

University of Windsor

Scholarship at UWindor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2009

Computational Investigations into Nucleic Acid-Related Chemistry

Haining Liu
University of Windsor

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

Recommended Citation

Liu, Haining, "Computational Investigations into Nucleic Acid-Related Chemistry" (2009). *Electronic Theses and Dissertations*. 8110.
<https://scholar.uwindsor.ca/etd/8110>

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

Computational Investigations into Nucleic Acid-Related Chemistry

by

Haining Liu

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

2009

© 2009 Haining Liu



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-82073-5
Our file *Notre référence*
ISBN: 978-0-494-82073-5

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 protègent 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.


Canada

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 3 and 6–8 are in collaboration with my supervisor Professor James Gauld.

Chapter 4 is in collaboration with Dr. Jorge Llano under the supervision of Professor James Gauld.

Chapter 5 is in collaboration with Jesse Robinet and Professor Sirinart Ananvoranich 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 4 original papers that have been previously published/submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Full citation	Publication status
<i>Chapter 3</i>	Liu, H.; Gauld, J. W. <i>J. Phys. Chem. B</i> 2008 , <i>112</i> , 16874.	<i>Published</i>
<i>Chapter 4</i>	Liu, H.; Llano, J.; Gauld, J. W. <i>J. Phys. Chem. B</i>	<i>Submitted</i>
<i>Chapter 5</i>	Liu, H.; Robinet, J. J.; Ananvoranich, S.; Gauld, J. W.; <i>J. Phys. Chem. B</i> 2007 , <i>111</i> , 439.	<i>Published</i>
<i>Chapter 6</i>	Liu, H.; Gauld, J. W. <i>Phys. Chem. Chem. Phys.</i> 2009 , <i>11</i> , 278.	<i>Published</i>

I certify that I have obtained a 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 a 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 or Institution.

Abstract

Nucleic acids are biopolymers of nucleotides, which are composed of a phosphate, nucleobase and ribose sugar. In addition to acting as the genetic carrier, nucleic acids play a variety of other important roles in biological systems. In this thesis, nucleic acid-related chemistry is investigated using computational methods.

Chapter 1 presents an overview of the problems addressed in this thesis, whereas Chapter 2 discusses various theoretical methods. Then, Chapter 3 investigates the feasibility of using the phosphate oxygens as the general base to catalyze the aminoacyl transfer reaction in histidyl-tRNA synthetase. Three possible mechanisms with different phosphate oxygens acting as the base to abstract the 3'-OH group of A76 were examined and compared. Chapter 4 elucidates the catalytic mechanism of the repair of an alkylated nucleobase by the enzyme AlkB. It was found that this mechanism consists of four stages and that our calculated barrier for the rate-controlling step is in good agreement with experimental studies. Chapter 5 addresses the catalytic mechanism of the HDV ribozyme. Both cytosine and hydrated Mg^{2+} ion were found to be involved in the reaction with the former acting as the acid and the latter as the base. Chapter 6 studies the protonation of guanine quartets and quartet stacks. Each quartet plane was found to be able to accept maximally two protons. Chapter 7 deals with the interactions of metal ions with ribose and locked ribose. Four metal ions, Na^+ , K^+ , Mg^{2+} and Cd^{2+} were chosen and their properties upon interacting with ribose and locked ribose were compared. Chapter 8 presents the influences of the selection of computational methods and chemical models on the amide bond formation as catalyzed by the ribosome. Two proton transfer processes involving four- and six-membered transition structures were systematically examined using a variety of methods. Finally, Chapter 9 summarizes the main conclusions and possible extensions of the current work.

Dedication

This work is dedicated to my family.

Acknowledgments

First and foremost, I would like to thank my supervisor Dr. James Gauld for his guidance, support and encouragement during the past several years. The skills I learned in his lab are invaluable for my future career. His suggestions and help on my research are greatly appreciated.

I am also grateful to all the Gauld group members: Dr. Jorge Llano, Wenjuan Huang and Eric Bushnell. Your discussions on computational chemistry and various other topics are acknowledged. In addition, I also thank you for proofreading this thesis. Furthermore, sincere gratitude also goes to all the past group members: Jesse Robinet, Robert Ty, Vikram Velker, Indervir Mann, Dr. Kyung-Bin Cho and Cristina Baciú. All of you made my life here enjoyable. In particular, I would like to thank Jesse Robinet for staying with me in the office for two years and making those ‘fantastic’ movies, which are probably still somewhere on the internet.

Thanks are also extended to Joe Lichaa and Shihab Choudhury for helping fix the Linux cluster and workstations whenever we encountered a problem. In addition, the email instructions from Dr. Donald Kinghorn at Parallel Quantum Solutions LLC are also greatly acknowledged.

Finally, I would like to thank my parents for their support and encouragement during my stay in Canada. Special thanks also go to Lili Yang for her patience, understanding and support during the past three years.

Table of Contents

Declaration of Co-Authorship/Previous Publication	iii
Abstract.....	vi
Dedication	vii
Acknowledgements	viii
List of Tables	xv
List of Figures.....	xvii
List of Schemes.....	xxiii
List of Appendices (on CD).....	xxv
List of Abbreviations and Symbols	xxvii
Chapter 1 Introduction	1
References	2
Chapter 2 Theoretical Methods.....	4
2.1 Introduction	4
2.2 Schrödinger Equation	5
2.3 Born-Oppenheimer Approximation	6
2.4 Orbital Approximation	7
2.5 Basis Set Expansion.....	8
2.6 Variational Theorem.....	9
2.7 Hartree-Fock Theory	9
2.8 Electron Correlation	12
2.9 Correlated Wave Function Based Methods	13
<i>2.9.1 Configuration Interaction.....</i>	13
<i>2.9.2 Møller-Plesset Perturbation Theory</i>	14

2.10 Density Functional Theory	14
2.10.1 <i>Hohenberg-Kohn Existence Theorem</i>	15
2.10.2 <i>Kohn-Sham Theorem</i>	15
2.10.3 <i>B3LYP Functional</i>	17
2.11 Basis Sets	17
2.11.1 <i>Split-Valence Basis Sets</i>	18
2.11.2 <i>Polarization Basis Sets</i>	18
2.11.3 <i>Diffuse Functions</i>	19
2.14.4 <i>Effective Core Potentials</i>	19
2.12 Hierarchy of Computational Methods	19
2.13 Basis Set Superposition Error	20
2.14 Atoms In Molecules	21
2.14.1 <i>Topology of the Electron Density</i>	21
2.14.2 <i>Chemical Bonds in AIM</i>	23
2.15 Potential Energy Surface	24
2.16 Computational Tools	25
2.16.1 <i>Geometry Optimizations</i>	25
2.16.2 <i>Frequency Calculations</i>	25
2.16.3 <i>Single Point Calculations</i>	26
2.16.4 <i>Electrostatic Potential Surface</i>	26
2.17 Solvation	27
2.18 Technical Aspects and Units	28
References	29
Chapter 3 Substrate-Assisted Catalysis in the Aminoacyl Transfer	
Mechanism of Histidyl-tRNA synthetase: A DFT Study	32
3.1 Introduction	32

3.2 Computational Methods	36
3.3 Results and Discussion	38
3.3.1 <i>Active Site with Bound Substrate</i>	38
3.3.2 <i>Direct Cleavage: Proton Transfer to O_b</i>	41
3.3.3 <i>O_{pro-R} as the Base</i>	44
3.3.4 <i>O_{pro-S} as the Base</i>	46
3.3.5 <i>Additional Comparisons of the Three Mechanisms</i>	51
3.4 Conclusions	53
References	54
Chapter 4 A DFT Study of Nucleobase Dealkylation by the DNA Repair	
Enzyme AlkB	58
4.1 Introduction	58
4.2 Computational Methods	62
4.3 Results and Discussion	64
4.3.1 <i>Binding of Dioxygen to Iron in the Active-Site Complex</i>	64
4.3.1.1 <i>Binding of O_2 without Electron Transfer</i>	65
4.3.1.2 <i>Binding of O_2 with Electron Transfer</i>	66
4.3.2 <i>Activation of Oxygen Through Ferryl-Oxo $Fe^{IV}=O$ and Ferric-Oxyl $Fe^{III}-O^{\bullet}$ Intermediates</i>	68
4.3.2.1 <i>Spin-Quintet Activation Pathway</i>	68
4.3.2.2 <i>Spin-Septet Activation Pathway</i>	73
4.3.2.3 <i>Spin-Triplet Activation Pathway</i>	75
4.3.3 <i>Conformational Reorientation of the High-Valency Iron-Oxygen Groups</i>	77
4.3.4 <i>Oxidative Dealkylation of the Methylated Nucleobase</i>	81
4.4 Conclusions	85

References	87
Chapter 5	A Density Functional Theory Investigation on the Mechanism of the Hepatitis Delta Virus Ribozyme
	92
5.1 Introduction	92
5.2 Computational Methods	96
5.3 Results and Discussion	97
5.3.1 <i>Proton Affinity of C75</i>	97
5.3.2 <i>The Hydrated Mg²⁺ Ion Species</i>	101
5.3.3 <i>The Uncatalyzed Mechanism</i>	103
5.3.4 <i>The Catalyzed Mechanism</i>	105
5.4 Conclusions	108
References	110
Chapter 6	Protonation of Guanine Quartets and Quartet Stacks: Insights From DFT Studies
	114
6.1 Introduction	114
6.2 Computational Methods	117
6.3 Results and Discussion	118
6.3.1 <i>Unprotonated G4's</i>	118
6.3.2 <i>Singly Protonated G4 [(G4-H⁺)]</i>	122
6.3.3 <i>Doubly Protonated G4 [(G4-2H⁺)]</i>	124
6.3.4 <i>Triply and Quadruply Protonated G4's [(G4-3H⁺)] and [(G4-4H⁺)]</i>	127
6.3.5 <i>Singly Protonated 2G4 [(2G4-H⁺)]</i>	129
6.3.6 <i>Doubly Protonated 2G4 [(2G4-2H⁺)]</i>	131
6.3.7 <i>Triply and Quadruply Protonated 2G4's [(2G4-3H⁺)] and</i>	

	$[(2G4-4H^+)]$	135
	6.4 Conclusions	138
	References	140
Chapter 7	Interactions of Metal Ions with Ribose and Locked Ribose: A	
	DFT Study	145
	7.1 Introduction	145
	7.2 Computational Methods	147
	7.3 Results and Discussion	148
	7.3.1 <i>Basis Set Assessment</i>	148
	7.3.2 <i>Complexes of Ribose with Metal Ions</i>	152
	7.3.2.1 <i>Na⁺-Ribose Complexes</i>	151
	7.3.2.2 <i>K⁺-Ribose Complexes</i>	155
	7.3.2.3 <i>Mg²⁺-Ribose Complexes</i>	157
	7.3.2.4 <i>Cd²⁺-Ribose Complexes</i>	157
	7.3.3 <i>Complexes of Locked Ribose with Metal Ions</i>	160
	7.3.3.1 <i>Metal-LNA(O) Complexes</i>	161
	7.3.3.2 <i>Metal-LNA(N) Complexes</i>	163
	7.3.3.3 <i>Metal-LNA(Se) Complexes</i>	164
	7.5 Conclusions	166
	References	167
Chapter 8	Influences of Computational Methods and Chemical Models on	
	Amide Bond Formation	171
	8.1 Introduction	171
	8.2 Computational Methods	173
	8.3 Results and Discussion	174

8.3.1	<i>Small Model Calculations</i>	174
8.3.2	<i>Catalytic Hydrogen Bonding Interactions</i>	178
8.3.3	<i>Large Active-Site Model Calculations</i>	179
8.4	Conclusions	180
	References	181
Chapter 9	Conclusions	184
	References	190
	Vita Auctoris	192

List of Tables

Table 2.1	The Characterization, Description and Nature of the Four Types of Critical Points.....	22
Table 3.1	Values of the Electron Density (ρ) and its Laplacian ($\nabla^2\rho$) (a.u.) at BCP's of Selected Hydrogen Bonds Between the Bound Substrate and Active Site in 1	40
Table 3.2	Calculated Mulliken and NBO Charges On O_{carb} and O_b of the Scissile $C_{\alpha}-O_b$ bond For All Optimized Structures In This Chapter ...	51
Table 4.1	Relative Total Electronic Energies (kJ mol^{-1}), Selected Distances (\AA), Bond Angles ($^{\circ}$), and Spin Densities for the Iron-Dioxygen Complexes in Their Four Possible Spin-States	65
Table 4.2	Selected Distances (\AA) and Spin Densities for the Complexes Involved in the Activation of Oxygen.....	70
Table 5.1	Proton Affinity (kJ mol^{-1}) of Cytosine's N3 Ring Centre in Environments of Increasing Dielectric Constant (ϵ) Modeled Using Various Solvation Methods.....	99
Table 5.2	Proton Affinity (kJ mol^{-1}) of Cytosine's N3 Ring Centre with Various Bases Hydrogen Bonded to Cytosine's Exocyclic Amine as in Figure 5.2, in the Gas-Phase and Polar Environments ($\epsilon = 4.0$) as Modeled Using PB-SCRF and IEF-PCM Solvation Methods	100
Table 7.1	Selected Optimized Distances (\AA), Binding Energies (ΔE) and Basis Set Superposition Errors (kJ mol^{-1}) Obtained Using the B3LYP Method with Various Basis Sets for the 2',3'-Coordinated M^{n+} -Ribose Complex Shown in Scheme 7.2	149
Table 7.2	Calculated ^a Binding Energies (ΔE) in kJ mol^{-1} of the Complexes of Ribose with Metal Ions (Na^+ , K^+ , Mg^{2+} , Cd^{2+}) Obtained in This Chapter	154

Table 7.3	Calculated ^a Binding Energies (ΔE) in kJ mol^{-1} of the Complexes of Locked Ribose with Metal Ions ($M^{n+} = \text{Na}^+, \text{K}^+, \text{Mg}^{2+}, \text{Cd}^{2+}$) Obtained in This Chapter	160
Table 8.1	Selected Interaction Distances (\AA) and Reaction Barriers (kJ mol^{-1}) of Amide Bond Formation in the Small and Large Active-Site Model via Six- and Four-Membered Ring Transition Structures	176

List of Figures

Figure 2.1	A flow chart of the self-consistent field procedure	11
Figure 2.2	A schematic illustration of the linear combination of three GTFs to give a good approximation of an STO	18
Figure 2.3	Illustration of the hierarchy of computational methods	20
Figure 2.4	An example of a potential energy surface that shows the reaction between A and B to give C and D	24
Figure 2.5	The electrostatic potential surface of water obtained at the B3LYP/6-311+G(d,p) level of theory. Red and blue regions indicate electron rich and electron poor regions respectively. Oxygen and hydrogen atoms are indicated by red and white respectively	27
Figure 2.6	A schematic illustration of the implicit continuum solvation model	28
Figure 3.1	Optimized structure of the active site with bound substrate. Distances shown are in angstroms. Atom color key: C (gray); O (red); N (blue); P (orange) and H (white).....	38
Figure 3.2	Optimized structures of the transition structure (TS1) and product complex (2^b) for aminoacyl transfer via direct cleavage of the C_α — O_bP bond. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white).....	43
Figure 3.3	Schematic PES for aminoacyl transfer via direct cleavage of the C_α — O_bP bond. The Glu270 has been omitted for clarity	43
Figure 3.4	Optimized structures of the transition structure (TS2) and product complex (2^R) for aminoacyl transfer in which the substrate's non-bridging pro- <i>R</i> oxygen acts as the base. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white)	45

Figure 3.5	Schematic PES for the aminoacyl transfer reaction in which the substrate's non-bridging pro- <i>R</i> oxygen acts as the base. The Glu270 has been omitted for clarity	46
Figure 3.6	Optimized structures of the transition structures (TS3 and TS4), intermediate (3) and product complex (2 ^S) for aminoacyl transfer in which the substrate's non-bridging pro- <i>S</i> oxygen acts as the base. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white).	47
Figure 3.7	Schematic PES for the aminoacyl transfer reaction in which the substrate's non-bridging pro- <i>S</i> oxygen acts as the base. The Glu270 has been omitted for clarity	48
Figure 4.1	Active site of <i>E. coli</i> AlkB complexed with (a) α-ketoglutarate (PDB accession code: 2FD8) and (b) succinate (PDB accession code: 2FDG). The 1-methyladenine fragment of the trinucleotide dT-(1-me-dA)-dT substrate is also displayed	60
Figure 4.2	Structural models of the active site: (a) Arrangement of the catalytically active residues of AlkB from the crystal structure (PDB accession code: 2FD8). (b) Catalytic groups replacing the residues and 3me4amPym ⁺ replacing 1meAde ⁺ in the computational model. Atoms in red and with an asterisk were held fixed in geometry optimizations	63
Figure 4.3	Optimized structures of the spin-septet active-site complexes of dioxygen: (a) Monohapto-coordinated dioxygen to iron ⁷ [Fe ^{III} (η ¹ -O ₂ ^{•-})] in a bent end-on binding mode, and (b) Dihapto-coordinated dioxygen to iron ⁷ [Fe ^{III} (η ² -O ₂ ^{•-})] in a side-on binding mode	66
Figure 4.4	Potential energy surfaces for activation of oxygen through formation of the ferryl-oxo Fe ^{IV} =O and ferric-oxyl Fe ^{III} -O [•] compounds	69

Figure 4.5	Optimized structures of the spin-quintet transition states (⁵ TS1, ⁵ TS3 and ⁵ TS4), intermediates (⁵ IC2 and ⁵ IC3) and product complex (⁵ IC4) for formation of the ferryl-oxo Fe ^{IV} =O compound.....	71
Figure 4.6	Optimized structures of the spin-septet transition states (⁷ TS1 and ⁷ TS2), intermediate complexes (⁷ IC1 and ⁷ IC2) and product complex (⁷ IC4) for formation of the ferric-oxyl Fe ^{III} -O [•] compound.....	74
Figure 4.7	Optimized structures of the spin-triplet intermediate ³ IC1 and product complex ³ IC4 for formation of the ferryl-oxo Fe ^{IV} =O compound.....	77
Figure 4.8	Potential energy surfaces for reorientation of the activated oxygen ligated to the iron centre from axial to equatorial position in the septet, quintet and triplet spin-states: (a) Concerted pathway, (b) Stepwise pathway.....	79
Figure 4.9	Optimized structures for the spin-quintet ferryl-oxo reorientation from axial (⁵ IC5) to equatorial (⁵ IC6) position.....	80
Figure 4.10	Potential energy surfaces for dealkylation of 3me4amPym ⁺ as a substrate model in the triplet and quintet spin-states.....	82
Figure 4.11	Optimized structures of the spin-quintet transition states (⁵ TS7, ⁵ TS8 and ⁵ TS9), intermediates (⁵ IC7 and ⁵ IC8) and product complex (⁵ PC) for dealkylation of 3me4amPym ⁺ as a substrate model.....	84
Figure 5.1	Plot of the effect of increasing the polarity (ϵ) of the environment on the proton affinity of the N3 centre of cytosine using the PB-SCRF (red circles) and IEF-PCM (green circles) methods. Common non-polar ($\epsilon = 1$) value indicated by a black circle.....	98
Figure 5.2	Plot of the proton affinity of cytosine's N3 centre with various bases hydrogen bonded to its <i>trans</i> amine group hydrogen in the gas-phase	

	(black circle) and in polar environments ($\epsilon = 4.0$) modeled using the PB-SCRF (red circle) and IEF-PCM (green circle) methods.....	100
Figure 5.3	Schematic potential energy surfaces for the isomeric <i>endo</i> (solid blue) and <i>exo</i> (dashed red) uncatalyzed phosphodiester cleavage mechanisms in solution (see text)	104
Figure 5.4	Optimized structures showing selected interaction distances (Å) for the (1b) initial reactant complex, (2b) transition structure to formation of the intermediate, (3b) reaction intermediate and (4b) transition structure to formation of the final products in the HDV ribozyme catalyzed phosphodiester cleavage mechanism	106
Figure 5.5	Schematic potential energy surface for the catalyzed mechanism of the HDV ribozyme in solution. Formal charges are indicated.....	107
Figure 6.1	Optimized structures (top- and side-views) and molecular graphs for the lowest energy bifurcated and Hoogsteen hydrogen bonded neutral G4 structures. Symmetries and relative energies are also shown while selected optimized distances are in angstroms. Red and yellow dots in the molecular graph indicate BCP's and RCP's respectively. Atom colour key: C (gray); O (red); N (blue) and H (white)	120
Figure 6.2	Optimized structures (top- and side-views) for (G4- $n\text{H}^+$) $n = 1-4$. Symmetries and relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white)	121
Figure 6.3	Molecular graphs for the lowest energy (G4- $n\text{H}^+$) $n=1-4$ complex of each protonation state. Red and yellow dots indicate BCP's and RCP's respectively	123

- Figure 6.4** Optimized structures (top-view) of all singly protonated **2G4**'s and the side-view and molecular graph for **(2G4-H⁺)₃**. For clarity, not all bond paths are shown. Relative energies are also shown while selected optimized distances are in angstroms. Red dots indicate BCP's. Atom colour key: C (gray); O (red); N (blue) and H (white) 130
- Figure 6.5** Optimized structures for the three lowest energy and two selected higher energy **(2G4-2H⁺)** complexes. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white)..... 132
- Figure 6.6** The three lowest energy optimized structures (top-views) for **(2G4-nH⁺)** *n* = 3 and 4. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white) 136
- Figure 6.7** Optimized structures for **(2G4-3H⁺)₄** and **(2G4-4H⁺)₄** in which three and four guanines respectively in one plane have been protonated. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white)..... 138
- Figure 7.1** Optimized structure [B3LYP/6-311+G(2df,p)] of the C3'-*endo* ribose sugar conformation, in which the 3'-carbon is on the same side as the 5'-carbon. Selected distances shown are in angstroms 152
- Figure 7.2** Optimized structures [B3LYP/6-311+G(2df,p)] of the complexes of ribose with metal ions (Na⁺, K⁺, Mg²⁺, Cd²⁺). Selected distances (angstroms) are shown without brackets for Na⁺ and in round, square and curly brackets for K⁺, Mg²⁺ and Cd²⁺, respectively..... 153
- Figure 7.3** Optimized structures [B3LYP/6-311+G(2df,p)] of the complexes of locked ribose with metal ions (Na⁺, K⁺, Mg²⁺, Cd²⁺). Selected

distances (angstroms) are shown without brackets for Na^+ and in round, square and curly brackets for K^+ , Mg^{2+} and Cd^{2+} , respectively.. 159

- Figure 8.1** The large active model used in this chapter. Color key: C (gray); O (red); N (blue); and H (white). 174
- Figure 8.2** The optimized structures of (a) TS_4 and (b) TS_6 in the large active-site model at the HF/3-21G level of theory. Selected distances are shown in angstroms 179

List of Schemes

Scheme 1.1	Schematic illustration of the chemical structures of deoxyribonucleotide, ribonucleotide and their nucleobase component .	1
Scheme 3.1	General mechanism for the first half-reaction catalyzed by aminoacyl-tRNA synthetases	33
Scheme 3.2	General mechanism for the second half-reaction catalyzed by class II aminoacyl-tRNA synthetases.	33
Scheme 3.3	Proposed interactions and aminoacylation mechanism as catalyzed by histidyl-tRNA synthetase	35
Scheme 3.4	The active site model used to investigate possible mechanisms for the second half-reaction as catalyzed by HisRS. Atoms in blue were held fixed during optimizations.....	37
Scheme 4.1	Examples of alkylated nucleobases known to be repaired by the AlkB class of enzymes.	59
Scheme 4.2	Proposed dealkylation mechanism catalyzed by AlkB with 1-methyladenine (Ade ⁺ CH ₃) as the substrate	62
Scheme 5.1	Proposed mechanism 1 for the HDV ribozyme in which cytosine acts as a general base while a hydrated Mg ²⁺ ion acts a general acid	94
Scheme 5.2	Proposed ¹⁸ mechanism 2 for the HDV ribozyme in which cytosine acts as a general acid while an ionized hydrated Mg ²⁺ ion acts as a general base	95
Scheme 5.3	Schematic illustration of the chemical model used to investigate the catalytic mechanism of the HDV ribozyme	97
Scheme 6.1	A guanine quartet bound via Hoogsteen hydrogen bonds (R ₁₋₄ = DNA strands).....	115
Scheme 6.2	Schematic overview of the proposed mechanism for G ₄ 's catalyzed Zn ²⁺ insertion into a porphyrin ring	117

Scheme 7.1	Schematic illustration of the ribose sugar in RNA and examples of some modified ribose	146
Scheme 7.2	Schematic illustration of the 2',3'-coordinated M^{n+} -ribose complex ($M^{n+}=Na^+, K^+, Mg^{2+}$).....	148
Scheme 7.3	Schematic illustration of locked ribose considered in this chapter	158
Scheme 8.1	Schematic illustration of the proposed mechanism for amide bond formation as catalyzed by the ribosome via (a) six-membered ¹⁵⁻²² and (b) four-membered ^{23,24} ring transition structure (see text).	172
Scheme 8.2	Schematic illustration of the P-site and A-site used in the small model.....	173

List of Appendices (on CD)

Appendix A

Table A.1 Optimized xyz-coordinates, ZPVE corrections and solution single point energies (h) for all species considered in Chapter 3.

Table A.2 The electron density (ρ) and Laplacian of the electron density ($\nabla^2\rho$) (a.u.) at selected BCPs of hydrogen bonds found within the small substrate-bound active site model (see text). Selected distances are in angstroms.

Appendix B

Table B.1 Optimized xyz-coordinates and solution single point energies (h) for all species considered in Chapter 4.

Appendix C

Table C.1 Optimized xyz-coordinates, single point energies and ZPVE corrections (h) for all species considered in Chapter 5.

Appendix D

Figure D.1 Electrostatic potential surfaces (EPS's), obtained using GaussView, for the most stable bifurcated and Hoogsteen hydrogen bonded neutral **G4**, the most stable (**G4-nH⁺**) $n=1-4$ complexes, and top- and side-views of (**2G4-H⁺**)₁, (**2G4-H⁺**)₃, (**2G4-2H⁺**)₂ and (**2G4-2H⁺**)₄ (see text). The EPS's of the neutral **G4** and protonated **G4** complexes were obtained at the B3LYP/6-31G(d,p) level of theory while those of the protonated **2G4** complexes were obtained at the B3LYP/6-31G(d) level of theory. Blue and red indicate electron-poor and -rich regions respectively.

Figure D.2 Optimized structures (top- and side-views) for (**2G4-nH⁺**) $n=2-4$. Selected distances are shown in angstroms.

Table D.1 Optimized xyz-coordinates and single point energies (h) for all species considered Chapter 6.

Table D.2 Electron density (ρ), Laplacian ($\nabla^2\rho$) in a.u., and ellipticity (ϵ) for selected BCP's.

Appendix E

Table E.1 Gaussian archives and ZPVE corrections (h) for all species considered in Chapter 7.

Appendix F

Table F.1 Optimized xyz-coordinates, ZPVE corrections and energies (h) for all species considered in Chapter 8.

List of Abbreviations and Symbols

aaRS	Aminoacyl-tRNA synthetase
A	adenine
AIM	Atoms In Molecules
Arg	Arginine
Asp	Aspartate
ATP	Adenosine Triphosphate
AMP	Adenosine Monophosphate
B	Becke's 88 exchange functional
BCP	Bond Critical Point
B3LYP	Becke's three parameter hybrid functional
BSSE	Basis Set Superposition Error
C	Cytosine
CI	Configuration Interaction
CID	Configuration Interaction with Doubles
CISD	Configuration Interaction with Singles and Doubles
CP	Counterpoise Correction
Cys	Cysteine
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
G4	Guanine Quartet
G4 _n	Guanine Quadruplex
GGA	Generalized Gradient Approximation
Gln	Glutamine
Glu	Glutamate
GTF	Gaussian-Type Functions

HDV	Hepatitis Delta Virus
HF	Hartree-Fock
His	Histidine
IC	Intermediate Complex
IEF	Integral Formalism Equation
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LYP	Lee-Yang-Parr correlation functional
MPn	<i>n</i> th order Møller-Plesset
NBO	Natural Bond Orbital
PA	Proton Affinity
PB	Poisson-Boltzmann
PC	Product Complex
Phe	Phenylalanine
PCM	Polarizable Continuum Method
PDB	Protein Data Bank
PES	Potential Energy Surface
PP _i	Pyrophosphate
RC	Reactant Complex
RCP	Ring Critical Point
RHF	Restricted Hartree-Fock
RNA	Ribonucleic Acid
SAC	Substrate-Assisted Catalysis
SCF	Self-Consistent Field
SCRf	Self-Consistent Reaction Field
STO	Slater-Type Orbitals
Trp	Tryptophan

TS	Transition Structure
Tyr	Tyrosine
U	Uracil
UHF	Unrestricted Hartree-Fock
VWN	Vosko-Wilk-Nusair
ZPVE	Zero Point Vibrational Energy
α -KG	α -Ketoglutarate

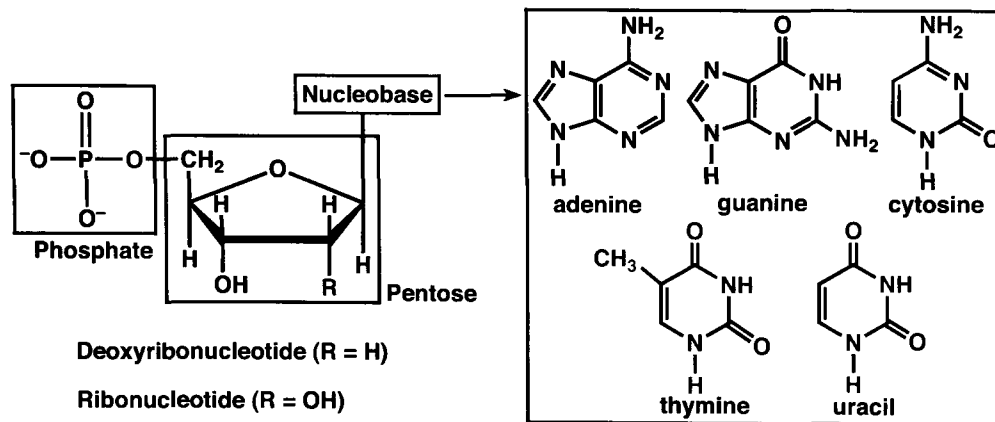
\hat{H}	Hamiltonian operator
E	Energy
Ψ	Wave function
r	Distance between two particles
χ_i	i th spin orbital
c_{ui}	Molecular orbital coefficient
ϕ	Basis function
F_{uv}	Fock matrix
S_i	i th spin function
S_{uv}	Overlap matrix
ε_i	Orbital energy for the i th molecular orbital
λ	Dimensionless parameter
ρ	Density
∇	Laplacian
E^T	Kinetic energy term
E^V	Potential energy term
E^J	Repulsion energy term
E^{xc}	Exchange-correlation energy term
ε	Dielectric constant
ε	Ellipticity
\AA	Angstrom
$^\circ$	Degree
a.u.	Atomic unit

Chapter 1

Introduction

Nucleic acids are the carriers of genetic information in biological systems. The basic component of these molecules is the nucleotide, which is composed of a pentose sugar, a nucleobase and a phosphate. The pentose sugar is either the deoxyribose or ribose, which classifies the nucleic acids into deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), respectively (**Scheme 1.1**). Another difference between DNA and RNA is the nucleobase attached to the pentose sugar. Although they have three common nucleobases: adenine, guanine and cytosine, the fourth one in DNA is thymine while in RNA it is uracil.¹

Scheme 1.1 Schematic illustration of the chemical structures of deoxyribonucleotide, ribonucleotide and their nucleobase component.



Nucleotides and nucleic acids play a variety of important roles in biological systems. For example, nucleotides are able to act as energy carriers, e.g. adenosine 5'-

triphosphate (ATP), messenger molecules, e.g. adenosine 3',5'-cyclic monophosphate (cAMP), key components of some enzyme cofactors, e.g. coenzyme A, and HIV drugs, e.g. 3'-azido-2',3'-dideoxythymidine (AZT).¹ In addition, nucleic acids are able to act as catalysts within cells.²⁻⁸ For instance, the active site of the ribosome is composed purely of RNA, in which proteins are synthesized.⁹⁻¹⁴ Furthermore, the human's telomere ends of DNA form quadruplex structures, which inhibits the enzyme telomerase in cancerous cells.^{15,16}

As a result, there is tremendous interest in understanding the chemistry and properties of nucleic acids. In addition to experimental techniques, computational chemistry is an alternative approach to provide invaluable insights into chemical problems. Such a method has the advantage of following the details of reaction pathways and to study highly reactive molecules as straightforwardly as stable ones. This thesis details our computational investigations on various aspects of nucleic acid-related chemistry.

References

- (1) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 4th ed.; W. H. Freeman and Company: New York, 2004.
- (2) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R. *Science* **2002**, *298*, 1421.
- (3) Das, S. R.; Piccirilli, J. A. *Nature Chem. Biol.* **2005**, *1*, 45.
- (4) Zhao, Z.-Y.; McLeod, A.; Harusawa, S.; Araki, L.; Yamaguchi, M.; Kurihara, T.; Lilley, D. M. J. *J. Am. Chem. Soc.* **2005**, *127*, 5026.
- (5) Bevilacqua, P. C.; Yajima, R. *Curr. Opin. Chem. Biol.* **2006**, *10*, 455.
- (6) Walter, N. G. *Mol. Cell* **2007**, *28*, 923.
- (7) Cochrane, J. C.; Strobel, S. A. *Acc. Chem. Res.* **2008**, *41*, 1027.

- (8) Doudna, J. A.; Lorsch, J. R. *Nat Struct Mol Biol* **2005**, *12*, 395.
- (9) Dorner, S.; Polacek, N.; Schulmeister, U.; Panuschka, C.; Berta, A. *Biochem. Soc. Trans.* **2002**, *30*, 1131.
- (10) Chamberlin, S. I.; Merino, E. J.; Weeks, K. M. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14688.
- (11) Dorner, S.; Panuschka, C.; Schmid, W.; Barta, A. *Nucleic Acids Res.* **2003**, *31*, 6536.
- (12) Schmeing, T. M.; Huang, K. S.; Kitchen, D. E.; Strobel, S. A.; Steitz, T. A. *Mol. Cell* **2005**, *20*, 437.
- (13) Beringer, M.; Rodnina, M. V. *Mol. Cell* **2007**, *26*, 311.
- (14) Huang, K. S.; Carrasco, N.; Pfund, E.; Strobel, S. A. *Biochemistry* **2008**, *47*, 8822.
- (15) O'Reilly, M.; Teichmann, S. A.; Rhodes, D. *Curr. Opin. Struct. Biol.* **1999**, *9*, 56.
- (16) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847.

Chapter 2

Theoretical Methods

2.1 Introduction

Chemistry has traditionally been recognized as an experimental science. For example, new chemical reactions and synthetic methods are developed and tested in the laboratory. Although great understanding of chemistry has been gained from the laboratory, not every chemical species can be easily studied experimentally. For instance, highly reactive molecules such as intermediates in the reaction pathway are often difficult to capture, and hence, their properties are sometimes unknown. An alternative approach to study such molecules is to model them using computers. This branch of chemistry is known as computational chemistry.

Computational chemistry can be defined as the application of mathematical methods to investigate chemical problems.¹ The methods used in this thesis originate from quantum chemistry. To better interpret the results obtained, a thorough understanding of quantum chemistry is needed. Detailed explanations of the fundamental concepts and theory of quantum chemistry can be found in several standard textbooks.²⁻⁴ In this chapter, only those related to this thesis are briefly discussed.

2.2 Schrödinger Equation

One of the fundamental equations in quantum chemistry is the Schrödinger equation. Generally, it is classified as time-dependent and time-independent. The time-dependent Schrödinger equation can be written as:⁵

$$i\hbar \frac{\partial \Psi(r,t)}{\partial t} = \hat{H}\Psi(r,t) \quad (2.1)$$

where the wave function $\Psi(r,t)$ is a function of the coordinates (r) and time (t), \hat{H} is the Hamiltonian operator and $\hbar = h/2\pi$, where h is Planck's constant.

Many chemical properties that we are interested in normally do not change with time. Thus, it is sufficient to use the time-independent Schrödinger equation:

$$\hat{H}\Psi(r) = E\Psi(r) \quad (2.2)$$

In the time-independent Schrödinger equation, the wave function is a function of the coordinates only and contains all the information about a system. Although the wave function cannot be experimentally observed, its probability density $|\Psi|^2$ represents the probability of finding the particles in certain locations and is experimentally measurable. Furthermore, this equation offers an approach to obtain the energy of a system.

For a chemical system, the Hamiltonian operator (field free, non-relativistic) in atomic units is:

$$\hat{H} = -\sum_{i=1}^{N_E} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{N_N} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N_E} \sum_{A=1}^{N_N} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_E} \sum_{j>i}^{N_E} \frac{1}{r_{ij}} + \sum_{A=1}^{N_N} \sum_{B>A}^{N_N} \frac{Z_A Z_B}{r_{AB}} \quad (2.3)$$

where i and j represent electrons, A and B represent nuclei, M is the mass of the nuclei, Z is the atomic number of the nuclei and r is the distance between two particles (electron–nucleus, electron–electron, or nucleus–nucleus). The first two terms are the kinetic

energies of the electrons and nuclei, respectively. The third term is the potential energy for the attraction between electrons and nuclei. The fourth and fifth terms are the potential energies for the repulsion between electrons and electrons, and nuclei and nuclei, respectively.

Due to the complexity of the Hamiltonian operator, the Schrödinger equation can only be solved exactly for hydrogen and hydrogen-like atoms, such as He^+ and Li^{2+} . For all the other atoms and molecules, approximations must be made.

2.3 Born-Oppenheimer Approximation

The first approximation to simplify the molecular Hamiltonian is the Born-Oppenheimer approximation. It assumes that the electrons are moving in a field of fixed nuclei because the nuclei are much heavier and hence move much slower than the electrons. Therefore, the kinetic energies of nuclei are now zero and the potential energies for nuclei repulsion are constants. Based on these assumptions, the Hamiltonian is simplified as:

$$\hat{H} = -\sum_{i=1}^{N_E} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N_E} \sum_{A=1}^{N_N} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_E} \sum_{j>i}^{N_E} \frac{1}{r_{ij}} + V_{NN} \quad (2.4)$$

where V_{NN} are the potential energies for nuclei repulsion. The first three terms are normally referred to as the electronic Hamiltonian operator \hat{H}_{el} :

$$\hat{H}_{el} = -\sum_{i=1}^{N_E} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N_E} \sum_{A=1}^{N_N} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_E} \sum_{j>i}^{N_E} \frac{1}{r_{ij}} \quad (2.5)$$

It should be noted that the first and second terms in \hat{H}_{el} involve only one electron. However, the third term $1/r_{ij}$ deals with two electrons. Thus, different approaches are taken to treat these two parts, which are discussed in **Section 2.7 — 2.9**.

Hence, the Schrödinger equation can now be rewritten as:

$$(\hat{H}_{\text{el}} + V_{NN})\psi_{\text{el}}(r_e; r_N) = E_{\text{total}}\psi_{\text{el}}(r_e; r_N) \quad (2.6)$$

where r_e and r_N are the coordinates of electrons and nuclei, respectively and E_{total} is the total electronic energy.

Practically, (2.6) is solved without the inclusion of V_{NN} . In this case, the obtained eigenvalue E is referred to as the ‘pure electronic energy’³ and V_{NN} is then added to obtain E_{total} . Importantly, E_{total} depends parametrically on the coordinates of the nuclei. This is the foundation for the concept of potential energy surface, which is discussed in **Section 2.15**.

2.4 Orbital Approximation

The orbital approximation provides a further simplification of the wave function. It assumes that the motions of electrons are independent from one another. Thus, each electron can be assigned to its own one-electron wave function or spin-orbital (χ_i):

$$\chi_i = \psi_i S_i \quad (2.7)$$

where ψ_i is the spatial function, which only depends on the coordinates of the i th electron, and S_i is the spin function. The spatial orbital is also known as a molecular orbital.

Thus, the wave function for an n -electron system can now be written in a product form, known as the Hartree-product:

$$\Psi_{\text{product}} = \chi_1(1)\chi_2(2)\cdots\chi_n(n) \quad (2.8)$$

In this way, the wave function is simplified to a series of one electron problems.

However, the Hartree-product is not an acceptable wave function. According to the Pauli Exclusion Principle, the wave function must change sign upon the interchange of the coordinates of any two electrons. That is, the wave function must be anti-symmetric. To satisfy this rule, the Hartree-product is written in a determinantal form, known as the Slater determinant.

$$\Psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_{n-1}(1) & \chi_n(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_{n-1}(2) & \chi_n(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \chi_1(n-1) & \chi_2(n-1) & \cdots & \chi_{n-1}(n-1) & \chi_n(n-1) \\ \chi_1(n) & \chi_2(n) & \cdots & \chi_{n-1}(n) & \chi_n(n) \end{vmatrix} \quad (2.9)$$

where $\frac{1}{\sqrt{n!}}$ is the normalization constant so that the wave function meets the normalization requirement:

$$\int \Psi^* \Psi d\tau = 1 \quad (2.10)$$

where Ψ^* is the complex conjugate of Ψ and the variable of integration is over all space.

2.5 Basis Set Expansion

The molecular orbital ψ_i can be further simplified by expressing it as a linear combination of one-electron functions, known as basis functions ϕ_μ .

$$\psi_i = \sum_{\mu=1}^N c_{\mu} \phi_{\mu} \quad (2.11)$$

where c_{μ} is the molecular orbital coefficient for the μ th basis function on the i th molecular orbital and ϕ_{μ} is the basis function. Ideally, an infinite number of basis functions should be used to form ψ_i . However, in practice only a limited number of basis functions are used. The accuracy of ψ_i can be improved by increasing the number of basis functions. In addition, if atomic orbitals are used as the basis functions, (2.11) is known as the linear combination of atomic orbitals (LCAO).

2.6 Variational Theorem

The variational theorem states that the energy obtained from any approximate wave function Ψ_{approx} , also known as the trial function, is always greater than or equal to the exact energy E_{exact} :

$$E_{\text{approx}} = \frac{\int \Psi_{\text{approx}}^* \hat{H} \Psi_{\text{approx}} d\tau}{\int \Psi_{\text{approx}}^* \Psi_{\text{approx}} d\tau} \geq E_{\text{exact}} \quad (2.12)$$

The variational theorem provides a practical approach to determine the molecular orbital coefficient $c_{\mu i}$ in (2.11). That is, these coefficients are varied until the energy obtained from the trial function Ψ_{approx} reaches the minimum. The resulting trial function is then the best approximate wave function. Importantly, it is found that E_{approx} approaches E_{exact} faster than Ψ_{approx} approaches Ψ_{exact} . Thus, a reasonable energy may be obtained from a relatively poor wave function. In addition, a computational method is classified as variational if the energy it yields is always an upper bound to the exact energy.

2.7 Hartree-Fock Theory

The Hartree-Fock method is the simplest *ab initio* method. The Latin word *ab initio* means ‘from the beginning’. In *ab initio* methods, only the fundamental physical constants and mathematical approximations are used. In the Hartree-Fock theory, only the average Coulombic electron-electron repulsions are taken into account. That is, the $1/r_{ij}$ term in \hat{H}_e is replaced by an integration of the repulsion between one electron and the charge density of all other electrons. Thus, the one-electron Hamiltonian operator \hat{h}_i for each orbital ψ_i is written as:³

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^{N_A} \frac{Z_A}{r_{iA}} + \sum_{j=1}^i \int \frac{\rho_j}{r_{ij}} d\mathbf{r} \quad (2.13)$$

where ρ_j is the charge density of electron j and can be calculated as

$$\rho_j = |\psi_j|^2 \quad (2.14)$$

Hence, the electron repulsion is not considered explicitly in the Hartree-Fock method and each electron feels an average effect of all the other electrons.

In Hartree-Fock theory, the molecular orbital coefficient c_μ is obtained from the Roothaan-Hall equations:

$$\sum_{\mu=1}^N c_\mu (F_{\mu\nu} - \epsilon_i S_{\mu\nu}) = 0 \quad (2.15)$$

where $F_{\mu\nu}$ is the Fock matrix which represents the effects of each electron in a field of the nuclei and all other electrons, ϵ_i is the orbital energy of molecular orbital ψ_i and $S_{\mu\nu}$ is the overlap matrix which represents the overlap between orbitals.

Because the Fock matrix $F_{\mu\nu}$ itself depends on the molecular orbital coefficient c_μ , the Roothaan-Hall equations (2.15) can only be solved iteratively. That is, the Fock matrix $F_{\mu\nu}$ is first formed using a set of guessed molecular orbital coefficients c_μ 's. This Fock matrix is then used to solve the Roothaan-Hall equations, which will generate an improved set of molecular orbital coefficients. From these improved coefficients, a new Fock matrix is formed and used to solve the Roothaan-Hall equations again. This process continues until there are no significant improvements of the coefficients and energies between the last two cycles. This procedure is known as the self-consistent field (SCF) procedure (**Figure 2.1**).

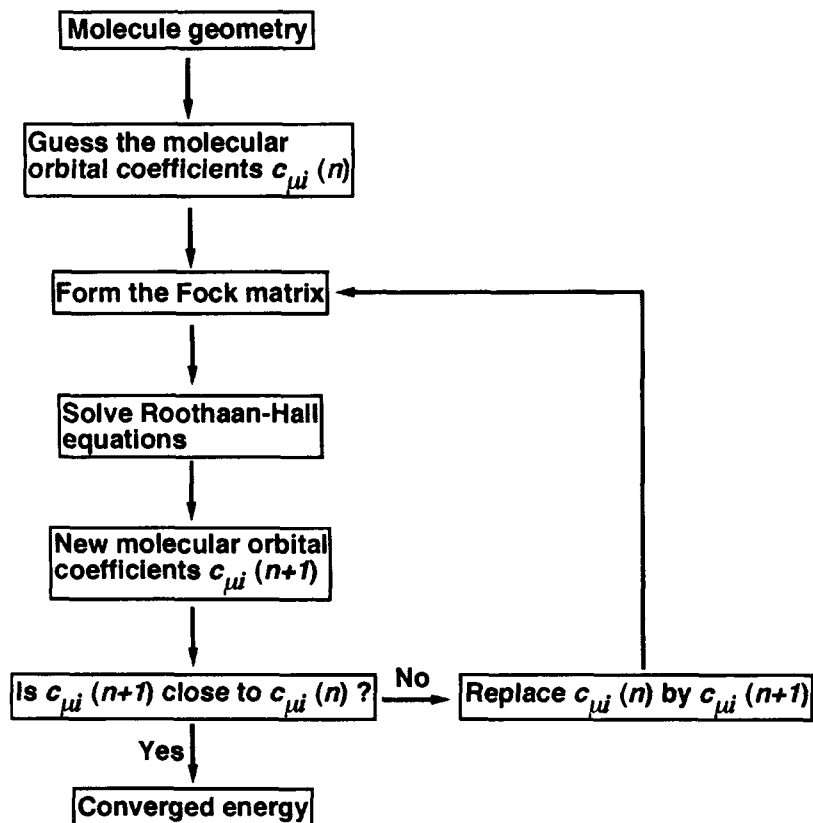


Figure 2.1 A flow chart illustrating the self-consistent field procedure.

