsuperscript{-1}) of the side chain of Lys204, modelled as ethylamine, within environments of different polarity...............................56

Table 4.1 B3LYP/6-31G(d) Second-Order Perturbation Estimates (in kJ mol\textsuperscript{-1}) of Hyperconjugative Interactions in Selected Equilibrium Structures of $\alpha$Glc1Me.....101

Table 4.2 B3LYP/6-31G(d) Second-Order Perturbation Estimates (in kJ mol\textsuperscript{-1}) of Hyperconjugative Interactions in Selected Equilibrium Structures of $\beta$Glc1Me.....105

Table 7.1 Values of the Electron Density ($\rho$) and its Laplacian ($\nabla^2\rho$) (a.u.) at BCPs of Selected Hydrogen Bonds Within the Fully Bound Reactive Complexes \textsuperscript{a}RC and \textsuperscript{b}RC.\textsuperscript{a}........................................................................................................................................198
List of Appendices (on CD)

Appendix A

Figure A1. Potential energy surface for the overall catalytic mechanism of UDPGlcDH.

Table A1. Full citation for Gaussian 03 (reference 21).

Table A2. Optimized cartesian coordinates obtained at the B3LYP/6-31G(d) level of theory and high level single point [$\Delta E^\text{SP} = \text{B3LYP/6-311+G}(2df,p)/\text{B3LYP/6-31G}(d) + \Delta E^\text{ZPVE}(\text{B3LYP/6-31G}(d))$] and solution energy corrections ($\Delta E^\text{env}$) calculated at the IEFPCM-B3LYP/6-311+G(2df,p)/B3LYP/6-31G(d) level of theory with $\varepsilon = 4$.

Appendix B

Table B1: Cremer–Pople Puckering Parameters for the Pyranoid Ring in the Active-Site Model Along the Reaction Coordinate.

Table B2: B3LYP/6-31G(d) Second-Order Perturbation Estimates (in kJ mol$^{-1}$) of Hyperconjugative Interactions in Selected Equilibrium Structures of $\beta$Glc1Me.

Reference B1: Full citation for the Gaussian 03 program.

Table B3: Cartesian coordinates of optimized structures at the B3LYP/6-31G(d) 5D level of theory. IEF-PCM($\varepsilon=4.0$)/B3LYP/6-311+G(2df,p)/B3LYP/6-31G(d) 5D single-point energies calculated for each equilibrium structure.

Appendix C

Table C1: MD annealing protocol.

Table C2: Cartesian coordinates of optimized structures at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory. Single point energies were obtained at ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory.
Appendix D

Figure D1: PES for the catalytic mechanism of UGM (with protonation His60) obtained at the IEFPCM ($\varepsilon=4.0$) B3LYP/6-311+G(2d,p)//B3LYP/6-31G(d) level of theory.

Figure D2: FADH$^+$ numbering.

Table D1: Active site structure differences between *Klebsiella pneumoniae* UGM (PDB ID: 3INT) and *Deinococcus radiodurans* UGM (PDB ID: 3HDQ)

Table D2: Cartesian coordinates of optimized structures at the B3LYP/6-31G(d) level of theory. Long-range effects due to the polarity of the protein environment were calculated with the IEF-PCM using a dielectric constant of 4.0 at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory for all equilibrium structures.

Table D3: Cartesian coordinates of optimized structures at the ONIOM(B3LYP/6-31G(d):AMBER96) level of theory. Single point energies were obtained at ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory. Long-range effects due to the polarity of the protein environment were calculated using electric embedding at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)=Embedcharge level of theory for all equilibrium structures.

Appendix E

Figure E1: Formation of the tetrahedral transition state and intermediate in the protonated His309-assisted model with Lys465.

Table E1: Cartesian coordinates of optimized structures at the B3LYP/6-31G(d,p) level of theory. Long-range effects due to the polarity of the protein environment were calculated with the IEF-PCM using a dielectric constant of 4.0 at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d,p) level of theory for all equilibrium structures.
# List of Abbreviations and Symbols

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaRS</td>
<td>Aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>AIM</td>
<td>Atoms In Molecules</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine Monophosphate</td>
</tr>
<tr>
<td>B</td>
<td>Becke’s 88 exchange functional</td>
</tr>
<tr>
<td>BCP</td>
<td>Bond Critical Point</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke’s three parameter hybrid functional</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalized Gradient Approximation</td>
</tr>
<tr>
<td>Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamate</td>
</tr>
<tr>
<td>GTF</td>
<td>Gaussian-Type Functions</td>
</tr>
<tr>
<td>HDV</td>
<td>Hepatitis Delta Virus</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>IC</td>
<td>Intermediate Complex</td>
</tr>
<tr>
<td>IEF</td>
<td>Integral Formalism Equation</td>
</tr>
<tr>
<td>LCAO</td>
<td>Linear Combination of Atomic Orbitals</td>
</tr>
<tr>
<td>LDA</td>
<td>Local Density Approximation</td>
</tr>
<tr>
<td>LYP</td>
<td>Lee-Yang-Parr correlation functional</td>
</tr>
<tr>
<td>NBO</td>
<td>Natural Bond Orbital</td>
</tr>
<tr>
<td>PA</td>
<td>Proton Affinity</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>PC</td>
<td>Product Complex</td>
</tr>
<tr>
<td>PCM</td>
<td>Polarizable Continuum Method</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PES</td>
<td>Potential Energy Surface</td>
</tr>
<tr>
<td>RC</td>
<td>Reactant Complex</td>
</tr>
<tr>
<td>RHF</td>
<td>Restricted Hartree-Fock</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>SAC</td>
<td>Substrate-Assisted Catalysis</td>
</tr>
<tr>
<td>SCF</td>
<td>Self-Consistent Field</td>
</tr>
<tr>
<td>STO</td>
<td>Slater-Type Orbitals</td>
</tr>
<tr>
<td>TS</td>
<td>Transition Structure</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>ZPVE</td>
<td>Zero Point Vibrational Energy</td>
</tr>
</tbody>
</table>
\[ \hat{H} \quad \text{Hamiltonian operator} \]

\[ E \quad \text{Energy} \]

\[ \Psi \quad \text{Wave function} \]

\[ r \quad \text{Distance between two particles} \]

\[ \chi_i \quad \text{ith spin orbital} \]

\[ c_{ui} \quad \text{Molecular orbital coefficient} \]

\[ \phi \quad \text{Basis function} \]

\[ F_{uv} \quad \text{Fock matrix} \]

\[ S_{uv} \quad \text{Overlap matrix} \]

\[ \varepsilon_i \quad \text{Orbital energy for the ith molecular orbital} \]

\[ \lambda \quad \text{Dimensionless parameter} \]

\[ \rho \quad \text{Density} \]

\[ \nabla \quad \text{Laplacian} \]

\[ T_s[\rho(\bar{r})] \quad \text{Kinetic energy term} \]

\[ J[\rho(\bar{r})] \quad \text{Potential energy term} \]

\[ E_{\text{ne}}[\rho] \quad \text{External potential} \]

\[ E_{\text{xc}}[\rho(\bar{r})] \quad \text{Exchange-correlation energy term} \]

\[ \varepsilon \quad \text{Dielectric constant} \]

\[ \varepsilon \quad \text{Ellipticity} \]

\[ \text{Å} \quad \text{Angstrom} \]

\[ \circ \quad \text{Degree} \]

\[ \text{a.u.} \quad \text{Atomic unit} \]
Carbohydrates, "poly carbon hydrate molecules",¹ are the most abundant biomolecules in the world.¹ This is due in part to their wide range of diverse and important biochemical roles such as structural components or as 'energy molecules'.¹ In addition, they are often found to be important components in many key metabolites including coenzymes (e.g., NAD⁺, FADH) and information molecules (e.g., DNA, RNA and blood type);¹ Figure 1.1. Furthermore, they are increasingly finding an array of alternate roles such as components of therapeutic drugs.²

![Figures 1.1](image)

**Figure 1.1** Examples of carbohydrate derivatives with the sugar component highlighted in red: (a) cellulose, (b) uridine diphosphate glucose (UDP), and (c) a fragment of deoxyribonucleic acid (DNA).

This diversity of carbohydrate bio-functions is due in large part to their multiple possible configurations and conformations.¹ For example, a hexose sugar such as glucose contains a six-carbon backbone. However, it may occur in a linear (non-cyclic) or cyclic form as shown in Figure 1.2. The latter is formed when a hydroxyl group oxygen attached to the carbon backbone attacks at the carbonyl
carbon (which then is known as the anomeric carbon). The resulting rings are commonly either a pyranose (6-membered ring) or furanose (5-membered ring) and may occur as α or β-anomers depending on the orientation of the hydroxyl group on the anomeric carbon.

Figure 1.2 Glucose conformational interconversion reactions.

Glycosidic bonds lie at the core of carbohydrate chemistry and biochemistry. They are formed via condensation reactions that involve the hydroxyl group on the anomeric carbon with an appropriate functional group in some other molecule. For instance, they link sugar monosaccharides together to form polysaccharides. Alternatively, they connect the nucleobases of DNA and RNA to their deoxyribose and ribose sugar moieties, respectively (see Figure 1.1). Most importantly, they are often said to be the chemically strongest biopolymer linkage found in nature; that is, compared to amide bonds between amino acid residues in proteins and phosphodiester bonds between nucleotides in DNA and RNA. This is due to the fact that they are highly resistant to hydrolysis. It has been reported that the half-life of uncatalyzed O-glycosidic bond hydrolysis of cellulose in water at SATP is around 5 million years. In general, one must use strong acids, medium or high temperature
(up to 200 °C) in order to decrease glycosidic bond hydrolysis reaction times to approximately 24 h or less.\(^5\)

For cells and organisms it is essential that the breaking and making of such bonds occur at life-sustainable rates. For example, glycogen is the major metabolic fuel source and glycogenolysis is the metabolic pathway for converting it into glucose such that it is available for energy production in many organisms.\(^1\) Fortunately, however, nature has evolved a group of enzymes called glycosidases, which have the ability to cleave glycosidic bonds under the mild conditions found within organisms. These enzymes are amongst some of the catalytically effective ones known with rates up to \(10^{17}\) times faster than for the corresponding uncatalyzed reaction.\(^4\) There are also numerous other enzymes involved in glycobiological metabolism including epimerases (interconvert sugar epimers), mutases (interconvert sugar rings, e.g., pyranose to furanose), and sugar oxidoreductases (perform redox chemistry on sugars). Thus, studies on the properties and mechanisms of such enzymes can provide tremendous insights into not only carbohydrate biological functions but also, for example, aid new therapeutic drug developments and lead to industrial applications (e.g., herbicides or paper processing).\(^2,6-9\)

Studies on such enzymes can also provide insights into the fundamental chemical principles that underlie biocatalysis and catalysis in general.\(^1\) For instance, multiple key factors have been proposed as the origin or main source(s) of catalysts rate-enhancing abilities, including: transition-state stabilization, enzyme-substrate binding entropy stabilization, desolvation of the substrate, steric effects, electrostatic pre-organization of the active site, enzyme conformational dynamics, quantum mechanical effects such as tunneling, ionic low-barrier hydrogen bonds, and donor-acceptor matching \(pK_a\)s.\(^10\) Furthermore, the interactions between substrate(s), enzyme active site residues, coenzymes and cofactors (such as metal
ions) can strongly influence the reactivity of enzymes and their specificity towards their target carbohydrates.\textsuperscript{9} Enzymes may employ more than one of these approaches. Thus, it is often unclear or even unknown which of these factors predominates and why.

The main focus of this thesis is the application of computational and theoretical methods to investigate biochemical problems, in particular biocatalysts involved in glycobiology. It is noted that the application of such methods to biochemical problems has previously been referred to as ‘quantum biochemistry’.\textsuperscript{11} More specifically, a broad range of methods including Docking, Molecular Dynamics, Molecular Mechanics, Quantum Mechanics/Molecular Mechanics and Density Functional Theory have been used to examine many different enzymes that exhibit a diverse variety of mechanisms and thus potentially, catalytic approaches. In addition to providing insights into their mechanisms and properties, such studies could lead to, for example, the development of new and novel carbohydrate-based inhibitors for specific enzymes. The results of our computational investigations on carbohydrate-related enzymology are detailed in this thesis.
Chapter 1. Introduction

References

Chapter 2

Theoretical Methods
2.1 Introduction

Traditionally, chemistry was based upon physical experimentation, observation, and deduction. In the early twentieth century, however, quantum mechanics, the mathematics and physical laws governing particles such as electrons and atoms, was discovered and developed.¹ This has enabled an alternate approach to investigating and understanding chemical phenomena based upon the study and solution of the underlying mathematical principles of chemistry. The understanding of atomic structures, molecular bonding and chemical reactions through the use of quantum mechanics is known as quantum chemistry.² Since its development, it has enabled revolutionary discoveries in chemistry, and by extension biochemistry, both on its own and when complementarily combined with experiment.

In theory, the application and solution of the equations of quantum mechanics could be used to exactly solve all chemical problems. However, the calculations are impossible to solve for all but the simplest of systems. It is now common to use computers to solve the quantum mechanical equations applied to a particular chemical system. This is known as computational chemistry. Furthermore, a variety of computational methods have been developed based upon the invocation of usually mathematically rigorous and/or chemically intuitive simplifications, all aimed at obtaining an accurate approximation to the exact solution.³ Continuing developments in computer technology are enabling chemists to investigate ever larger systems. For example, currently, chemical models as large as thousands of atoms can be investigated using appropriate theoretical methods on parallel workstations. Importantly, these calculations can often accurately and reliably describe geometries and reactivities for both stable and reactive systems.³ As a result, computational chemistry has established itself as a valuable approach to the study of chemical and biochemical phenomena.
Because the computational methods used in this thesis are based upon quantum mechanics, a brief overview of the relevant underlying theories and approximations are presented in this chapter. It is noted, that more comprehensive discussions of the concepts and theories of quantum chemistry can be found in many excellent textbooks.²⁻ʰ

### 2.2 The Schrödinger Equation

The Schrödinger equation lies at the very heart of quantum mechanics and hence, computational chemistry. It should be noted that as most applications of quantum mechanics to chemistry problems involves obtaining approximate solutions to the time-independent Schrödinger equation, herein we simply begin from this equation. In its simplest expression, the Schrödinger equation can be written as shown in **equation (eq.) 2.1:**

\[
\hat{H} \psi(r) = E \psi(r)
\]  

where \(\hat{H}\) is the Hamiltonian operator, \(\psi(r)\) is the wave function of the system of interest with coordinates \((r)\), and \(E\) is the energy of the system described by the wave function. It should be noted that it is a fundamental postulate of quantum chemistry that \(\psi(r)\) contains all information we may want to know or could ever know about a chemical system; that is, it completely describes the system.

The Hamiltonian operator is the sum of two operators; one for the kinetic energy \((\hat{T})\) and the other for the potential energy \((\hat{V})\). This leads to expression of the Schrödinger equation shown in **eq. 2.2:**

\[
(\hat{T} + \hat{V}) \psi(r) = E \psi(r)
\]  

(2.2)
For molecules this can then be expanded even further to arrive at the expression (in atomic units) shown in eq. 2.3:

\[
\left( \sum_{i=1}^{N_e} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{N_n} \frac{1}{2M_A} \nabla_A^2 \right) + \left( \sum_{i=1}^{N_e} \sum_{A=1}^{N_n} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_e} \sum_{j=1, j \neq i}^{N_n} \frac{1}{r_{ij}} + \sum_{A=1}^{N_n} \sum_{B=1, B \neq A}^{N_n} \frac{Z_A Z_B}{r_{AB}} \right) \psi(r) = E \psi(r)
\] (2.3)

The first two single-summations are the kinetic energy of the electrons and nuclei respectively. The latter three double-summations are the potential energy terms for electron-nuclei, electron-electron, and nuclei-nuclei interactions, respectively. It should be noted that conventionally, \(i\) and \(j\) denote electrons, \(A\) and \(B\) denote nuclei, \(M_A\) is the mass of nuclei \(A\), \(Z_A\) is the charge of nuclei \(A\), and \(r_{xy}\) is the distance between the particles \(x\) and \(y\).

Except for hydrogen and hydrogen-like systems, the Schrödinger equation shown in eq. 2.3 is impossible to solve exactly. Therefore, as previously noted, approximations are required in order to simplify the equation such that it can be applied to most chemical systems.

### 2.3 The Variational Theorem

For a given chemical system described by \(\psi(r)\), if eq. 2.3 was exactly solved, the eigenvalue \((E)\) obtained would be the exact energy of the system; that is, \(E_{\text{exact}}\). Unfortunately, however, the exact wave function can rarely be known. Thus, usually, one must guess an approximate wave function \((\psi_{\text{approx}})\). Consequently, instead of obtaining \(E_{\text{exact}}\) we instead will calculate an approximate energy \((E_{\text{approx}})\) of the system. The relationship between \(E_{\text{exact}}\), \(\psi_{\text{approx}}\), and \(E_{\text{approx}}\) is expressed mathematically by the Variational Theorem and shown in eq. 2.4. Importantly, it shows that for any \(\psi_{\text{approx}}\), \(E_{\text{approx}}\) will be an upper bound to \(E_{\text{exact}}\).
$E_{\text{approx}} = \frac{\int \psi^*_{\text{approx}} \hat{H} \psi_{\text{approx}} d\tau}{\int \psi^*_{\text{approx}} \psi_{\text{approx}} d\tau} \geq E_{\text{exact}} \quad (2.4)$

Equation 2.4 is also referred to as the variational integral. The lower the energy obtained from the variational integral, the closer the trial variational function ($\psi_{\text{approx}}$) approaches the exact wave function of the system. In practice it has been observed that $E_{\text{approx}}$ approaches $E_{\text{exact}}$ much faster than $\psi_{\text{approx}}$ approaches $\psi_{\text{exact}}$. Computational methods which obey the variational theorem, i.e., give $E_{\text{approx}}$ as an upper bound to $E_{\text{exact}}$, are said to be variational methods. Unfortunately, however, despite the obvious power of this theorem, it does not help to simplify the intractable problem presented in eq. 2.3.

2.4 Further Simplification of the Schrödinger Equation

2.4.1 The Born-Oppenheimer (B-O) Approximation

In general, in chemistry we are interested in those properties of a molecular system determined by its electrons. The Born-Oppenheimer approximation can be stated as: the nuclei are much more massive than the electrons and consequently move much more slowly and therefore, from the perspective of the electrons, the nuclei appear stationary. This approximation is physically intuitive and can be shown mathematically to introduce little error to molecular structures and energies in most cases. Importantly, it enables us to effectively separate electronic and nuclear motions in a chemical system. As a result, the wave function of the chemical system, which is dependent on the coordinates of both the electrons ($q_i$) and the nuclei ($q_a$) can be written as a product of an electronic ($\psi_{\text{el}}$) and nuclear ($\psi_{\text{nu}}$) wave functions as shown in eq. 2.5:
\[ \psi(q_i, q_\alpha) = \psi_e(q_i) \psi_N(q_\alpha) \]

(2.5)

It should be noted that \( \psi_{el} \) is parametrically dependent on the nuclear coordinates. Thus, we need now focus our attention on only obtaining solutions for the electronic wave function.

The B-O approximation also helps to simplify the molecular Hamiltonian shown in eq. 2.3. In particular, as the nuclei are stationary, the kinetic energy term for nuclei equals zero and the potential energy term of nuclei-nuclei interactions is a constant \( (V_{NN}) \) for any nuclear configuration. Hence, \( V_{NN} \) can be omitted from the Hamiltonian without affecting the wave functions. The resulting simplified Hamiltonian operator shown in eq. 2.6 is known as the electronic Hamiltonian \( (\hat{H}_{el}) \) and is used to obtain the electronic energy \( (E_{el}) \) of a system.

\[ \hat{H}_{el} = -\sum_{i=1}^{N_e} \frac{1}{2} V_i^2 - \sum_{i=1}^{N_e} \sum_{\alpha=1}^{N_N} Z_\alpha \frac{1}{r_{i\alpha}} + \sum_{i=1}^{N_e} \sum_{j>i} \frac{1}{r_{ij}} \]

(2.6)

Thus, the overall energy of a chemical system \( (U) \) becomes the sum of \( E_{el} \) and the internuclear repulsion energy \( V_{NN} \) as shown in eq. 2.7.

\[ U = E_{el} + V_{NN} \]

(2.7)

While this approximation has simplified the molecular Hamiltonian to a three-term electronic Hamiltonian, the corresponding electronic Schrödinger equation is still impossible to solve for multi-electron chemical systems due to the electron-electron interaction term, whereas the other two terms only ever depend upon one-electron at a time.
2.4.2 The Orbital Approximation

While the variational theorem provides a way to establish the quality of a wave function, in reality, it is impossible to search all acceptable N-electron wave functions to solve for the energy of an N-electron system using the Schrödinger equation. The electron-electron repulsion term causes the electronic Schrödinger equation to be inseparable for multi-electronic systems. In the orbital approximation, this term is essentially ignored. More specifically, it states that the motions of the electrons are independent of each other. As a result, the best approximate wave function of the N-electron system ($\psi_{\text{total}}$) can be written in as a product of N one-electron orbitals $\psi_i$. The simplest form of this type of wave function is a Hartree-product shown in eq. 2.8.

$$\psi_{\text{total}} = \psi_1 \psi_2 \cdots \psi_N \tag{2.8}$$

However, the Hartree-product is not an acceptable wave function due to the fact that electrons are not distinguishable between each other. This is a consequence of the uncertainty principle. Furthermore, the Pauli principle states that an electronic wave function must be antisymmetric with respect to interchange of any two electrons; that is, it must change sign. As a result, the electronic wave function ($\psi_{\text{total}}$) of an N-electron system must in fact be expressed as an antisymmetric linear combination of products of N one-electron orbitals. This is commonly written in the form of a determinant of a matrix known as a Slater determinant or $\Phi_{\text{sd}}$ (eq. 2.9):
Chapter 2. Theoretical Methods

\[ \psi_{\text{total}} = \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & \cdots & \chi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & \cdots & \chi_N(\vec{x}_N) \end{bmatrix} \] (2.9)

The \( \frac{1}{\sqrt{N!}} \) factor is known as a normalization constant. When spin is also taken into account, we use one-electron spin-orbitals (\( \chi_i(\vec{x}_i) \)) as shown in eq. 2.9. For a molecular system these orbitals are the product of a one-electron spatial orbital \( \psi_i \) (in this case a one-electron molecular orbital (MO)) and a spin function. It should be noted that for an atomic system the MO is simply replaced by an atomic orbital (e.g., s, p, d etc.).

The spatial orbital (\( \psi_i \)) components in the Slater determinant can be further represented as linear combinations of a complete set of known basis functions, \( \phi_\mu \) (eq. 2.10). Roothaan proposed this expansion in 1951. When the atomic orbitals of the constituent atoms of a molecular system are used as the basis functions, eq. 2.10 is called a linear combination of atomic orbitals (LCAO).

\[ \psi_i = \sum_{\mu=1}^{N} c_\mu \phi_\mu \] (2.10)

2.5 The Hartree-Fock Self-Consistent Method

The Hartree-Fock (HF) self-consistent-field (SCF) method is the most common procedure in computational chemistry for finding the best approximate wave function for a multi-electron system. At the core of this method is the approximation that each electron only ‘sees’ a smeared electronic charge density resulting from the average effects of all other electrons. The initial guess wave function for say the ground-state of the chemical system of interest is written as an
antisymmetrized Slater determinant of spin-orbitals, in which the spatial components have themselves been written as a linear combination of basis functions. Based upon the variational theorem, we can then find the set of coefficients that describe the spatial orbitals and together give the best \( \Phi_{SD} \). That is, we will find the \( \Phi_{SD} \) that gives closest agreement to the exact ground-state of the target system. This calculated approximate energy is commonly referred to as the HF energy (\( E_{HF} \)) of the system. Mathematically, for a closed-shell system in which all electrons are paired, it is commonly expressed as shown in eq. 2.11 (\( V_{NN} \) has been included for completeness).

\[
E_{HF} = 2 \sum_{i=0}^{n/2} H_{ii}^{\text{core}} + \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} (2J_{ij} - K_{ij}) + V_{NN}
\]

The \( J_{ij} \) and \( K_{ij} \) terms in the above equation are referred to as the coulombic and exchange energy terms, respectively. The former results from the coulombic interactions of each electron with all of the others in the system. The latter, however, has no classical analogy and results from the fact that the wave function is a linear combination of products that differ from each other by exchange of electrons. However, in the HF procedure these two energy terms arise from the \( \hat{J} \) and \( \hat{K} \) operators, respectively. Notably, these operators themselves depend upon the very wave functions for which we seek to solve.

**2.5.1 Self-Consistent-Field Process**

As noted above, in order to feasibly calculate accurate SCF wave functions, the spatial MOs in \( \Phi_{SD} \) are expanded as linear combinations of a complete set of one-electron basis functions (eq. 2.10). In the HF method, for a closed-shell system this leads to the Roothaan-Hall equation shown in eq. 2.12:
\[ \sum_{\mu=1}^{n} c_{\mu}(F_{\mu\nu} - \varepsilon_{\nu} S_{\mu\nu}) = 0 \quad (2.12) \]

The \( c_{\mu} \) terms are the coefficients in eq. 2.10, for which we are trying to solve. The \( \varepsilon_{\nu} \) term represents the orbital energies of the chemical system we are considering, while the \( F_{\mu\nu} \) and \( S_{\mu\nu} \) terms represent the Fock matrix elements and the overlap matrix elements, respectively. In particular, it should be noted that the Fock matrix contains the coulombic and exchange terms within it and also depends on \( c_{\mu} \).

In order to solve this equation, we begin by guessing a set of coefficients for each spatial orbital (eq. 2.10). This initial set is used to construct the \( F_{\mu\nu} \) elements and to calculate the corresponding orbital energies, \( \varepsilon_{\nu} \). This enables us to solve eq. 2.12 to obtain an improved set of coefficients. This improved set is then used to compute an improved Fock operator and corresponding orbital energies. Then, using these improved values eq. 2.12 is then solved to obtain a further improved set of coefficients, and so on. This iterative process continues until the changes in the coefficients is zero or, more practically, negligible between two subsequent cycles. That is, it continues until a self-consistent field is achieved.

It may be noticed that the basis functions used in the iteration are also very important for the optimization process. Ideally, a complete set of atomic orbitals (AOs) basis functions should be used in the LCAO process. In reality, we must consider a balance between accuracy and computational resources. A complete basis set is usually not necessary in order to obtain accurate energies or geometries for models in computational calculations. Please refer to the basis set section for more information.
2.5.2 Electron Correlation

A key issue with the HF method, however, is that we ignored the instantaneous interactions between electrons, with each electron only feeling the average effects of the other electrons. It is said that the HF method ignores electron correlation. The energy difference between exact energy ($E_0$) of an $N$-electron system and the HF energy ($E_{HF}$) is called the correlation energy ($E_{corr}$) as illustrated in eq. 2.13:

\[ E_{corr} = E_0 - E_{HF} \]  \hspace{1cm} (2.13)

The effects of electron correlation can be important for accurately describing, for example, bond forming and breaking processes in chemical reactions. Therefore, a variety of computational approaches have been developed that aim to include electron correlation, e.g., Møller-Plesset second order perturbation theory (MP2), configuration interaction (CI), configuration interaction single and double excitations (CISD), and coupled cluster method (CC).

2.6 Basis Sets

A basis set is the set of functions, referred to as basis functions, used to construct the molecular orbitals of the chemical system. Two common types of basis functions are Slater-type atomic orbitals (STOs) and Gaussian-type functions (GTFs). STOs provide more accurate descriptions of atomic orbitals and hence the total wave function than GTFs. However, computationally they are more expensive than GTFs; i.e., requiring more time and computer resources. GTFs also have the added benefit that a combination of GTFs generates a new GTF. Hence, a combination of such functions can be used to obtain a good approximation to a corresponding STO,
and still remain computationally cheaper than using a single STO. One can re-express eq. 2.10 to highlight the relationship between an 'improved' GTF, or contracted basis function, and a linear combination of 'primitive' GTF's (eq. 2.14):

\[
\phi_{\mu} = \sum_k d_{\mu k} g_k
\]

In eq. 2.14 \( \phi_{\mu} \) is the GTF contracted basis function, \( g_k \) denotes the primitive GTFs, and \( d_{\mu k} \) is the coefficient, or weighting, of the corresponding primitive in the combination for the \( \mu \)th contracted basis function. When the number of basis functions used in a given computational model are discussed, it usually refers to the number of contracted Gaussian functions used.

In computational chemistry, there are several common types of defined basis sets. A number of these have been used throughout this thesis and are briefly discussed below.

### 2.6.1 Minimal Basis Set

In a minimal basis set only those atomic orbitals that are occupied in the constituent atoms of the chemical system are included. For example, in an STO-3G basis set each occupied atomic orbital is described by a single STO-type orbital which is itself approximated by 3 GTF's. An advantage of using such a small basis set is the speed of calculations. However, the reduced accuracy and reliability of the optimized geometries and thermochemical data obtained limits its use.

### 2.6.2 Split-Valence Basis Sets

In split-valence basis sets, two or more basis functions are used to describe each valence orbital. This enables one to more accurately allow for the fact that the
radial size of the atomic valence orbitals are influenced by the environment in which the atom is situated. Common examples of split-valence basis sets include the double-zeta- and triple-zeta-valence basis sets 6-31G and 6-311G, respectively. In both cases the core orbitals are described by a single basis function constructed from a linear combination of 6 primitive GTFs. However, for the 6-31G basis set the valence atomic orbitals are described by two basis functions; one constructed from 3 primitives and the other from 1 primitive. For the 6-311G basis set, each atomic valence orbital is described by 3 basis functions; one constructed from 3 primitives and the other two from 1 primitive each.

2.6.3 Polarization Basis Sets

In many chemical environments, e.g., within a molecule, the atomic environment leads to an anisotropic 'polarizing' of its electrons. For example in a C—N bond, the electrons on the carbon will be displaced towards the nitrogen. In order to accurately describe this effect, one must include unoccupied orbitals of higher-angular momentum. For example, for carbon one typically includes $d$- and/or $f$-functions, while for hydrogen one includes $p$- and/or $d$-functions. The resulting basis sets are known as polarization basis sets. A common example of such a basis set is 6-31G(d,p), which includes a set of $d$-functions on each heavy atom such as carbon, and a set of $p$-functions on each hydrogen.

2.6.4 Diffuse Functions

Diffuse functions are spatially large functions that are used to better describe electrons moving to regions that are far from the nucleus. Such functions can be important for describing, for example, anions and long-range interactions. In Pople basis sets, their inclusion in a basis set is denoted by a "+" symbol. For example, in
the 6-31++G(d,p) basis set, a set of diffuse functions has been included on each heavy-atom (e.g., C) and each hydrogen or helium atom.

### 2.7 Hierarchy of *ab initio* Methods

In *ab initio* theory, the exact solution of the Schrödinger equation representing a chemical system can in theory be achieved by increasing both the extent of inclusion of electron correlation and the size of the basis set used. This relationship is nicely illustrated in the Pople diagram shown in Figure 2.1. Naturally, as the size of basis set or the extent of inclusion of electron correlation is increased, so too does the cost of the calculation. Thus, when undertaking a computational investigation, we must make an acceptable compromise between accuracy and computational cost.

![Diagram](image)

**Figure 2.1** A Pople diagram illustrating the relationship between basis set size, extent of inclusion of electron correlation, and the exact solution of the Schrödinger equation.
Chapter 2. Theoretical Methods

The 'cheapest' wave function-based electron correlation method is the Møller-Plesset second order perturbation theory (MP2) method. Unfortunately, it can generally only be routinely applied to chemical systems containing up to approximately 20 non-hydrogen atoms. However, for the study of biochemical systems, it is often necessary to use chemical models that include several hundred non-hydrogen atoms. Fortunately, there are alternative methods now available that include electron correlation effects.

2.8 Density Functional Theory

As previously noted, the wave function of a chemical system can provide all information about the ground state of a particular system. However, accurate calculations that include the effects of electron correlation can become prohibitively expensive.

Density functional theory (DFT) provides an alternative approach for investigating multi-electron systems. Unlike wave function-based methods, DFT methods are based upon the electron density of a system. Such an approach has several mathematical, physical and chemical advantages. First, unlike the wave function which depends on 4N variables (3 spatial and one spin variable) per electron, the density is only dependent on 3 variables \( \{x, y, z\} \) regardless of the number of electrons. In addition, unlike the wave function, the electron density can be measured experimentally. Hence, potentially, the accuracy of a calculated density can be experimentally measured. Furthermore, DFT calculations inherently include electron correlation effects. Thus, in theory, it is able to provide reliable and accurate results that are often comparable to high level wave function-based \textit{ab initio} correlation methods, yet at the same time, at significantly reduced computational cost.
2.8.1 Hohenberg-Kohn Theorem

The Hohenberg-Kohn theorem provides the connection for a given chemical system between its ground state energy and its electron density $\rho(\vec{r})$. It states that for ground state of a chemical system, the external potential, also denoted by $\rho_0(\vec{r})V_{\text{Ne}}\,d\vec{r}$, is a unique functional of the electron density $\rho(\vec{r})$. That is, from the density we can determine all properties of the ground-state of a molecule. For example, the energy of a system ($E_0$) is thus a functional (a function of a function) of its density. Mathematically, this relationship can be expressed as in eq. 2.15:

$$E_0 = E_0[\rho_0] = \int \rho_0(\vec{r})V_{\text{Ne}}\,d\vec{r} + T[\rho_0] + E_{ee}[\rho_0]$$ (2.15)

In the above equation, $T[\rho_0]$ and $E_{ee}[\rho_0]$ represent the kinetic energy and the electron-electron energy functionals, respectively. Both of them are universally valid terms; that is, their mode of calculation is independent of the number of electrons ($N$), the positions of electrons with respect to the nuclei $R_A$, and the nuclear charge ($Z_A$) of a chemical system. In contrast, the $\int \rho_0(\vec{r})V_{\text{Ne}}\,d\vec{r}$ term is a system dependent term.

Importantly, this theorem also obeys the variational theorem. That is, for any guess electron density function ($\tilde{\rho}$) the calculated energy will always be greater than or equal to the exact ground-state energy as shown in eq. 2.16:

$$E_0 \leq E[\tilde{\rho}] = \int \tilde{\rho}(\vec{r})V_{\text{Ne}}\,d\vec{r} + T[\tilde{\rho}] + E_{ee}[\tilde{\rho}]$$ (2.16)

While the Hohenberg-Kohn Theorem demonstrates the theoretical possibility of density functional theory, it does not know how to fully solve the required equations.
2.8.2 The Kohn-Sham Approach

Fortunately, the Kohn-Sham approach shows how one may approach a solution of the underlying equations of density functional theory. Similar to the orbital approximation used in HF theory and associated wave function-based methods (Section 2.4.2), it shows that one can separate a molecules N-electron density into N one-electron density packets or orbitals. These are known as Kohn-Sham (KS) orbitals and are herein denoted by $\psi_{i}^{KS}(\mathbf{x})$. The total density calculated from the summation of the square of all KS orbitals should equal to the ground state density of the real system with interacting electrons and is shown in eq. 2.17:

$$\rho_{s}(\mathbf{r}) = \sum_{i}^{N} \sum_{s} |\psi_{i}^{KS}(\mathbf{r},s)|^{2} = \rho_{0}(\mathbf{r}) \quad \text{(2.17)}$$

where $\psi_{i}^{KS}(\mathbf{r},s)$ is spin orbital obtained from the Kohn-Sham orbital, $\rho_{s}(\mathbf{r})$ is the electron density of the non-interacting system, and $\rho_{0}(\mathbf{r})$ is the ground state density.

Again, similar to HF theory, one can construct a Slater determinant $\Phi_{SD}$ (Section 2.4.2) from the set of Kohn-Sham orbitals. Furthermore, however, it shows that the ground state energy $E[\rho(\mathbf{r})]$ of a molecular system can be represented as shown in eq. 2.18:

$$E[\rho(\mathbf{r})] = T_{s}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + E_{xc}[\rho] \quad \text{(2.18)}$$

where $T_{s}[\rho(\mathbf{r})]$ represents the kinetic energy of the non-interacting system, $J[\rho(\mathbf{r})]$ represents the electron-electron repulsion energy, and $E_{xc}[\rho]$ is the external potential. Importantly, $E_{xc}[\rho(\mathbf{r})]$, the exchange-correlation energy functional,
includes all terms that cannot be accurately obtained. In order to simplify the
calculation, it is usually calculated as the sum of an exchange term $E_x[\rho(\mathbf{r})]$ and
correlation term $E_c[\rho(\mathbf{r})]$ as shown in eq. 2.19.

$$E_{xc}[\rho(\mathbf{r})] = E_x[\rho(\mathbf{r})] + E_c[\rho(\mathbf{r})] \quad (2.19)$$

Unfortunately, as noted above, we do not know the exact form of the
$E_{xc}[\rho(\mathbf{r})]$ term. As a result, a variety of approximate functionals for the exchange
and correlation terms have been developed based upon different approximations,
e.g., inclusion of local (e.g., the value of the density at a given point) and/or non-local
(e.g., the gradient of the electron density) effects.