Importantly, the Hartree-Fock method is variational. That is, the energy that it yields is always larger than the exact ground-state energy. Thus, with the inclusion of more basis functions, the energy obtained from Hartree-Fock method tends to approach a limit, known as the Hartree-Fock limit (E_{HF}).

So far, electrons are assigned to molecular orbitals in pairs and the same spatial orbital is used for both the α and β spin electrons. However, this is only applicable for the systems without unpaired electrons (i.e., closed-shell systems). This type of Hartree-Fock method is called the spin-restricted Hartree-Fock (RHF) method. For the systems with one or more unpaired electrons (i.e., open-shell systems), a different approach must be used.

A commonly used method to describe open-shell systems is the spin-unrestricted Hartree-Fock (UHF) method. In the UHF method, the α and β spin electrons are treated differently. That is, they are assigned to separate spatial orbitals with different sets of molecular orbital coefficients.

$$\psi_i^\alpha = \sum_{\mu=1}^N c_{\mu}^\alpha \phi_\mu \quad \psi_i^\beta = \sum_{\mu=1}^N c_{\mu}^\beta \phi_\mu \quad (2.16)$$

The resulting Roothaan-Hall equations, now known as the Pople-Nesbet equations, are also solved separately for the α and β spin electrons.

$$\sum_{\mu=1}^N c_{\mu}^\alpha (F_{\mu\nu}^\alpha - \epsilon_i^\alpha S_{\mu\nu}) = 0 \quad \sum_{\mu=1}^N c_{\mu}^\beta (F_{\mu\nu}^\beta - \epsilon_i^\beta S_{\mu\nu}) = 0 \quad (2.17)$$

Due to the different treatment of the α and β spin electrons, the UHF wave function is no longer an eigenfunction of the total spin operator. This introduces an error by mixing higher spin states, known as spin contamination. Fortunately, it is not significant for most systems.

2.8 Electron Correlation

The Hartree-Fock method is referred to as a non-correlated method, as the electron repulsions are not considered explicitly but rather taken into account in an average way. However, the motion of an electron is influenced by all the other electrons. This effect is called electron correlation. The difference between the Hartree-Fock limit energy (E_{HF}) and the exact energy (E_{exact}) is called the correlation energy (E_{corr}).

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \quad (2.18)$$

There are two approaches to account for electron correlation. The first is to use the correlated wave function based methods, which are mathematically complicated and

computationally expensive. An alternative way is to use density functional theory methods, which have the electron correlation inherently built in the potential and are less computational demanding. These two methods are discussed below.

2.9 Correlated Wave Function Based Methods

These methods add extra terms to the Hartree-Fock wave function in order to account for electron correlations. Two common approaches are discussed below.

2.9.1 Configuration Interaction

In the configuration interaction (CI) methods, new determinants that promote electrons from the occupied to virtual orbitals (i.e., excited states), are added to the Hartree-Fock wave function.

$$\Psi_{\text{CI}} = a_0 \Psi_{\text{HF}} + \sum_i^{\text{occ}} \sum_r^{\text{vir}} a_i^r \Psi_i^r + \sum_{i < j}^{\text{occ}} \sum_{r < s}^{\text{vir}} a_{ij}^{rs} \Psi_{ij}^{rs} + \dots \quad (2.19)$$

where i and j are occupied molecular orbitals in Ψ_{HF} and r and s are virtual orbitals in Ψ_{HF} .

If all the possible excitation states are included, this gives the full configuration interaction (Full-CI) method. By using a basis set of infinite size, Full-CI is able to solve the Schrödinger equation exactly. However, this is extremely computationally expensive. In practice, only a limited number of excitations are included. For example, the configuration interaction, singles (CIS) and configuration interaction, doubles (CID) include the single and double excitations respectively.

2.9.2 Møller-Plesset Perturbation Theory

Perturbation theory takes another approach to account for electron correlation. It adds a perturbation λV to the Hartree-Fock Hamiltonian \hat{H}_0 in order to obtain the complete Hamiltonian \hat{H} .

$$\hat{H} = \hat{H}_0 + \lambda V \quad (2.20)$$

where λ is a dimensionless parameter.

The wave function and energy of a system described by the Hamiltonian \hat{H} can be further expanded as a Taylor series:

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \quad (2.21)$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (2.22)$$

The Møller-Plesset perturbation method is referred by the order at which the Taylor series are truncated. For example, MP2 and MP4 indicate the inclusion of the second and fourth order perturbations respectively.

2.10 Density Functional Theory

Instead of a wave function, density functional theory (DFT) methods are based on the use of the electron density. This results in several advantages compared with the conventional *ab initio* methods. First, the electron density is an experimentally measurable quantity. In contrast, the wave function lacks any physical significance and is not experimentally observable. In addition, DFT methods are much cheaper than correlated *ab initio* methods, as the electron density is a three-dimensional function regardless of the number of electrons in a system. Furthermore, DFT methods inherently include electron correlation, which makes the accuracy of DFT calculations generally

comparable to high level *ab initio* calculations.¹ Due to these advantages, DFT methods have been widely used to study large chemical systems, such as enzymatic reactions.⁶⁻⁸

2.10.1 Hohenberg-Kohn Existence Theorem

The first question one may have for DFT methods is: can we obtain the properties of a molecular system purely from the electron density? This is answered by the Hohenberg-Kohn existence theorem, which states that the ground-state molecular energy, wave function and all other molecular electronic properties are uniquely determined by the ground-state electron probability density of the molecule. Hence, the ground-state energy of a molecule can be written as:

$$E_0 = E_0[\rho_0] \quad (2.23)$$

where E_0 is the ground-state energy and ρ_0 is the ground-state electron density. It should be noted that the electron density ρ_0 itself is a function of the x, y, z coordinates. Thus, DFT methods use a functional, which is a function of a function, to obtain the molecular energy.

Although the Hohenberg-Kohn existence theorem provides a theoretical proof to use the electron density instead of wave function to obtain the energy of a molecule, it does not tell us how this can be done.

2.10.2 Kohn-Sham Theorem

Kohn and Sham showed that the ground-state electron density can be written as the sum of one-electron orbitals (ψ_i^{KS}), known as the Kohn-Sham orbitals.

$$\rho = \sum_{i=1}^n |\psi_i^{\text{KS}}|^2 \quad (2.24)$$

These Kohn-Sham orbitals are expressed as a linear combination of basis functions similar to the HF method. However, the Kohn-Sham orbitals have no physical significance.

In addition, they showed that the exact total electronic energy (E_{exact}) is composed of four terms, each of which is a functional of the electron density:

$$E_{\text{exact}}[\rho] = E^{\text{T}}[\rho] + E^{\text{V}}[\rho] + E^{\text{J}}[\rho] + E^{\text{XC}}[\rho] \quad (2.25)$$

where E^{T} is the kinetic energy of electrons, E^{V} is the potential energy for the attraction between electrons and nuclei, E^{J} is the potential energy for the repulsion between electrons and electrons, and E^{XC} is the exchange-correlation energy.

Unfortunately, the exact exchange-correlation energy term E^{XC} is not known and approximations must be used. In practice, this term is usually separated into the exchange energy term (E^{X}) and the correlation energy term (E^{C}).

$$E^{\text{XC}} = E^{\text{X}} + E^{\text{C}} \quad (2.26)$$

Recently, a variety of exchange and correlation functionals have been developed. Generally, they are classified as the local density approximation (LDA) functional, generalized gradient approximation (GGA) functional, meta generalized gradient approximation (meta-GGA) functional, and hybrid functional. The LDA functional, such as the Vosko-Wilks-Nusair (VWN) correlation functional,⁹ only includes the electron density at local points. The GGA functional, such as the Becke 86 (B86) exchange functional,¹⁰ include both the electron density and its gradient. The meta-GGA functional, such as the Minnesota series of functionals M06-L,^{11,12} further includes the non-interacting kinetic energy density. The hybrid functional, such as the Becke's 3-parameter exchange functional^{13,14} and Lee-Yang-Parr correlation functional¹⁵ (B3LYP), mixes the

functional from other methods. The key to the success of a DFT calculation is to choose these functionals properly.

2.10.3 B3LYP Functional

The B3LYP method¹³⁻¹⁵ is one of the most popular DFT methods and has been used throughout this thesis. It is a hybrid functional formed by mixing the LDA and GGA functionals as well as the Hartree-Fock exchange, and is mathematically expressed as:

$$E_{xc}^{\text{B3LYP}} = (1-a)E_x^{\text{LSDA}} + aE_x^{\text{HF}} + b\Delta E_x^{\text{B}} + (1-c)E_c^{\text{LSDA}} + cE_c^{\text{LYP}} \quad (2.27)$$

The values of a , b and c are optimized to best reproduce experimental data, such as the atomization energies, ionization potentials and proton affinities.

2.11 Basis Sets

The basis set is the set of basis functions used to describe the orbitals in a molecule. There are two main types of basis functions: the Slater-type orbitals (STOs) and Gaussian-type functions (GTFs). The STOs are able to accurately describe atomic orbitals. However, they are very computationally demanding. In contrast, the GTFs are less accurate. However, they are very computationally efficient. In practice, one can use a linear combination of GTFs to obtain a good approximation to an STO (**Figure 2.2**) and still computationally cheaper than using STOs directly. Each of the Gaussian functions is called a primitive function and the resulting basis function is called a contracted basis function.³ When the number of basis functions increases, the atomic orbital is generally better described. Several types of basis sets are discussed below.

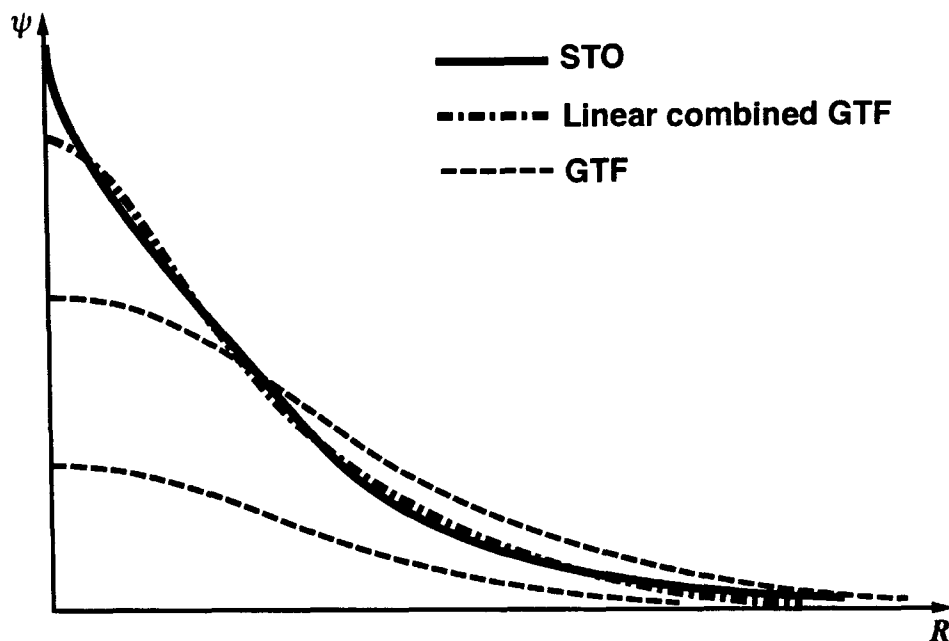


Figure 2.2 A schematic illustration of the linear combination of three GTFs to give a good approximation of an STO.

2.11.1 Split-Valence Basis Sets

Better descriptions of the valence orbitals may be gained by the split-valence basis sets, which use more than one basis functions to describe the valence orbitals. Thus, they allow a greater flexibility of the radial size of the valence orbitals. The double- ζ basis set, such as 6-31G, uses two basis functions per valence orbital while the triple- ζ basis set, such as 6-311G, uses three.

2.11.2 Polarization Basis Sets

These basis sets add basis functions of higher orbital angular-momentum quantum numbers. Thus, they change the shape of an orbital by displacing the electron density away from the nuclear centers. Two polarization basis sets used in this thesis are LACVP* and 6-31G(d,p). The former includes d functions on heavy atoms while the latter includes both d functions on heavy atoms and p functions on hydrogen.

2.11.3 Diffuse Functions

To allow the orbitals to occupy the region far from the nuclei, one can use diffuse functions, which are spatially large atomic orbitals. These functions are denoted by “+” in the Pople basis sets, such as 6-31+G(d,p) and 6-31++G(d,p). The former includes diffuse functions on heavy atoms while the latter includes diffuse functions on both heavy atoms and hydrogen.

2.11.4 Effective Core Potentials

Effective core potentials (ECPs) are used for atoms beyond the third row in the periodic table. These atoms are associated with a large number of electrons. However, most of them are core electrons. To save computational costs, these core electrons are replaced with analytic functionals, known as ECPs.³ In addition, for heavy elements ECPs are also able to take into account the relativistic effects. An example of ECP used in this thesis is the LACVP basis set developed at Los Alamos National Laboratory.¹⁶

2.12 Hierarchy of Computational Methods

The Pople diagram (**Figure 2.3**) shows the hierarchy of computational methods and the relation of increasing the electron correlation, size of basis set and exact solution. Generally, the accuracy of a calculation is improved with increasing electron correlation and basis set size. The exact solution of the Schrödinger equation is obtained if the Full-CI method with an infinite size of basis set is used. However, this can never be done except for some small molecules due to the huge computational costs associated with this procedure. Thus, we must make a compromise between accuracy and computational cost. Normally, a higher level of theory is needed to obtain accurate energies. This is discussed in **Section 2.16.3**.

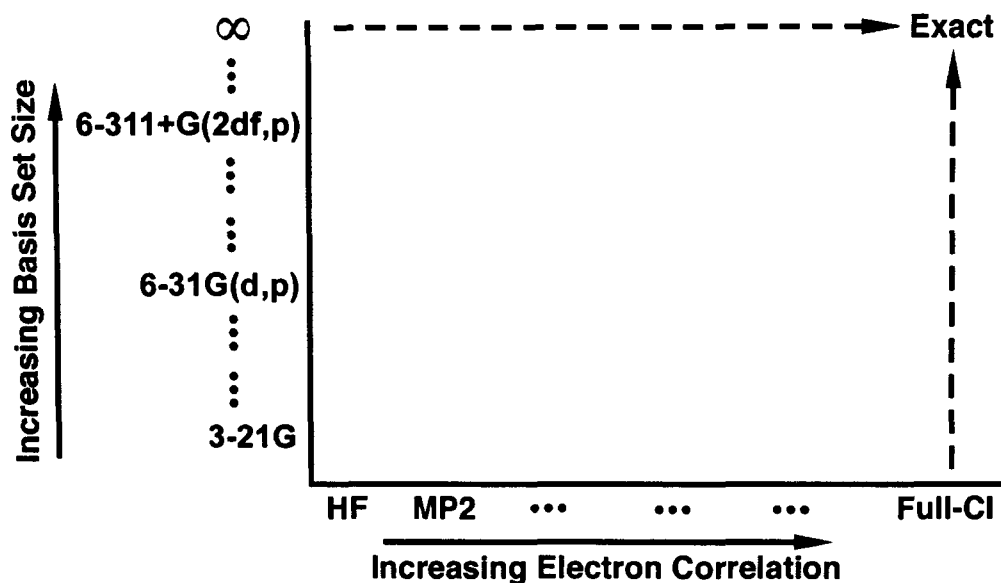


Figure 2.3 Illustration of the hierarchy of computational methods.

2.13 Basis Set Superposition Error

The binding energy of a bimolecular complex (AB) can be obtained by calculating the energy difference between AB (E_{AB}) and the sum of its monomers A (E_A) and B (E_B).

$$\Delta E_{\text{binding}} = E_{AB}^{ab} - (E_A^a + E_B^b) \quad (2.28)$$

where the superscripts refer to the basis functions on the corresponding species. If an incomplete basis set is used, the complex AB contains more basis functions than both the monomers A and B. As a result, the complex is better described, as both the monomers in AB tend to ‘borrow’ the basis functions from each other. This thus results in an artificial lowering of the energy of AB, which in turn increases the binding energy. This error is known as the basis set superposition error (BSSE).³

A commonly used approach to correct for BSSE is the counterpoise (CP) method proposed by Boys and Bernardi.¹⁷ In this method, one of the monomers (such as B) in the complex is replaced by a set of ghost orbitals in order to determine the artificial lowering

of the other monomer's energy (such as A) in the presence of the basis functions of both A and B. Thus,

$$CP = (E_A^{ab} - E_A^a) + (E_B^{ab} - E_B^b) \quad (2.29)$$

where E_A^{ab} is the energy of A based on the optimized geometry in the complex with the basis functions of B present. Similarly, E_B^{ab} is the energy of B based on the optimized geometry in the complex with the basis functions of A present. In addition, E_A^a and E_B^b are the energies of A and B in the optimized complex geometry with only their own basis functions present.

As a result, the CP corrected binding energy can now be written as:

$$\Delta E_{\text{binding}}^{\text{CP}} = E_{\text{AB}}^{ab} - (E_A^a + E_B^b) - CP \quad (2.30)$$

2.14 Atoms In Molecules Theory

The Atoms In Molecules (AIM) method, developed by Bader, investigates the topology of the electron density. It provides a new tool to analyze chemical bonds. Detailed explanations of the AIM theory can be found in several reviews by Bader^{18,19} and some other excellent books.^{20,21} Only those related to this thesis are briefly discussed.

2.14.1 Topology of the Electron Density

In AIM, the topology of the electron density is investigated by analyzing its gradient vector $\nabla\rho(\mathbf{r})$. Within a molecule, there exist certain points where $\nabla\rho(\mathbf{r})$ equals zero.

$$\nabla\rho(\mathbf{r}) = \frac{\partial\rho}{\partial x}\mathbf{i} + \frac{\partial\rho}{\partial y}\mathbf{j} + \frac{\partial\rho}{\partial z}\mathbf{k} = \bar{\mathbf{0}} \quad (2.31)$$

These points are referred as the critical points and they can be a minimum, maximum or a saddle point.

To further determine the nature of the critical points, the second derivative of $\rho(\mathbf{r})$ at these critical points must be considered, which are usually written as a Hessian matrix:

$$H = \begin{pmatrix} \rho''_{xx} & \rho''_{xy} & \rho''_{xz} \\ \rho''_{yx} & \rho''_{yy} & \rho''_{yz} \\ \rho''_{zx} & \rho''_{zy} & \rho''_{zz} \end{pmatrix} \xrightarrow{\text{diagonalize}} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (2.32)$$

where λ_1 , λ_2 and λ_3 are the eigenvalues of the Hessian.

The Laplacian of the electron density $\nabla^2\rho(\mathbf{r})$ at the critical point is defined as the sum of the three eigenvalues of the Hessian.

$$\nabla^2\rho(\mathbf{r}) = \lambda_1 + \lambda_2 + \lambda_3 \quad (2.33)$$

Table 2.1 The Characterization, Description and Nature of the Four Types of Critical Points.

(r, s)	Description	Nature
(3, -3)	$\rho(\mathbf{r})$ is a local maximum.	Nucleus
(3, -1)	$\rho(\mathbf{r})$ is a saddle point: a maximum in the plane defined by the axes of the two negative eigenvalues and a minimum in the third axis perpendicular to the plane.	Bond critical point (BCP)
(3, +1)	$\rho(\mathbf{r})$ is a saddle point: a minimum in the plane defined by the axes of the two positive eigenvalues and a maximum in the third axis perpendicular to the plane.	Ring critical point (RCP)
(3, +3)	$\rho(\mathbf{r})$ is a local minimum.	Cage critical point (CCP)

The critical points are characterized by the rank r and signature s of the diagonalized matrix in (2.32). The former is the number of non-zero eigenvalues while the latter is the algebraic sum of the signs of eigenvalues. The critical points are usually expressed as (r, s) and there are four chemically important types of critical points, which are explained in Table 2.1.

2.14.2 Chemical Bonds in AIM

AIM offers a complementary yet powerful tool to characterize chemical bonds. Recall from above that the gradient of the electron density $\nabla\rho(\mathbf{r})$ is a vector, which points in the direction of steepest ascent of $\rho(\mathbf{r})$. A series of infinitesimal gradient vectors constitute a gradient path. If two gradient paths start from the same BCP and both end at a different nucleus, they constitute a bond path. In AIM, the existence of a bond path indicates a chemical interaction.

The electron density $\rho(\mathbf{r})$ and Laplacian of the electron density $\nabla^2\rho(\mathbf{r})$ at a BCP can be used to characterize the nature of a chemical bond.²² Closed-shell interactions, such as ionic bonds, hydrogen bonds and Van der Waals interactions, are associated with small values of $\rho(\mathbf{r})$ and positive $\nabla^2\rho(\mathbf{r})$. Open-shell interactions, such as covalent bonds, are associated with large values of $\rho(\mathbf{r})$ and negative $\nabla^2\rho(\mathbf{r})$. In addition, for hydrogen bonds the values of $\rho(\mathbf{r})$ and $\nabla^2\rho(\mathbf{r})$ at the BCP are typically in the range of 0.002–0.040 and 0.024–0.139 a.u., respectively.^{23,24} Furthermore, for similar type of bonds, $\rho(\mathbf{r})$ at the BCP can be used as a measure of the bond strength.

In this thesis, four parameters obtained from the AIM calculation are used:

- (1) The electron density $\rho(\mathbf{r})$ at BCP;
- (2) The Laplacian of the electron density $\nabla^2\rho(\mathbf{r})$ at BCP;

- (3) The ellipticity (ϵ) at BCP, which is defined as:

$$\epsilon = \frac{\lambda_1}{\lambda_2} - 1 \quad (2.34)$$

where λ_1 , and λ_2 are the eigenvalues of the Hessian in (2.32). A larger ellipticity indicates the structural instability of a bond;²⁵

- (4) The distance between BCP and its nearest RCP. A small distance indicates the structural instability of a bond.²⁵

2.15 Potential Energy Surface

As discussed in Section 2.3, based on the Born-Oppenheimer approximation, the energy of a system is a function of the nuclear coordinates. A surface that shows the relation between the energy and nuclear coordinates is called the potential energy surface (PES). An example of PES is shown in Figure 2.4.

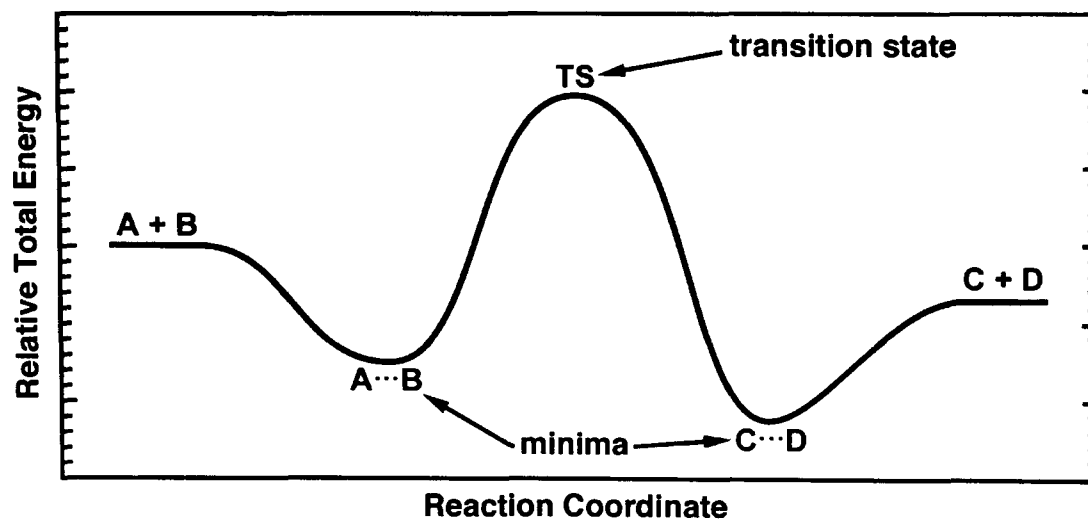


Figure 2.4 An example of a potential energy surface that shows the reaction between A and B to give C and D.

Generally, we are only interested in certain points that are mechanistically important. These include minima, i.e., equilibrium structures, and first-order saddle points, i.e., transition states. Hence, only these points are drawn on the PES.

2.16 Computational Tools

This section discusses a variety of computational tools to obtain the properties of a chemical system, such as structures and energies. Topological analysis has been discussed in **Section 2.14** and is omitted here.

2.16.1 Geometry Optimizations

Generally, the starting point to investigate the properties of a molecule is to find its lowest energy structure through geometry optimization. In the optimized structures, all the first derivatives of the energy with respect to the nuclei coordinates are zero. That is, there are no forces acting on the nuclei.

$$\frac{\partial E}{\partial r_i} = 0 \quad (2.35)$$

The optimized structures correspond to stationary points on PES. However, they can be either minima or saddle points. The nature of the stationary points is determined by frequency calculations.

2.16.2 Frequency Calculations

The second derivatives of the energy with respect to the nuclei coordinates are obtained from frequency calculations. This is particularly useful when determining the nature of the stationary points. Specifically, there are no imaginary frequencies for minima and only one imaginary frequency for transition states.

In addition, the zero-point vibrational energy (ZPVE) is also obtained from frequency calculations. The ZPVE corrections originate from the fact that even at 0 K the nuclei vibrate, which are associated with certain amount of energy. One can add ZPVE to the optimized electronic energy.

2.16.3 Single Point Calculations

Generally, a reasonable geometry can be obtained at a relatively lower level of theory with a moderate size of basis set. However, to obtain an accurate energy, a higher level theory with a larger basis set must be used. Thus, a common procedure is to run a single point energy calculation with a large basis set based on the geometry obtained from a smaller basis set. This is based on the assumption that the geometry at the large basis set is very similar as that obtained at the small basis set.

The above procedure is usually denoted as:

$$\text{Method 1/Basis Set 1//Method 2/Basis Set 2} \quad (2.36)$$

For example, one that is used in this thesis is:

$$\text{B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d,p)} \quad (2.37)$$

It indicates that the single point calculation is performed using the B3LYP method in conjunction with the 6-311+G(2df,p) basis set based on the geometry obtained using the B3LYP method in conjunction with the 6-31G(d,p) basis set.

2.16.4 Electrostatic Potential Surface

The charge distributions in a molecule can be visually represented in the form of an electrostatic potential surface (EPS). In an EPS, the electron rich and electron poor regions are shown in different colors, such as red and blue, respectively. An example of the EPS of water calculated at the B3LYP/6-311+G(d,p) level of theory is shown in **Figure 2.5**.



Figure 2.5 The electrostatic potential surface of water obtained at the B3LYP/6-311+G(d,p) level of theory. Red and blue regions indicate electron rich and electron poor regions, respectively. Oxygen and hydrogen atoms are shown as red and white, respectively.

2.17 Solvation

By default, calculations are performed in the gas phase. However, the chemical reactions studied in this thesis occur in the polar environment within enzymes. Thus, the gas phase calculations may not be able to accurately describe these reactions. An approach to take into account the general environment within enzymes is to use the implicit solvation model. In this model, the environment is represented by a continuum dielectric constant (**Figure 2.6**). Such an electric field is called the reaction field. The applicability of such an approach to study enzymatic reactions has been previously reviewed.^{7,8}

Two implicit solvation models are used in this thesis: the Integral Equation Formalism Polarizable Continuum Model (IEF-PCM) and Poisson-Boltzmann Self-Consistent Reaction Field (PB-SCRF) Model. Both of them put a series of interlocking spheres around each atom of the solute. The charge distribution of the solute polarizes the

dielectric continuum, which in turn polarizes the solute charge distribution.²⁶ However, they differ in how the sphere radii and point charges are chosen. In the IEF-PCM method, the sphere radii used in this thesis are determined from the United Atom Topological Model applied on atomic radii of the UFF force field while the point charges are obtained by fitting the electrostatic potential according to the Green function in the theory of integral equations.²⁷ In the PB-SCRF method, the sphere radii are chosen from the CHARMM and OPLS force field and parameterized against experimental data while the point charges are obtained by fitting the electrostatic potential according to the least-squares criterion.²⁸

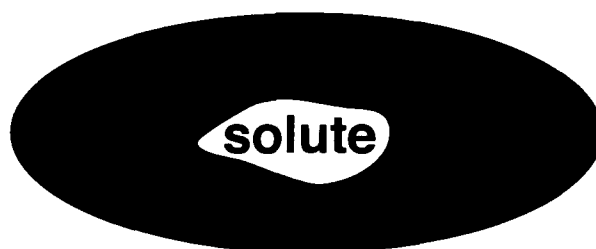


Figure 2.6 A schematic illustration of the implicit continuum solvation model.

2.18 Technical Aspects and Units

All the geometry optimizations, frequency calculations and single point calculations were performed using the Gaussian 03²⁹ and Jaguar 5.5³⁰ programs. Electrostatic potential surfaces were generated using GaussView.³¹ Topological analyses were performed using AIM2000.³²

The bond distances are reported in angstroms (Å) and bond angles are reported in degrees (°). The relative energies are expressed in kilojoules per mole (kJ mol⁻¹) using the following conversion factors from hartrees:

$$1 \text{ hartree} = 2625.5 \text{ kJ mol}^{-1}$$

References

- (1) Young, D. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*; John Wiley & Sons, Inc.: New York, 2001.
- (2) Levine, I. N. *Quantum Chemistry*, 4 ed.; Prentice-Hall, Inc.: New Jersey, 1991.
- (3) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons Ltd: England, 2002.
- (4) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: New York, 1989.
- (5) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons Ltd: West Sussex, England, 1999.
- (6) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.
- (7) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232.
- (8) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515.
- (9) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.
- (10) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524.
- (11) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
- (12) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (15) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.
- (16) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
- (17) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (18) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893.
- (19) Bader, R. F. W. *J. Phys. Chem. A* **2007**, *111*, 7966.
- (20) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1990.

- (21) Matta, C. F.; Boyd, R. J. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Wiley-VCH: Weinheim, 2007.
- (22) Bader, R. F. W.; Essén, H. *J. Chem. Phys.* **1984**, *80*, 1943.
- (23) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747
- (24) Popelier, P. L. A. *J. Phys. Chem. A* **1998**, *102*, 1873.
- (25) Cremer, D.; Kraka, E.; Slee, T. S.; Bader, R. F. W.; Lau, C. D. H.; Nguyen-Dang, T. T.; MacDougall, P. J. *J. Am. Chem. Soc.* **1983**, *105*, 5069.
- (26) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
- (27) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.
- (28) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
- (29) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.
- (30) Jaguar. 5.5; Schrodinger, L. L. C.: Portland, OR, 1991-2003.

-
- (31) GaussView. Dennington II, R., Keith, T., Millam, J., Eppinnett, K., Hovell, W. L., Gilliland, R., Eds.; Semichem, Inc.: Shawnee Mission, KS, 2003.
- (32) Biegler-König, F.; Schönbohm, J. AIM2000; Version 2.0; University of Applied Science: Bielefeld, Germany, 2002.

Chapter 3

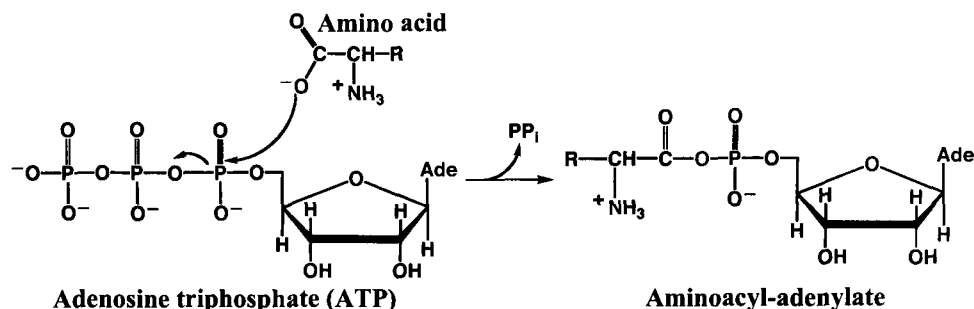
Substrate-Assisted Catalysis in the Aminoacyl Transfer Mechanism of Histidyl-tRNA Synthetase: A DFT Study

3.1 Introduction

Proteins play a range of diverse key roles in cells and organisms. For instance, catalytic proteins are essential for ensuring that reactions within cells occur at life-sustainable rates and often with great stereospecificity. Meanwhile others have, for example, important structural or signaling functions. As a result there is considerable interest in understanding all facets of their biochemistry. Within cells they are synthesized via a multi-step process. In the initial step the amino acid is 'activated' by its cognate transfer-RNA by forming the corresponding aminoacyl-tRNA. This reaction is catalyzed via two half-reactions by the family of enzymes known as the aminoacyl-tRNA synthetases (aaRS's).¹⁻³

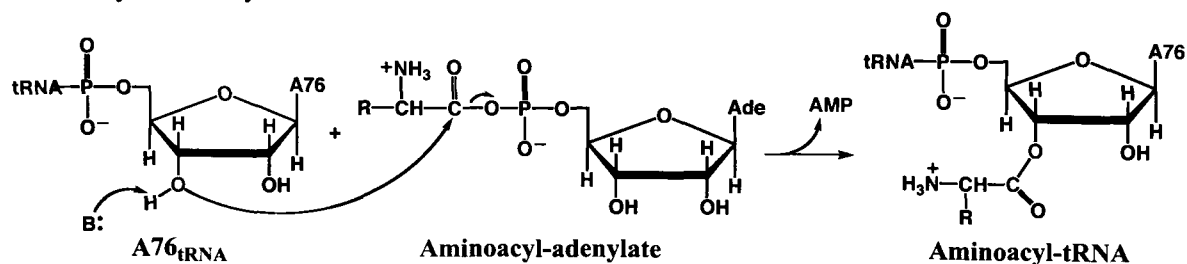
In the first half-reaction an adenosine triphosphate (ATP) and amino acid are bound by the appropriate aaRS. The latter then catalyzes their reaction to give the aminoacyl-adenylate and a pyrophosphate (PP_i) as shown in **Scheme 3.1**. The binding of these initial substrates to aaRS and the catalytic mechanism of this first half-reaction have been studied in detail both experimentally³⁻¹⁷ and theoretically¹⁸⁻²⁰

Scheme 3.1 General mechanism for the first half-reaction catalyzed by aminoacyl-tRNA synthetases.



In the second half-reaction, however, the same aaRS now catalyzes the transfer of the aminoacyl moiety from the aminoacyl-adenylate on to its cognate tRNA. Specifically, a base is thought to deprotonate the 2'-OH (class I)³ or 3'-OH (class II³ except PheRS^{21,22}) of the terminal adenosine (A76) of the tRNA, thus generating an R-O⁻ nucleophile. This alkoxy anion then attacks the aminoacyl-adenylate's α -carboxylate carbon (C_α) resulting in transfer of the aminoacyl moiety to the 2'- or 3'-O_{A76} with release of AMP (**Scheme 3.2**). Unfortunately, however, several important details of the mechanism of this second half-reaction remain unknown. In particular, the identity of the catalytic base (**B:** in **Scheme 3.2**) remains unclear as well as the nature of the reaction, i.e., whether it is concerted or step-wise.

Scheme 3.2 General mechanism for the second half-reaction catalyzed by class II aminoacyl-tRNA synthetases.

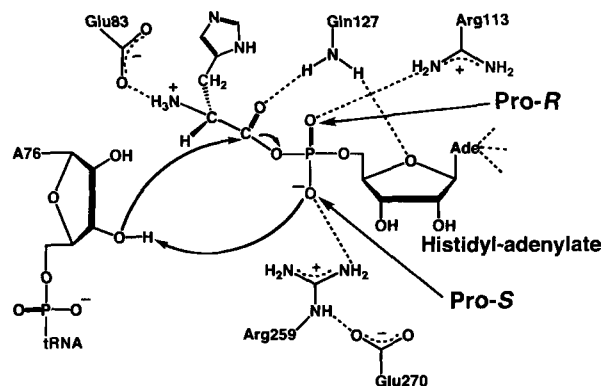


In an early study, Steitz and co-workers⁸ obtained crystal structures of the class I aaRS GlnRS with ATP and its cognate tRNA (tRNA^{Gln}) bound. They noted that in the active site the key reactive components of these substrates are within hydrogen bonding distance of each other. As a result, they suggested that a non-bridging phosphate oxygen of the glutamyl-adenylate substrate may be the base that deprotonates the 2'-OH group of A76. That is, GlnRS uses a substrate-assisted catalysis (SAC) mechanism, a class of reactions that are themselves a subject of increasing interest.²³ Later, Moras and co-workers²⁴ obtained crystal structures of several complexes of a class II AspRS with bound substrate and analogues. Subsequently, they²⁵ also obtained a crystal structure of an AspRS, tRNA^{Asp} and aspartyl-adenylate complex. From these structures they concluded that the 3'-OH of A76 is favorably positioned for proton transfer to a non-bridging phosphate oxygen. However, in each of these studies it remained unclear as to which of the two non-bridging phosphate oxygens was the general base.

More recently, Francklyn and co-workers²⁶ performed a detailed examination of a class II HisRS (**Scheme 3.3**). They first ruled out an active site glutamate residue (Glu83) as the catalytic base due in part to its ability to neutralize the substrate's histidine α -amino group. Rather, they concluded that its role was most likely in electrostatic binding of the substrate. However, they observed that when the substrate's non-bridging phosphate pro-*S* oxygen ($O_{\text{pro-}S}$) was replaced by sulfur, the aminoacyl transfer rate decreased by at least a factor of 10000. In addition, they also reexamined the available crystal structure of HisRS with histidyl-adenylate bound (PDB: 1KMM). They noted that $O_{\text{pro-}S}$ hydrogen bonds with an arginine (Arg259) that is itself hydrogen bonded to an adjacent glutamate (Glu270) as shown in **Scheme 3.3**. In contrast, the non-bridging phosphate pro-*R* oxygen ($O_{\text{pro-}R}$) hydrogen bonds with an arginine (Arg113) that is not hydrogen bonded to any other active site residues. Thus, they proposed that the

Arg259 \cdots O_{pro-S} bond may be weaker than Arg113 \cdots O_{pro-R} and therefore, O_{pro-S} may in fact be more basic and thus better able to act as the base.

Scheme 3.3 Proposed interactions and aminoacylation mechanism²⁶ as catalyzed by histidyl-tRNA synthetase.



With respect to the nature of the second half-reaction, from their work on GlnRS Steitz and co-workers^{8,27} proposed that it is concerted. That is, it proceeds via a “tetrahedral transition structure” in which formation of the 2'-O_{A76}—C_α bond is simultaneous with cleavage of the substrate's scissile C_α—O-P bond. First and co-workers^{28,29} used mutagenesis studies to investigate the roles of active site residues of TyrRS in transition structure stabilization. Based on their findings they also concluded that the reaction is concerted. In contrast, Moras and coworkers²⁵ proposed that the mechanism of AspRS involves the formation of a “tetrahedral intermediate”, implying that it proceeds in a step-wise manner. Furthermore, they suggested that in eubacteria a conserved active site arginine and glutamine are both involved in formation and stabilization of its oxyanion centre.

In the above same study by Francklyn and co-workers,²⁶ they further examined the role of the active site glutamine (Gln127) of HisRS. However, they found the reaction rate to be independent of hydrogen bonding between Gln127 and the substrate's α -

carboxylate oxygen (Scheme 3.3). Hence, they suggested that this did not support a step-wise associative mechanism (i.e., $3'\text{-O}_{A76}\text{—C}_\alpha$ bond formation occurs prior to cleavage of the $\text{C}_\alpha\text{—O-P}$ bond) as such an interaction would be expected to help stabilize the tetrahedral oxyanion intermediate. However, they also noted an apparent lack of residues to stabilize the increasing charges on the C_α and oxygen of the scissile $\text{C}_\alpha\text{—O-P}$ bond during cleavage. Hence, a step-wise dissociative mechanism was also concluded to be unlikely. Consequently, they proposed that the reaction follows a substrate-assisted concerted mechanism. In 2006, Perona and co-workers³⁰ performed a mutational analysis on GlnRS and concluded that it follows the same mechanism as proposed for HisRS. Indeed, it has been suggested that given the mechanistic use of a substrate's functional group and the dissimilar active sites of GlnRS and HisRS, a SAC mechanism may be common in tRNA synthetases.^{26,30}

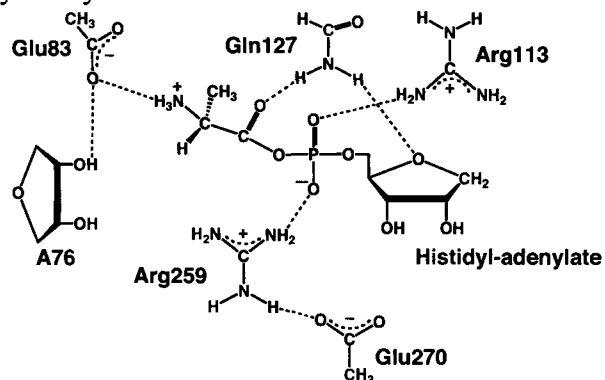
In order to gain insights into the role(s) of active site residues such as Gln127 and the mechanism of the second half-reaction as catalyzed by aaRS's, we have performed a density functional theory (DFT) study on HisRS. It was chosen as it has been experimentally studied in detail and its mechanism is thought to have broader general implications for aaRS's.²⁶

3.2 Computational Methods

All optimization and single point calculations were performed using Gaussian 03.³¹ Topological analysis of the electron densities at selected bond critical points were performed using AIM2000.³² Optimized gas-phase structures were obtained using the B3LYP functional³³⁻³⁵ in conjunction with the 6-31G(d,p) (5D) basis set. Frequency calculations were also performed in order to confirm the nature of the stationary points and obtain zero-point vibrational energy (ZPVE) corrections. Relative energies within the

enzyme's environment were obtained by performing single point calculations at the B3LYP/6-311+G(2df,p) level using the integral equation formalism polarized continuum solvation model (IEFPCM)³⁶ with a dielectric constant (ϵ) of 4.0 on the above optimized geometries and ZPVE corrected, i.e., IEFPCM-B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d,p) (5D) + ZPVE.

Scheme 3.4 The active site model used to investigate possible mechanisms for the second half-reaction as catalyzed by HisRS. Atoms in blue were held fixed during optimizations.



The active site model for HisRS was derived from the crystal structure PDB: 1KMM and is shown in **Scheme 3.4**. The key amino acid residues were modeled as follows: Glu83 and Glu270 as acetates, Arg259 and Arg113 as guanidinium cations, and Gln127 as a formamide. The imidazole and adenine of the histidyl-adenylate substrate were replaced by hydrogens. There is no crystal structure available for the tRNA^{His}-HisRS-histidyl-adenylate complex or an analogue. Hence, A76 of tRNA^{His} was modeled as a ribose sugar and positioned in the active site as suggested by experimental studies on HisRS^{26,37-39} and the available tRNA^{Asp}-AspRS-aspartyl-adenylate complex (PDB: 1C0A).²⁵ To maintain the integrity of the active site, selected atoms were held fixed at their crystal structure positions and are highlighted in blue in **Scheme 3.4**. As the histidyl-adenylate's adenine moiety hydrogen bonds and stacks with a number of residues, the substrate's C1' centre was also held fixed. The applicability and limitations of the above

computational methods and approach for investigating enzymatic reactions have been previously reviewed in detail.^{40,41}

3.3 Results and Discussion

3.3.1 Active Site with Bound Substrate

The optimized structure obtained for the complex comprising the initial substrates bound within the active site (1) is shown in **Figure 3.1**.

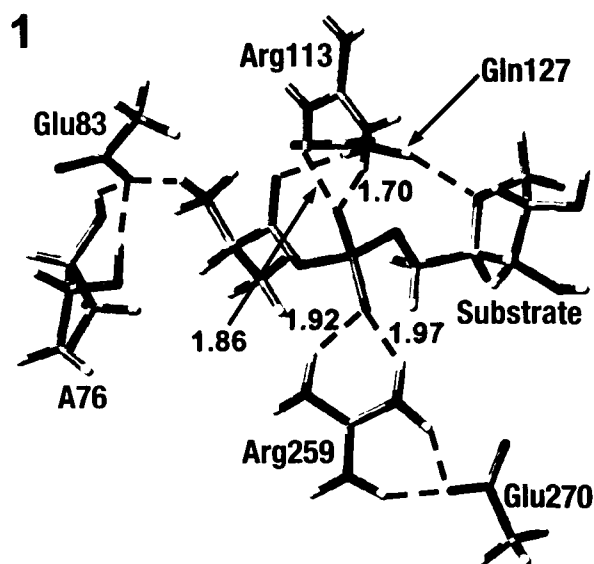


Figure 3.1 Optimized structure of the active site with bound substrate. Distances shown are in angstroms. Atom color key: C (gray); O (red); N (blue); P (orange) and H (white).

Upon binding, the substrate's $O_{\text{pro-R}}$ forms two moderately short hydrogen bonds (1.70 and 1.86 Å) with the HisRS's Arg113 residue, one to each of its terminal $-\text{NH}_2$ groups. Similarly, the substrate's $O_{\text{pro-S}}$ moiety forms two hydrogen bonds with Arg259, one to each of its terminal $-\text{NH}_2$ groups. However, they are decidedly longer interactions at 1.92 and 1.97 Å. This is likely due to the fact that Arg259 also forms two short hydrogen bonds (1.73 and 1.67 Å) via a terminal $-\text{NH}_2$ and its $-\text{NH}-$ group with the

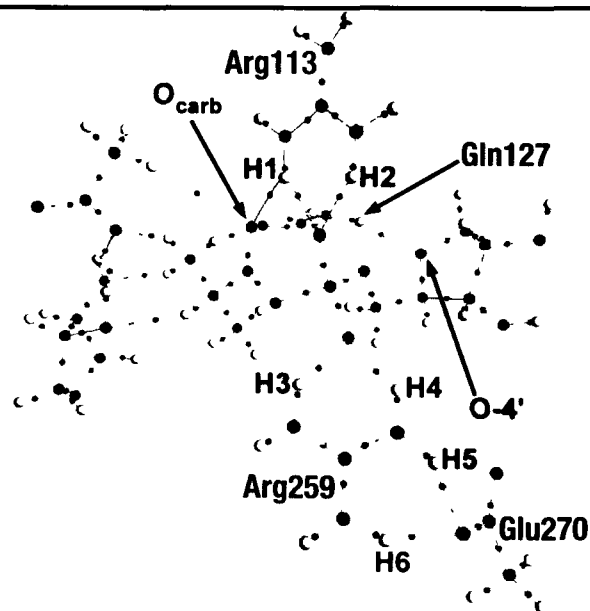
carboxyl of Glu270. Thus, as previously suggested,²⁶ Glu270 appears to weaken the Arg259...O_{pro-S} hydrogen bonds compared to the Arg113...O_{pro-R} bonds.

In addition, it is noted that the side-chain -NH₂ of Gln127 hydrogen bonds with both the histidyl-adenylate's α -carboxylate oxygen (O_{carb}) and 4'-oxygen with distances of 2.12 and 2.07 Å respectively. Furthermore, the side-chain carboxyl of Glu83 hydrogen bonds with the substrate's α -amino -NH₃⁺ group and the 2'- and 3'-OH's of tRNA's A76 sugar. Thus, Glu83 may play a role in positioning the A76 sugar as suggested by previous experimental studies.^{26,37-39} Interestingly, however, the resulting 3'-OH...O_{pro-R} and 3'-OH...O_{pro-S} distances are 6.16 and 6.75 Å respectively. That is, the more tightly bound O_{pro-R} is considerably closer to the A76 proton to be abstracted by 0.59 Å.

It has been previously proposed²⁶ that the basicities of O_{pro-R} and O_{pro-S} may be determined by the strengths of the hydrogen bonds in which they are involved. Topological analysis is a powerful complementary tool for characterizing interactions between atoms. In Bader's Atoms In Molecules (AIM) theory,⁴² the presence of a bond critical point (BCP) and bond path connecting two nuclei indicates an interaction. In particular, the electron density (ρ) at the BCP can be used as a measurement of its strength while the Laplacian of the electron density ($\nabla^2\rho$) reflects its nature. It has been stated that for hydrogen bonds the values of ρ and $\nabla^2\rho$ are typically in the ranges of 0.002–0.040 and 0.024–0.139 a.u., respectively^{43,44}

Table 3.1 Values of the Electron Density (ρ) and its Laplacian ($\nabla^2\rho$) (a.u.) at BCP's of Selected Hydrogen Bonds Between the Bound Substrate and Active Site in **1**.^a

BCP	ρ	$\nabla^2\rho$
Arg113-H1 \cdots O _{pro-R}	0.027	0.102
Arg113-H2 \cdots O _{pro-R}	0.047	0.137
Arg259-H3 \cdots O _{pro-S}	0.029	0.083
Arg259-H4 \cdots O _{pro-S}	0.022	0.082
Arg259-H5 \cdots O-Glu270	0.053	0.133
Arg259-H6 \cdots O-Glu270	0.039	0.134
Gln127-NH ₂ \cdots O _{carb}	0.017	0.054
Gln127-NH ₂ \cdots O-4'	0.020	0.061



^a Red points indicate BCP's. Color key: C (gray); O (red); N (blue); P (maroon) and H (white).

The values of ρ and $\nabla^2\rho$ at the BCP's of selected hydrogen bonds within **1** are listed in **Table 3.1**. As can be seen, all $\nabla^2\rho$ values lie within the above noted range typical for such interactions. Importantly, the ρ values of the Arg113-H1 \cdots O_{pro-R} and Arg113-H2 \cdots O_{pro-R} interactions are 0.027 and 0.047 a.u. respectively. Indeed, the latter value is in

fact higher than the range proposed for typical hydrogen bonds. In contrast, the ρ values for the Arg259-H3 \cdots O_{pro-S} and Arg259-H4 \cdots O_{pro-S} interactions are 0.022 and 0.029 a.u. respectively. Notably, the former value is only slightly greater than that of the much longer and neutral Gln127-NH₂ \cdots O-4' interaction. In addition, the *lowest* ρ value for either Arg113 \cdots O_{pro-R} bonds is approximately as large as the *highest* ρ value of either of the Arg259 \cdots O_{pro-S} interactions. Hence, consistent with the above structural data, the Arg113 \cdots O_{pro-R} bonds are predicted to be on average stronger due in large part to the fact that Arg259 also hydrogen bonds with the anionic Glu270 residue.

Although we were unable to directly calculate the proton affinities (PA's) of O_{pro-R} and O_{pro-S} using the large active site model, they could be obtained using a model consisting of Arg113, Arg259 and Glu270 with the substrate modeled as (CH₃O)₂PO₂⁻ (Table A.2 in Appendix A). It should be noted that the optimized bond distances, ρ and $\nabla^2\rho$ values of this smaller model showed the same trends as observed for the larger model (Table A.2 in Appendix A). The calculated PA's of O_{pro-R} and O_{pro-S} are 1074.5 and 1084.6 kJ mol⁻¹ respectively. Thus, while it is predicted initially to be further away from the 3'-OH proton, O_{pro-S} is indeed slightly more basic than O_{pro-R}.

3.3.2 Direct Cleavage: Proton Transfer to O_b

In addition to O_{pro-R} and O_{pro-S}, the 3'-OH proton of A76 may also transfer to the bridging phosphate oxygen (O_b) of the scissile C_α—O_bP bond. This can be considered as a 'direct cleavage' while those involving 3'-OH proton transfers to O_{pro-R} or O_{pro-S} are substrate-assisted. Examination of the former can help provide insights into any advantages or enhancements gained by employing a SAC mechanism, hence all three reaction possibilities were considered. The optimized structures and potential energy surface (PES) obtained for the direct cleavage mechanism are given in **Figures 3.2** and **3.3** respectively.

It was found that transfer of the 3'-OH proton to O_b occurs simultaneously with attack of the 3'-OH oxygen at the C_α centre of the histidyl-adenylate substrate. This proceeds via the four-membered ring transition structure (TS1) at a considerable cost of $191.7 \text{ kJ mol}^{-1}$. It is noted that previous studies on related reactions have found four-membered ring TS's to be generally associated with relatively high barriers.⁴⁵⁻⁴⁷ Interestingly, in TS1 the 3'-O—H bond has lengthened to 1.06 \AA while the corresponding 3'-OH $\cdots O_b$ distance is still comparatively long at 1.44 \AA . Hence, the proton is still much more closely associated with the 3'-oxygen than O_b . In contrast, the 3'-O $\cdots C_\alpha$ distance is relatively short at just 1.59 \AA while the C_α — O_b bond has lengthened dramatically from 1.33 \AA in **1** to 1.93 \AA in TS1 and thus seems to be almost cleaved. This suggests that the reaction is driven at least initially by formation and cleavage of the 3'-O— C_α and C_α — O_b bonds respectively rather than the proton transfer which instead occurs late in the reaction. Intriguingly, the C_α = O_{carb} bond has shortened slightly from 1.22 \AA in **1** to 1.20 \AA in TS1 (Table A.1 in Appendix A). That is, it appears to have gained some positive charge and can be characterized as $O=C^+$. There is no formation of an oxyanion-type C_α — O_{carb}^- . We note that while there is no proton donation from an external source, this reaction has some similarities to the mechanism of acid catalyzed unimolecular ester hydrolysis⁴⁸ which similarly proceeds via formation of an acyl carbocation intermediate. In addition, the O_{carb} centre forms just one hydrogen bond with the enzyme via a terminal $-\text{NH}_2$ group of Arg113 with a modest length of 2.11 \AA . The Gln127 residue instead now forms hydrogen bonds via its side-chain $-\text{NH}_2$ with both the 4'-O and $O_{\text{pro-S}}$ centres of the substrate.

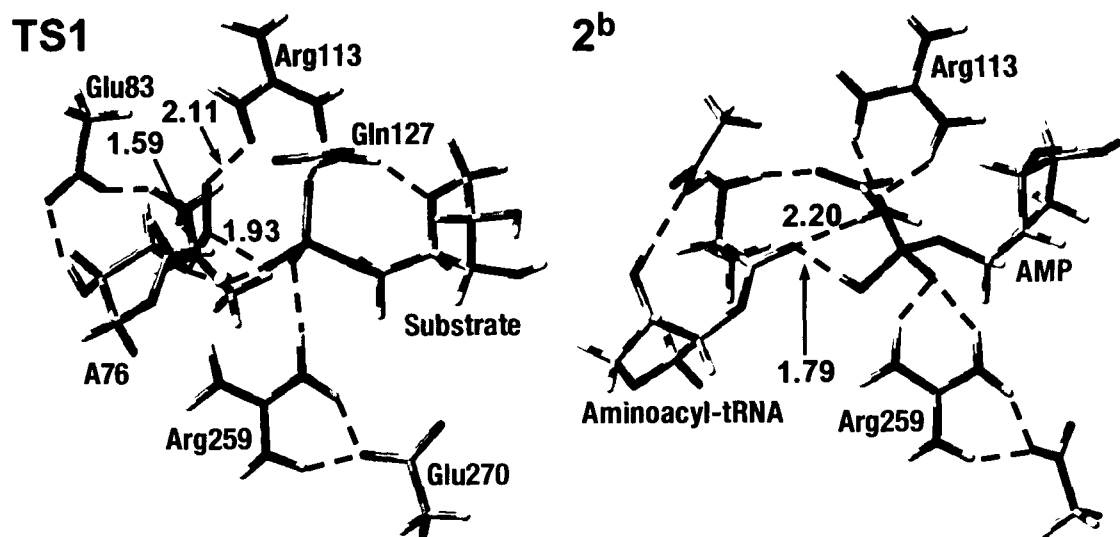


Figure 3.2 Optimized structures of the transition structure (**TS1**) and product complex (**2^b**) for aminoacyl transfer via direct cleavage of the C_{α} — O_bP bond. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white).

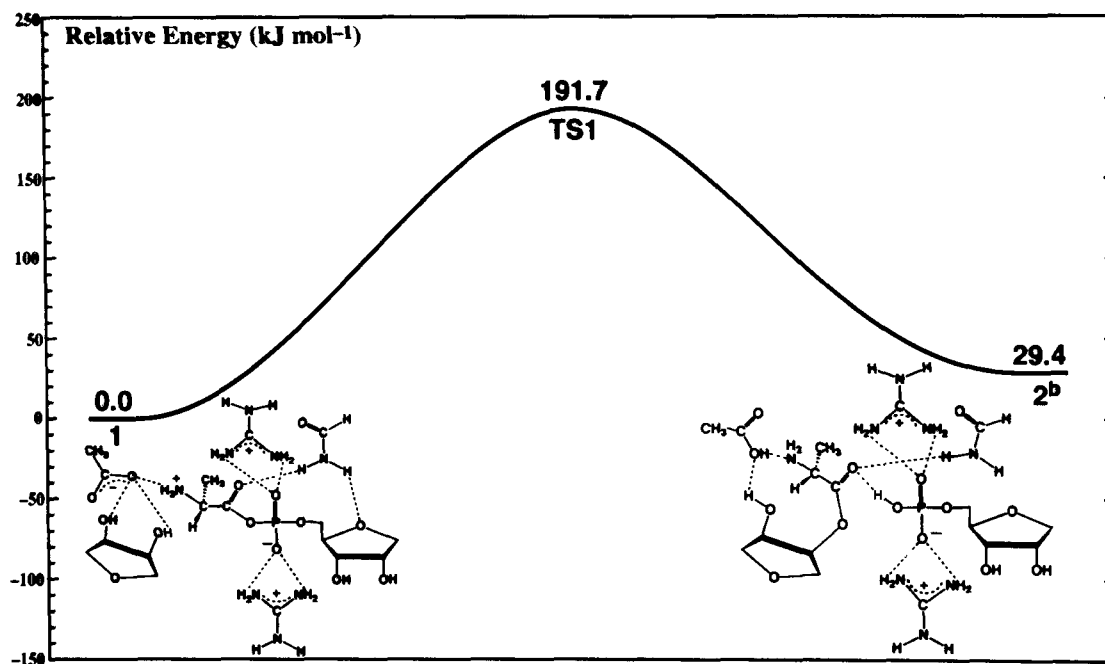


Figure 3.3 Schematic PES for aminoacyl transfer via direct cleavage of the C_{α} — O_bP bond. The Glu270 has been omitted for clarity.

The collapse of **TS1** directly leads to the formation of the product-bound complex **2^b** which lies 29.4 kJ mol^{-1} higher in energy than **1**. Thus, the ‘direct cleavage’ mechanism is a one-step or concerted reaction. In **2^b** the $\text{C}_\alpha\text{—O}_b$ bond is completely cleaved while O_b is now protonated and forms a moderately short hydrogen bond (1.79 \AA) with the O_{carb} centre. The two non-bridging phosphate oxygens $\text{O}_{\text{pro-R}}$ and $\text{O}_{\text{pro-S}}$ continue to form two hydrogen bonds with the nearby Arg113 and Arg259 residues respectively.

3.3.3 $\text{O}_{\text{pro-R}}$ as the Base

Possible aminoacyl transfer pathways in which the substrate's $\text{O}_{\text{pro-R}}$ moiety acts as a base and deprotonates the 3'-OH group were then considered. The optimized structures and PES obtained are given in **Figures 3.4** and **3.5** respectively.

Similar to the ‘direct cleavage’ mechanism, the lowest energy pathway involving $\text{O}_{\text{pro-R}}$ is also found to be a one-step or concerted reaction (**Figure 3.5**). Now, however, the 3'-OH proton transfers to $\text{O}_{\text{pro-R}}$ with concomitant nucleophilic attack of the 3'-OH oxygen at the substrate's C_α centre. This occurs via the six-membered ring transition structure **TS2** lying $145.8 \text{ kJ mol}^{-1}$ higher in energy than **1** (**Figure 3.5**). This is 45.9 kJ mol^{-1} lower in energy than the barrier for the direct cleavage mechanism. However, it is still higher than the generally held upper thermodynamic limit for an enzymatic reaction of approximately 120 kJ mol^{-1} ⁴⁹ In **TS2**, the 3'-O—H bond has significantly lengthened to 1.30 \AA while the corresponding 3'-OH... $\text{O}_{\text{pro-R}}$ distance has shortened considerably to 1.14 \AA . It is noted that in **TS1** the angle of the hydrogen bond of the proton transfer, $\angle \text{O}_3\text{HO}_b$, is just 127.5° while in **TS2** the corresponding $\angle \text{O}_3\text{HO}_{\text{pro-R}}$ is markedly closer to linearity at 159.5° . We note that in related systems it has been found^{50,51} that an increased linearity of a proton transfer is generally associated with a lower barrier.

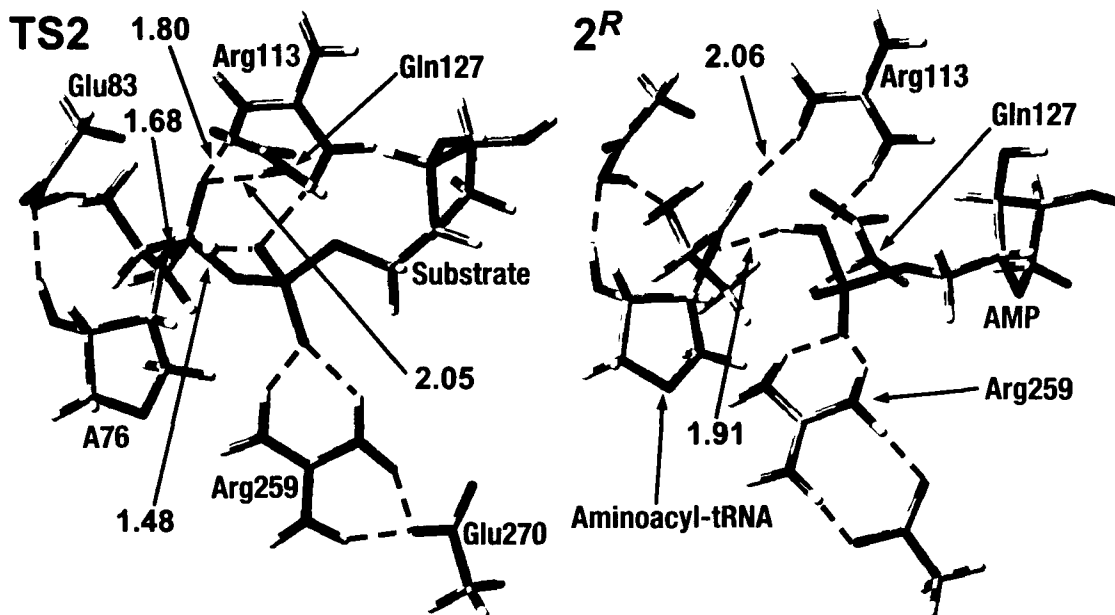


Figure 3.4 Optimized structures of the transition structure (TS2) and product complex (2^R) for aminoacyl transfer in which the substrate's non-bridging pro-*R* oxygen acts as the base. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white).

In TS2 the length of the forming $3'\text{-O}\cdots\text{C}_\alpha$ bond is 1.68 Å, 0.09 Å longer than in TS1. In addition, the cleaving $\text{C}_\alpha\text{—O}_b$ bond has lengthened significantly by 0.15 Å to 1.48 Å. However, this is still considerably less by 0.45 Å than observed in TS1. These differences are due in part to the fact that unlike in TS1, the $\text{C}_\alpha\text{—O}_{\text{carb}}$ bond now has reduced double bond character as illustrated by its lengthening from 1.22 Å in **1** to 1.26 Å in TS2 (Table A.1 in Appendix A). That is, the O_{carb} centre now has oxyanion character. In TS1 O_{carb} only formed one weak hydrogen bond with an active site residue, Arg113 (cf. Figure 3.2). In contrast, in TS2 it forms hydrogen bonds with the terminal side-chain —NH_2 groups of both Arg113 and Gln127 with distances of 1.80 and 2.05 Å respectively (Figure 3.4). These are notably shorter than their respective distances in **1** of 2.13 and 2.12 Å. This suggests that both Arg113 and, to a lesser extent, Gln127 play an important role in stabilizing the oxyanion in this reaction pathway.

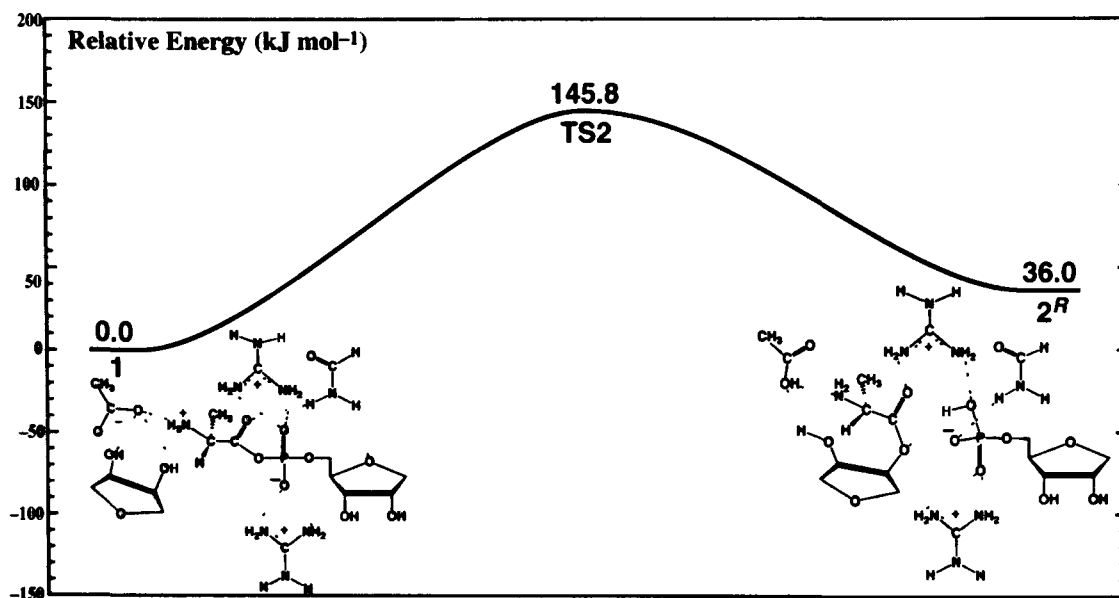


Figure 3.5 Schematic PES for the aminoacyl transfer reaction in which the substrate's non-bridging pro-*R* oxygen acts as the base. The Glu270 has been omitted for clarity.

The collapse of TS2 leads directly to the product-bound complex 2^R , which lies 36.0 kJ mol^{-1} higher in energy than **1**. The $C_\alpha\text{---}O_b$ bond is now completely cleaved while the newly formed $3'\text{-O---}C_\alpha$ bond has a length of 1.35 \AA . The $C_\alpha\text{---}O_{\text{carb}}$ bond has also recovered its double bond character with its length now back to 1.22 \AA . The O_{carb} centre still forms a hydrogen bond (2.06 \AA) with Arg113. However, Gln127 now forms a short hydrogen bond with the oxyanion centre O_b (1.81 \AA) instead of O_{carb} . In addition, the newly formed $\text{HO}_{\text{pro-}R}$ group hydrogen bonds (1.91 \AA) with the 3'-oxygen of A76.

3.3.4 $O_{\text{pro-}S}$ as the base

The optimized structures and corresponding PES obtained for the mechanism in which the substrate's $O_{\text{pro-}S}$ centre deprotonates the 3'-OH group during the aminoacyl transfer are given in **Figures 3.6** and **3.7** respectively.

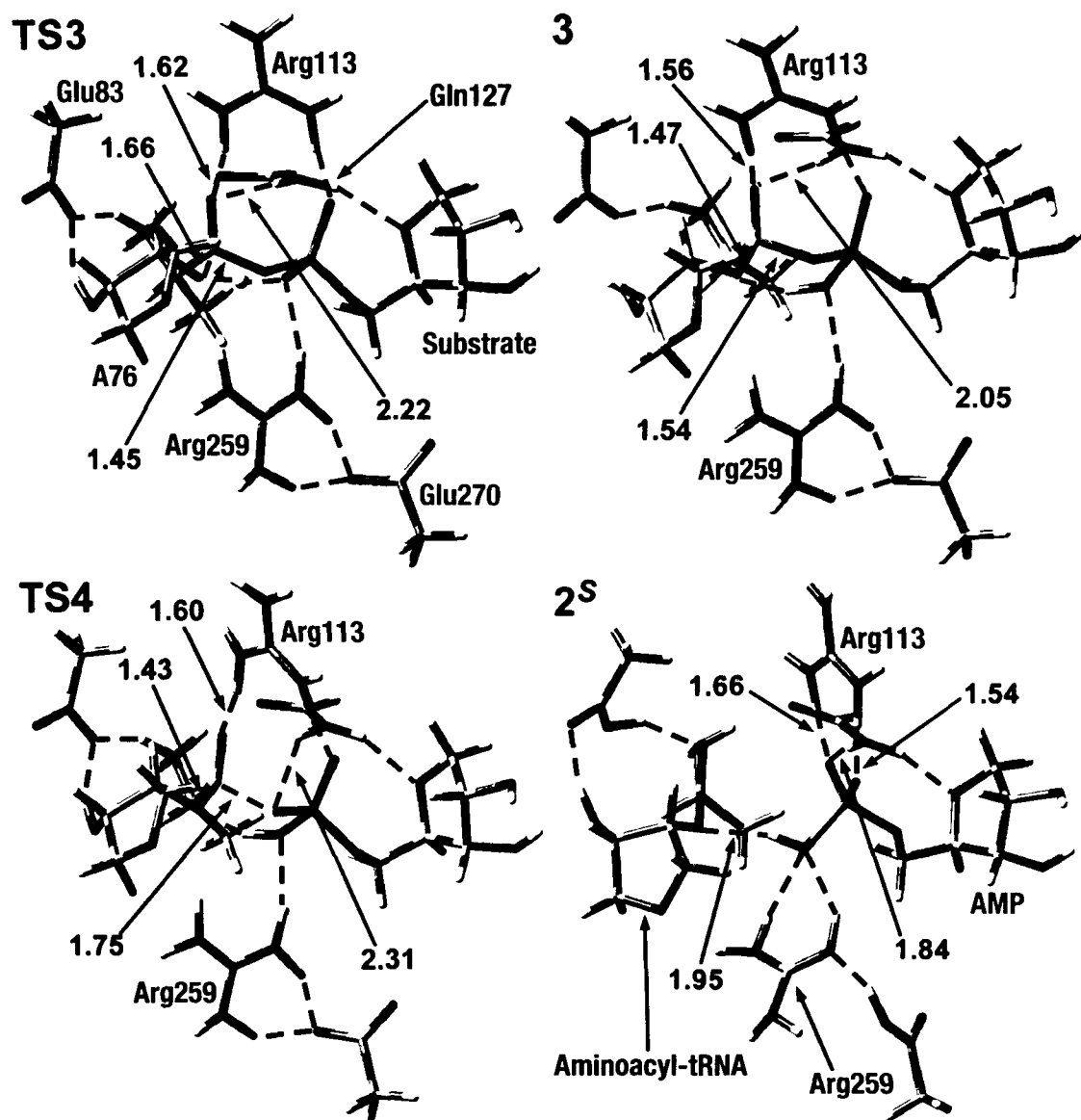


Figure 3.6 Optimized structures of the transition structures (TS3 and TS4), intermediate (3) and product complex (2^S) for aminoacyl transfer in which the substrate's non-bridging pro-S oxygen acts as the base. Distances shown are in angstroms. Color key: C (gray); O (red); N (blue); P (orange) and H (white).

Similar to the above mechanisms, transfer of the 3'-OH proton to O_{pro-S} occurs with concomitant attack of the 3'-oxygen at C_α. This proceeds via the six-membered ring transition structure TS3 lying 109.2 kJ mol⁻¹ higher in energy than 1 (**Figure 3.7**).

Notably, this is 36.6 kJ mol^{-1} lower in energy than the barrier obtained when $\text{O}_{\text{pro-R}}$ acted as the base that deprotonates the 3'-OH group (cf. **Figure 3.5**). Importantly, it is now decidedly lower than the generally held upper limit for an enzymatic mechanism (see above).⁴⁹

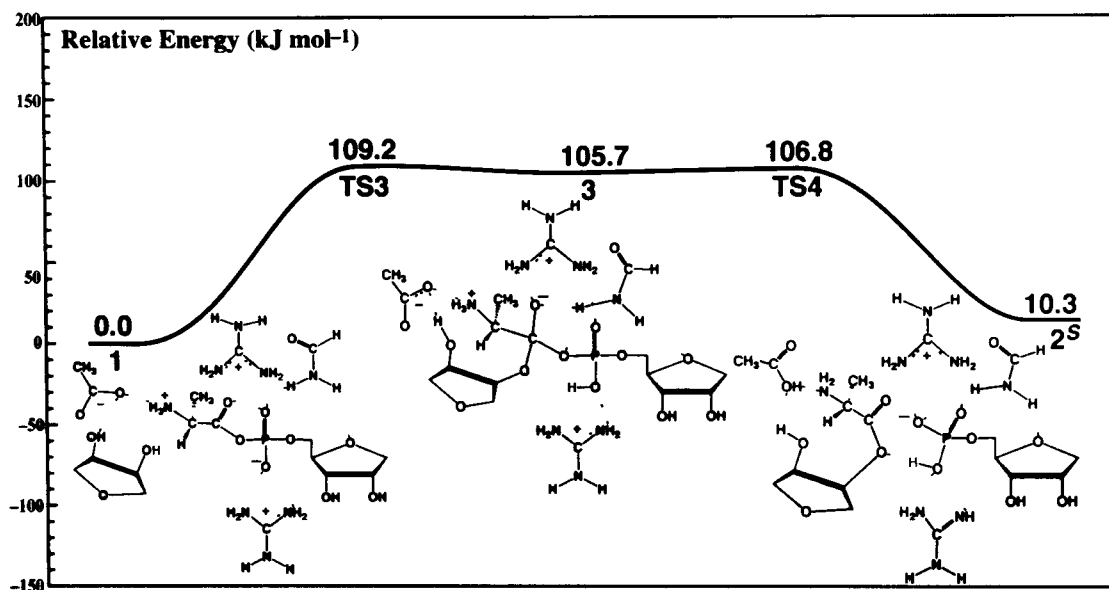


Figure 3.7 Schematic PES for the aminoacyl transfer reaction in which the substrate's non-bridging pro-*S* oxygen acts as the base. The Glu270 has been omitted for clarity.

Structurally, **TS3** has several key differences from the **TS**'s obtained above (**TS1** and **TS2**) for proton transfer from the 3'-OH group to one of the other substrate's phosphate oxygens. For instance, in **TS3** the $3'\text{-O}\cdots\text{H}$ and $3'\text{-OH}\cdots\text{O}_{\text{pro-S}}$ distances of the proton transfer from 3'-O to $\text{O}_{\text{pro-S}}$ are 1.23 and 1.19 Å respectively. These transfer distances are closest to equidistant of the three **TS**'s listed above. In addition, the proton transfer angle $\angle\text{O}_3\text{HO}_{\text{pro-S}}$ is 165.6° which is slightly closer to linearity than the corresponding $\angle\text{O}_3\text{HO}_{\text{pro-R}}$ angle in **TS2**. This may also contribute to **TS3** having a lower relative energy than **TS2**. Furthermore, the forming $3'\text{-O}\cdots\text{C}_\alpha$ and cleaving $\text{C}_\alpha\text{—O}_b$ bond distances are 1.66 and 1.45 Å respectively. Both of these are shorter by 0.02–0.03 Å than

the corresponding distances in **TS2**. This may be due in part to the fact that in **TS3** the $C_\alpha-O_{\text{carb}}$ bond has again elongated markedly though now to 1.27 Å, which is slightly longer than observed in **TS2**. The O_{carb} centre now forms a much shorter and stronger hydrogen bond (1.62 Å) with an $H_2N\text{-Arg113}$ group. In contrast, the $O_{\text{carb}}\cdots H_2N\text{-Gln127}$ distance has instead increased to 2.22 Å and is in fact longer than observed in the initial reactant complex **1** (2.12 Å). Thus, the oxyanion is now quite clearly stabilized primarily by the $O_{\text{carb}}\cdots H_2N\text{-Arg113}$ hydrogen bond.

The collapse of **TS3** leads to formation of the tetrahedral intermediate **3** lying just 3.5 kJ mol⁻¹ lower in energy than **TS3**, i.e., 105.7 kJ mol⁻¹ higher in energy than **1**. In **3** the proton has now fully transferred to $O_{\text{pro-S}}$ [$r(H\cdots O_{\text{pro-S}})=1.00$ Å] while the newly formed $3'\text{-O}-C_\alpha$ bond now resembles a standard C—O single bond as indicated by its length of 1.47 Å. As a result the $C_\alpha-O_b$ bond has elongated further to 1.54 Å. Significantly, however, it is not yet completely cleaved. Thus, aminoacyl transfer proceeds via a stepwise associative mechanism: the $3'\text{-O}-C_\alpha$ bond forms before cleavage of the $C_\alpha-O_b$ bond. Similar to that observed in **TS3**, O_{carb} forms two $O_{\text{carb}}\cdots H_2N\text{-}$ hydrogen bonds with Gln127 and Arg113 with lengths of 2.05 and 1.56 Å respectively. Both of these interactions are shorter than observed in **TS2** suggesting even greater stabilization of the oxyanion, again particularly by Arg113. This is also seen in the $C_\alpha-O_{\text{carb}}$ bond which has now lengthened further to 1.28 Å, indicating increased C—O⁻ character. We note that the effects of mutations at the Arg113 position were not investigated in the experimental study by Francklyn and co-workers.²⁶ However, such effects should perhaps be examined in light of the possible role of Arg113 in stabilizing the oxyanion intermediate.

The second and final step is cleavage of the $C_\alpha-O_b$ bond. This occurs via **TS4** at a cost of only 1.1 kJ mol⁻¹ with respect to **3** (**Figure 3.7**). Thus, it also lies 2.4 kJ mol⁻¹

lower in energy than TS3 indicating that the first-step is rate-determining. The second step's quite small barrier also suggests that it may not be kinetically or thermodynamically important and the reaction may experimentally appear concerted as observed by Francklyn and co-workers.²⁶ In TS4 the C_α—O_b bond has lengthened considerably to 1.75 Å with concomitant but more modest shortenings of both the 3'-O—C_α and C_α—O_{carb} bonds to 1.43 and 1.25 Å respectively. In addition, the O_{carb}···H₂N—Gln127 and O_{carb}···H₂N—Arg113 interactions that stabilized the oxyanion moiety of **3** have also lengthened to 2.32 and 1.60 Å respectively. In fact, the Gln127 amine group is now slightly closer to O_b than O_{carb} with a distance of 2.31 Å. Indeed, AIM calculations on TS4 give a bond path between the Gln127—NH₂ hydrogen and O_b centre with ρ and ∇²ρ values at its BCP of 0.013 and 0.044 a.u. respectively. These values are within the ranges proposed for hydrogen bonding interactions.^{43,44} Thus, while in the first step Gln127 appears to play a role in stabilizing the increasing negative charge on O_{carb} in TS3 and **3**, in the second step (TS4) it may help to stabilize the increasing negative charge on O_b as indicated in both the Mulliken and NBO charge models (Table 3.2).

The final active site-bound product complex **2^S** is directly formed from TS4 and lies just 10.3 kJ mol⁻¹ higher in energy than **1**. It is noted that **2^S** is the lowest relative energy product of the three aminoacyl transfer mechanisms considered and furthermore, that this is due solely to differences in their hydrogen bond networks (see below). In **2^S** the C_α···O_b distance is 3.44 Å, indicating it has fully cleaved, while both the 3'-O—C_α and C_α—O_{carb} bonds have concomitantly shortened markedly to 1.37 and 1.21 Å respectively (Table A.1 in Appendix A). Similar to that observed in the alternate product complexes **2^b** and **2^R**, the newly formed P—O_{pro-S}H moiety hydrogen bonds with the 3'-oxygen of A76 with a distance of 1.95 Å. In addition it continues to maintain two hydrogen bonds with Arg259. The O_b centre from the now cleaved C_α—O_b bond forms quite short hydrogen bonds with the terminal —NH₂ groups of Arg113 and Gln127 with distances of 1.66 and

1.84 Å respectively. The remaining phosphate 'pro-*R*' oxygen forms a quite short hydrogen bond (1.54 Å) with the other side-chain –NH₂ group of Arg113 (see **Figure 3.6**).

Table 3.2 Calculated^a Mulliken and NBO Charges On O_{carb} and O_b of the Scissile C_α—O_b bond For All Optimized Structures In This Chapter.

Proton Acceptor	Structure	Mulliken		NBO	
		O _{carb}	O _b	O _{carb}	O _b
O _b	1	-0.47	-0.55	-0.65	-0.81
	TS1	-0.43	-0.73	-0.62	-1.06
	2 ^b	-0.47	-0.61	-0.65	-1.05
O _{pro-R}	TS2	-0.60	-0.60	-0.82	-0.89
	2 ^R	-0.46	-0.66	-0.65	-1.15
O _{pro-S}	TS3	-0.60	-0.62	-0.83	-0.90
	3	-0.64	-0.63	-0.86	-0.93
	TS4	-0.57	-0.66	-0.79	-1.01
	2 ^S	-0.44	-0.71	-0.61	-1.17

^a B3LYP/6-31G(d,p) (5D)

3.3.5 Additional Comparisons of the Three Mechanisms

A number of differences in the aminoacyl transfer mechanisms with protonation of three phosphate oxygens O_b, O_{pro-R} and O_{pro-S} were described above. One of the most obvious is that concerted mechanisms are obtained when the 3'-OH proton transfers to O_b and O_{pro-R}. In contrast, an associative step-wise mechanism is obtained when O_{pro-S} acts as the base and furthermore, this pathway has the lowest barrier of all three obtained. An explanation for these differences in HisRS may be obtained by considering some key hydrogen bonding interactions and distances between active site residues and O_{carb} and O_b in the transition structures and intermediate. This is because as the 3'-oxygen of A76

attacks C_α these oxygen centers bear increasing negative charges, except for O_{carb} in **TS1** (see **Table 3.2**). Hence, the stabilization of these charges by the active site residues is an important factor that may determine the nature of the reaction.

For the direct cleavage mechanism no active-site residues are found to be able to stabilize the increasing negative charge on O_b in **TS1**. Indeed, this is the largest change in charge observed for either O_{carb} or O_b in any of the three mechanisms (see **Table 3.2**). The nearest hydrogen bond donor to O_b is the side-chain $-\text{NH}_2$ of Gln127 at a distance of 3.16 Å. Furthermore, four-membered ring transition states are known to be typically associated with higher barriers.⁴⁵⁻⁴⁷

When $O_{\text{pro-}R}$ acts as the base the O_{carb} centre of **TS2** hydrogen bonds with both the Arg113 and Gln127 residues with distances of 1.80 and 2.05 Å respectively. However, the nearest hydrogen bond donor to O_b is the side-chain $-\text{NH}_2$ of Gln127 at a distance of 2.77 Å. Thus, the increased negative charge on O_b is not being greatly stabilized. In contrast, when $O_{\text{pro-}S}$ acts as the base, while O_{carb} again interacts with both the Arg113 and Gln127 residues, the side-chain $-\text{NH}_2$ of Gln127 is now much closer to O_b . Indeed, in **TS3** and the resulting intermediate **3** the $r(\text{Gln127}-\text{NH}_2 \cdots O_b)$ distances are 2.52 and 2.56 Å respectively. These distances are considerably shorter by 0.25 and 0.21 Å than observed in **TS2**. Such an interaction further systematically shortens to 2.31 and 1.84 Å in **TS4** and the product complex **2^S** respectively (see **Figure 3.6**). However, we note that, as detailed above, these are not strong interactions, as is also reflected in the marginal stabilization of **3** relative to **TS3** and subsequent quite small barrier corresponding to **TS4**.

Lastly, as detailed above, A76 was manually positioned in the active site model used based on available experimental evidence from tRNA^{His} 26,37-39. Notably, it is in a

similar position in the related tRNA^{Asp}-AspRS-aspartyl-adenylate complex (PDB: 1C0A).²⁵ Should the crystal structure of the tRNA^{His}-HisRS-histidyl-adenylate complex become available, the exact position of A76 may be elucidated and provide additional new insights into its binding.

3.4 Conclusions

In this chapter, we have employed density functional theory methods to investigate the possible mechanisms of the second half-reaction catalyzed by histidyl-tRNA synthetase, in which the aminoacyl moiety at the histidyl-adenylate transfers to A76 of tRNA^{His}.

In the initial active site-bound substrate complex **1**, the two Arg113-NH₂···O_{pro-R} hydrogen bonds are found to be shorter than the two Arg259-NH₂···O_{pro-S} bonds, suggesting that the former interactions are stronger. This is supported by results from a topological analysis in which the ρ values at the BCP's of the two Arg113-NH₂···O_{pro-R} bonds are on average significantly larger, and thus stronger, than those of the two Arg259-NH₂···O_{pro-S} bonds. As a result, the substrate's O_{pro-S} centre is found to be slightly more basic than the other non-bridging phosphate oxygen O_{pro-R}.

The three possible mechanisms for aminoacyl transfer as catalyzed by HisRS were then considered. In each case the substrate's oxygen of the scissile C_a—O_b bond or one of its non-bridging (O_{pro-R} and O_{pro-S}) phosphate oxygens acted as the mechanistic base that deprotonates the key 3'-OH group of the A76_{tRNA} residue.

The lowest energy pathway was found to be an associative step-wise mechanism in which the substrate's non-bridging phosphate oxygen O_{pro-S} acts as the base. This is

concluded to be due to several reasons: (i) a higher-energy and more constrained four-membered ring transition structure (TS1) with a highly bent proton transfer is found when O_b acts as the base. In contrast, six-membered ring transition structures (TS2 and TS3) are found when $O_{\text{pro-R}}$ and $O_{\text{pro-S}}$ act as the base. The latter, in particular, involves the most linear path for the proton transfer. (ii) The $O_{\text{pro-S}}$ centre is found to be slightly more basic than $O_{\text{pro-R}}$ due to the fact that it hydrogen bonds more strongly with a nearby active site arginine residue. (iii) No active site residue is found to be able to stabilize the increasing negative charges on the cleaving $C_\alpha-O_b$ bond when O_b or $O_{\text{pro-R}}$ act as the mechanistic base. In contrast, such a role is played by Gln127 when $O_{\text{pro-S}}$ acting as a base in TS4. It is found, however, that this stabilization is only slight and thus experimentally the reaction may appear concerted.

References

- (1) Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D. *Nature* **1990**, *347*, 203.
- (2) Ibba, M.; Soll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617.
- (3) Arnez, J. G.; Moras, D. *Trends Biochem. Sci.* **1997**, *22*, 211.
- (4) Leatherbarrow, R. J.; Fersht, A. R.; Winter, G. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 7840.
- (5) Fersht, A. R. *Biochemistry* **1987**, *26*, 8031.
- (6) Leatherbarrow, R. J.; Fersht, A. R. *Biochemistry* **1987**, *26*, 8524.
- (7) Fersht, A. R. *Biochemistry* **1988**, *27*, 1577
- (8) Perona, J. J.; Rould, M. A.; Steitz, T. A. *Biochemistry* **1993**, *32*, 8758.
- (9) Belrhali, H.; Yaremchuk, A.; Tukalo, M.; Berthet-Colominas, C.; Rasmussen, B.; Bosecke, P.; Diat, O.; Cusack, S. *Structure* **1995**, *3*, 341.
- (10) Arnez, J. G.; Augustine, J. G.; Moras, D.; Francklyn, C. S. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 7144.

- (11) Desogus, G.; Todone, F.; Brick, P.; Onesti, S. *Biochemistry* **2000**, *39*, 8418.
- (12) Bovee, M. L.; Pierce, M. A.; Francklyn, C. S. *Biochemistry* **2003**, *42*, 15102.
- (13) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881.
- (14) Giegé, R.; Sissler, M.; Florentz, C. *Nucleic Acids Res.* **1998**, *26*, 5017–5035.
- (15) Onesti, S.; Desogus, G.; Brevet, A.; Chen, J.; Plateau, P.; Blanquet, S.; Brick, P. *Biochemistry* **2000**, *39*, 12853–12861.
- (16) Briand, C.; Poterszman, A.; Eiler, S.; Webster, G.; Thierry, J. C.; Moras, D. *J. Mol. Biol.* **2000**, *299*, 1051.
- (17) Retailleau, P.; Weinreb, V.; Hu, M.; Carter Jr, C. W. *J. Mol. Biol.* **2007**, *369*, 108.
- (18) Zurek, J.; Bowman, A. L.; Sokalski, W. A.; Mulholland, A. J. *Structural Chem.* **2004**, *15*, 405.
- (19) Kapustina, M.; Carter Jr, C. W. *J. Mol. Biol.* **2006**, *362*, 1159.
- (20) Kekenes-Huskey, P. M.; Vaidehi, N.; Floriano, W. B.; Goddard, W. A. *J. Phys. Chem. B* **2003**, *107*, 11549.
- (21) Fraser, T. H.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3044.
- (22) Sprinzl, M.; Cramer, F. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3049.
- (23) For reviews, see: (a) Dall'acqua, W.; Carter, P. *Protein Sci.* **2000**, *9*, 1. (b) Kosloff, M.; Selinger, Z. *Trends Biochem. Sci.* **2001**, *26*, 161 and references therein.
- (24) Cavarelli, J.; Eriani, G.; Rees, B.; Ruff, M.; Boeglin, M.; Mitschler, A.; Martin, F.; Gangloff, J.; Thierry, J. C.; Moras, D. *EMBO J.* **1994**, *13*, 327
- (25) Eiler, S.; Dock-Bregeon, A. C.; Moulinier, L.; Thierry, J. C.; Moras, D. *EMBO J.* **1999**, *18*, 6532.
- (26) Guth, E.; Connolly, S. H.; Bovee, M.; Francklyn, C. S. *Biochemistry* **2005**, *44*, 3785.
- (27) Rath, V. L.; Silvian, L. F.; Beijer, B.; Sproat, B. S.; Steitz, T. A. *Structure* **1998**, *6*, 439.
- (28) Xin, Y.; Li, W.; First, E. A. *J. Mol. Biol.* **2000**, *303*, 299.

- (29) Xin, Y.; Li, W.; Dwyer, D. S.; First, E. A. *J. Mol. Biol.* **2000**, *303*, 287.
- (30) Uter, N. T.; Perona, J. J. *Biochemistry* **2006**, *45*, 6858–6865.
- (31) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.
- (32) Biegler-König, F.; Schönbohm, J. AIM2000; Version 2.0; University of Applied Science: Bielefeld, Germany, 2002.
- (33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (34) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (35) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (36) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.
- (37) Hawko, S. A.; Francklyn, C. S. *Biochemistry* **2001**, *40*, 1930.
- (38) Connolly, S. A.; Rosen, A. E.; Musier-Forsyth, K.; Francklyn, C. S. *Biochemistry* **2004**, *43*, 962.
- (39) Guth, E. C.; Francklyn, C. S. *Mol. Cell* **2007**, *25*, 531.

- (40) For reviews, see: (a) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459. (b) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232. (c) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 115.
- (41) (a) Velichkova, P.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 16. (b) Hopmann, K. H.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 21299. (c) Hopmann, K. H.; Guo, J. D.; Himo, F. *Inorg. Chem.* **2007**, *46*, 4850. (d) Chen, S. L.; Marino, T.; Fang, W. H.; Russo, N.; Himo, F. *J. Phys. Chem. B* **2008**, *112*, 2494-2500.
- (42) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893.
- (43) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747
- (44) Popelier, P. L. A. *J. Phys. Chem. A* **1998**, *102*, 1873.
- (45) Chalmet, S.; Harb, W.; Ruiz-López, M. F. *J. Phys. Chem. A* **2001**, *105*, 11574.
- (46) Gorb, L.; Asensio, A.; Tuñón, I.; Ruiz-López, M. F. *Chem. Eur. J.* **2005**, *11*, 6743.
- (47) Qiao, Q.-A.; Yang, C.; Qu, R.; Jin, Y.; Wang, M.; Zhang, Z.; Xu, Q.; Yu, Z. *Biophys. Chem.* **2006**, *122*, 215.
- (48) (a) Sykes, P. *A Guidebook to Mechanisms in Organic Chemistry*, 6th edition; Longman Scientific and Technical: Singapore, 1986. pp 241-242. (b) Smith, M. B.; March, J. *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 6th edition; John Wiley & Sons, Inc.: Hoboken, NJ, 2007. pp 1402-1403.
- (49) (a) Lim, V. L.; Curran, J. F.; Garber, M. B. *J. Mol. Biol.* **2005**, *351*, 470. (b) Lim, V. L.; Kljashtorny, V. G. *Mol. Biol.* **2006**, *40*, 507.
- (50) Duan, X.; Scheiner, S. *J. Am. Chem. Soc.* **1992**, *114*, 5849.
- (51) Rangelov, M. A.; Vayssilov, G. N.; Yomtova, V. M.; Petkov, D. D. *J. Am. Chem. Soc.* **2006**, *128*, 4964.