Some of the perhaps simplest DFT methods applied in chemistry are based
upon the local density (LDA) and local spin-density approximations (LSDA).\textsuperscript{8} In this
approximation, the electron density is essentially treated as a homogeneous electron
gas; it is constant, or slowly varying, everywhere. Obviously, this approximation is
far from the situation that occurs in many chemical systems. However, it is the only
system whose exchange and correlation energy functionals are known exactly.\textsuperscript{8}

A set of functionals has been developed based upon the generalized gradient
approximation (GGA).\textsuperscript{8} More specifically, they take into account the gradient of
electron density $\nabla \rho(\mathbf{r})$. As a result, this enables them to describe the non-
homogeneity of the true electron density. However, compared to LDA, it leads to the
loss of most of the physical meaning of the exchange-correlation hole. Furthermore,
it can give calculation results that are in poorer agreement with experiment than
LDA.\textsuperscript{8}

An alternative approach to enhance the accuracy of exchange-correlation
functionals is the development of hybrid functionals. These are based upon the
observation that in the overall exchange-correlation functional, the exchange
contribution has a much larger absolute value than the corresponding correlation
effects. Hence, a strategy that has proved highly successful for obtaining more
accurate and meaningful results is to include the exact HF exchange functional as
illustrated in eq. 2.20.\(^8\)

\[ E_{XC} = E_{X}^{HF}_{exact} + E_{C}^{KS} \]  \hspace{0.5cm} (2.20)

The correlation component of eq. 2.20, \(E_{C}^{KS}\), is always an approximate functional
expressing all missing parts in calculations. One of the most widely used hybrid
functionals in DFT is the Becke 3-parameter Lee-Yang-Parr (B3LYP) exchange-
correlation hybrid functional,\(^8\) which is often expressed (and implemented in
Gaussian 03 and 09)\(^9,10\) as shown in eq. 2.21:

\[ E_{XC}^{B3LYP} = (1 - a)E_{C}^{LDA} + aE_{X}^{HF}_{exact} + bE_{X}^{BS88} + cE_{C}^{LYP} + (1 - c)E_{C}^{LDA} \]  \hspace{0.5cm} (2.21)

This functional has been proven to successfully predict the properties of
many chemical and biochemical systems and as a result, is the most common DFT
functional used in this thesis.\(^11,12\)

### 2.9 Molecular Mechanics

Molecular mechanics (MM) methods can be simply described as the use of
classical mechanical ball-spring-type chemical models to calculate the energies and
geometries of chemical systems.\(^3\) Empirical parameter-based force fields are used in
the construction of the associated equations. Most force fields include four terms:
- bond stretching, angle bending, bond rotation, and non-bonded interactions
  (including electrostatic interactions and Van der Waals interactions). For example, a
commonly used expression for the energy is shown in \textbf{eq. 2.22} in both a simplified and expanded format.

\[ E_{\text{total}} = \sum_{\text{bonds}} E_{\text{stretch}} + \sum_{\text{angles}} E_{\text{bend}} + \sum_{\text{dihedrals}} E_{\text{torsion}} + \sum_{\text{pairs}} E_{\text{non-bonded}} \]  \hspace{1cm} (2.22)

\[ E_{\text{total}} = \sum_{\text{bonds}} \frac{1}{2} k_b (l - l_{eq})^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{\text{pairs}} \left[ \frac{A_{ij}}{R_{ij}^2} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{eR_{ij}} \right] \]

The first term in both expressions calculates the energy component that arises from distortion of a bond of length \( l \) from its ideal equilibrium value (\( l_{eq} \)). The parameters \( k_b \) (the force constant for the bond stretch) and \( l_{eq} \) are commonly derived from \textit{ab initio} calculations or experimental data. Similarly, for solution of the other expressions an appropriate mathematical formula is chosen and empirical parameters used. MM calculations can provide, for example, reliable relative conformational energies at a fraction of computer time of \textit{ab initio} or DFT calculations.\(^3\)

In addition to the four terms noted above, some force fields also include cross terms, terms that model the coupling between internal coordinates (e.g., bond stretch with angle bending), to get more accurate results. In addition, the polarization of the charge density change due to surrounding polar environment or molecules, also called many-body effect, can also be taken into account.

As with any such empirical-based methods, it is important that one choose a force field that has been parameterized for the types of molecules you are investigating.\(^3\) The more your molecule or system differs from those used in parameterization of the force field, the more likely it is fail to provide accurate and reliable structures and energies. A list of commonly used force fields and the type of
chemical systems to which they can generally be reliably applied is shown in Table 2.1.

A major drawback of MM methods, however, is their inability to describe electronic properties. As a result, their calculated relative energies are usually not as accurate as those obtained from QM methods. Furthermore, they also cannot be used to investigate bond forming and breaking processes nor, by extension, transition structures.³

**Table 2.1 Force field examples.⁵**

<table>
<thead>
<tr>
<th>Name</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM</td>
<td>Biomolecules</td>
</tr>
<tr>
<td>CHARMm</td>
<td>Biomolecules and organics</td>
</tr>
<tr>
<td>MM2</td>
<td>Organics</td>
</tr>
<tr>
<td>MM3</td>
<td>Biomolecules and organics</td>
</tr>
<tr>
<td>MM4</td>
<td>Hydrocarbons</td>
</tr>
<tr>
<td>MMFF</td>
<td>Biomolecules and organics</td>
</tr>
<tr>
<td>PEP95SAC</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td>UFF</td>
<td>General</td>
</tr>
<tr>
<td>AMBER</td>
<td>Proteins and nucleic acids</td>
</tr>
</tbody>
</table>

The AMBER force field is specifically developed for proteins and nucleic acids.¹³ It has been successfully used in investigations of a variety of enzymatic systems.¹⁴ In this thesis, it is used in docking and molecular dynamics (MD) simulations for obtaining reliable initial enzyme and enzyme-substrate structures. In addition, we have also used it as the MM method of choice in the ONIOM QM/MM calculations, as it often able to reliably describe the protein environment. The
mathematical expression for the energy as calculated using the AMBER force field is shown in eq. 2.22.

### 2.10 Molecular Dynamics

In molecular dynamics (MD), Newton’s laws of motion under specified thermodynamic conditions, are used to express the position and velocities of particles in the computational model with respect to time.\(^\text{15}\) That is, one can examine time-dependent dynamic behavior of a chemical system over a chosen period of time, and under different thermodynamic conditions. In particular, it can be used to enable chemical models to sample conformational space and, for example, help in determination of the most likely lowest energy conformer.

In MD simulations, one generally chooses from one of 3 common types of models for describing the chemical system which can be summarized as:\(^\text{15}\)

(a) simple: the atoms are treated as hard-sphere particles in a vacuum that move in straight lines between collisions.

(b) continuous potential: the motion of particles are coupled with each other and influenced by a continuous potential in the space.

(c) motion-dependent model: a particle’s position and velocity are related to those of all other particles in the system.

For MD simulations on large protein systems, the continuous potential model is often used as it provides a suitable compromise between accuracy and computational cost.

When doing MD simulations, it is important to choose an appropriate time-step in order to obtain clear particle trajectories and to provide an adequate sample of conformational space. For chemical systems it is common to use a time-step of 1-2 ps, one order of magnitude smaller than the time for a C-H bond vibration.\(^\text{15}\) By
iteratively repeating this process over a period of simulating time, e.g., several nanoseconds,\textsuperscript{16} the chemical system is able to gradually reach its equilibrium conformation. This equilibrium state can be defined as when the potential and kinetic energies fluctuate in equal and opposite directions in the plot of their root-mean-square deviations against the time step.\textsuperscript{15}

Once a stable state has been reached, one can then analyze the MD simulations and conformations sampled in order to obtain insights into the equilibrium structure of the chemical system and such processes as substrate binding, hydrogen bond interactions, and solvation effects.\textsuperscript{16,17}

In this thesis, MD simulations have generally been performed in order to obtain important insights into initial substrate-bound enzyme active site conformations as well as substrate-enzyme interactions. For example, it has enabled us to use X-ray crystallographic structures in order to construct and obtain equilibrium structures of catalytically competent solvated enzyme-substrate complexes. In other words, it enables us to obtain a more reliable model of the protein in aqueous solution than might otherwise be provided by the X-ray crystal structure. Such MD structures are often a reasonable and even preferred choice for constructing initial chemical models for use in DFT or QM/MM studies on enzymes.\textsuperscript{6,15,16}

### 2.11 Quantum Mechanics/Molecular Mechanics (QM/MM) Methods

As noted above, while the MM method is suited for use with very large chemical models, it is not able to describe bond making and breaking processes. Conversely, QM methods can only be applied to comparatively modestly sized chemical models; 100-200 atoms for DFT methods. But, they can be used to describe
bond making and breaking processes and electronic properties (e.g., spectra). The strengths of these two methods have been exploited in the development of the hybrid quantum mechanics/molecular mechanics (QM/MM) methods. Such methods are well-suited to provide a balance between accuracy and computational cost for large biochemical systems. A major advantage of this approach is that one can use different computational methods to describe different chemical regions, within one model. This is illustrated in Figure 2.2.

![Figure 2.2 Illustration of a two-layer QM/MM model: the atoms shown in ball and stick format are in the high-layer (QM region) while those shown in wire format are in the low-layer (MM region).](image)

This approach has shown itself to be particularly useful in the study of enzymatic processes. The reactive region within an active site, the region in which bonds are being made and broken, is often fairly localized to several adjacent functional groups or molecules. Hence, it is well suited to be described using a QM method and is commonly referred to as the high-layer. In contrast, the protein environment surrounding the active site is often important for, for example,
maintaining the positioning of the substrate(s) and active site residues, and providing a non-homogeneous polar field. Thus, environmental effects from these outer residues can be modeled using a lower level of theory such as MM, and is commonly referred to as the low-layer.

For such hybrid methods, there are two common approaches for calculating the energy of the system: (a) additive and (b) subtractive. In the additive approach the energies of the high-layer (QM), the low-layer (MM), and the interface between these two regions is calculated and then summed together. In the alternate subtractive approach, mathematically represented in eq. 2.23, the QM energy of the high-layer and the MM energy of the overall system are calculated and added together. From this sum is subtracted the MM energy of chemical model used in the QM region. An example of the energy subtraction approach is the ONIOM QM/MM method as available in the Gaussian09 software package. This method has been used in this thesis.

\[
E^{\text{ONIOM}} = E^{\text{Model, High}} + E^{\text{Real, Low}} - E^{\text{Model, Low}}
\]  

(2.23)

where \( E^{\text{Model, High}} \) refers to the energy of the high layer calculated by a QM method, \( E^{\text{Real, Low}} \) refers to the energy of the full system calculated by a MM method, and \( E^{\text{Model, Low}} \) refers to the high layer calculated by a MM method.

In QM/MM methods, the question of how to properly modify and describe the boundary between the QM and MM layers has been a challenging problem. Several general rules have been developed to minimize the influence of boundary choice on the accuracy of the computational investigations: (i) the boundary can occur at non-polar single bonds, (ii) the boundary should be at least 3 bonds away from the reaction center, and (iii) the use of a link atom approach.
In the ONIOM approach, if the boundary occurs at a non-polar bond, the bond is simply replaced by an X–H bond in the QM layer. This ensures that there are no open valences in the QM region. However, if one is not careful, this can cause unrealistic energy changes in the QM/MM calculations.\textsuperscript{18,19} For example, the 'artificial' X–H bonds can cause over-polarization and large displacement problems. Fortunately, if a model is chosen properly (the QM region fully describes the bond breaking/forming processes while the MM region is not involved in any changes in the number and type of bonds), these systematic errors will cancel out in relative energy calculations of stationary points along a potential energy surface (PES).

### 2.12 Natural Bond Orbital (NBO) Method

Natural bond orbital (NBO) population analysis is a technique that allows one to study many-electron molecular systems using localized electron-pair bonding units.\textsuperscript{20} For example, the electron delocalization information obtained from an NBO analysis, can be used to provide insights into molecular stabilities and driving forces of reaction mechanisms. A molecular orbital in NBO analysis is the combination of the localized occupied Lewis orbitals $\sigma_{AB}$ (eq. 2.24) and unoccupied Lewis orbitals $\sigma_{AB}^*$ (eq. 2.25):\textsuperscript{21}

$$\sigma_{AB} = c_A h_A + c_B h_B \quad (2.24)$$

$$\sigma_{AB}^* = c_A h_A - c_B h_B \quad (2.25)$$

where $h_A$ and $h_B$ denote orthonormal hybrid orbitals. It should be noted that in the NBO approach, unlike in the case of the virtual (unoccupied) MO's in SCF-MO theory, the anti-bonding orbital has non-zero occupancy. Furthermore, the idealized natural Lewis structure commonly accounts for \(\geq99\%\) of the total electron density and so it
includes the preponderant portion of the correlation contributions from electron delocalization in a molecule.\textsuperscript{21,22} In a molecule the local molecular orbital (LMO) associated with a localized A–B bond may be written in NBO form as given in eq. \textbf{2.26}:

\[
\phi_{AB}^{\text{LMO}} \approx \sigma_{AB} + \lambda \sigma_{CD}^* + \cdots
\]  

(2.26)

The small contribution of the anti-bond $\sigma_{CD}^*$ represents the irreducible delocalization of $\sigma_{AB}$ from an idealized localized form due to hyper-conjugative non-covalent interactions.\textsuperscript{23}

\[
E = E_{\sigma\sigma} + E_{\sigma\sigma^*}
\]  

(2.27)

\[
\Delta E_{\sigma\sigma^*}^{(2)} = -2 \frac{\left\langle \sigma | \hat{F} | \sigma^* \right\rangle}{\epsilon_{\sigma^*} - \epsilon_{\sigma}}
\]  

(2.28)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.3}
\caption{Schematic illustration of the energies of the perturbative donor-acceptor interaction involving the occupied and unoccupied orbitals in the NBO method.}
\end{figure}

The total bond energy is calculated as the sum of the covalent energy of a localized covalent bond ($E_{\sigma\sigma}$) and the energy of non-covalent contributions ($E_{\sigma\sigma^*}$) as shown in eq. \textbf{2.27}. The latter $E_{\sigma\sigma^*}$ term is calculated using a second order
Chapter 2. Theoretical Methods

perturbation method and the energy of the occupied and unoccupied Lewis orbitals (eq. 2.28) and is schematically illustrated in Figure 2.3.21

2.13 Atoms In Molecules (AIM)

Atoms In Molecules (AIM) method can be used to analyse the topology of the computed electron density.24 For example, it is commonly used to identify inter- and intra-molecular interactions, and to characterize the nature of such interactions. There are several wonderful reviews and books that explain the details of the AIM theory.24-26 Hence, for the sake of brevity, only key aspects of AIM as related to my thesis are discussed herein.

2.13.1 Topology of the Electron Density

The electron density \( \rho(r) \) of a molecule changes as one moves around the molecule. Of particular interests are those points in space at which the first derivative of \( \rho(r) \), the gradient vector \( \nabla \rho(r) \) (eq. 2.29), equals zero. These correspond to points at which the electron density is a minimum, maximum, or a saddle point and are referred to as critical points.

\[
\nabla \rho(r) = \frac{\partial \rho}{\partial x} i + \frac{\partial \rho}{\partial y} j + \frac{\partial \rho}{\partial z} k = 0
\]

(eq. 2.29)

The second derivatives of \( \rho(r) \), \( \nabla^2 \rho(r) \), also known as the Laplacian of the density, are used to determine the nature of such critical points. The Laplacian of the electron density is obtained using the eigenvalues of the Hessian matrix. In its simplest form, \( \nabla^2 \rho(r) \) is obtained by adding three such eigenvalues together (eq. 2.30).
\[ \nabla^2 \rho(r) = \lambda_1 + \lambda_2 + \lambda_3 \quad (2.30) \]

Based on the nature and sign of the three eigenvalues, four types of critical points in a molecule can be defined by their rank \((r)\) and signature \((s)\): \(r\) is the number of non-zero \(\lambda\) and \(s\) is the algebraic sum of the signs of the eigenvalues. These values are then commonly expressed in the form \((r, s)\) and can tell us the nature of the critical point. In Table 2.2 the relationship between a specific \((r, s)\) expression and the nature of a critical point is listed.

<table>
<thead>
<tr>
<th>((r, s))</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, -3)</td>
<td>Nucleus</td>
</tr>
<tr>
<td>(3, -1)</td>
<td>Bond critical point (BCP)</td>
</tr>
<tr>
<td>(3, +1)</td>
<td>Ring critical point (RCP)</td>
</tr>
<tr>
<td>(3, +3)</td>
<td>Cage critical point (CCP)</td>
</tr>
</tbody>
</table>

2.13.2 Criteria for Defining Hydrogen Bond Interactions in AIM

The values of \(\rho(r)\) and \(\nabla^2 \rho(r)\) at a BCP can be used to identify the character of an interaction. For example, for hydrogen bonds the values of \(\rho(r)\) are usually within the range of 0.002–0.040 a.u. while the values of \(\nabla^2 \rho(r)\) are typically in the range 0.024–0.139 a.u. In addition, the strength of these interactions can be determined by calculating the ellipticity \((\varepsilon)\) via eq. 2.31, of the density at the BCP.

\[ \varepsilon = \frac{\lambda_3}{\lambda_2} - 1 \quad (2.31) \]
The values $\lambda_1$ and $\lambda_2$ are two of the eigenvalues used in eq. 2.30. The larger the value of $\epsilon$, the weaker the interaction.

2.14 Potential Energy Surfaces

As noted previously (section 2.4.1), the Born-Oppenheimer Approximation leads to an electronic wave function ($\psi_{el}$) of the chemical system that is parametrically dependent on the nuclear coordinates of the system. That is, for each nuclear configuration there are a set of $\psi_{el}$'s corresponding to the ground-state, first excited state and so on. This enables us to construct potential energy surfaces (PESs) for a set of nuclei and provides a visual expression of, most commonly, the relationship of the energy of system as a function of the geometry. However, these are multidimensional surfaces with one-dimension per internal coordinate and one for energy!

Fortunately, in chemistry we are generally only interested in structures that correspond to energy minima such as reactants, products, intermediates or first-order transition structures (reaction transition states; energy minima in all directions except one, along the multidimensional PES). Collectively, such points are known as stationary points. Using the relative energy of these points one can construct a two-dimensional (2-D) PES, energy versus reaction coordinate, as shown in Figure 2.4.
Figure 2.4 Schematic illustration of a 2-D PES for a chemical reaction generated by plotting energy versus reaction coordinate: RC = reactant complex, TS = transition state, and PC = product complex.

In some cases, however, it may be necessary to more extensively calculate the energy of a system with respect to its nuclear configuration. This is often the case in reactions in which the motion of one coordinate is strongly coupled with that of another. For example, proton transfer from an active site residue onto the substrate may occur with concomitant cleavage of a bond within the substrate. We may wish to better understand which motion initiates the reaction step, e.g., does proton transfer lead to bond cleavage or vice versa. In such cases one can generate a 3-dimensional (3-D) PES of energy versus two structural parameters of which an example is shown in Figure 2.5.
Figure 2.5 Schematic illustration of a 3-D PES for a chemical reaction generated by plotting energy versus two structural coordinates.

For some mechanisms, it is possible that there is more than one possible pathway that interconnects the reactants to the products or even intermediate complexes (ICs) as illustrated in Figure 2.6. The preferred pathway of a reaction process generally corresponds to that which proceeds via the lowest activation energies.

Figure 2.6 Schematic illustration of multiple reaction pathways interconnecting the reactant (RC) and product (PC) complexes via different possible intermediates (ICs).
2.14.1 Geometry Optimizations

To obtain a potential energy surface and other properties of a chemical system, chemical models must be optimized in order to obtain the lowest energy structure. In reaction stationary points, see above, the first derivative of the energy with respect to nuclear coordinates equals zero. The nature of the stationary point can then be determined via a frequency calculation (see below).

In this thesis all optimized geometries solely using a DFT method were obtained at the B3LYP/6-31G(d) or B3LYP/6-31G(d,p) level of theory, while all ONIOM QM/MM optimized geometries were obtained at the ONIOM(B3LYP/6-31G(d):AMBER) or ONIOM(B3LYP/6-31G(d,p):AMBER) level of theory.

2.14.2 Frequency Calculations

The characterization of a stationary point as an energy minimum or first-order transition state is achieved by performing a harmonic vibrational frequency calculation on the optimized geometry. More specifically, the harmonic vibrational frequencies are obtained by calculating the second derivative of the energy with respect to the nuclear coordinates. If all vibrational frequencies are real then the structure is an energy minimum, e.g., reactant complex, an intermediate, or reaction product. However, if one and only one of the vibrational frequencies is imaginary, the structure corresponds to a first-order transition state.

In addition, various energy corrections can also be obtained via harmonic vibrational frequency calculations including zero-point vibrational energy (ZPVE), enthalpy, and Gibbs free energy corrections. It should be noted that such corrections are only truly valid if the chemical model contains no fixed points or constrained parameters. In practice it has been noted that if the number of fixed points is kept
consistent along a reaction PES then any errors in, for example, the ZPVE, will generally cancel in calculations of relative energies.

### 2.14.3 Single Point Energy Calculations

In order to obtain reliable optimized geometries it has been found that one generally only requires modest-sized well-balanced basis sets, e.g., 6-31G(d,p). In contrast, the calculation of accurate relative energies generally requires the use of much larger and more extensive basis sets, e.g., 6-311+G(2df,p). Unfortunately, the use of large basis sets for geometry optimizations is often computationally unfeasible. However, it has also been noted that in many cases the optimized geometry obtained using a more modest-sized well-chosen basis set does not energetically significantly differ from that obtained using the much larger basis set. Hence, a common cost-saving technique used throughout computational chemistry is to obtained optimized geometries at a suitable lower level of theory and then perform single point energy calculations on the structures obtained at a higher level of theory. An example of this composite approach from this thesis is:

\[
\text{B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d,p)} \quad (2.32)
\]

The nomenclature used in eq. 2.32 denotes that optimized geometries were obtained at the B3LYP/6-31G(d,p) level of theory while single-point energy calculations were performed on these structures at the B3LYP/6-311+G(2df,p) level of theory.
2.15 Environmental Effects

By default, geometry optimizations and single point energy calculations are performed in the gas-phase; i.e. the environment dielectric constant ($\varepsilon$) = 1. However, in reality, most chemical and biochemical reactions occur in a solvent or environment with $\varepsilon > 1$. This polarity can have a tremendous influence on chemistry and properties of a chemical system. In order to take environmental effects into account, different computational approaches have been developed. For the purposes of this thesis, we will briefly describe those approaches used herein.

2.15.1 Solvation Models

A common approach in computational chemistry for inclusion of general polar environment effects is to use an implicit solvation model. One well-developed and widely used such approach is the polarizable continuum model (PCM). A variety of PCM-based methods are available; however, one available in Gaussian 09 and used throughout this thesis is the integral equation formalism polarized continuum model (IEF-PCM).

In PCM based methods, a 'solute cavity' is created that is fitted to the shape of the solute (Figure 2.7). For instance, spheres may be placed on each atom with radii that match the van der waals radii of the given atom. This cavity is surrounded by a polarized medium with a selected dielectric constant ($\varepsilon$) that mimics the polarity of the solvent we wish to consider, e.g., water, toluene. The reciprocity polarization effect between solute molecule (inside the cavity) and solvent (outside the cavity) modifies the electron density of the solute. For chemical systems in isotropic media, the PCM approach has been demonstrated to be able to accurately describe solvent effects in computational investigations. For those investigations in this present thesis that use a DFT-cluster approach to examine reactions within a protein
environment, the IEF-PCM solvation model was used with $\varepsilon = 4.0$. That latter value has been shown to reliably describe the polarity of the protein environment surrounding an enzyme active site.\(^\text{11}\)

![Figure 2.7](image)

**Figure 2.7** Graphic description of a polarized continuum model.

### 2.15.2 Electronic Embedding (EE) in QM/MM

The above PCM-based approach uses a continuum around the solute of homogenous polarity. However, in proteins or other complicated chemical systems, the environment can be distinctly non-homogeneous. For example, the residues around an enzyme active site can vary from highly polar or charged (e.g., arginyl) to non-polar (e.g., phenylalanyl). In addition, these differing polarities of different residues can have significant impacts on the active site reactions and their thermochemistry. In such cases, an ONIOM QM/MM method with use of the electronic embedding formalism can be a better choice for investigating enzymatic mechanisms.\(^\text{19,30}\) In the ONIOM scheme as used in this thesis, the MM layer can be treated using either mechanical embedding (ME) or electronic embedding (EE). In the former, the residues are essentially treated as steric entities; their effect on the active site arises from their size and shape. In EE, one takes into account the point
Chapter 2. Theoretical Methods

Charges on each atom, which are often based on the selected force field being used for the MM layer. The electrostatic interactions between the MM layer and QM layer; how the electron density distribution in the QM region is influenced by the MM region, can then be taken into account. Computationally, EE calculations are much more expensive than ME calculations. Therefore, similar to the single-point energy calculation approach, QM/MM optimized geometries and corresponding frequency calculations have been performed within the ME formalism. QM/MM relative energies have then been obtained via single-point energy calculations, often at higher level of theory, within the EE formalism.

2.16 Technical Aspects and Units

Gaussian 03\textsuperscript{9} and Gaussian 09\textsuperscript{10} software packages were used for geometry optimizations, frequency calculations and single point energy calculations. The Molecular Operating Environment (MOE)\textsuperscript{31} software package was used for Docking, MM minimization, and MD simulations. The AIM2000\textsuperscript{32} program was used in performing topological analyses. The NBO3.0\textsuperscript{20} program was used for natural bond orbital population analyses.

Throughout this thesis, bond distances are given angstroms (Å) and angles are reported in degrees (°). Relative energies are in kJ mol\textsuperscript{-1} and determined using the standard conversion from hartrees (h) of 1 h = 2625.5 kJ mol\textsuperscript{-1}. 
Chapter 2. Theoretical Methods

References


Chapter 2. Theoretical Methods


(20) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. NBO 3.1; Theoretical Chemistry Institute, University of Wisconsin: Madison, 1996.


Chapter 3

A DFT Study on the Catalytic Mechanism of UDP–Glucose Dehydrogenase
3.1 Introduction

Carbohydrates have key roles in cellular biochemistry including biosignalling, energy storage, membrane structural assemblage and metabolic intermediation as nucleotides, glycolipids and glycoproteins. In fact, several carbohydrate-containing precursors are common to unrelated metabolic pathways in different species.\(^1\,^2\) For instance, uridine 5'-diphosphate glucuronic acid (UDPGlcUA in Scheme 3.1) is an essential precursor in the syntheses of many polysaccharides. In mammalian cells, UDPGlcUA bonds to certain waste products and toxic substances, thereby increasing their water solubility, so that they may be excreted from the body.\(^3\,^4\) UDPGlcUA is also an essential precursor in the making of antiphagocytic capsular polysaccharides of pathogenic bacteria. In *Streptococcus pneumoniae*, for example, UDPGlcUA is a key intermediary in the synthesis of the polysaccharide capsule. This capsule acts as an antiphagocytic protector and is responsible for the toxicity and extent of the antibiotic resistance observed in many organisms.\(^5\,^6\,^7\) UDPGlcUA is synthesized by uridine diphosphate glucose dehydrogenase (UDPGlcDH),\(^3\,^8\) which regulates the glycoprotein biosynthesis and other pathways. Thus, UDPGlcDH is a logical target for the development of new antibacterial drugs.

![UDPGlcUA](image)

**Scheme 3.1** Uridine 5'-diphosphate glucuronic acid, abbreviated UDPGlcUA.

The catalytic performance of UDPGlcDH is unique in several aspects. First, it should be noted that the chemical oxidation of glucose is not specific, as the terminal 1-
formyl and 6-hydroxymethyl groups of the open-chain form can be oxidized to carboxylic groups in the presence of moderate oxidizing agents such as transition metal ions, halogens and warm dilute nitric acid. However, within a single active-site, UDPGlcDH is able to perform an overall four-electron oxidation by successively transforming the 6-hydroxymethyl group into a formyl group, and the latter into the final carboxylic function.\(^9,10\) In addition, UDPGlcDH uses two NAD\(^+\) molecules as the final electron acceptors, and this type of redox reaction path is rarely found in the enzymatic chemistry of carbohydrates.\(^11\) Finally, UDPGlcDH belongs to the family of sugar nucleotide-modifying enzymes, and hence, various molecular details of the UDPGlcDH enzymatic mechanism may be common to other enzymes of the family.\(^12,13\)

The first breakthrough in the elucidation of the catalytic mechanism of UDPGlcDH was obtained by Oppenheimer and Handlon who found that the catalytic pathway was a NAD\(^+\)-dependent two-fold oxidation.\(^8\) Thereafter, based on their X-ray crystallography, site-directed mutagenesis and kinetic isotope effect studies, Ge et al.\(^10\) proposed two possible catalytic pathways with the difference being the identity of mechanistically required acid-base active site residues. More specifically, they proposed that either an active site aspartate or possibly a neutral lysyl residue is the mechanistic base. While the latter is directly adjacent to the substrate, the aspartate interacts indirectly via a possible H\(_2\)O bridge with substrate. Recently, studies combining gene encoding and kinetics,\(^3,14\) and the crystal structures of UDPGlcDH from several species (e.g., human, cows, bacteria and plants) have been reported.\(^15\)

The proposed mechanism of UDPGlcDH involves multiple steps and is shown in **Scheme 3.2.** The first step is the oxidation of the 6-hydroxymethyl to a formyl group via hydride transfer to NAD\(^+\) with concomitant deprotonation of O6 by a general base B (Lys204–NH\(_2\) or Asp264–COO\(^-\)). Then, Cys260 in concurrence with Glu145 attacks the carbonyl group to form a covalent thiohemiacetal adduct, while the conjugated acid HB donates the proton back to O6. In the second stage, NADH leaves to allow a new NAD\(^+\)
to enter the active site. The latter, assisted by the same general base B, further oxidizes the thiohemiacetal adduct intermediate to form a thioester. In the final step, the thioester is hydrolyzed to give the final UDPGlcUA product.

![Diagram of catalytic mechanism]

**Scheme 3.2** The proposed catalytic mechanism of UDP-glucose dehydrogenase (UDPGlcDH).\(^\text{10}\)

While the proposed mechanism outlined in **Scheme 3.2** shows fair agreement with the experimental evidence, unsatisfactorily there remains the question of the identity of the base that initiates the acid–base catalytic pathway.\(^\text{8,10,11,16,17}\) In this chapter, the potential energy surfaces of the catalytic mechanism of UDPGlcDH are explored. Specifically, density functional theory methods are applied in order to assess the feasibility of these pathways and to clarify the role of active site residues.

### 3.2 Computational Methods

The hybrid density functional B3LYP,\(^\text{18-20}\) as implemented in the *Gaussian 03* program,\(^\text{21}\) was used in all calculations. This functional combines Becke’s three-parameter hybrid exchange functional and the correlation functional of Lee, Yang and Parr. The B3LYP/6-31G(d) level of theory was used for geometry optimizations,
harmonic vibrational frequencies and zero-point vibrational energy (ZPVE) corrections. Single-point calculations at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level were done to compute relative total energies. In addition, the long-range dielectric effect of the protein environment that surrounds the active site was estimated by performing single-point calculations on the optimized geometries using IEF-PCM at the B3LYP/6-31G(d) level of theory. The dielectric constant $\varepsilon$ was set to 4.0 in these calculations.

Our active-site model was extracted from the crystallographic structure of the mutated UDPGlcDH (with Ser replacing Cys260) complexed with glucuronic acid (PDB ID: 1DLJ).\textsuperscript{10} Key amino acid residues were replaced as follows: Asp264 and Glu145 by acetate ions, Cys260 by ethanethiol, Asn208 by acetamide, Thr118 by methanol. In the case of Lys204, the side chain of the residue was replaced by the ethylammonium ion in Model A and by ethylamine in Model B, as illustrated in Figure 3.1. The substrate UDP-glucose was modelled as $\alpha$-D-glucopyranose, and NAD$^+$, as an $N$-hydroxyethylnicotinamidium ion. The two water molecules present in the crystal structure were kept in the active-site model. One H$_2$O forms a hydrogen-bond bridge between Asp264 and the 6-hydroxymethyl of glucose, and the other H$_2$O, between Glu145 and the thiol of Cys260. For both active-site models, NADH leaves the catalytic cleft during the last step, allowing Glu141 (modeled as butyrate) to approach the substrate. In order to keep the spatial arrangement of catalytic residues in the active site, selected atoms were held at their positions in the crystal structure (cf. Figure 3.1). The applicability of the present computational approach to the study of enzymatic reactions has been previously reviewed in detail.\textsuperscript{22-27}
3.3 Results

3.3.1 Protonation state of the active site of UDPGlcDH in the Michaelis complex

The proposed catalytic mechanism of UDPGlcDH is initiated by oxidation of the 6-hydroxymethyl to a formyl group by a hydride transfer to NAD$^+$ with concomitant deprotonation of O6 by either Lys204 or Asp264. In the case of Asp264, the experimental evidence suggests that it occurs as an aspartate rather than an aspartic group. However, there is no unambiguous experimental evidence for the side chain of Lys204 occurring as either a neutral amine or a positively charged ammonium group. As a result, we began our investigation by constructing the two alternate models of the active site with bound substrate and cofactor, the
Michaelis complex, shown in Figure 3.1. Then, we estimated the proton affinities and basicities of the side chain of Lys204 under four different electrostatic conditions, in order to investigate the effect of the local electrostatic environment of the enzyme on the side chain of Lys204.

The proton affinities (PA’s) were estimated as the negative change in the internal energy of protonation of the side chain at 0 K with the proton coming from the vacuum state, i.e., \( PA = -[E_0(BH^+) - E_0(B) - E_0(H^+)] \), where \( E_0(H^+) = 0 \). The basicities were estimated as the negative change in the internal energy of protonation of the side chain at 0 K with the proton coming from an aqueous ideal-dilute solution at pH 7. The standard energy of bulk solvation of the proton in aqueous solution was taken to be \(-1097.9 \text{ kJ mol}^{-1} \) because this value is the only one consistent with the Born–Haber-type cycles of both hydrogen and the electron.\(^{28,29}\) That is, Basicity = PA – 1097.9 kJ mol\(^{-1}\).

The proton affinities and basicities of ethylamine, the model chosen for the side chain of Lys204, within four different electrostatic environments are reported in Table 3.1. In a first approximation, the protonation of isolated ethylamine was considered in the vacuum and in a homogeneous medium of dielectric constant of 4. The latter value being chosen as it is typically used as an appropriate estimate of the average electrostatic dielectric constant found within a bulk protein environment.\(^{30,31}\) The proton affinity of 941.7 kJ mol\(^{-1}\) indicates that protonation of the side chain in the vacuum is thermodynamically feasible. However, the side chain in a homogeneous medium of dielectric constant of 1 cannot abstract a proton from the bulk of an ideal-dilute aqueous solution, according to the basicity of \(-156.1 \text{ kJ mol}^{-1} \). However, the situation changes if the side chain of lysine is placed under conditions simulating the average effect of the bulk protein environment. In particular, the proton affinity of the lysine increases such that its estimated basicity now becomes positive at 43.1 kJ mol\(^{-1}\). This suggests that when embedded in the
average protein environment, the lysine side chain amino group is thermodynamically able to gain a proton from the bulk of an ideal-dilute aqueous solution at pH 7.

We then placed the model of lysine within our chosen active site model and re-calculated its proton affinity and basicity (Table 3.1). First, we only considered the anisotropic electrostatic environment generated by the first shell of residues found in the catalytic cleft with the surrounding electrostatic environment set to a dielectric constant (ε) of 1 (i.e., the vacuum state). It is found that the protonation of the side chain of Lys204 becomes even more feasible, as indicated by the now larger proton affinities and basicities of 1280.4 and 182.4 kJ mol⁻¹, respectively. In contrast to that observed for isolated lysine, however, increasing the polarity of the surrounding environment to that resembling the bulk protein (i.e., increasing ε from 1 to 4) decreases the proton affinity and basicity of the lysine. Indeed, its PA decreases by approximately 80 kJ mol⁻¹ to 1208.2 kJ mol⁻¹ with the same decrease in basicity to 110.5 kJ mol⁻¹.

Thus, in summary, the three electrostatic conditions simulating the electrostatic environment felt by the side chain of Lys204 in the active site of UDPGlcDH unambiguously show that Lys204 is very likely to be protonated. Therefore, the protonation state of the active site of UDPGlcDH in the enzyme-substrate-cofactor complex is most likely to correspond to the structural Model A (cf. Figure 3.1), with Asp264 bound to water acting as the general base in the initial step of the catalytic mechanism.
Table 3.1 Proton affinities and basicities (in kJ mol⁻¹) of the side chain of Lys204, modelled as ethylamine, within environments of different polarity.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Proton Affinity (α = vac)</th>
<th>Basicity (α = aqu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B + H⁺ → BH⁺</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Isolated</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys (ε = 1) + H⁺ (α) → LysH⁺ (ε = 1)</td>
<td>941.7</td>
<td>-156.1</td>
</tr>
<tr>
<td>Lys (ε = 4) + H⁺ (α) → LysH⁺ (ε = 4)</td>
<td>1141.1</td>
<td>43.1</td>
</tr>
<tr>
<td><strong>Within the active site model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys (ε = 1) + H⁺ (α) → LysH⁺ (ε = 1)</td>
<td>1280.4</td>
<td>182.4</td>
</tr>
<tr>
<td>Lys (ε = 4) + H⁺ (α) → LysH⁺ (ε = 4)</td>
<td>1208.2</td>
<td>110.5</td>
</tr>
</tbody>
</table>

3.3.2 Substrate binding in the active-site of UDPGlcDH

Since Model A (cf. Figure 3.1) represents the most likely Michaelis complex for UDPGlcDH, it was selected as the reactant complex (RC, cf. Figure 3.2) for our mechanistic study. Model A was extracted from a crystallographic structure of the enzyme (PDB accession code: 1DLJ),¹⁶ with β-glucopyranose as the substrate.¹⁰,¹⁶ This chemical model was then carefully optimized in order to preserve the hydrogen-bonded network observed in the crystallographic model, as it has been suggested that this network is essential for the enzyme to carry out its catalytic mechanism.¹⁰ The optimized structure of our chemical model shows that the 6-hydroxymethyl group of glucose hydrogen-bonds to the oxygen of an H₂O (1.64 Å), and this water, to Asp264 and Thr118 (1.66 and 1.71 Å, respectively), forming a ring. As in the crystal structure, Lys204–NH₃⁺ forms a hydrogen-bonded bridge to the 6-hydroxymethyl and the carbonyl group of Asn208. In addition, one of the 6-
position C–H bonds points directly to the 4-position of the pyridine ring of NAD$^+$ at a distance of 2.78 Å.

On the opposite side of the cleft, water forms hydrogen bonds that enable the bridge Glu145–COO$^-$···H$_2$O···HS–Cys260 at 1.97 and 2.08 Å, respectively. It has been debated whether Cys260 is in the thiol (Cys260–SH) or thiolate (Cys260–S$^-$) form in the active site.$^{16}$ Our computational model of the enzyme–substrate complex shows that the thiolate form of Cys260 causes the orientation of Glu145 to change such that it flips into the binding region of the substrate, hydrogen-bonding to glucose and NAD$^+$. Hence, the thiol form of Cys260 is the only alternative that makes possible a stable Glu145–water–Cys260 bridge and that is in agreement with the crystallographic structure.

**Figure 3.2** Optimized structure of the reactant complex (RC): substrate and cofactor bound to the active-site model of UDPGlcDH.
3.3.3 First catalytic step: Oxidation of glucose to form glucuronaldehyde

The first catalytic step of UDPGlcDH (cf. Figure 3.3) is the oxidation of the 6-hydroxymethyl group of glucopyranose to form the 6-formyl terminus. This process begins with a proton transfer from the substrate’s 6-hydroxymethyl group to the $\text{H}_2\text{O}$ bound to the carboxylate of Asp264. Concomitantly, this bridging water transfers a proton to Asp264 (RC in Figure 3.2). This deprotonation has a barrier of 28.2 kJ mol$^{-1}$, with the resulting 6-oxyethyl anion intermediate IC1 (Figure 3.4) lying 35.5 kJ mol$^{-1}$ above both RC and IC1 (cf. Figure 3.3). It is noted that the energy of TS1 is lower than that of IC1 due to the inclusion of ZPVE. This is a commonly occurring artifact for flat PES’s and merely indicates that the reaction proceeds without any barrier. Consequently, the deprotonation brings about a C6–O bond-length shortening from 1.43 to 1.39 Å and a disruption of the hydrogen-bonded network originally found in the Michaelis complex. More specifically, the original hydrogen-bonded chain Asp264–COO$^-$–$\text{H}_2\text{O}$–HO–CH$_2$–Glc in RC changes directionality to form the Asp264–COOH–OH$_2$–O–CH$_2$–Glc chain in IC1, with hydrogen-bonding distances of 1.59 and 1.44 Å, respectively. In addition, the distance between the 6-position C–H bond that points to the 4-position of the pyridine ring decreases from 2.78 to 2.67 Å (cf. IC1 in Figure 3.4). The 6-oxyethyl anion is further stabilized by accepting a hydrogen bond from Lys204–NH$_3^+$, which keeps bridging to the carbonyl group of Asn208.

Then, the 6-oxyanion undergoes a formal hydride transfer from the methyleneoxyanion of IC1 to NAD$^+$ through TS2 with a barrier of 106.5 kJ mol$^{-1}$, leading to formation of glucuronaldehyde, i.e., the 6-formyl intermediate IC2 (cf. Figure 3.4). Thermodynamically, IC2 lies 33.1 kJ mol$^{-1}$ above RC (Figure 3.3). Moreover, the C–O bond has shortened from 1.39 Å in IC1 to 1.24 Å in IC2, indicating the formation of the 6-formyl derivative. The oxidation process to give
IC2 does not alter the hydrogen-bonding pattern relating to the active-site residues, i.e., Asp264–COOH…OH₂…O=CH–Glc. In addition, the formyl terminus accepts a hydrogen bond (1.65 Å) from Lys204–NH₃⁺, which keeps the bridge to the carbonyl group of Asn208 (1.90 Å).

**Figure 3.3** Potential energy surface for the oxidation of the 6-hydroxymethyl group of glucopyranose (RC) to form glucuronaldehyde (IC2) through the 6-oxymethyl anion intermediate (IC1).
3.3.4 Second catalytic step: Nucleophilic addition to form the 6-thiohemiacetal intermediate

Glucuronaldehyde, i.e., the 6-formyl intermediate IC2, then undergoes nucleophilic addition of thiol from Cys260 to form a rather stable 6-thiohemiacetal intermediate (cf. Figure 3.5). This is achieved through attack of the sulfur centre of Cys260 at the carbonyl carbon of IC2 with concomitant formal transfer of the thiol proton of Cys260–SH to the carboxylate of Glu145. This process occurs at a cost of just 11.0 kJ mol\(^{-1}\) to give the thiohemiacetal anion intermediate IC3, lying 32.1 kJ mol\(^{-1}\) lower in energy than IC2. In fact, the formation of IC3 is essentially thermoneutral with respect to the initial active-site bound substrate complex RC···NAD\(^{+}\).
Figure 3.5 Potential energy surface for the nucleophilic addition of the Cys260 thiol group to the 6-formyl end of glucurononaldehyde in order to form a 6-thiohemiacetal intermediate (IC4).

In **IC3**, the S–C bond forms at 1.94 Å, whereas the C6–O bond lengthens to 1.36 Å (cf. **Figure 3.6**). Both the bond lengthening and increase of negative charge on O6 (i.e., from +0.05 in **IC2** to −0.55 in **IC3**) indicate the formation of the highly basic oxyanion structure C6–O⁻. This oxyanion is further stabilized by accepting a hydrogen bond from Lys204–NH₃⁺, which has now broken the bridge to the carbonyl group of Asn208. Then, the oxyanion of **IC3** abstracts a proton back from Asp264 through the bridging water molecule (at a cost of 5.3 kJ mol⁻¹) to form the neutral thiohemiacetal intermediate **IC4** [Glc–C6H(S–Cys260)OH], lying 31.8 kJ mol⁻¹ below **IC2**. Predictably, the proton transfer causes the C6–O bond to elongate to 1.41 Å and the C6–S bond to shorten to 1.86 Å in **IC4** with respect to **IC3**. Finally, the hydrogen-bonding network around the O6–H group in **IC4** (**Figure 3.6**) is restored to the pattern found in the enzyme–substrate complex **RC** (**Figure 3.2**). However, even
when the C6-OH group in IC4 is stabilized by accepting a hydrogen bond from Lys204–NH₃⁺, the bridge to the carbonyl group of Asn208 remains broken in the optimized structure of this intermediate.

![Figure 3.6 Optimized structures of the thiohemiacetal anion (IC3) and neutral (IC4) intermediates.](image)

### 3.3.5 Third catalytic step: Second oxidation to form the 6-thioester intermediate

The first oxidation step produces an enzyme–intermediate–cofactor complex, which, for the sake of simplicity, we represent as IC2···NADH. An experiment with the Cys260→Ala mutant complexed with labelled UDP-[6",6"-di-²H]-d-glucose in a solution containing NAD⁺ and excess NADH showed that the NAD²H formed in the first oxidation step did not exchange with free NADH.¹⁰ Thus, according to this evidence, the reduced cofactor remains bound to the active site until the thiohemiacetal neutral intermediate IC4···NADH is formed. Then, the NADH can be
exchanged for NAD\textsuperscript{+}, which then binds and continue oxidizing the 6-position of glucose.\textsuperscript{5,10,11}

**Figure 3.7** Optimized structures of the 6-thiohemiacetal complex with both NADH and NAD\textsuperscript{+}.

The optimized geometries of the 6-thiohemiacetal complex **IC4** with the reduced and the replacing oxidized cofactor molecules (i.e., **IC4**⋯NADH and **IC4**⋯NAD\textsuperscript{+}, respectively, in **Figure 3.7**) show high structural similarity, including the hydrogen-bonding pattern to Lys204. Furthermore, it was found that the **IC4**⋯NADH complex together with an infinitely separated NAD\textsuperscript{+} ion is only 4.0 kJ mol\textsuperscript{-1} more stable than the **IC4**⋯NAD\textsuperscript{+} complex together with an infinitely separated NADH molecule.

Once the fresh, oxidized cofactor molecule is appropriately bound to the active site, the **IC4**⋯NAD\textsuperscript{+} complex is ready to proceed with the second oxidation reaction (cf. **Figure 3.8**). This begins with formation of a 6-thiohemiacetal anion intermediate by way of a proton transferring from the 6-hydroxyl group to Asp264–COO\textsuperscript{−} through the bridging water molecule. The barrier of this step is just 1.9 kJ mol\textsuperscript{-1}.
1, with the 6-thiohemiacetal oxyanion intermediate IC5 lying 0.3 kJ mol⁻¹ lower in energy with respect to IC4···NAD⁺. The proton transfer reverses the directionality of the hydrogen-bonded chain from Glc–C6H(S–Cys260)OH···OH₂···OOC–Asp264 in IC4 to Glc–C6H(S–Cys260)O⁻···H₂O···HOOCC–Asp264 in IC5 (Figure 3.9), the hydrogen-bonding distances in the latter being 1.60 and 1.65 Å. The 6-oxyanion is further stabilized by accepting a hydrogen bond from Lys204–NH₃⁺, which itself keeps the bridge to the carbonyl group of Asn208 broken. Moreover, the C6–H bond in IC5 is left directed more precisely to the reduction centre of the cofactor, with the hydrogen located 2.48 Å apart from the 4-position of the NAD⁺ pyridine ring. In this geometrical arrangement, the two-electron one-proton transfer occurs from a C6–H group to NAD⁺, to give the 6-thioester intermediate IC6. In the thioester intermediate, the C6=O link has a typical carbonyl bond distance of 1.23 Å, while the C6–S bond is shortened to 1.78 Å. This formal hydride transfer has a barrier of 72.3 kJ mol⁻¹ with respect to IC4···NAD⁺, and the 6-thioester intermediate IC6···NADH lies 19.3 kJ mol⁻¹ lower in energy than IC4···NAD⁺. In the IC6···NADH complex, the 6-carbonyl group accepts a hydrogen bond from Lys204–NH₃⁺, and the latter restores the bridge to the carbonyl group of Asn208.
Chapter 3. A DFT Study on UDP Mechanism

Figure 3.8 Potential energy surface for the oxidation of the 6-thiohemiacetal intermediate (IC4) to the 6-thioester intermediate (IC6) through a 6-thiohemiacetal oxyanion intermediate (IC5).

Figure 3.9 Optimized structures of the 6-thiohemiacetal oxyanion intermediate (IC5) and the 6-thioester intermediate (IC6).
3.3.6 Fourth catalytic step: Hydrolysis of the 6-thioester intermediate to form glucuronic acid

It has been proposed that hydrolysis of the 6-thioester intermediate IC6 is catalyzed by the strictly conserved Glu141 residue.\textsuperscript{16,17} However, the crystallographic structure of the UDPGlcDH Cys260→Ser mutant with bound UDP-GlcUA and NADH shows that Glu141 lies behind the nicotinamide ring, which blocks the interaction of Glu141 with the 6-position of the thioester intermediate.\textsuperscript{16} It has been hypothesized that, in the absence of NADH, Glu141 could extend into the active site and deliver an activated water molecule to the 6-position of the thioester intermediate.\textsuperscript{16} In view of these ideas, we studied the hydrolysis of the thioester in both the NADH-bound active-site complex and in a complex in which NADH was exchanged by a glutamate ion bound to two water molecules.

First, we probed into the feasibility of hydrolysis of the 6-thioester within the IC6···NADH complex in the absence of nearby negatively charged groups that polarize the attacking water molecule. To this end, we devised chemical models of IC6···NADH that included one water molecule into the active site and computed 4-, 6- and 8-membered ring transition structures. The calculated barrier heights to hydrolysis (i.e., 189.0, 197.6 and 188.2 kJ mol\textsuperscript{-1}, respectively) showed a better agreement with a nonenzymatic thioester hydrolysis reaction. For example, at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory, the barrier to hydrolysis of CH\textsubscript{3}(C=O)SCH\textsubscript{3} with a single water molecule is 184.7 kJ mol\textsuperscript{-1}. It is noted that our barrier height is in disagreement with that of 44.3 kJ mol\textsuperscript{-1} reported for the same chemical model but calculated at the HF/6-31++G** level of theory.\textsuperscript{32} In contrast, the MP2/6-311++G(2d,2p)//B3LYP/6-31++G(d,p) barrier to the OH\textsuperscript{-} catalyzed hydrolysis of CH\textsubscript{3}(C=O)SCH\textsubscript{3} including four explicit water molecules and the bulk solvent IPCM correction was found to be 51.5 kJ mol\textsuperscript{-1}.\textsuperscript{33}
Next, our chemical model of the 6-thioester intermediate was modified by removing NADH from the **IC6**⋯NADH complex and by incorporating glutamate and two water molecules to generate the **IC6**⋯Glu141•2H2O complex, which is illustrated in **Figure 3.10**. The optimized structure of the **IC6**⋯Glu141–COO−⋯2H2O complex consists of an eight-membered hydrogen-bonded ring, in which the Glu141 carboxylate accepts one hydrogen bond from each water molecule at 1.84 and 1.99 Å respectively. The two water molecules are hydrogen-bonded to each other at 1.82 Å (cf. **Figure 3.11**). The displacement of NADH from the **IC6**⋯NADH complex by the eight-membered hydrogen-bonded ring complex Glu141–COO−⋯2H2O to form the **IC6**⋯Glu141–COO−⋯2H2O complex is highly exergonic by 122.6 kJ mol⁻¹. In other words: the Glu141–COO−⋯2H2O complex binds stronger than NADH to the intermediate–active-site complex, and this provides a thermodynamic pull favoring the displacement of NADH from the active site.

**Figure 3.10** Potential energy surface for hydrolysis of the 6-thioester intermediate (**IC6**) assisted by Glu141 and two water molecules, and subsequent formation of a product complex containing GlcUA (**PC**).
Figure 3.11 Optimized structures of the 6-thioester intermediate incorporating glutamate and two water molecules (IC6···Glu141•2H2O), the 6-thioether-6,6-diol anion intermediate (IC7) and the product complex with GlcUA (PC).

In the geometric arrangement of the IC6···Glu141–COO–•2H2O complex, the 6-carbonyl group accepts a hydrogen bond from the Lys204–NH3⁺, which itself also hydrogen-bonds to the carbonyl group of Asn208. Importantly, the nearest water oxygen is initially located 3.28 Å from C6 of the thioester intermediate. This H2O
then nucleophilically adds to the carbonyl carbon of the thioester intermediate to form the 6-thioether-6,6-diol anion intermediate IC7. The formation of IC7 is endoergic by 80.9 kJ mol⁻¹, with a barrier of 102.1 kJ mol⁻¹ with respect to the IC6···Glu141•2H₂O complex (cf. Figure 3.10). In IC7, the four-centre hydrogen-bonded ring changes slightly with the newly formed 6-hydroxyl group accepting a hydrogen bond from Glu141–COOH, while the other two hydrogen bonds remain the same as in the parent IC6···Glu141–COO⁻•2H₂O complex. The 6-oxyanion is further stabilized by accepting a hydrogen bond from Lys204–NH₃⁺, which has now broken the bridge to the carbonyl group of Asn208. The arrangement of the substituents around the 6-position of the 6-thioether-6,6-diol anion intermediate IC7 is tetrahedral, with typical bond lengths of 1.44 and 1.33 Å for the C6–OH and C6–O⁻ bonds, respectively. Predictably, the C6–S bond lengthens from 1.79 Å in IC6 to 1.96 Å in IC7, because of formation of the geminal diol intermediate (cf. Figure 3.11).

Finally, the formation of GlcUA occurs by scission of the C6–S bond in concert with two proton transfers mediated by bridging water molecules: one occurring from Asp264–COOH to the C6–O⁻ group, and the other occurring from Glu145–COOH to the S atom of Cys260 (cf. Figure 3.10). This step has a barrier of 69.6 kJ mol⁻¹ and leaves the product complex (PC) 14.2 kJ mol⁻¹ lower in energy than the IC6···Glu141•2H₂O complex. In PC, it was found that the geometry around C6 is trigonal planar with typical bond lengths for the carboxylate group of 1.23 and 1.32 Å for the C=O and C–OH bonds, respectively (Figure 3.11). Lys204–NH₃⁺ forms a bridge to the 6-carbonyl group of GlcUA and the carbonyl group of Asn208 in the optimized product complex. Notably, the hydrogen-bonded chain Glu145–COO⁻···H₂O···HS–Cys260 is regenerated for the next catalytic cycle.
3.4. Conclusions

The potential energy surface (PES) for the catalytic mechanism of UDPGlcDH was investigated for an active-site model of this enzyme and calculated with the B3LYP density functional method in conjunction with the IEF-PCM self-consistent reaction field, in order to correct for the long-range electrostatic effect of the protein environment. An overall PES for the catalytic mechanism of UDPGlcDH, with energies referenced to our active-site model of the enzyme–substrate–cofactor complex is included in the supplementary data for this paper (Appendix Figure A.1).

This study made it possible to give microscopic insight into the catalytic roles of the residues along the reaction path, and mainly, unravel the roles of the Lys204 and Asp264 residues. To this end, we applied three electrostatic conditions simulating the electrostatic environment felt by the side chain of Lys204 in the active site of UDPGlcDH to unambiguously show (by the calculated basicities of the side-chain of lysine) that: (i) Lys204 is very likely to be protonated in the active site; and (ii) the reactive form of the enzyme–substrate–cofactor complex most likely contains Asp264–COO\(^{-}\) bound to water, which acts as the general base in the initial step of the catalytic mechanism. Moreover, the stationary structures along the reaction path show that Lys204–NH\(_3\)\(^{+}\) assists in decreasing the rotational freedom of the 6-terminus of glucopyranose and in stabilizing the 6-oxyanion intermediates and transition states.

The catalytic mechanism consists of two NAD\(^{+}\)-dependent oxidation stages. The calculations show that the first two-electron one-proton transfer from the activated 6-oxymethyl anion of glucopyranose to NAD\(^{+}\) forming glucuronaldehyde is the rate-controlling step, with a barrier of 106.5 kJ \(\text{mol}^{-1}\). Within the computational error for an active-site model,\(^{27}\) this barrier is in good agreement with the experimental activation energy of 72 kJ \(\text{mol}^{-1}\), which was estimated from the catalytic constant for wild-type UDPGlcDH with UDP-glucose under saturating
conditions \( k_{\text{cat}} = 1.8 \pm 0.1 \text{ s}^{-1} \).\(^{10}\) Moreover, we showed that in order for the 6-thioester intermediate to be hydrolyzed and form GlcUA in the last catalytic step, hydrated Glu141–COO\(^-\) has to displace the NADH molecule formed in the second oxidation step. This statement is supported by:

(i) Calculated high barriers to hydrolysis of the 6-thioester intermediate within the NADH-bound active-site complex larger than 180 kJ mol\(^{-1}\), which are unusual in enzymatic reaction mechanisms.

(ii) The exoergic character \((-122.6 \text{ kJ mol}^{-1}\) of displacement of NADH from the 6-thioester intermediate–active-site complex by a Glu141–COO\(^-\)•2H\(_2\)O complex, which suggests that the Glu141–COO\(^-\) dihydrate binds stronger than NADH to the intermediate–active-site complex.

(iii) The lower barrier height (102.1 kJ mol\(^{-1}\)) for the hydrolysis of the activated 6-thioester intermediate within the active-site complex including the Glu141–COO\(^-\) dihydrate residue.
Chapter 3. A DFT Study on UDP Mechanism

References

Chapter 3. A DFT Study on UDP Mechanism


Chapter 4

Redox Mechanism of Glycosidic-Bond Hydrolysis Catalyzed by 6-Phospho-α-Glucosidase (GlvA): A DFT Study
4.1 Introduction

Glycosidic bonds are ubiquitous in living matter, occurring not only between sugars in oligosaccharides and polysaccharides such as starch, cellulose and glycogen, but also in natural glycoconjugates such as nucleosides, glycoproteins and glycolipids. A biochemically important property of such linkages is that they are remarkably resistant to cleavage via hydrolysis. In fact, to hydrolyze half of the glycosidic bonds of an β-glycoside in contact with water within 24 hours requires temperatures over 200 °C or an acidic medium, or both. For example, the half-life of such bonds for the uncatalyzed hydrolysis of cellulose at SATP has been estimated at approximately 5 million years, with a pseudo first-order rate constant of hydrolysis that increases as the pH decreases.

However, the extreme reaction conditions and small rate constants of chemical glycoside hydrolysis are incompatible with the cellular environment and metabolic needs of hydrolysis of glycosidic bonds. Thus, the group of enzymes known as glycoside hydrolases emerged to catalyze their selective hydrolysis in oligosaccharides, polysaccharides and glycoconjugates. In fact, glycosidases are amongst some of the most efficient biocatalysts known, with rate constants of up to 1000 s⁻¹.

Typically, in the course of hydrolysis, wild-type glycosidases can invert or retain the stereochemical configuration at the anomeric carbon by following either a nonredox associative reaction pathway or a net redox-neutral dissociative reaction pathway. In the nonredox reaction mechanism, inverting glycosidases perform a direct, single acid–base-catalyzed displacement of the aglycone by water. Importantly, the resulting product saccharide has the opposite stereochemical configuration at the anomeric carbon to that of the parent glycoside. In contrast,
retaining glycosidases follow a nonredox double-displacement catalytic mechanism. Its first step begins with a nucleophilic group (belonging to an enzyme’s residue or to the substrate itself) attacking at the glycoside’s anomic carbon to form a glycosyl–nucleophile intermediate. As a result, the parent glycoside and the glycosyl–nucleophile intermediate have inverted configurations at the anomic carbon. In the second step, the glycosyl–nucleophile intermediate is hydrolyzed; specifically, a water molecule, activated by a general base, substitutes the aglycone by a hydroxyl group at the anomic carbon. Consequently, the resulting product saccharide recovers the stereochemical configuration of its parent glycoside at C1. In summary, nonredox glycosidases are highly specific and follow cationic associative mechanisms involving oxocarbenium-like intermediates.\(^2,14,15\)

A distinct group of glycoside hydrolases are gathered in the glycosidase family GH4.\(^23\) They catalyze a net redox-neutral NAD\(^+\)-dependent glycosidic-bond hydrolysis in a divalent metal complex and follow an anionic dissociative mechanism involving enolate-like intermediates.\(^13,18-20,24,25\) Thus, some of the GH4 glycosidases have shown reduced stereospecificity, being able to hydrolyze both \(\alpha\)- and \(\beta\)-glycosidic bonds. For instance, \textit{B. subtilis} 6-phospho-\(\alpha\)-glucosidase (GlvA),\(^26-28\) which has high primary-sequence similarity to other GH4 glycosidases such as 6-phospho-\(\beta\)-glucosidase (BglT) and \(\alpha\)-glucosidase (AglA),\(^17,29-31\) preferentially hydrolyzes \(\alpha\)-glycosidic bonds but also cleaves their alternate \(\beta\)-anomers having suitable leaving groups, following a pathway that always retains the stereochemical configuration at the anomic carbon.\(^21,32\) Based on crystallographic and kinetic studies, Withers and coworkers proposed a NAD\(^+\)-dependent redox pathway of hydrolysis catalyzed by GlvA and related enzymes (\textit{Scheme 4.1}).\(^18,19,21,30\) Further experimental support for such a pathway has been obtained from kinetic,\(^16,17,27,33\) NAD\(^+\) affinity,\(^27,34\)
It should be noted that, in carbohydrate biochemistry, many enzymes exploit the oxidizing power of cofactors (e.g., NAD$^+$, NADP$^+$ and FAD$^+$) to generate transient keto intermediates at different positions in the hexose ring over the course of their catalytic pathways, as a means of activating the ring in order to achieve the intended chemical transformation.$^{14,15,22,36,37}$ A wide variety of sugar deoxygenation,$^{14,15,36-38}$ monosaccharide epimerization,$^{14,15,38}$ glycoside hydrolysis$^{16,17,19-22,30}$ and glycosyl transfer$^{14,15,38}$ reactions are catalyzed by enzymes that employ the transient oxidation strategy, as for instance, sugar deoxygenases, reductases, dehydratases, epimerases, hydrolases and transferases. In the particular case of GlvA, the glycosidic-bond hydrolysis is initiated by oxidation of the glucopyranose ring at the 3-position to a keto group by the NAD$^+$ molecule bound to the active site.
Scheme 4.1 The NAD⁺-Dependent Redox Pathway of Glycosidic-Bond Hydrolysis Catalyzed by GlvA.¹⁶,¹⁷,¹⁹-²²,³⁰
The overall proposed pathway for GlvA effectively occurs in two half-reactions. The first half-reaction is initiated by transfer of the substrate's C3–OH proton to the Mn$^{2+}$-bound OH$^-$. Concomitantly, a hydride transfers from the substrate's C3–H group to NAD$^+$ to form a C3=O carbonyl intermediate. Subsequently, the C2–H proton is abstracted by an adjacent active-site tyrosine residue to give a C3–O$^-$ enolate derivative (Scheme 4.1). Then, an active-site aspartic group (Asp172) protonates the glycosidic oxygen, which leads to C1–OCH$_3$ bond dissociation and loss of CH$_3$OH. The second half-reaction is then essentially the reverse of the first half-reaction except that now an H$_2$O replaces the CH$_3$OH molecule. Indeed, it begins with the H$_2$O nucleophilically attacking the substrate's C1 centre while donating a proton to the Asp172–COO$^-$ residue to regenerate a C3–O$^-$ enolate (Scheme 4.1). The Tyr265–OH group then acts as an acid, protonating the C2 carbon to regenerate a C3=O carbonyl intermediate. The final step is a concomitant protonation of the C3=O oxygen by the Mn$^{2+}$-bound H$_2$O with a formal hydride transfer from the NADH cofactor to the substrate's C3, thereby generating the anomerically hydroxylated saccharide product.

The aim of this study is to investigate the NAD$^+$-dependent redox mechanism of glycosidic-bond hydrolysis as catalyzed by GlvA, through the combined application of density functional theory and a self-consistent reaction field to a large active-site model obtained from the crystallographic structure of the enzyme. In addition, we have used natural bond orbital and second-order perturbation analyses to monitor the electron flow and change in oxidation state on each atomic centre along the reaction coordinate, in order to rationalize the energetics and regioselectivity of this catalytic mechanism.
4.2 Computational Methods

All calculations were performed using the Gaussian 03 suite of programs. The density functional method B3LYP, a combination of Becke’s three-parameter hybrid exchange functional with Lee, Yang and Parr’s correlation functional, was used in all calculations. Optimized geometries and corresponding harmonic frequencies were obtained with the 6-31G(d) 5D basis set. Diffuse functions were not included in geometry optimizations because the negative charges in the model are highly localized in the functional groups. In addition, there are not dipole-bound anions within the functional groups. In such cases, the inclusion of diffuse functions in the basis set will not significantly change the optimized structures with respect to optimized geometries in a basis set excluding diffuse functions. General effects due to the polarity of the internal protein environment were modelled using the Polarizable Continuum Model in the framework of the Integral Equation Formalism (IEF-PCM) solved at the B3LYP/6-311+G(2df,p) level of theory with a dielectric constant of 4.0. Consequently, all relative energies reported throughout this study were obtained at the IEF-PCM(ε=4.0)/B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) 5D level, unless otherwise noted. The electron flow along the reaction coordinate was tracked by using the natural bond orbital and second-order perturbation analyses of the electron densities of selected equilibrium structures. In each equilibrium structure, the NBO basis spans more than 97 % of the electron density. These calculations were performed at the B3LYP/6-31G(d) level of theory with the NBO 3.1 program.

A large active-site modelling approach in the framework of density functional theory (specifically applying the B3LYP functional) was used for this mechanistic study. Such approaches are widely used in computational studies on metalloenzymes, and their applicability and reliability have been reviewed in detail.
elsewhere.\textsuperscript{43,44,46-51} Our active-site model (Figure 4.1b) was obtained by extracting the coordinates of the active-site residues from the crystal structure of GlvA complexed with the substrate analog 6-phospho-\(\beta\)-glucose (Figure 4.1a, PDB accession code: 1U8X).\textsuperscript{30} Since 6-phospho-\(\alpha\)-glycosides are the preferred substrates of GlvA, we replaced the substrate analog with 1-methyl-\(\alpha\)-glucopyranoside (\(\alpha\)Glc1Me) in our chemical model. Inspection of the crystallographic model shows that the 6-phosphate is complexed by four arginine residues, i.e., Arg88, Arg95, Arg285 and Arg293. Hence, the role of the phosphate group is very likely to be related to binding and positioning of the substrate within the active site and is not likely to be directly involved in the catalytic mechanism of redox glycosidic-bond hydrolysis.

All catalytically relevant species were included: Asn149, Glu111, Thr112, Tyr265 and an active-site H\(_2\)O; the Mn\(^{2+}\) ion and its coordinated ligands (His202, OH\(^-\) and Cys171); the NAD\(^+\) cofactor; and the protein backbone along the carboxylic carbon of Ile170 and C\(\alpha\) of Asp172. Protonation states of active-site residues were assigned according to the experimental evidence provided by the Brønsted correlation analyses and \(^2\)H kinetic isotope effect (KIE) data obtained for GlvA.\textsuperscript{19,21} These species were replaced in the active-site model as follows: Glu111 by acetate, Asp172 by acetic acid, Cys171 by ethanethiol, Thr112 by methanol, Tyr265 by phenol, His202 by imidazole, Asn119 by acetamide, and NAD\(^+\) by the 1-hydrionicotinamidium ion. A minimal number of atoms remote from the reaction site were held fixed at their crystal structure positions (marked with asterisks in Figure 4.1b) in order to maintain spatial integrity of the active site. The puckering of the substrate’s pyranoid ring within the active-site model along the reaction coordinate was characterized by the Cremer–Pople parameters,\textsuperscript{52} which are listed in Appendix Table B.1.
Figure 4.1 Structural models of the active site: (a) Arrangement of the catalytically active residues of GlvA and the 6-phospho-β-glucopyranose analog substrate in the crystal structure (PDB accession code: 1U8X); (b) Chemical model for the mechanistically important species, with 1-methyl α-glucopyranoside replacing the 6-phospho-β-glucopyranose analog substrate.

4.3 Results

4.3.1 Substrate Binding in the Active Site

The optimized structure of the initial substrate bound within the active site (RC, pyranoid ring conformation $^4C_1$) with selected distances is shown in Figure 4.2. Overall, it is in good agreement with the experimentally determined crystal structure of the enzyme with a 6-phospho-β-glucose analog bound within the active site (PDB: 1U8X).\textsuperscript{30}
For example, the NAD$^+$ cofactor hydrogen-bonds via its carboxyl group with the backbone amide group between residues Cys171 and Asp172 as also found in the crystal structure. Furthermore, an active-site water sits between the side chains of Tyr265, Thr112 and Glu111. Specifically, it acts as a hydrogen-bond acceptor from the hydroxyl groups of both Tyr265 and Thr112, with distances of approximately 1.8 Å, while simultaneously acting as a hydrogen-bond donor to the carboxylate of Glu111 at a distance of 1.60 Å. Importantly, and as previously suggested$^{21}$, the water acts as a hydrogen-bonding bridge between the Glu111 carboxylate and the Tyr265 hydroxyl group.

Moreover, in agreement with the crystal structure, the Mn$^{2+}$ ion remains pentacoordinate in our active-site model. Mn$^{2+}$ is bidentately chelated by the oxygen atoms at the 2- and 3-positions of glucose with distances of 2.42 and 2.23 Å, respectively.
It also ligates the catalytic OH\(^{-}\) ion with an Mn\(^{2+}\)\(\cdots\)OH\(^{-}\) distance of 1.94 Å. Finally, it is also directly coordinated to the enzyme via the imidazole of His202 and the thiol sulfur of Cys171, with Mn\(^{2+}\)\(\cdots\)N–His202 and Mn\(^{2+}\)\(\cdots\)S(H)–Cys171 distances of 2.10 and 2.73 Å, respectively. It is worth remembering that, in our model and in line with Withers's et al. proposal,\(^{21}\) we chose Cys171–SH to be neutral. We note that we built an active-site model that contained the anion Cys171–S\(^{-}\) and obtained an Mn\(^{2+}\)\(\cdots\)S–Cys171 distance of 2.34 Å, which is still compatible with the crystal structure. However, such a thiolate-containing active-site model brings about a change in the ligation pattern of Mn\(^{2+}\), which now forms a tetracoordinate centre because of breakage of the dative Mn\(^{2+}\)\(\cdots\)O3 bond. Therefore, this active-site model was not considered further.

Crucially, two mechanistically relevant interactions are observed in RC, indicating an appropriate positioning of the substrate to initiate the reaction. First, the neutral carboxylic group in Asp172–COOH forms a relatively tight hydrogen bond (1.62 Å) to the glycosidic oxygen O1 (see Figure 4.2), although this only causes a marginal lengthening of 0.02 Å in the C1–OMe bond with respect to the free substrate (Appendix Table B.3). Second, the \(\_\text{O3H}\) group of glucose hydrogen-bonds to the oxygen of the Mn\(^{2+}\)-bound OH\(^{-}\) at 1.77 Å, and thus is prepared for the first step of the catalytic mechanism.

4.3.2 3-Oxidation of Glucopyranose to Form the C2–C3=O Carbonyl Intermediate

As outlined by Withers, the step proposed to initiate the catalytic mechanism of GlvA is formation of a C3=O carbonyl intermediate as a result of (i) a proton transferring from the C3–OH group to the hydroxido ligand in the Mn\(^{2+}\)\(\cdots\)OH\(^{-}\) moiety, and (ii) a formal hydride transferring from the C3–H group to NAD\(^{+}\).\(^{19,21,30}\) Our
calculations show that the proton and hydride transfers occur stepwise with the proton transfer occurring first followed by the hydride transfer. The potential energy surface (PES) obtained for the initial stepwise catalytic step is displayed in **Figure 4.3**, while the corresponding optimized structures are illustrated in **Figure 4.4** (see also Appendix Table B.2).

![Figure 4.3 Potential energy surface for the first-half reaction (3-oxidation and deglycosylation) in the catalytic mechanism of redox hydrolysis of αGlc1Me by GlvA.](image)

### 4.3.2.1. Proton Transfer from the C3–OH Group to the Mn$^{2+}$···OH$^-$ Moiety

Abstraction of the C3–OH proton by the hydroxido ligand in Mn$^{2+}$···OH$^-$ proceeds via **TS1** (pyranoid ring conformation $^4C_1$) with a quite low barrier of just 22.7 kJ mol$^{-1}$. The resulting 3-oxyanion intermediate **IC1** lies 46.3 kJ mol$^{-1}$ lower in energy than **RC**, i.e., the initial active site with bound substrate complex.

In the intermediate complex **IC1** (pyranoid ring conformation $^4C_1$), the hydroxido ligand in the Mn$^{2+}$···OH$^-$ moiety has now become an aqua ligand (Mn$^{2+}$···H$_2$O), with the Mn$^{2+}$···O distance increasing by 0.19 Å to 2.13 Å. Although the substrate’s C3–O3 bond has shortened markedly by 0.04 Å to 1.40 Å, its length and partial charge of $-0.76$ indicate that the O3 centre still maintains considerable
oxyanion character. This charge is stabilized by the ligating interaction of O3 with Mn$^{2+}$, as suggested by the fact that in IC1 the Mn$^{2+}$⋯O3 distance has shortened to 1.96 Å compared to 2.42 Å in RC (cf. Figure 4.2). It is also noted that the distance between the C3–H hydrogen and the hydride-accepting C4 centre of NAD$^+$ has now decreased significantly from 3.81 Å in RC to 3.22 Å in IC1. Furthermore, the C3–H bond has lengthened slightly to 1.11 Å.

**Figure 4.4** Optimized structures for the transition states TS1 and TS2, the 3-oxyanion intermediate IC1 and the 3-keto intermediate IC2.
4.3.2.2. Hydride Transfer from the C3–H Group to NAD$^+$

As noted above, following the initial proton transfer, the next step in the overall mechanism of GlvA is a hydride transfer from the C3–H group of IC1 to the pro-R face of the 1$H$-nicotinamidium ion (i.e., NAD$^+$) at C4 to give 1$H$,4$H$-nicotinamide (i.e., NADH) and the 3-keto intermediate IC2. Our model shows that the hydride transfer to the pro-R face of the 1$H$-nicotinamidium ion is favoured over that to the pro-S face in agreement with the structural and kinetic evidence.$^{16,19,29,30}$ This step proceeds via TS2 (pyranoid ring conformation 4C1) with a barrier of 80.8 kJ mol$^{-1}$ with respect to IC1, or just 34.5 kJ mol$^{-1}$ with respect to RC. The resulting 3-keto intermediate IC2 lies 39.1 kJ mol$^{-1}$ lower in energy than RC (see Figure 4.3).

In IC2 (pyranoid ring conformation 4C1), the C3–O bond has now shortened significantly to 1.23 Å, typical for a C=O double bond. Simultaneously, the dative Mn$^{2+}$...O3 bond has weakened, as indicated by its lengthening to 2.21 Å. As a result, the Mn$^{2+}$ ion now forms slightly stronger interactions with the oxygens of the adjacent water and NADH, and the sulfur centre of Cys171 (Appendix Table B.3).

At this stage, the largest structural changes in the substrate are fairly localized to the C3=O moiety. Indeed, only quite minor changes are observed in the sugar ring. For instance, the C2–C3 and C3–C4 bonds have shortened in IC2 by just 0.01 and 0.02 Å relative to their distances in RC. In addition, the C1–C2 and C4–C5 bonds have both now lengthened by just 0.01 Å. Furthermore, the C1–O glycosidic bond in IC2 is calculated to be 0.01 Å shorter than in RC. Importantly however, it is also noted that in IC2 the C2–H proton has a slightly greater positive charge (0.23) compared to that in RC (0.22) or IC1 (0.20). The charge increase is in line with the increased acidity suggested by the measured KIE at C2.$^{17,21}$
4.3.3 2-Deprotonation of the C2–C3=O Carbonyl Intermediate to Form the C2=C3–O⁻ Enolate

In agreement with the proposed mechanism, the next step is the proton abstraction from the C2–H group of IC2 by Tyr265–OH, which has been suggested to act as a base according to the pH-dependent activity studies. This transfer proceeds via TS3 (pyranoid ring conformation ⁴C₁) with a relatively high barrier of 93.1 kJ mol⁻¹ with respect to IC2 to give the resulting enolate (i.e., the 2-en-3-oxyanion) derivative IC3, which lies 24.5 kJ mol⁻¹ higher in energy than IC2 (Figure 4.3).