Chapter 4

A DFT Study of Nucleobase Dealkylation by the DNA Repair Enzyme AlkB

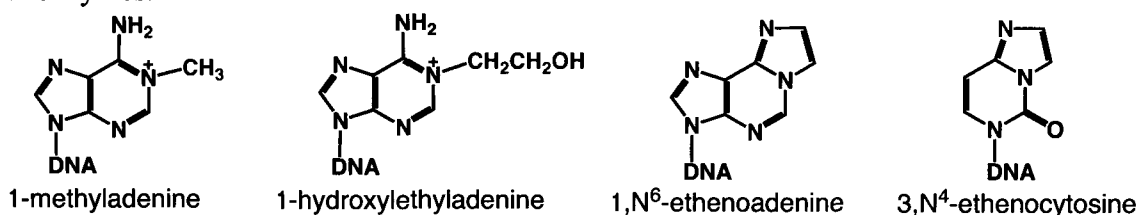
4.1 Introduction

The genetic information of cells is encoded within their DNA by way of the specific sequence of its nucleobase components. Thus, maintaining its structural integrity is essential for ensuring fidelity of this information. Unfortunately, many chemical and physical agents can react with the nucleobases and cause damage to them in the form of depurination, oxidation or deamination.¹⁻⁸ Alkylation is another particularly common damage-type and can be caused by either endogenous (e.g., *S*-adenosylmethionine)⁹ or exogenous (e.g., *N*-nitroso compounds of tobacco smoke) agents.¹⁰ Furthermore, it can occur at any of the oxygen or nitrogen atoms of the nucleobase and may involve more than one such site.¹¹

As a result, cells have developed several mechanisms to repair the damage caused by alkylation.^{12,13} One approach is to simply excise the affected nucleobase via use of enzymes known as DNA glycosylases. For example, alkyladenine DNA glycosylase has been found to excise 3-methyladenine, 1,*N*⁶-ethenoadenine and others via an acid-base catalytic mechanism.^{14,15} In another approach, repair enzymes remove the alkyl group itself by non-redox mechanisms. For example, *O*⁶-methylguanine-DNA methyltransferase catalyzes removal of the methyl from *O*⁶-methylguanine by transferring it to one of its

own active-site cysteine residues.¹⁶ Recently, however, a third strategy has been discovered in which the AlkB class of enzymes oxidatively dealkylate such damaged nucleobases. Furthermore, *E. coli* AlkB and its human homologues ABH2 and ABH3 have been found to be able to repair a variety of both methylated and larger alkylated nucleobases (Scheme 4.1).¹⁷⁻²⁰

Scheme 4.1 Examples of alkylated nucleobases known to be repaired by the AlkB class of enzymes.



The AlkB proteins have been included in the α -ketoglutarate–Fe(II)-dependent dioxygenase superfamily as the only class of enzymes with an oxidative dealkylation catalytic mechanism in this group.^{21,22} This is due to the fact that, as in other members of this superfamily, the Fe(II) ion is ligated in AlkB by a facial triad of residues consisting of two histidines (His131, His187) and one aspartate (Asp133).^{18,23} Secondly, AlkB requires Fe(II) and α -ketoglutarate as a cosubstrate for its catalytic activity.^{17,21,22} Finally, AlkB is able to activate dioxygen by forming a high-valency ferryl-oxo ($\text{Fe}^{\text{IV}}=\text{O}$) intermediate through conversion of α -ketoglutarate into succinate.^{24,25}

Crystal structures of *E. coli* AlkB complexed with substrate and both α -ketoglutarate and succinate (Figure 4.1) have given insight into the coordination geometry of the iron centre and the enzymes catalytic mechanism.²⁶ In particular, these structures provided an atomic level approximation to the enzyme-cosubstrate complex to which dioxygen binds (i.e., α -ketoglutarate-AlkB) and the oxidizing ferryl-oxo

intermediate complex (i.e., succinate-AlkB). In the case of the ϵ -ketoglutarate-ligated ferrous complex (**Figure 4.1a**), the Fe(II) ion was found to be monodentately and facially coordinated by Asp133, His187 and His131, and bidentately chelated by ϵ -ketoglutarate on its equatorial plane. In addition, a water is bound to the Fe(II) ion *trans* to the residue His131 along the reference axis. Hence, it was suggested that dioxygen should also bind at this position²⁶ as observed in related enzymes.^{27,28}

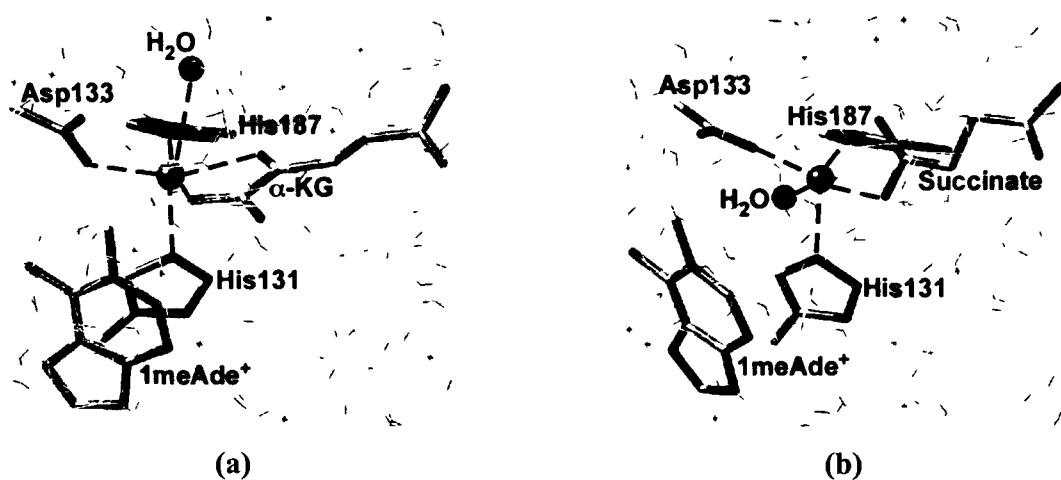


Figure 4.1 Active site of *E. coli* AlkB complexed with (a) ϵ -ketoglutarate (PDB accession code: 2FD8) and (b) succinate (PDB accession code: 2FDG). The 1-methyladenine fragment of the trinucleotide dT-(1-me-dA)-dT substrate is also displayed.

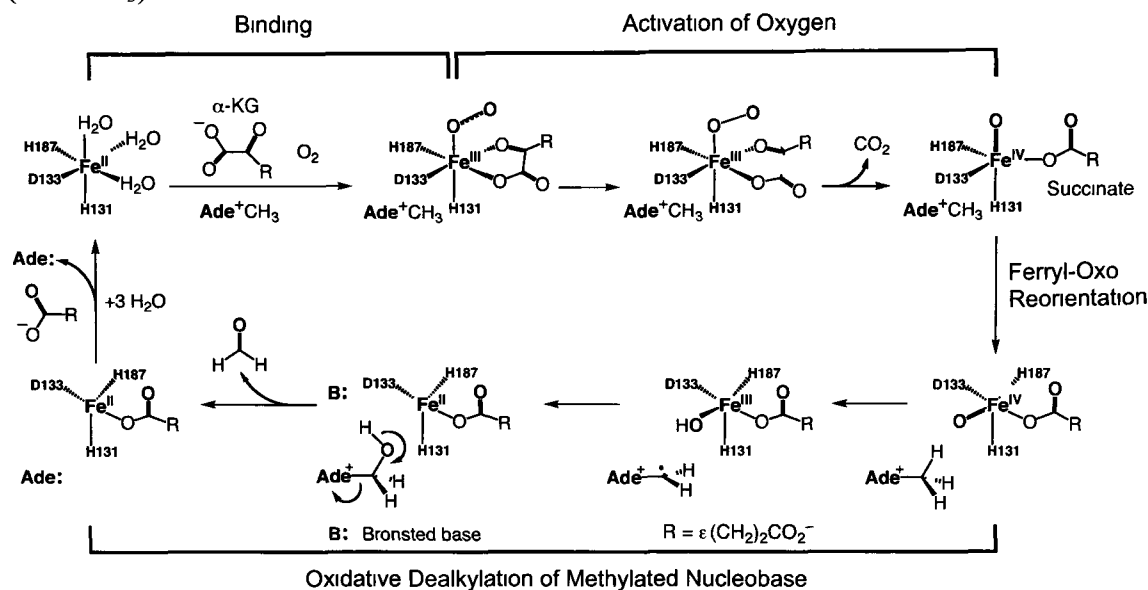
However, if the resulting catalytically active ferryl-oxo ($\text{Fe}^{\text{IV}}=\text{O}$) ligand lies along the major axis, the oxygen centre would be almost 6 Å from the target methyl carbon of the 1-methyladenine substrate. Hence, the spatial arrangement arising from the initial ϵ -ketoglutarate-ligated ferrous complex is unfavourable for the subsequent demethylation reaction. However, in the crystal structure of the succinate-ligated ferrous complex (**Figure 4.1b**) the Fe(II)-bound water is now *trans* to His187. In this orientation, the $\text{Fe}^{\text{IV}}=\text{O}$ oxygen would be approximately only 3 Å from the target methyl carbon of 1-methyladenine. Hence, it was proposed that after the activation of oxygen, the resulting

ferryl-oxo ligand reorients from the axial to an equatorial position in order to initiate the oxidative dealkylation of the substrate.²⁶ Previous studies have suggested that the reorientation of ferryl-oxo ligand may also occur in other enzymes of this superfamily, such as the clavamate synthase.²⁹

Based on all this information, the following catalytic mechanism of AlkB has been proposed (**Scheme 4.2**). In the first step, α -ketoglutarate chelates Fe(II) by displacing two water molecules from the equatorial plane. The bonding of dioxygen then displaces a third Fe(II)-bound water sitting *trans* to His131 along the long axis. The resulting ferrous-dioxygen $\text{Fe}^{\text{II}}\text{-O}_2^{2\cdot}$ moiety readily rearranges in a self-redox process to form the isoelectronic ferric-superoxido $\text{Fe}^{\text{III}}\text{-O}_2^{\cdot-}$ species.^{30,31} This latter anion then nucleophilically attacks the α -keto group of α -ketoglutarate. This step is followed by or concerted with decarboxylation of the α -ketoglutarate to form succinate with concomitant generation of a ferryl-oxo ($\text{Fe}^{\text{IV}}\text{=O}$) intermediate. The ferryl-oxo ligand then reorients from its initial axial position *trans* to His131 to the equatorial plane *trans* to His187. In this arrangement, it is able to abstract a hydrogen atom from the methyl group of the 1-methyladenine substrate. This is followed by a radical rebound process to yield 1-hydroxymethyladenine, which subsequently decomposes to form formaldehyde and the native adenine. Finally, the succinate is released, and the active site is regenerated.

Density functional theory (DFT) methods have been used previously to investigate the catalytic mechanisms of enzymes of the α -ketoglutarate-Fe(II)-dependent dioxygenase superfamily.³²⁻³⁶ However, detailed atomic level insights into the chemistry and catalytic mechanism of the AlkB class of enzymes remain unclear. In this chapter, we have used DFT methods to investigate the catalytic mechanism of AlkB. In particular, we have examined the processes of binding of dioxygen, oxygen activation, the ferryl-oxo ligand reorientation, and the oxidative demethylation of 1-methyladenine.

Scheme 4.2 Proposed dealkylation mechanism catalyzed by AlkB with 1-methyladenine (Ade^+CH_3) as the substrate.



4.2 Computational Methods

DFT calculations were performed with the B3LYP functional^{37 39} as available in the programs Gaussian 03⁴⁰ and Jaguar 5.5.⁴¹ Geometry optimizations were performed at the B3LYP/LACVP* level of theory including the Poisson-Boltzmann self-consistent reaction field (PB-SCRF) method implemented in Jaguar 5.5 with a dielectric constant of 4.0. This was done in order to take into account effects of the polar environment within the enzyme. All structures were characterized as minima or transition states on the potential energy surface (PES) by a gas-phase harmonic frequency analysis using Gaussian 03. Relative energies of all stationary structures along the reaction coordinate were obtained by single-point energy calculations at the PB-SCRF/B3LYP/LACV3P+** level of theory based on the above geometries.

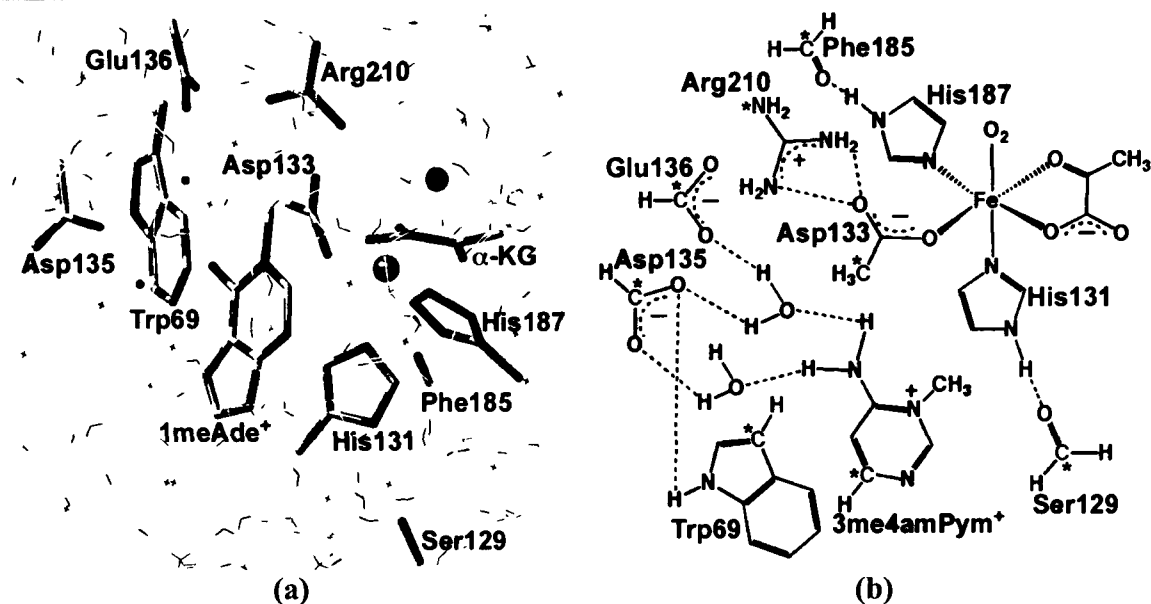


Figure 4.2 Structural models of the active site: (a) Arrangement of the catalytically active residues of AlkB from the crystal structure (PDB accession code: 2FD8). (b) Catalytic groups replacing the residues and 3me4amPym⁺ replacing 1meAde⁺ in the computational model. Atoms in red and with an asterisk were held fixed in geometry optimizations.

The coordinates of the groups belonging to the active site of AlkB were obtained from a crystal structure (PDB accession code: 2FD8).²⁶ This structure is an anaerobic complex of the enzyme and the single-stranded trinucleotide dT-(1-me-dA)-dT substrate. The key amino acid residues were replaced as follows (cf. **Figure 4.2**): His131 and His187 by two imidazole rings, Asp133 by acetate, Asp135 and Glu136 by two formate anions, Trp69 by indole, Arg210 by a guanidinium cation, and the backbone of Ser129 and Phe185 by two formaldehyde molecules. The substrate 1-methyladenine was modelled as the 3-methyl-4-amino pyrimidinyl cation (3me4amPym⁺) and the α -ketoglutarate (α -KG) cosubstrate was modelled as pyruvate. Dioxygen was bound to the Fe(II) ion *trans* to His131 along the reference axis. In the crystal structure, this position is occupied by a ligated water. Two water molecules that were hydrogen bonded to 1-methyladenine, Trp69, Asp135 and Glu136 in the crystal structure were also included. In

order to preserve the spatial arrangement of catalytic residues in the active site, selected atoms were held fixed at their positions in the crystal structure (Figure 4.2). The reliability and accuracy of these computational methods and protocols in the investigation of catalytic mechanisms of metalloenzymes have been tested and reviewed in depth elsewhere.⁴²⁻⁴⁶

4.3 Results and Discussion

4.3.1 Binding of Dioxygen to Iron in the Active-Site Complex

In AlkB, the Fe(II) ion is initially ligated to the enzyme via two histidines and an aspartate residue and to three water molecules in a nearly-octahedral arrangement. The binding of the cosubstrate α -KG to the iron centre displaces two of the equatorial water molecules while maintaining the overall octahedral arrangement. This latter complex is hereafter referred to as the α -KG–Fe(II) active-site complex. In related enzymes with the same Fe(II) coordination environments, it has been found that the Fe(II) is in the high-spin state with four unpaired, parallel spin electrons, which corresponds to the spin-quintet state ($S = 2$).^{47,48} The ligation of O₂ to the Fe(II) centre of AlkB displaces the remaining coordinated water molecule that lies along the reference axis.

There exist several possible O₂-bound active-site complexes as the dioxygen ligand may be linear or bent end-on monohapto-coordinated (η^1 -) or side-on dihapto-coordinated (η^2 -) to the iron centre. Furthermore, this binding may occur with or without electron transfer from the iron. Four overall spin combinations are thus possible for the dioxygen-bound α -KG–Fe(II) active-site complex: singlet, triplet, quintet and septet. It should be noted that the binding energies of O₂ to the iron cannot be accurately obtained from the active-site model as the coordination of O₂ is further stabilized by van der Waals

interactions and the cage effect of the protein environment by about 12–16 and 16–20 kJ mol⁻¹, respectively, according to ONIOM calculations in related enzymes.⁴⁹

Table 4.1 Relative Total Electronic Energies (kJ mol⁻¹), Selected Distances (Å), Bond Angles (°), and Spin Densities for the Iron-Dioxygen Complexes in Their Four Possible Spin-States.

Structure	ΔE	Distance			Angle	Spin density		
		Fe–O ₁	Fe–O ₂	O ₁ –O ₂	Fe–O ₁ –O ₂	O ₁	O ₂	Fe
⁷ [Fe ^{III} (η^1 -O ₂ ⁻)]	-12.5	2.37	3.24	1.23	125	0.85	0.89	3.98
⁷ [Fe ^{III} (η^2 -O ₂ ⁻)]	-7.0	2.33	2.40	1.25	77	0.86	0.83	3.99
³ [Fe ^{II} (η^1 -O ₂ ²⁺)]	0.0	2.30	3.15	1.22	124	-0.92	-0.99	3.78
⁵ [Fe ^{III} (η^1 -O ₂ ⁻)]	5.7	2.25	3.10	1.25	122	-0.16	-0.36	4.12
⁵ [Fe ^{III} (η^2 -O ₂ ⁻)]	12.4	2.19	2.25	1.28	76	-0.30	-0.24	4.14
¹ [Fe ^{II} (η^1 -O ₂ ²⁺)]	57.1	1.90	2.74	1.29	117	-0.51	-0.58	1.15

4.3.1.1 Binding of O₂ without Electron Transfer

Two dioxygen-bound active-site complexes were found to be formed without electron transfer: an overall spin-singlet, ¹[Fe^{II}(η^1 -O₂²⁺)] or ¹RC, consisting of a singlet-excited dioxygen (¹ Σ_g^+ , $S = 0$) bound to a low-spin singlet Fe(II) centre ($S = 0$), and an overall spin-triplet, ³[Fe^{II}(η^1 -O₂²⁺)] or ³RC, consisting of triplet dioxygen (³ Σ_g^- , $S = 1$) bound to a high-spin quintet Fe(II) centre ($S = 2$). In both overall states, the spin-density on the iron centre is typical for an Fe(II) ion (cf. Table 4.1).

Of these two, the lowest energy complex is the overall spin-triplet compound ³RC, which was therefore chosen to be the reference energy level throughout this chapter. In particular, ¹RC lies 57.1 kJ mol⁻¹ higher than ³RC, both compounds with dioxygen in the bent end-on binding mode (cf. Table 4.1). The relatively larger stability of the spin-triplet

over the spin-singlet ferrous-dioxygen compound could be due to a more favourable antiferromagnetic alignment of the unpaired electrons of $^3\text{O}_2$ ($^3\Sigma_g^-$) with the high-spin Fe(II) centre in ^3RC , when compared to $^1\text{O}_2$ ($^1\Sigma_g^+$) with the low-spin Fe(II) centre in ^1RC . We note that no η^2 -coordinated species were found for O_2 binding without charge transfer.

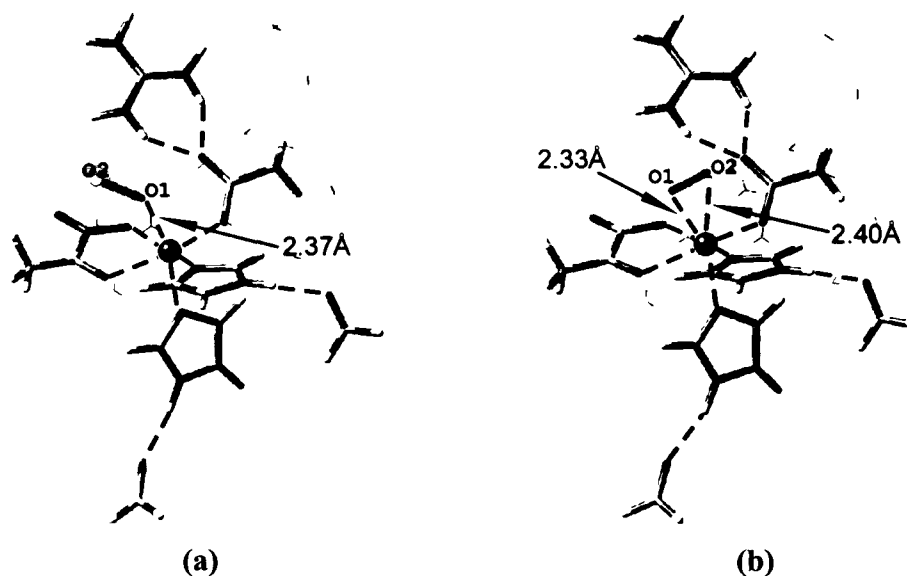


Figure 4.3 Optimized structures of the spin-septet active-site complexes of dioxygen: (a) Monohapto-coordinated dioxygen to iron $^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$ in a bent end-on binding mode, and (b) Dihapto-coordinated dioxygen to iron $^7[\text{Fe}^{\text{III}}(\eta^2\text{-O}_2^-)]$ in a side-on binding mode.

4.3.1.2 Binding of O_2 with Electron Transfer

Dioxygen binding in the overall quintet and septet spin-states was found to occur associated with a charge transfer that oxidized the iron to a high-spin sextet Fe(III) ion ($S = 5/2$) and reduced the $^3\text{O}_2$ (valence electron configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*1} \pi_z^{*1} \sigma_x^{*0}$ in a fragment orbital picture) to O_2^- (valence electron configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \varepsilon_y^{*2} \varepsilon_z^{*1} \sigma_x^{*0}$). The lowest energy complex found in which, overall, dioxygen binding occurs with charge transfer was the spin-septet η^1 -coordinated compound $^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$, also denoted ^7RC . The change in the formal oxidation state of the iron centre is indicated by the increase in

its spin density to 3.98. Importantly, ${}^7\text{RC}$ lies 12.5 kJ mol^{-1} lower in energy than the above ${}^3\text{RC}$ species.

As illustrated in **Figure 4.3**, the ${}^7\text{RC}$ species occurs as both a monohapto-, ${}^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$, and a dihapto-coordinated complex, ${}^7[\text{Fe}^{\text{III}}(\eta^2\text{-O}_2^-)]$, with the former lying 5.5 kJ mol^{-1} lower in energy (cf. **Table 4.1**). It is noted that the η^1 -coordinated superoxido ligand in ${}^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$ binds in both a bent and linear end-on mode with Fe–O1–O2 angles of 122° and 170° respectively. However, these two modes are found to differ in energy by just 1 kJ mol^{-1} , suggesting that the superoxido ligand may in fact have a broad range of angular mobility. The O–O bond both in η^1 - ${}^7\text{RC}$ and η^2 - ${}^7\text{RC}$ is elongated slightly (1.23 and 1.25 \AA , respectively) relative to ${}^3\text{O}_2$ but is shortened significantly relative to O_2^- . The shorter O–O bond of ${}^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$ could be related to an increase in the bond strength of the η^2 -coordinated O_2^- favored by a more effective ferromagnetic alignment of its unpaired electron with those of the sextet Fe(III) ion, thereby stabilized by electron exchange.

In contrast, the overall spin-quintet η^1 - and η^2 - O_2 -bound active-site complexes lie higher in energy than ${}^3\text{RC}$ by 5.7 and 12.4 kJ mol^{-1} respectively (**Table 1**). As for the overall spin-septet state, the bent η^1 -binding mode is more stable than the η^2 -binding mode by ca 6 kJ mol^{-1} . It is noted that the linear end-on binding mode was not a stationary point on the PB-SCRF/B3LYP/LACVP* potential energy surface. It is remarked that the O–O bond distance (1.28 \AA) in ${}^5[\text{Fe}^{\text{III}}(\eta^2\text{-O}_2^-)]$ is the arithmetic mean of those of isolated ${}^3\text{O}_2$ and O_2^- .

4.3.2 Activation of Oxygen Through Ferryl-Oxo $Fe^{IV}=O$ and Ferric-Oxyl $Fe^{III}-O^{\cdot}$ Intermediates

The potential energy surfaces obtained for activation of oxygen to form the highly reactive oxo- and oxyl-type intermediates (IC4) are shown in **Figure 4.4**. All relative energies are given with respect to the overall spin-triplet η^1-O_2 -bound active site complex 3RC . The calculated spin densities of the different oxidation states of iron have previously been found to generally have distinct ranges: Fe(II) = 3.7 – 3.8; Fe(III) = 4.0 – 4.2; Fe(IV) = 3.0 – 3.3.^{32,49-51} Thus, the iron centre's oxidation state in each complex was in part determined by calculating its spin density (**Table 4.2**).

In order to provide further insights into the electron transfer processes along the reaction paths, for each complex, a stylized molecular representation of the iron coordination environment is also shown in **Figure 4.4**. These structures are based on the geometric and spin density data given in **Table 4.2**. Furthermore, for clarity, the electron distribution of the *d*-subshell shown for each Fe centre is simplified to that of a 3*d*-block element in an idealized octahedral electrostatic field.

4.3.2.1 Spin-Quintet Activation Pathway

It has been previously demonstrated experimentally⁵²⁻⁵⁵ and theoretically^{32,56-58} that the ferric-superoxido group $Fe^{III}-O_2^{\cdot-}$ can act as a strong nucleophile. Indeed, $^5[Fe^{III}(\eta^1-O_2^{\cdot-})]$ -type complexes (denoted 5RC in **Figure 4.4**) have been found to be the most reactive. Hence, the first step of the current spin-quintet path is found to be nucleophilic attack of the $^5[Fe^{III}(\eta^1-O_2^{\cdot-})]$ group to the α -carbon (C2) of pyruvate. This occurs concomitantly with decarboxylation of the pyruvate and the reduction of both Fe(III) to Fe(II) and $O_2^{\cdot-}$ to O_2^{2-} , respectively. This reaction proceeds via 5TS1 with a barrier of 32.4 kJ mol⁻¹ to give the energetically low lying intermediate 5IC2 . In 5TS1 (**Figure 4.5**), the peroxy bond has lengthened to 1.38 Å while both O2–C2 and Fe–O1 distances have

shortened considerably to 1.44 and 2.01 Å, respectively. Moreover, the pyruvate's carboxyl group has been partially heterolytically cleaved as indicated by its increased C1-C2 distance of 1.78 Å and an O-C1-O angle of 136°

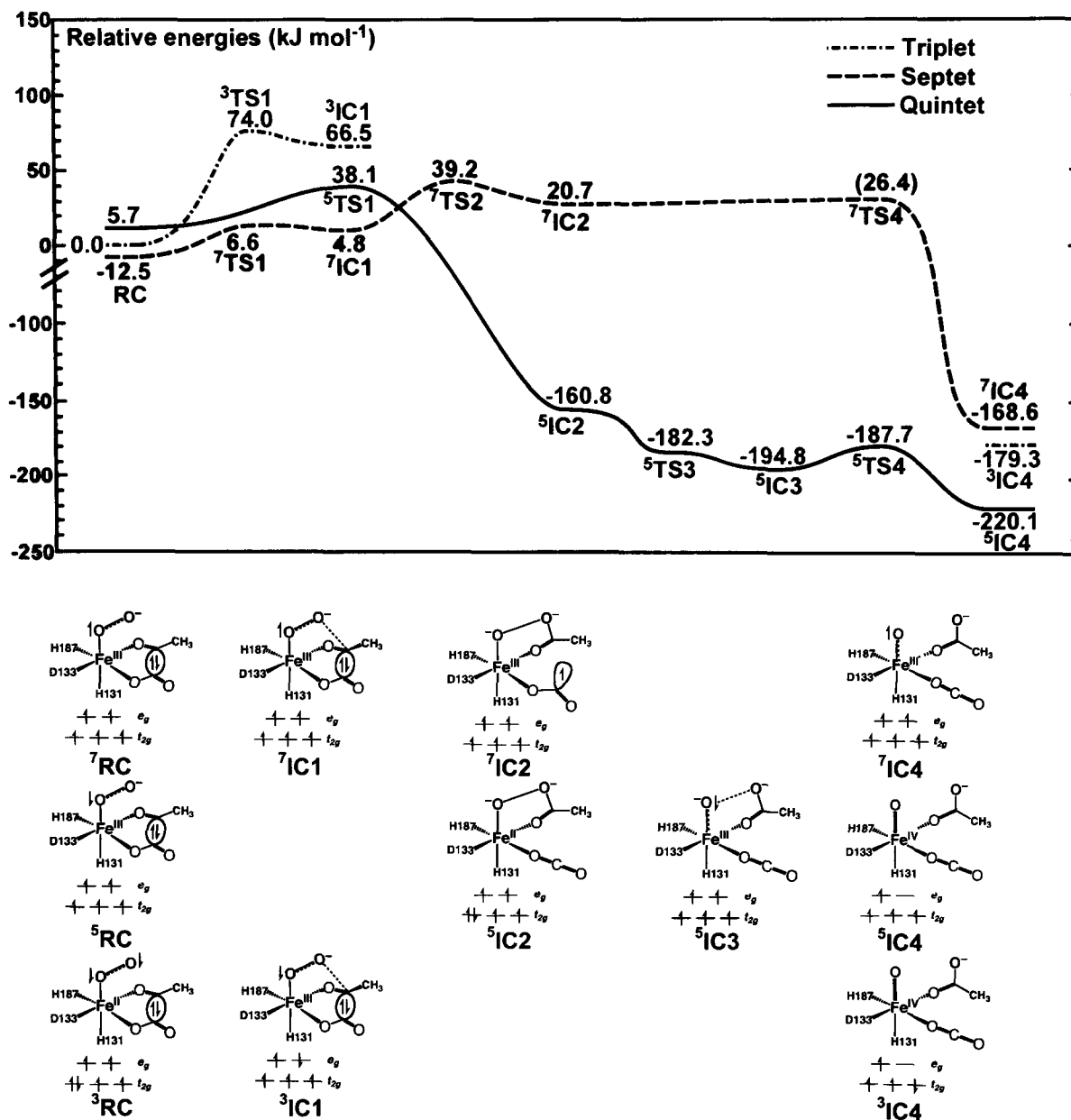


Figure 4.4 Potential energy surfaces for activation of oxygen through formation of the ferryl-oxo $\text{Fe}^{\text{IV}}=\text{O}$ and ferric-oxyl $\text{Fe}^{\text{III}}-\text{O}^\bullet$ compounds.

Table 4.2 Selected Distances (Å) and Spin Densities for the Complexes Involved in the Activation of Oxygen.

Structure	Distance (Å)				Spin density		
	Fe-O ₁	O ₁ -O ₂	O ₂ -C ₂	C ₁ -C ₂	O ₁	O ₂	Fe
Quintet							
⁵ RC – Fe(III)	2.25	1.25	2.83	1.55	-0.16	-0.36	4.12
⁵ TS1 – Fe(III)	2.01	1.38	1.44	1.78	0.09	-0.16	4.21
⁵ IC2 – Fe(II)	2.00	1.45	1.32	3.75	0.06	0.00	3.82
⁵ TS3 – Fe(III)	1.75	2.02	1.27	3.61	-0.09	-0.30	4.13
⁵ IC3 – Fe(III)	1.75	2.17	1.25	3.44	-0.21	-0.24	4.15
⁵ TS4 – Fe(III)	1.78	2.40	1.24	3.47	-0.39	-0.11	4.20
⁵ IC4 – Fe(IV)	1.65	2.71	1.22	3.60	0.72	0.04	3.11
Septet							
⁷ RC – Fe(III)	2.37	1.23	2.92	1.55	0.85	0.89	3.98
⁷ TS1 – Fe(III)	2.29	1.29	1.78	1.55	0.65	0.47	4.24
⁷ IC1 – Fe(III)	2.25	1.31	1.60	1.58	0.64	0.40	4.25
⁷ TS2 – Fe(III)	1.96	1.40	1.42	1.88	0.33	0.21	4.23
⁷ IC2 – Fe(III)	1.95	1.43	1.32	3.99	0.22	0.01	4.25
⁷ IC4 – Fe(III)	1.86	2.32	1.25	3.39	1.24	0.17	4.19
Triplet							
³ RC – Fe(II)	2.30	1.22	3.19	1.55	-0.92	-0.99	3.78
³ TS1 – Fe(III)	2.36	1.30	1.79	1.53	-0.55	-0.42	2.93
³ IC1 – Fe(III)	2.47	1.31	1.61	1.55	-0.60	-0.36	2.94
³ IC4 – Fe(IV)	1.59	2.72	1.23	3.56	0.49	0.02	1.52

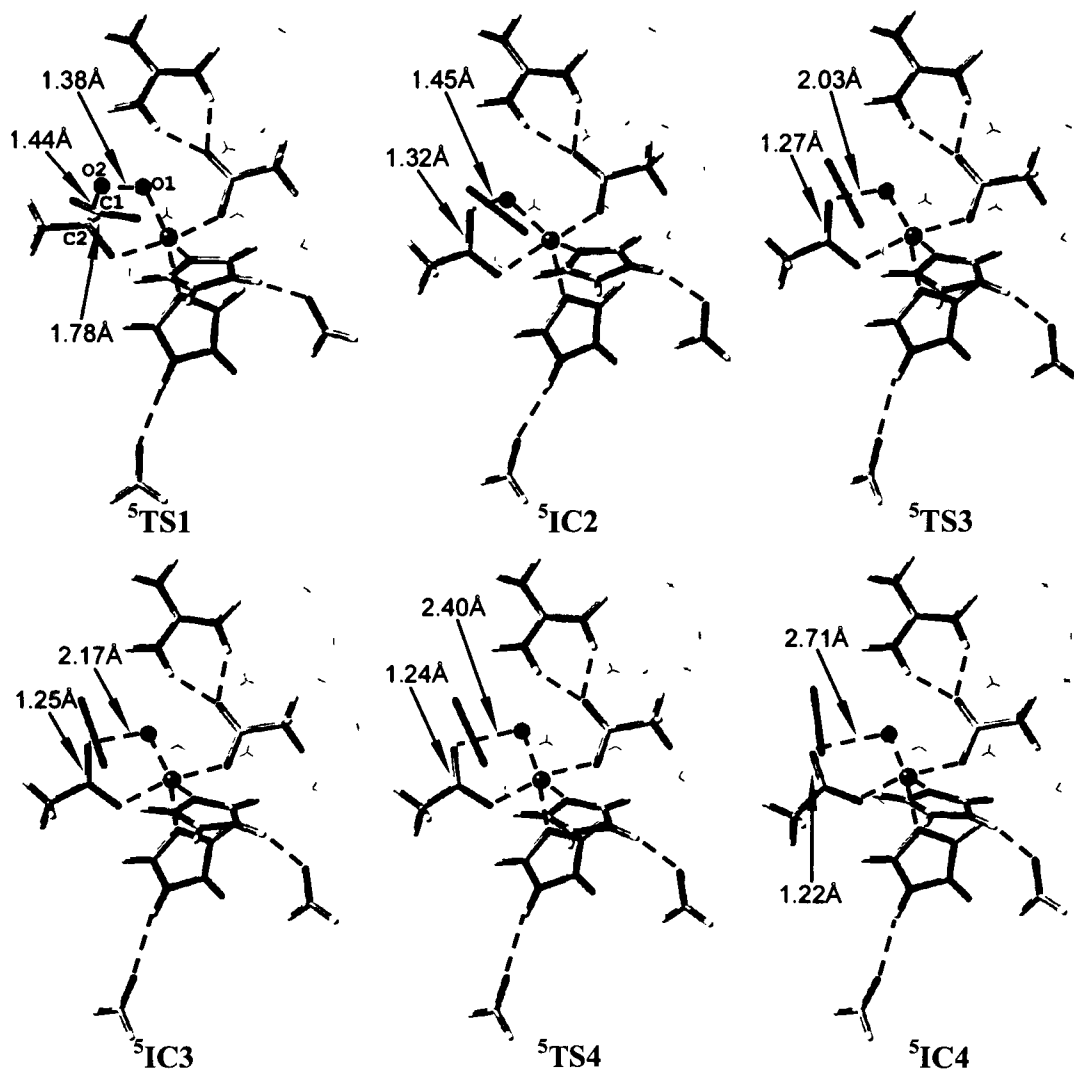


Figure 4.5 Optimized structures of the spin-quintet transition states (${}^5\text{TS1}$, ${}^5\text{TS3}$ and ${}^5\text{TS4}$), intermediates (${}^5\text{IC2}$ and ${}^5\text{IC3}$) and product complex (${}^5\text{IC4}$) for formation of the ferryl-oxo $\text{Fe}^{\text{IV}}=\text{O}$ compound.

The conversion of ${}^5\text{RC}$ into ${}^5\text{IC2}$ is significantly exoergic by $-166.5 \text{ kJ mol}^{-1}$. This is likely due in part to the extensive charge transfer from the cleaving C–C bond to both the peroxy unit and the Fe centre, in combination with reorganization of the ligands around the iron. The ferrous-peroxido intermediate ${}^5\text{IC2}$ is a complex in which both acetate and CO_2 are monodentately ligated to the Fe(II) centre (**Figure 4.5**). The peroxy-bridge has further lengthened to 1.45 Å while the C1–C2 bond is now fully cleaved with

a distance of 3.75 Å. In short, the formation of ${}^5\text{IC2}$ from ${}^5\text{RC}$ can be considered as the heterolytic cleavage of a C–C bond with one electron being promoted to reduce Fe(III) to Fe(II). The other electron is promoted to reduce the superoxido ligand to the peroxido unit. This is, in a fragment orbital picture, from O_2^- with electron configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*2} \pi_z^{*1} \sigma_x^{*0}$ to O_2^{2-} with electron configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*2} \pi_z^{*2} \sigma_x^{*0}$. These results are in line with those of previous computational studies on related enzymatic^{32,58} and biomimetic systems.⁵⁶

Subsequent conversion of the ferrous-peroxido intermediate ${}^5\text{IC2}$ into the ferryl-oxo compound ${}^5\text{IC4}$ is kinetically and thermodynamically favored, occurring via the stepwise transfer of two electrons from Fe(II) to the O–O unit. Specifically, ${}^5\text{IC2}$ readily undergoes a one-electron transfer via ${}^5\text{TS3}$ to give the ferric-oxyl intermediate ${}^5\text{IC3}$ (cf. **Figures 4.4** and **4.5**). That is, one electron from Fe(II) in ${}^5\text{IC2}$ is promoted to a $\sigma_{\text{O-O}}^*$ orbital, leaving the peroxy unit in the highly unstable electronic configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*2} \pi_z^{*2} \sigma_x^{*1}$. As a result, in ${}^5\text{IC3}$ the Fe–O1 distance has shortened to just 1.75 Å, while the O1–O2 distance has increased to 2.17 Å (**Table 4.2**). Thus, the $\sigma_{\text{O-O}}^*$ orbital in which the unpaired electron lies is chiefly localized on the $2p_x$ orbital of the coordinated oxygen O1.

In the last step, the ferric-oxyl intermediate ${}^5\text{IC3}$ undergoes a one-electron transfer via ${}^5\text{TS4}$ at a cost of 7.1 kJ mol⁻¹ to give the ferryl-oxo compound ${}^5\text{IC4}$ lying 25.3 kJ mol⁻¹ lower in energy. More specifically, the one-electron oxidation of the Fe(III) centre to Fe(IV) promotes the electron to the $\sigma_{\text{O-O}}^*$. This results in full scission of the O–O linkage since the O–O unit has now reached the highly unstable electron configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*2} \pi_z^{*2} \sigma_x^{*2}$. Intermediate ${}^5\text{IC4}$ is an octahedral complex in which the ferryl-oxo group [$r(\text{Fe–O}) = 1.65$ Å] lies along the reference axis, while both the acetate and CO_2 formed remain monodentately ligated to the now Fe(IV) centre (**Figure 4.5**).

Furthermore, for intermediate **IC4**, the quintet state is found to be significantly lower in energy than the corresponding triplet or septet spin-state complexes (**Figure 4.4**).

4.3.2.2 Spin-Septet Activation Pathway

Similar to the spin-quintet pathway, the first step of the spin-septet path is found to be a nucleophilic attack of the ferric-superoxido group to the pyruvate's α -carbon. Now, however, nucleophilic attack and decarboxylation of pyruvate occur via a stepwise mechanism. Attack of the ${}^7[\text{Fe}^{\text{III}}(\eta^1\text{-O}_2^-)]$ group in ${}^7\text{RC}$ to the pyruvate's α -carbon (C2) proceeds via ${}^7\text{TS1}$ with a barrier of 19.1 kJ mol^{-1} (**Figure 4.4**). The O2–C2 distance has now decreased significantly to 1.78 \AA (**Figure 4.6**) while concomitantly, the Fe–O1 bond has shortened markedly to 2.29 \AA and the O1–O2 bond has lengthened to 1.29 \AA (**Table 4.2**). The resulting intermediate complex formed, ${}^7\text{IC1}$, lies just 1.8 kJ mol^{-1} lower in energy than ${}^7\text{TS1}$ (**Figure 4.4**). Interestingly, in this complex the oxygen atoms form a peroxy bridge between Fe(III) and the pyruvate's C2 centre (**Figure 4.6**). Specifically, while the Fe–O1 and O2–C2 bonds have both markedly shortened further to 2.25 and 1.60 \AA , the O1–O2 bond has lengthened only slightly to 1.31 \AA . Perhaps more importantly, however, the pyruvate cosubstrate has not been decarboxylated as indicated by the fact that its C1–C2 bond has only increased to 1.58 \AA (cf. **Table 4.2**). It should be noted that, owing to their structural similarity, ${}^7\text{IC1}$ could be regarded as the reaction-coordinate equivalent of ${}^5\text{TS1}$.

The complex ${}^7\text{IC1}$ is then able to react further by homolytic cleavage of the pyruvate's C1–C2 bond. This reaction proceeds via ${}^7\text{TS2}$ with a barrier of 34.4 kJ mol^{-1} to give the ferric-peroxido intermediate ${}^7\text{IC2}$ lying 15.9 kJ mol^{-1} higher than ${}^7\text{IC1}$ (**Figure 4.4**). In ${}^7\text{IC2}$ the C1–C2 distance is now 3.99 \AA while the Fe–O1 and O2–C2 bonds have shortened to 1.95 and 1.32 \AA , respectively. This C1–C2 bond rupture produces a bent (bond angle of 130°) carbon-centered CO_2^- that remains monodentately

coordinated to the Fe(III) centre through one of its oxygens (**Figure 4.6**). Remarkably, while the O1 η O2 bond has lengthened to 1.43 Å, it is not yet cleaved (**Table 4.2**). It should be noted, however, that in a fragment orbital picture the second electron from the C–C bond rupture has been transferred to one of the π_{O-O}^* orbitals to leave the peroxy unit in the closed-shell electron configuration $\sigma_x^2 \varepsilon^2 \pi_z^2 \pi_y^{*2} \pi_z^{*2} \sigma_x^{*0}$

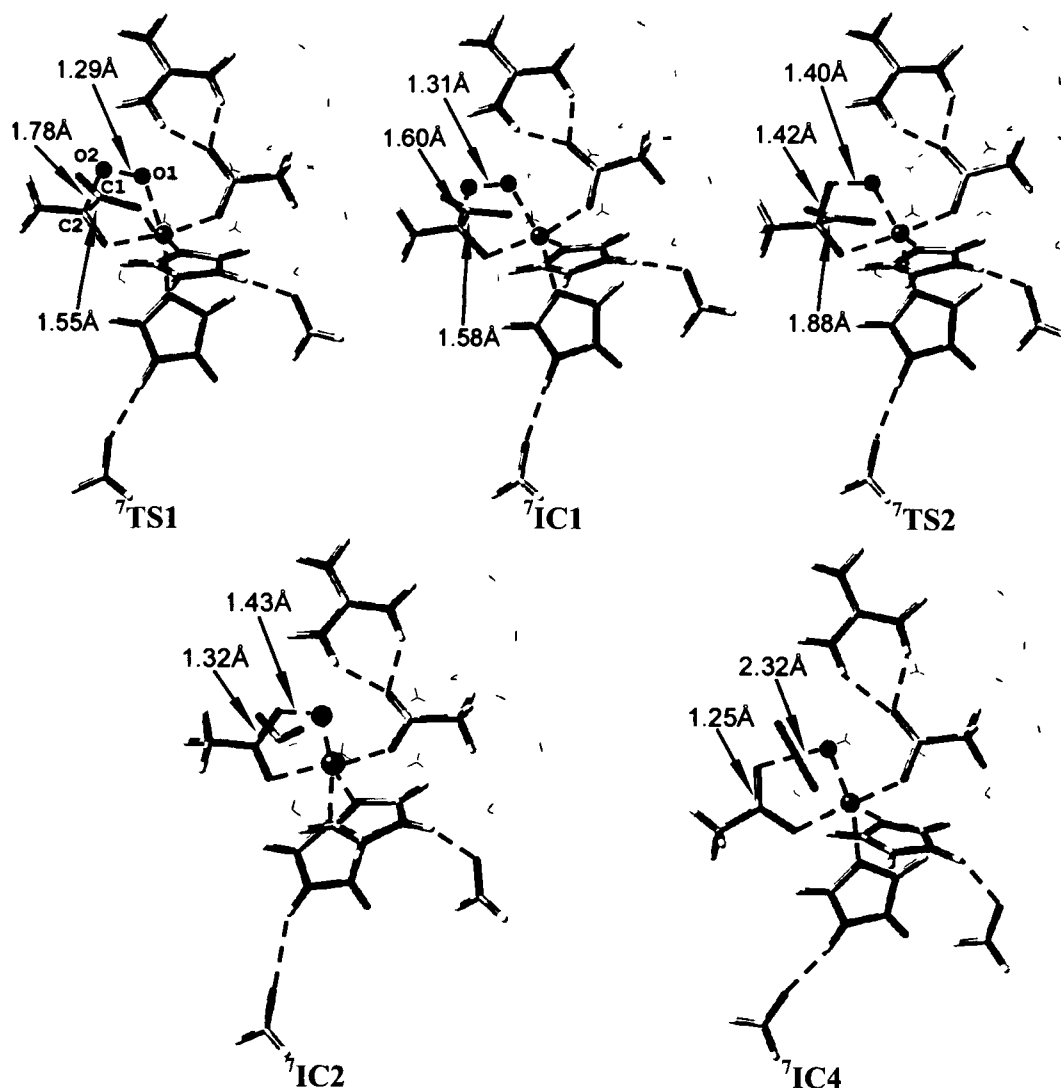


Figure 4.6 Optimized structures of the spin-septet transition states (${}^7\text{TS1}$ and ${}^7\text{TS2}$), intermediate complexes (${}^7\text{IC1}$ and ${}^7\text{IC2}$) and product complex (${}^7\text{IC4}$) for formation of the ferric-oxyl $\text{Fe}^{\text{III}}\text{-O}^\bullet$ compound.

The next and final step is formation of CO₂ and scission of the O1–O2 peroxy link. This is achieved by promotion of the unpaired electron localized on the CO₂^{•-} in ⁷IC2 to the $\sigma_{\text{O-O}}^*$ orbital on the peroxy unit (electronic configuration $\sigma_x^2 \pi_y^2 \pi_z^2 \pi_y^{*2} \pi_z^{*2} \sigma_x^{*1}$) resulting in it being cleaved. The barrier for this electron transfer via ⁷TS4 is approximately 5.7 kJ mol⁻¹ and leads to formation of ⁷IC4, which lies significantly lower in energy than ⁷IC2 by 189.3 kJ mol⁻¹. The newly formed CO₂ and acetate remain monodentately coordinated to the Fe(III) centre. It should be noted that the electron transfer to the $\sigma_{\text{O-O}}^*$ orbital ends up localized on the 2p_x atomic orbital of O1 in ⁷IC4. Consequently, in order to maintain the formally six unpaired electrons in the spin-septet state, no further electron transfer occurs. That is, in ⁷IC4 a ferric-oxyl (Fe^{III}-O[•]) radical moiety is formed with $r(\text{Fe-O}) = 1.86 \text{ \AA}$, rather than the ferryl-oxo (Fe^{IV}=O) species.

Previous computational studies have examined O₂ activation with decarboxylation of an α -KG cosubstrate on spin-septet PES for other α -KG–Fe(II)-dependent enzymes and a related biomimetic non-heme iron complex.^{32,56,58} Interestingly, in contrast to the above mechanism, they found that the reaction was concerted with decarboxylation of α -KG occurring concomitantly with nucleophilic attack by the Fe^{III}-O₂^{•-} group. It is noted, however, that these previous studies used the slightly smaller B3LYP/LACVP level of theory. Importantly, we also observed that ⁷IC1 is not a stationary point at this level of theory but instead collapses directly to ⁷IC4. In contrast, it continues to exist when larger basis sets with polarization functions were used. Therefore, the use of polarization functions on heavy atoms may be critical for more detailed descriptions of the reaction mechanism of related enzymes on the spin-septet PES.

4.3.2.3 Spin-Triplet Activation Pathway

As for the spin-septet path, nucleophilic attack and decarboxylation of pyruvate proceeds through a stepwise mechanism. More specifically, initially, the ³[Fe^{II}(η^1 -O₂²⁻)]

group in ${}^3\text{RC}$ attacks the pyruvate's α -carbon (C2) via ${}^3\text{TS1}$ with a barrier of 74.0 kJ mol^{-1} (Figure 4.4). This step gives the ferric-superoxido bound intermediate ${}^3\text{IC1}$ (cf. Figure 4.7), which is 66.5 kJ mol^{-1} higher in energy than ${}^3\text{RC}$. On the spin-triplet potential energy surface, ${}^3\text{IC1}$ is the reaction-coordinate equivalent to ${}^5\text{TS1}$ and ${}^7\text{IC1}$ on the quintet and septet surfaces respectively. In ${}^3\text{IC1}$, the Fe(III) centre is a high-spin quartet ($S = 3/2$) rather than a high-spin sextet ($S = 5/2$), which may explain why it is higher in energy than the equivalent ${}^5\text{TS1}$ and ${}^7\text{IC1}$ structures. Importantly, the barrier for this step almost doubles that of the barriers for oxygen activation via the alternate spin-quintet or -septet pathways. Consequently, at least for this part of the overall mechanism, a spin-triplet pathway is unlikely to compete with the more thermodynamically and kinetically favored initiation via the spin-quintet and septet surfaces. Hence, it is not discussed in detail herein. Nonetheless, it is noted that the spin-triplet ferryl-oxo intermediate ${}^3\text{IC4}$ is only 40.8 kJ mol^{-1} higher in energy than the quintet ferryl-oxo species ${}^5\text{IC4}$. In contrast, the septet ferric-oxyl intermediate ${}^7\text{IC4}$ is 51.5 kJ mol^{-1} higher in energy than ${}^5\text{IC4}$. This suggests that the subsequent reorientation of the ferryl-oxo ligand and oxidative dealkylation may proceed on either the triplet or quintet or both potential energy surfaces.

In summary, the redox conversion of the ferrous-dioxygen $\text{Fe}^{\text{II}}\text{-O}_2^{2\cdot}$ active-site complex into the ferric-superoxido $\text{Fe}^{\text{III}}\text{-O}_2^{\cdot-}$ intermediate on the overall spin-triplet surface seems thermodynamically unfavoured relative to the low-energy redox conversion on the overall septet and quintet surfaces. In fact, the initiation of oxygen activation can be mechanistically competing on the septet and quintet potential energy surfaces. This competitive initiation contrasts with the preference for the spin-quintet activation pathway previously found in computational studies on related enzymes.^{32,51,58,59} The origin of this discrepancy is likely due at least in part to (i) inclusion of polarization functions on the heavy atoms in the present study compared to (ii) their non-inclusion in previous studies.

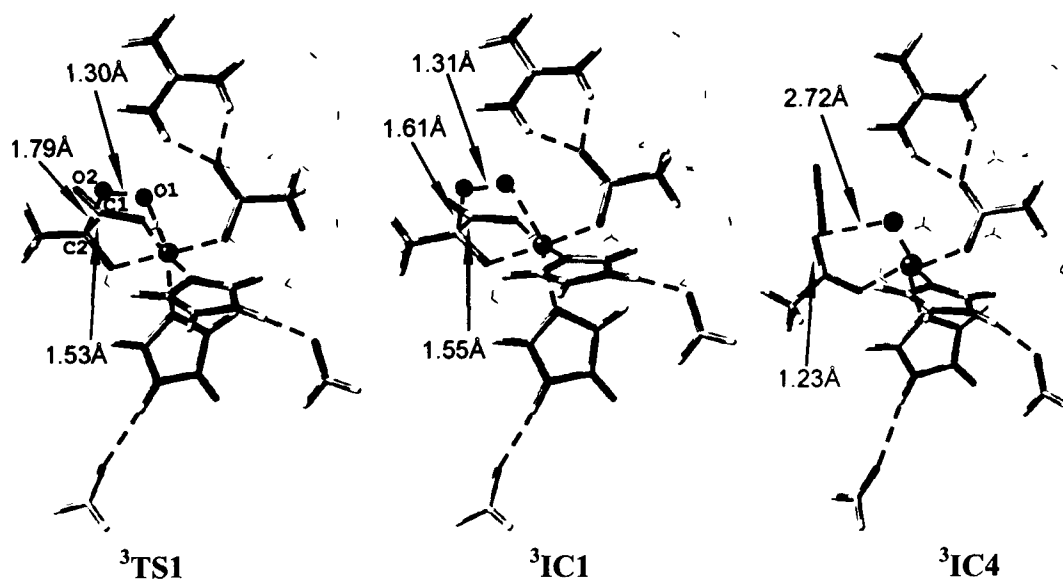


Figure 4.7 Optimized structures of the spin-triplet intermediate ${}^3\text{IC1}$ and product complex ${}^3\text{IC4}$ for formation of the ferryl-oxo $\text{Fe}^{\text{IV}}=\text{O}$ compound.

4.3.3 Conformational Reorientation of the High-Valency Iron-Oxygen Groups

Once oxygen has been activated as the highly reactive ferryl-oxo or ferric-oxyl species, the CO_2 molecule leaves the active site allowing the ferryl-oxo or ferric-oxyl group to reorient from the axial position to the equatorial plane of the iron centre. Release of CO_2 has been suggested to be a required step prior to conformational reorientation.²⁶ Therefore, the structures of the CO_2 -bound intermediates **IC4** with the CO_2 molecules eliminated were optimized to give stationary structures for the corresponding CO_2 -free complexes **IC5**.

The release of CO_2 frees up an equatorial position, thereby enabling rotation of the acetate group so that it now bidentately ligates to the iron centre. Importantly, the order of stability found for the CO_2 -bound complexes (**IC4**) is preserved in the CO_2 -free complexes (**IC5**), with relative energies again increasing in the order of quintet < triplet < septet. However, in the CO_2 -free complexes, the triplet ferryl-oxo ${}^3\text{IC5}$ and the septet ferric-oxyl ${}^7\text{IC5}$ compounds are now 27.9 and 100.5 kJ mol^{-1} higher in energy than the

quintet ferryl-oxo species $^5\text{IC5}$. That is, the triplet now lies lower and the septet higher in energy with respect to the quintet species, when compared with the CO_2 -bound complexes ($^3\text{IC4}$ and $^7\text{IC4}$ are 40.8 and 51.5 kJ mol^{-1} higher in energy than $^5\text{IC4}$, respectively). This also suggests that CO_2 release is energetically least favored on the spin-septet surface, but most favored on the spin-triplet surface.

Mechanistically, the subsequent conformational reorientation may occur in either a concerted or stepwise manner. The potential energy surfaces obtained for each of these possibilities are shown in **Figure 4.8**.

In $^5\text{IC5}$ and $^3\text{IC5}$, the concerted reorganization of the ligands occurs through $^5\text{TS6}$ and $^3\text{TS6}$ with barriers of 47.1 and 93.5 kJ mol^{-1} to yield $^5\text{IC6}$ and $^3\text{IC6}$, respectively (**Figure 4.8a**). Notably, $^5\text{IC6}$ lies only 0.4 kJ mol^{-1} higher in energy than $^5\text{IC5}$, whereas $^3\text{IC6}$ is 4.9 kJ mol^{-1} lower than $^3\text{IC5}$ and thus, is now just 22.6 kJ mol^{-1} higher in energy than the corresponding spin-quintet $^5\text{IC6}$. In contrast, the reorientated spin-septet complex $^7\text{IC6}$ lies 106.1 kJ mol^{-1} higher in energy than $^5\text{IC6}$.

For the concerted rearrangement on the spin-quintet surface, the $\text{Fe}^{\text{IV}}=\text{O}$ group is oriented relative to the equatorial plane with approximate angles ($\angle\text{N}_{\text{His187}}\text{FeO}$) in $^5\text{IC5}$, $^5\text{TS6}$ and $^5\text{IC6}$ of 94° , 140° and 178° respectively. Concomitantly, the Fe–O bond slightly lengthens from 1.62 Å in $^5\text{IC5}$ to 1.65 and 1.63 Å in $^5\text{TS6}$ and $^5\text{IC6}$ respectively. On the spin-triplet surface the corresponding orientations of the $\text{Fe}^{\text{IV}}=\text{O}$ group are 93° , 145° and 177° in $^3\text{IC5}$, $^3\text{TS6}$ and $^3\text{IC6}$, respectively, while the Fe–O bond length varies from 1.62 Å in $^3\text{IC5}$ to 1.60 and 1.62 Å in $^3\text{TS6}$ and $^3\text{IC6}$. That is, the Fe–O bond shortens slightly in the pathways transition structure.

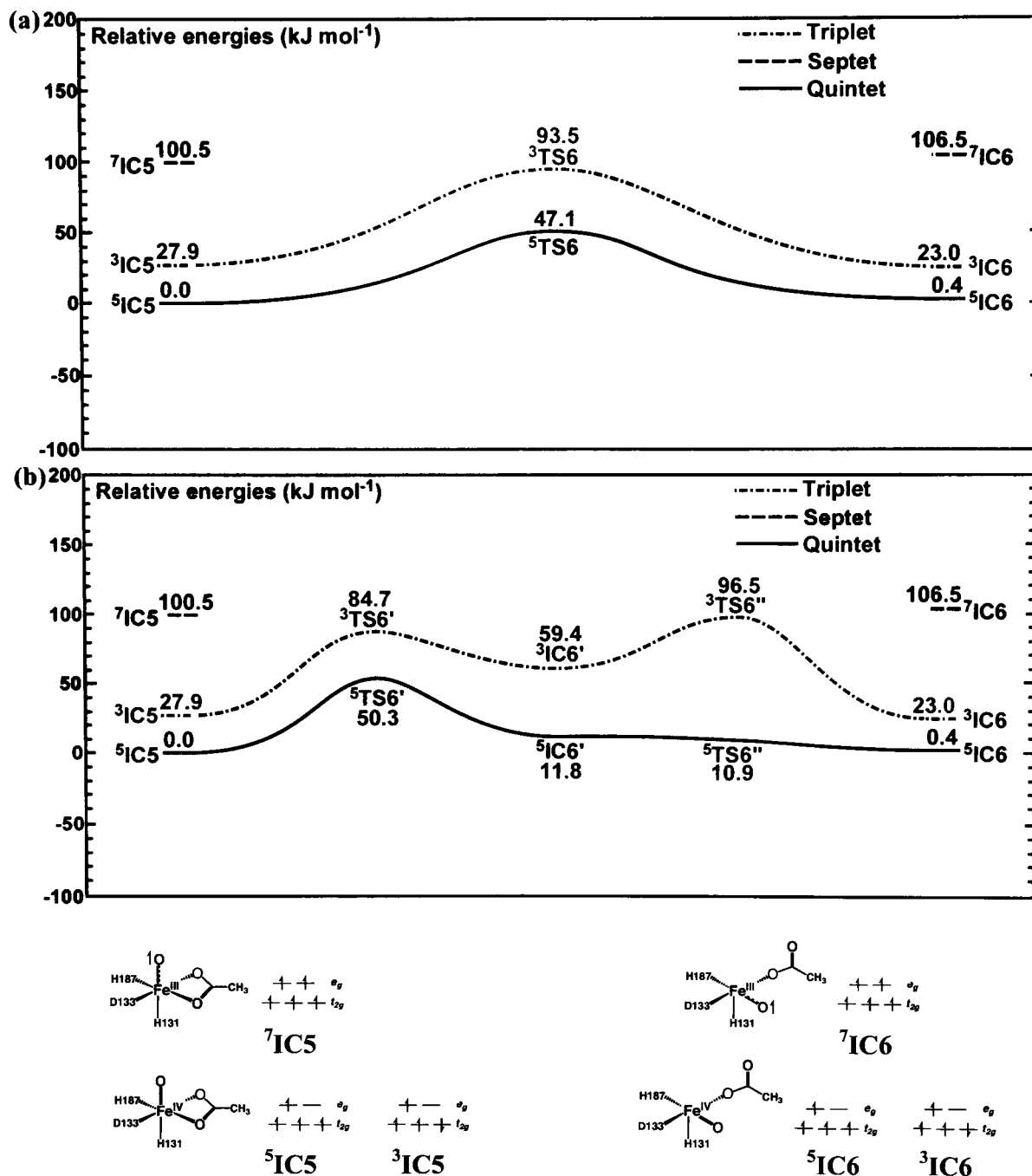


Figure 4.8 Potential energy surfaces for reorientation of the activated oxygen ligated to the iron centre from axial to equatorial position in the septet, quintet and triplet spin-states: (a) Concerted pathway, (b) Stepwise pathway.

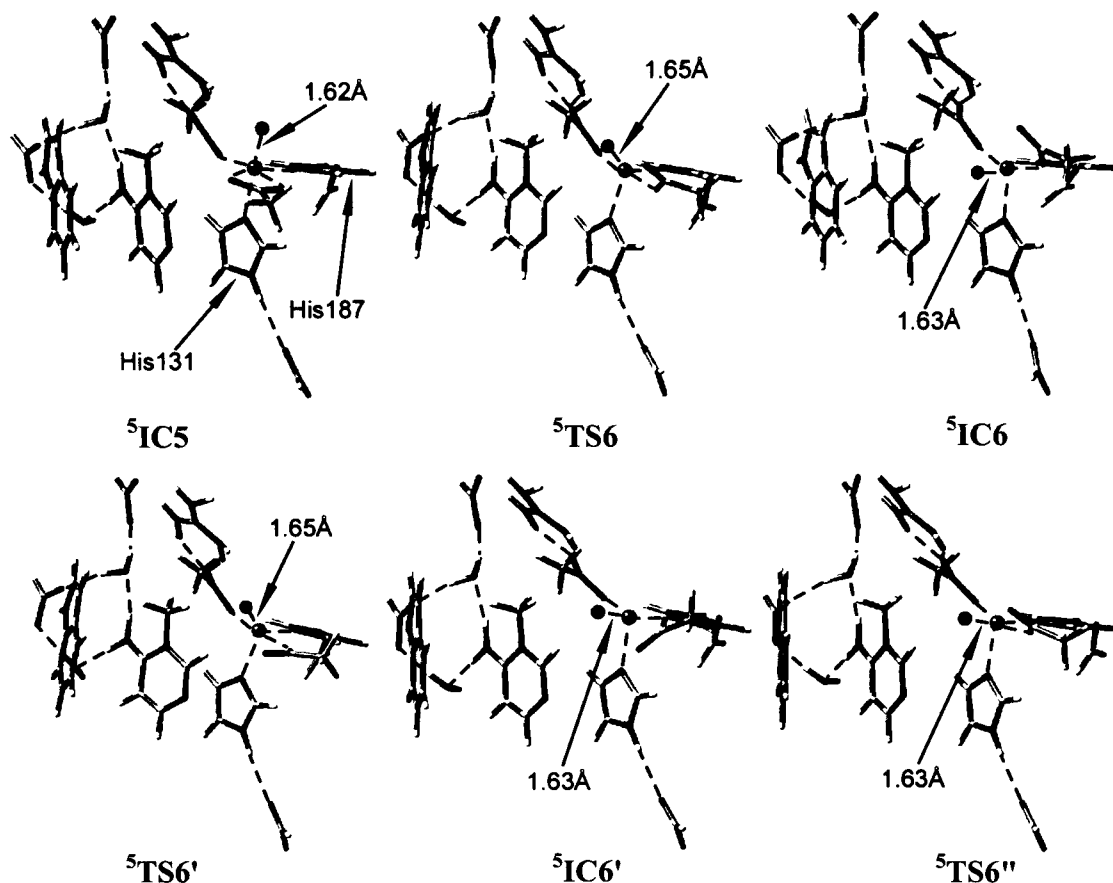


Figure 4.9 Optimized structures for the spin-quintet ferryl-oxo reorientation from axial ($^5\text{IC5}$) to equatorial ($^5\text{IC6}$) position.

The stepwise conformational reorientation is slightly less favoured than the concerted step for both the spin-quintet and triplet states (**Figure 4.8b** and **4.9**). On the quintet surface, the first step is rotation of the $\text{Fe}^{\text{IV}}=\text{O}$ ligand from the axial to equatorial position *trans* to His187 via $^5\text{TS6}'$, with a barrier of 50.3 kJ mol^{-1} . This leads to formation of intermediate $^5\text{IC6}'$ which lies 11.8 kJ mol^{-1} higher in energy than $^5\text{IC5}$. The carboxylate group then rotates via $^5\text{TS6}''$, effectively without a barrier, to give $^5\text{IC6}$. On the triplet surface, the rotation of $\text{Fe}^{\text{IV}}=\text{O}$ ligand from the axial to the equatorial via $^3\text{TS6}'$ has a barrier of 56.8 kJ mol^{-1} . This is significantly lower than the barrier (65.6 kJ mol^{-1}) for the alternate concerted pathway, and leads to the formation of the distorted

intermediate ${}^3\text{IC6}'$ lying 59.4 kJ mol^{-1} higher in energy than ${}^3\text{IC5}$. However, the subsequent rotation of the carboxylate group to give ${}^3\text{IC6}$ proceeds via the high energy transition structure ${}^3\text{TS6}''$ at an overall cost of 96.5 kJ mol^{-1} .

Finally, on the spin-septet surface it was found that during either the concerted or stepwise reorientation, the highly electrophilic ferric-oxyl radical ($\text{Fe}^{\text{III}}\text{-O}'$) abstracts a proton from the nearby active site residue Arg210 to form the ferric-hydroxyl ($\text{Fe}^{\text{III}}\text{-OH}$) radical group. As a result, the ferric-oxyl group loses its catalytic activity because it is now disabled from initiating the oxidative dealkylation of the damaged nucleobase.

Thus, overall, migration of the activated oxygen from the axial to equatorial position *trans* to His187 facilitates cleavage of the dative carboxylate–Fe bond at the equatorial position and subsequent rotation of the carboxylate. Furthermore, the latter group now sits almost upright with respect to the equatorial plane, as in the CO_2 -bound complexes (**IC4**). Thus, conformational reorientation alters the geometry of the Fe centre from a six-coordinate octahedral to a five-coordinate square-pyramidal arrangement, with moderate reorganization of the ligands on the basal plane.

4.3.4 Oxidative Dealkylation of the Methylated Nucleobase

Once the activated oxygen of the $\text{Fe}^{\text{IV}}=\text{O}$ group has occupied the position *trans* to His187 in the active site, it is now sufficiently near to the methylated nucleobase to be able to remove the alkylation damage. This process is initiated by hydrogen abstraction from the methyl group, followed by a radical rebound hydroxylation of the methylene group. The subsequent decomposition of the hydroxymethyl nucleobase yields formaldehyde and the native nucleobase. The potential energy surfaces obtained for oxidative removal of the methyl group (the alkyl damage) from the model compound 3me4amPym^+ are presented in **Figure 4.10**. Unlike the activation stage of the catalytic

pathway, the oxidative dealkylation stage follows the same reaction mechanism on both the spin-quintet and -triplet potential energy surfaces, with the triplet surface lying significantly higher in energy than the quintet surface. This could follow from the electron configurations of the Fe centre being more stable in the overall quintet spin-state (cf. Figure 4.10).

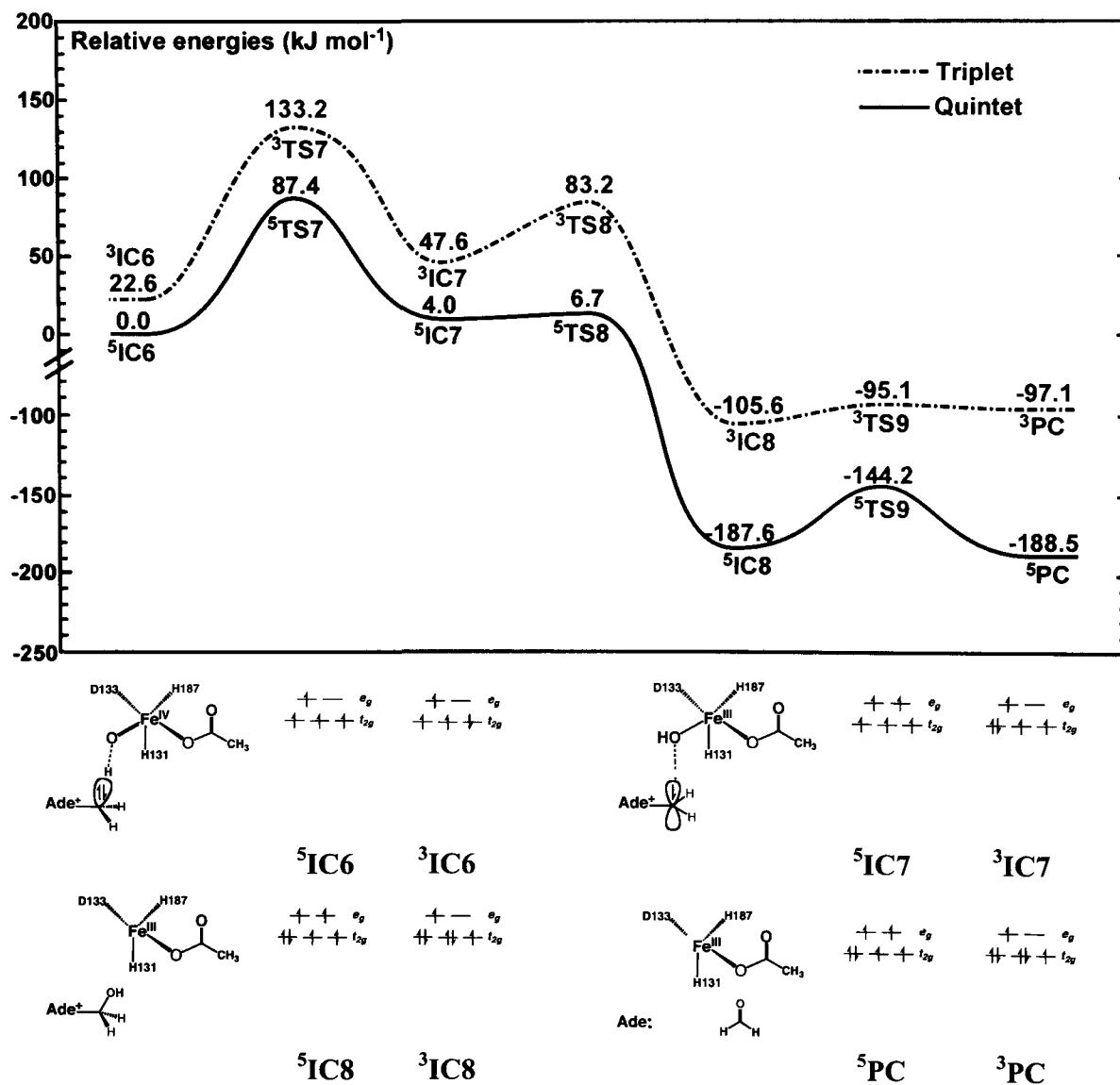


Figure 4.10 Potential energy surfaces for dealkylation of 3me4amPym⁺ as a substrate model in the triplet and quintet spin-states.

Hydrogen atom (H^\bullet) abstraction from the substrate is found to be the rate-controlling step in the catalytic mechanism of AlkB. In this step the highly electrophilic ferryl-oxo ($\text{Fe}^{\text{IV}}=\text{O}$) oxygen in the **IC6** intermediates abstracts the axial H^\bullet from the methyl of $3\text{me}4\text{amPym}^+$. This transfer results in formation of the intermediates **IC7** in which the $\text{Fe}^{\text{IV}}=\text{O}$ has now been reduced to a ferric-hydroxido ($\text{Fe}^{\text{III}}-\text{OH}$) group with concomitant conversion of the substrate to a 3-methylene-4-amino pyrimidinyl radical cation (cf. **Figures 4.10** and **4.11**). Although for both the spin-quintet and -triplet surfaces this process is only relatively slightly endoergic by 4.0 and 25.0 kJ mol^{-1} , it proceeds with quite high barriers of 87.4 and 133.2 kJ mol^{-1} , respectively (**Figure 4.10**). Indeed, the latter barrier is above that generally considered to be the upper thermodynamic limit for enzymatic reactions.^{60,61}

It is noted that a previous DFT computational study examined the related α -KG-Fe(II) non-heme enzyme TauD and the mechanism by which it hydroxylates propene.³³ Similar to the present findings for AlkB, the H^\bullet abstraction by the $\text{Fe}^{\text{IV}}=\text{O}$ group was also found to be the rate-limiting step. However, the barrier was markedly lower. This difference may be due in part to the fact that, at the level of theory employed herein, the $-\text{CH}_2-\text{H}$ bond dissociation energy (BDE) for 1meAde^+ (437.3 kJ mol^{-1}) is markedly higher than that of propene (380.5 kJ mol^{-1}). Indeed, for related H^\bullet abstraction reactions in heme iron enzymes, it has been previously found that the barrier is proportional to the appropriate homolytic C-H BDE of the substrate.^{62,63} This may also help explain why in cells the highly oxidant α -KG-Fe(II)-dependent non-heme enzymes are involved in oxidative dealkylation of damaged nucleobases rather than their more moderately oxidant heme counterparts.³³ Furthermore, AlkB is a relatively slow enzyme with $k_{\text{cat}} = 4.5 \text{ min}^{-1}$ at 37°C.^{20,26} This implies that the activation energy for the catalyzed reaction is ca 83 kJ mol^{-1} , which is in good agreement with the calculated spin-quintet barrier for the rate-controlling H^\bullet abstraction (87.4 kJ mol^{-1}).

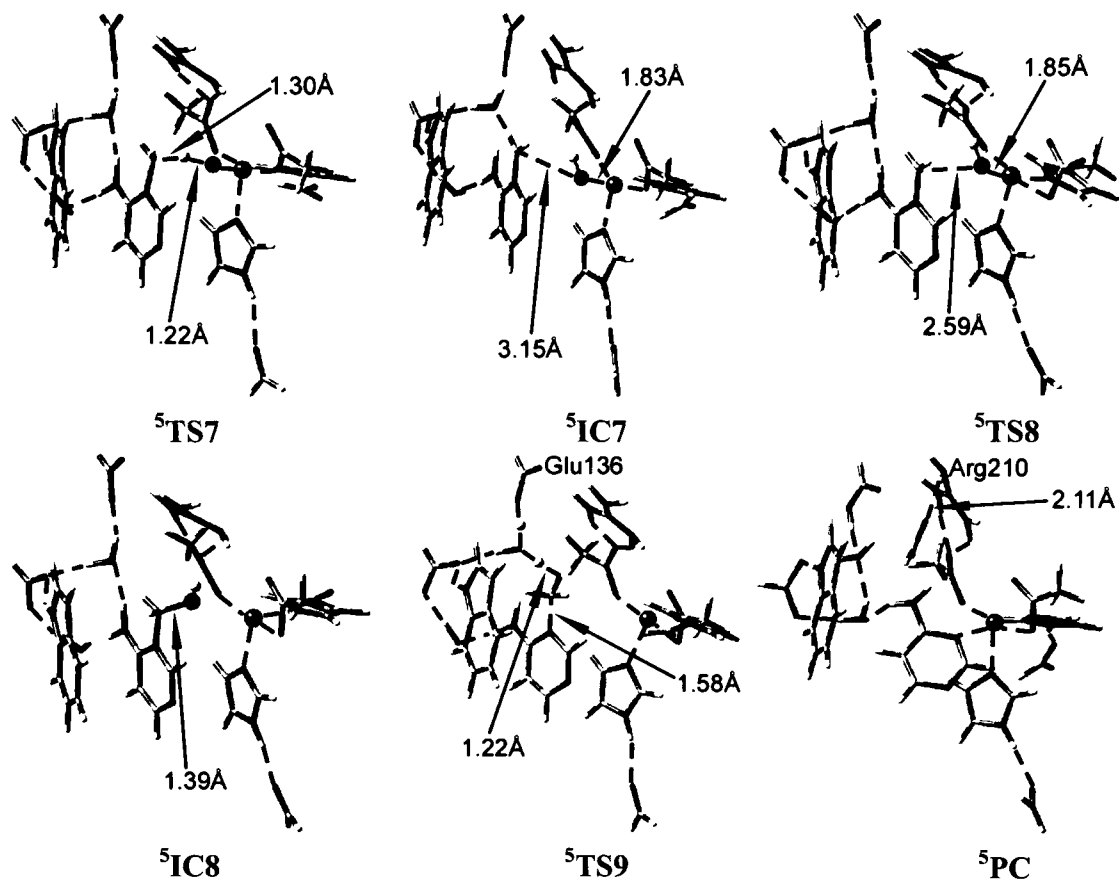


Figure 4.11 Optimized structures of the spin-quintet transition states (⁵TS7, ⁵TS8 and ⁵TS9), intermediates (⁵IC7 and ⁵IC8) and product complex (⁵PC) for dealkylation of 3me4amPym⁺ as a substrate model.

In the next step, the Fe^{III}-OH group in the intermediate IC7 electrophilically attacks the methylene group in a radical rebound process in which Fe(III) is reduced to Fe(II) and a hydroxyl radical OH[•] is transferred from the FeηOH group to the substrate to form the 3-hydroxymethyl-4-amino pyrimidinyl cation in IC8. This step is highly exoergic by 191.6 and 153.2 kJ mol⁻¹ and occurs via TS7 with barriers of 2.7 and 35.6 kJ mol⁻¹ on the quintet and triplet surfaces, respectively. The significant stabilization energies observed in this step is due in large part to the structural reorganization that occurs during reduction of the penta-coordinated Fe(III) to the tetra-coordinated Fe(II).

In the final step, the 3-hydroxymethyl-4-amino pyrimidinyl cation (intermediates $^5\text{IC8}$ and $^3\text{IC8}$) decomposes to yield formaldehyde and 4-amino pyrimidine, the regenerated nucleobase based on the substrate model used. Specifically, the spin-quintet and -triplet product complexes ^5PC and ^3PC lie -0.9 and 8.5 kJ mol^{-1} in energy relative to their corresponding intermediates $^5\text{IC8}$ and $^3\text{IC8}$. The barriers to decomposition are quite low at just 43.4 and 10.5 kJ mol^{-1} on the quintet and triplet surfaces, respectively. Our active-site model suggests that the Glu136 residue acts as the final proton acceptor in the water-catalyzed deprotonation of the 3-hydroxymethyl OH group and the concomitant rupture of the $\text{N3}-\text{C}^3$ bond to form formaldehyde. For example, in $^5\text{TS9}$ both the $\text{N3}-\text{C}^3$ and $\text{O}-\text{H}$ bonds lengthen to 1.58 and 1.22 Å, indicating that this decomposition is concerted. In the product complexes the reduced 4-amino pyrimidine is ligated to the Fe(II) through N3 at the basal position left vacant by the hydroxyl group (e.g., in ^5PC the Fe-N distance is 2.39 Å). The newly formed formaldehyde instead hydrogen bonds to Arg210 with a distance of 2.11 Å. Finally, the repaired nucleobase, formaldehyde and succinate leave the cleft and three water molecules ligate to the now Fe(II) centre (not shown), thereby regenerating the initial active site for a new catalytic cycle.

4.4 Conclusions

This chapter presents a thorough mechanistic study of the catalytic pathway in the AlkB class of enzymes using an active-site model in conjunction with DFT and a self-consistent reaction field. The initial binding of O_2 to the active site Fe centre is found likely to occur coupled with a fast electron transfer. Complexes in the overall septet and quintet spin-states are thermodynamically stable in the ferric-superoxido $\text{Fe}^{\text{III}}-\text{O}_2^{\cdot-}$ form regardless of the binding mode of dioxygen, because Fe(II) readily reduces $^3\text{O}_2$ to yield the highly reactive $\text{O}_2^{\cdot-}$. In contrast, the overall triplet and singlet spin-states are stable in the ferrous-dioxygen $\text{Fe}^{\text{II}}-\text{O}_2^{2\cdot}$ form, and charge transfer from the Fe(II) to $^3\text{O}_2$ moiety

does not take place. In addition, no η^2 -coordinated species were found without charge transfer while for all η^1 -coordinated structures, the bent binding mode is always preferred over any linear arrangement.

The activation of oxygen through formation of a spin-quintet ferryl-oxo $\text{Fe}^{\text{IV}}=\text{O}$ intermediate is thermodynamically preferred over activation through formation of a spin-septet ferric-oxyl $\text{Fe}^{\text{III}}-\text{O}^{\bullet}$ intermediate, even though both pathways are exoergic. In fact, it was found that the initiation of oxygen activation can be mechanistically competing on the septet and quintet potential energy surfaces. This is in contrast to the preference for the spin-quintet activation pathway previously found in computational studies on related enzymes.^{32,51,58,59} This is found to likely be due to (i) inclusion of polarization functions on the heavy atoms in this chapter compared to (ii) their non-inclusion in previous studies.

Reorientation of the activated oxygen ferryl-oxo ($\text{Fe}^{\text{IV}}=\text{O}$; spin-quintet and -triplet pathways) and ferric-oxyl ($\text{Fe}^{\text{III}}-\text{O}^{\bullet}$, spin-septet pathway) species is necessary for the oxidative dealkylation to proceed. This step is found to be nearly thermoneutral on all three spin-state potential energy surfaces, with the lowest barrier (ca 50 kJ mol^{-1}) being found for the quintet surface. For the spin-septet pathway it was found that during reorientation the $\text{Fe}^{\text{III}}-\text{O}^{\bullet}$ group, it accepts a proton from the nearby Arg210 active site residue to form $\text{Fe}^{\text{III}}-\text{OH}$. As a result, it is rendered catalytically inactive.

Finally, the rate-controlling step for both the spin-triplet and -quintet pathways was located at the oxidative dealkylation stage. Specifically, it is the abstraction of the axial hydrogen atom from the methyl group (the alkyl damage modeled in this chapter) of the alkylated nucleobase. The lowest barrier for this step (87.4 kJ mol^{-1}) occurs on the spin-quintet surface and is in good agreement with the experimental activation energy for the

catalyzed reaction (ca 83 kJ mol⁻¹). The final undamaged nucleobase is then generated via a typical radical-rebound-type mechanism. Specifically, the Fe^{III}-OH group resulting from the H[•] abstraction, transfers its -OH to the substrate to form a -C³H₂OH group. The final step is cleavage of the N-C³ bond, which occurs in concert with deprotonation of the -C³H₂OH moiety to give H₂CO and the undamaged nucleobase.

References

- (1) De Bont, R.; van Larebeke, N. *Mutagenesis* **2004**, *19*, 169.
- (2) Wang, Y. *Chem. Res. Toxicol.* **2008**, *21*, 276.
- (3) Taylor, J. S. *Acc. Chem. Res.* **1994**, *27*, 76.
- (4) Kamiya, H.; Iwai, S.; Kasai, H. *Nucleic Acids Res.* **1998**, *26*, 2611.
- (5) Lysetska, M.; Knoll, A.; Boehringer, D.; Hey, T.; Krauss, M.; Krausch, G. *Nucleic Acids Res.* **2002**, *30*, 2686.
- (6) Neeley, W. L.; Essigmann, J. M. *Chem. Res. Toxicol.* **2006**, *19*, 491.
- (7) Friedberg, E. C.; McDaniel, L. D.; Schultz, R. A. *Curr. Opin. Gene. Dev.* **2004**, *14*, 5.
- (8) Lindahl, T. *Nature* **1993**, *362*, 709.
- (9) Rydberg, B.; Lindahl, T. *EMBO J.* **1982**, *1*, 211.
- (10) Hecht, S. S. *Mutat. Res.* **1999**, *424*, 127.
- (11) Drabløs, F.; Feyzi, E.; Aas, P. A.; Vaagbø, C. B.; Kavli, B.; Bratlie, M. S.; Peña-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. *DNA repair* **2004**, *3*, 1389.
- (12) Mishina, Y.; Duguid, E. M.; He, C. *Chem. Rev.* **2006**, *106*, 215.
- (13) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. *DNA repair* **2007**, *6*, 429.
- (14) McCullough, A. K.; Dodson, M. L.; Lloyd, R. S. *Annu. Rev. Biochem.* **1999**, *68*, 255.

- (15) O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418.
- (16) Lindahl, T.; Sedgwick, B.; Sekiguchi, M.; Nakabeppu, Y. *Annu. Rev. Biochem.* **1988**, *57*, 133.
- (17) Mishina, Y.; Yang, C. G.; He, C. *J. Am. Chem. Soc.* **2005**, *127*, 14594.
- (18) Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16660.
- (19) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 855.
- (20) Koivisto, P.; Duncan, T.; Lindahl, T.; Sedgwick, B. *J. Biol. Chem.* **2003**, *278*, 44348.
- (21) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174.
- (22) Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. *Nature* **2002**, *419*, 178.
- (23) Aravind, L.; Koonin, E. V. *Genome Biol.* **2001**, *2*, research0007.1.
- (24) Clifton, I. J.; McDonough, M. A.; Ehrismann, D.; Kershaw, N. J.; Granatino, N.; Schofield, C. J. *J. Inorg. Biochem.* **2006**, *100*, 644.
- (25) Kovaleva, E. G.; Lipscomb, J. D. *Nat. Chem. Biol.* **2008**, *4*, 186.
- (26) Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F. *Nature* **2006**, *439*, 879.
- (27) Wilmouth, R. C.; Turnbull, J. J.; Welford, R. W. D.; Clifton, I. J.; Prescott, A. G.; Schofield, C. J. *Structure* **2002**, *10*, 93.
- (28) Valegård, K.; van Scheltinga, A. C. T.; Lloyd, M. D.; Hara, T.; Ramaswamy, S.; Perrakis, A.; Thompson, A.; Lee, H. J.; Baldwin, J. E.; Schofield, C. J.; Hajdu, J.; Andersson, I. *Nature* **1998**, *394*, 805.
- (29) Zhang, Z.; Ren, J.; Harlos, K.; McKinnon, C. H.; Clifton, I. J.; Schofield, C. J. *FEBS Lett.* **2000**, *517*, 7.

- (30) Nakajima, R.; Yamazaki, I. *J. Biol. Chem.* **1987**, *262*, 2576.
- (31) Berglund, G. I.; Carlsson, G. H.; Smith, A. T.; Szöke, H.; Henriksen, A.; Hajdu, J. *Nature* **2002**, *427*, 463.
- (32) Borowski, T.; Bassan, A.; Siegbahn, P. E. M. *Chem. Eur. J.* **2004**, *10*, 1031.
- (33) de Visser, S. P. *J. Am. Chem. Soc.* **2006**, *128*, 9813.
- (34) de Visser, S. P. *Chem. Commun.* **2007**, 171.
- (35) Borowski, T.; de Marothy, S.; Broclawik, E.; Schofield, C. J.; Siegbahn, P. E. M. *Biochemistry* **2007**, *46*, 3682.
- (36) Godfrey, E.; Porro, C. S.; de Visser, S. P. *J. Phys. Chem. A* **2008**, *112*, 2464.
- (37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (38) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (39) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (40) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.

- (41) Jaguar 5.5; Schrodinger, L. L. C.: Portland, OR, 1991–2003.
- (42) Robinet, J. J.; Gault, J. W. *J. Phys. Chem. B* **2008**, *112*, 3462.
- (43) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.
- (44) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232.
- (45) Siegbahn, P. E. M.; Borowski, T. *Acc. Chem. Res.* **2006**, *39*, 729.
- (46) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515.
- (47) Solomon, E. I. *Inorg. Chem.* **2001**, *40*, 3656.
- (48) Solomon, E. I.; Decker, A.; Lehnert, N. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3589.
- (49) Lundberg, M.; Morokuma, K. *J. Phys. Chem. B* **2007**, *111*, 9380.
- (50) Nemukhin, A. V.; Grigorenko, B. L.; Topol, I. A.; Burt, S. K. *Int. J. Quant. Chem.* **2006**, *106*, 2184.
- (51) Lundberg, M.; Siegbahn, P. E. M.; Morokuma, K. *Biochemistry* **2008**, *47*, 1031.
- (52) Filippo Jr, J. S.; Romano, L. J.; Chern, C. I.; Valentine, J. S. *J. Org. Chem.* **1976**, *41*, 586.
- (53) Filippo Jr, J. S.; Chern, C. I.; Valentine, J. S. *J. Org. Chem.* **1976**, *41*, 1077
- (54) Chiou, Y. M.; Que Jr, L. *J. Am. Chem. Soc.* **1995**, *117*, 3999.
- (55) Mehn, M. P.; Fujisawa, K.; Hegg, E. L.; Que Jr, L. *J. Am. Chem. Soc.* **2003**, *125*, 7828.
- (56) Borowski, T.; Bassan, A.; Siegbahn, P. E. M. *Inorg. Chem.* **2004**, *43*, 3277
- (57) Bassan, A.; Borowski, T.; Siegbahn, P. E. M. *Dalton Trans.* **2004**, 3153.
- (58) Borowski, T.; Bassan, A.; Siegbahn, P. E. M. *Biochemistry* **2004**, *43*, 12331.
- (59) Bassan, A.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Chem. Eur. J.* **2003**, *9*, 106
- (60) Lim, V. L.; Curran, J. F.; Garber, M. B. *J. Mol. Biol.* **2005**, *351*, 470.
- (61) Lim, V. L.; Kljashtorny, V. G. *Mol. Biol.* **2006**, *40*, 507.
- (62) Kaizer, J.; Klinker, E. J.; Oh, N. Y.; Rohde, J.-U.; Song, W. J.; Stubna, A.; Kim, J.; Münck, E.; Nam, W.; Que Jr, L. *J. Am. Chem. Soc.* **2004**, *126*, 472.

- (63) de Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. *J. Am. Chem. Soc.* **2004**, *126*, 8362.

Chapter 5

A Density Functional Theory Investigation on the Mechanism of the Hepatitis Delta Virus Ribozyme

5.1 Introduction

Recently, there has been increasing interest in understanding the properties and mechanisms of ribozymes, catalytic RNAs.¹⁻⁴ This is due in part to the fact that they have been found to play physiologically important roles within many cells and viruses.¹⁻⁵ In addition, they are also now known to catalyze a diverse variety of reactions including phosphodiester cleavage¹ and amide bond formation.⁶ Yet in comparison to their protein-based enzyme counterparts, they possess a quite limited and more homogeneous repertoire of chemical functionality. For example, the only chemical difference between the four component nucleotides of RNA is the nucleobases themselves (adenine, guanine, cytosine and uracil) which all possess similar hydrogen bond donors and acceptors, and extended π -systems. Thus, ribozymes offer challenging new insights into catalysis and the chemistry of their components.