The pH-dependent activity studies also suggest that for Tyr265–OH to act as a base, it needs to be deprotonated, and hence, the suggestion of a nearby carboxylate to act as the final proton acceptor. The crystallographic structural model for GlvA suggests that a Tyr265–OH···OH₂···OOC–Glu111 hydrogen-bonding network is likely to perform a Grothuss proton-transfer step. Accordingly, we have adopted such a hydrogen-bonding network in our computational model. As can be seen in the optimized structure of IC2 (Figure 4.4), the Tyr265–OH group acts as a proton donor to an adjacent water, which is itself a hydrogen-bond donor to the carboxylate of Glu111. Thus, as previously suggested, this network increases the basicity of the Tyr265–OH oxygen. As seen in the optimized structure of TS3 (Figure 4.5), this makes it possible for the Tyr265–OH group to abstract the proton from the C2–H group. This occurs concomitantly with a proton transfer along the Tyr265–OH···OH₂···OOC–Glu111 hydrogen-bonding network (Figure 4.5). As a result, a neutral Glu111–COOH is formed, while Tyr265–OH remains effectively unchanged in IC3.
Chapter 4. DFT Study on GlvA Mechanism

Figure 4.5 Optimized structures of the transition state TS3 and the C2=C3−O− enolate intermediate IC3.

The enolic character of the fragment C2=C3−O− in IC3 (pyranoid ring conformation 0H5) is confirmed by the shortening of both the C3−O3 and C2−C3 linkages with respect to the C2−C3−O− segment in the 3-oxyanion intermediate IC1. The changes in length of the C3−O3 and C2−C3 linkages in the sequence IC1, IC2 and IC3 are also an indication of charge delocalization in the latter. For instance, the C3−O3 bond distance of 1.31 Å in the fragment C2=C3−O− of IC3 corresponds to an elongation of 0.08 Å with respect to the fragment C2=C3=O of IC2, but to a shortening of 0.09 Å with respect to the segment C2−C3−O− of IC1. Similarly, the C2−C3 bond distance of 1.36 Å in IC3 represents a shortening of bond length with respect to both IC2 (by 0.16 Å) and IC1 (by 0.18 Å). Furthermore, as illustrated in Figure 4.5, the chair conformation 4C1 of the ring turns into a half-chair conformation 0H5 with O5 lying out of the ring pseudo-plane (see Appendix Table B.1). This conformational change brings the Mn2+ ion and the heavy atoms of the segment −O=C3=C2−OH into the same plane.
Chapter 4. DFT Study on GlvA Mechanism

The enolic character of the segment C2=C3−O\textsuperscript{−} and the change in ring puckering in **IC3** also seem, first, to strengthen the chelating effect of the HO−C2=C3−O\textsuperscript{−} fragment with respect to HO−C2−C3=O in **IC2**, and second, to modify the chelating interaction balance with respect to the segment HO−C2−C3−O\textsuperscript{−} in **IC1**. For instance, the Mn\textsuperscript{2+}−O3 and Mn\textsuperscript{2+}−O2 lengths of 2.06 and 2.30 Å in **IC3** represent shortenings of 0.15 and 0.07 Å with respect to **IC2**.

Notably, the **IC3** system is ready to proceed with the glycosidic bond cleavage, now having the C2−OH group appropriately oriented to convert the 2-en-3-oxyanion into a 1-en-3-keto-2-oxyanion intermediate. This is supported by the fact that the Mn\textsuperscript{2+}−O2−H interaction is optimal to increase the acidity of O2 and to assist in stabilizing the resulting 1-en-3-keto-2-oxyanion through dative bonding. Moreover, the C2−OH strongly hydrogen-bonds to the hydroxyl oxygen of Asp172−COOH (1.93 Å), and this hydroxyl itself forms a short and strong hydrogen bond (1.75 Å) with the glycosidic O1 (**Figure 4.5**).

### 4.3.4. Glycosidic bond cleavage

The next catalytic step is the conversion of the 2-en-3-oxyanion into a 1-en-3-keto-2-oxyanion intermediate, in conjunction with cleavage of the glycosidic bond (cf. **Figure 4.3** and **4.6**). Specifically, this is achieved by donation of the proton from the Asp172−COOH group to the glycosidic oxygen. Concomitantly, the very same aspartic oxygen accepts a proton from the substrate’s C2−OH group (**Figure 4.6**). Thus, the transition structure **TS4** (pyranoid ring conformation E\textsubscript{5}) contains a seven-membered ring involving the substrate, Asp172 and the two protons being transferred. Moreover, the energy barrier to break the glycosidic bond in the 2-en-3-oxyanion **IC3** is only 58.9 kJ mol\textsuperscript{−1}, which is just 44.3 kJ mol\textsuperscript{−1} with respect to the
initial enzyme–substrate–cofactor complex **RC**. Remarkably, these barriers are less than one-half of the experimental activation energy of 126.8 kJ mol⁻¹ reported for the uncatalyzed hydrolysis of αGlc1Me in water.¹ Finally, the 1-en-3-keto-2-oxyanion intermediate with bound methanol **IC4**⋯MeOH lies 15.3 kJ mol⁻¹ lower in energy than **IC3**.

Figure 4.6 Optimized structures for the transition state TS4 and the 1-en-3-keto-2-oxyanion intermediate complex IC4⋯MeOH.

The bond cleavage with loss of methanol leaves the substrate in the **IC4**⋯MeOH complex (pyranoid ring conformation **E₅**) as a highly π-conjugated O=C₃−C₂(−O⁻)=C₁ system. Hence, these atoms along with Mn²⁺ lie in a plane (cf. Appendix Table B.3). The shift of the enolate functional group from the formal resonance structure −O=C₃=C₂(−OH)=C₁ to O=C₃−C₂(−O⁻)=C₁ is evidenced by the shortening of the C₁−C₂ bond to 1.37 Å, lengthening of the C₂−C₃ bond to 1.45 Å, and the relative shortening of both the C₂−O and C₃−O linkages to 1.33 and 1.25 Å, respectively. The dominant O=C₃−C₂(−O⁻)=C₁ resonance structure is stabilized by chelation to the Mn²⁺ ion with Mn²⁺⋯O₂ and Mn²⁺⋯O₃ bond lengths of 2.12 and 2.23
Å, respectively. These distances also indicate that in IC4 the Mn²⁺⋯O–C2 interaction is tighter and the Mn²⁺⋯O=C3 interaction is weaker than the respective Mn²⁺⋯O(H)–C2 and Mn²⁺⋯O–C3 interactions in IC3.

In IC4⋯MeOH, the resulting methanol molecule hydrogen-bonds to the carbonyl group of Asp–COOH (Figure 4.6). Nevertheless, for the enzyme to proceed with the generation of α-glucopyranose, methanol must leave the reaction site in order to let a water enter. The displacement of MeOH from the IC4⋯MeOH complex by a water molecule to form IC4⋯H₂O complex is exoergic by 6.4 kJ mol⁻¹, which can be ascribed to the stronger hydrogen bonding abilities of H₂O compared to MeOH. The resulting IC4 complex, hereafter denoted as IC4⋯H₂O, has a very similar overall structure to IC4⋯MeOH (cf. Appendix Table B.3).

4.3.5. Hydrolysis of the C1=C2–O⁻ Enolate and Subsequent 2-Protonation to Form the C2–C3=O Carbonyl Intermediate

The second half-reaction is essentially the reverse of the first half, although now, as noted above, water has replaced methanol in the active site. The kinetic and structural data indicate that, even if the C1=C2–O⁻ enolate could in principle be attacked by water from either side of the plane, hydrolysis can only occur from the side of the plane in which the H₂O can be hydrogen-bonded to and activated by Asp172–COOH.¹⁶,¹⁹,³⁰ As a result, the configuration at the anomeric carbon is retained. The second stage of the catalytic mechanism starts with nucleophilic attack of H₂O at the anomeric carbon C1, as shown in Figure 4.7. This occurs via the 7-membered ring transition structure TS5 (pyranoid ring conformation E₅) in which the H₂O transfers a proton to the hydroxyl oxygen of Asp172–COOH, which itself concomitantly donates its proton to the C2–O⁻ group. The aspartic-catalyzed
addition of water has a barrier of 77.7 kJ mol\(^{-1}\) and leads to formation of the 2-en-3-oxyanion intermediate \textbf{IC5} (pyranoid ring conformation \(^0\text{H5}\)), which lies 17.8 kJ mol\(^{-1}\) higher in energy than \textbf{IC4}···H\(_2\)O (pyranoid ring conformation \textit{E}\(_5\)). It is noted that the barrier for hydrolysis (74.2 kJ mol\(^{-1}\)) is only marginally higher than that for the reverse of the glycosidic bond cleavage (cf. \textbf{Figure 4.3}). Moreover, the addition of water brings about a shift of the enolate functional group from the resonance structure O=C3–C2(–O–)–C1–H in \textbf{IC4} to –O–C3=C2(–OH)–C1(–H)–OH in \textbf{IC5}. This is corroborated by the shortening of the C2–C3 linkage to 1.36 Å and the lengthening of the C3–O bond to 1.31 Å (cf. Appendix \textbf{Table B.3}). It should be noted that the –O–C3=C2(–OH)–C1(–H)–OH moiety within \textbf{IC5} is homologous to –O–C3=C2(–OH)–C1(–H)–OMe in \textbf{IC3} from the first half of the mechanism, with both pyranoid rings in the envelope conformation \textit{E}\(_5\).

\textbf{Figure 4.7} Potential energy surface for the second-half reaction (1-hydroxylation and 3-reduction) in the catalytic mechanism of redox hydrolysis of \(\alpha\)Glc1Me by GlvA.

The next reaction step is the axial reprotonation of C2 mediated by Tyr265–OH. Specifically, the addition of water brings about the disruption of the \(\pi\)-conjugated system as the geometry around C1 changes from trigonal planar to
tetrahedral. This, in turn, causes C2 to pucker slightly out of the ring pseudoplane, which consequently increases its Brønsted basicity. Consequently, C2 is now able to abstract the proton from Tyr265–OH, and concertedly, the protons shuffle along the hydrogen-bonding network Tyr265–(H)O····H2O···HOOC–Glu111 to form glutamate. This step proceeds via TS6 (pyranoid ring conformation $^4C_1$) at a cost of 79.5 kJ mol$^{-1}$ to form the 3-keto intermediate IC6 (Figure 4.7). Similar to the previous step, it is noted that this barrier is slightly higher than that required for the reverse deprotonation of the C2–H group by Tyr265 in the first half reaction. In IC6, the C3=O bond has shortened to 1.23 Å, while the C2–C3 bond shows a typical single bond length of 1.52 Å. Furthermore, the sugar ring itself has now re-adopted a chair conformation $^4C_1$.

4.3.6. 3-Reduction of the C2–C3=O Carbonyl Intermediate and Subsequent Reprotonation to Form Glucopyranose

The 3-keto intermediate IC6 undergoes reduction by the NADH molecule formed in the first half-reaction, through axial hydride transfer from the 4-position of the 1,4-dihydronicotinamide to the 3-position of the pyranose ring. Our model shows that the hydride transfer from the pro-R face of the 1H,4H-nicotinamide is favoured in agreement with the structural and kinetic evidence. $^{15,19,29,30}$ This step proceeds via TS7 (pyranoid ring conformation $^4C_1$) with a barrier of 72.2 kJ mol$^{-1}$ and leads to formation of the 3-oxyanion intermediate IC7 (pyranoid ring conformation $^4C_1$), which is 6.9 kJ mol$^{-1}$ lower in energy than IC6. Again, the electronic structure of the pyranose ring in IC7 is equivalent to that in IC1. As expected, a significant lengthening of the C3–O bond to 1.40 Å is observed, indicative of increased oxyanion character. The 3-oxyanion is stabilized by the chelating effect
of Mn$^{2+}$, with Mn$^{2+}$−O2 and Mn$^{2+}$−O3 bond lengths of 2.34 and 1.96 Å, respectively. This rearrangement facilitates the proton transfer from the Mn$^{2+}$-bound water to the C3−O$^-$ group, with a barrier of 98.6 kJ mol$^{-1}$, through TS8 (pyranoid ring conformation $^4C_1$). Finally, glucose has been formed and the cofactor−active-site complex is fully regenerated for the next catalytic cycle in PC (pyranoid ring conformation $^4C_1$, Figure 4.8).

![Figure 4.8](image)

**Figure 4.8** Optimized structures for the 2-en-3-oxyanion intermediate IC5, the 3-keto intermediate IC6, the 3-oxyanion intermediate IC7 and the product complex consisting of glucose in the regenerated active site (PC).
4.3.7. Natural Population and Energy Analyses of the Electron Densities of Equilibrium Structures Along the Reaction Coordinate

Weinhold’s natural bond orbital (NBO) population technique is a formalism designed to analytically map uncorrelated or correlated, atomic or molecular electron densities into a few-centred maximum-occupancy orbital basis. This representation makes the multicentre quantum formulation of molecular electronic structure mathematically compatible with both the conceptual electron-pair pictures of bonding and the electron donor–acceptor ideas of interaction.\textsuperscript{45,53-57} The mapping of the electron density or total wave function of a polyelectronic system into the set of natural bond orbitals (NBOs) is exact as, in principle, is any monocentre basis set. But unlike the latter, the orthonormal NBO basis incorporates the key idea of chemical functional group, and thus enables evaluation and comparison of the behaviour of bonding units (i.e., bonds, antibonds and lone-pairs) within distinct molecular environments.\textsuperscript{58,59} In the NBO approach, natural hybrid orbitals (NHOs, e.g., $sp^\lambda$, where $\lambda$ is a positive real number) and NBOs (e.g., n, $\sigma$, $\pi$) are the respective one-centred and $N$-centred ($N = 1, 2, 3$) linear combinations of natural atomic orbitals (NAOs) which maximize the electron occupancies in particular directed regions in space. The electronic structure around each nucleus embedded in its local molecular environment is described by an orthonormal set of NAOs. This set contains as many spatial functions as required to accommodate an atomic number of electrons in (n,$l$)-subshells around the respective nucleus. The sizes and orientations of such functions are optimized to fulfil maximum electron occupancies, and as a result, the atom is often left in a 'promoted' electron configuration within the local molecular environment.\textsuperscript{58,59} In essence, while retaining the information stored in a many-electron density or wave function, the NBO formalism can
optimally describe the directional localization of electrons (in pairs for closed-shell systems, and in pairs and singles for open-shell systems), in certain zones around atoms, and in the bonding regions within molecules (usually in between adjacent pairs of nuclei).

Importantly, the NBO, second-order perturbation, and deletion analyses of many-electron densities and wave functions have proved to be effective approaches to explain and estimate the hyperconjugative origin of the generalized anomic effect and its role in reaction mechanisms occurring in closed- and open-shell systems.\textsuperscript{60-65} In particular, the generalized anomic effect in a methyl pyranoside is the preference for the synclinal (sc) or gauche arrangement over the antiperiplanar (ap) or trans arrangement in the C5–O5–C1–O1–CH\textsubscript{3} linkage. The overall effect in this segment consists of two components: the \textit{endo}-anomic effect (\textit{endo}-An.) related to the C5–O5–C1–O1 fragment, and the \textit{exo}-anomic effect (\textit{exo}-An.) related to the O5–C1–O1–CH\textsubscript{3} fragment.\textsuperscript{60,66-70} Moreover, the existence of a pyranose ring pseudo-plane allows us to further decompose the \textit{endo}-anomic component into two contributions: the axial (\textit{a-endo}-An.) and the equatorial (\textit{e-endo}-An.). In the NBO scheme, each component of the generalized anomic effect is expressed as a set of delocalization interactions from bonds $i$ into antibonds $j^\ast$, which cause charge to transfer, usually from one atomic centre to another. The extent and strength of mixing orbitals $i$ and $j^\ast$ (and the resultant electron transfer) can be evaluated through the stabilization energy $\varepsilon_{i \rightarrow j^\ast}$ associated with superimposing the electron population $n_i$ that occupies bond $i$ with energy $\varepsilon_i^{(0)}$ onto the antibond $j^\ast$ with energy $\varepsilon_j^{(0)}$. To a good approximation, $\varepsilon_{i \rightarrow j^\ast}$ can be estimated by means of two-state second-order perturbation theory as: $\varepsilon_{i \rightarrow j^\ast}^{(2)} = -n_i \langle \hat{F} \rangle_{j^\ast}^2 \left/ \left( \varepsilon_j^{(0)} - \varepsilon_i^{(0)} \right) \right.$, with $\hat{F}$ being either the Kohn–Sham or the Fock operator.\textsuperscript{58,59}
Chapter 4. DFT Study on GlvA Mechanism

Estimates of the B3LYP/6-31G(d) second-order delocalization interactions within some relevant equilibrium structures along the catalytic pathway are reported in Table 4.1. First, it is noted that the placement of the αGlc1Me substrate within RC gives rise to a favourable $p_{\text{OH}} \rightarrow \sigma_{\text{O3-H}}^*$ interaction (79.2 kJ mol$^{-1}$) that leads to deprotonation of the 3-hydroxyl group. In addition, the hyperconjugative interactions in RC can be used as our reference level to assess the influence of the various anomeric components along the pathway. In this case, the e-endo-, a-endo-, and total endo- and exo-anomeric components amount to 13.5, 80.4, 93.9 and 61.4 kJ mol$^{-1}$ respectively, to give an overall anomeric stabilization of 155.3 kJ mol$^{-1}$.

Following a series of elementary steps, the enzyme replaces the axial C2–H group by an axial electron pair with a centre at C2. This exerts a stronger hyperconjugative interaction on the axial $\sigma_{\text{C1-O1}}^*$ orbital, and by this, the glycosidic bond is weakened. The first of these steps, i.e., proton transferring from the C3–OH group to the hydroxido ligand in the Mn$^{2+} \cdots \text{OH}^{-}$ moiety (cf. Figure 4.4), implies strengthening the $p_{\text{O3}} \rightarrow \sigma_{\text{O3-H}}^*$ interaction by about three times from 11.7 kJ mol$^{-1}$ in RC to 33.9 kJ mol$^{-1}$ in IC1. This interaction polarizes the axial $\sigma_{\text{C3-H}}^*$, easing the subsequent transfer of H$^-$ to the empty non-conjugated $p_{C4}^*$ orbital in NAD$^+$. The resulting $\pi_{\text{C3-O}}$ orbital in IC2 weakly delocalizes into the $\sigma_{\text{O2-H}}^*$ antibond, with a $\pi_{\text{C3-O}} \rightarrow \sigma_{\text{C2-H}}^*$ delocalization energy amounting to just 5.6 kJ mol$^{-1}$. Hence, the $\sigma_{\text{C2-H}}$ in IC2 remains unpolarized as it is in IC1 and RC.

Then, a proton abstraction from the C2–H group by Tyr265 relocalizes the electron pair in the $p_{\text{O3}}$ orbital, and the electron pair from the C2–H group localizes into the $\pi_{\text{C2-C3}}$ orbital in IC3 (cf. Figure 4.5). It is noted that the a-endo-anomeric component rises to 95.3 kJ mol$^{-1}$ because the replacement of $\sigma_{\text{O2-H}}$ in RC by $\pi_{\text{C2-C3}}$ in IC3 strengthens the axial hyperconjugation from C2 to $\sigma_{\text{C1-O1}}^*$ by about a factor of
two, i.e., from 19.5 kJ mol\(^{-1}\) in the \(\sigma_{c2-H} \rightarrow \sigma_{c1-o1}^*\) interaction within \textbf{RC} to 39.0 kJ mol\(^{-1}\) in the \(\pi_{c2-c3} \rightarrow \sigma_{c1-o1}^*\) interaction within \textbf{IC3}.

Thus, the resulting intermediate \textbf{IC3} is activated and can reach the transition structure for the heterolytic breakage of the glycosidic bond (cf. \textbf{Figure 4.3}) through the combined influence of several factors: \(i\) thermal motion, \(ii\) the slight weakening and elongation of the \(\sigma_{c1-o1}\) because of the increased \textit{a-endo}-anomeric effect, and \(iii\) hydrogen bonding from Asp172–COOH to O1. Specifically, the elongation of the C1–O1 distance from equilibrium dramatically increases the strength of the \textit{a-endo}-anomeric effect. Hence, very strong \(p_{05} \rightarrow p_{c1}^*\) and \(\pi_{c2-c3} \rightarrow p_{c1}\) delocalizations, which amount to 286.4 and 306.1 kJ mol\(^{-1}\) respectively, are seen in \textbf{TS4}. More important, however, is the fact that the \textit{a-endo}-anomeric effect amounting to 592.5 kJ mol\(^{-1}\) overcomes the combined effects of the two components that give the glycosidic bond its atypical enhanced strength: on one hand, the \(sp_{01,a}^2 \rightarrow p_{c1}^*\) delocalization interaction amounts to 429.2 kJ mol\(^{-1}\) and favours the formation of the \(\sigma_{c1-o1}\), and on the other, the exo-anomeric effect is weakened to just 5.9 kJ mol\(^{-1}\). Finally, \textbf{TS4} leads to the stable electron configuration of \textbf{IC4}···MeOH, in which all delocalization interactions of MeOH with the pyranose ring are essentially negligible.

In short, the data reveals that the rationale behind a redox mechanism of hydrolysis is the gradual strengthening of the \textit{a-endo}-anomeric component along the reaction coordinate to facilitate the heterolytic cleavage of the \(\sigma_{c1-o1}\) bond. Specifically, when the mechanism reaches the transition state for glycosidic-bond cleavage (\textbf{TS4}), the electron pairs lying in the axial \(\pi_{c2-c3}\) and \(p_{05}\) orbitals now push the bonding pair lying along the C1–O1 axis by 169.2 kJ mol\(^{-1}\) \textit{stronger} than the pair occupying the \(sp_{01,a}^2\) orbital which is trying to remake the \(\sigma_{c1-o1}\) bond.
**Table 4.1** B3LYP/6-31G(d) Second-Order Perturbation Estimates (in kJ mol\(^{-1}\)) of Hyperconjugative Interactions in Selected Equilibrium Structures of αGlc1Me.

<table>
<thead>
<tr>
<th>Component</th>
<th>RC</th>
<th>IC1</th>
<th>IC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>i (\rightarrow j^*)</td>
<td>(\varepsilon_{i\rightarrow j^*})</td>
<td>i (\rightarrow j^*)</td>
<td>(\varepsilon_{i\rightarrow j^*})</td>
</tr>
<tr>
<td>(p_{03} \rightarrow \sigma_{03-H})</td>
<td>11.7</td>
<td>(p_{03} \rightarrow \sigma_{03-H})</td>
<td>33.9</td>
</tr>
<tr>
<td>(p_{01} \rightarrow \sigma_{03-H})</td>
<td>79.2</td>
<td>(\sigma_{03-H} \rightarrow p_{c4-NAD^+})</td>
<td>0.4</td>
</tr>
<tr>
<td>e-endo-An.</td>
<td>(sp_{05} \rightarrow \sigma_{c1-c2})</td>
<td>9.3</td>
<td>(sp_{05} \rightarrow \sigma_{c1-c2})</td>
</tr>
<tr>
<td>a-endo-An.</td>
<td>(\sigma_{c2-o} \rightarrow \sigma_{c1-o5})</td>
<td>4.2</td>
<td>(\sigma_{c2-o} \rightarrow \sigma_{c1-o5})</td>
</tr>
<tr>
<td>(p_{05} \rightarrow \sigma_{c1-o1})</td>
<td>60.9</td>
<td>(p_{05} \rightarrow \sigma_{c1-o1})</td>
<td>67.6</td>
</tr>
<tr>
<td>(\sigma_{c1-o1} \rightarrow \sigma_{c2-o})</td>
<td>19.5</td>
<td>(\sigma_{c1-o1} \rightarrow \sigma_{c1-o1})</td>
<td>21.1</td>
</tr>
<tr>
<td>exo-An.</td>
<td>(p_{01} \rightarrow \sigma_{c1-o5})</td>
<td>36.2</td>
<td>(sp_{01} \rightarrow \sigma_{c1-o5})</td>
</tr>
<tr>
<td>(sp_{01} \rightarrow \sigma_{c1-H})</td>
<td>17.8</td>
<td>(sp_{01} \rightarrow \sigma_{c1-H})</td>
<td>17.7</td>
</tr>
<tr>
<td>(\sigma_{c1-c2} \rightarrow \sigma_{c1-c2})</td>
<td>7.4</td>
<td>(\sigma_{c1-c2} \rightarrow \sigma_{c1-c2})</td>
<td>7.0</td>
</tr>
</tbody>
</table>
### Chapter 4. DFT Study on GlvA Mechanism

#### IC3

- $\text{sp}^2_{O5} \rightarrow \sigma^*_{\text{C1-C2}}$: 14.8 eV
- $\sigma_{C2-O} \rightarrow \sigma^*_{\text{C1-O5}}$: 4.8 eV
- $\pi_{\text{C2-C3}} \rightarrow \sigma^*_{\text{ClO1}}$: 39.0 eV

#### TS4

- $\text{sp}^2_{O5} \rightarrow p^*_\text{C1}$: 429.2 eV
- $sp^2_{O5} \rightarrow \sigma^*_{\text{C1-C2}}$: 23.0 eV
- $\alpha_{C2-O} \rightarrow \sigma^*_{\text{C1-O5}}$: 7.4 eV
- $\pi_{\text{C2-C3}} \rightarrow p^*_\text{C1}$: 306.1 eV

#### IC4••MeOH

- $p_{O1} \rightarrow \sigma^*_{\text{C1-O5}}$: 47.0 eV
- $sp^2_{O1} \rightarrow \sigma^*_{\text{C1-H}}$: 14.6 eV
- $\sigma_{O1-C} \rightarrow \sigma^*_{\text{C1-C2}}$: 6.2 eV

---

<table>
<thead>
<tr>
<th>Component</th>
<th>$i \rightarrow j^*$</th>
<th>$\epsilon_{i \rightarrow j^*}$</th>
<th>$i \rightarrow j^*$</th>
<th>$\epsilon_{i \rightarrow j^*}$</th>
<th>$i \rightarrow j^*$</th>
<th>$\epsilon_{i \rightarrow j^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-endo-An.</td>
<td>$\text{sp}^2_{O5} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>14.8 eV</td>
<td>$\text{sp}^2_{O5} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>23.0 eV</td>
<td>$sp^2_{O5} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>12.4 eV</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{C2-O} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>4.8 eV</td>
<td>$\sigma_{C2-O} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>7.4 eV</td>
<td>$\sigma_{C2-O} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>3.7 eV</td>
</tr>
<tr>
<td>a-endo-An.</td>
<td>$p_{O5} \rightarrow \sigma^*_{\text{C1-O1}}$</td>
<td>56.3 eV</td>
<td>$p_{O5} \rightarrow p^*_{\text{C1}}$</td>
<td>286.4 eV</td>
<td>$p_{O5} \rightarrow \pi^*_{\text{C1-C2}}$</td>
<td>0.0 eV</td>
</tr>
<tr>
<td></td>
<td>$\pi_{\text{C2-C3}} \rightarrow \sigma^*_{\text{C1-O1}}$</td>
<td>39.0 eV</td>
<td>$\pi_{\text{C2-C3}} \rightarrow p^*_{\text{C1}}$</td>
<td>306.1 eV</td>
<td>$\pi_{\text{C2-C3}} \rightarrow p^*_{\text{C1}}$</td>
<td>0.0 eV</td>
</tr>
<tr>
<td>exo-An.</td>
<td>$p_{O1} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>47.0 eV</td>
<td>$p_{O1} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>4.1 eV</td>
<td>$\alpha_{O1-H} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>0.0 eV</td>
</tr>
<tr>
<td></td>
<td>$sp^2_{O1} \rightarrow \sigma^*_{\text{C1-H}}$</td>
<td>14.6 eV</td>
<td>$sp^2_{O1} \rightarrow \sigma^*_{\text{C1-O5}}$</td>
<td>1.8 eV</td>
<td>$sp^2_{O1} \rightarrow \sigma^*_{\text{C1-H}}$</td>
<td>0.0 eV</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{O1-C} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>6.2 eV</td>
<td>$\sigma_{O1-C} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>0.0 eV</td>
<td>$\sigma_{O1-C} \rightarrow \sigma^*_{\text{C1-C2}}$</td>
<td>0.0 eV</td>
</tr>
</tbody>
</table>
Chapter 4. DFT Study on GlvA Mechanism

It has been reported that GlvA hydrolyzes glycosides regioselectively, i.e., the redox hydrolyses of $\alpha$-anomers are significantly favoured over $\beta$-anomers.\textsuperscript{21,32} In order to explain the origin of such regioselectivity, we also computed the equilibrium structures for conversion of the 2-en-3-oxyanion intermediate (IC3) into a 1-en-3-keto-2-oxyanion intermediate (IC4...MeOH) with cleavage of the glycosidic bond for methyl $\beta$-glucopyranose ($\beta$Glc1Me). The optimized structures of these complexes are not shown herein but their coordinates are included in Appendix Table B.3. Estimates of the B3LYP/6-31G(d) second-order delocalization interactions for this catalytic step with the $\beta$-anomer are reported in Table 4.2. First, it is noted that the placement of the $\beta$Glc1Me substrate within the polar active-site environment of RC gives rise to a stronger anomeric stabilization (in comparison to $\alpha$Glc1Me) with e-endo-, a-endo-, and total endo- and exo-anomeric components amounting to 51.0, 26.8, 77.8 and 65.8 kJ mol\textsuperscript{-1} respectively, to result in an overall anomeric stabilization of 221.4 kJ mol\textsuperscript{-1} (cf. Appendix Table B.2).

Basically, the series of elementary steps by which GlvA replaces the axial C2-H group by an axial electron pair with a centre at C2 is essentially the same whether the substrate is an $\alpha$- or $\beta$-glycoside. However, the activated $\beta$-substrate intermediate IC3 experiences a reduced driving influence of such activation with respect to its $\alpha$-counterpart. The second-order perturbation analysis (cf. Appendix Table B.2) shows that in IC3 only that part of the e-endo-anomeric component consisting of $sp_{o5}^{1.5} \rightarrow \sigma_{c1-o1}^*$ and $\pi_{c2-c3} \rightarrow \sigma_{c1-o1}^*$ delocalization interactions contribute to the elongation of the C1–O1 bond by 27.9 kJ mol\textsuperscript{-1}, while the exo-anomeric effect markedly contributes against the breakage of the $\beta$-glycosidic bond by 84.4 kJ mol\textsuperscript{-1}. Once the activated $\beta$-substrate reaches TS4, the e-endo-anomeric $p_{o5} \rightarrow p_{c1}^*$ and $\pi_{c2-c3} \rightarrow p_{c1}^*$ delocalizations, which amount to 315.5 kJ mol\textsuperscript{-1}, are not strong enough to overcome the interactions favouring the making of the C1–O1 bond, which
amount to 872.9 kJ mol\(^{-1}\). Specifically, the latter are the \(sp^4_{O1} \rightarrow p^*_{C1}\) interaction and the exo-anomeric effect, each contributing by 778.0 and 94.9 kJ mol\(^{-1}\), respectively. Again, \(TS4\) leads to the stable electron configuration of \(IC4\)--MeOH, in which all delocalization interactions of MeOH with the pyranose ring are essentially negligible.

In summary, the second-order perturbation analysis shows that the gradual strengthening of the \(a\)-endo-anomeric component along the reaction coordinate does not facilitate the heterolytic cleavage of the \(\sigma_{C1-O1}\) bond in a \(\beta\)-glycoside. Specifically, when the mechanism reaches the transition state for glycosidic-bond cleavage (\(TS4\)), the electron pairs lying in the axial \(\pi_{C2-C3}\) and \(p_{O5}\) orbitals cannot push the bonding pair lying along the C1–O1 axis stronger than the combined thrust brought about by the exo-anomeric effect and the pair occupying the \(sp^4_{O1}\) orbital which is trying to remake the \(\sigma_{C1-O1}\) bond. However, a highly electron-withdrawing aglycone would significantly decrease the strength of the \(sp^4_{O1} \rightarrow p^*_{C1}\) delocalization along the C1–O1 axis, thereby shifting the balance of these effects and facilitating the cleavage of a \(\beta\)-glycosidic bond. In fact, this orbital effect explains the decrease in the observed specificity constant \((k_{cat}/K_m)\) of GlvA for a series of 6-phospho \(\alpha\)- and \(\beta\)-\(D\)-glycosides with the decreasing electron-withdrawing effect of the aglycone in the sequence of substituents 3,4-dinitrophenyl, 4-nitrophenyl, 4-\(O\)-\(D\)-glucopyranosyl and methyl groups.\(^{21}\)
Table 4.2 B3LYP/6-31G(d) Second-Order Perturbation Estimates (in kJ mol\(^{-1}\)) of Hyperconjugative Interactions in Selected Equilibrium Structures of βGlc1Me.

<table>
<thead>
<tr>
<th>Component</th>
<th>( i \rightarrow j^* )</th>
<th>( \varepsilon_{i\rightarrow j^*} )</th>
<th>( i \rightarrow j^* )</th>
<th>( \varepsilon_{i\rightarrow j^*} )</th>
<th>( i \rightarrow j^* )</th>
<th>( \varepsilon_{i\rightarrow j^*} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-endo-An.</td>
<td>( sp_{O5} \rightarrow \sigma_{C1-C2} )</td>
<td>10.3</td>
<td>( sp_{O5} \rightarrow \sigma_{C1-C2} )</td>
<td>27.9</td>
<td>( sp_{O5} \rightarrow \sigma_{C1-C2} )</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td>( p_{O3} \rightarrow \pi_{C1-O1} )</td>
<td>10.6</td>
<td>( p_{O5} \rightarrow p_{C1} )</td>
<td>114.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \sigma_{C2-O} \rightarrow \sigma_{C1-O5} )</td>
<td>3.4</td>
<td>( \sigma_{C2-O} \rightarrow \sigma_{C1-O5} )</td>
<td>5.4</td>
<td>( \sigma_{C2-O} \rightarrow \sigma_{C1-O5} )</td>
<td>7.5</td>
</tr>
<tr>
<td>a-endo-An.</td>
<td>( sp_{O1} \rightarrow \sigma_{C1-O1} )</td>
<td>16.4</td>
<td>( \pi_{C2-C1} \rightarrow p_{C1} )</td>
<td>200.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \pi_{C2-C1} \rightarrow \sigma_{C1-O1} )</td>
<td>16.4</td>
<td>( \pi_{C2-C1} \rightarrow \sigma_{C1-O1} )</td>
<td>200.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>exo-An.</td>
<td>( p_{O1} \rightarrow \sigma_{C1-O5} )</td>
<td>59.8</td>
<td>( sp_{O1} \rightarrow p_{C1} )</td>
<td>44.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p_{O1} \rightarrow \sigma_{C1-H} )</td>
<td>14.3</td>
<td>( sp_{O1} \rightarrow p_{C1} )</td>
<td>44.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( sp_{O1} \rightarrow \sigma_{C1-H} )</td>
<td>10.3</td>
<td>( sp_{O1} \rightarrow p_{C1} )</td>
<td>44.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \sigma_{O1-C} \rightarrow \sigma_{C1-C2} )</td>
<td>0.0</td>
<td>( \sigma_{O1-C} \rightarrow p_{C1} )</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.4. Conclusions

The catalytic mechanism of GlvA was investigated using an active-site model of this enzyme and calculated with the B3LYP density functional method in conjunction with the IEF-PCM self-consistent reaction field, in order to correct for the long-range electrostatic effect of the protein environment. An overall PES for the catalytic mechanism of GlvA, with energies referenced to our active-site model of the enzyme–substrate–cofactor complex is displayed in Figures 4.3 and 4.7.

The overall PES shows that the rate-controlling step for this multi-step catalytic mechanism is the 2-protonation of the C2=C3–O- enolate intermediate IC5 to form the C2–C3=O carbonyl intermediate IC6, and hence, our calculated activation energy for GlvA is 61.1 kJ mol\(^{-1}\). This value is in line with the standard activation Gibbs energy of 78.3 kJ mol\(^{-1}\) estimated from the catalytic constant \((k_{\text{cat}} = 0.42 \text{ s}^{-1})\) for wild-type GlvA with methyl 6-phospho-\(\alpha\)-D-glucopyranoside as the substrate under saturating conditions at pH 7.5 and temperature of 37 °C.\(^{21}\) Our calculations support the hypothesis by Yip et al. that the rate-controlling step in the catalytic mechanism of GlvA is not the glycosidic-bond dissociation, nor the subsequent step, but a step further.\(^{21}\) The calculations also support the hypothesis by Yip et al. that the deprotonation of the 3-keto intermediate IC2 could be partially rate-controlling,\(^{21}\) as this step is the rate-controlling step for the first half-reaction, with an activation energy of 54.0 kJ mol\(^{-1}\). Importantly, the overall hydride transfer steps are fast and decoupled from any proton transfer occurring at C2, and are unlikely to be rate-limiting. It is noted that following a redox hydrolytic mechanism of catalysis, GlvA is able to lower the activation energy of the heterolytic \(\alpha\)-glycosidic-bond cleavage to up to 44.3 kJ mol\(^{-1}\), which is as much as three times lower than the experimental activation energy of 126.8 kJ mol\(^{-1}\) reported for the uncatalyzed hydrolysis of \(\alpha\)Glc1Me in water.\(^{1}\) Such a
lowering of the activation energy for catalyzed cleavage of the α-glycosidic bond leaves the mechanism to be controlled by a typical enzymatic proton-transfer step.

The natural population and second-order perturbation analyses of the electron densities of equilibrium structures along the reaction coordinate demonstrated that the rationale behind a redox mechanism of hydrolysis is the gradual strengthening of the α-endo-anomeric component within the substrate along the reaction coordinate in order to facilitate the heterolytic dissociation of the axial \( \sigma_{C_1-O_1} \) bond. Electrons lying in orbitals perpendicular to the glucopyranose pseudo plane are able to increase polarization of the axial \( \sigma_{C_1-O_1} \) bond and facilitate its heterolytic cleavage. However, the push of the very same electrons lying in axial orbitals is less effective, and thus weaker, if the \( \sigma_{C_1-O_1} \) bond is in the equatorial position. Hence, the heterolytic breakage of the β-glycosidic bond becomes unfavourable. These effects fully explain the regioselectivity of GlvA, which favours the redox hydrolysis of α- over β-glycosides, and in the case of the latter, GlvA increasingly favours the hydrolysis of both α- and β-anomers with increasingly electron-withdrawing aglycones, in agreement with the kinetic evidence.\textsuperscript{21}
References

(13) Rempel, B. P.; Withers, S. G. Glycobiology 2008, 18, 570.


Chapter 4. DFT Study on GlvA Mechanism


(45) Glendenning, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO 3.1*, Theoretical Chemistry Institute, University of Wisconsin, 1996.


Chapter 5

An Active Site Water Broadens Substrate Specificity in S-ribosylhomocysteinase (LuxS): A Docking, MD and QM/MM Study
5.1. Introduction

Traditionally, antibiotics are used in the treatment of bacterial infections in order to kill or inhibit the growth of the pathogen. However, their extensive use has had a significant role in the emergence of antibiotic-resistant bacteria.\(^1\) This is increasingly being recognized as a potential major public-health problem. Consequently, there is tremendous interest in identifying new therapeutic approaches to treating such pathogens. One promising approach is to render the pathogens harmless or reduce their efficacy, and allowing the host’s immune system the time to adequately respond naturally to these pathogens.\(^2\)

Type 2 quorum sensing (QS-2) is a cell-to-cell signaling process that regulates gene expression and ultimately population behavior via culture density.\(^3\)-\(^5\) The signaling molecule central to this process is the borate diester of cyclic 4,5-dihydroxy-2,3-petadione (DPD), otherwise known as autoinducer 2 (AI-2).\(^5\),\(^6\) Detection of AI-2 by the microbes allows their populations to regulate diverse processes such as virulence, sporulation, motility and biofilm formation.\(^2\),\(^7\) The latter, in particular, has been shown to be involved in approximately 80% of all microbial infections, and plays a role in anti-bacterial resistance.\(^2\),\(^7\),\(^8\) As QS-2 is not critical to the life-cycle of the bacteria, it is thought that it is not subject to the same degree of evolution-selective pressures.\(^7\),\(^9\) Hence, inhibition of its biosynthesis has been suggested as a promising alternative approach for treating a broad range of pathogens.\(^2\),\(^10\),\(^11\)

The biosynthesis of AI-2 is a multi-enzymatic process (Scheme 5.1).\(^5\) It can be said to begin with demethylation of \textit{S}-adenosylmethionine (SAM) by methyltransferase thereby generating \textit{S}-adenosylhomocysteine (SAH). \textit{S}-adenosylhomocysteine nucleosidase (Pfs) then removes the adenine moiety from
SAH to produce S-ribosylhomocysteine (SRH) which is catalytically converted by S-ribosylhomocysteinase (LuxS) into homocysteine and 4,5-dihydroxy-2,3-pentadione (DPD). Importantly, the latter enzyme LuxS is found in a diverse array of pathogenic bacteria species' but is not expressed in humans.\textsuperscript{12} Hence, it represents a potential target for the inhibition of AI-2 biosynthesis and by extension QS-2.\textsuperscript{13}

\begin{center}
\textbf{Scheme 5.1} Illustration of the enzymatic pathway of AI-2 biosynthesis.
\end{center}

There have been a number of detailed experimental studies on the nature and mechanism of LuxS. In particular, it has been shown to be a dimeric non-heme iron metalloenzyme with its active sites located at the interface between its two monomers; both contributing residues to each active site.\textsuperscript{9,10,13-15} Furthermore, the divalent Fe(II) ion is bound to the enzyme in a tetrahedral-like arrangement via two histidyl (His54, His58) and a cysteinyl (Cys126) residue.\textsuperscript{15-17} In the absence of substrate, based on an X-ray crystal structure of a Zn(II)-containing inactive form of the enzyme (PDB ID: 1J6V),\textsuperscript{10} the fourth ligand is thought to be an H\textsubscript{2}O.\textsuperscript{15} It should be noted that the numbering used herein is taken from that of \textit{Bacillus subtilis} LuxS (PDB ID: 1YCL).\textsuperscript{15} UV-Vis spectroscopy has been used to examine ionization states of various active site residues. Importantly, they concluded that an active site cysteine (Cys84) is in its anionic form (i.e., Cys84-S\textsuperscript{-}) while a nearby arginyl residue (Arg39) is protonated and likely involved in stabilizing the thiolate.\textsuperscript{9} In addition, X-ray crystallographic structures have been obtained of furanose-containing SRH bound within the active site of a LuxS in which Cys84 has been oxidized to sulfonic acid.
while the metal ion is Zn(II).\textsuperscript{17} Furthermore, others have been obtained of the catalytically viable 2-keto derivative (2-keto-SRH), which contains a linear ribosyl moiety, bound within the active site of a Cys84Ala mutated LuxS containing a Co(II) ion. In the former the substrate's cyclic ribosyl moiety was positioned "adjacent to" the Zn(II) ion,\textsuperscript{17} while in the latter the non-cyclic ribosyl was ligated to the Co(II) via its C2=O carbonyl oxygen.\textsuperscript{15}

The kinetics of its mechanism have been investigated via \textsuperscript{13}C NMR and UV-Vis spectroscopy using the SRH substrate itself and the catalytically viable substrate derivatives 2- and 3-keto-SRH.\textsuperscript{18,19} It was concluded that overall it proceeds in three distinct reaction stages.\textsuperscript{19} Site-directed mutagenesis studies have examined the catalytic importance of several active site residues including Ser6, His11, Arg39, Cys84, and Glu57.\textsuperscript{9,16,18-20} It was observed that while the Cys84Asp mutation decreased the rate of reaction, a Cys84Ala substitution completely inhibited its activity.\textsuperscript{18,19} In addition, Glu57 was shown to be important throughout the reaction but more-so for catalytic turnover starting from the putative 3-keto-SRH intermediate, i.e., for the later stages of the mechanism.\textsuperscript{19} Kinetic isotope studies into the stereo- and regio-chemistry of mechanistic proton transfers suggested that reaction of 2-keto-SRH to give the subsequent 3-keto-SRH intermediate was partially rate-limiting.\textsuperscript{20} It has also been suggested that His11 may help transfer the C2-OH proton onto the C1-O oxygen with or without the help of other active-site residues and solvent molecules in the initial stage of the mechanism.\textsuperscript{15,19}

Based on such observations a catalytic mechanism for LuxS has been proposed and is shown in Scheme 5.2.\textsuperscript{9,15-20} Importantly, while it was noted that in solution ribose preferably exists in its furanose form rather than as a linear aldose,\textsuperscript{15,21} it was suggested that the ribosyl moiety in SRH must be linear in order for catalysis to occur.\textsuperscript{9} In this linear aldose configuration, SRH is proposed to bind to the Fe(II) ion.
via its C1=O oxygen (1). The Cys84 thiolate then abstracts a proton from the substrates \(-\text{C2H}^\cdot\) group and transfers it onto the adjacent C1 carbon with concomitant transfer of the \(-\text{O2H}\) proton onto the O1 centre (Scheme 5.2: 1 \(\rightarrow\) 2 \(\rightarrow\) 3 \(\rightarrow\) 4). The resulting 2-keto-SRH derivative (4) formed ligates via its C2=O oxygen to the Fe(II), as observed in the corresponding X-ray crystal structure (PDB ID: 1YCL).\(^{15}\) This is then followed by the analogous reaction in which the Cys84 thiolate, having been regenerated upon protonation of the C1 centre, facilitates a 1,2-proton transfer from \(-\text{C3H}^\cdot\) onto the adjacent \(-\text{C2}\) carbon (i.e., 4 \(\rightarrow\) 5 \(\rightarrow\) 6 \(\rightarrow\) 7). Similarly, this occurs with transfer of the \(-\text{O3H}\) proton onto O2 to give the 3-keto-SRH derivative (Scheme 5.2). Subsequent cleavage of the substrate's thioester C—S bond (i.e., 7 \(\rightarrow\) 8 \(\rightarrow\) 9) is achieved via a Glu57 assisted 1,3-proton transfer. Specifically, it abstracts a proton from \(-\text{C4H}^\cdot\) in 7 and transfers it onto the sulfur of the cleaving C—S bond, i.e., the homocysteine leaving group. This results in formation of homocysteine (Hcys) and DPD, the latter ultimately being converted into AI-2.

Unfortunately, many key fundamental questions still remain including the exact nature of the initial bound SRH substrate and how LuxS may facilitate conversion of the \(\alpha\)- and/or \(\beta\)-furanose-containing SRH anomers into their corresponding linear aldose forms for further reaction as proposed.\(^{19,22}\) Knowledge of substrate binding is highly pertinent for elucidating and understanding an enzyme mechanism and, for example, inhibitory studies. In particular, it can provide invaluable insights into key catalytic interactions between the substrate and the enzyme’s active site functionalities (e.g., residues, waters and cofactors) and consequently, potential initial reaction steps. Hence, in this study several complementary computational methods have been synergistically applied to investigate the initial stages of the proposed catalytic mechanism of LuxS. Specifically, docking, molecular dynamics (MD) simulations and an ONIOM QM/MM
approach have been used to examine the nature of the initial active site-bound substrate and the initial steps of the catalytic mechanism of LuxS leading to formation of the proposed 2-keto-SRH intermediate.

Scheme 5.2 The proposed catalytic mechanism of LuxS.\textsuperscript{19}

5.2. Computational Methods

Docking and Molecular dynamics (MD) simulations: These calculations were performed using the Molecular Operating Environment (MOE) software package.\textsuperscript{23} The initial LuxS-substrate complex was obtained by appropriately modifying the X-ray crystal structure of catalytically inactive Cys84Ala Co(II)-LuxS (PDB ID: 1YCL) with the putative 2-keto-ribose (2-keto-SRH) intermediate bound within its active site.\textsuperscript{15} Specifically, for one of the active sites the Ala84 residue and Co(II) were
replaced by a cysteiny1 and Fe(II) ion respectively, thus regenerating the wild type active site. The bound 2-keto-SRH moiety in the crystal structure was used as a template to generate the structure of a 'linear aldose SRH' (Scheme 5.3: linear-SRH). For binding of the α- and β-furanose containing anomers of SRH (α-SRH and β-SRH), the 2-keto-SRH and all water molecules were removed from the crystal structure and the substrate analogue 5-S-ethyl-5-thio-D-β-ribofuranose (Scheme 5.3: β-SRH’) was then docked using a Proxy Triangle placement approach in conjunction with the London dG scoring function. The top 30 complexes were then refined using the AMBER99 force field and rescored with the London dG scoring function. The top scoring complex was then selected for further analysis. Initial structures of the enzyme—α-/β-SRH complexes were generated by appropriately replacing β-SRH’ in the above optimized structure with α- and β-SRH, respectively.

The four chemical models of LuxS with α-SRH, β-SRH, linear-SRH, and 2-keto-SRH (Scheme 5.3) bound were solvated in a 7Å spherical layer of water molecules. The damping functional factor included in the electrostatic and Van der Waals potentials were set to decay smoothly beyond 8 – 10 Å. A spherical potential wall with a scaling constant of 2 was placed around each complex. The resulting solvated structures were then minimized using the AMBER99 force field until the root mean square gradient of the total energy fell below 1 kcal mol⁻¹ Å⁻¹. They were then annealed at the same level of theory under constrained pressure and temperature with the equations of motion coupled to a Nosé-Poincaré-Andersen thermostat and a time step for numerical integration of 2 fs. The complexes were then heated and cooled using our previously published protocol of gradually raising the temperature from 150 to 400 K then lowering to 300 K (Appendix Table C.1).


Scheme 5.3 Illustration of the structures of the 2-keto-ribose intermediate and different substrate and substrate-analogues (see text).

The final four annealed structures obtained were then used as the starting points for subsequent 3.5 ns MD simulations. However, now only those residues within 13.5 Å of the substrates or intermediates ribosyl moiety were free to move. The simulations were performed using the AMBER99 force field\textsuperscript{25} within the same parameters and convergence criteria as noted above but at a constant temperature of 300K.

**QM/MM calculations:** The Gaussian 09 program\textsuperscript{26} was used for all two-layer ONIOM QM/MM calculations. Optimized structures and their corresponding harmonic vibrational frequencies were obtained at B3LYP/6-31G(d,p) level of theory for the QM layer and AMBER96 force field for the MM layer, within the mechanical embedding (ME) formalism; ONIOM(B3LYP/6-31G(d,p)//AMBER96)-ME\textsuperscript{25,27-30} Relative energies were obtained via single point energy calculations at the ONIOM(B3LYP/6-311+G(2df,p)//AMBER96)-ME level of theory on the above optimized structures\textsuperscript{30-32}.
Figure 5.1 Schematic illustration of the ONIOM QM/MM active site-containing chemical model with 2-keto-SRH bound: the inner circle indicates those moieties included in the QM region while the outer circle indicates those residues and H$_2$O’s included in the MM layer.

The chemical model used was taken from the X-ray crystal structure of a mutated LuxS with the catalytically viable 2-keto-SRH putative intermediate bound (PDB ID: 1YCL). The chemical model used, with 2-keto-SRH bound, is shown in Figure 5.1. More specifically, the QM layer included the substrate’s 5-S-ethyl-5-thio-D-ribofuranose moiety, the Fe(II) ion and the R-groups of His54A, His58B, and Cys126A to which it is ligated. In addition, the R-groups of the active site residues Cys84B, Glu57A, Ser6B were included as well as an active site water molecule. The residues for the MM layer were chosen, and included in their entirety, on the basis that together they fully surround the QM layer. Several H$_2$O molecules that interact electrostatically with QM layer constituents were also included. In order to ensure
the integrity of the model, a minimal number of select atoms were held fixed at their 
X-ray crystal structure position (Appendix Table C.2).

5.3. Results and Discussion

5.3.1 Docking and MD Simulations

Due in part to the fact that there is an experimental X-ray crystal structure 
(PDB ID: 1YCL) available for comparison, we began the Docking and MD simulation 
studies on substrate binding by examining the enzyme-2-keto-SRH intermediate 
complex. It should be noted that rather than consider the entire complex we have 
focused on the dynamic behavior of the substrate/intermediate itself and several 
key active site residues and the active site water. Specifically, we considered the root 
mean square deviations (RMSD) of the heavy atoms of those chemical functionalities 
in the QM region of the QM/MM model (see Figure 5.1) during the 1 – 3.5 ns time 
period of the simulations.

5.3.1.1 The active site-bound 2-keto-SRH intermediate complex

Within the bound active site complex it is possible that the enzyme residues 
and the substrate are not equally fluxional in their behavior. That is the residues 
may be quite rigidly oriented while the substrate is more mobile, or vice versa. This 
behaviour can be seen in their respective RMSDs over the course of the simulation. 
Hence, in Figure 5.2, for the active site-bound 2-keto-SRH complex, time is plotted 
versus the RMSDs of the: (i) entire QM region (RMSD-QM); (ii) the residues and 
active site water in the QM region (RMSD-QM-residues); and (iii) the bound 2-keto-
SRH intermediate (RMSD-Substrate) itself. It is noted that the potential energy (U) of 
the system was very stable during the 3.5 ns simulation. It should be noted that the
RMSD’s shown are relative to the initial structure (see Computational Methods) of the bound complex at time $t = 0.0$ ps.

In the plot of RMSD-QM versus time (Figure 5.2a) it can be seen that the RMSD-QM’s generally lie in the reasonably broad range of $0.15 - 0.5$ Å and the bound active site as a whole shows a notable degree of fluxional behavior. That is, there is marked mobility in the positions of at least some atoms of the QM region during the course of the MD simulation. However, it does not tell us whether this observed mobility arises from the residues or possibly the 2-keto-SRH intermediate itself.

In Figure 5.2b the RMSDs of only the non-substrate moieties in QM region (RMSD-QM-Residues), which includes the active site residues and water molecule, is plotted against time while in Figure 5.2c the RMSDs of only the 2-keto-SRH moiety are plotted against time. In general, the RMSD-QM-Residues lie in the relatively narrow range of $0.14 - 0.32$ Å, indicating that they are relatively stably positioned during the simulation. In contrast, however, the 2-keto-SRH moiety shows significant fluctuations in its calculated RMSDs over the course of the simulation, ranging from $\sim 0.08 - 0.85$ Å. This clearly indicates that the large fluctuations observed in Figure 5.2a, i.e., the RMSDs of the entire QM region, are due predominantly to the 2-keto-SRH intermediate itself.

In order to gain further insight into which atom(s) within the 2-keto-SRH moiety are most mobile, and by extension possible active site---intermediate interactions, we performed a cluster analysis of the RMSD-QM, i.e., of the entire 'QM region'. Specifically, the RMSD-QM results were grouped into five clusters and an average structure for each cluster obtained. These structures were then overlaid with each other and are shown in Figure 5.3. It should be noted that the average
structures obtained were in good overall agreement with the experimentally
determined X-ray crystal structure (PDB ID: 1YCL).

Figure 5.2 Plots of RMSD’s calculated during the MD simulations for the enzyme···2-keto-SRH intermediate complex: (a) RMSD-QM versus time (ps); (b) RMSD-active site residues versus time (ps); (c) RMSD-substrate versus time (ps).
Figure 5.3 Overlay of select residues and active site moieties of the five (5) average structures obtained from cluster analysis of RMSD-QM’s of the active site bound-2-keto-SRH complex. For clarity, only selected hydrogen atoms are shown. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

The overlaid structures shown in Figure 5.3 clearly show those atoms or groups that vary least and those that vary most in their positioning during the MD simulations, and thus provide insight into the above plots (see Figure 5.2). One can see that the active site residues do not vary greatly in their positioning. However, large variability is seen in several groups within the 2-keto-SRH moiety. More specifically, in approximately half of the structures its –O1H group hydrogen bonds to the imidazole of His11, while in the other half it is instead intra-molecularly hydrogen bonded with the 2-keto-SRH’s own –O4H hydroxyl. It is noted that these two alternative binding modes of –O1H give rise to two average distances between O1 and the nearest imidazole nitrogen of His11 (His11-N) of 2.8 and 3.6 Å. These values bracket the corresponding distance observed in the experimental X-ray crystal structure of 3.04 Å. The 2-keto-SRH –O4H hydroxyl is itself quite consistently
directed towards the sulfur of the homocysteine moiety. However, the variability of its positioning may reflect in part the mobility of the sulfur (see Figure 5.3).

Perhaps more importantly it can be seen that the 2-keto-SRH’s O2 and O3 centres are consistently ligated to the Fe(II) ion over the period of the simulation with average distances of 2.24 and 2.30 Å, respectively. Meanwhile, the –O3H hydroxyl maintains a hydrogen bond with the carboxylate of Glu57. It is noted that the average Glu57-COO···O3 distance of 2.67 Å is in good agreement with that observed in the X-ray crystal structure (PDB ID: 1YCL) of 2.51 Å. Furthermore, the average distance between the thiolate of Cys84 and the substrate’s C3, C2 and C1 atoms is 4.47, 3.75 and 3.48 Å, respectively. Given \( r(\text{CH3}—\text{H}) = 1.33 \text{ Å} \),\(^{33}\) it appears that both supposedly mechanistically important acid/base groups are positioned for their respective proposed proton abstraction roles. A positional change or reaction of the 2-keto-SRH may be required to facilitate proton transfer from -C3H- to the Cys84 thiolate in the next stage of the mechanism.

5.3.1.2 The active site-bound \( \beta \)- and \( \alpha \)-SRH complexes

We then proceeded to examine the active site-bound substrate complexes in which the ribosyl sugar within the SRH substrate was in its either its \( \alpha \)- or \( \beta \)-furanose configuration; \( \alpha \)- and \( \beta \)-SRH respectively.

For the \( \beta \)-SRH···enzyme complex the plot of RMSD’s for the active site-bound substrate region versus time is given in Figure 5.4a. In contrast to that observed for the above active site-bound 2-keto-SRH intermediate complex (cf. Figure 5.2), over the period of the simulation the majority of the points now lie within a much narrower range: approximately 0.17 - 0.38 Å. This suggests that the configuration of
the 'QM-region' of the enzyme is quite consistent with only relatively modest changes observed in the positions of atoms and/or groups.

**Figure 5.4** For the active site-bound b-SRH substrate complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

The five average structures obtained by performing a cluster analysis on the RMSD-QM's are shown overlaid in Figure 5.4b. In agreement with that indicated by Figure 5.4a it can be seen that there is much less variation in the positions and orientations of various active site and substrate functional groups. Notably, both the O2 and O3 ribosyl oxygens of β-SRH are consistently ligated to the Fe(II) centre with average Fe(II)--O distances of 2.30 and 2.26 Å, respectively. These lengths are similar to that obtained above for the corresponding interactions in the 2-keto-SRH intermediate complex. In addition, the -O1H hydroxyl is hydrogen bonded to the R-group carboxylate of an aspartyl residue (Asp37) via a bridging water molecule while the R-group hydroxyl of Ser6 hydrogen bonds to the substrate's sulfur (Figure 5.4b). Importantly, the carboxylate of Glu57 again maintains a consistent strong
hydrogen bonding interaction with the substrates -O3H group. This is illustrated by the fact that in all five structures the -O3H proton is directed towards an oxygen of the carboxylate of Glu57 with an average Glu57-COO⁻--O3 distance of just 2.52 Å. Furthermore, Cys84 is located near the ribose’s -C2H- group with an average Cys84-S⁻--C2 distance of 2.70 Å. Hence, both proposed mechanistic acid/base groups, Glu57 and Cys84, appear suitably positioned for proton abstraction from the -O3H and -C2H- groups respectively.

For the alternate enzyme--α-SRH complex the plot of RMSD’s for the active site-bound substrate region versus time is given in Figure 5.5a. Similar to that observed for the enzyme--β-SRH complex, the plot of RSMD-QM versus time clearly indicates that the active site and bound substrate are again quite consistent in their positioning throughout the simulation with the majority of points lying within the quite narrow range of 0.15 - 0.3 Å.

The five average structures obtained from a cluster analysis of the RMSD-QM analysis are shown overlaid in Figure 5.5b. As can be seen the active site residues are quite static in their positioning while the substrates ribosyl moiety is consistently ligated to the Fe(II) centre via both its O2 and O3 oxygens with average distances of 2.27 Å and 2.25 Å, respectively. As observed for the active site bound alternate anomer β-SRH these distances are in reasonable agreement with those calculated for the active site-bound 2-keto-SRH intermediate complex. Similarly, the -O3H hydroxyl in all structures is strongly hydrogen bonded to the Glu57 R-group carboxylate as indicated by its very short average Glu57-COO⁻--O3 distance of 2.49 Å. In addition, the thiolate sulfur of the other proposed active site base/acid Cys84 is situated near the ribose moieties -C2H- proton with an average Cys84-S⁻--HC2 distance of 2.50 Å.
Figure 5.5 For the active site-bound α-SRH substrate complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

As might be expected the main differences seen between binding of the α- and β-anomers of the SRH substrate involves the groups on C1, in particular the –O1H hydroxyl group. In contrast to that observed for the bound β-SRH substrate, on binding of α-SRH the –O1H group hydrogen bonds to the basic nitrogen of the R-group imidazole of His11 with an average distance of 2.20 Å. In addition, the protonated R-group guanidinium of Arg39 also strongly interacts with the –O1H oxygen; average \( r(\text{Arg39-NH}_2^+\cdots\text{O1}) = 2.76 \) Å.

5.3.1.3 The active site-bound linear aldose-SRH complex

It is proposed\(^{19}\) that the SRH ribosyl moiety must be in its linear aldose configuration (i.e., contain a –C1HO group; see Scheme 5.3) in order for the catalytic mechanism to proceed (see Scheme 5.2). Potentially, this may be formed via ring-opening of the above furanose-containing α- and β-SRH upon their interaction with
LuxS or may be formed within the aqueous solution and then binds within the LuxS active site.\textsuperscript{22} Thus, we also examined the binding of such a 'linear-SRH' moiety. The RMSD-QM's obtained from the MD simulation of the active site-bound linear-SRH complex are plotted against time in Figure 5.6a. It can be seen that the resulting complex is much more dynamic with two distinct regions observed in the measured RMSDs. The first exists in the period of 1 – 1.8 ns with a range of RMSDs of 0.15-0.35 Å while the second begins at 1.8 ns and lasts for the remainder of the simulation. For this latter region the RMSD-QM's now mostly lie in the approximate range of 0.35-0.55 Å (Figure 5.6a).

As for the previous complexes, we then performed a cluster analysis, as described above, of the RMSD-QM’s shown in Figure 5.6a. For each of the five clusters an average structure was obtained and then overlaid with each other as shown in Figure 5.6b. As for the above possible bound substrate and 2-keto-SRH intermediate complexes, the active site residues are quite localized with little positional movement observed over the course of the simulation. Similarly, the ribosyl moieties O2 and O3 oxygens are ligated to the Fe(II) ion with average distances of 2.36 and 2.28 Å respectively while the R-group carboxylate of Glu57 forms a short strong hydrogen bond with the ribose’s –O3H group; average $r$(Glu57-COO\textsuperscript{-}⋯O3-) = 2.52 Å. In contrast to that observed in both the bound α- and β-SRH complexes (see above), the Cys84 thiolate no longer sits above the ribose moiety and the -C2H- group. In contrast, it forms a strong hydrogen bond with the ribose’s –O2H hydroxyl with an average $r$(Cys84-S\textsuperscript{-}⋯HO2-) = 2.05 Å. This altered mechanistic base-substrate interaction could affect the ability of Cys84-S\textsuperscript{-} to act as a base to deprotonate the -C2H- group, the initial reaction in the proposed mechanism. It is likely due in part to this altered Cys84⋯substrate interaction that, unlike in the above enzyme⋯substrate complexes containing a furanose moiety, neither Arg39 or
His11 interact with the substrate. The C1=O carbonyl oxygen instead hydrogen bonds with the water ligated to the Fe(II) ion.

**Figure 5.6** For the active site-bound linear aldose SRH substrate (Linear-SRH) complex: (a) plot of RMSD-QM versus time (ps), and (b) overlay of the five average structures obtained from cluster analysis of the RMSD-QM. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (aqua).

The jump in RMSD-QM’s observed at ~1.8 ns in Figure 5.6a, however, is principally due to the distinct conformational change that occurs in the -CH2CH2S-component of the homocysteine moiety of the substrate (Figure 5.6b). This change is not observed in the active site-bound β- and α-SRH complexes (see Figures 5.4b and 5.5b) nor the bound 2-keto-SRH intermediate (see Figure 5.3c). This may be due in part to the linearity of the substrate and its fewer substrate-enzyme interactions and hence its greater flexibility than in the above three complexes.

### 5.3.2 QM/MM investigation on the catalytic mechanism

#### 5.3.2.1 Active site-bound substrate complexes
The above MD simulations suggest that the SRH substrate may bind within the active site when its ribosyl moiety occurs as either an α- or β-furanose or as a linear aldose. Using a QM/MM approach we then examined the structures and energies of four possible substrate-bound active site complexes: SRH in its α- or β-furanose (α-RC and β-RC) forms, and the corresponding linear aldose configurations (α-RC' and β-RC') which can be thought of as resulting from ring opening of their respective furanose anomers. The optimized structures obtained with selected bond lengths given in Angstroms are shown in Figure 5.7.

The optimized structures of the α-RC and β-RC complexes share a number of similarities, but also exhibit several notable differences. For instance in both, as in their average MD structures, the sugar moiety of the SRH coordinates to the Fe(II) via its O2 and O3 centres (Figure 5.7). In α-RC the Fe(II)···O2/O3 distances are 2.19/2.12 Å while in β-RC the sugar is slightly more tightly ligated with interaction distances of 2.16/2.12 Å respectively. In addition, in both complexes the Ser6 R-group hydroxyl hydrogen bonds to the SRH’s sugar ring oxygen (O4). However, in β-RC it is markedly weaker with a Ser6-OH···O4 distance of 2.84 Å compared to just 2.03 Å in α-RC. In fact in the former the Ser6-OH is positioned closer to the substrate’s sulfur.

Notably, in both complexes the acid/bases, Cys84 and Glu57, appear well positioned for their proposed mechanistic roles. For instance, the thiolate of Cys84 interacts with the ribose’s C2–H proton with Cys84-S···HC2 distances of 2.35 and 2.55 Å in α-RC and β-RC, respectively. Meanwhile, the carboxylate of Glu57 forms very short and strong hydrogen bonds with the ribose’s C3–OH hydroxyl proton with Glu57-COO···HOC3 distances of 1.46 Å (α-RC) and 1.48 Å (β-RC). A water molecule is also found in the active sites of both α- and β-RC. In α-RC, it is hydrogen bonded to the sugar’s O1 centre with r(OH2···O1) = 1.83 Å while in β-RC it is
hydrogen bonded to both the sugar's O2 and O4 centers with distances of $r(\text{OH}_2\cdots\text{O2}) = 1.72 \text{ Å}$ and $r(\text{OH}_2\cdots\text{O4}) = 1.80 \text{ Å}$, respectively.

It is noted that these differences in substrate-enzyme interactions lead to modest differences within the ribose ring. Namely, in $\alpha$-RC the O4–C1 and C1–O1 bonds have lengths of 1.41 and 1.43 Å, respectively. In contrast, in $\beta$-RC these bond lengths are now 1.46 and 1.39 Å, respectively. Thus, the former appears to have a more strongly bound ring while in the latter, the C1–O bond has partial double bond character and consequently, the O4–C1 bond is lengthened.

Significantly, due to these binding and structural differences the $\alpha$-RC active site-bound substrate complex is thermodynamically preferred, lying 47.5 kJ mol$^{-1}$ lower in energy than $\beta$-RC.

The two possible initial 'linear-SRH substrate' complexes were then considered. In $\alpha$-RC' the C1=O oxygen is directed down analogous to the C1-OH group in $\alpha$-RC, while in $\beta$-RC' it is directed upwards analogous to $\beta$-RC. Both complexes exhibit similar enzyme···substrate interactions as observed in their corresponding complexes $\alpha$-RC and $\beta$-RC, respectively. For instance, in $\alpha$- and $\beta$-RC' the SRH moiety is again ligated via its O2 and O3 oxygens to the Fe(II) ion while its O4 centre is hydrogen bonded to the R-group hydroxyl of Ser6 (Figure 5.7). The latter, however, are now markedly shorter with Ser6-OH···O4 distances of 2.02 Å ($\alpha$-RC') and 1.96 Å ($\beta$-RC'). In $\alpha$-RC' the -O2H group is intra-molecularly hydrogen bonded to the C1=O carbonyl oxygen. In contrast, in $\beta$-RC' it is hydrogen bonded via the active site H2O to O4 (Figure 5.7).

In $\alpha$- and $\beta$-RC' the Cys84 thiolate again interacts with the -C2H- proton, though now with markedly shorter distances of 2.06 and 2.41 Å, respectively. Similarly, the R-group carboxylate of Glu57 again strongly hydrogen bonds with their C3–OH protons with Glu57-COO−···HOC3 distances of 1.51 Å ($\alpha$-RC') and 1.56 Å
(β-RC'). Importantly, however, α-RC' and β-RC' lie decidedly higher in energy than α-RC by 50.4 and 107.4 kJ mol⁻¹, respectively! Furthermore, it should be noted that α-RC' and β-RC' are potentially able to interconvert via rotation about their C1–C2 bond.
Figure 5.7 Optimized structures with selected bond distances (angstroms) of the active site-bound SRH substrate complexes in which the substrates ribose sugar is in its α– (α-RC) or β-furanose (β-RC) anomer or corresponding linear aldose configurations (α-RC' and β-RC'), and transition structures for ring opening α-TS and β-TS (see text). Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (purple).

5.3.2.2 Interconversion of active site-bound substrate complexes

The above results suggest that LuxS preferentially binds SRH containing an α- or β-furanose ring rather than the linear aldose (i.e., contain a C1=O group) form. Yet, as previously noted, it has been proposed that the SRH substrate must exist as the latter in order to undergo the subsequent reactions (Scheme 5.2). This suggests that α-RC and β-RC are able to interconvert with their analogous linear complexes α-RC’ and β-RC’, respectively; that is, α- and β-RC are able to undergo a ring-opening process.

There exists two possible approaches for formal transfer of the C1–OH proton onto the ring oxygen O4: (i) direct or unassisted transfer, or (ii) assisted transfer involving a non-substrate active site component. The potential energy surfaces
(PESs) obtained for the former are shown in Figure 5.8. In the unassisted ring opening process the C1-OH proton directly transfers onto the ring oxygen O4. This results in ring-opening with formation of a -C1H=O aldehyde and an -O4H hydroxyl group. Such a reaction step necessarily requires a four-membered ring transition structure (TS) and which are known to often be quite high in energy. Indeed, the ring opening of α-RC to give α-RC' proceeds via α-TS at a cost of 207.6 kJ mol\(^{-1}\) (Figure 5.8). Similarly, for the analogous conversion of β-RC to β-RC' via β-TS the barrier is quite high at 155.6 kJ mol\(^{-1}\). Although this is lower than for α-RC/α-RC', reflecting the structural differences noted in the initial anomeric furanose ring moieties above, it is still significantly higher than that considered the upper thermodynamic limit for an enzyme catalysed reaction; \(\sim\)125 kJ mol\(^{-1}\).\(^{34}\) Thus, it appears that neither 'unassisted ring-opening' process is likely to occur in LuxS.

**Figure 5.8** Schematic PESs for unassisted inter-conversion of (a) α-RC and α-RC', and (b) β-RC and β-RC'.
Possible alternate pathways in which either an active site residue or water assists the ring-opening process were also examined. In both cases, however, there appears to be no suitable residue positioned to facilitate the proton transfer. Alternatively, the active site water near the ribosyl moiety observed in \( \alpha \)- and \( \beta \)-RC (Figure 5.7) may be able to assist transfer of the C1-OH proton onto O4. Unfortunately, no such pathway could be located for either system; all efforts leading to either the above four-membered ring TS or to TS’s for an alternate ring-opening process (see below).

5.3.2.2 The proposed mechanism for formation of the 2-keto-SRH intermediate

In the proposed mechanism, the formation of the 2-keto-SRH intermediate from the bound linear aldose isomer occurs via involvement of Glu57. More specifically, it abstracts a proton from the ribosyl −O2H and transfers it onto the C1=O oxygen (O1). It has been alternatively suggested that this process may involve the −O2H transferring its proton via several residues onto His11, which then protonates O1.\(^{15,19}\) In addition, Cys84-S- assists transfer of a proton from the ribosyl -C2H- group onto the adjacent C1 (see Scheme 5.2). However, in the above structures of \( \alpha \)- and \( \beta \)-RC’, the −O2H proton is directed away from Glu57. For completeness, and as the linear aldose SRH conformers may be bound by LuxS, possible pathways for their reaction to give a bound 2-keto-SRH intermediate were examined. The PES surface obtained for \( \alpha \)-RC’ is shown in Figure 5.9 while optimized structures of stationary points along the pathway with selected bond lengths in Angstroms are presented in Figure 5.10.
Chapter 5. Docking, MD and QM/MM Investigations on LuxS

Figure 5.9 Calculated PES for reaction of the aldehyde-containing active site-bound substrate complex α-RC' to give the 2-keto containing derivative IC2 (see text).

Due to the fact that in α-RC' the ribose's –O2H group is hydrogen bonded directly to the O1 oxygen, proton transfer can occur via the five-member ring TS α-TS1 at a cost of just 6.4 kJ mol⁻¹ with respect to α-RC'1 (Figure 5.9). This occurs with concomitant abstraction of the -C2H- by Cys84–S⁻. This leads to formation of the –O2⁻ oxyanion intermediate α-IC1 lying significantly lower in energy than α-RC' by 199.9 kJ mol⁻¹; -149.5 relative to α-RC (Figure 5.7). In α-IC1, due to the increased oxyanion character of O2, the Fe(II)···O2 interaction has shortened significantly to 1.94 Å. However, while the C1–C2 bond has shortened (1.36 Å) as expected, a decrease in the C2–O2 length to 1.33 Å is also observed (Figure 5.10). This suggests that there is a degree of electron delocalization occurring amongst these bonds.
Figure 5.10 Optimized structures with selected bond distances (angstroms) of $\alpha$-TS1, $\alpha$-IC1, $\alpha$-TS2, IC2; see text. Color key: C (gray); O (red); N (blue); S (yellow); H (white); Fe (purple).

Formation of the corresponding 2-keto-SRH derivative from $\alpha$-IC1 then occurs in one step by transfer of the Cys84-SH thiol proton onto the ribose’s C1 centre via $\alpha$-TS2 with a markedly high relative cost to $\alpha$-IC1 of 121.1 kJ mol$^{-1}$ ($\approx$28.4 kJ mol$^{-1}$ relative to $\alpha$-RC). That is, the barrier for this single reaction step is near the upper thermodynamic limit for an enzymatic reaction.$^{34,35}$ The resulting bound 2-keto-SRH complex IC2 lies considerably lower in energy than $\alpha$-RC’ by 301.6 kJ mol$^{-1}$
or -251.2 kJ mol\(^{-1}\) relative to \(\alpha\)-RC! It is noted that in IC2 the anionic Cys84-S\(^{-}\) sulfur sits quite close to the now quite electrophilic C2 centre with \(r(\text{Cys84-S} \cdots \text{C2}) = 2.06\) Å. It should also be noted that a strong hydrogen bond is formed between -O1H and the basic nitrogen of the R-group imidazole of His11 (His11-N) with an average \(r(\text{His11-N} \cdots \text{HO1}) = 1.68\) Å and \(r(\text{His11-N} \cdots \text{O1}) = 2.64\) Å. This His11-N\cdots O1 distance is in good agreement with that observed in the crystal structure (3.04 Å; PDB ID: 1YCL) and the average distance (2.80 Å) observed in the MD simulation of the enzyme\cdots 2-keto-SRH complex.

Conversion of \(\beta\)-RC′ to its corresponding 2-keto-SRH derivative is not feasible via a pathway analogous to that above for \(\alpha\)-RC′. This is because in \(\beta\)-RC′ the aldehyde oxygen O1 is trans to the ribose’s -O2H group and thus, direct proton transfer from -O2H onto O1 can not occur. However, \(\beta\)-RC′ can first undergo a conformational rearrangement via rotation about its C1\cdots C2 bond to give \(\alpha\)-RC′ and then react as detailed above.

**5.3.2.2 Alternative mechanisms for formation of the 2-keto-SRH intermediate**

The above results suggest that LuxS preferably binds the SRH substrate in which the ribose is in its furanose form, with the \(\alpha\)-anomer being preferred. However, for both \(\alpha\)- and \(\beta\)-RC the reaction barriers for formation of the bound linear-SRH complex (i.e., ring-opening) were enzymatically unfeasible (see Figure 5.8). Hence, possible alternative mechanisms for reaction of \(\alpha\)- and \(\beta\)-RC to form a 2-keto-SRH intermediate were then examined.