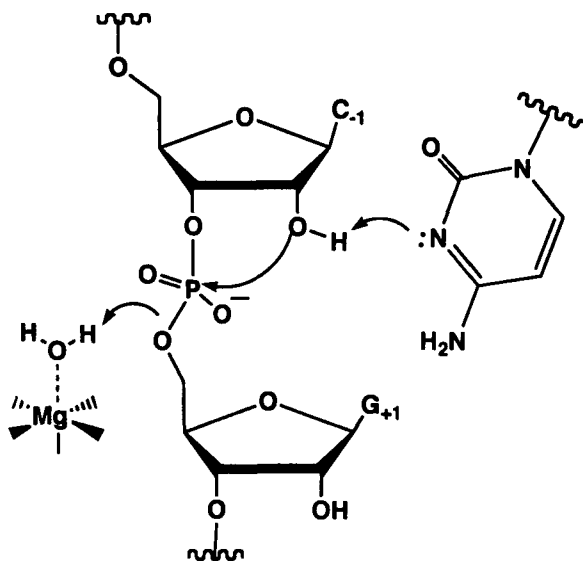
In the hepatitis delta virus (HDV), a ribozyme simply known as the HDV ribozyme plays an essential role in its reproductive cycle.⁷ In particular, it catalyses its own cleavage at a select backbone phosphodiester bond at a rate of $>1 \text{ s}^{-1}$ ⁸ This is slower than that recently measured for a natural hammerhead ribozyme which catalyses the same reaction, but at a rate of $>14.5 \text{ s}^{-1}$ ⁹ However, significantly, it is one of an increasing

number of remarkable phosphodiester-bond-cleaving ribozymes that are proposed to directly employ an active site nucleobase in their catalytic mechanisms.¹⁰⁻¹² Unfortunately, it is unclear how this nucleobase may participate in the mechanism of the HDV ribozyme.

In 1998 and 2000 Doudna and coworkers^{8,13} reported several crystal structures of the HDV ribozyme. On the basis of these, they suggested that a cytosine (C75) is important in the catalytic mechanism and, moreover, that it acts as a general base using its N3 ring nitrogen to deprotonate the 2'-OH group of the ribose sugar as shown in **Scheme 5.1 (Mechanism 1)**. They suggested that this ability to act as a catalytic base was possibly due in part to it being surrounded by several polar ligands such as phosphate groups as well as a negative electrostatic potential resulting from the general environment of the active site. At approximately the same time, Been and coworkers¹⁴ showed that an imidazole buffer could rescue the activity of a mutant in which the key cytosine was replaced by a uracil. In particular, they showed that when C75 was replaced by adenine which also contains a basic N3 ring centre, the HDV ribozyme only exhibited decreased activity. However, it lost all catalytic ability when C75 was replaced by uracil (C75U) or guanine, both of which contain a ring $-NH-$ group at the N3 position. This was interpreted as additional evidence of C75 acting as a general base. More recently, Doudna and coworkers¹⁵ concluded that they had obtained further support for C75 acting as a base from crystal structures they reported of wild-type HDV ribozyme in its pre-cleaved state and of several variants resulting from mutagenesis studies. In addition, they proposed that a hydrated divalent magnesium ion acts a catalytic general acid by donating a proton from one of its waters to the appropriate 5'-oxygen thus cleaving the 5'-O—P bond (**Scheme 5.1**). Recently, Spomer and coworkers^{16,17} used molecular dynamics to investigate structures of possible precursor and product states of the HDV ribozyme. In part on the basis of the formation of transient interactions between the mechanistically important

ribose sugar and cytosine they also suggested that C75 may act as a catalytic base via its N3 centre.

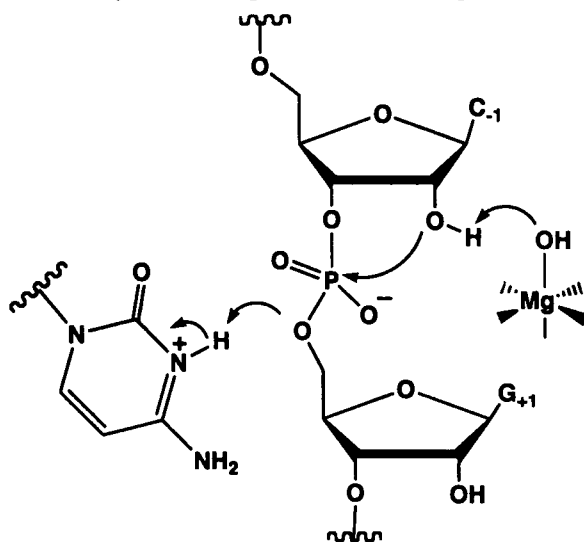
Scheme 5.1 Proposed^{14,15} mechanism 1 for the HDV ribozyme in which cytosine acts as a general base while a hydrated Mg^{2+} ion acts a general acid.



However, almost simultaneously an alternative mechanism was proposed. In 2000, Bevilacqua and coworkers¹⁸ investigated the reactivity-pH profiles of the HDV ribozyme in solutions of particular monovalent and divalent metal cations. On the basis of their observations they proposed that C75 in fact acts as a general acid in the catalytic mechanism. Furthermore, they also proposed that when the hydrated Mg^{2+} ion binds in the active site one of its waters is deprotonated. Thus, the resulting hydrated Mg^{2+} hydroxide now in fact acts as a catalytic general base in HDV's mechanism. In addition, they suggested that the rescue of activity by imidazole of the C75U mutant could alternatively support the proposal of C75 as a catalytic general acid. The resulting alternative proposed mechanism (Mechanism 2) is shown in **Scheme 5.2**. More recently, Piccirilli and coworkers^{19,20} considered a modified HDV ribozyme in which the key scissile phosphodiester (P—O) bond was replaced by a phosphorothiolate (P—S) linkage.

They found that the C75U mutant could in fact cleave the P—S but not wild-type P—O bond. In addition, they also found that the catalytic rate of the HDV ribozyme was reduced at higher pH's, when the amount of available protonated C75 is expected to be less. Consequently, they also concluded that the active site cytosine is protonated and hence acts as a general acid to protonate the leaving 5'-oxygen (**Scheme 5.2**).

Scheme 5.2 Proposed¹⁸ mechanism 2 for the HDV ribozyme in which cytosine acts as a general acid while an ionized hydrated Mg^{2+} ion acts as a general base.



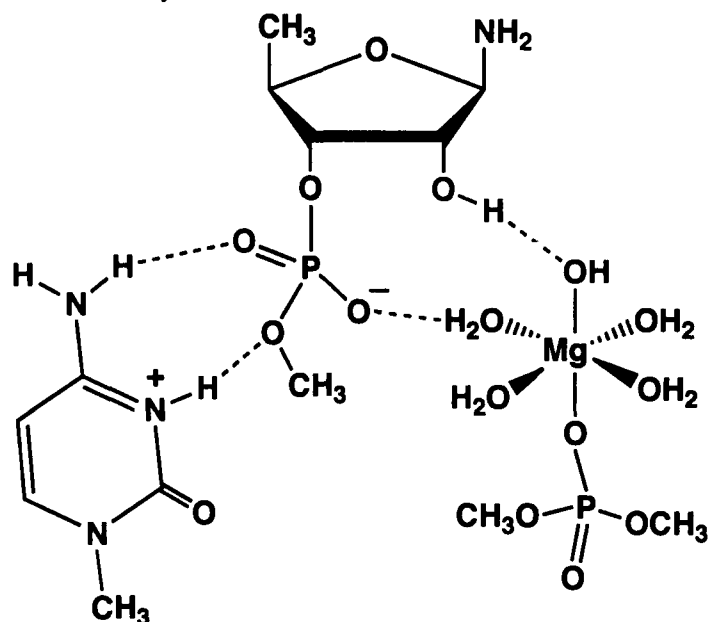
Clearly, it has proven experimentally difficult to discern whether cytosine acts as the base or acid in the HDV ribozyme. In this chapter, we have employed density functional theory methods in order to investigate its catalytic mechanism. The potential of cytosine to act either as a catalytic base or acid also was considered by examining the ability of the general environment and specific hydrogen bonding species to influence its basicity.

5.2 Computational Methods

All density functional theory (DFT) calculations were performed using the Gaussian 03²¹ and Jaguar 5.5²² suite of programs. The B3LYP method was used, a combination of Becke's three-parameter hybrid exchange functional^{23,24} as implemented²⁵ in these programs and the correlation functional of Lee, Yang and Parr.²⁶ For investigations into the effects of the polarity of the environment and/or specific hydrogen bonding interactions on the proton affinity of cytosine, optimized structures were obtained at the B3LYP/6-311G(d,p) level in the gas-phase and with inclusion of general environmental effects via the Integral Equation Formalism polarizable-continuum model (IEF-PCM) or Poisson-Boltzmann self-consistent reaction field (PB-SCRF) methods of Gaussian 03²¹ and Jaguar 5.5²² respectively. Harmonic vibrational frequencies and zero-point vibrational energy (ZPVE) corrections were calculated for the optimized gas-phase geometries. Relative energies were obtained by performing single point calculations on these geometries at the B3LYP/6-311+G(d,p) level using the same continuum method as appropriate, with inclusion of the corresponding ZPVE correction.

For the mechanism studies, due to the size of the models employed, optimized structures and harmonic vibrational frequencies were obtained at the B3LYP/6-31G(d,p) level of theory. The ribozyme active site model was derived from the crystal structures PDB: 1VC5 and 1SJ3 (**Scheme 5.3**). Relative energies were obtained by performing single point calculations on these geometries at the B3LYP/6-311+G(2df,p) level of theory using the IEF-PCM solvation method ($\epsilon = 4.0$), with inclusion of the appropriate ZPVE corrections. We note that similar computational approaches have been used to study related systems,²⁷⁻²⁹ including the related catalytic mechanism proposed for the hammerhead ribozyme.^{30,31} The applicability of such methods to the study of enzymatic mechanisms has been previously reviewed.³²

Scheme 5.3 Schematic illustration of the chemical model used to investigate the catalytic mechanism of the HDV ribozyme.



5.3 Results and Discussion

In order to gain greater insights into the possible roles of the proposed active site cytosine and the hydrated magnesium ion, we began by considering their individual properties. In particular, the effect of a polar environment and specific hydrogen bond interactions, or coordination environments respectively on their basicity was examined.

5.3.1 Proton Affinity of C75

It has been noted previously⁸ that in order for cytosine to act as a general base and deprotonate the ribose sugar's 2'-OH group (Scheme 5.1) the pK_a of its N3 centre must be approximately 7. However, in free cytidine it has been experimentally measured to be 4.2.³³ The required pK_a shift has been suggested⁸ to possibly occur by formation of a hydrogen bond between the cytosine's amine group and a nearby backbone phosphate anion, although it has been suggested^{15,34} that such an interaction may not occur until

after the reaction has occurred. Alternatively, this may be achieved via electrostatic stabilization of protonated cytosine by the active site environment.

First, the effect of a polar environment on the affinity of cytosine's N3 centre for a proton (PA_{N3}) was considered. Specifically, the dielectric constant (ϵ) was increased from 1.0 to 2.0, 10.0, 45.0 and 78.39 (water) for both IEF-PCM and PB-SCRF methods used in order to consider the effect of an increasingly polar environment and to compare the applicability of these solvation models. The results obtained are shown in **Figure 5.1** and summarized in **Table 5.1**. It should be noted that at $\epsilon = 1.0$ PA_{N3} is calculated to be $964.4 \text{ kJ mol}^{-1}$, significantly lower than that calculated for a $2'\text{-O}^-$ ribose anion, $1819.1 \text{ kJ mol}^{-1}$. We note that the ribose model used for this calculation consisted of the active site sugar model shown in **Scheme 5.3**.

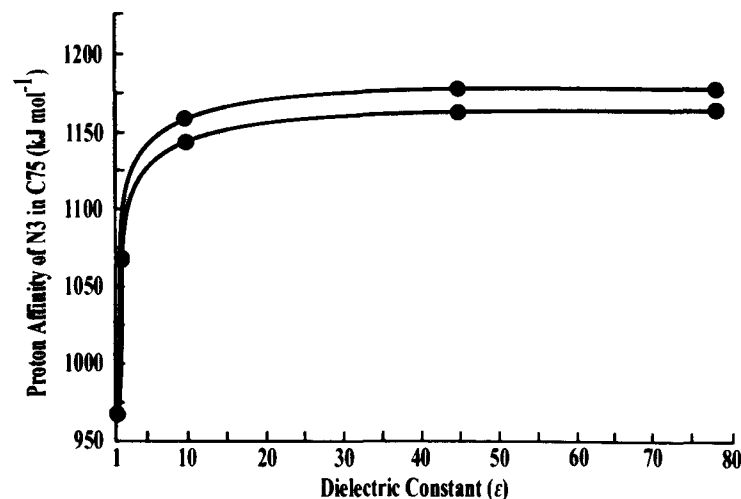


Figure 5.1 Plot of the effect of increasing the polarity (ϵ) of the environment on the proton affinity of the N3 centre of cytosine using the PB-SCRF (red circles) and IEF-PCM (green circles) methods. Common non-polar ($\epsilon = 1$) value indicated by a black circle.

For both of the solvation methods PA_{N3} dramatically increases when the dielectric constant is raised from 1.0 to 2.0. Specifically, the PB-SCRF and IEF-PCM approaches give increases of 102.7 and 102.6 kJ mol^{-1} to 1067.1 and 1067.0 kJ mol^{-1} , respectively. Both again calculate PA_{N3} to increase considerably, though to a lesser extent, upon raising ϵ further to 10.0; the PB-SCRF and IEF-PCM methods now giving increases of 74.6 and 80.5 kJ mol^{-1} to 1141.7 and 1147.5 kJ mol^{-1} respectively. For $\epsilon = 45.0$ and 78.39 both methods predict much more moderate further increases in PA_{N3} . Indeed, as can be seen in **Figure 5.1** they are now essentially converged to values of about 1156 kJ mol^{-1} (PB-SCRF) and 1165 kJ mol^{-1} (IEF-PCM). It is noted, however, that despite these large increases, for all ϵ PA_{N3} is still decidedly lower than that of the above model 2'-O⁻ ribose anion which, for example, using IEF-PCM with $\epsilon = 4.0$ and 78.39 is calculated to be 1429.5 and 1295.0 kJ mol^{-1} respectively.

Table 5.1 Proton Affinity (kJ mol^{-1}) of Cytosine's N3 Ring Centre in Environments of Increasing Dielectric Constant (ϵ) Modeled Using Various Solvation Methods.

Solvation Method	Dielectric Constant (ϵ)				
	1.0	2.0	10.0	45.0	78.39
PB-SCRF ^a	964.4	1067.1	1141.7	1154.6	1156.4
IEF-PCM ^b	964.4	1067.0	1147.5	1162.8	1164.7

^a Poisson-Boltzmann Self-Consistent Reaction Field Method. ^b Integral Equation Formalism Polarizable Continuum Method.

Bases were then hydrogen bonded to cytosine via its exocyclic amine as shown in **Figure 5.2** in order to consider their ability to influence the affinity of the N3 centre for a proton. Both H₂O ($PA^{35} = 681.5 \text{ kJ mol}^{-1}$) and NH₃ ($PA^{35} = 847.0 \text{ kJ mol}^{-1}$) were used so as to also consider the effects of varying the strength of the base. In addition, the influence of a polar environment ($\epsilon = 4.0$) on such interactions was also included using

the above solvation methods. The results obtained are shown in **Figure 5.2** and summarized in **Table 5.2**.

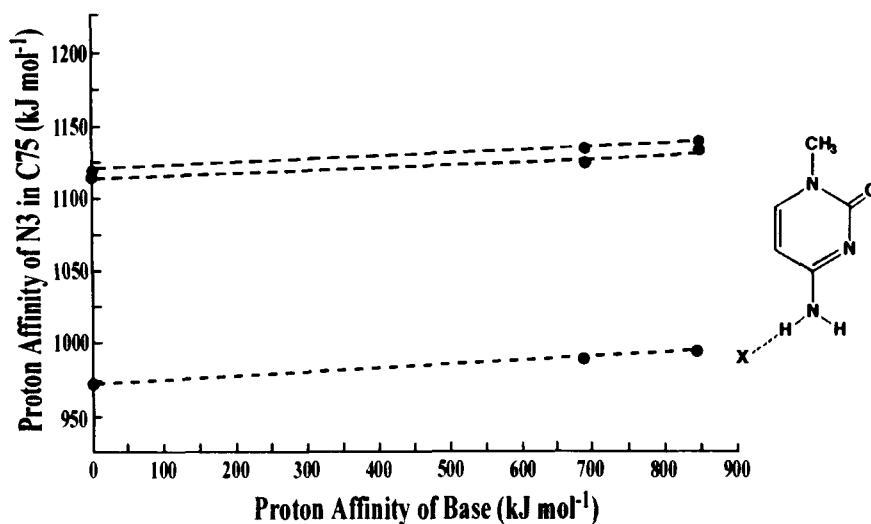


Figure 5.2 Plot of the proton affinity of cytosine's N3 centre with various bases hydrogen bonded to its *trans* amine group hydrogen in the gas-phase (black circle) and in polar environments ($\epsilon = 4.0$) modeled using the PB-SCRF (red circle) and IEF-PCM (green circle) methods.

Table 5.2 Proton Affinity (kJ mol^{-1}) of Cytosine's N3 Ring Centre with Various Bases Hydrogen Bonded to Cytosine's Exocyclic Amine as in Figure 5.2, in the Gas-Phase and Polar Environments ($\epsilon = 4.0$) as Modeled Using PB-SCRF and IEF-PCM Solvation Methods.

Solvation Method	Base		
	None	H ₂ O	NH ₃
Gas-Phase ($\epsilon = 1.0$)	964.4	998.6	1006.1
PB-PCM ^a	1114.7	1122.8	1130.2
IEF-PCM ^b	1117.5	1128.9	1134.8

^a Poisson-Boltzmann Polarizable Self-Consistent Reaction Field Method. ^b Integral Equation Formalism Polarizable Continuum Method.

In a non-polar environment ($\epsilon = 1.0$) PA_{N3} increases by 34.2 and 41.7 kJ mol^{-1} when H_2O and NH_3 , respectively, hydrogen bond to cytosine as in **Figure 5.2** (see also **Table 5.2**). However, both solvation methods predict a polar environment to moderate the effects of hydrogen bonding on cytosine's N3 proton affinity. Indeed, the PB-SCRF and IEF-PCM methods predict H_2O to now increase PA_{N3} by just 8.1 and 11.4 kJ mol^{-1} while for NH_3 the observed increases are 15.5 and 17.3 kJ mol^{-1} , respectively. As noted above, it has been suggested^{8,13} that the cytosine may be hydrogen bonded to a phosphate anion via its exocyclic amine as shown in **Figure 5.2**. The resulting ionic hydrogen bond formed would be expected to have an even greater effect on the affinity of cytosine's N3 centre for a proton than either H_2O or NH_3 . Hence, PA_{N3} also was calculated using IEF-PCM and $\epsilon = 4.0$ when the model phosphate anion $(\text{CH}_3\text{O})_2\text{PO}_2^-$ was hydrogen bonded to the cytosine. The resulting value of 1208.1 kJ mol^{-1} , is indeed higher than with the above neutral bases. However, it is still markedly lower than that of the 2'- O^- ribose anion (see above). This appears to suggest that regardless of any active site interactions or environmental effects, C75 is unlikely to be able to act as a base that simply deprotonates the 2'-OH group of the ribose sugar to initiate phosphodiester bond cleavage as has been previously suggested.^{8,13-15}

5.3.2 The Hydrated Mg^{2+} Ion Species

As described in the introduction, it has been suggested that a hydrated Mg^{2+} may in fact act as the catalytic base. Unfortunately, the exact coordination environment of the Mg^{2+} ion itself is unknown, but several possibilities have been suggested. Bevilacqua and coworkers¹⁸ found that the ribozyme is inhibited by $[\text{Co}(\text{NH}_3)_6]^{3+}$ even in the presence of Mg^{2+} . Hence, they suggested that the ' Mg^{2+} moiety' is outer-sphere coordinated to the active site and is deprotonated, i.e., forms $[\text{Mg}(\text{H}_2\text{O})_5\text{OH}]^+$, thus enabling it to act as a base. A recent theoretical gas-phase study,³⁶ however, suggested that deprotonating $[\text{Mg}(\text{H}_2\text{O})_6]^{2+}$ causes the resultant ion to become 5-coordinate, i.e., $[\text{Mg}(\text{H}_2\text{O})_4\text{OH}]^+$

Further experiments by Bevilacqua and coworkers^{37,38} suggested that there are two types of metal ion sites, one structural and the other catalytic, that are inner- and outer-sphere coordinated respectively. However, more recent crystal structures of the precursor form of the ribozyme^{15,34} suggest that the catalytic Mg^{2+} species is in fact outer-sphere coordinated to both the 5'-O and a non-bridging oxygen of the scissile phosphate, but inner-sphere coordinated to a non-bridging phosphate oxygen of U23. Thus, two models were chosen in order to consider the basicity of possible hydrated Mg^{2+} species which correspond to outer- and inner-sphere coordination to the active site phosphate of U23: $[\text{Mg}(\text{H}_2\text{O})_5\text{OH}]^+$ and $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$. Furthermore, the surrounding environment was included using the IEF-PCM method with $\epsilon = 4.0$.

Unlike that previously observed in the gas-phase,³⁶ within a polar environment the $[\text{Mg}(\text{H}_2\text{O})_5\text{OH}]^+$ ion is found to be stable. Furthermore, its ligated OH^- group has a reasonably high affinity for a proton of $1099.0 \text{ kJ mol}^{-1}$. However, due in part to greater electron donation to the Mg^{2+} centre by the anionic phosphate compared to H_2O , the OH^- ligand of the inner-sphere coordinated $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$ species has an even higher affinity of $1188.7 \text{ kJ mol}^{-1}$. This is slightly lower than that calculated for the cytosine's N3 centre for the complex when the phosphate anion was hydrogen bonded to the cytosine. Interestingly, the ordering is reversed if one increases the polarity of the environment. For instance, using the IEF-PCM method at $\epsilon = 10.0$ gives the affinities of $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$ and $[\text{Cytosine}\cdots(\text{CH}_3\text{PO}_4\text{CH}_3)^-]$ to be 1225.4 and $1189.4 \text{ kJ mol}^{-1}$ respectively. Regardless, however, the above results appear to suggest that the most basic Mg^{2+} species considered above will be unable to simply deprotonate the sugar's 2'-OH group. However, this does not preclude the possibility that it acts in concert with another active site component or via some proton shuttle mechanism.

Cleavage of the phosphodiester backbone of RNA does occur naturally though at an appreciably slower rate. Indeed, it has been reported that the half-life of solvent catalyzed (H_2O , H^+ , OH^-) or free solution hydrolysis of RNA is about 110 years at pH 7 and 25°C in the absence of divalent metal ions.³⁹ In order to gain further insights into the catalytic effects of the HDV ribozyme, we began by considering uncatalyzed phosphodiester cleavage.

5.3.3 The Uncatalyzed Mechanism

Isolated phosphodiester cleavage has been previously studied though not with the same chemical model and level of theory employed in this present study. For example, Lopez and coworkers⁴⁰ recently performed an extensive computational study on the reverse phosphodiester hydrolysis reaction. Two isomerically related mechanisms were identified and classified as *exo* or *endo* as determined by the orientation of the protonated phosphate oxygen with respect to the ribose ring. Herein, both pathways are considered. Within the context of the HDV ribozyme, the retained sugar is modeled by the $-\text{OCH}_3$ group attached to the phosphorus centre. Similar models have been used in previous computational studies on related reactions.^{27,28,30,40} The schematic potential energy profiles for the *exo* and *endo* mechanisms in solution are shown in **Figure 5.3**. Mechanistically the two pathways are the same with the *exo* reaction being slightly lower in energy. Thus, for simplicity the *endo* reaction is not discussed unless otherwise noted.

Cleavage of the phosphodiester bond begins with nucleophilic attack of the leaving sugar's 2'-oxygen at the phosphorus centre with concomitant transfer of the 2'-OH proton to an equatorial oxygen of the phosphate moiety. This reaction proceeds via transition structure (TS) **2a** at a cost of $118.2 \text{ kJ mol}^{-1}$ to give the high energy phosphorane intermediate **3a** lying 99.7 kJ mol^{-1} higher in energy than the initial reactant **1a**. We note that the *endo* **2a** lies marginally lower in energy by 0.3 kJ mol^{-1} and is the only species

for which the *endo* isomer is lower in energy. The next step in either pathway is rotation about the newly formed P—OH group such that its proton now interacts with the 5'-oxygen of the 'retained sugar'. This proceeds via TS **4a** at a cost of just 33.0 kJ mol⁻¹ for the *exo* reaction to give the slightly higher energy (112.6 kJ mol⁻¹) phosphorane intermediate **5a**. In the third and final step the P—OH proton of **5a** is transferred to the 5'-oxygen of the 'retained sugar' resulting in cleavage of the 5'-O—P bond. This occurs via TS **6a** at a cost of 29.9 kJ mol⁻¹ with respect to **5a** and an overall barrier of 142.5 kJ mol⁻¹ with respect to **1a**. It is noted that this step has a lower barrier of 22.3 kJ mol⁻¹ in the *endo* pathway, however, with respect to **1a** it has an even higher barrier of 144.6 kJ mol⁻¹. This is also the overall rate-limiting step in both isomeric pathways. The final product complex in which the resultant 2',3'-cyclic phosphate ester is hydrogen bonded to the now released methanol, **7a**, lies marginally higher in energy than the initial reactant **1a** by just 0.9 kJ mol⁻¹. The separated products (**8a**) lie slightly higher in energy at 19.1 kJ mol⁻¹.

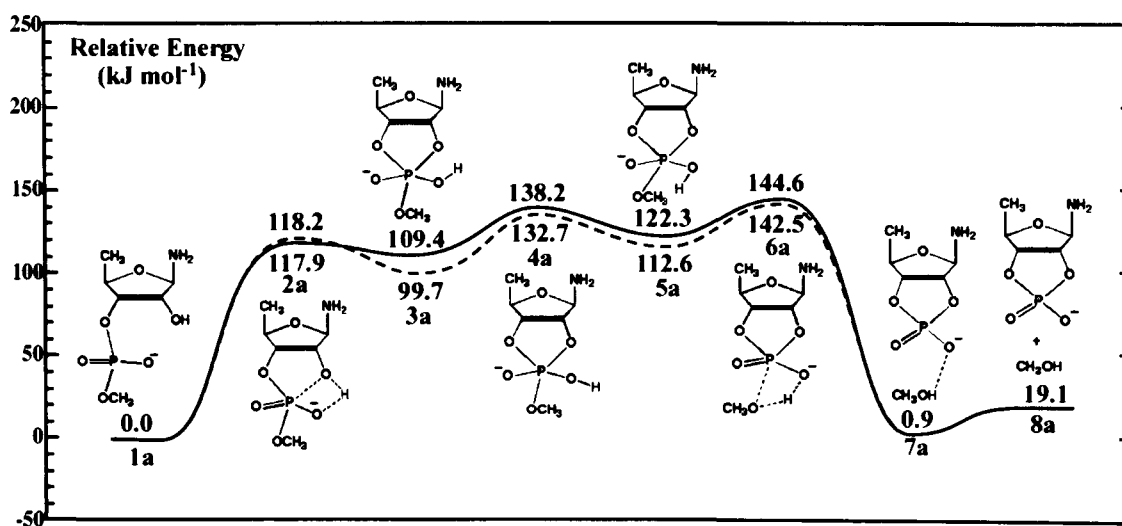


Figure 5.3 Schematic potential energy surfaces for the isomeric *endo* (solid blue) and *exo* (dashed red) uncatalyzed phosphodiester cleavage mechanisms in solution (see text).

5.3.4 The Catalyzed Mechanism

The active site model (**Scheme 5.3**) was chosen based upon previous experimental studies and incorporating the above results. The substrate is the same as in the uncatalyzed mechanism while the proposed^{15,34} inner-sphere U23 phosphate coordinated Mg^{2+} species, herein modeled as $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$, was used. It has been experimentally suggested^{37,41-44} that C75's exocyclic amine can interact with a non-bridging phosphate oxygen and that its N3 centre is within hydrogen bonding distance of the retained sugar's 5'-oxygen.^{8,13} Thus, in the present model the cytosine was protonated and allowed to interact with the scissile phosphate in a similar manner. Key optimized structures and the schematic potential energy surface for the HDV ribozyme catalyzed mechanism are shown in **Figures 5.4** and **5.5** respectively.

In the optimized structure of the initial reactant complex **1b** the 'leaving' ribose sugar was rotated thus enabling its 2'-OH group to form a short hydrogen bond (1.397 Å) with the -OH ligand on the Mg^{2+} centre. This strong interaction also causes a lengthening of the 2'-O—H bond to 1.072 Å. It should be noted that this arrangement is consistent with previous proposals that a conformational change, specifically a rotation of the sugar, occurs to initiate the catalytic mechanism.¹⁵ Two waters ligated to the Mg^{2+} also form moderately strong hydrogen bonds with the substrate (see **Figure 5.4**); one with the 2'-oxygen of the ribose sugar (1.675 Å) and the other with a non-bridging oxygen of the scissile phosphate (1.610 Å). The protonated cytosine's N3—H moiety forms a moderately strong hydrogen bond (1.712 Å) with the leaving groups 5'-O while its exocyclic amine forms a strong hydrogen bond (1.373 Å) with the remaining non-bridging phosphate oxygen.

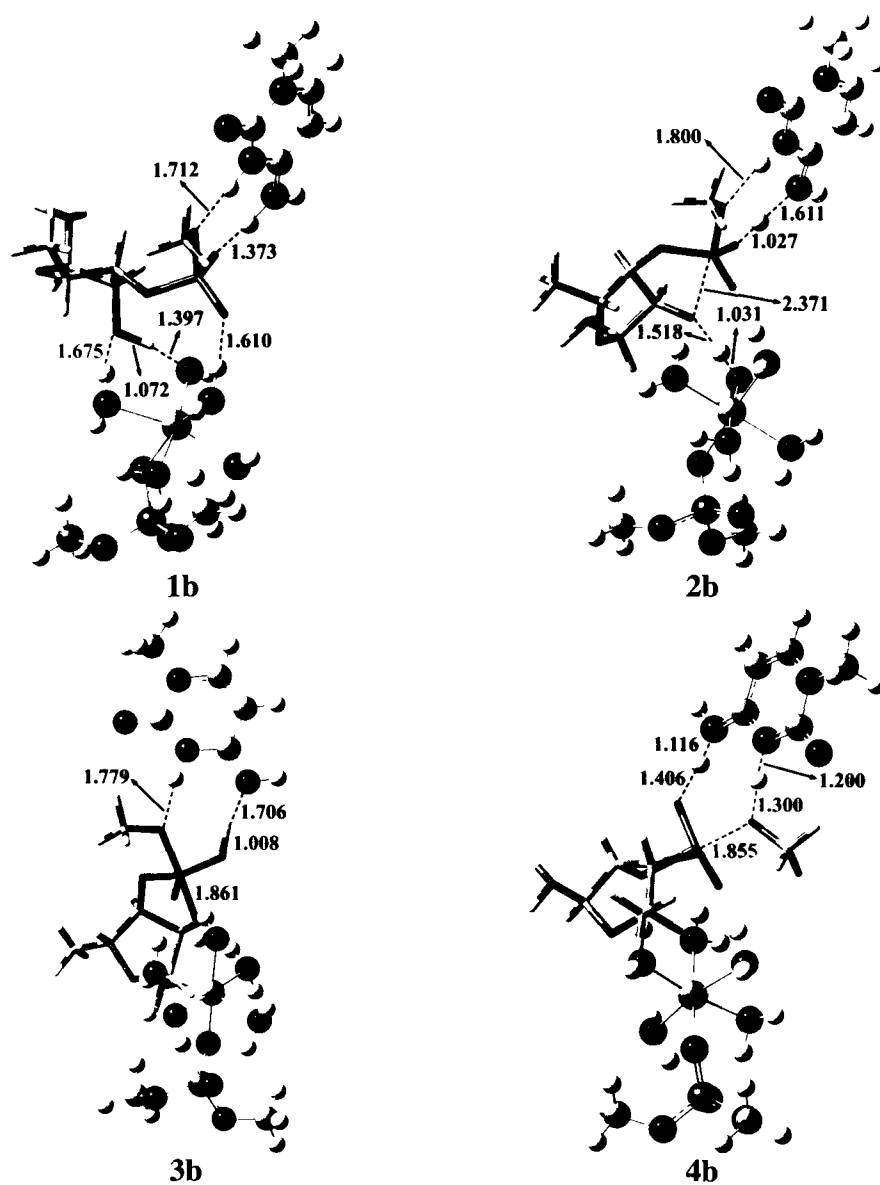


Figure 5.4 Optimized structures showing selected interaction distances (Å) for the (1b) initial reactant complex, (2b) transition structure to formation of the intermediate, (3b) reaction intermediate and (4b) transition structure to formation of the final products in the HDV ribozyme catalyzed phosphodiester cleavage mechanism.

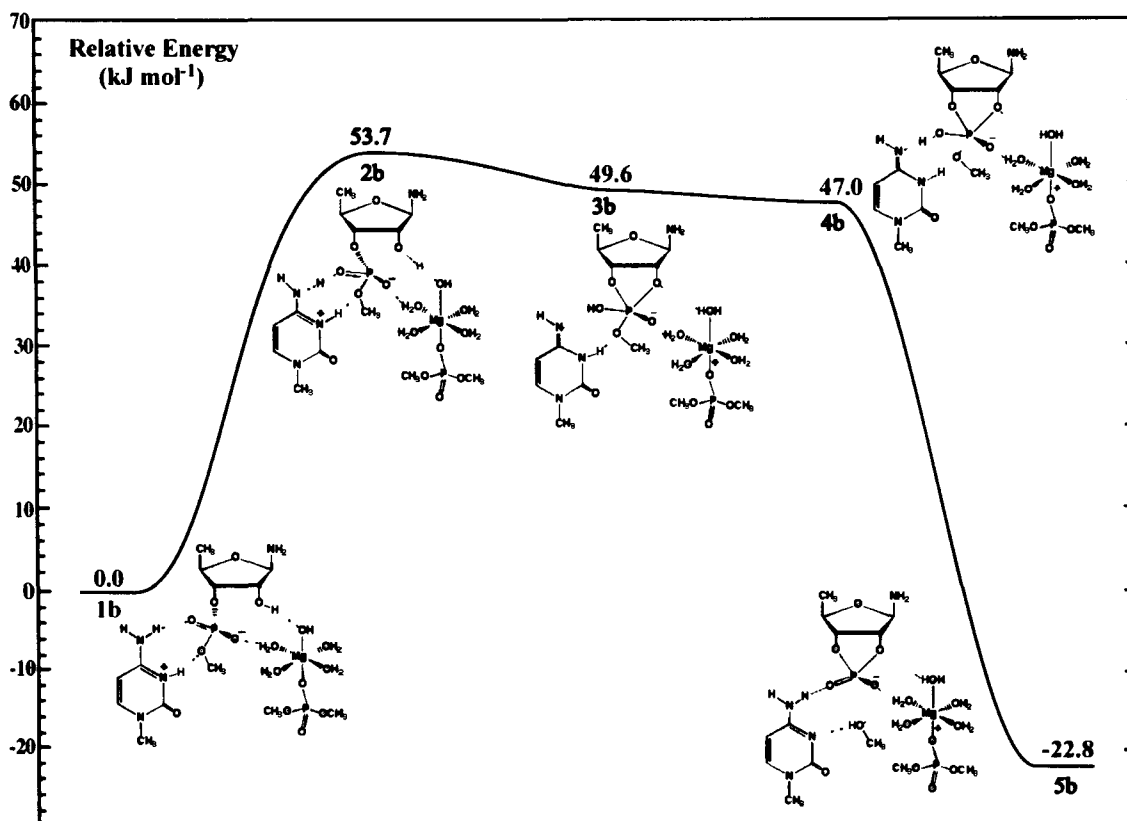


Figure 5.5 Schematic potential energy surface for the catalyzed mechanism of the HDV ribozyme in solution. Formal charges are indicated.

As in the uncatalyzed mechanism, the first step in the reaction is nucleophilic attack of the 2'-oxygen at the phosphorus centre. However, now it proceeds via TS **2b** at a much reduced cost of 53.7 kJ mol^{-1} (**Figure 5.5**). In addition, the 2'-OH proton is concomitantly transferred to the Mg^{2+} ligated $-\text{OH}$ while the protonated cytosine transfers a proton from its exocyclic amine to a non-bridging phosphate oxygen. We note that this P—O corresponds to the *exo* oxygen in the uncatalyzed mechanism. Furthermore, the protonated cytosine has now formed the neutral tautomer of cytosine. This step is also found to be the rate limiting step in the overall solution reaction. Interestingly, previous experimental studies^{15,18,43-45} have suggested a similar possible role for tautomerization of the active site cytosine C75. In contrast to that observed for the uncatalyzed mechanism the resulting phosphorane intermediate (**3b**) now sits just 49.6 kJ mol^{-1} higher in energy

than the initial reactant complex **1b**. It is stabilized by hydrogen bonds between its now bridging 2'-O and the Mg^{2+} bound H_2O derived from the initial $-\text{OH}$ ligand, and the non-bridging phosphate $\text{P}-\text{O}^-$ moiety and a second Mg^{2+} bound H_2O (**Figure 5.4**). The tautomeric cytosine also helps to stabilize the phosphorane by forming a moderately strong hydrogen bond (1.706 Å) between the nitrogen centre of its exocyclic imine and the $\text{PO}-\text{H}$ moiety while its $\text{N3}-\text{H}$ group remains hydrogen bonded to the substrate's bridging 5'-oxygen.

The second and final step is cleavage of the substrate's 5'-O—P bond leading to formation of the final products. This step proceeds via TS **4b** with essentially no barrier with respect to **3b** to give the hydrogen bonded product complex **5b** lying 22.8 kJ mol^{-1} lower in energy than **1b**. Interestingly, in TS **4b** the tautomeric cytosine's imine group deprotonates the $\text{P}-\text{OH}$ moiety of **3b**. Simultaneously, however, its $\text{N3}-\text{H}$ group protonates the 5'-oxygen thus cleaving the 5'-O—P bond. Throughout this step the hydrated Mg^{2+} ion plays a more passive role, remaining hydrogen bonded to the bridging 2'-oxygen and the $\text{P}-\text{O}^-$ moiety, thus helping to stabilize the negative charge.

Thus, this overall mechanism is a two-step reaction in which the $[\text{Mg}(\text{OH})(\text{H}_2\text{O})_4(\text{U23 phosphate})]$ is able to act as a base to deprotonate the ribose's 2'-OH group due to *cooperation* with the protonated cytosine which also helps to stabilize the negative charge on the phosphate. The resulting cytosine tautomer then acts as both an acid and base to cleave the desired 5'-O—P bond in the second step.

5.4 Conclusions

Density functional theory methods have been employed in order to undertake a detailed systematic investigation on the HDV ribozyme and its proposed catalytic

mechanism. Specifically, we considered the: (i) effect of a polar environment and hydrogen bonding, separately or combined, on cytosine's N3 ring centre proton affinity (PA_{N3}); (ii) basicity of various previously suggested hydrated Mg^{2+} bases; and (iii) uncatalyzed phosphodiester bond cleavage and the proposed HDV ribozyme catalyzed phosphodiester bond cleavage mechanism in which the active site cytosine acts as a general acid or base.

Both PB-SCRF and IEF-PCM methods suggest that a polar environment can considerably increase the affinity of cytosine's N3 ring centre for a proton (PA_{N3}). In low polarity environments (e.g., $\epsilon = 2.0$) the PB-SCRF and IEF-PCM methods predict similar environmental effects on PA_{N3} . However, as ϵ is increased the IEF-PCM predicts modestly larger effects. For example, PA_{N3} is calculated to be 1156.4 and 1164.7 kJ mol^{-1} at $\epsilon = 78.39$ with the PB-SCRF and IEF-PCM methods, respectively.

The effects of bases (H_2O , NH_3 and $(CH_3O)_2PO_2^-$) hydrogen bonding to cytosine's exocyclic amine were also considered. It was found that while such an interaction does indeed increase PA_{N3} , though not as effectively as the polarity of the environment itself, its effect is tempered by a polar environment. The present results suggest that neither the environment nor hydrogen bonding, individually or combined, are able to enhance the affinity of the active sites cytosine N3 centre such that it can directly deprotonate the 2'-OH of the leaving ribose sugar.

The basicity of various suggested hydrated and coordinated $Mg(OH)^+$ species was also examined. Specifically, the species $[Mg(H_2O)_5OH]^+$ and $[Mg(H_2O)_4(CH_3PO_4CH_3)OH]$ were used in order to model outer- and inner-sphere active site coordination via a phosphate oxygen of U23, respectively. The latter species was

found to be the more basic of the two. However, as for cytosine, this is still decidedly lower than that of the relevant 2'-O⁻ ribose anion.

An appropriate chemical model was then chosen in order to investigate the proposed HDV ribozyme catalyzed phosphodiester cleavage. The reaction pathway was found to have the active site cytosine (C75) and the species [Mg(H₂O)₄(CH₃PO₄CH₃)OH] interact with different non-bridging phosphate oxygen and act as the general acid and base, respectively. As in the uncatalyzed mechanism the initial step involves nucleophilic attack of the 2'-oxygen at the phosphorus centre to give a phosphorane intermediate. However, this now occurs at a cost of just 53.7 kJ mol⁻¹ and concomitantly involves transfer of the 2'-OH proton to the Mg²⁺ bound OH⁻ and protonation of a non-bridging phosphate oxygen by C75's exocyclic amine. Furthermore, the present results suggest that both active site components are necessary for this initial step and that the neutral cytosine tautomer is formed. In the final step the cytosine tautomer acts in a proton shuttle manner: deprotonating the phosphoranes P—OH moiety while simultaneously protonating the 5'-oxygen at a cost of just 47.0 kJ mol⁻¹, cleaving the 5'-O—P bond to give the final products.

References

- (1) Kuimelis, R. G.; McLaughlin, L. W. *Chem. Rev.* **1998**, *98*, 1027.
- (2) Doherty, E. A.; Doudna, J. A. *Annu. Rev. Biochem.* **2000**, *69*, 597.
- (3) Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001**, *29*, 1815.
- (4) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.
- (5) Macnaughton, T. B.; Shi, S. T.; Modahl, L. E.; Lai, M. M. C. *J. Virol.* **2002**, *76*, 3920.

- (6) Steitz, T. A.; Moore, P. B. *Trends Biochem. Sci.* **2003**, *28*, 411.
- (7) Taylor, J. M. *Virology* **2006**, *344*, 71.
- (8) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567.
- (9) Canny, M. D.; Jucker, F. M.; Kellogg, E.; Khvorova, A.; Jayasena, S. D.; Pardi, A. *J. Am. Chem. Soc.* **2004**, *126*, 10848.
- (10) Zhao, Z. Y.; McLeod, A.; Harusawa, S.; Araki, L.; Yamaguchi, M.; Kurihara, T.; Lilley, D. M. J. *J. Am. Chem. Soc.* **2005**, *127*, 5026.
- (11) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferre-D'Amare, A. R. *Science* **2002**, *298*, 1421.
- (12) Han, J.; Burke, J. M. *Biochemistry* **2005**, *44*, 7864.
- (13) Ferre-D'Amare, A. R.; Doudna, J. A. *J. Mol. Biol.* **2000**, *295*, 541.
- (14) Perrotta, A. T.; Shih, I.; Been, M. D. *Science* **1999**, *286*, 123.
- (15) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H. D.; Doudna, J. A. *Nature* **2004**, *429*, 201.
- (16) Krasovska, M. V.; Sefcikova, J.; Spackova, N.; Sponer, J.; Walter, N. G. *J. Mol. Biol.* **2005**, *351*, 731.
- (17) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. *Biophys. J.* **2006**, *91*, 626.
- (18) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493.
- (19) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45.
- (20) Strobel, S. A. *Nat. Chem. Biol.* **2005**, *1*, 5.
- (21) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin,

- A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.
- (22) Jaguar 5.5, Schrodinger, L. L. C., Portland, OR, 1991-2003.
- (23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (24) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (25) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (26) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.
- (27) Liu, Y.; Lopez, X.; York, D. M. *Chem. Commun.* **2005**, 3909.
- (28) Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 19987.
- (29) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. *Biochemistry* **2006**, *45*, 10043.
- (30) Torres, R. A.; Himo, F.; Bruice, T. C.; Noodleman, L.; Lovell, T. *J. Am. Chem. Soc.* **2003**, *125*, 9861.
- (31) Leclerc, F.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 3395.
- (32) For recent reviews, see: (a) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459. (b) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232 and references therein.
- (33) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (34) Tinsley, R. A.; Harris, D. A.; Walter, N. G. *Biochemistry* **2004**, *43*, 8935.

- (35) Relative energies obtained at the B3LYP/6-311+G(d,p)//B3LYP/6-311G(d,p) level of theory.
- (36) Kluge, S.; Weston, J. *Biochemistry* **2005**, *44*, 4877.
- (37) Nakano, S.; Proctor, D. J.; Bevilacqua, P. C. *Biochemistry* **2001**, *40*, 12022.
- (38) Nakano, S.; Cerrone, A. L.; Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982.
- (39) Williams, N. H.; Takasaki, B.; Wall, M.; Chin, J. *Acc. Chem. Res.* **1999**, *32*, 485.
- (40) Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 11525.
- (41) Shih, I.; Been, M. D. *Annu. Rev. Biochem.* **2002**, *71*, 887.
- (42) Luptak, A.; Ferre-D'Amare, A. R.; Zhou, K.; Zilm, K. W.; Doudna, J. A. *J. Am. Chem. Soc.* **2001**, *123*, 8447.
- (43) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* **2004**, *73*, 90.
- (44) Oyelere, A. K.; Kardon, J. R.; Strobel, S. A. *Biochemistry* **2002**, *41*, 3667.
- (45) Nakano, S.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2001**, *123*, 11333.