In particular, it is noted that in the optimized structures of \(\alpha\)- and \(\beta\)-RC the substrate’s -O2H oxygen is ligated to the Fe(II) ion and as a result, will be more
acidic. Furthermore, in both complexes the active site water is positioned 'under' the ribose ring and could potentially act to facilitate proton transfer from −O2H onto the furanose ring oxygen O4. Such a process could potentially be independent of the anomeric character of the C1 carbon, i.e., α or β. The PESs obtained for the resulting 'water-assisted' pathways for α- and β-RC are shown in Figure 5.11, while optimized structures with selected bond lengths in Angstroms are given in Figure 5.12.

For α-RC the transfer of the −O2H proton via the water onto O4 occurs with concomitant abstraction of the -C2H- proton by the Cys84 thiolate. This reaction step proceeds via α-TS1' with a barrier of 87.8 kJ mol⁻¹ and directly gives the energetically low lying ring-opened 2-enolate intermediate α-IC1 described above. Importantly, it has been formed without the need to pass through the linear aldose complex α-RC'. The subsequent formation of the 2-keto-SRH complex IC2 then proceeds via α-TS2 as detailed above.

Similarly, for β-RC transfer of the −O2H proton via the active site water onto O4 occurs with concomitant abstraction of the -C2H- proton by Cys84-S⁻ to give the analogous ring-opened 2-enolate intermediate β-IC1'. This step proceeds via β-TS1' at a cost of just 60.6 kJ mol⁻¹ relative to β-RC while β-IC1' lies considerably lower in energy than β-RC by 215.8 kJ mol⁻¹. Structurally, β-IC1 is similar to that of α-IC1 in that both the ribose O3 oxygen and the O2, which is formally an oxyanion, are ligated to the Fe(II) centre with distances of 2.23 and 1.97 Å respectively (Figure 5.12). Furthermore, again, electron delocalization is observed across the C2–C1 and C2–O2 bonds as indicated by their bond lengths of 1.36 and 1.34 Å, respectively. The marked double bond character of the C1–C2 bond hinders rotation about the bond, that is, inhibits interconversion of β-IC1 and α-IC1.
Figure 5.11 Calculated PESs for the alternate 'water assisted' pathway for reaction of (a) α-RC to give the 2-keto derivative IC2 and (b) β-RC to give the 2-keto derivative IC2.

However, proton transfer from the now neutral Cys84-SH onto the substrate's C1 centre can proceed via β-TS2' with a barrier of only 82.2 kJ mol$^{-1}$ with
respect to β-IC1'. The resulting 2-keto-SRH intermediate complex (IC2) is the same as that formed by reaction of the alternate initial reactant complex α-RC. This is due to the fact that upon protonation of the C1 centre via either α-TS2 or β-TS2, the C1–C2 bond formally becomes a single bond as indicated by its length of 1.54 Å in IC2 (Figure 5.12). Conversely, the double bond has been localized to the C2=O moiety which now has a length of 1.33 Å in IC2. As a result the intermediate can undergo less hindered rotation about the C1–C2 bond and thus, both possible initial bound anomers (α-RC and β-RC) can convert to the low lying common intermediate IC2. It is noted that in α-TS2 and β-TS2 the anionic Cys84–S–thiolate hydrogen bonds with the R-group guanidinium of Arg37. This would help to stabilize Cys84–S–, as has been previously suggested,⁹ while also enhancing the acidity of a neutral Cys84-SH.

The PESs for reaction of both α-RC and β-RC via the above water-assisted ring-opening processes to form the common 2-keto-SRH mechanistic intermediate IC2, are overlaid and shown in Figure 5.13. As can be seen, despite β-RC having a lower barrier for the initial ring-opening reaction step, it is fact higher than obtained for α-RC due to the higher relative energy of β-RC. Specifically, relative to α-RC, β-TS1' lies 20.3 kJ mol⁻¹ higher in energy than α-TS1' (Figure 5.13). Importantly, however, for both α- and β-RC this initial ring-opening reaction is now enzymatically feasible with barriers of 87.8 and 108.1 kJ mol⁻¹ respectively relative to α-RC.⁴⁴,⁴⁵ Furthermore, this step represents the highest energy barriers for both systems relative to α- or β-RC encountered along the pathway to formation of the IC2.
The resulting 2-oxyanion intermediates formed, α-IC1 and β-IC1', both lie significantly lower in energy than α-RC with β-IC1' being 18.8 kJ mol⁻¹ lower in energy than α-IC1. The subsequent barriers for formation of the 2-keto-SRH intermediate via α-TS2 (−28.4 kJ mol⁻¹) and β-TS2' (−86.1 kJ mol⁻¹) both lie lower in relative energy than α- and β-RC, respectively. However, for both anomers the barriers for this reaction step are distinctly higher than for the preceding reaction.
step via $\alpha$-TS1 and $\beta$-TS1. This suggests that for either $\alpha$-SRH (i.e., $\alpha$-RC) and $\beta$-SRH (i.e., $\beta$-RC) formation of the 2-keto-SRH intermediate (IC2) from the preceding 2-oxyanion intermediate ($\alpha$-IC1 and $\beta$-IC1') is the kinetically limiting step in this initial stage of the overall reaction.\(^{35}\)

**Figure 5.13** Comparison of the PESs for reaction of $\alpha$-RC and $\beta$-RC via a 'water assisted' ring opening process alternative mechanisms.

### 5.4. Conclusions

For S-ribosylhomocysteinate (LuxS), Docking, molecular dynamics (MD) simulations and ONIOM QM/MM methods have been synergistically applied to investigate binding of the S-ribosylhomocysteine (SRH) substrate and possible catalytic pathways for formation of a putative linear 2-keto-SRH mechanistic intermediate.
The Docking and MD simulations showed that both the α- and β-furanose containing substrates could bind within the active site of LuxS to the Fe(II) ion via their ribose’s O2 and O3 oxygens. Furthermore, in both cases a H₂O was also consistently positioned beneath the ribose ring. The main difference in binding of these anomers results from the different positions of the C1-O1H group. In the β-anomer it is hydrogen bonded to the R-group carboxylate of an aspartyl residue (Asp37) via a bridging water while in the alternate α-anomer it instead hydrogen bonds with the R-group imidazole of His11. An alternative linear aldose-containing SRH substrate was shown to also be able to bind within the active site, again, ligating to the Fe(II) via its ribose O2 and O3 oxygens.

Four possible active site-bound substrate complexes were then optimized using an ONIOM QM/MM approach. Specifically, structures were obtained of the active site-bound β- (β-RC) and α-furanose (α-RC) containing SRH complexes. In addition, the active site-bound linear aldose containing SRH complexes β-RC’ and α-RC’, which can be considered as arising from ring opening of β- and α-RC respectively, were obtained. These latter complexes were previously experimentally proposed to be the substrate configuration upon which the enzyme acts.²² Importantly, however, the energetically preferred bound-substrate complex was calculated to in fact be α-RC with β-RC lying 47.5 kJ mol⁻¹ higher in energy. Notably, the bound-aldose containing SRH complexes α-RC’ and β-RC’ lie even higher in energy than α-RC by 50.4 and 107.4 kJ mol⁻¹, respectively.

Significantly, the barriers for interconversion of β-RC with β-RC’, and α-RC with α-RC’ were calculated to occur via four-membered ring transition structures with relative energies of 155.6 and 207.6 kJ mol⁻¹, respectively. Due to their very high barriers these processes are concluded to be enzymatically unfeasible.
However, an alternate mechanism for both β- and α-RC is elucidated in which an active site H₂O facilitates proton transfer from the ribose's Fe(II)-ligated –O₂H group onto its ring oxygen O4. This occurs with concomitant abstraction of the -C₂H-proton by the thiolate of Cys84 and results in opening of the furanose ring with formation of the –O₂⁻ oxyanion-containing intermediates β-IC1’ and α-IC1 respectively. Importantly, the relative energy barriers for this ring opening process are now only 60.6 (β-RC to β-IC1’) and 87.8 (α-RC to α-IC1) kJ mol⁻¹ with respect to β- and α-RC respectively and consequently, are now enzymatically feasible. Notably, both β-IC1’ and α-IC1 are then able to undergo proton transfer from the now neutral Cys84-SH onto C1 with barriers of 82.2 (β-TS2’) and 121.1 (α-TS2) kJ mol⁻¹ respectively. Both processes lead to formation of a common mechanistic 2-keto-SRH intermediate (IC2), which lies significantly lower in energy than α-RC by 251.2 kJ mol⁻¹.

That is, LuxS is able to bind all three possible configurations of the SRH substrate, i.e., containing either an α- or β-furanose ring, the predominate forms in solution, or the linear aldose, with the former α-anomer preferred. The presence of the active site H₂O, however, enables LuxS to be able act upon any of these three possibilities and catalyse their subsequent reaction to form the common mechanistic intermediate 2-keto-SRH.
References


(33) Calculated at the B3LYP/6-31G(d,p) level of theory.


Chapter 6

Tautomerization in the UDP-Galactopyranose Mutase Mechanism: A DFT-Cluster and QM/MM Investigation
6.1 Introduction

Tuberculosis (TB), the well-known disease caused by the pathogenic bacterium *Mycobacterium tuberculosis*, has long afflicted humans. It has been estimated that in 2011 it infected 8.8 million people and resulted in 1.1 million deaths.\(^1\) Unfortunately, the increasing prevalence of multidrug-resistant TB presents major challenges to its diagnosis and present treatments.\(^1\)-\(^3\) Galactofuranose (Gal\(f\)) is a key component of a complex peptidoglycan essential for the cell walls of *M. tuberculosis*.\(^4\),\(^5\) Its precursor, UDP-galactofuranose (UDP-Gal\(f\)), is synthesized by the flavoenzyme UDP-galactopyranose mutase (UGM) which catalyzes its formation from UDP-galactopyranose (UDP-Gal\(p\)) via the overall reaction shown in Scheme 6.1.\(^6\) Importantly, UGM has been shown to be essential for the growth and survival of *M. tuberculosis*\(^1\),\(^7\) and many other pathogenic bacteria such as *Escherichia coli*\(^8\) and *Klebsiella pneumoniae*.\(^9\)-\(^11\) Furthermore, Gal\(f\), does not occur in human cells.\(^5\),\(^12\) Therefore, UGM represents a logical target for new TB therapeutic drugs\(^1\),\(^13\) that are potentially more effective and less toxic to human cells.\(^5\),\(^7\),\(^14\)-\(^21\)

Scheme 6.1 The overall mutase reaction catalyzed by UDP-galactopyranose mutase (UGM).

A number of experimental investigations on UGM have appeared in the literature including X-ray crystallographic, spectroscopic and kinetic analyses, and mutagenesis studies.\(^6\),\(^9\),\(^17\),\(^18\),\(^21\)-\(^26\) However, a clear understanding of its catalytic mechanism is elusive due in part to the fact that UGM binds both NADP(H) and the
highly redox-sensitive cofactor FAD. The latter provides further complications due to its complex redox and protonation states as a result of its isoalloxazine ring that make it unique amongst coenzymes. For example, in 1996 Nassau et al. appeared to show a requirement for NAD(P)H in order to synthesize UDP-Gal; meanwhile they noted that UGM is a flavoenzyme. Liu and coworkers later suggested that it instead uses an oxidized FAD to abstract a hydride from the substrate. Furthermore, they also proposed that the mechanism involves an oxygen-bridged cage-like galactose intermediate. In contrast, Sanders et al. alternatively concluded that the flavin cofactor is in fact fully-reduced and transfers an electron to and from the substrate’s sugar during the course of the reaction. Based in part on thermodynamic analyses, however, it has also been put forth that UGM employs a neutral semiquinone radical (FADH•)-assisted mechanism.

Most recently, Soltero-Higgin et al. concluded that UGM utilizes the fully reduced anionic FAD cofactor, i.e., FADH−. Furthermore, they proposed that its catalytic mechanism proceeds via a covalently bonded cofactor-substrate intermediate. More specifically as shown in Scheme 6.2, the reduced cofactor’s N5 centre (N5FAD) is proposed to nucleophilically attack the anomeric carbon of the substrate sugar (C1Gal), via an SN1 or SN2 mechanism, to give a novel N5FAD—C1Gal covalently bonded intermediate 2. This occurs concomitantly with the cleavage of the glycosidic C1Gal—O-UDP bond and it has been suggested that cleavage of this glycosidic bond "could be" the rate-limiting step. This is followed by opening of the pyranose sugar ring with the heterolytic cleavage of the C1Gal—O5Gal bond while the C1Gal—N5FAD link formally changes from a single to a double bond forming a Schiff base.
Scheme 6.2 The catalytic mechanism of UGM as proposed by Soltero-Higgin et al. and Gruber et al. involving formation of a covalent bond between the N5FAD centre of FAD and the anomeric carbon (CGal) of UDP-galactopyranose.

A furanose ring intermediate 4 is then formed by nucleophilic attack of the sugars –O4GalH oxygen at the CGal position. Finally, a phosphate oxygen of the UDP moiety (OUDP) attacks at the CGal centre to give the desired UDP−Galp product (5) with regeneration of the reduced FADH− cofactor.

Additional experimental evidence supporting this proposed mechanism has been obtained. In particular, kinetic isotope studies have also suggested that the CGal−OUDP bond is indeed cleaved during the mechanism. Furthermore, saturation transfer difference (STD) NMR was used to examine the binding of the FAD cofactor, as well as the effects of redox state in binding was examined. It was concluded that, compared to the oxidized state, the reduced state has the higher binding affinity to the enzyme-product complex UGM−Galp−UDP. Recently, X-ray crystal structures of the enzyme-substrate complex was co-crystallized with either the reduced or oxidized form of FAD for different species, including Klebsiella pneumoniae, Deinococcus radiodurans, and Aspergillus fumigatus. These studies discussed enzyme overall fold, flavin binding, substrate binding, and enzyme-substrate interactions in details. Importantly, it was observed that the UDP-Galp and FAD moieties are positioned towards each other when the latter is in its reduced state. In particular, the CGal−N5FAD distance is 8.0 Å (PDB ID: 3INR) when the flavin is in an oxidized state but only 3.6 Å (PDB ID: 3INT) when the
flavin is reduced. In addition, using the weak reducing agent NaCNBH$_3$ an imine covalently bonded flavin-intermediate complex was obtained.$^{6,10}$

Several computational investigations on UGM have also appeared in the literature.$^{17,18,35}$ These have focused on the application of docking and molecular dynamics (MD) methods to provide insights into the bound substrate and enzyme···substrate interactions. More specifically, the docking studies$^{28}$ considered possible bound conformations of the substrate. Later, MD studies$^{17,18}$ examined substrate binding and showed that a mobile peptidase loop containing an arginyl residue (Arg174) closes over the bound active site. As a result, Arg174 is then able to interact with the substrate's phosphate attached to the sugar ring. One of these also considered binding of the product analogue UDP-[3,F]Galp and its hydrogen bonding interactions with a flavin cofactor and active site residues. Importantly, however, none of these previous investigations examined the catalytic mechanism of UGM.

In this study, density functional theory (DFT) and quantum mechanics/molecular mechanics (QM/MM) methods have been complementarily used, in combination with large active site models, to investigate the mechanism by which UDP-galactopyranose mutase (UGM) catalyses the conversion of UDP-galactopyranose (UDP–Galp) to UDP-galactofuranose (UDP–Galf). In particular, we have examined the feasibility of the flavin neucleophilically attacking mechanism proposed by Soltero-Higgin et al. as shown in Scheme 6.2.$^{6,10}$ From the present study it was found that while FADH$^-$ nucleophically attacks the substrate, a key step assisting the ring opening process involves a tautomerization reaction. In particular, this tautomerization allows for a proton transfer between the coenzyme and substrate sugar moiety during ring opening. This investigation provides insights into
the physiologically important flavoenzymes, in particular, those that involve a
covalent modification between substrate and flavin cofactor.

6.2 Computational methods

The Gaussian 09 suite of programs was used for all calculations.\textsuperscript{29} The DFT-
large active site cluster model calculations utilized the hybrid exchange-correlation
functional B3LYP.\textsuperscript{30-32} Optimized structures and their harmonic vibrational
frequencies were obtained with the 6-31G(d) basis set. Relative energies, with
solvation effects included, were then obtained by performing single-point
calculations on the above geometries at the B3LYP/6-311+G(2df,p) level within an
Integral Equation Formalism Polarizable Continuum Model (IEF-PCM). As is
common when including the general polarity of the protein environment around an
enzyme active site, a dielectric constant ($\varepsilon$) of 4.0 was used.\textsuperscript{33,34} That is, relative
energies were obtained at the IEF-PCM($\varepsilon=4.0$)-B3LYP/6-311+G(2df,p)//B3LYP/6-
31G(d) level of theory. All transition states in the DFT-cluster model were validated
by frequency calculations with a single imaginary frequency to indicate their nature
as a transition state.

For these DFT-based studies a large active site cluster model was obtained
from an X-ray crystal structure of \textit{Klebsiella pneumoniae} UGM with the substrate
"UDP-galactose" bound within its active site (PDB ID: 3INT).\textsuperscript{6} It should be noted that
Partha \textit{et al.} published the structure of \textit{Deinococcus radiodurans} UGM (PDB ID:
3HDQ) with the same substrate "UDP-galactose".\textsuperscript{35} However, the binding mode of
the substrate within the respective active sites is consistent. The differences that
exist involve the distances between the substrate and key active site moieties, the
results of which are shown in the supporting information (Appendix \textbf{Table D.1}). The
largest difference in distance was measured to be 0.5 Å, while the average was found to be 0.3 Å. Thus, the resulting model based on the crystal structure of *Klebsiella pneumoniae* UGM is shown in Figure 6.1. Specifically, the substrate was modeled by 1-methylphosphate-a-Galp while the reduced flavin was represented by lumiflavin. In addition, the \( R \)-groups of the key active site residues Arg174 and Arg280, Tyr314 and His60 were included and modeled by methylguanidinium, 4-methylphenol and 4-methyl-3H-imidazole, respectively. To ensure the structural integrity of the model, atoms marked by an asterisk in Figure 6.1 were held fixed at their X-ray crystal structure position. In aqueous solution histidine’s imidazole group has a \( pK_a \) of approximately 6.0,\(^3\) but this value can differ within a protein environment. Hence, we considered His60 as neutral or protonated (Appendix Figure D.1), which are identified by the superscripts a and b in corresponding potential energy surfaces and for each stationary point. For the latter the resulting mechanism obtained is less favourable than with a neutral His60 (TS 148.1 kJ/mol). Thus, for simplicity, hereafter only the mechanism obtained for when His60 is neutral is discussed in detail, unless otherwise noted. The DFT-cluster approach has been successfully applied to study enzymatic mechanisms and reviewed in detail.\(^3\)\(^-\)\(^4\)

QM/MM calculations were performed using the two-layer ONIOM method as available in Gaussian 09.\(^4\)\(^2\) Optimized structures and their harmonic vibrational frequencies were obtained at the ONIOM(B3LYP/6-31G(d):Amber)-ME level of theory.\(^4\)\(^2\)\(^4\)\(^3\) More accurate relative energies, based on these structures, were obtained by performing single point energy calculations at the ONIOM(B3LYP/6-311+G(2df,p):AMBER) level of theory within the electronic embedding (EE) formalism.\(^4\)\(^2\) The QM/MM models of TS1, TS5, and TS6 were obtained via systematic scans of the PES, while all other TSs were validated by frequency calculations where a single imaginary frequency was obtained to indicate their transition state nature.
The QM/MM substrate-bound active site model was also obtained from the above X-ray crystal structure (PDB ID: 3INT) and is shown in Figure 6.2. Specifically, the QM region included the substrate, Arg174 and Arg280 and the flavin cofactor modeled by 1-methylpyrophosphate-a-Galp, methylguanidinium and lumiflavin, respectively. The MM layer included a number of residues modeled either in their entirety or only including their backbone or R-group (see Figure 6.2). In addition, several H2O observed in the X-ray crystal structure were included in the MM layer. Certain atoms in each residue was held fixed at their X-ray crystal structure position so as to maintain the integrity of the model (Appendix Table D.2).

![Figure 6.1](image-url) The large fully-bound active site model used in the DFT-cluster calculations. Atoms marked by an asterisk (*) were fixed at their X-ray crystal structure (PDB ID: 3INT) position.
Figure 6.2 The QM/MM substrate-bound active site model: the inner circle indicates those moieties placed in the QM layer while the outer circle indicates those placed in the MM layer. Residue color code: included in their entirety (red); backbone was included with R-group modeled by a methyl (blue); only R-groups included (black).

6.3 Results and Discussion

6.3.1 Deprotonation of the initial flavin cofactor

An issue with the mechanism proposed by Soltero-Higgin et al.\textsuperscript{6,10} is that no insight into the loss of the proton from the N\textsubscript{5}FAD–H group of the fully reduced anionic FADH\textsuperscript{-} cofactor is given.\textsuperscript{6,10} In the X-ray crystal structure of UGM with bound flavin cofactor and substrate (PDB ID: 3INT), His\textsubscript{60} is the only polar-basic residue that may be within hydrogen bonding distance of N\textsubscript{5}FAD–H; \( r(\text{N5}_{\text{FAD}} \cdots \text{NHHis}_{60}) = 3.70 \, \text{Å} \), and thus able to act as a base. Experimentally and computationally, the pK\textsubscript{a}'s of histidine and N5 of FADH\textsuperscript{-} in aqueous solution have been predicted to be \(~6 and
greater than 14, respectively.\textsuperscript{36,44,45} Within a protein-like environment the proton affinity of the imidazole of His60 (1161.9 kJ mol\textsuperscript{-1}) is calculated to still be markedly less than that of FAD\textsuperscript{2\-} (1521.1 kJ mol\textsuperscript{-1}).\textsuperscript{46} Thus, His60 is unlikely to be able to deprotonate the FADH\textsuperscript{-} cofactor. Another possibility is that a phosphate of the UDP moiety resulting from the proposed transient dissociation of the substrate may act as the base. However, within the above same X-ray crystal structure (PDB ID: 3INT) the shortest distance between a UDP phosphate oxygen (O_{UDP}) and N\textsubscript{5}_{FAD} is approximately 5.00 Å. Furthermore, these two groups are positioned on almost opposite sides of the substrate's galactose ring. Hence, the UDP moiety is also unlikely to be able to deprotonate N\textsubscript{5}_{FAD}−H. A final alternative is that the proton is lost to solution, however, no water is observed in the UGM active site in the available crystal structures near N\textsubscript{5}_{FAD} (within a distance of 4Å), and therefore, the isoalloxazine ring of the cofactor is sequestered from the surrounding solvent.\textsuperscript{6,35} Consequently, and given the high pK\textsubscript{a} of N\textsubscript{5}_{FADH\textsuperscript{-}}, it would seem unlikely that the N\textsubscript{5}_{FAD}−H proton is lost to an active site water or the aqueous solvent.

![Figure 6.3 PES for the catalytic mechanism of UGM (with a neutral His60) obtained at the IEFPCM(ε=4.0) B3LYP/6-311+G(2d,p)//B3LYP/6-31G(d) level of theory.](image)
6.3.2 Catalytic mechanism obtained using the DFT-cluster model

Alternatively, however, the nitrogen of the FADH− cofactor's N5_FAD−H group may directly nucleophilically attack the substrate's anomeric C1 carbon; that is, without deprotonation of the N5_FAD−H group. The potential energy surface (PES) obtained for the resulting mechanism is shown in Figure 6.3.
6.3.2.1 Formation of a flavin-substrate ($N5_{\text{FAD}}-C1_{\text{Gal}}$) covalent bond

In the fully-bound active site ($^b\text{RC}$) the $N5_{\text{FAD}}$ centre of the cofactor is positioned 3.55 Å from the substrate’s $C1_{\text{Gal}}$ centre, almost opposite the substrate’s phosphate group (Figure 6.4). In addition, it is slightly pyramidal as indicated by a $\angle H5-N5-C4a-C4$ dihedral angle of $-15.0^\circ$ (see Appendix Figure D.2 for FAD$^-\text{numbering}$).

The first step of the UGM mechanism is formation of a covalent bond between the FAD$^-\text{isoalloxazine ring and the substrate’s Galp ring. More specifically, the substrate Galp ring is located underneath the } re \text{ face of the coenzyme isoalloxazine ring, while the imidazole moiety of His60 blocks the } si \text{ face of FAD}^-\text{. Therefore, the cofactor’s } N5_{\text{FAD}} \text{ centre inverts and nucleophilically attacks the substrate’s anomic carbon (} C1_{\text{Gal}} \text{) from the top of the Galp ring while concomitantly the } C1_{\text{Gal}}—O_{\text{UDP}} \text{ bond is cleaved. This } S_N2\text{-type reaction step proceeds via the transition structure (TS) } ^b\text{TS1 with a barrier of } 119.1 \text{ kJ mol}^{-1} \text{ relative to the reactant complex } ^b\text{RC. The resulting covalently bonded intermediate } ^b\text{IC1 is higher in energy than } ^b\text{RC by only } 55.8 \text{ kJ mol}^{-1}. \text{ It is noted that the negative charge has been transferred from the FAD}^-\text{moiety onto the cleaved UDP moiety.}

In $^b\text{IC1}$, an elongated $N5_{\text{FAD}}—C_{\text{Gal}}$ single-bond of length 1.58 Å has now been formed. In addition, the $\angle H5_{\text{FAD}}-N5_{\text{FAD}}-C4a_{\text{FAD}}-C4_{\text{FAD}}$ dihedral angle has increased to $42.9^\circ$, indicating that $N5_{\text{FAD}}$ has increased $sp^3$ character. Meanwhile, the cleaved methyl-phosphate group has been stabilized and neutralized by accepting a proton from Arg280. However, it continues to interact with Arg280 via a quite short and strong hydrogen bond (1.56 Å), slightly shorter than that observed in $^b\text{RC}$ (1.77 Å; Figure 6.4). It should also be noted that the $N5_{\text{FAD}}—H$ bond has lengthened marginally from that observed in $^b\text{RC}$ (1.02 Å) to 1.03 Å (Appendix Table D.2).
Chapter 6. DFT and QM/MM Investigations on UGM Mechanism

It is noted that for the corresponding fully-bound active site model in which His60 is protonated the N$_{\text{FAD}}$--C$_{\text{Gal}}$ distance in the initial reactant complex ($^a$RC; Appendix Table D.2) is slightly shorter at 3.41 Å. However, the barrier for this first reaction step is significantly higher at 140.9 kJ mol$^{-1}$ (Appendix Figure D.1). Importantly, this is markedly higher than what we obtained in the catalytic process with neutral His60.

6.3.2.2 Substrate sugar ring opening via tautomerization

It has been noted that sugar ring opening processes are usually initiated by ring oxygen protonation reactions.$^{36,47,48}$ In $^b$IC1, the now tetrahedral N$_{\text{FAD}}$--H could potentially act as the proton donor. However, direct transfer onto the pyranose ring oxygen (O$_{\text{Gal}}$) would necessarily require a strained four-member ring transition structure. These are generally inherently high in energy. Indeed, the barrier for such a reaction process is calculated to be 189.7 kJ mol$^{-1}$ with respect to $^b$RC (Appendix Table D.2). These are significantly higher than the upper activation energy limit of enzymes$^{41,49}$ and thus, are not feasible.

However, in $^b$IC1 the N$_{\text{FAD}}$--H group forms a weak intra-molecular hydrogen bond with the C$_{\text{FAD}}$=O oxygen; $r$(N$_{\text{FAD}}$--H...O$_{\text{FAD}}$) = 2.51 Å (Figure 6.4). In fact, the FAD–sugar moiety can tautomerize by an intra-molecular proton transfer from N$_{\text{FAD}}$--H to the C$_{\text{FAD}}$=O carbonyl oxygen. This proceeds via the five-membered ring transition structure $^b$TS2 (Figure 6.5) at a cost of only 56.1 kJ mol$^{-1}$ with respect to $^b$IC1; 111.9 kJ mol$^{-1}$ with respect to $^b$RC (Figure 6.3). This barrier is less than that for the preceding nucleophilic attack of N$_{\text{FAD}}$ at C1 of the substrate. Furthermore, it is also markedly lower than for direct proton transfer from N$_{\text{FAD}}$--H onto the pyranose’s O$_{\text{Gal}}$ center via a four-membered ring TS (see above) and in fact is now enzymatically feasible.
Chapter 6. DFT and QM/MM Investigations on UGM Mechanism

The resulting tautomeric intermediate bIC2 formed lies quite low in energy: 33.8 kJ mol\(^{-1}\) lower in energy than bIC1, or just 22.0 kJ mol\(^{-1}\) above that of the initial bound active site \(^{\text{b}}\)RC. It is noted that this relative energy trend for the tautomers of the covalently bonded intermediates is the opposite of that observed for the isolated initial FADH\(^{-}\) cofactor (Appendix Table D.2). In bIC2 the N5\(_{\text{FAD}}\)–C1\(_{\text{Gal}}\) bond formed in the first step has shortened significantly by 0.15 Å to 1.43 Å, and is now slightly shorter than a typical C—N single bond (e.g., \(r(H_3C—NH_2) = 1.44\) Å).\(^{50}\) The change in hybridization of the N5\(_{\text{FAD}}\) center (sp\(^3\) to sp\(^2\)) is further highlighted by the decrease in the angle \(\angle N5_{\text{FAD}}—C5a_{\text{FAD}}—C4a_{\text{FAD}}—C1_{\text{Gal}}\) from 27.6\(^{\circ}\) (bIC1) to 13.2\(^{\circ}\) (bIC2). Related shortenings in the N5\(_{\text{FAD}}\)–C4a\(_{\text{FAD}}\) and N5\(_{\text{FAD}}\)–C5a\(_{\text{FAD}}\) bonds are also observed from 1.47 and 1.48 Å respectively to 1.42 Å each, while the C4\(_{\text{FAD}}\)–O4\(_{\text{FAD}}\) bond has lengthened to 1.35 Å (Figure 6.5).

Notably, in bIC2 the newly formed −O4\(_{\text{FAD}}\)H group forms a short strong hydrogen bond (1.71 Å) with the sugar ring’s 05\(_{\text{Gal}}\) center. Thus, the proton is now well-placed to transfer onto the substrate’s sugar ring oxygen. Indeed, such a process proceeds via the seven-membered ring transition structure bTS3 with a barrier of only 30.1 kJ mol\(^{-1}\) relative to bIC2; 52.1 kJ mol\(^{-1}\) with respect to bRC. It is noted that in bTS3 the C1\(_{\text{Gal}}\)—05\(_{\text{Gal}}\) bond has lengthened significantly by 0.31 Å to 1.77 Å. Concomitantly, the −O4\(_{\text{FAD}}\)H proton is essentially wholly transferred onto 05\(_{\text{Gal}}\) as illustrated by the fact that \(r(05_{\text{Gal}}—\cdot—H) = 1.10\) Å (Figure 6.5), thus helping to stabilize the charge build-up on 05\(_{\text{Gal}}\). Notably, the N5\(_{\text{FAD}}\)–C1\(_{\text{Gal}}\) bond has shortened by 0.12 Å to 1.35 Å indicating a marked increase in double-bond character. Additional changes are also observed in the cofactor’s structure such as shortening of the C4—O bond and lengthening of both the C4a—C4 and N5—C4a bonds (Figure 6.5). This also seems to suggest that proton transfer from the cofactor tautomer onto
O$_5$$_{\text{Gal}}$ and the resulting associated changes within the cofactor may in fact help facilitate ring opening of the sugar.

![Diagrams of TS2, IC2, TS3, and IC3](image)

**Figure 6.5** Optimized structures with selected bond lengths (Angstroms) of $^b$TS2, $^b$IC2, $^b$TS3 and $^b$IC3 (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).

The resulting imine covalently bonded linear sugar-flavin tautomer $^b$IC3 lies just 13.9 kJ mol$^{-1}$ higher in energy than $^b$RC (Figure 6.4). That is, it is 8.1 kJ mol$^{-1}$ lower in energy than the preceding glycosidic covalently bonded cyclic sugar-flavin tautomer $^b$IC2. In $^b$IC3 the C$_1$$_{\text{Gal}}$---O$_5$$_{\text{Gal}}$ distance has increased markedly to 2.48Å, that is, the bond has broken and the galactose is now in a linear conformation.
(Figure 6.5). In addition, the N5\textsubscript{FAD}=C1\textsubscript{Gal} distance has decreased even further to just 1.31 Å. The newly formed –O5\textsubscript{Gal}H group, however, remains hydrogen bonded to the cofactor’s C4\textsubscript{FAD}=O oxygen. It is noted that sugar ring opening with formation of a neutral –O5\textsubscript{Gal}H group rather than the proposed anionic –O5\textsubscript{Gal}\textsuperscript{−} moiety\textsuperscript{6,10} may enable nucleophilic attack of the –O4\textsubscript{Gal}H oxygen at the C1\textsubscript{Gal} center to be competitive with re-attack by O5\textsubscript{Gal}. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{optimized_structures.png}
\caption{Optimized structures with selected bond lengths (Angstroms) of \textsuperscript{b}TS4, \textsuperscript{b}IC4, \textsuperscript{b}TS5 and \textsuperscript{b}IC5 (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).}
\end{figure}
The above observed reaction steps also provide insights into the experimental\textsuperscript{6,9} observation that the catalytic mechanism of UGM requires a fully reduced flavin cofactor; the oxidized form rendering the enzyme inactive.\textsuperscript{6,24,26,51} In particular, the reduced form is required for the formation of the N5\textsubscript{FAD}=C1\textsubscript{Gal} double bond. Furthermore, the N5\textsubscript{FAD}−H proton, lacking in oxidized FAD, is required for the ring opening process and neutralization of the O5\textsubscript{Gal} centre.

### 6.3.2.3 Conformational rearrangement of the linear sugar moiety

In \textsuperscript{b}IC3 the \textsuperscript{−}O4\textsubscript{Gal}H group is also hydrogen bonded with the C4\textsubscript{FAD}=O moiety, r(O4\textsubscript{Gal}H···OC4) = 1.80 Å. More importantly, however, it is not suitably positioned for nucleophilic attack at C1\textsubscript{Gal} to form the desired Galf ring. Hence, the linear sugar moiety must first undergo an appropriate conformational rearrangement. Such a process can be achieved via two sequential rotations.

First, the \textsuperscript{−}O5\textsubscript{Gal}H group rotates such that rather than being hydrogen bonded to the C4\textsubscript{FAD}=O oxygen, it instead forms a hydrogen bond with a phosphate oxygen of the initial substrate’s cleaved but still active site-bound UDP moiety. More specifically, this conformational change occurs via \textsuperscript{b}TS4 with a barrier of only 26.1 kJ mol\textsuperscript{−}1 relative to \textsuperscript{b}IC3 to give \textsuperscript{b}IC4 lying just 5.4 kJ mol\textsuperscript{−}1 higher in energy than \textsuperscript{b}IC3. Importantly, structurally in \textsuperscript{b}IC4 the \textless O4\textsubscript{FAD}···C5\textsubscript{Gal}·O5\textsubscript{Gal}·H5\textsubscript{Gal} dihedral angle has changed from \textminus13.2° (\textsuperscript{b}IC3) to 137.0°. This indicates that the \textsuperscript{−}O5\textsubscript{Gal}H group is now directed away from the C4\textsubscript{FAD}=O and towards a UDP phosphate oxygen with which it now forms a short strong hydrogen bond (1.73 Å). It is also noted that this rotation also induces a slight shift in the positioning of the sugar moieties \textsuperscript{−}O4\textsubscript{Gal}H group. In particular, the \textsuperscript{−}O4\textsubscript{Gal}H···OC4\textsubscript{FAD} hydrogen bond has shortened and strengthened slightly by 0.04 Å to 1.76 Å (Figure 6.6).
However, as can be seen in Figure 6.6, the O4_Gal oxygen in bIC4 is still not ideally positioned to nucleophilically attack the C1_Gal centre, necessary for formation of the GalF ring. Hence, a further shift in the positioning of the −O4_GalH hydroxyl group is required. More specifically, it undergoes a rotation around the sugar moieties C3—C4 bond via the transition structure bTS5 to give the alternative conformer bIC5 lying just 24.8 kJ mol⁻¹ higher in energy than bIC4 (Figure 6.3). It is noted that the relative energy of bTS5 is slightly lower than that of bIC4. This is a common artifact of single-point energy calculations with corrections (e.g., solvation). It simply indicates that within the present DFT-cluster model, bIC5 is predicted to be able to rearrange essentially without a barrier back to bIC4. Structurally, in bIC5 the −O4_GalH group still maintains a short strong hydrogen bond with the C4_FAD=O group (1.77 Å). More importantly, however, O4_Gal is now orientated over the C1_Gal centre with an O4_Gal···C1_Gal distance of only 2.49 Å (Figure 6.6). Thus, it is now better situated for the nucleophilic attack at C1_Gal than O5_Gal, which has further shifted position slightly but remains hydrogen bonded to a UDP phosphate oxygen.

### 6.3.2.4 Formation of the galactofuranose ring

This third and final stage of the overall catalytic mechanism is in essence the analogous reverse of the initial substrate reaction with the FADH⁻ cofactor and the subsequent ring opening process (see 6.3.2.1 and 6.3.2.2).

For instance, it is initiated by nucleophilic attack of the −O4_GalH hydroxyl's oxygen at the C1_Gal center with concomitant transfer of its proton to the cofactor moieties C4_FAD=O carbonyl oxygen. This reaction occurs via the seven-membered ring transition structure bTS6 and requires just 9.3 kJ mol⁻¹ with respect to bIC5 to give the furanose-containing intermediate bIC6. The latter species lies quite low in
relative energy at 17.2 kJ mol$^{-1}$ above that of $^b$RC (Figure 6.3). Notably, as can be seen in Figure 6.7, in $^b$IC6 the cofactor moiety has now regenerated its C4–OH containing tautomer. Furthermore, the protonation of the cofactor’s C4$_{FAD}$–O oxygen has caused a significant lengthening in its bond by 0.08 Å to 1.34 Å. This concomitantly induces modest shortenings in both the C4a$_{FAD}$—C4$_{FAD}$ and N5$_{FAD}$—C4a$_{FAD}$ bonds. Consequently, the C1$_{Gal}$—N5$_{FAD}$ cross-link has lengthened significantly to 1.43 Å, such that it now resembles a more typical N-glycosidic single bond. Importantly, the C1$_{Gal}$—O4$_{Gal}$ distance has also decreased markedly by 1.02 Å to just 1.47 Å; a typical C—O single bond length. It is noted that the furanose-flavin tautomer complex $^b$IC6 is marginally lower in energy by 4.8 kJ mol$^{-1}$ lower in energy than the analogous pyranose-containing complex $^b$IC2 (Figure 6.3).

The subsequent reaction step is formation of the N5$_{FAD}$—H containing tautomer, in which the flavin moiety is now covalently bonded with the furanose. That is, transfer of the proton from the C4$_{FAD}$–OH group onto the spatially adjacent N5$_{FAD}$ centre. This occurs via the five-membered ring transition structure $^b$TS7 at a cost 120.0 kJ mol$^{-1}$ relative to the initial reactant complex $^b$RC (Figure 6.3). This is the highest reaction barrier calculated overall for the UGM catalytic mechanism using the present DFT-cluster model, and is close to the generally held upper thermodynamic limit of enzymatic mechanisms.$^{41,49}$ This likely reflects in part the computational model (see below), the constrained ring involved in the transfer, and structural changes such as slight deformation of the flavin and breaking of the O4$_{Gal}$···HOC4$_{FAD}$ hydrogen bond in $^b$IC6. In the resulting intermediate $^b$IC7, lying 33.5 kJ mol$^{-1}$ above that of $^b$RC, the N5$_{FAD}$ center has once again become tetrahedral (Figure 6.7). This causes the N5$_{FAD}$—C1$_{Gal}$ linkage to weaken and hence lengthen considerably by 0.19 Å to 1.62 Å. As a result, the C1$_{Gal}$ center is now again susceptible to substitution. In particular, it is noted that the UDP phosphate oxygen
(UDP-O\textsuperscript{2}) initially cleaved from C\textsubscript{1Gal} in the first step of the overall mechanism is situated 3.89 Å from C\textsubscript{1Gal} in \textsuperscript{b}IC\textsubscript{7}, hydrogen bonded to the guanidinium of Arg174.

Scheme 6.3 Schematic summary of the reactant, intermediates and product complexes obtained for the overall mechanism of UGM using the DFT-cluster approach (see text).

Indeed, the final step in the overall mechanism is nucleophilic attack by a phosphate oxygen of the UDP moiety at the furanose's C\textsubscript{1Gal} centre to form the final UDP-galactofuranose product complex (\textsuperscript{b}PC). This occurs via \textsuperscript{b}TS\textsubscript{8} with a barrier of 44.2 kJ mol\textsuperscript{-1} with respect to \textsuperscript{b}IC\textsubscript{7}, or 77.7 kJ mol\textsuperscript{-1} relative to \textsuperscript{b}RC (Figure 6.3). In \textsuperscript{b}PC, lying only 11.8 kJ mol\textsuperscript{-1} above \textsuperscript{b}RC, the N\textsubscript{5FAD}···C\textsubscript{1Gal} distance has now increased considerably to 3.53 Å, that is, the N-glycosidic sugar-cofactor cross-link has been cleaved (Figure 6.7). Meanwhile, a UDP phosphate oxygen–C\textsubscript{1Gal} bond has been re-formed with a length of 1.47 Å.
Figure 6.7 Optimized structures with selected bond lengths (Angstroms) of \( ^b\)TS6, \( ^b\)IC6, \( ^b\)TS7, \( ^b\)IC7, \( ^b\)TS8 and \( ^b\)PC (see text). For clarity, only some residues are shown in ball and stick format. Color key: C(grey); O(red); N(blue); P(orange); H(white).
The reactant, product and intermediate complexes observed along the above overall mechanism using the above DFT-cluster approach are summarized in Scheme 6.3. It is noted that for the corresponding DFT-cluster model in which His60 is protonated, the lowest energy pathway obtained followed the same analogous sequence (Appendix Figure D.1). Similarly, the highest overall barrier again occurred for the second last step; tautomerization of the reformed FADH\(^-\) covalently bonded cofactor (i.e., \(\text{b}^\text{TS7}: \text{bIC6} \rightarrow \text{bIC7}\)).

### 6.3.3 Catalytic mechanism obtained using the QM/MM model

In the DFT-cluster approach, the residues and protein environment surrounding the active site are simply and implicitly modeled as a homogenous polar continuum. Thus, in order to explicitly consider the effects of these residues, their sterics and the non-homogenous polarity of the surrounding environment on the above catalytic mechanism, an ONIOM QM/MM approach was used (see Computational Methods). The catalytic pathway obtained is shown in Figure 6.8 and overall, is the same as obtained using the DFT-cluster approach. However, several key differences in the optimized structures of the various complexes and in particular, their relative energies were observed.

For instance, in contrast to the single Arg280···substrate interaction observed in \(\text{bRC}\), in the QM/MM optimized reactant complex RC, Arg280 is now hydrogen bonded to both a phosphate oxygen of the substrate’s UDP moiety and its pyranose ring oxygen (O\(_5\)). In addition, Arg174 hydrogen bonds with both phosphates of the UDP.
Figure 6.8 PES for the catalytic mechanism of UGM (with a neutral His60) obtained at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96)-EE//ONIOM(B3LYP/6-31G(d): Amber96)-ME level of theory.

These additional phosphate···arginyl interactions help stabilize the negative charge on the substrate’s phosphate groups. Indeed, the barrier for the first step, the S_N2 nucleophilic attack of N5_FAD at C1_Gal to form an N5_FAD(H)···C1_Gal cross link with concomitant dissociation of the UDP moiety, occurs via TS1 with a barrier now of only 79.1 kJ mol^{-1} (Figure 6.8); a decrease of 40.0 kJ mol^{-1} from that observed for ^\text{b}TS1 (cf. Figure 6.3). The comparatively "earlier" occurrence of TS1 in the reaction step than ^\text{b}TS1 is also illustrated by the fact that in the former the \text{UDP}O···C1_Gal distance is 0.03 Å shorter at 2.02 Å while the N5_FAD···C1_Gal distance is significantly longer by 0.77 Å at 3.37 Å (Appendix Table D.3). Again, a shift towards planarity of the C1_Gal centre and slight increased pyramidalization of N5_FAD is observed (Appendix Table D.3). It is also noted that during this step four tyrosyl residues in the MM layer also hydrogen bond with the substrate’s pyrophosphate group, helping to further stabilize its anionic charge. The resulting flavin-Galp covalently bonded complex IC1 has an elongated N-glycosidic bond with r(N5_FAD···C1_Gal) = 1.61 Å.
Furthermore, it lies 40.3 kJ mol\(^{-1}\) higher than RC (Figure 6.8), a decrease of 15.5 kJ mol\(^{-1}\) compared to the DFT-cluster results (cf. Figure 6.3).

The barrier for the subsequent tautomerization to give the C4\(_{\text{FAD}}\)-OH containing complex IC2 occurs via TS2 and is similarly reduced by approximately 37.2 kJ mol\(^{-1}\) to 74.7 kJ mol\(^{-1}\) relative to RC (Figure 6.8). This may reflect in part the lower relative energy of IC2 which, unlike that observed on the DFT-cluster PES (cf. Figure 6.3), now lies below that of RC by \(-3.3\) kJ mol\(^{-1}\). It should also be noted that no suitably positioned residues were observed in IC2 that may be able to help stabilize a build-up of negative charge on the O5\(_{\text{Gal}}\) centre. The nearest residue hydrogens were those of an R-group methyl of Ile61 (4.10 Å) and a -CH\(_2\)- of Pro59 (3.72 Å); see Appendix Table D.3. Consequently, opening of the pyranose ring, cleavage of the O5\(_{\text{Gal}}\)—C1\(_{\text{Gal}}\) bond, again occurs with concomitant transfer of the proton from the C4\(_{\text{FAD}}\)-OH group onto O5\(_{\text{Gal}}\). This proceeds via TS3 at a cost of 47.1 kJ mol\(^{-1}\) relative to RC (Figure 6.8), only 5.0 kJ mol\(^{-1}\) lower than obtained using the DFT-cluster approach. The resulting imine covalently bonded flavin-linear sugar intermediate IC3 is 30.6 kJ mol\(^{-1}\) higher in energy than RC (Figure 6.8). This is an increase of 16.7 kJ mol\(^{-1}\) compared to that obtained for the analogous reaction on the DFT-cluster PES (cf. Figure 6.3). It is noted that of the three "linear" sugar containing intermediates encountered along the QM/MM mechanism pathway (IC3, IC4 and IC5), the IC3 conformation is the highest in energy (Figure 6.8). This is the opposite of that obtained using the DFT-cluster method (cf. Figure 6.3). As observed in the analogous complex \(^{1}\)IC3 (cf. Figure 6.5), the newly formed −O5\(_{\text{Gal}}\)H group is hydrogen bonded to the flavin's regenerated C4\(_{\text{FAD}}\)=O carbonyl oxygen with a length of 1.90 Å (Appendix Table D.3). Thus, again, the substrate-cofactor complex must undergo a conformational change so as to favourably position the −O4\(_{\text{Gal}}\)H oxygen for nucleophilic attack at the C1\(_{\text{Gal}}\)=N5\(_{\text{FAD}}\) carbon centre.
In fact, the rate-limiting step on the QM/MM PES is calculated to be the conformational change in which the \( -05_{\text{Gal}} \) group rotates so that it instead hydrogen bonds to a phosphate oxygen of the cleaved UDP moiety rather than the flavin’s C4=O carbonyl oxygen. This occurs via TS4 at a cost of 99.2 kJ mol\(^{-1}\) relative to RC. This is 59.2 kJ mol\(^{-1}\) higher than that obtained using the DFT-cluster approach (cf. Figure 6.1)! This markedly higher barrier is likely due to the inclusion of additional active site residues in the larger QM/MM active site model. Furthermore, the residues in the vicinity of the \( -05_{\text{Gal}} \) group in TS4 are predominantly hydrophobic (Appendix Table D.3). Thus, they do not facilitate the conformational change through, for instance, favourable polar or hydrogen bonding interactions. In the resulting complex IC4, lying only slightly higher in energy than RC by 6.9 kJ mol\(^{-1}\), the \( -05_{\text{Gal}} \) and \( -04_{\text{Gal}} \) groups are hydrogen bonded with a UDP phosphate oxygen and the C4\(_{\text{FAD}}\)=O oxygen, respectively.

Similar to that observed using the DFT-cluster approach, a further slight conformational change is required in order to reposition the \( -04_{\text{Gal}} \) oxygen for nucleophilic attack at the C1\(_{\text{Gal}}\)=N5\(_{\text{FAD}}\) carbon. This step occurs via TS5 with a barrier of 56.4 kJ mol\(^{-1}\) relative to RC to give the alternate conformer IC5 (Figure 6.8). In contrast to that observed with the above DFT-cluster approach, the latter complex is quite low in energy at 14.8 kJ mol\(^{-1}\) with respect to RC. Additionally, it is no longer calculated to be able to undergo a thermodynamically barrierless rearrangement back to IC4. Notably, the O4\(_{\text{Gal}}\) oxygen is now similarly positioned to O5\(_{\text{Gal}}\) in IC3; almost above and in proximity to the C1\(_{\text{Gal}}\) centre.

Indeed, nucleophilic attack of O4\(_{\text{Gal}}\) at C1\(_{\text{Gal}}\) occurs via TS6 with a low cost of only 26.2 kJ mol\(^{-1}\) relative to RC (Figure 6.8). In contrast to that observed on the DFT-cluster PES (cf. Figure 6.3), the resulting N-glycosidic covalently bonded flavin tautomer-furanose intermediate IC6 lies 10.9 kJ mol\(^{-1}\) lower in energy than RC. This
is likely due, however, to enhanced stabilization of mechanistic intermediates and products in general using the extended QM/MM model. For instance, at the QM/MM level the relative energies of all intermediates, except IC3, is lowered by 12.7 – 63.2 kJ mol\(^{-1}\) with respect to that observed on the DFT-cluster PES (cf. Figure 6.3).

Using the DFT-cluster method the subsequent tautomerization reaction via \(^{b}TS7\), i.e., conversion of the C4\(_{\text{FAD}}\)-OH containing tautomer to that having a tetrahedral protonated N5\(_{\text{FAD}}\) centre, had the overall highest barrier at 120.0 kJ mol\(^{-1}\) (Figure 6.3). At the QM/MM level, however, this barrier is reduced by 44.4 kJ mol\(^{-1}\) to 75.6 kJ mol\(^{-1}\) (TS7) relative to RC and is no longer the rate-limiting step. The tautomeric intermediate IC7 lies markedly lower in energy than RC by 29.7 kJ mol\(^{-1}\). Furthermore, it contains a tetrahedral N5\(_{\text{FAD}}\) centre with an N5\(_{\text{FAD}}\)—C1\(_{\text{Gal}}\) glycosidic bond length of 1.56 Å, which is 0.06 Å shorter than observed in \(^{b}IC7\), i.e., at the DFT-cluster level (cf. Figure 6.7). However, it is markedly lower in energy than the corresponding pyranose-cofactor complex IC1 by 70.0 kJ mol\(^{-1}\). At the DFT-cluster level \(^{b}IC7\) was only modestly lower in energy than \(^{b}IC1\) by 22.3 kJ mol\(^{-1}\) (cf. Figure 6.3). This noticeably lower relative energy for IC7 may in part reflect a reduction in intra-molecular strain. For instance, in the optimized structures of \(^{b}IC7\) and IC1 there is an intra-molecular hydrogen bond between O4\(_{\text{FAD}}\) and the sugar’s -O3\(_{\text{Gal}}\)H or -O4\(_{\text{Gal}}\)H groups, respectively. In contrast, in the optimized structure of IC7 the C4\(_{\text{FAD}}\)O and -O3\(_{\text{Gal}}\)H groups instead form inter-molecular hydrogen bonds with the backbone amide of Leu252 and the R-group of Asp351 via a water molecule, respectively. This helps to reduce the strain in the sugar-cofactor complex as illustrated by the fact that the N5\(_{\text{FAD}}\)—C1\(_{\text{Gal}}\) bond length in IC7 (1.70 Å) is 0.09 and 0.08 Å longer than observed in IC1 or \(^{b}IC7\), respectively.

The final step is again nucleophilic attack of an oxygen of the terminal phosphate of the UDP moiety at the furanose’s C1\(_{\text{Gal}}\) centre with concomitant
cleavage of the N5\textsubscript{FAD}—C1\textsubscript{Gal} bond. For this reaction the barrier via TS8 is only 33.2 kJ mol\(^{-1}\) higher than IC7 or 3.5 kJ mol\(^{-1}\) with respect to RC. This is 69.6 kJ mol\(^{-1}\) lower than obtained using the DFT-cluster model (Figure 6.3). The "earlier" occurrence of TS8 compared to \(^{b}\)TS8 in the reaction is illustrated by the fact the forming \(^{\text{UDP}}\text{O} \cdots\text{C1 Gal}\) bond is 0.29Å longer at 2.38 Å while the cleaving C1\textsubscript{Gal}—N5\textsubscript{FAD} bond is 0.03 Å shorter at 2.52 Å (Appendix Table D.3). Furthermore, the cleaved flavin moiety formally regenerates the FADH\(^{-}\) cofactor. The resulting final active site bound-product complex PC is significantly lower in energy than RC by \(-74.2\) kJ mol\(^{-1}\) (Figure 6.8). Hence, unlike the overall endergonic mechanism obtained using the DFT-cluster approach (Figure 6.3), using the QM/MM approach it is now thermodynamically favoured, being notably exergonic. This may be partly due to the fact that unlike for RC, all of the sugar hydroxyl groups and the non-bridging phosphate oxygen’s in PC form hydrogen bonds with active site residues, in particular the four Tyr residues (314, 349,185, 198), Asp351, His63 and Asn84.

It is noted that experimentally, the effectiveness of several anti-mycobacterial agents have been examined as potential inhibitors of UGM including sugar substrate analogues and some small cell permeable compounds that can bind in the enzyme active site.\(^{5,7,15,19,34}\) However, as of yet, no sugar-substrate-analogues have been found to effectively inhibit UGM.\(^{7}\) Furthermore, some of the small cell permeable inhibitors have been found to be toxic.\(^{5}\) Therefore, most of these trials have proven less successful.

### 6.4 Conclusions

The catalytic mechanism of the flavoenzyme UDP-galactopyranose mutase (UGM) has been investigated via the complementary use of DFT-cluster and ONIOM
QM/MM approaches. This has enabled not only examination of the reaction between the substrate UDP-galactopyranose (UDP-Galp) and the fully reduced flavin cofactor (FADH\textsuperscript{-}), but also to consider the role of various active site residues and the protein environment in general.

The first step in the overall mechanism is nucleophilic attack of the flavin’s N5\textsubscript{FAD} centre at the anomeric carbon C1\textsubscript{Gal} of the substrate’s sugar moiety with concomitant heterolytic cleavage of the substrate’s UDP phosphate oxygen–C1\textsubscript{Gal} bond. This result is in agreement to the S\textsubscript{n}2 reaction process proposed by Soltero-Higgin \textit{et al.} (Scheme 6.2).\textsuperscript{6,10} This is then followed by a series of tautomerization reactions involving transfer of the N5\textsubscript{FAD}−H proton that lead to and result in opening of the Galp ring. Specifically, the flavin moiety first undergoes a tautomerization with the proton being transferred from N5\textsubscript{FAD} onto the C4\textsubscript{FAD}=O oxygen to give a C4\textsubscript{FAD}−OH containing complex (\textsuperscript{b}IC2/IC2). The proton is then transferred onto the ring oxygen O5\textsubscript{Gal} resulting in formation of a linear galactose sugar (\textsuperscript{b}IC3/IC3); a chain-ring tautomerization process. Such a result, clearly points to the necessity of FADH\textsuperscript{-} and not the oxidized form. The latter of which has no proton to neutralize the charge build-up on O5\textsubscript{Gal} during O5\textsubscript{Gal}—C1\textsubscript{Gal} bond cleavage.

The existence of the linear sugar intermediate obtained by both DFT-cluster and QM/MM models agrees with the corresponding intermediate structure directly or indirectly proposed by UV-Vis spectroscopy,\textsuperscript{10} mass spectral analysis,\textsuperscript{10} and \textsuperscript{1}H NMR spectrum experiments.\textsuperscript{6} Furthermore, it also enables the later nucleophilic attack of the −O4\textsubscript{Gal}H oxygen at C1\textsubscript{Gal} to form the furanose ring, to be competitive with re-attack of O5\textsubscript{Gal} at C1\textsubscript{Gal}. Suitable positioning of O4\textsubscript{Gal} for such a nucleophilic attack is achieved via a two-step conformational rearrangement, \textsuperscript{b}IC3/IC3 → \textsuperscript{b}IC4/IC4 → \textsuperscript{b}IC5/IC5, that results in positioning of the −O4\textsubscript{Gal}H group over the C1\textsubscript{Gal} centre. Furthermore, it forms a hydrogen bond to the C4\textsubscript{FAD}=O oxygen. Furanose ring
and final product formation is then achieved via the analogous reverse of the ring opening process. Specifically, $^b$IC5/IC5 undergoes chain-ring tautomerization to form the furanose ring containing complex $^b$IC6/IC6. The latter complex now also contains a C4$_{\text{FAD}}$−OH group and can then tautomerize to give the tetrahedral-N5$_{\text{FAD}}$ intermediate $^b$IC7/IC7. Formation of the final desired product UDP-galactofuranose is achieved via nucleophilic attack of a phosphate oxygen of the UDP moiety at C1$_{\text{Gal}}$ with concomitant cleavage of the C1$_{\text{Gal}}$—N5$_{\text{FAD}}$ glycosidic bond.

In the DFT-cluster approach used herein, the highest barrier in the overall mechanism at 120.0 kJ mol$^{-1}$ ($^b$TS7) is calculated to be tautomerization of the imine covalently bonded furanose ring containing intermediate. In other words, for transfer of the proton from C4$_{\text{FAD}}$−OH once the furanose ring has been formed ($^b$IC6), onto N5$_{\text{FAD}}$ to give the tetrahedral-N5$_{\text{FAD}}$ intermediate $^b$IC7. This barrier is only slightly higher by 0.9 kJ mol$^{-1}$ than that calculated for the initial nucleophilic attack of the flavin's N5$_{\text{FAD}}$ at the substrate's C1$_{\text{Gal}}$ centre; i.e., RC $\rightarrow$ $^b$IC1.

More extensive inclusion of the residues and protein surrounding the active site and its non-homogeneous polarity were modeled using a QM/MM approach. In contrast, to that observed on the DFT-cluster PES, the highest overall barrier was obtained for the required conformational changes in TS4. More specifically, the barrier for rotation of the −O$_{\text{GalH}}$ group from hydrogen bonding to the C4$_{\text{FAD}}$=O oxygen to a phosphate oxygen of the initial substrate-derived UDP moiety was calculated to be 99.2 kJ mol$^{-1}$. Furthermore, unlike that obtained at the DFT-cluster level, the overall reaction was calculated to be quite exergonic with the final product complex (PC) lying markedly lower in energy than the initial reactant complex (RC) by 74.2 kJ mol$^{-1}$.

These results provide new insights into the catalytic mechanism of a novel flavoenzyme but also into the remarkable possible chemistry of flavin cofactor. In
addition, they also help to explain why an oxidized flavin cofactor appears to be a catalytic inhibitor to UGM. These mechanistic insights can also be used to help develop potential new inhibitors that exploit the tautomerization chemistry in the UGM mechanism.
References

(14) Müller, F. Chemistry and biochemistry of flavoenzymes; CRC Press, Inc., 1990; Vol. I.


(46) All acidity results were calculated at the IEF-PCM($\varepsilon=4.0$)-B3LYP/6-31G(d) level of theory.


(50) The bond length was calculated at the B3LYP/6-31G(d) level of theory in the gas phase.

The $\alpha$-Amino Group of the Threonine Substrate As The General Base During tRNA Aminoacylation: A New Version of Substrate Assisted Catalysis Predicted by Hybrid DFT
7.1 Introduction

Proteins are the 'work horses' of cells being involved in almost all aspects of their physiology. Typically, their synthesis within cells occurs via a multi-step process of transcription and translation (i.e. initiation, elongation and termination).\(^1\) However, prior to their inclusion in an elongating protein, the intended constituent amino acids must first be attached to their cognate tRNA.\(^2\)-\(^4\) This process occurs via two half-reactions that are catalysed by the class of enzymes known as aminoacyl-tRNA synthetases (aaRS's).\(^2\)-\(^6\)

The first half-reaction is activation of the amino acid by reacting it with adenosine triphosphate (ATP) to give the corresponding aminoacyl-adenylate (aa-AMP) plus pyrophosphate (PP\(_i\)). This reaction has previously been studied both experimentally and computationally in detail.\(^6\)-\(^14\)

In the second-half reaction, the aminoacyl moiety is transferred from the aa-AMP onto its cognate tRNA to give the desired aminoacyl-tRNA product.\(^2\)-\(^6\) Remarkably, this reaction is catalysed by the same aaRS as for the first half-reaction. More specifically, the 2' (class I aaRS's) or 3'OH (most class II aaRS's) oxygen of the A76 residue of the tRNA substrate nucleophilically attacks the carbonyl carbon of the aa-AMP co-substrate.\(^4\),\(^7\),\(^15\),\(^16\) That is, it attacks the amino acid residue's carboxylic carbon.\(^2\),\(^3\) This ultimately results in cleavage of the phosphoester link in aa-AMP with formation of an A76-(2' or 3')O—C(0)R ester bond; i.e., transfer of the aminoacyl moiety onto the A76 residue of the cognate tRNA.\(^16\),\(^17\) It is generally held that aaRS's utilize a base catalysed mechanism in which an appropriate basic active site group deprotonates the nucleophilic A76 hydroxyl group.\(^17\)-\(^19\) However, the identity of the actual mechanistic base is often unclear.
Experimentally, X-ray crystal structures have been obtained for class I GlnRS by Perona et al.,\textsuperscript{17} and several class II aaRS's including HisRS by Guth et al.,\textsuperscript{16} and AspRS by Eiler et al.\textsuperscript{18-20} In each structure it was observed that a non-bridging phosphate oxygen of the aminoacyl-adenylate substrate was in close proximity to the sugar hydroxyl of the tRNA cosubstrate's A76 to which the aminoacyl moiety is to be transferred. In addition, Francklyn and coworkers\textsuperscript{16} performed a detailed experimental mutagenesis study on a non-metal containing HisRS. They observed that substitution of the \textit{pro}-R (O_{pro-R}) and \textit{pro}-S (O_{pro-S}) non-bridging aa-AMP phosphate oxygens by sulfur resulted in a 50- and 10000-fold decrease in the rate of reaction, respectively.\textsuperscript{16,21} Shortly thereafter, Perona et al. performed a mutational analysis on GlnRS and concluded that it followed the same mechanism.\textsuperscript{22} Thus, it was proposed that in the catalytic mechanism of HisRS O_{pro-S} acted as a general base and deprotonated the hydroxyl of the A76 residue of the tRNA cosubstrate as shown in \textbf{Scheme 7.1}.\textsuperscript{16} Furthermore, based on their observed dissimilar active sites and apparent common utilization of a non-bridging phospho-oxygen of the substrate as the mechanistic base, it has been suggested that for aaRS's such a substrate-assisted catalysis (SAC) mechanism may be a general feature of these presumed ancient enzymes.\textsuperscript{5,16,22-26}

\textbf{Scheme 7.1} General catalytic mechanism of aaRS's in which the substrates pro-S non-bridging oxygen acts as the mechanistic base.\textsuperscript{16}
Following these experimental studies, I performed a detailed density functional theory (DFT)-based computational investigation on the catalytic mechanism of HisRS. In particular, the ability of the aa-AMP’s bridging and pro-R and pro-S non-bridging oxygens to act as a base and the mechanism by which they may catalyse the aminoacylation reaction was systematically examined. It was found that the $O_{pro-S}$ oxygen could indeed act as the required base. Furthermore, in the resulting stepwise reaction pathway the rate limiting step was the initial proton transfer from the 3’OH group to $O_{pro-S}$ with a barrier of approximately 109.2 kJ mol$^{-1}$ (see Scheme 7.1). It was also noted that the tetrahedral intermediate was only stable with respect to conversion to the reactants or products by less than 5 kJ mol$^{-1}$. Thus, in vivo, the overall mechanism may effectively be concerted as had been experimentally suggested.$^{16,17}$

Recently, there have been several experimental studies on a metal containing ThrRS.$^{21,27-29}$ In contrast to that observed for HisRS,$^{16}$ mutagenesis studies by Francklyn and coworkers found that substitution of either non-bridging phosphate oxygens of the threonyl-adenylate substrate by sulfur resulted in only quite minor, three-fold or less, decreases in the rate of reaction.$^{21}$ However, substitution of a histidyl residue (His309),$^{21}$ thought to be positioned near the A76 2’OH group, decreased both $k_{cat}$ (aminoacylation) and $k_{trans}$ (aminoacyl transfer) by 34- and 242-fold, respectively. Consequently, a catalytic mechanism for ThrRS was proposed in which the His309 residue plays the role of the mechanistic base as shown in Scheme 7.2.$^{21,28}$ Specifically, it acts by deprotonating the 2’-OH group of the tRNA’s A76 residue. It was further suggested that this may occur directly or indirectly via a bridging water molecule between the 2’OH group and His309. The resulting 2’O$^-$ then deprotonates the adjacent 3’OH group enabling the now negatively charged 3’O
to act as a nucleophile and attack the carboxylic carbon of the threonyl-adenylate co-substrate, leading to transfer of the aminoacyl moiety.

Scheme 7.2 Proposed mechanism of ThrRS.\textsuperscript{21}

It is noted that the active site of ThrRS also contains a Zn(II) ion coordinated to the enzyme via two histidines (His385 and 511) and a cysteine (Cys334) residue.\textsuperscript{21,27,30} X-ray crystal structures\textsuperscript{27,30-33} have been obtained of ThrRS with and without various substrate analogues bound within the active site. Based on the observed differences, it has been suggested that the role of the Zn(II) may simply bind the substrate, possibly with displacement of a Zn(II)-bound water.\textsuperscript{30} Indeed, in all available X-ray crystal structures of ThrRS bound with threonine or a threonyl-AMP analogue such as those obtained by Torres-Larios et al.,\textsuperscript{33} Sankaranarayanan et al.,\textsuperscript{32} and Dock-Bregeon et al.,\textsuperscript{30} the threonyl or its analogue bidentately ligate via their R-group hydroxyl and neutral α-amino group to the Zn(II) ion. In fact, it has been suggested that this bidentate coordination of the threonyl to the Zn(II) is an essential characteristic of the active site of ThrRS that allows the enzyme to discriminate against the isosteric valine.\textsuperscript{33}

In this present study we have investigated the aminoacyl transfer mechanism as catalysed by threonyl-RNA synthetase using density functional theory-based methods in combination with large chemical models. For completeness and to provide additional insights into the roles of active site residues and functional group
moieties, we have examined the ability of the active site His309 residue to act as either a catalytic acid or base.

### 7.2. Computational Methods

All calculations were performed using the Gaussian 09 program\textsuperscript{34} and employed the hybrid density functional theory B3LYP method.\textsuperscript{35-37} Optimized geometries of all species detailed herein were obtained at the B3LYP/6-31G(d,p) level of theory as were their corresponding harmonic vibrational frequencies in order to confirm the nature of the stationary points, unless otherwise noted. In the "neutral His309" mechanism the pathway connectivity of all transition structures was confirmed by IRC calculations. The general effects of the surrounding polar protein environment were included via the integral equation formalism polarizable continuum model (IEF-PCM)\textsuperscript{38} with a dielectric constant (\( \varepsilon \)) of 4.0, a value commonly used for proteins.\textsuperscript{39} Relative energies were obtained by performing single point calculations at the IEF-PCM(\( \varepsilon=4.0 \))-B3LYP/6-311+G(2df,p) level of theory on the above optimized structures.
Scheme 7.3 Schematic illustration of the two ThrRS active site models used in this study in which His309 (a) is neutral and acts as a base and (b) is protonated and acts as an acid.

The two active site chemical models used in this present study were derived from our previous MD study\(^{40}\) on the fully bound active site. While the full details of this study are not repeated herein, those most pertinent to the present model selection are briefly summarized. In particular, the MD simulations examined the structure of the fully bound enzyme-ThrAMP/Thr-tRNA complex. These were based in part on the experimental\(^{27}\) X-ray crystal structure (PDB: 1QF6) in which both AMP and Thr-tRNA are bound within the active site while the threonyl moiety was added in accordance with several available X-ray crystal structures containing threonyl or similar groups.\(^{30,32,33}\) The resulting complex was then solvated and annealed for 500 ps followed by 10 ns production simulations in which the His309 residue and ThrAMP’s α-NH\(_2\) group was either neutral or protonated. It was concluded that the most likely initial states of the bound active site have a neutral α-NH\(_2\) group and either a neutral or protonated (His309-H\(^+\)) residue. A clustered based approach was
then used with the distance between the attacking \( A76 \) of the threonyl chosen as the clustering parameter. The average structure from the most populated cluster for each simulation was then retained, and then minimized using the AMBER99 MM method. The optimized structures were then used to derive the bound active site models used in the present hybrid DFT investigations and shown in Scheme 7.3. The key active site residues and substrates were modeled as follows: Thr-AMP by methyl monophosphate threonyl; A76 of the cognate tRNA by 1',5'-deoxyribose; Arg363 by ethylguanidinium cation; Glu484 acetamide; Glu383 by acetate; His385 and His511 by imidazole; and Cys334 by methylthiolate. His309 was modeled as either neutral or protonated 1-ethyl-imidazole. In addition, the Zn(II) cation and an active site water were also included. It is noted that the latter has been observed within the X-ray crystal structures.\(^{27}\) Selected atoms in the models were fixed at their final MM minimized positions in order to maintain integrity in the spatial arrangement of active site residues (Scheme 7.3). This computational approach has been successfully applied to related enzymes and has been previously reviewed in detail.\(^{39,41-44}\) It has been noted that this approach can commonly gives errors of approximately 12 kJ mol\(^{-1}\).\(^{39}\)

In both bound active site models employed herein the threonyl’s \( \alpha \)-amino group (Thr-NH\(_2\)) has been modeled as initially neutral (i.e., \( \alpha \)-NH\(_2\)). This is due to the fact that, as previously noted in the Introduction, in all available X-ray crystal structures\(^{30,32,33}\) in which a threonine or threonyl-AMP analogue is bound within the active site of ThrRS, the threonyl moiety at least initially coordinates via its R-group hydroxyl and \( \alpha \)-NH\(_2\) group to the Zn(II) centre. In addition, in most class II aaRS, it has been observed that the \( \alpha \)-amino group forms an ion pair with an active site glutamyl or aspartyl residue.\(^{16,20,45,46}\) Such an interaction is absent in ThrRS,\(^{21}\) however, while within the active site of ThrRS the Thr-NH\(_2\) group does lie near the
side chain of Tyr462, which is itself hydrogen bonded to the \( A_{76}2'0 \), Gln484, as well as those of several hydrophobic residues. Furthermore, in a previous QM/MM computational investigation on the first half-reaction as catalyzed by ThrRS, Zurek et al.\(^8\) observed that if the threonyl's \( \alpha \)-amino was protonated, the threonyl substrate rotated within the active site such that its \( \alpha \)-carboxylate instead ligated to the Zn(II) ion.\(^8\) It should also be noted that we did optimize the structure of the fully bound active site (using the method detailed above) in which the threonyl's \( \alpha \)-amino group was modeled as protonated (i.e., \( \alpha \)-NH\(_3^+\)). However, it was found that the \( \alpha \)-NH\(_3^+\) group shifted away from the Zn(II) ion and instead formed a strong hydrogen bond with \( A_{76}3'0 \).

In addition, a topological analysis of the electron densities at selected bond critical points was performed on the above optimized structures at the B3LYP/6-31G(d) level of theory using the AIM2000 program.\(^{47}\)

### 7.3. Results

#### 7.3.1. Substrate-bound active site

We began by considering the initial fully bound active sites in which the His309 residue is either protonated (His309–H\(^+\)) and thus may potentially act as an acid, or neutral and thus may act as a base. For both systems, two complexes lying close in energy were obtained: a 'pre-reactive' complex (PRC) in which the threonyl's \( \alpha \)-NH\(_2\) group is coordinated to the Zn(II) centre, which is consequently five-coordinate; and a 'reactive' complex in which the same \( \alpha \)-NH\(_2\) is no longer ligated to the Zn(II) centre, which is, as a result, four-coordinate. It is noted that unlike that observed in some aaRS's,\(^{16,48}\) in ThrRS there is no adjacent residue carboxylate with which the threonyl's \( \alpha \)-amino can interact.
Figure 7.1 Optimized structures with selected bond distances (Angstroms) of the fully bound active site pre-reactive and reactive complexes in which His309 is protonated (aPRC and aRC respectively) or neutral (bPRC and bRC respectively).

7.3.1.1. 'Pre-reactive' complexes (PRC's)

The optimized structures, with selected bond and interaction distances of the PRC's obtained in which His309 is protonated (aPRC) or neutral (bPRC) are shown.
in Figure 7.1. Both overall structures are in reasonable agreement with experimentally obtained X-ray crystal structures. For example, in both complexes one of the substrates non-bridging phosphate oxygens binds to the guanidinium of an arginyl (Arg363) in agreement with that observed\textsuperscript{27} experimentally. Furthermore, the threonyl moiety coordinates to the Zn(II) ion via both its R-group hydroxyl (Thr–OH) oxygen and α-NH\textsubscript{2} nitrogen (N\textsubscript{Thr}) centres. More specifically, the Thr–OH oxygen forms a short strong interaction with the Zn(II) centre with Zn...O(H)\textsubscript{Thr} distances of 1.98 and 1.99 Å in \textsuperscript{a}PRC and \textsuperscript{b}PRC, respectively. This is likely due in part to the fact that it is also simultaneously hydrogen bonded to the carboxylate of an adjacent aspartate (Asp383), thus enhancing the nucleophilicity of the Thr–OH oxygen (Figure 7.1). The Zn...N\textsubscript{Thr} coordination bonds in \textsuperscript{a}PRC and \textsuperscript{b}PRC are both slightly longer at 2.24 and 2.28 Å, respectively. Thus, as noted above, the Zn(II) centre in the PRC’s is five-coordinate as it is also ligated to the enzyme via two histidines (His511 and His385) and a cysteine (Cys334). Both optimized structures are in agreement with X-ray crystal structure of, for example, a bound ThrRS...SerAMP complex, which has a similar but slightly longer Zn...N\textsubscript{Thr} distance of 2.33 Å.

7.3.1.2. 'Reactive' complexes (RC's)

Notably, alternate substrate-bound active site complexes were also obtained in which the threonyl moiety is \textit{only} coordinated to the Zn(II) centre via its R-group oxygen; that is the Zn...N\textsubscript{Thr} coordination bond has been cleaved and the Zn(II) centre is now, as a result, four-coordinate. The optimized structures with selected bond lengths of the corresponding resultant reactive complexes, \textsuperscript{a}RC and \textsuperscript{b}RC respectively, are shown in Figure 7.1. Importantly, these complexes lie only
marginally higher in energy than their corresponding pre-reactive complexes \(^*\text{PRC}\) and \(^b\text{PRC}\) by 13.6 and 7.5 kJ mol\(^{-1}\), respectively. Thus, similar to that experimentally observed for other Zn(II)-containing complexes,\(^{49-52}\) the Zn···N\(_{\text{Thr}}\) coordination bond appears to be quite labile with the Zn(II) center able to easily interconvert between five and four coordinate with little cost in energy (see below).