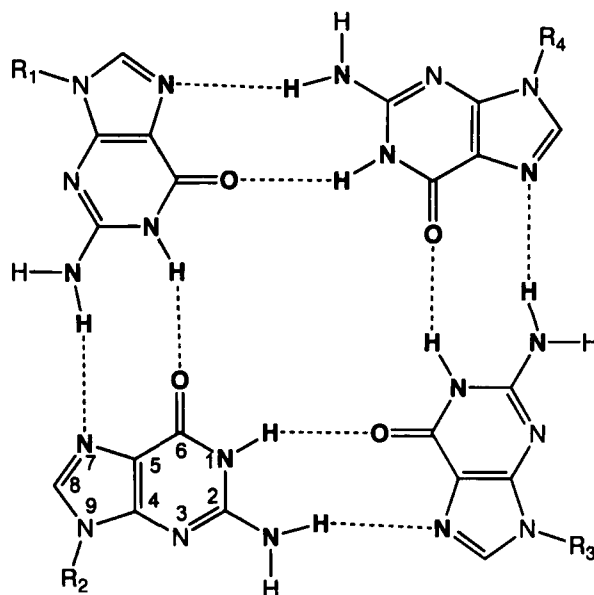
Chapter 6

Protonation of Guanine Quartets and Quartet Stacks: Insights from DFT Studies

6.1 Introduction

Nucleobases are now known to be able to self-assemble into highly organized planar structures.^{1,2} Indeed, high-order polyads such as pentads and hexads have been observed.³⁻⁷ Of particular interest have been quartets, especially those composed solely of guanine (**Scheme 6.1**). This is due in part to the fact that they have been observed *in vivo*,^{2,8-14} typically as multi-layer stacks or quadruplexes. For example, crystalline guanine plates have been found in some deep sea-fish and spiders.⁹⁻¹¹ In humans, the guanine-rich single stranded telomere ends of DNA are thought to be able to ‘wrap’ into guanine quartets and quadruplexes.¹²⁻¹⁴ Indeed, the formation of such structures is believed to regulate the activity of telomerase, a DNA repair enzyme most commonly found in cancerous cells.¹⁵⁻¹⁷ Thus, due in large part to the diverse and important roles of guanine quartets (G4’s) and quadruplexes (G4_n’s), there is considerable interest in obtaining a greater understanding of their chemistry.

Scheme 6.1 A guanine quartet bound via Hoogsteen hydrogen bonds (R_{1-4} = DNA strands).



In particular, it has long been known that $G4_n$'s are stabilized by alkali metal ions such as K^+ and Na^+ via coordination with the guanine oxygens (O6's).^{18,19} As a result, the interactions of $G4_n$'s with a range of cationic species have been experimentally investigated.²⁰⁻²⁶ For example, using both X-ray crystallography and NMR techniques, Strobel and co-workers²⁰ found that Tl^+ can promote the formation and stabilization of $G4_n$'s in the same manner as K^+ . Similarly, using extended X-ray adsorption fine structure (EXAFS) methods, Shafer and co-workers²¹ found that Pb^{2+} can also bind with $G4_n$'s by sitting between two $G4$'s via coordination with all eight O6's. Furthermore, some cationic-²² and copper-porphyrins,²³ nickel(II)-²⁴ and platinum(II)²⁵-based complexes have also been shown to be able to bind and stabilize $G4_n$'s.

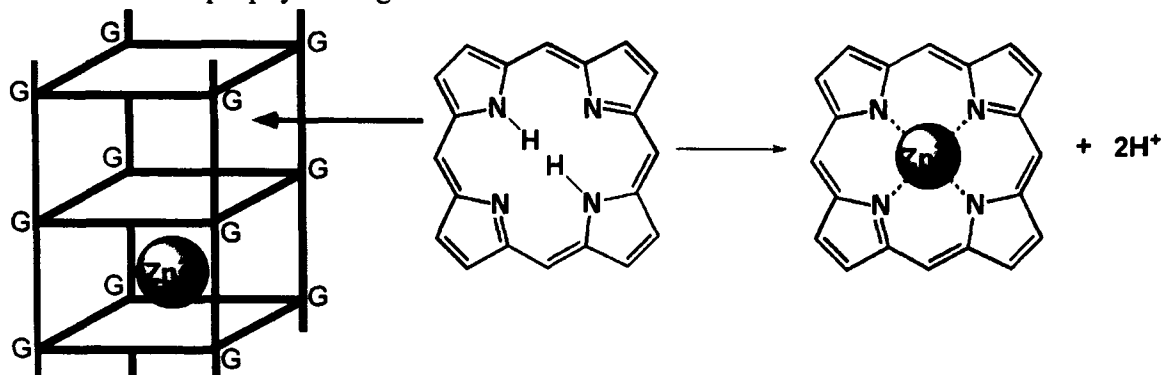
In addition to experimental studies, there have also been a number of computational studies on the interactions of $G4$'s and $G4_n$'s with cations, in particular various metal ions.²⁶⁻³¹ For example, Leszczynski and co-workers²⁷ used both Hartree-Fock (HF) and

density functional theory (DFT) methods to investigate the preferential coordination of K^+ over Na^+ and Li^+ to G4's while Meyer and co-workers²⁸ employed DFT methods to examine the complexation of a range of metal ions with guanine and isoguanine quartets. In addition, the interactions of Na^+ and K^+ with a range of two-plane nucleobase quartet stacks have been studied by several groups using HF and DFT methods.^{28b,29} Furthermore, the interactions of metal ions with G4_n's have also been investigated using molecular dynamics methods.^{30e,30} In each of these studies, for G4's the metal ions were found to generally sit in the centre and interact with all four O6's while for G4_n's, the metal ions may also sit between planes and coordinate with all eight O6's.

The simplest yet arguably most fundamentally important cation within biochemical systems is the proton.³¹ Indeed, the structure and hence proper functioning of many important biomolecules such as proteins and nucleic acids are often strongly influenced by their protonation states.^{31,32} It is now known that some nucleobase quartets are in fact stabilized by protonation. For example, i-motifs are formed at low pH by intercalation of two cytosine duplexes, with one cytosine in each pair protonated.^{33,34} Furthermore, G4_n's are also found to be able to form at low pH.³⁵ Unfortunately, however, the resulting structures are not known. Intriguingly, it has been shown that G4_n's can catalyse the insertion of metal ions, specifically zinc, into porphyrin rings.³⁶ Sugimoto and co-workers³⁷ proposed that the Zn^{2+} initially binds between two G4's while the porphyrin intercalates between the next two (Scheme 6.2). The Zn^{2+} then passes through the dividing G4 and inserts into the porphyrin ring. They further proposed that the G4_n simply provides a 'suitable reaction field' However, it is unclear as to where the two porphyrin protons are at least initially transferred. Interestingly, ferrochelatase is an enzyme that is also able to catalyze porphyrin metallation.^{38,39} It has been suggested that a histidine residue in their active sites may help deprotonate the porphyrin.³⁸ Relating this to the G4_n-catalyzed reaction raises the intriguing questions as to whether a G4 may also

act as a base to abstract or accept one or both porphyrin protons and furthermore what are the possible resulting structures.

Scheme 6.2 Schematic overview of the proposed mechanism³⁷ for G4_n's catalyzed Zn²⁺ insertion into a porphyrin ring.



In order to gain insights into the effects of protonation and the possible ability of G4's to act as catalytic bases or proton acceptors, DFT methods have been used in this chapter to examine the effects of up to quadruple protonation on the structure and properties of a guanine quartet and a two-plane guanine quartet stack.

6.2 Computational Methods

All gas-phase geometry optimization and single point calculations were performed using Jaguar 5.5⁴⁰ and Gaussian 03.⁴¹ The DFT method B3LYP, a combination of Becke's three-parameter hybrid exchange functional^{42,43} and the correlation functional of Lee, Yang and Parr,⁴⁴ was used throughout the study. It is noted that the B3LYP functional has been used previously to investigate structures and properties of G4's, 2G4's and related nucleobase quartets.²⁸⁻³⁰ For G4's, optimized geometries were obtained at the B3LYP/LACVP** level of theory. Relative energies were obtained by performing single point calculations at the B3LYP/6-311+G(2df,p) level on the above geometries,

i.e., B3LYP/6-311+G(2df,p)//B3LYP/LACVP** The sites at which each guanine would be attached to the DNA backbone were replaced by a methyl group.

For the two-plane guanine quartet stack (**2G4**), due to the size of the model, optimized geometries were obtained at the B3LYP/LACVP* level. Relative energies were obtained by performing single point calculations at the B3LYP/6-311+G(d,p) level on the above geometries, i.e., B3LYP/6-311+G(d,p)//B3LYP/LACVP* The **2G4** model was derived from the outer-most quartet-stack in the crystal structure PDB: 1K8P. They represent the two planes most exposed to the solution environment and thus include the **G4** most likely to be initially protonated and the adjacent **G4**. The sites attached to the DNA backbone were replaced by H, with N9 of each guanine held fixed at their crystal structure positions. Topological analyses were performed using the AIM2000 program.⁴⁵

6.3 Results and Discussion

6.3.1 Unprotonated *G4*'s

Unprotonated **G4**'s have been previously studied in detail using a variety of computational methods.⁴⁶ Thus, only those of most relevance to the protonated complexes investigated in this chapter are herein briefly described. Similar to that previously observed,^{28a,47a} in the absence of cations **G4**'s can be bound via either bifurcated or Hoogsteen hydrogen bonds and lie close in energy. Hence, unless otherwise noted, only the lowest energy **G4**'s for each of these hydrogen bonding patterns obtained at the current level of theory is described below. All of the optimized structures, including the previously unreported *C_i* structure, are given in Table D.1 in Appendix D.

The overall preferred structure was found to be the planar bifurcated complex **G4₁** with C_{4h} symmetry, in which the O6 of each guanine forms hydrogen bonds with both the N1–H and exocyclic -NH₂ of the same adjacent guanine with lengths of 1.956 and 1.972 Å respectively (**Figure 6.1**). The AIM calculated values of the electron density (ρ) and Laplacian of the electron density ($\nabla^2\rho$) at the bond critical points (BCP's) of these two interactions are 0.025 and 0.074, and 0.023 and 0.080 a.u. respectively (Table D.2 in Appendix D). These all lie within the ranges for ρ (0.002–0.040 a.u.) and $\nabla^2\rho$ (0.024–0.139 a.u.) previously suggested as typical for hydrogen bonding interactions.^{47,48} Due to the fact that each oxygen now participates in two moderately strong hydrogen bonds, the C=O bond of each guanine moiety has lengthened from that observed in monomeric guanine (1.219 Å) to 1.242 Å (Table D.1 in Appendix D). The AIM topological analysis also identified a bond path between N7 of each guanine and an exocyclic -NH₂ hydrogen of the adjacent guanine (**Figure 6.1**). A previous study^{47c} also observed such an interaction and characterized it as a weak hydrogen bond. However, based on a recent re-examination of the AIM criteria⁴⁹ it is perhaps better characterized as an electrostatic interaction due to the large ellipticity (ϵ) at its BCP (1.067; Table D.2 in Appendix D).

The lowest energy Hoogsteen bound **G4** obtained was the non-planar S_4 symmetric complex **G4₂** lying just 3.0 kJ mol⁻¹ higher in energy than **G4₁** (see **Figure 6.1**). Unlike **G4₁**, each O6 now forms only one much shorter and stronger hydrogen bond (1.763 Å) with the N1–H group of an adjacent guanine while the N7 centre now forms a longer and weaker hydrogen bond (1.977 Å) with its exocyclic -NH₂ group. The relative strengths of these bonds are further illustrated by the larger value of ρ at the BCP of the former (0.032; Table D.2 in Appendix D) than the latter (0.017 a.u.; Table D.2 in Appendix D). The C=O bond length of each guanine moiety in **G4₂** (1.233 Å; Table D.1 in Appendix D) has again increased compared to that in monomeric guanine, although to a lesser extent

than that observed in G4₁. It is noted that as each O6 is involved in only one hydrogen bond, the central cavity of G4₂ is slightly smaller and more electron rich than that of G4₁ (Figure D.1 in Appendix D).

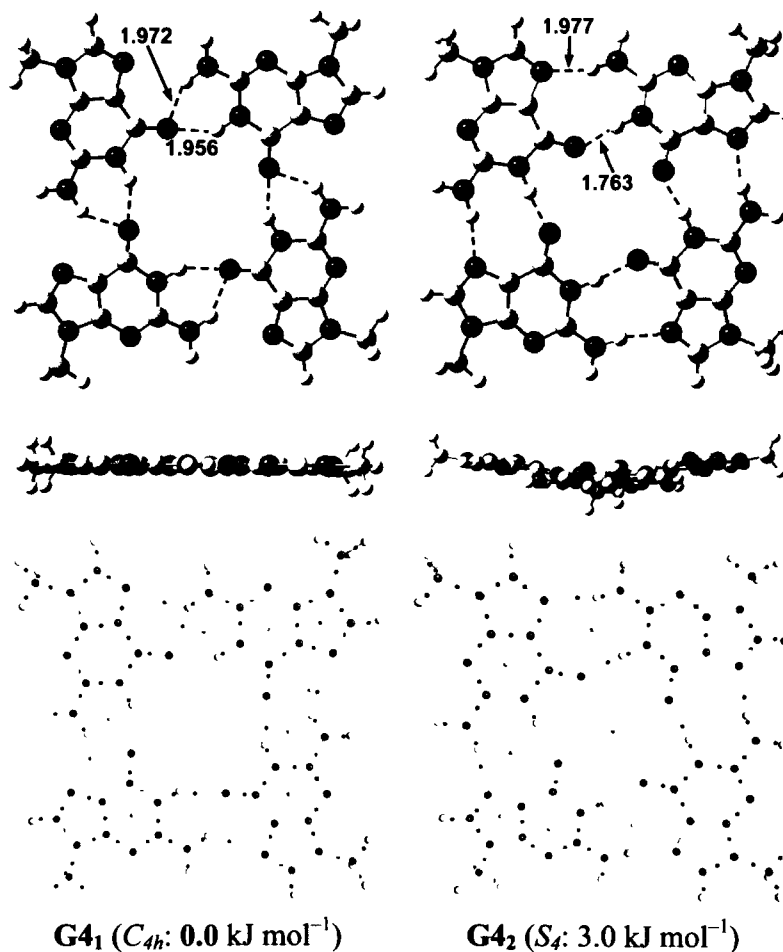


Figure 6.1 Optimized structures (top- and side-views) and molecular graphs for the lowest energy bifurcated and Hoogsteen hydrogen bonded neutral G4 structures. Symmetries and relative energies are also shown while selected optimized distances are in angstroms. Red and yellow dots in the molecular graph indicate BCP's and RCP's respectively. Atom colour key: C (gray); O (red); N (blue) and H (white).

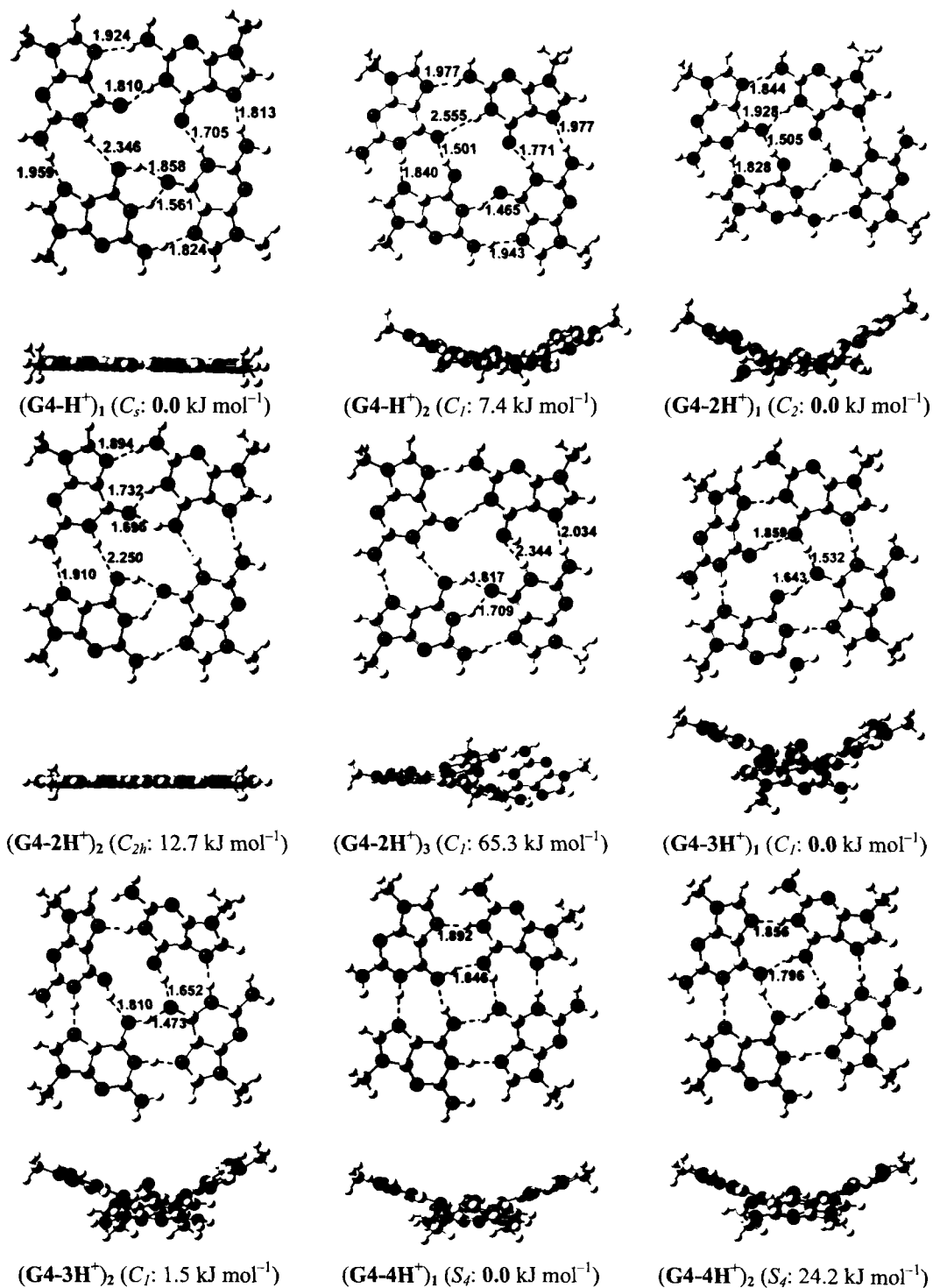


Figure 6.2 Optimized structures (top- and side-views) for $(G4-nH^+)$ $n = 1-4$. Symmetries and relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white).

6.3.2 Singly Protonated G4 [(G4-H⁺)]

Guanine's O6 has two non-equivalent lone-pairs. One participates in its Watson-Crick (WC) hydrogen bonding motif while the other partakes in its Hoogsteen motif. Their proton affinities (PA's) at the level of theory used herein are 943.5 and 979.6 kJ mol⁻¹ respectively. That is, the Hoogsteen lone-pair is notably more basic. Interestingly, however, in the lowest energy singly protonated complex, (G4-H⁺)₁, an O6 has been protonated at its WC-face (**Figure 6.2**). Furthermore, the relative energy of (G4-H⁺)₁ with respect to G4₁ corresponds to a PA of 1079.5 kJ mol⁻¹, an increase of 136.0 kJ mol⁻¹ in the PA of O6's WC lone-pair! In contrast, the alternate Hoogsteen protonated complex (G4-H⁺)₂, lying 7.4 kJ mol⁻¹ higher in energy than (G4-H⁺)₁, corresponds to a lesser increase of 92.5 kJ mol⁻¹ in the PA of O6's Hoogsteen lone-pair. In both cases, however, the increased PA's are likely due to stabilization of the positive charge via hydrogen bonds and charge delocalization from the protonated to the adjacent neutral guanine, as well as the presence of the nearby electron rich unprotonated O6's (**Figure D.1** in Appendix D). Indeed, it is noted that the N1—H bond of the protonated guanine in (G4-H⁺)₁ is 1.067 Å, significantly longer than observed in the guanines of G4₁ (1.022 Å; **Table D.1** in Appendix D). This lengthening will assist in charge transfer to the adjacent guanine. It should also be noted that in the molecular graph of (G4-H⁺)₁ a bond path was obtained between the protonated O6 and Hoogsteen-adjacent O6 lying 3.388 Å away (**Figure 6.3**). The value of $\nabla^2\rho$ at its BCP is 0.009 a.u. (**Table D.2** in Appendix D) suggesting a closed-shell bonding O...O interaction.⁵⁰⁻⁵² This appears to be a weak interaction as indicated by the BCP's large ϵ (0.646 a.u.; **Table D.2** in Appendix D) and its short separation (0.4 Å) from the nearest ring critical point (RCP).⁵³

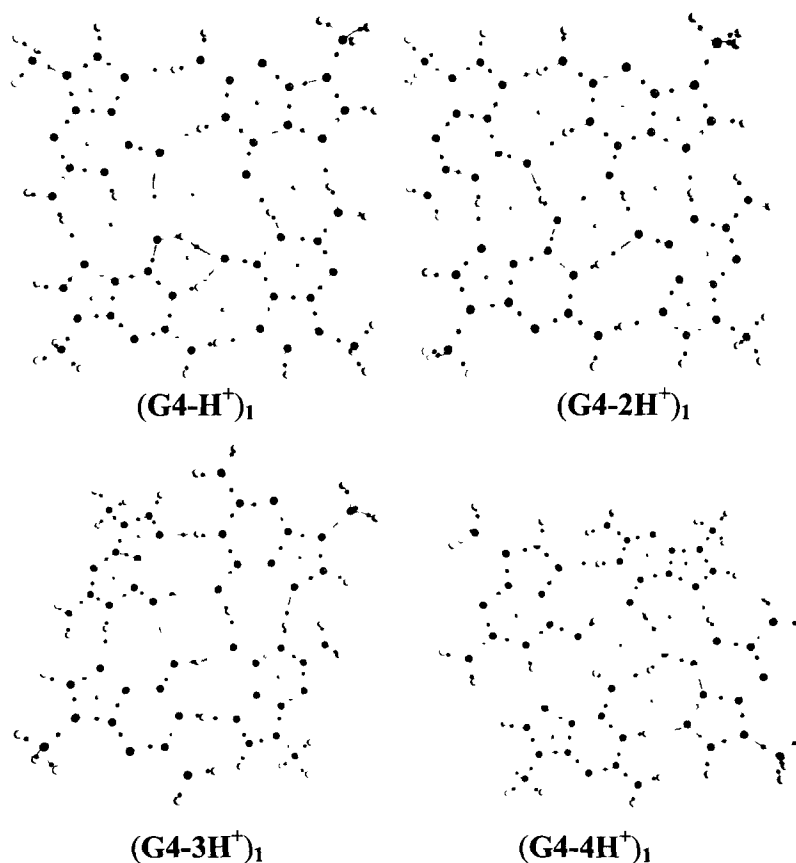


Figure 6.3 Molecular graphs for the lowest energy $(\mathbf{G4-nH}^+)$ $n=1-4$ complex of each protonation state. Red and yellow dots indicate BCP's and RCP's respectively.

In the planar WC-protonated complex $(\mathbf{G4-H}^+)_1$ all guanines interact in the same manner as in $\mathbf{G4}_2$ except that the newly formed C=OH^+ group now also forms a hydrogen bond with the WC-adjacent O6. As a result, this latter neutral O6 now forms a bifurcated hydrogen bond with both the N1-H and C=OH^+ groups of the protonated guanine. Furthermore, both of these interactions appear to be reasonably strong with quite short lengths of 1.858 and 1.561 Å respectively (**Figure 6.2**). Indeed, the value of ρ at the BCP of the former N1-H...O6 hydrogen bond (0.070 a.u. Table D.2 in Appendix D) is much larger than that obtained for the same interaction in $\mathbf{G4}_2$ (0.032 a.u. Table D.2 in Appendix D), further indicating a considerable increase in its strength. In contrast, the N1-H...O6(H^+) bond has weakened significantly and now

has a length of 2.346 Å and a much smaller value of ρ (0.005 a.u. Table D.2 in Appendix D) at its BCP. Due to protonation the C=OH⁺ bond has significantly lengthened to 1.316 Å. In addition, however, the C=O bond for the neutral oxygen involved in the bifurcated but still quite strong hydrogen bonds has also as a result concomitantly increased markedly to 1.264 Å (Table D.1 in Appendix D).

In contrast, the alternative Hoogsteen-protonated complex (**G4-H⁺**)₂ is non-planar and contains no bifurcated hydrogen bonds (**Figure 6.2**). Rather, the protonated guanine shifts its position such that its C=OH⁺ group and N7 centre form strong hydrogen bonds with the adjacent neutral guanine's O6 and N1-H groups with C=OH⁺...O6 and N7...N1-H distances of 1.501 and 1.840 Å respectively (see **Figure 6.2**). Such an arrangement of the hydrogen bonding interactions seems to cause the added proton to sit away from the electron rich central cavity, which may contribute to the preference for WC-protonation. In addition, two notable changes in the inter-guanine hydrogen bonding are also observed. The N1-H...O6 interaction between the protonated guanine and the WC-adjacent neutral guanine has shortened markedly from 1.763 Å in **G4**₂ to 1.465 Å while the diagonally opposite O6...H-N1 bond has increased considerably to 2.555 Å (cf. **Figure 6.1**).

Comparison of (**G4-H⁺**) to the complexes formed by **G4** binding with Li⁺ and other metal ions highlights the unique role of H⁺. For example, in contrast to that observed for H⁺, even the smallest metal ion Li⁺ sits in the middle of the quartets central cavity binding equally with all four oxygens.²⁸

6.3.3 Doubly Protonated **G4** [(**G4-2H⁺**)]

Three complexes were found upon double protonation (**Figure 6.2**). Interestingly, they are all protonated at diagonally opposite O6's but differ with

respect to which face the O6's are protonated and orientation of the resulting -OH group.

In contrast to that observed for (**G4-H⁺**) the lowest energy complex (**G4-2H⁺**)₁ corresponds to both O6's being protonated at their Hoogsteen faces. Consequently, the resulting hydrogen bonding arrangement resembles that of (**G4-H⁺**)₂. For instance, each protonated guanine forms two modified Hoogsteen C=OH⁺...O6 and N7...H-N1 hydrogen bonds with the adjacent neutral guanine while the two neutral guanines retain their standard O6...H-N1 and N7...H₂N- Hoogsteen hydrogen bonds. Interestingly, the protonated guanine's C=OH⁺...O6 and N7...H-N1 hydrogen bond lengths differ only slightly by +0.004 and -0.012 Å respectively from those in (**G4-H⁺**)₂. In contrast, however, the neutral guanine's O6...H-N1 and N7...H₂N- bonds differ dramatically by +0.463 and -0.099 Å respectively. Remarkably, despite these changes and the increased charge all inter-guanine hydrogen bond lengths in (**G4-2H⁺**)₁ lie within a much narrower range (1.505–1.928 Å) than those in (**G4-H⁺**)₂ (1.465–2.555 Å).

Interestingly, the value of ρ at the BCP of the C=OH⁺...O6 bond is 0.066 a.u. (Table D.2 in Appendix D). This value is much larger than for the same interaction in (**G4-H⁺**)₁ (0.017 a.u., Table D.2 in Appendix D) and in fact, is the largest value obtained for such an interaction in all of the (**G4-nH⁺**)₁ complexes. As observed in (**G4-H⁺**)₁ the positive charges are found to not only delocalize over the protonated guanine but also to a degree onto the adjacent neutral guanines (Figure D.1 in Appendix D). Similarly, the protonated guanine's N1-H bond has again lengthened, although now to a lesser extent, to 1.034 Å. Interestingly, the C=OH⁺ bond is also shorter than (**G4-H⁺**)₁ at 1.296 Å. This is likely due in part to the fact that the two added protons are shielded from each other by the formally protonated oxygens.

Indeed a bond path is found between the two protonated O6's (**Figure 6.3**). Notably, they lie 2.718 Å apart, approximately twice the ionic radius of oxygen⁵⁴ which has been previously suggested to be a requirement for such an interaction.^{51,52} The values of ρ and $\nabla^2\rho$ at its BCP are 0.012 and 0.048 a.u. respectively (Table D.2 in Appendix D), indicating a closed-shell bonding interaction. In fact, this $\text{O}(\text{H}^+)\cdots(\text{H}^+)\text{O}$ interaction has the largest ρ and smallest ϵ of all $\text{O}\cdots\text{O}$ interactions found in the $(\text{G4-}n\text{H}^+)_1$, $n = 1-3$ complexes (Table D.2 in Appendix D).

Comparing the relative energy of $(\text{G4-2H}^+)_1$ with respect to $(\text{G4-H}^+)_1$ gives the affinity of G4 for a second proton to be 867.6 kJ mol⁻¹ which is 211.9 kJ mol⁻¹ lower than the first PA. However, it should be noted that at the same level of theory the PA's of H₂O and NH₃ are 714.8 and 883.4 kJ mol⁻¹ respectively. Thus, the PA of $(\text{G4-H}^+)_1$ is higher than water, suggesting that it appears to still be a quite reasonable base.

In the alternate planar complex $(\text{G4-2H}^+)_2$, lying just 12.7 kJ mol⁻¹ higher in energy than $(\text{G4-2H}^+)_1$, both O6's are protonated at the WC-faces. Analogous to that observed for $(\text{G4-2H}^+)_1$, it also resembles its corresponding singly protonated quartet $(\text{G4-H}^+)_1$ with regards to its inter-guanine interactions (**Figure 6.2**). Specifically, each neutral guanine's O6 forms bifurcated hydrogen bonds with the C=OH⁺ and N1-H groups of an adjacent protonated guanine. In addition, similar to $(\text{G4-2H}^+)_1$, the range of hydrogen bond lengths in $(\text{G4-2H}^+)_2$ is also reduced (1.696–2.250 Å) compared to that found in $(\text{G4-H}^+)_1$ (1.561–2.346 Å).

The third complex $(\text{G4-2H}^+)_3$ found lies considerably higher in energy than $(\text{G4-2H}^+)_1$ by 65.3 kJ mol⁻¹. Importantly, it exhibits a significantly different protonation arrangement from all the above singly and doubly protonated complexes. Specifically, while one O6 is protonated at its WC-face in the same manner as in $(\text{G4-}$

H⁺)₁, the –OH⁺ group of the second protonated guanine is twisted slightly out-of-plane with a ∠HOC6N1 dihedral angle of 155.4° (Table D.1 in Appendix D). Such a rotation results in increased single bond character of the C=OH⁺ bond as well which lengthens markedly to 1.343 Å. Indeed, it is now considerably longer than the same bond in both (G4-2H⁺)₁ (1.296 Å) and (G4-2H⁺)₂ (1.322 Å) (Table D.1 in Appendix D). Remarkably, however, this second protonated guanine continues to maintain standard O6⋯H–N1 and N7⋯H₂N– Hoogsteen bonds with the adjacent neutral guanine, although both have now lengthened considerably to 2.344 and 2.034 Å respectively (see **Figure 6.2**).

6.3.4 Triply and Quadruply Protonated G4's [(G4-3H⁺)] and [(G4-4H⁺)]

Two complexes, both non-planar, were obtained upon adding a third proton (**Figure 6.2**). These two complexes share a number of similarities. In particular, the O6 of the sole remaining unprotonated guanine forms two O6⋯⁺HO=C hydrogen bonds, one each with the C=OH⁺ groups of the two adjacent protonated guanines. This is made possible because these two protonated guanines are in fact protonated at opposite faces of their O6 moieties (see **Figure 6.2**). In addition, the neutral guanine also maintains N1–H⋯N7 and N7⋯H–N1 hydrogen bonds with these two adjacent protonated guanines. As a result, unlike in all previous complexes that involve WC-protonation, bifurcated hydrogen bonds do not exist in triply protonated complexes. However, where these two complexes differ from each other is with respect to the face at which the third guanine, diagonally opposite the neutral guanine, is protonated. The preferred complex (G4-3H⁺)₁ corresponds to protonation at the Hoogsteen-face of its O6 moiety. The alternate WC-face protonated complex (G4-3H⁺)₂, however, lies only quite marginally higher in energy by 1.5 kJ mol⁻¹. In both complexes the third protonated guanine forms just three inter-guanine hydrogen bonds, two via its protonated face. The AIM analysis, however, found an O⋯O interaction between the

third protonated guanines $C=OH^+$ and an adjacent $^+HO6=$ group (see **Figure 6.3**). This $(H^+)O6\cdots O6(H^+)$ distance in $(G4-3H^+)_1$ is 2.909 Å (Table D.1 in Appendix D). In addition, the ρ and $\nabla^2\rho$ values at its BCP are 0.007 and 0.028 a.u. respectively (Table D.2 in Appendix D), again indicating a weak closed-shell bonding interaction. Comparing the relative energy of $(G4-3H^+)_1$ with respect to $(G4-2H^+)_1$ gives the affinity for addition of a third proton to be 563.7 kJ mol⁻¹. This is 303.9 kJ mol⁻¹ lower than the second PA and importantly, it is now also markedly lower than the PA of H₂O (see above).

Two complexes were obtained, again both non-planar, upon adding a fourth proton (**Figure 6.2**). The lowest energy complex $(G4-4H^+)_1$ corresponds to all four O6's being Hoogsteen-face protonated. The second complex $(G4-4H^+)_2$ has all four O6's protonated at their WC-faces and lies 24.2 kJ mol⁻¹ higher in energy. In both complexes all guanines interact via N1–H \cdots N7 and $(H^+)O6\cdots^+HO=C$ hydrogen bonds. Notably, despite the large positive charge now on the quartets, all hydrogen bond lengths are moderately short and fall within the reasonably narrow range of 1.796–1.892 Å. This is again possible due in part to significant charge delocalization over the guanines which, as a result, helps to decrease the positive charge around the central cavity (Figure D.1 in Appendix D). This delocalization is illustrated by the lengthening of both the N1–H and $C=OH^+$ bonds to, for example 1.047 and 1.340 Å respectively in $(G4-4H^+)_1$ (Table D.1 in Appendix D). The continued preference for Hoogsteen protonation in $(G4-3H^+)$ and $(G4-4H^+)$ may indicate that it is more efficient for delocalizing the charges. The calculated PA for the fourth proton is quite low at 301.0 kJ mol⁻¹, which is also significantly lower than the PA of H₂O (see above).

To gain further insights into possible structures and effects of protonation of guanines within a G4_n stack, up to quadruple protonation of a two-G4 stack (2G4) was then considered. Due to the number of structures obtained, unless otherwise noted only the lowest energy structures for each protonation state are discussed. All of the complexes found, however, are given in Table D.1 in Appendix D and Figure D.2 in Appendix D.

6.3.5 Singly Protonated 2G4 [(2G4-H⁺)]

The three structures shown in **Figure 6.4** were obtained upon addition of a single proton to 2G4. While they share a number of similarities to the related (G4-H⁺) complexes some important differences also were observed.

For example, the lowest energy complex (2G4-H⁺)₁ again corresponds to protonation of a guanine at the WC-face of its O6 moiety. As in the analogous single-plane complex (G4-H⁺)₁, the O6 of a neutral guanine within the same plane forms short bifurcated hydrogen bonds with the protonated guanine's C=OH⁺ (1.779 Å) and N1-H (1.643 Å) groups (see **Figure 6.4**). Meanwhile, all other guanines in both planes interact via standard Hoogsteen N1-H...O6 and -NH₂...N7 bonds. The alternate Hoogsteen-protonated complex (2G4-H⁺)₂ again lies higher in energy, though now by a slightly larger margin of 23.6 kJ mol⁻¹. As for (2G4-H⁺)₁, its protonated plane also has the same hydrogen bonding arrangement as in the analogous single plane complex (G4-H⁺)₂. For example, the protonated guanine shifts and forms O6...⁺HO=C (1.531 Å) and N1-H...N7 (2.375 Å) bonds with an adjacent guanine. Similarly, the guanines of the unprotonated plane bind via standard Hoogsteen interactions.

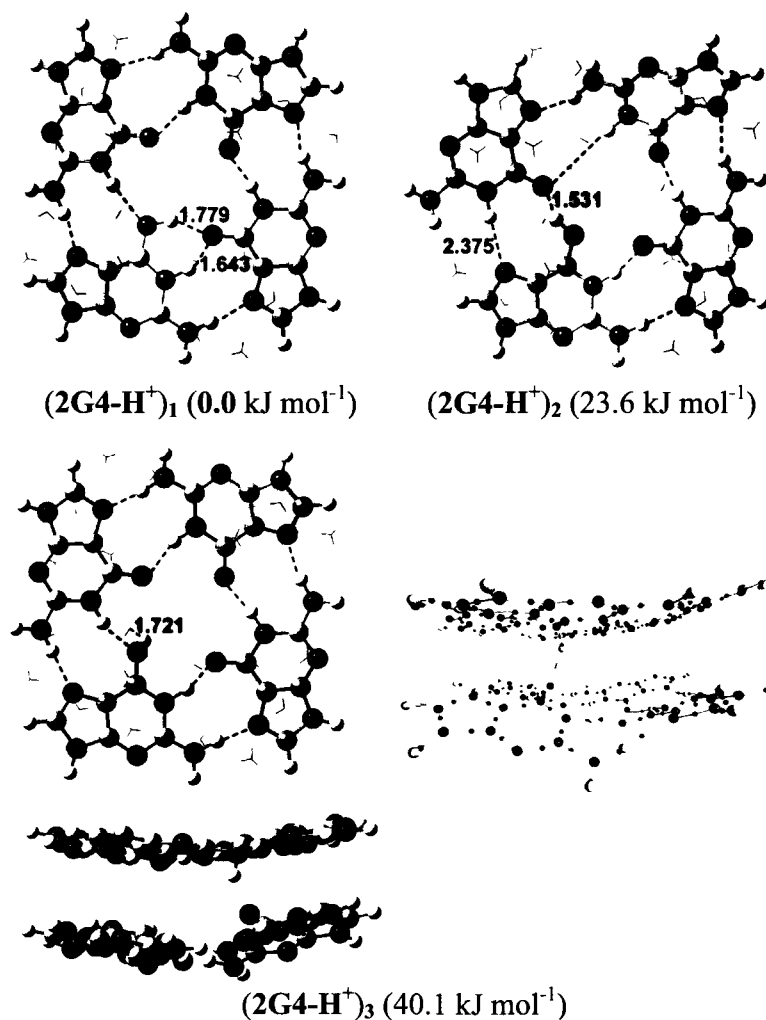


Figure 6.4 Optimized structures (top-view) of all singly protonated **2G4**'s and the side-view and molecular graph for **(2G4-H⁺)₃**. For clarity, not all bond paths are shown. Relative energies are also shown while selected optimized distances are in angstroms. Red dots indicate BCP's. Atom colour key: C (gray); O (red); N (blue) and H (white).

A third complex was also obtained in which, unlike that observed for either (**G4-H⁺**) quartet, the proton is now directed out of the plane. In this complex **(2G4-H⁺)₃**, lying 40.1 kJ mol⁻¹ higher in energy than **(2G4-H⁺)₁**, the C=OH⁺ group has rotated such that its proton is directed towards the lower neutral plane and thus now forms a short hydrogen bond (1.721 Å) with the nearest O6 in the lower plane. That is, the

proton now sits between the planes. This rotation is facilitated by a lengthening of the C=OH⁺ bond to 1.318 Å (Table D.1 in Appendix D), which is indeed now slightly longer than observed in both (2G4-H⁺)₁ (1.314 Å) and (2G4-H⁺)₂ (1.293 Å). Formation of this inter-plane hydrogen bond also results in a lengthening of the lower plane's C=O bond that is involved (1.252 Å; Table D.1 in Appendix D) as well as a slight upwards rotation of this guanine (see **Figure 6.4**). Interestingly, this inter-plane hydrogen bond *does not disrupt* the Hoogsteen inter-guanine bonding within *either* plane. The bond path for this interaction is shown in the molecular graph of (2G4-H⁺)₃ given in **Figure 6.4**. The values of ρ and $\nabla^2\rho$ at its BCP are 0.035 and 0.132 a.u. respectively (Table D.2 in Appendix D), indicative of a strong hydrogen bonding-type interaction.^{47,48}

It is noted that in each of the above (2G4-H⁺) complexes, the formally neutral planes have standard Hoogsteen and not bifurcated hydrogen bonding arrangements. This is the reverse of that observed for the isolated G4 quartets. However, as mentioned above, the central region of the Hoogsteen-bound neutral G4 is more electron rich and thus perhaps better able to electrostatically stabilize the protonated central cavity of the (G4-H⁺) plane. Indeed, the calculated Mulliken charges of the =OH⁺ moieties of (2G4-H⁺)₁ and (G4-H⁺)₁ are 0.25 and 0.07 respectively. That is, there is a reduced delocalization of the charge over the guanine moiety in (2G4-H⁺)₁.

6.3.6 Doubly Protonated 2G4 [(2G4-2H⁺)]

While these are the simplest multi-protonated 2G4 complexes, there are many possible isomers and indeed, a large number were obtained (Table D.1 in Appendix D and Figure D.2 in Appendix D). However, many simply correspond to stacking two (G4-H⁺)'s or (G4-2H⁺) with G4 and furthermore, exhibit the same general features.

Hence, for simplicity, only the lowest energy complexes or those exhibiting features unique to a multi-protonated 2G4 are discussed herein and shown in **Figure 6.5**.

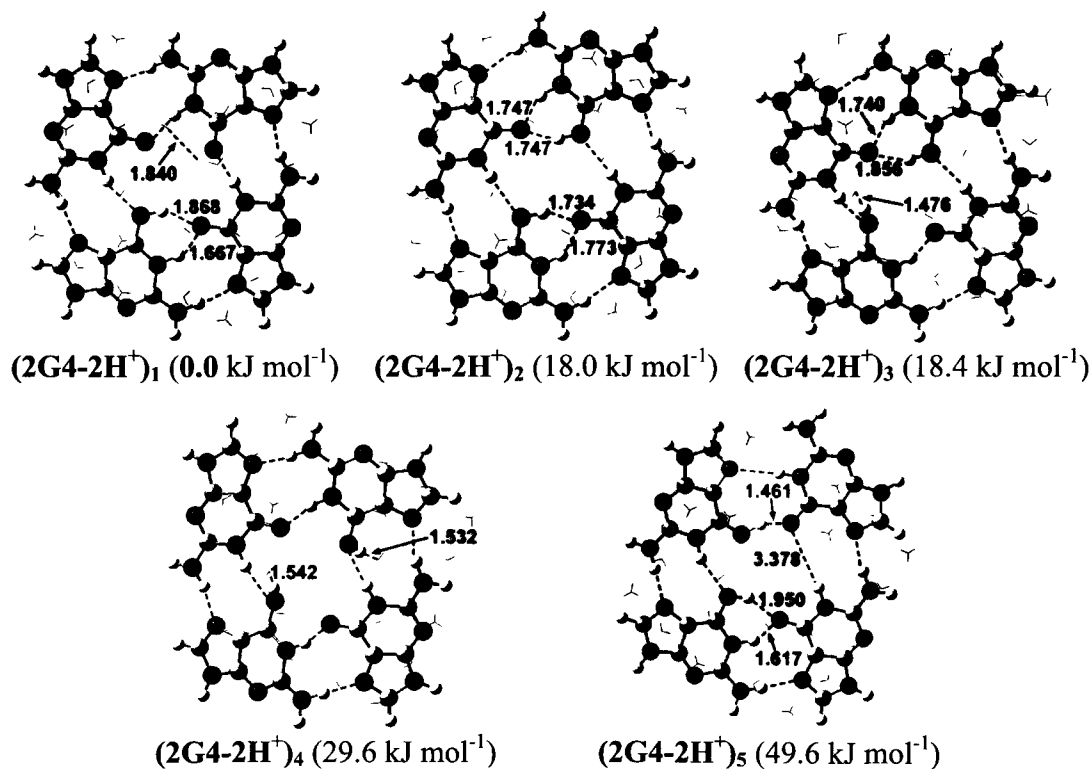


Figure 6.5 Optimized structures for the three lowest energy and two selected higher energy (2G4-2H⁺) complexes. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white).

The lowest energy complex, (2G4-2H⁺)₁, corresponds to the stacking of two of the singly WC-protonated quartet (G4-H⁺)₁, i.e., the most stable (G4-H⁺) complex. This also suggests that the charge prefers to be delocalized over both planes. Furthermore, the protonated guanines are diagonally opposite each other. The analogous alternate complexes in which they are diagonally adjacent or stacked upon each other lie 6.8 and 15.1 kJ mol⁻¹ higher in energy respectively (Figure D.2 in Appendix D). Despite small differences in hydrogen bond lengths, the same overall

trends as $(\mathbf{G4-H}^+)_{1}$ are observed in $(\mathbf{2G4-2H}^+)_{1}$. For example, bifurcated hydrogen bonds are again observed between a neutral O6 and the N1–H and C=OH⁺ groups of each protonated guanine. In addition, all other neutral guanines interact via standard Hoogsteen hydrogen bonds.

The next distinct lowest energy complex $(\mathbf{2G4-2H}^+)_{2}$ lies 18.0 kJ mol⁻¹ higher in energy and has both protons in the same plane. However, it does not correspond to a simple stacking of the lowest energy **G4** and $(\mathbf{G4-2H}^+)$ quartets but rather stacking of their alternate higher energy complexes, i.e., the doubly WC-protonated $(\mathbf{G4-2H}^+)_{2}$ on the Hoogsteen-bound neutral **G4**₂. The preference reversal for the protonated plane, i.e., double WC- vs. Hoogsteen-protonation, may be due in part to the fact that the WC-protonated complex $(\mathbf{G4-2H}^+)_{2}$ is more planar (cf. **Figure 6.1**) and thus perhaps better able to π - π stack with the neutral **G4**. Indeed, the analogous complexes in which one or both planes are instead Hoogsteen-protonated lie 19.8 and 40.8 kJ mol⁻¹ higher in energy respectively than $(\mathbf{2G4-2H}^+)_{1}$ (Figure D.2 in Appendix D). Meanwhile, as observed in the $(\mathbf{2G4-H}^+)$ complexes, the guanines of the neutral plane interact via standard Hoogsteen hydrogen bonds. In fact, this binding was observed in all formally neutral **G4** planes within $(\mathbf{2G4-2H}^+)$ complexes (Figure D.2 in Appendix D). The more electron rich central cavity of such bound neutral planes again helps to electrostatically stabilize the charged central region of the protonated plane.

Interestingly, lying only very marginally higher in energy than $(\mathbf{2G4-2H}^+)_{2}$ by just 0.4 kJ mol⁻¹ is the alternate complex $(\mathbf{2G4-2H}^+)_{3}$. Again, two diagonally opposite guanines in the same plane have been protonated. Now, however, while one O6 has been protonated at its WC-face the second C=OH⁺ group has rotated out-of-the-plane to form a quite short (1.476 Å) inter-plane hydrogen bond with the nearest O6 of the lower formally neutral **G4** quartet. This latter interaction is similar to that observed in

(2G4-H⁺)₃ and shows the same general features. For example, it causes the lower neutral G4 guanine involved to rotate slightly upwards, yet without disrupting the Hoogsteen bonding patterns within *either* plane. It is noted that the energy difference of 18.4 kJ mol⁻¹ between (2G4-2H⁺)₃ and (2G4-2H⁺)₁ is decidedly less than between the corresponding singly protonated complexes (2G4-H⁺)₃ and (2G4-H⁺)₁ which differ by 40.1 kJ mol⁻¹ (cf. **Figure 6.4**). It is also possible to have both C=OH⁺ groups rotated and forming inter-plane C=OH⁺...O6 bonds. The resulting complex, (2G4-2H⁺)₄, lies a further 11.2 kJ mol⁻¹ higher in energy (**Figure 6.5**).