In both \(^*\text{RC}\) and \(^b\text{RC}\) the distance between the Zn(II) ion and the threonyl’s amino nitrogen (N\(_{\text{Thr}}\)) centre has lengthened markedly with \(r(\text{Zn···N}_{\text{Thr}})\) distances of 3.07 and 3.43 Å, respectively. In fact, in both cases the N\(_{\text{Thr}}\) centre now forms a moderately strong hydrogen bond (H-bond) with the A\(_{76}\)3’OH of the tRNA cosubstrate with \(\text{Thr} \text{N···H}3’\) distances of 1.91 and 1.94 Å respectively. The properties of the density at the bond critical point (BCP) of a particular interaction can give additional insights. For hydrogen bonds (H-bonds), the typical values of the electron density (\(\rho\)) and its Laplacian (\(\nabla^2 \rho\)) are said to be 0.002-0.040 and 0.024-0.139 respectively.\(^{53,54}\) As can be seen from the results of the topological analyses of \(^*\text{RC}\) and \(^b\text{RC}\) given in Table 7.1, these two \(\text{Thr} \text{N···H}3’\) interactions are calculated to have very similar \(\rho\) values of 0.034 and 0.033 that are towards the upper end of the range typically noted for such bonds (i.e., are relatively strong H-bonds).\(^{53,54}\) In addition, in \(^*\text{RC}\) and \(^b\text{RC}\) the crystallographically observed active site H\(_2\)O now hydrogen bonds to the threonyl’s carbonyl oxygen (C\(_{\text{carb}}\)=O) with C\(_{\text{carb}}\)O···HOH distances of 1.92 and 1.83 Å, respectively. The comparative strength of these two interactions is reflected in their BCP \(\rho\) and \(\nabla^2 \rho\) values of 0.026 and 0.073, and 0.032 and 0.097 respectively, i.e., the C\(_{\text{carb}}\)O···HOH H-bond in \(^b\text{RC}\) is stronger.

As a result of the dissociation of the Zn···N\(_{\text{Thr}}\) coordination bond, the threonyl’s R-group hydroxyl oxygen is now more strongly ligated to the Zn(II) ion than in the corresponding PRC’s. Hence, the Zn···O(H)\(_{\text{Thr}}\) distances are now markedly shorter at 1.92 and 1.91 Å in \(^*\text{RC}\) and \(^b\text{RC}\) respectively (Figure 7.1). This shortening
is likely assisted by Asp383 which is still hydrogen bonded via its carboxylate to the Thr-OH group.

In addition, in both $^a$RC and $^b$RC there is a hydrogen bonding bridge network between the imidazole of His309 and the threonyl’s carbonyl oxygen (Figure 7.1). However, due to their different protonation states of the His309 residue, they exhibit several notable differences. Specifically, in $^a$RC the protonated imidazole of His309 acts as a hydrogen bond donor to the 2’O of the Thr-tRNA moiety. Due to the fact that this interaction involves a charged moiety, the resulting $^{His309NH\cdots O2’}_A^{76}$ interaction is quite short and strong with a length of just 1.53 Å and a BCP $\rho$ value (0.070; Table 7.1), that is well above the range of typical H-bonds (see above).$^{53,54}$ The $^{A76}2’OH$ group itself then acts as a H-bond donor to the active site H$_2$O with an $r(A762’OH\cdots OH_2)$ length of 1.66 Å. Again this is also a quite short and strong hydrogen bond as is also reflected in the relatively high value of $\rho$ (0.051) at its BCP. As noted above, the H$_2$O moiety also forms a moderately strong interaction with the threonyl’s $C_{\text{carb}}=O$ oxygen.
Table 7.1 Values of the Electron Density ($\rho$) and its Laplacian ($\nabla^2 \rho$) (a.u.) at BCPs of Selected Hydrogen Bonds Within the Fully Bound Reactive Complexes $^{a}$RC and $^{b}$RC. $^{a}$

<table>
<thead>
<tr>
<th>BCP</th>
<th>$\rho$</th>
<th>$\nabla^2 \rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{a}$RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{a}$B1</td>
<td>$A_{76}^{2'}O\cdots H N_{\text{His309}}$</td>
<td>0.070</td>
</tr>
<tr>
<td>$^{a}$B2</td>
<td>$A_{76}^{3'}O H \cdots N_{\text{Thr}}$</td>
<td>0.034</td>
</tr>
<tr>
<td>$^{a}$B3</td>
<td>$A_{76}^{2'}O H \cdots O H_2$</td>
<td>0.051</td>
</tr>
<tr>
<td>$^{a}$B4</td>
<td>OH$<em>2\cdots O=C</em>{\text{carb}}$</td>
<td>0.026</td>
</tr>
<tr>
<td>$^{b}$RC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{b}$B1</td>
<td>$A_{76}^{2'}O H \cdots N_{\text{His309}}$</td>
<td>0.031</td>
</tr>
<tr>
<td>$^{b}$B2</td>
<td>$A_{76}^{3'}O H \cdots N_{\text{Thr}}$</td>
<td>0.033</td>
</tr>
<tr>
<td>$^{b}$B3</td>
<td>OH$<em>2\cdots A</em>{76}^{2'}O$</td>
<td>0.028</td>
</tr>
<tr>
<td>$^{b}$B4</td>
<td>OH$<em>2\cdots O=C</em>{\text{carb}}$</td>
<td>0.032</td>
</tr>
</tbody>
</table>

$^{a}$ Small red points indicate BCP’s. Color key: C (gray); O (red); N (blue); P (maroon); S (yellow); H (grey) and Zn (white).
In contrast, in \textsuperscript{b}RC the now unprotonated imidazole N centre of His309 (N_{His309}) acts as a hydrogen bond acceptor towards the \textit{A76}2’\textit{OH} group (Figure 7.1). It is noted that this is now a neutral interaction, the resulting \text{His309N}••\text{HO2}’_{A76} distance (1.93 Å) is 0.40 Å longer than the corresponding interaction in \textsuperscript{a}RC. Its now more typical H-bond length is also illustrated by the fact that at its BCP (\textsuperscript{b}B1) the values of \( \rho \) (0.031) and \( \nabla^2 \rho \) (0.074) now lie within the standard range of such interactions (Table 7.1). As a consequence, the 2’-oxygen now forms a longer though still moderately strong hydrogen bond to the active site H\textsubscript{2}O with a 2’O••H\textsubscript{2}O length of 1.90 Å (\( \rho=0.028 \); Table 7.1). As noted above, the H\textsubscript{2}O in \textsuperscript{b}RC is also simultaneously hydrogen bonded to the threonyl’s C\textsubscript{carb}=O oxygen. In addition, however, it also forms a weak hydrogen bond with the threonyl’s \( \alpha \)-NH\textsubscript{2} group; \( r(\text{H\textsubscript{2}O}••\text{HN}_{\text{Thr}}) = 2.44 \) Å. This is also reflected in its \( \rho \) (0.011) and \( \nabla^2 \rho \) (0.045) values which lie towards the lower end of the ranges usually observed for hydrogen bonds.\textsuperscript{53,54} Thus, in \textsuperscript{b}RC the His309 residue is involved in a hydrogen bond bridge with both the threonyl’s C\textsubscript{carb}=O and \( \alpha \)-NH\textsubscript{2} groups via \textit{A76}2’\textit{OH} and the active site H\textsubscript{2}O (Figure 7.1).

It is noted that we also considered larger active site models that also included Tyr462, thought to also interact with \textit{A76}2’\textit{OH}, and Gln484 which was found to hydrogen bond via its R-group oxygen with the threonyl’s \( \alpha \)-NH\textsubscript{2} (not shown). These models were found to give overall structures in close agreement with that of \textsuperscript{a}RC and \textsuperscript{b}RC above.

In our previous MD study\textsuperscript{40} on the fully bound active site of ThrRS in which His309 was either neutral or protonated, the structures of both resulting complexes were observed to be in reasonable agreement with experiment. Indeed, as detailed above, the major difference is the hydrogen bond network centered on the \textit{A76}2’\textit{OH} group. In addition, at the level of theory used in the present study the R-group of
His309 is calculated to have similar proton affinities whether it is within the active site (1195.0 kJ mol\(^{-1}\)) or in aqueous solution (1192.9 kJ mol\(^{-1}\)). Thus, for completeness and to obtain greater insights we have considered possible pathways in which His309 is protonated (Section 7.3.2) or neutral (Section 7.3.3).

### 7.3.2. The protonated His309-assisted mechanism

The potential energy surface (PES) obtained for the catalytic mechanism of ThrRS when His309 is protonated is shown schematically in Figure 7.2. The optimized structure of each intermediate and transition structure (TS) along the pathway is provided in Figure 7.3. As can be seen, for the chemical model used herein the overall mechanism can be thought of as occurring in three stages: (i) formation and (ii) rearrangement of a tetrahedral intermediate complex and (iii) formation of the product complex.

![Figure 7.2](image.png) **Figure 7.2** Calculated PES for the protonated His309-assisted mechanism of ThrRS.

#### 7.3.2.1. Formation of the tetrahedral intermediate complex

This initial 'stage' of the overall mechanism occurs in two steps. First, the His309-H\(^+\) residue essentially transfers its proton via the A762’OH and active site H\(_2\)O
onto the threonyl’s C=O oxygen. This process occurs via transition structure $^a$TS1 (obtained via detailed scans) with a moderately low barrier of 71.6 kJ mol$^{-1}$ to give the intermediate complex $^a$IC1 lying only 0.2 kJ mol$^{-1}$ lower in energy at 71.4 kJ mol$^{-1}$. These almost equivalent energies are reflected in their optimized structures with the largest difference in the bond lengths for $^a$TS1 and $^a$IC1 shown in Figure 7.3 being just 0.02 Å.

In $^a$IC1 the His309 residue is neutral and forms a strong hydrogen bond chain to the threonyl’s now protonated C=OH$^+$ group via the $^{A76}2’$OH and H$_2$O moieties. This protonation causes a slight increase in the positive charge on threonyl’s C$_{\text{carb}}$ centre from 0.63 ($^a$RC) to 0.68, thus enhancing its electrophilicity. It is noted that while the C=OH$^+$ bond has lengthened from 1.23 ($^a$RC) to 1.27 Å, the C$_{\text{carb}}$—OP bond has shortened by almost the same amount from 1.32 ($^a$RC) to 1.28 Å due to the now greater positive charge on C$_{\text{carb}}$. 

![Diagram of aTS1 and aIC1](image-url)
Chapter 7. A DFT Study on Thr-tRNA synthetase

![Images of molecular structures with distances labeled (e.g., Zn 1.70 Å, A76 1.84 Å, His309 2.02 Å).]
Figure 7.3 Optimized structures with selected bond lengths (Ångstroms) of the intermediates (IC), transition structures (TS) and product complex (PC) of the protonated His309-assisted mechanism of ThrRS. For clarity, only some residues are shown in ball and stick format. Color key: C (gray); O (red); N (blue); P (orange); S (yellow), Zn (lavender) and H (white).

Indeed, this step is followed by nucleophilic attack of the $^{3'}$O at the threonyl's $C_{\text{carb}}$ centre. This proceeds via $^a$TS2 at a cost of 70.0 kJ mol$^{-1}$ with respect to $^a$IC1. More importantly, however, the overall barrier of this step is 141.4 kJ mol$^{-1}$ with respect to the initial bound active site complex $^a$RC; 155.0 kJ mol$^{-1}$ with respect to the pre-reactive complex $^a$PRC! As seen in Figure 7.2, this is the rate-limiting step in the overall protonated His309-assisted mechanism. Importantly, however, it is significantly higher than that generally considered to be the upper thermodynamic limit for an enzyme catalysed reaction: $\sim$125 kJ mol$^{-1}$.$^{39}$ Furthermore, the formation of the threonyl-tRNA $C_{\text{carb}}$—$O3'$ bond occurs with concomitant transfer (Figure 7.3: $^a$TS2) of the $^{3'}$OH proton to the threonyl's $\alpha$-NH$_2$ group. The resulting tetrahedral intermediate complex $^a$IC2 lies 86.0 kJ mol$^{-1}$ higher in energy than $^a$RC.
In \textbf{aIC2} the $C_{\text{carb}}$—O3'$_{A76}$ crosslink formed has a length (1.43 Å) that is typical for C—O single bonds while both the $C_{\text{carb}}$—OH and $C_{\text{carb}}$—OP bonds have concomitantly lengthened significantly by 0.12 and 0.10 Å to 1.39 and 1.38 Å, respectively (Appendix \textbf{Table E.1}). Moreover, the threonyl's now $\alpha$-NH$_3^+$ group hydrogen bonds via the H$_2$O to the $A_{76}$2'OH, which itself still interacts with the His309 residue (\textbf{Figure 7.3}).

As a note here, we did consider the direct attack of the 3'O to $C_{\text{carb}}$ prior to the proton transfer to the carbonyl oxygen in \textbf{aRC}. The barrier for this process was found to be 174.0 kJ mol$^{-1}$ with respect to \textbf{aRC} (not shown). Thus while the initial proton transfer from \textsubscript{His309}NH$^+$ has lowered the barrier by $\sim$30kJ mol$^{-1}$, it remains too high to be enzymatically feasible.

\textbf{7.3.2.2. Rearrangement of the tetrahedral intermediate complex}

Prior to cleavage of the PO—$C_{\text{carb}}$ bond (completing the transfer of the threonyl moiety), within the present computational model, the tetrahedral intermediate must undergo a two-step rearrangement. This rearrangement guarantees the regeneration of His309-H$^+$ within the mechanism. It begins with rotation about the $C_{\text{carb}}$—OH bond that no longer hydrogen bonds with the H$_2$O but instead now interacts with a non-bridging phosphate oxygen (\textbf{Figure 7.2}). This step proceeds via \textbf{aTS3} (obtained via detailed scans) at a low cost of approximately 8.6 kJ mol$^{-1}$. The alternate tetrahedral complex formed (\textbf{aIC3}) lies markedly lower in energy than \textbf{aIC2} by 40.6 kJ mol$^{-1}$; just 45.4 kJ mol$^{-1}$ higher than \textbf{aRC}. The formation of a short $C_{\text{carb}}$OH—OP hydrogen bond (1.66 Å) results in minor changes to the $C_{\text{carb}}$—OP and $C_{\text{carb}}$—OH bonds of +0.02 to 1.40 Å and $\sim$0.01 to 1.38 Å, respectively.
In addition, however, the threonyl’s \( \alpha \text{-NH}_3^+ \) group now directly interacts with the \( A_76' \text{OH} \) group; \( r(\text{NH} \cdots \text{O}2'_{A76})=1.93 \text{ Å} \) (Figure 7.3).

Subsequently, the \( \alpha \text{-NH}_3^+ \) moiety transfers a proton via \( A_76' \text{OH} \) to the imidazole of His309, thus regenerating His309-H\( ^+ \), i.e., protonated His309. This step proceeds via \( ^a\text{TS4} \) at a cost of 48.8 kJ mol\(^{-1} \) with respect to \( ^a\text{IC3} \) to give a higher energy tetrahedral intermediate \( ^a\text{IC4} \) having a relative energy of 71.0 kJ mol\(^{-1} \) with respect to \( ^a\text{RC} \). In \( ^a\text{IC4} \) the \( C_{\text{carb}}\text{OH} \cdots \text{OP} \) hydrogen bond has shortened slightly to 1.64 Å, resulting in a minor increase (0.01 Å) in the \( C_{\text{carb}} \cdots \text{OP} \) bond to 1.41 Å, while the His309 residue is now once again protonated (Figure 7.3).

### 7.3.2.3. Formation of the product complex

The last step in the aminoacyl transfer process is cleavage of the \( C_{\text{carb}} \cdots \text{OP} \) bond to give the final corresponding product complex (\( ^a\text{PC} \)). This step occurs via the six-membered ring transition structure \( ^a\text{TSS5} \) at a quite low cost of only 12.2 kJ mol\(^{-1} \) with respect to \( ^a\text{IC4} \). The final product-bound active site complex \( ^a\text{PC} \) lies slightly higher in energy than the initial substrate-bound active site complex \( ^a\text{RC} \) by 7.0 kJ mol\(^{-1} \). In \( ^a\text{PC} \) the cleaved \( C_{\text{carb}} \cdots \text{OP} \) distance is now 3.29 Å while both the \( C_{\text{carb}}=0 \) and \( C_{\text{carb}}-O3'_{A76} \) bonds have both shortened markedly to 1.22 and 1.34 Å, respectively.

Thus, with His309 initially protonated, the preferred pathway occurs via an acid catalysed esterification.\(^{55} \) His309-H\( ^+ \) acts as the initial acid protonating the substrates carbonyl oxygen while the \( A_{76}3' \text{OH} \) loses its proton via donation to the substrates \( \alpha \text{-NH}_2 \) group. However, the barrier for the rate-limiting step (141.4 kJ mol\(^{-1} \)) for the formation of the tetrahedral intermediate \( ^a\text{IC2} \), is above the generally held thermodynamic ‘upper-limit’ for \textit{in vivo} enzymatic processes (\( \sim 125 \text{ kJ mol}^{-1} \)).\(^{39} \)
7.3.3. The neutral His309-assisted mechanism

The PES obtained for the catalytic mechanism of ThrRS with a neutral His309 is shown schematically in Figure 7.4 while the optimized structure of each intermediate and TS involved, with selected bond lengths is given in Figure 7.5.

![Figure 7.4](image)

**Figure 7.4** Calculated PES for the neutral His309-assisted mechanism of ThrRS.

### 7.3.3.1. Formation of a tetrahedral intermediate complex

In contrast to that observed in the protonated His309-assisted mechanism, the tetrahedral intermediate is formed from the reactive complex \( ^b\text{RC} \) in a single step. This occurs via \( ^b\text{TS1} \) at a cost of 106.8 kJ mol\(^{-1} \) or 114.3 kJ mol\(^{-1} \) with respect to the pre-reactive complex \( ^b\text{PRC} \). It is noted that this barrier is close to that previously obtained for HisRS where the analogous nucleophilic attack of 3’OH to give a tetrahedral intermediate was calculated to be approximately 109.2 kJ mol\(^{-1} \). More importantly, in contrast to that observed in the protonated His309 assisted mechanism, this barrier is lower than that generally held to be the upper thermodynamic enzymatic limit. It should be noted that as observed in the analogous TS of the ‘protonated His309’ pathway, \( ^a\text{TS2} \) (cf. Figure 7.3), in \( ^b\text{TS1} \) the
A763’OH oxygen nucleophilically attacks at the substrates’ C_carb centre with concomitant transfer of the 3’OH proton to the threonyl substrates α-NH₂ group. Now, however, the 3’OH proton is almost fully transferred onto the Thr-NH₂ group as indicated by the 3’O⋯H and 3’OH⋯NThr distances of 1.61 and 1.08 Å respectively (Figure 7.5), while the forming C_carb⋯O3’ bond (1.96 Å) is now much longer (cf. Figure 7.3).

The resulting oxyanion tetrahedral intermediate **bIC1** lies 85.6 kJ mol⁻¹ higher in energy relative to **bRC**. As can be seen in Figure 7.5 the C_carb—O3’ A76 bond has now been formed with a length of 1.51 Å while the former C_carb=O and C_carb—OP bonds have significantly lengthened to 1.30 and 1.43 Å respectively (Appendix Table E.1), and the threonyl’s amino group is now protonated (i.e., α-NH₃⁺). The C_carb—O⁻ oxyanion is stabilized in part via a strong hydrogen bond (1.47 Å) with the H₂O which is itself hydrogen bonded to both the α-NH₃⁺ (1.87 Å) and A762’OH (2.40 Å) groups. It is noted that the α-NH₃⁺ also forms weak hydrogen bonds with both the A763’O and A762’O with distances of 2.17 and 2.08 Å, respectively.

We did also examine stepwise and concerted pathways in which His309 may directly act as a base by accepting a proton from A762’OH as previously proposed. However, within the present computational model as well as a larger model (not shown) that also included the residues Tyr462 and Gln484, no stable intermediate in which threonyl’s amino group remains neutral while His309 is fully or partially protonated was obtained. That is, **bIC1**-type intermediates were the only stable oxyanionic intermediate obtained.
Chapter 7. A DFT Study on Thr-tRNA synthetase
Chapter 7. A DFT Study on Thr-tRNA synthetase

![Figure 7.5 Optimized structures with selected bond lengths (Angstroms) of the intermediates (IC), transition structures (TS) and product complex (PC) of the neutral His309-assisted mechanism of ThrRS. For clarity, only key residues are shown in ball and stick format. Color key: C (gray); O (red); N (blue); P (orange); S (yellow), Zn (lavender) and H (white).]

7.3.3.2. Formation of the 'neutral' product complex

Similar to that observed in the 'protonated His309'-assisted mechanism, within the present computational model, the tetrahedral intermediate bIC1 must then undergo a two-step rearrangement prior to formation of the final product complex. The first step occurs with a proton transfer from the α-NH$_3^+$ moiety via the H$_2$O to the oxyanion C$_{carb}$—O$^-$ centre. This process is essentially barrierless (bTS2) to give the alternate tetrahedral intermediate bIC2 lying 3.5 kJ mol$^{-1}$ lower in energy than bIC1 (Figure 7.4). In bIC2 the now 'neutralized' C$_{carb}$—OH bond has lengthened markedly by 0.08 Å to 1.38 Å while both the C$_{carb}$—OP and C$_{carb}$—O$^-$A76 bonds have shortened to 1.39 and 1.44 Å, respectively (Appendix Table E.1). It is noted that in
Chapter 7. A DFT Study on Thr-tRNA synthetase

bIC2 the active site H₂O forms a hydrogen bond bridge between the α-NH₂ (1.77 Å), A76'OH (1.96 Å) and C_carb OH (1.83 Å) as shown in Figure 7.5.

The second step, again is similar to that observed in the 'protonated His309'-assisted mechanism; rotation about the C_carb—OH bond. However, this rotation proceeds via bTS3 with an approximately 16 kJ mol⁻¹ lower barrier of only 32.5 kJ mol⁻¹; 114.6 kJ mol⁻¹ with respect to bRC (Figure 7.4). Furthermore, the resulting alternate tetrahedral conformer bIC3 formed lies just 5.2 kJ mol⁻¹ higher in energy than bIC2 (cf. Figure 7.2). In bIC3 the C_carb OH group now forms a moderately short hydrogen bond (1.77 Å) to a non-bridging phosphate oxygen (Figure 7.5). Importantly, this distance is longer than observed in aIC4 (cf. Figure 7.3) causing a slight lengthening in the C_carb—OP bond to 1.43 Å (Appendix Table E.1).

With the rotation of the hydroxyl group, as observed in the 'protonated His309'-assisted mechanism, the final step is then the concomitant transfer of the proton from the C_carb—OH to the non-bridging phosphate oxygen and cleavage of the PO—C bond. This occurs via the six-membered ring transition structure bTS4 at a cost of only 27.0 kJ mol⁻¹, 114.3 kJ mol⁻¹ with respect to the initial substrate-bound active site complex bRC. In contrast, however, the resulting product-bound active site complex bPC is calculated to lie lower in energy than the initial substrate-bound active site complex bRC by 24.4 kJ mol⁻¹ (Figure 7.5). In bPC the C_carb···OP distance is now 5.73 Å while the C_carb—O3'₆₇₆ and C_carb=O bonds have shortened significantly to 1.35 and 1.22 Å, respectively. The threonyl has been transferred onto the tRNA₆₇₆ at its 3'-oxygen.

In the above 'neutral His309' assisted mechanism, the highest overall mechanism barriers are the rotation about the C_carb—OH bond and proton transfer from C_carb—OH to a non-bridging phosphate oxygen. With regards to bRC, as noted above, these barriers are 114.6 and 114.3 kJ mol⁻¹ while with regards to the pre-
reactive complex \textsuperscript{b}PRC these barriers are 122.1 and 121.8 \textit{kJ mol\textsuperscript{-1}} respectively. These barriers are still enzymatically feasible, though they are near the generally-held upper enzyme-catalysed thermodynamic limit.\textsuperscript{39} However, a possible alternate pathway is direct cleavage of the \textit{C}_{\text{carb}}—OP bond in the tetrahedral intermediate \textbf{IC1} to give a product complex consisting of bound AMP (modeled as methylphosphate) and aminoacylated\textsuperscript{\textit{tRNA}A76} (modeled by dexoxyribose threonyl); i.e., a two-step process. However, in order to model this route, a larger chemical model was required that also included \textit{Lys465} (modeled as an unconstrained protonated methylamine) in order to help stabilize the charge on the phosphate.

![Relative Energy (kJ mol\textsuperscript{-1})](image)

**Figure 7.6** Calculated PES for the neutral His309-assisted mechanism of ThrRS in which the lysyl residue (\textit{Lys465}) is included in the chemical model (see text).

The resulting PES obtained is shown in **Figure 7.6**. It should be noted that the optimized structures of all stationary points along the pathway were very similar to
their corresponding counterparts in Figures 7.4 and 7.5, the main differences being relatively minor changes in the non-bridging P—O bond lengths and that in the product complex where the threonyl amino group remains protonated. As can be seen, inclusion of the lysyl has little affect on the barrier for formation of the tetrahedral intermediate $bIC1'$ which proceeds via $bTS1'$ with an almost negligibly lower barrier of 105.1 kJ mol$^{-1}$ (cf. Figure 7.6). Indeed, in mutagenesis studies a Lys465Ala (K465A) substitution had only minor effects on the reaction kinetics.$^{21}$ In this step the 3’OH proton is again transferred directly to the Thr-NH$_2$ group. Furthermore, the product complex $bPC'$ still lies markedly lower in energy than the initial reactive complex $bRC'$ by 15.8 kJ mol$^{-1}$. Importantly, however, the tetrahedral complex $bIC1'$ lies 52.8 kJ mol$^{-1}$ higher in energy than $bRC'$ and cleavage of the C$_{carb}$—OP bond, formal transfer of the aminoacyl moiety, now occurs directly via $bTS2'$ without a barrier to give the product $bPC'$. Thus, the rate-limiting step is formation of the tetrahedral intermediate as observed in related computational studies.$^{5}$

7.4. Conclusions

In this current investigation, density functional theory-based methods in combination with large chemical models have been applied to the second half-reaction catalyzed by Thr-tRNA synthetase. Specifically, transfer of the threonyl aminoacyl moiety from Thr-AMP onto the ribose 3’O of the A76 residue of the cognate Thr$_2$RNA. Furthermore, we have considered both cases in which an active site histidyl residue (His309) is either protonated or neutral.
Chapter 7. A DFT Study on Thr-tRNA synthetase

For each resulting initial substrate bound-active site complex two alternate binding modes for the threonyl moiety were obtained. In the lowest energy complexes (\textsuperscript{a}PRC and \textsuperscript{b}PRC), the threonyl moiety was bidentate-coordinated to the Zn(II) ion via its R-group hydroxyl oxygen and \(\alpha\)-amino nitrogen \((N_{\text{Thr}})\) centre. In the corresponding alternate reactive complexes, \textsuperscript{a}RC and \textsuperscript{b}RC the Zn(II)--\(N_{\text{Thr}}\) bond is cleaved and the threonyl is now only monodentately ligated to the Zn(II) centre via its R-group hydroxyl oxygen. Furthermore, they lie only slightly higher in energy than \textsuperscript{a}PRC and \textsuperscript{b}PRC by 13.6 and 7.5 \(\text{kJ mol}^{-1}\) respectively. Thus, the Zn(II)--\(N_{\text{Thr}}\) bond appears to be reasonably labile, particularly in the neutral His309 system.

In the aminoacyl transfer pathway obtained for when the His309 is initially protonated, the rate-limiting step is formation of the tetrahedral intermediate. That is, the step in which the 3’OH oxygen nucleophilically attacks the threonyl’s carbonyl carbon centre. However, the barrier is 141.4 (155.0) \(\text{kJ mol}^{-1}\) with respect to \textsuperscript{a}RC (\textsuperscript{a}PRC). Consequently, it is not enzymatically feasible. Notably, in this step, the 3’OH proton is not transferred to the His309 residue, but instead to the Thr-NH\(_2\) group.

Two models were considered in which the His309 residue was initially neutral, the difference being whether Lys465 was included, simply in order to counterbalance the negative charge on the phosphate in the model. For both models the rate-limiting step was found to be below the enzymatic limit of 125 \(\text{kJ mol}^{-1}\). In particular, for the model containing Lys465 the rate limiting step was formation of the tetrahedral intermediate. The barrier being approximately 105 \(\text{kJ mol}^{-1}\) with respect to \textsuperscript{b}RC. For the model which did not include Lys465, the rate limiting step was rotation about the \(C_\text{carb}--\text{OH}\) bond. The barrier being approximately 114.6 \(\text{kJ mol}^{-1}\) with respect to \textsuperscript{b}RC; 122.1 \(\text{kJ mol}^{-1}\) with respect to \textsuperscript{b}PRC. However, while the rate limiting steps differed for the two models, the barriers for formation of the tetrahedral intermediate were very similar and differed by only 1.7 \(\text{kJ mol}^{-1}\). More
importantly, however, is that for both models, the 3’OH proton is not transferred to the His309 residue. Rather, during formation of the tetrahedral intermediate it directly transfers from the 3’OH onto the threonyl moieties α-NH$_2$ group.

Thus, the current results suggest that as observed for other aminoacyl-tRNA synthetases,$^{16,17,20}$ ThrRS utilizes a substrate-assisted catalytic mechanism in the aminoacyl transfer for ThrRS. However, it is not the threonyl-AMP substrates phosphate group that acts as the base, but rather its α-NH$_2$ group. The presence of the Zn(II) ion and lability of the Zn(II)···N$_{Thr}$ bond appears to allow the α-NH$_2$ to act as the mechanistic base.
References

(1) Zaher, H. S.; Green, R. Cell 2009, 136, 746.
Chapter 7. A DFT Study on Thr-tRNA synthetase


Chapter 8

Conclusions
This work has primarily focused on the properties and mechanisms of enzymes involved in carbohydrate-related metabolism. Furthermore, a very broad range of computational and theoretical methods have been applied in order to elucidate their mechanisms and to gain invaluable insights into the catalytic approaches they utilize in order to achieve their remarkable rate-enhancing power. A number of key points from each study are summarized in this chapter.

Chapter 3 describes the synthesis of uridine 5’-diphosphate glucuronic acid (UDPGlcUA) from UDP-glucose (UDPGlc) by uridine diphosphate glucose dehydrogenase (UDPGlcDH) via a two-fold two-electron-one-proton oxidation. This process successively transforms the 6-hydroxymethyl of glucopyranose into a formyl group and into the final carboxylic acid functional group. The computational method used in this project was a DFT cluster approach with IEF-PCM SCF. The overall mechanism consists of four steps: (i) NAD⁺-dependent oxidation of glucose to glucuronaldehyde, (ii) nucleophilic addition of Cys260–SH to glucuronaldehyde to form a 6-thiohemiacetal intermediate, (iii) NAD⁺-dependent oxidation of the 6-thiohemiacetal to form a 6-thioester intermediate, and finally, (iv) hydrolysis of the 6-thioester to give glucuronic acid. In addition, this study provides insight into the debated roles of Lys204 and Asp264, and the most likely protonation state of a reactive Michaelis complex of UDPGlcDH. The two-step redox reactions in this enzyme are dependent on two NAD⁺ coenzymes with the redox ability and the transient association, which were lack from regular enzyme residues. The highest energy barrier in the mechanism is the hydride transfer processes from the substrate to the NAD⁺, demonstrating the driving force in this enzyme mechanism is based on two coenzymes.

Chapter 4 describes the glycosidic bond cleavage reaction catalyzed by the enzyme GlvA, a member of the glycoside hydrolases superfamily. The PESs in this
enzyme mechanism were obtained with DFT-cluster approach. The mechanism follows a non-redox catalytic pathway or a net redox-neutral catalytic pathway using NAD$^+$ and divalent metal ion as cofactor. This is a regioselective redox-neutral mechanism of glycosidic bond hydrolysis that favours α- over β-glycosides. Its proposed catalytic mechanism can be divided into two half-reactions. The first one activates the glucopyranose ring by successively forming intermediates that are oxidized at the 3-, 2- and 1-positions of the ring, which ultimately facilitate the heterolytic deglycosylation. The second half-reaction is essentially the reverse of the first half-reaction, beginning with the pyranose ring hydroxylation at the anomeric carbon. It is followed by 3-reduction and regeneration of the active forms of the catalytic site and its cofactors. Both DFT and NBO analyses were applied to this enzymatic mechanistic study. Specifically, NBO analyses were used to monitor the electron flow and change in oxidation state on each atomic centre along reaction coordinates in order to rationalize the energetics and regioselectivity of this catalytic mechanism. Although both NAD$^+$ and Fe(II) ion cofactors have important catalytic functions in the mechanism, it has been found that the redox hydrolytic catalytic mechanism is driven by the gradual strengthening of the axial endo-anomeric component within the hexose ring along the reaction coordinates, in order to facilitate the heterolytic dissociation of the axial C1–O bond. In addition, the combined influence of specific components of the generalized anomeric effect fully explains the regioselectivity observed in the catalytic activity of GlvA.

Substrate binding and possible pathways leading to formation of a 2-keto intermediate (2-keto-SRH), experimentally observed to be catalytically viable and for which an X-ray crystal structure has been obtained of it bound within the active site of a mutated LuxS (PDB ID: 1YCL), were examined for the enzyme S-ribosylhomocysteinease (LuxS) in Chapter 5. The combination of Docking, MD, and
QM/MM methods were used to investigate its catalytic process. In particular it was shown that LuxS preferably binds the α-furanose (α-RC) and β-furanose (β-RC) containing S-ribosylhomocysteine substrate with the former preferred by 47.5 kJ mol⁻¹. The binding of the alternate linear aldose substrate was examined, but the two possible bound configurations were both found to lie significantly higher in energy than α-RC by at least 50.4 kJ mol⁻¹. Importantly, an active site water is found to play an integral role in the mechanism. Specifically, the ribosyl moiety of the SRH binds to the active site Fe(II) via both its -O2H and -O3H groups. The former is then able to transfer its proton via the H₂O onto the ribosyl ring O4 oxygen. This results in ring opening and occurs with concomitant transfer of the ribosyl -C2H- proton onto the thiolate of Cys84. Remarkably, because of the presence and position of the water and increased acidity of the -O2H group by binding to Fe(II), both the α- and β-anomers of SRH can utilize the same ring-opening mechanism! In both cases this step leads to formation of a corresponding -O2⁻ oxyanionic enolate intermediate. The subsequent step, transfer of a proton from Cys84-SH onto the ribosyl C1 centre, results in formation of a common active site-bound 2-keto containing intermediate. That is, regardless of the initial bound furanose-containing substrate, LuxS is able to catalyse their conversion to the same mechanistic intermediate.

Chapter 6 provides details of the UDP-galactopyranose mutase (UGM) catalyzed the inter-conversion of UDP-galactopyranose (UDP-Galp) to UDP-galactofuranose (UDP-Galf). The synergistically DFT and ONIOM QM/MM hybrid investigation were used to elucidate the mechanism of this important flavin enzyme and to provide insight into its uncommon mechanism. It is shown that the flavin must initially be in its fully reduced form as it must undergo redox during the course of the overall mechanism. Furthermore, it requires an N₅FAD–H proton which, through a series of tautomerizations is transferred onto the ring oxygen of the
substrate's Galp moiety to facilitate ring opening with concomitant Schiff base formation. Conversely, Galf formation is achieved via a series of tautomerizations involving proton transfer from the galactose's -O4GalH group ultimately onto the flavin's N5FAD center. With the DFT-cluster model, the overall rate limiting step is inter-conversion of two Galf-flavin tautomers: one containing a C4FAD-OH group and the other a tetrahedral protonated-N5FAD center. In contrast, in the QM/MM model a considerably more extensive chemical model was used that included all of the residues surrounding the active site, and modeled both their steric and electrostatic effects. In this approach, the overall rate-limiting step occurs during conformational rearrangement of the Schiff base cross-linked linear galactose-flavin complex. This appears due in part to a lack of suitable hydrophilic functional groups to facilitate the rearrangement in the enzyme active site.

Chapter 7 examines a substrate assisted mechanism of Thr-tRNA synthetase. DFT-based methods in combination with large chemical models have been used to investigate the mechanism of the second half-reaction catalyzed by Thr-tRNA synthetase; the aminoacyl transfer of Thr-AMP onto the A763'OH of the cognate tRNA. In the substrate assisted catalysis, the substrate actively participates in the catalytic process and supplies key functional groups for the activity of the enzyme. In this mechanism, the AMP-Thr takes advantage of binding with the metal ion to keep its amine group at a neutral protonation state. This allows the neutral amine moiety on the substrate to act as the general base to initiate the mechanism. Calculation results also indicate that this is the rate limiting step in the mechanism. A major difference between this enzyme and other tRNA synthetases is the presence of a neutral amine rather than a positively charged amine moiety. This is caused from the interaction between the substrate and the positively charged Zn(II) ion in the enzyme active site. The enzymatic mechanism of Thr-tRNA synthetase provides a very good
example that illustrates how different substrate structures and chemical properties can allow different catalytic pathways in a group of enzymes.

The above discussions demonstrate that the inherent properties of the substrate and the enzyme active site collectively affect the choice of the reaction pathway to produce the desired product. For instance, in substrate-assisted catalysis, at least initially, the enzyme interacts electrostatically with the substrate. By providing a suitable reaction environment, the enzyme is able to enhance the inherent chemistry of the substrate's functional groups, e.g., the basicity of the substrate's α-amino group in Thr-tRNA synthesis by ThrRS. However, as they work at the interface between nucleic acid and protein biochemistry, aaRS's are believed to be very ancient enzymes. In contrast, in enzymes that are thought to have evolved more recently such as the glycosidases, the enzyme's active site components are more directly involved in the catalytic mechanism. As result, the reactions and interplay between them and those of the substrates appear more complicated. It is noted, however, that their roles and functions still reflect the nature of the substrate upon which they act. Therefore, in a study of an enzymatic mechanism it is important to also be aware of the inherent chemistry and properties of the substrate as this can provide invaluable insights into an enzyme's choice of catalytic approaches. As the work presented in this thesis has shown, different enzymes are remarkably versatile in their catalytic approaches, thus enabling them to achieve outstandingly high rate enhancements end efficiencies.
**Vita Auctoris**

Name: WenJuan Huang  
Place of Birth: Chongqing, China  
Year of Birth: 1984

**Education**

2007 – 2012  
Ph.D. in Chemistry  
University of Windsor, Windsor, Ontario, Canada

2002 – 2006  
B.Sc. in Chemistry  
Lanzhou University, Lanzhou, Gansu Province, P. R. China

**Publications**


**Conference Presentations and Workshops (Presenter’s name in bold)**