A number of high energy (2G4-2H⁺) complexes were found which, like those described above, can be considered as mixtures of two (G4-H⁺)'s or a (G4-2H⁺) and a G4 (**Figure D.2** in Appendix D). However, one particular complex, (2G4-2H⁺)₅, lying 49.6 kJ mol⁻¹ higher in energy than (2G4-2H⁺)₁, was found with a unique protonation arrangement. Specifically, one guanine has again been protonated at the WC-face of its O6 moiety. Now, however, the Hoogsteen-adjacent guanine's O6 has also been protonated at its Hoogsteen face (see **Figure 6.5**). That is, two adjacent guanines in the same plane have been protonated. This adjacent proximity of protonated guanines and mixed protonated faces was previously only seen in the very highly protonated (G4-nH⁺) $n \geq 3$ quartets, i.e., where they had no choice but to be next to each other (cf. **Figure 6.2**). This further illustrates the stabilizing effects of a second and neutral G4 plane. Each of the protonated guanines in (2G4-2H⁺)₅ forms typical hydrogen bonding arrangements: the WC-protonated guanine forms =O6H⁺...O6 and N1-H...O6 bonds with the adjacent neutral O6 while the Hoogsteen-protonated guanine shifts and forms N7...H-N1 and =O6H⁺...O6 bonds with the other neutral guanine. However, as a result the N1-H...O6 bond between the two neutral guanines is disrupted having now lengthened dramatically to 3.378 Å (see **Figure 6.5**).

6.3.7 Triply and Quadruply Protonated 2G4's [(2G4-3H⁺)] and [(2G4-4H⁺)]

Upon further addition of one and two protons to (2G4-2H⁺), a large number of complexes were obtained (Figure D.2 in Appendix D). However, as was observed for (2G4-2H⁺) many of these complexes simply correspond to the stacking of two protonated G4 planes. Hence, unless otherwise noted, for simplicity only the lowest energy (2G4-3H⁺) and (2G4-4H⁺) complexes obtained are discussed herein and are shown in Figure 6.6.

For both protonation states the lowest energy complexes, (2G4-3H⁺)₁ and (2G4-4H⁺)₁, have *no more than* two added protons per plane. Furthermore, all protonation of the O6's has occurred at their WC-faces. In addition, each protonated guanine sits above/below a neutral guanine, thus also maximising the distance between their positive charges. The series of complexes in which one or more of the O6's are instead protonated at their Hoogsteen-faces lie systematically higher in energy (see also Figure D.2 in Appendix D). For example, in the next lowest energy complexes for both protonation states, (2G4-3H⁺)₂ and (2G4-4H⁺)₂, one guanine O6 is now protonated at its Hoogsteen-face while all others are protonated at their WC-face. Notably, in (2G4-3H⁺)₂ the Hoogsteen protonated guanine is found in the singly protonated plane. (2G4-3H⁺)₂ and (2G4-4H⁺)₂ are 7.1 and 4.0 kJ mol⁻¹ higher in energy than (2G4-3H⁺)₁ and (2G4-4H⁺)₁ respectively. These relative energy differences are less than observed between the analogous (2G4-2H⁺) complexes (see above).

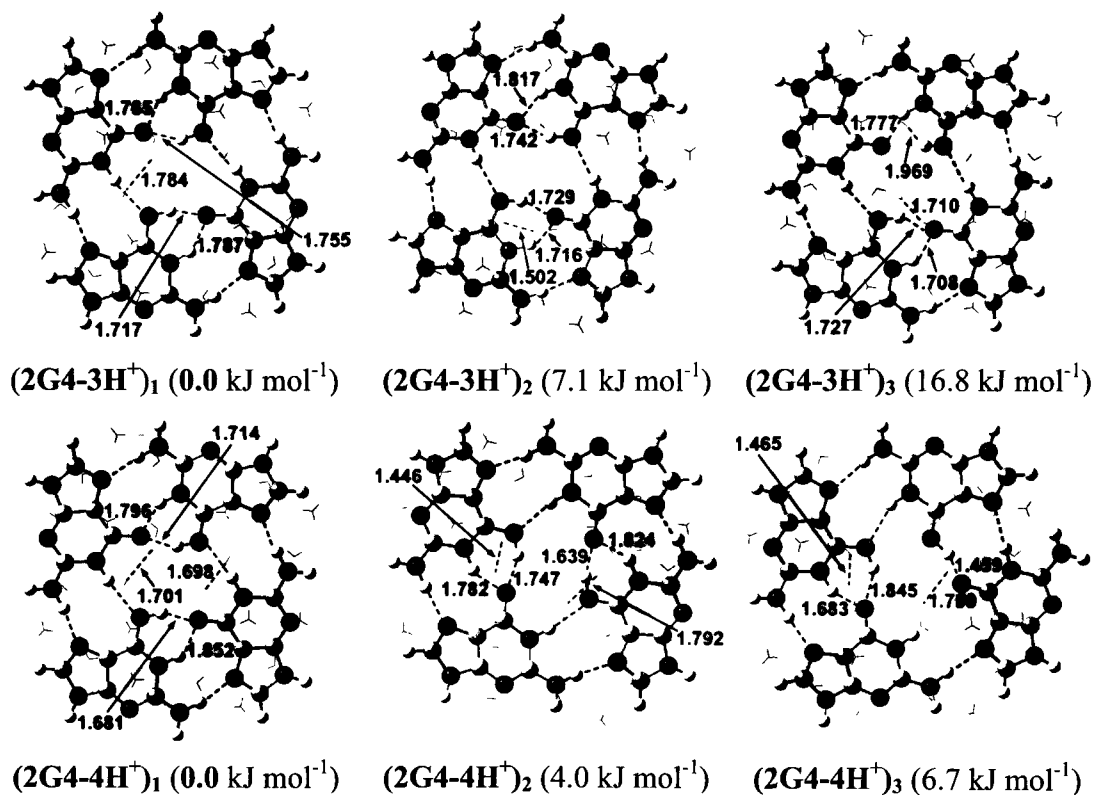


Figure 6.6 The three lowest energy optimized structures (top-views) for $(2\text{G4-}n\text{H}^+)$ $n = 3$ and 4. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white).

However, some differences between the triply and quadruply protonated 2G4 stacks are also observed. For the former, the next lowest energy complex $(2\text{G4-3H}^+)_3$ has one O6 in each plane protonated at its WC-face. Furthermore, these are diagonally adjacent to each other. However, this enables the second protonated guanine of the upper plane, which is diagonally opposite the first, to form a short inter-plane hydrogen bond (1.542 Å) with a neutral guanine's O6 of the lower singly-protonated plane (see **Figure 6.6**). The upper plane is thus analogous to the doubly protonated plane in $(2\text{G4-2H}^+)_3$ (cf. **Figure 6.5**). The relative energy of $(2\text{G4-3H}^+)_3$ is 12.8 kJ mol^{-1} , which is less than that of $(2\text{G4-2H}^+)_3$.

In contrast, for $(2G4-4H^+)$ complexes involving inter-plane hydrogen bonds are at least 37.2 kJ mol^{-1} higher in energy than $(2G4-4H^+)_1$ (Figure D.2 in Appendix D). This may be due to the fact that the formation of such inter-plane interactions causes greater non-planarity of both planes and thus, decreases π - π stacking stabilization effects. Thus instead, for $(2G4-4H^+)$ the third lowest energy complex, $(2G4-4H^+)_3$, has adjacent WC- and Hoogsteen-protonated guanines in each plane that are stacked above/below the other plane's two neutral guanines (Figure 6.6). This is similar to that observed in $(2G4-2H^+)_5$ (cf. Figure 6.5). However, unlike the high relative energy of $(2G4-2H^+)_5$, $(2G4-4H^+)_3$ is just 6.7 kJ mol^{-1} higher in energy than $(2G4-4H^+)_1$.

It is noted that the same general trends in C=O bond lengths were observed for all of the above complexes as was previously described for their analogous singly and doubly protonated complexes (Table D.1 in Appendix D). For example, increased protonation typically causes greater lengthening of the C=OH⁺ bonds. Again, the increases are generally not as large as observed for their corresponding $(G4-3H^+)$ and $(G4-4H^+)$ counterparts due to decreased charge delocalization over the guanine rings.

It is noted that several $(2G4-3H^+)$ and $(2G4-4H^+)$ complexes were found in which three or more guanines in one plane are protonated. However, all lie significantly higher in energy (Figure D.2 in Appendix D). Furthermore, in all such cases at least one of the protonated guanines forms an inter-plane hydrogen bond. Two examples of such complexes are shown in Figure 6.7. In $(2G4-3H^+)_4$ there are two inter-plane hydrogen bonds between the triply protonated upper plane and the neutral lower plane. In $(2G4-4H^+)_4$, all four protonated guanines form inter-plane hydrogen bonds. This is also the highest energy quadruply protonated 2G4 stack.

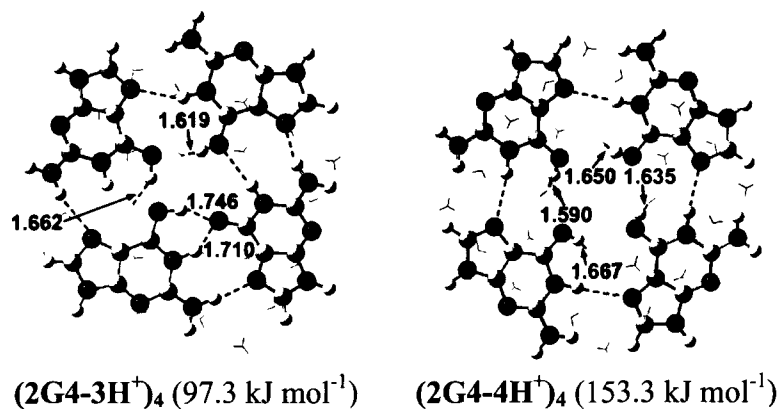


Figure 6.7 Optimized structures for $(2\text{G4-3H}^+)_4$ and $(2\text{G4-4H}^+)_4$ in which three and four guanines respectively in one plane have been protonated. Relative energies are also shown while selected optimized distances are in angstroms. Atom colour key: C (gray); O (red); N (blue) and H (white).

6.4 Conclusions

Density functional theory methods have been used to investigate the single to quadruple protonation of a one-plane guanine quartet (**G4**) and a two-plane guanine quartet stack (**2G4**), particularly at their oxygen (O6) centres.

G4 quartets: for the singly protonated (**G4-H**⁺) complexes protonation at the Watson-Crick face of an O6 is found to be preferred. This leads to formation of a bifurcated hydrogen bond between the O6 moiety of the adjacent neutral guanine with both the C=OH⁺ and N1-H groups of the protonated guanine. In contrast, as protonation at the Hoogsteen-face of an O6 alters the Hoogsteen motif, the resulting protonated guanines shifts and forms N7...H-N1 and C=OH⁺...O6 hydrogen bonds with the adjacent neutral guanine. In contrast, for all multi-protonated **G4** complexes protonation at the Hoogsteen-faces of the O6 moieties is found to be preferred. In addition, the resulting C=OH⁺ bonds are increasingly lengthened upon systematic protonation of the **G4** due to systematically greater delocalisation of the positive

charge from the =OH⁺ groups onto the guanine rings. As a result, even with the addition of four protons the protonated guanine quartets can maintain moderately short and strong hydrogen bonds between the protonated guanine moieties. For example, in (G4-4H⁺)₁, the longest hydrogen bond is the N7···H–N1 bond with a length of 1.892 Å.

The calculated proton affinities (PA's) for addition of one, two, three and four protons to a G4 are 1079.5, 867.5, 563.8 and 300.9 kJ mol⁻¹ respectively. The first is significantly higher than the PA of both the Hoogsteen and WC lone-pairs of O6 of monomeric guanine by at least 92.5 kJ mol⁻¹. Thus, G4 formation enhances the basicity of the guanine O6's. Notably, the first and second gas-phase PA's of G4 are also higher than that of H₂O as determined at the same level of theory.

2G4 stacks: for single to quadruple protonation of the two guanine stack 2G4 several general trends were observed. In addition, some important differences from the (G4-*n*H⁺) *n* = 1-4 complexes were observed. First, unlike for the G4 quartets, for *all* protonation states [i.e., (2G4-*n*H⁺) *n* = 1-4] the most stable complexes correspond to solely WC-face protonation of the O6 centres. Furthermore, the protonation is distributed over both planes with maximally two protons per plane. This is suggested to be due a combination of several factors including electrostatic stabilization of the formally =OH⁺ groups of one plane by the electron rich central cavity of the other plane. Indeed, less charge delocalization over the guanine rings is observed in the 2G4 stacks than in the lone G4 quartets. In addition, singly and doubly WC-protonated G4, i.e., (G4-*n*H⁺) *n* = 1–2, are more planar and thus also better able to π-π overlap with the other plane. Complexes were also found in which one or more protons 'sit' between planes due to formation of inter-plane hydrogen bonds. The lowest relative energy complexes that form such bonds occur for the doubly and triply protonated

2G4's, i.e., (2G4-2H⁺) and (2G4-3H⁺). This suggests that proton transfer between planes may be more efficient at an appropriate H⁺ concentration.

These findings provide interesting initial insights into the effects of a fundamental cation of biochemistry (H⁺) on guanine quartets and quadruplexes. In addition, they also provide new insights into the abilities and role of guanine quadruplexes, particularly as a base for porphyrin metallation.

References

- (1) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013.
- (2) Max, A. K. *Biopolymers* **2000**, *56*, 123.
- (3) Seela, F.; Kroschel, R. *Bioconjugate Chem.* **2001**, *12*, 1043.
- (4) Cai, M.; Sidorov, V.; Lam, Y. F.; Flowers, R. A.; Davis, J. T. *Org. Lett.* **2000**, *2*, 1665.
- (5) Cai, M.; Marlow, A. L.; Fettinger, J. C.; Fabris, D.; Haverlock, T. J.; Moyer, B. A.; Davis, J. T. *Angew. Chem.* **2000**, *112*, 1339.
- (6) Shi, X.; Fettinger, J. C.; Cai, M.; Davis, J. T. *Angew. Chem. Int. Ed.* **2000**, *39*, 3124.
- (7) Kettani, A.; Gorin, A.; Majumdar, A.; Hermann, T.; Skripkin, E.; Zhao, H.; Jones, R.; Patel, D. J. *J. Mol. Biol.* **2000**, *297*, 627.
- (8) Davis, J. T. *Angew. Chem. Int. Ed.* **2004**, *43*, 668.
- (9) Douglas, R. H.; Partridge, J. C.; Marshall, N. J. *Prog. Retinal Eye Res.* **1998**, *17*, 597.
- (10) Oxford, G. S. *Proceedings of the 17th European Colloquium of Arachnology, Edinburgh, 1997*, ed. Selden, P. A., British Arachnological Society, Burnham Beeches, Bucks, 1998, pp. 121.

- (11) Oxford, G. S.; Gillespie, R. G. *Annu. Rev. Entomol.* **1998**, *43*, 619.
- (12) O'Reilly, M.; Teichmann, S. A.; Rhodes, D. *Curr Opin. Struct. Biol.* **1999**, *9*, 56.
- (13) Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275.
- (14) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847.
- (15) Neidle, S.; Read, M. A. *Biopolymers* **2001**, *56*, 195.
- (16) Saretzki, G. *Cancer Lett.* **2003**, *194*, 209.
- (17) Reddel, R. R. *Cancer Lett.* **2003**, *194*, 155.
- (18) Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. *J. Am. Chem. Soc.* **1978**, *100*, 3625.
- (19) Rovnyak, D.; Baldus, M.; Wu, G.; Hud, N. V.; Feigon, J.; Griffin, R. G. *J. Am. Chem. Soc.* **2000**, *122*, 11423.
- (20) a) Basu, S.; Szewczak, A. A.; Cocco, M.; Strobel, S. A. *J. Am. Chem. Soc.* **2000**, *122*, 3240. b) Gill, M. L.; Strobel, S. A.; Loria, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 16723. c) Gill, M. L.; Strobel, S. A.; Loria, J. P. *Nucleic Acids Res.* **2006**, *34*, 4506.
- (21) Smirnov, I. V.; Kotch, F. W.; Pickering, I. J.; Davis, J. T.; Shafer, R. H. *Biochemistry* **2002**, *41*, 12133.
- (22) a) Wei, C.; Jia, G.; Yuan, J.; Feng, Z.; Li, C. *Biochemistry* **2006**, *45*, 6681. b) Wang, P.; Ren, L.; He, H.; Liang, F.; Zhou, X.; Tan, Z. *ChemBioChem* **2006**, *7*, 1155. c) Shelton, A. H.; Rodger, A.; McMillin, D. R. *Biochemistry* **2007**, *46*, 9143.
- (23) a) Keating, L. R.; Szalai, V. A. *Biochemistry* **2004**, *43*, 15891. b) Evans, S. E.; Mendez, M. A.; Turner, K. B.; Keating, L. R.; Grimes, R. T.; Melchoir, S.; Szalai, V. A. *J. Biol. Inorg. Chem.* **2007**, *12*, 1235.
- (24) Reed, J. E.; Arnal, A. A.; Neidle, S.; Vilar, R. *J. Am. Chem. Soc.* **2006**, *128*, 5992.
- (25) a) Reed, J. E.; Neidle, S.; Vilar, R. *Chem. Commun.* **2007**, 4366. b) Kieltyka, R.; Fakhoury, J.; Moitessier, N.; Sleiman, H. F. *Chem. Eur. J.* **2008**, *14*, 1145.
- (26) Mourik, T.; Dingley, A. J. *Chem. Eur. J.* **2005**, *11*, 6064.

- (27) a) Gu, J. D.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 6308. b) Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 529.
- (28) a) Meyer, M.; Steinke, T.; Brandl, M.; Sühnel, J. *J. Comput. Chem.* **2001**, *22*, 109. b) Meyer, M.; Sühnel, J. *J. Phys. Chem. A* **2003**, *107*, 1025.
- (29) a) Meyer, M.; Hocquet, A.; Sühnel, J. *J. Comput. Chem.* **2005**, *26*, 352. b) Meyer, M.; Sühnel, J. *J. Phys. Chem. A* **2008**, *112*, 4336. c) Meyer, M.; Steinke, T.; Sühnel, J. *J. Mol. Model.* **2007**, *13*, 335. d) Meng, F.; Xu, W.; Liu, C. *Internet Electron. J. Mol. Des.* **2006**, *5*, 79. e) Clay, E. H.; Gould, I. R. *J. Mol. Graph Model.* **2005**, *24*, 138.
- (30) a) Štefl, R.; Špačková, N.; Berger, I.; Koča, J.; Šponer, J. *Biophys. J.* **2001**, *80*, 455. b) Fadrná, E.; Špačková, N.; Štefl, R.; Koča, J.; Cheatham, T. E.; Šponer, J. *Biophys. J.* **2004**, *87*, 227. c) Cavallari, M.; Calzolari, A.; Garbesi, A.; Di Felice, R. *J. Phys. Chem. B* **2006**, *110*, 26337.
- (31) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 4th edition; W H. Freeman and Company: New York, 2004.
- (32) a) Tollinger, M.; Crowhurst, K. A.; Kay, L. E.; Forman-Kay, J. D. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4545. b) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167. c) Russo, N.; Toscano, M.; Grand, A.; Jolibois, F. *J. Comput. Chem.* **1998**, *19*, 989. d) Kubarenko, A. V.; Sergiev, P. V.; Bogdanov, A. A.; Brimacombe, R.; Dontsova, O. A. *Nucleic Acids Res.* **2001**, *29*, 5067. e) Rios-Font, R.; Rodríguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071.
- (33) Gehring, K.; Leroy, J.; Gueron, M. *Nature* **1993**, *363*, 561.
- (34) Diederichsen, U. *Angew. Chem. Int. Ed.* **1998**, *37*, 2273.
- (35) Phan, A. T.; Mergny, J. L. *Nucleic Acids Res.* **2002**, *30*, 4618.
- (36) a) Li, Y.; Sen, D. *Nat. Struct. Biol.* **1996**, *3*, 743. b) Li, Y.; Sen, D. *Biochemistry* **1997**, *36*, 5589. c) Li, Y.; Sen, D. *Chem. Biol.* **1998**, *5*, 1.

- (37) Sugimoto, N.; Toda, T.; Ohmichi, T. *Chem. Commun.* **1998**, 1533.
- (38) Hansson, M. D.; Karlberg, T.; Rahardja, M. A.; Al-Karadaghi, S.; Hansson, M. *Biochemistry* **2007**, *46*, 87.
- (39) Dailey, H. A.; Wu, C. K.; Horanyi, P.; Medlock, A. E.; Najahi-Missaoui, W.; Burden, A. E.; Dailey, T. A.; Rose, J. *Biochemistry* **2007**, *46*, 7973.
- (40) Jaguar 5.5, Schrodinger, L. L. C., Portland, OR, 1991-2003.
- (41) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.
- (42) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (43) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (44) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.
- (45) Biegler-König, F.; Schönbohm, J. AIM2000; Version 2.0; University of Applied Science: Bielefeld, Germany, 2002.

- (46) a) Gu, J.; Leszczynski, J.; Bansal, M. *Chem. Phys. Lett.* **1999**, *311*, 209. b) Meyer, M.; Brandl, M.; Sühnel, J. *J. Phys. Chem. A* **2001**, *105*, 8223. c) Louit, G.; Hocquet, A.; Ghomi, M.; Meyer, M.; Sühnel, J. *PhysChemComm* **2002**, *5*, 94. d) Mourik, T.; Dingley, A. J. *Chem. Eur. J.* **2005**, *11*, 6064.
- (47) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747.
- (48) Popelier, P. L. A. *J. Phys. Chem. A* **1998**, *102*, 1873.
- (49) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413.
- (50) Bianchi, R.; Gervasio, G.; Marabello, D. *Inorg. Chem.* **2000**, *39*, 2360.
- (51) Zhurova, E. A.; Tsirelson, V. G.; Stash, A. I.; Pinkerton, A. A. *J. Am. Chem. Soc.* **2002**, *124*, 4574.
- (52) Pakiari, A. H.; Eskandari, K. *J. Mol. Struct: THEOCHEM* **2007**, *806*, 1.
- (53) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893.
- (54) Shannon, R. D. *Acta Crystallogr.* **1976**, *A32*, 751.

Chapter 7

Interactions of Metal Ions with Ribose and Locked Ribose: A DFT Study

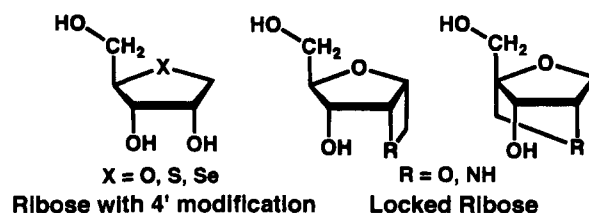
7.1 Introduction

Metal ions play a variety of key roles in biochemical systems, for example, as key components of enzyme active sites or in maintaining the three dimensional structure of proteins and nucleic acids.¹ Hence, there is tremendous interest in understanding their interactions with important biomolecules.

For example, nucleotides are well known as the monomers of DNA and RNA. However, along with their dephosphorylated derivatives (nucleosides) they are also found as key components of a range of important biomolecules including enzyme cofactors (e.g., CoA and NADH), energy molecules (e.g., ATP) and in therapeutic drugs (e.g., AZT).¹ As a result a number of experimental and theoretical studies on the interaction of various metal ions including alkaline metals,²⁻⁶ alkaline earth metals,⁵⁻¹⁴ Al(III),^{15,16} and transition metals such as Cu(I and II),^{11,17,18} Zn(II)^{6,7,11,12} and Cd(II)¹¹, with nucleotides and -sides have been reported. These studies, however, have generally focused on the interactions of the metal ions with their phosphate and/or nucleobase components. Yet, metal-ribose sugar interactions have been proposed to play, for example, important roles in the properties and catalytic mechanisms of some ribozymes. For instance, in the hammerhead ribozyme the binding of an Mg²⁺ ion with a ribose's 2'-oxygen was thought to lower the pK_a of the 2'-OH group and facilitate formation of a mechanistically

important 2'-O⁻ nucleophile.^{19,20} In addition, in the group I intron ribozyme the binding of two Mg²⁺ ions with the 2' and 3'-oxygens of a ribose sugar was proposed to play a role in stabilizing the transition structure.^{21,22}

Scheme 7.1 Schematic illustration of the ribose sugar in RNA and examples of some modified ribose.



Recently, there is also increasing interest in the construction of modified synthetic nucleotides. One commonly exploited modification is the 'locking' of a nucleotide by introducing a methylene linkage between a ring carbon and a hydroxyl oxygen of the ribose sugar (**Scheme 7.1**). Indeed, a variety of such locked nucleic acids (LNA) have been synthesized such as the four-membered 1',2'-azetidine and -oxetane,^{23,24} and five-membered 2'-N,4'-C- and 2'-O,4'-C-methylene bridged riboses.²⁵⁻²⁸ Some LNA's of these types have shown high affinity when hybridized with complementary DNA or RNA²⁹ and were also found to be able to affect the duplex-quadruplex equilibrium in some genes.³⁰ Hence, they have potential use in therapeutic applications, in particular as cancer drugs.²⁹ Another possible modification of ribose that has garnered increasing interest is the substitution of their ring 4'-oxygen by other heteroatoms such as sulphur or selenium (see **Scheme 7.1**).³¹⁻³⁴ Significantly, the resulting nucleosides and nucleotides have shown different properties compared with their 4'-oxoribose counterparts. For example, 4'-seleno- and 4'-oxo-nucleosides prefer different conformations³² while incorporation of 4'-selenonucleosides into oligonucleotides increases their thermal stabilities.³³ However,

how such sugar modifications influence their binding properties with biochemically important metal ions remains unclear.

In this chapter, we aim to determine how a range of biochemically important metal ions may interact with selected locked riboses and, in particular, any difference compared to their binding with native ribose. Four metal ions Na^+ , K^+ , Mg^{2+} and Cd^{2+} were chosen in our study. In addition to their biological importance, Na^+ , K^+ and Mg^{2+} were chosen in order to facilitate comparison of their behaviours due to the fact that they are adjacent to each other in the periodic table. In addition, Cd^{2+} has similar size as Na^+ but a larger charge, and hence is also included.

7.2 Computational Methods

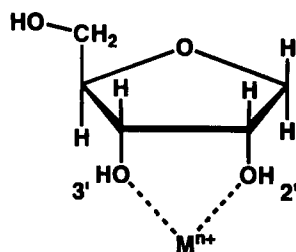
All calculations were performed using the Gaussian 03 suite of programs.³⁵ Optimized geometries were obtained using the density functional theory (DFT) method B3LYP: a combination of Becke's three-parameter hybrid exchange functional^{36,37} as implemented in Gaussian 03,³⁸ and the Lee-Yang-Parr correlation functional.³⁹ Initially, basis sets ranging from 6-311G(d,p) to 6-311++G(3df,3pd) were assessed for their ability to provide reliable structures and binding energies (ΔE 's) for a typical M^{n+} -ribose complex ($\text{M}^{n+}=\text{Na}^+$, K^+ , Mg^{2+}). The ΔE 's were corrected for zero point vibrational energy (ZPVE) while the basis set superposition error (BSSE) was determined using the counterpoise method.⁴⁰ An appropriate level of theory was then chosen and applied to all further complexes, with ΔE 's corrected for ZPVE only as calculated at the appropriate level of theory (see text).

7.3 Results and Discussion

7.3.1 Basis Set Assessment

We began our study by considering the complexes of native ribose with metal ions. Previous computational studies on related metal ion–biomolecule complexes have used a variety of methods and basis sets. In particular, however, the hybrid DFT method B3LYP has been applied successfully to the study of the binding of such ions in related systems.^{3-5,7,9-15,17,18} Furthermore, it has been found to generally give smaller BSSE's than conventional *ab initio* methods.⁴¹⁻⁴⁶ In order to determine an appropriate basis set for obtaining reliable structures *and* binding energies (ΔE 's) of the presently considered metal ion–ribose complexes, optimized structures and ΔE 's were determined for a typical bidentate complex using the B3LYP method in conjunction with basis sets ranging from 6-311G(d,p) to 6-311++G(3df,3pd). It should be noted that this range also includes some basis sets previously employed for related systems.

Scheme 7.2 Schematic illustration of the 2',3'-coordinated M^{n+} –ribose complex ($M^{n+} = Na^+, K^+, Mg^{2+}$).



We first considered an appropriate basis set choice for obtaining reliable geometries, in particular the key $r(M^{n+} \cdots O-2'/-3')$ interaction distances (see **Scheme 7.2**). The values obtained are given in **Table 7.1** with those at the B3LYP/6-311++G(3df,3pd) level being taken as the benchmark values.

Table 7.1 Selected Optimized Distances (Å), Binding Energies (ΔE) and Basis Set Superposition Errors (kJ mol^{-1}) Obtained Using the B3LYP Method with Various Basis Sets for the 2',3'-Coordinated M^{n+} -Ribose Complex Shown in Scheme 7.2.

	Basis set	$r(M^{n+} \dots O-3')$	$r(M^{n+} \dots O-2')$	ΔE_{uncorr}^a	BSSE	ΔE_{corr}^b
Na ⁺	6-311G(d,p)	2.214	2.243	192.0	13.1	178.9
	6-311+G(d,p)	2.223	2.253	176.7	5.2	172.5
	6-311G(df,p)	2.215	2.244	192.7	15.6	177.1
	6-311G(2d,p)	2.215	2.243	192.7	17.1	175.6
	6-311+G(2df,p)	2.225	2.254	174.3	5.4	168.9
	6-311+G(2df,2p)	2.225	2.255	173.8	4.6	169.2
	6-311++G(3df,3pd)	2.221	2.250	174.3	3.5	170.8
K ⁺	6-311G(d,p)	2.581	2.625	133.9	11.2	122.7
	6-311+G(d,p)	2.596	2.643	122.4	2.0	120.4
	6-311G(df,p)	2.581	2.624	133.8	8.7	125.1
	6-311G(2d,p)	2.584	2.628	134.2	12.8	121.4
	6-311+G(2df,p)	2.602	2.647	119.6	1.5	118.1
	6-311+G(2df,2p)	2.602	2.648	119.2	1.2	118.0
	6-311++G(3df,3pd)	2.600	2.646	118.8	1.0	117.8
Mg ²⁺	6-311G(d,p)	1.913	1.950	724.7	16.2	708.5
	6-311+G(d,p)	1.915	1.952	700.4	8.0	692.4
	6-311G(df,p)	1.908	1.942	730.3	19.3	711.0
	6-311G(2d,p)	1.898	1.933	733.5	20.6	712.9
	6-311+G(2df,p)	1.897	1.933	707.8	5.8	702.0
	6-311+G(2df,2p)	1.898	1.932	706.7	4.6	702.1
	6-311++G(3df,3pd)	1.892	1.927	712.3	3.3	709.0

^a Binding energies without BSSE correction. ^b Binding energies with BSSE correction.

The Na⁺- and K⁺-ribose complexes exhibit similar trends. Specifically, when basis sets containing diffuse functions on heavy atoms are used [i.e., 6-311+G(d,p), 6-311+G(2df,p) and 6-311+G(2df,2p)], they give structures in close agreement with each other. More importantly, however, they are also in good agreement with the benchmark

values. For example, for the above three basis sets the $r(\text{Na}^+\cdots\text{O-3}')$ and $r(\text{Na}^+\cdots\text{O-2}')$ distances differ only negligibly from each other by 0.001–0.002 Å. In addition, they are only quite marginally longer than the corresponding benchmark lengths by 0.002–0.005 Å. Similarly, for those basis sets without diffuse functions on heavy atoms [i.e., 6-311G(d,p), 6-311G(df,p), 6-311G(2d,p)], the $r(\text{M}^+\cdots\text{O-3}')$ and $r(\text{M}^+\cdots\text{O-2}')$ distances are in close agreement with each other. However, they are now distinct from those with diffuse functions included. For example, the $r(\text{Na}^+\cdots\text{O-3}')$ and $r(\text{Na}^+\cdots\text{O-2}')$ distances among these three basis sets are 0.009–0.012 Å shorter than those obtained with diffuse functions. This effect is even more pronounced in the K^+ –ribose complex. Indeed, the $r(\text{K}^+\cdots\text{O-3}')$ and $r(\text{K}^+\cdots\text{O-2}')$ distances are now 0.012–0.024 Å shorter than obtained when using basis sets that include diffuse functions on heavy atoms (see **Table 7.1**).

The Mg^{2+} –ribose complex, however, shows quite different trends. Specifically, the most accurate results are obtained for those basis sets that contain a second set of *d*-functions on heavy atoms. For example, when the basis sets 6-311G(2d,p), 6-311+G(2df,p) or 6-311+G(2df,2p) are used the $r(\text{Mg}^{2+}\cdots\text{O-3}')$ and $r(\text{Mg}^{2+}\cdots\text{O-2}')$ distances obtained differ from each other by just 0.001 Å. Furthermore, they are only 0.005–0.006 Å longer than the benchmark values (see **Table 7.1**). In contrast, for all other basis sets [i.e., 6-311G(d,p), 6-311G(df,p) and 6-311+G(d,p)] these key distances are in close agreement with each other, however, they are 0.015–0.025 Å longer than the benchmark values.

We then considered an appropriate basis set for obtaining accurate binding energies (ΔE). The BSSE corrected binding energies (ΔE_{corr}) obtained at the B3LYP/6-311++G(3df,3pd) level of theory were used as the benchmark energies (**Table 7.1**). All three metal ion–ribose complexes are observed to follow the same trends. For instance, in all three species when basis sets with diffuse functions on heavy atoms are used [i.e., 6-

311+G(d,p), 6-311+G(2df,p) and 6-311+G(2df,2p)], the resulting uncorrected binding energies (ΔE_{uncorr}) are within -2.3–8.6 kJ mol⁻¹ of the corresponding BSSE corrected (ΔE_{corr}) benchmark energy. Furthermore, the calculated BSSE corrections for these basis sets are also quite small, all lying in the range of 1.2–8.0 kJ mol⁻¹ (see **Table 7.1**). In contrast, the ΔE_{uncorr} values obtained when using basis sets that do not include diffuse functions on heavy atoms [i.e., 6-311G(d,p), 6-311G(df,p) and 6-311G(2d,p)] are all significantly higher than their corresponding benchmark ΔE_{corr} values by 15.7–24.5 kJ mol⁻¹ (**Table 7.1**). In addition, the calculated BSSE correction for each of these basis sets is now larger, ranging from 8.7–20.6 kJ mol⁻¹. It is also noted that the calculated ΔE_{corr} energies are less basis set dependent than the ΔE_{uncorr} energies. Indeed, the absolute ranges for the $\Delta E_{corr}/\Delta E_{uncorr}$ energies for the Na⁺, K⁺ and Mg²⁺ complexes are 18.9/10.0, 15.4/7.3 and 33.1/20.5 kJ mol⁻¹, respectively.

Thus, a suitable basis set that can provide reliable structures *and* binding energies for each of the above three metal ion complexes must include both diffuse and a second set of *d*-functions on heavy atoms. In addition, it is also noted that the largest difference between ΔE_{uncorr} obtained at the B3LYP/6-311+G(2df,p) level of theory and the corresponding benchmark ΔE_{corr} occurs for the Na⁺-ribose complex and is only 3.5 kJ mol⁻¹! Therefore, all optimized geometries and binding energies for Na⁺, K⁺ and Mg²⁺ complexes reported in this chapter were obtained at the B3LYP/6-311+G(2df,p) level of theory *without* inclusion of BSSE corrections. For simplicity, all binding energies discussed below are abbreviated as ΔE .

It is noted that the 6-311+G(2df,p) basis set is not defined for Cd. For consistency, however, the basis set used in this chapter for Cd²⁺ includes the effective core potential (ECP) of Stevens *et. al.* (SBKJ)⁴⁷ further augmented with diffuse *d*- ($\alpha_d=0.075$) and *2fg*-

($\alpha_{1f}=2.0$, $\alpha_{2f}=0.3$, $\alpha_g=0.775$) polarization functions. We note that such a basis set has been previously used successfully to study related Cd^{2+} -biomolecule complexes.⁴⁸⁻⁵⁰

Finally it should be noted that the ribose sugar is able to adopt several possible conformations. Previously, the C2'- and C3'-*endo* conformers have been found to be preferred in nucleosides and nucleotides.⁵¹ At the B3LYP/6-311+G(2df,p) level of theory the latter conformer was found to be slightly more stable (**Figure 7.1**). Unless otherwise noted, this conformation is found to still be preferred upon binding of a metal ion.

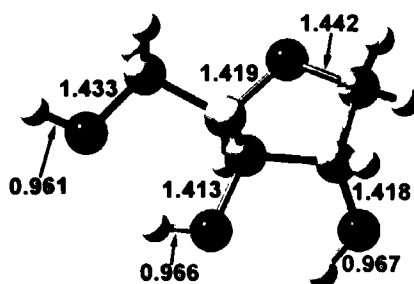


Figure 7.1 Optimized structure [B3LYP/6-311+G(2df,p)] of the C3'-*endo* ribose sugar conformation, in which the 3'-carbon is on the same side as the 5'-carbon. Selected distances shown are in angstroms.

7.3.2 Complexes of Ribose with Metal Ions

7.3.2.1 Na^+ -Ribose Complexes

A number of complexes were obtained upon ligation of Na^+ with a ribose sugar, all of which involved binding of the metal ion via various numbers of the ribose's oxygen centres. The optimized structures for each of these complexes obtained at the B3LYP/6-311+G(2df,p) level of theory are given in **Figure 7.2** while the corresponding binding energies are listed in **Table 7.2**.

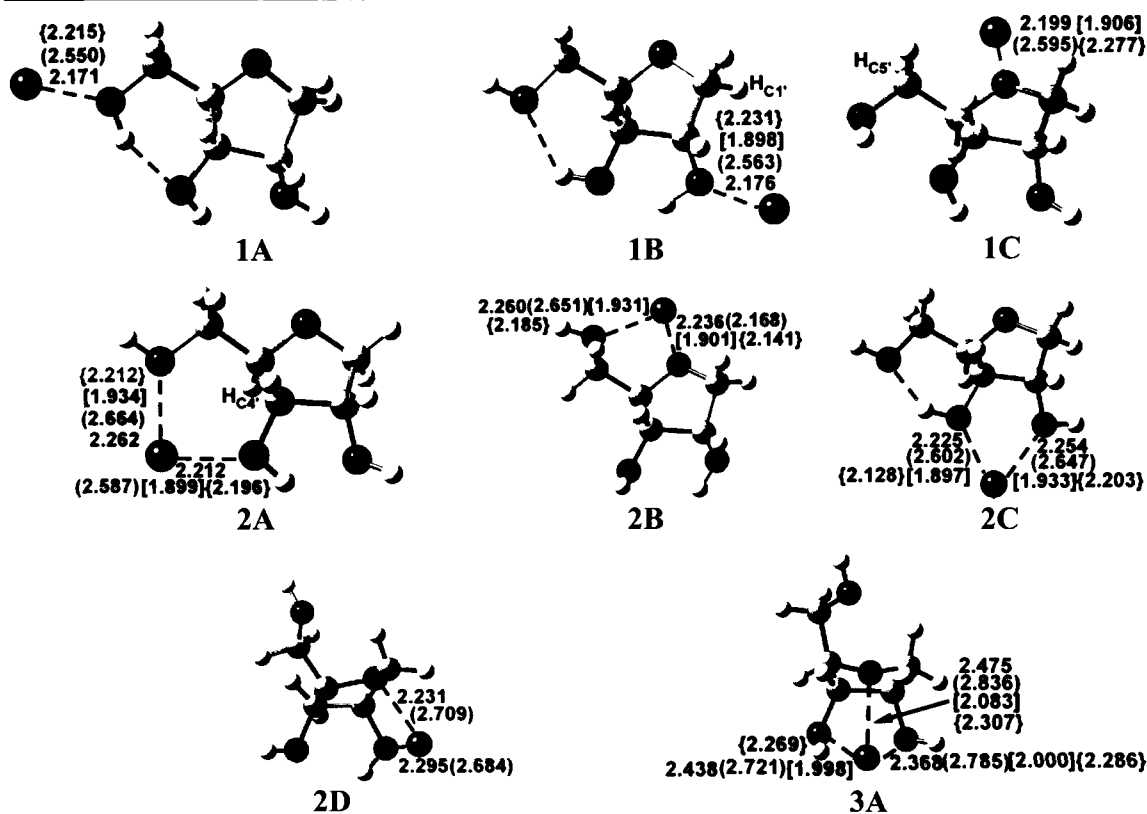


Figure 7.2 Optimized structures [B3LYP/6-311+G(2df,p)] of the complexes of ribose with metal ions (Na⁺, K⁺, Mg²⁺, Cd²⁺). Selected distances (angstroms) are shown without brackets for Na⁺ and in round, square and curly brackets for K⁺, Mg²⁺ and Cd²⁺, respectively.

Three monodentate complexes were obtained. Specifically, the Na⁺ ion is coordinated via the sugar's 5'-O (**1A**_{Na⁺}), 2'-O (**1B**_{Na⁺}) or ring 4'-O (**1C**_{Na⁺}) oxygen centres with $r(\text{Na}^+\cdots\text{O})$ distances of 2.171, 2.176 and 2.199 Å, respectively. The inverse of this trend in interaction distances is observed in the calculated binding energies. That is, ηE for **1A**_{Na⁺} (129.0 kJ mol⁻¹) is only slightly higher than that of **1B**_{Na⁺} (125.7 kJ mol⁻¹), while that of **1C**_{Na⁺} is noticeably less at 104.6 kJ mol⁻¹. In all three complexes, however, the binding of Na⁺ at a particular oxygen causes both the corresponding C—O and O—H (if appropriate) distances to be lengthened. For example, in **1A**_{Na⁺} the 5'-O—C and 5'-O—H distances increase from 1.433 to 1.455 Å and from 0.961 to 0.980 Å, respectively.

(Table E.1 in Appendix E). It should be noted that no monodentate metal ion–ribose complexes, for any metal ion considered herein, were found in which the ion was only coordinated to the sugar via its 3'-O centre.

Table 7.2 Calculated^a Binding Energies (ΔE) in kJ mol^{-1} of the Complexes of Ribose with Metal Ions (Na^+ , K^+ , Mg^{2+} , Cd^{2+}) Obtained in This Chapter.

Structure	ΔE	Structure	ΔE	Structure	ΔE
1A_{Na⁺}	129.0	1B_{Na⁺}	125.7	1C_{Na⁺}	104.6
2A_{Na⁺}	180.3	2B_{Na⁺}	177.8	2C_{Na⁺}	174.3
2D_{Na⁺}	159.8	3A_{Na⁺}	162.7		
1A_{K⁺}	91.9	1B_{K⁺}	88.6	1C_{K⁺}	67.5
2A_{K⁺}	123.2	2B_{K⁺}	124.8	2C_{K⁺}	119.6
2D_{K⁺}	108.3	3A_{K⁺}	110.4		
1A_{Mg²⁺}	—	1B_{Mg²⁺}	550.1	1C_{Mg²⁺}	540.0
2A_{Mg²⁺}	727.4	2B_{Mg²⁺}	712.2	2C_{Mg²⁺}	707.8
2D_{Mg²⁺}	—	3A_{Mg²⁺}	735.6		
1A_{Cd²⁺}	583.9	1B_{Cd²⁺}	600.9	1C_{Cd²⁺}	601.1
2A_{Cd²⁺}	704.5	2B_{Cd²⁺}	682.6	2C_{Cd²⁺}	679.2
2D_{Cd²⁺}	—	3A_{Cd²⁺}	697.2		

^a B3LYP/6-311+G(2df,p) + ZPVE

Four bidentate Na^+ –ribose complexes were found (**2A_{Na⁺}**–**2D_{Na⁺}**) and are shown in **Figure 7.2**. In **2A_{Na⁺}** the metal ion coordinates via the riboses 5'- and 3'-oxygens while in the **2B_{Na⁺}**, **2C_{Na⁺}** and **2D_{Na⁺}** complexes it is coordinated via their 4'-/5'-, 2'-/3'- and 2'-/4'-oxygens respectively. As can be seen, the $\text{Na}^+\cdots\text{O}$ distances are all longer compared to those observed in the monodentate complexes. This is due to the fact that the metal ion now interacts with the sugar via two oxygen centres. However, as observed for the monodentate complexes, the C—O and O—H bonds for those oxygens ligated to the Na^+ are again lengthened, though now to a lesser extent (Table E.1 in Appendix E). For

example, in $2A_{Na^+}$ the 3'-O—C and 5'-O—C distances increase from 1.413 to 1.428 Å and 1.433 to 1.451 Å, respectively. For all of these complexes, ΔE is found to be higher than that of the most strongly bound monodentate complex $1A_{Na^+}$ by 30.8—51.3 kJ mol⁻¹. Indeed, the largest binding energy is obtained for $2A_{Na^+}$ for which $\Delta E=180.3$ kJ mol⁻¹. It should be noted, however, that the calculated binding energy of $2A_{Na^+}$, $2B_{Na^+}$ and $2C_{Na^+}$ all lie within 6 kJ mol⁻¹ while that of $2D_{Na^+}$ is decidedly lower by at least 14.5 kJ mol⁻¹. The noticeably lower ΔE of this latter complex likely reflects the less favorable combined positioning of the 2'- and 4'-oxygens for binding of the Na⁺ ion.

In addition, one complex ($3A_{Na^+}$) was found in which the Na⁺ was tri-coordinated to the ribose via the 2'- and 3'-hydroxyl oxygens and the 4'-ring oxygen with distances of 2.368, 2.348 and 2.475 Å respectively. These are the longest Na⁺...O interactions observed of all the Na⁺-ribose complexes. Furthermore, the ribose itself has now adopted an O4'-*exo* conformation. While this change likely enables the ring oxygen to better interact with the metal ion, the binding energy of $3A_{Na^+}$ is just 162.7 kJ mol⁻¹. That is, it is 17.6 kJ mol⁻¹ *less* than that of the lowest energy bidentate complex $2A_{Na^+}$. Furthermore, it is just 2.9 kJ mol⁻¹ higher than that for the weakest bidentate complex $2D_{Na^+}$. This surprisingly low binding energy is probably due in large part to the fact that the ribose is no longer in its more preferred C2'- or C3'-*endo* conformations.

7.3.2.2 K⁺-Ribose Complexes

In general, the complexation of K⁺ with ribose showed trends similar to those observed for Na⁺. Structurally, for example, three mono- ($1A-C_{K^+}$), four bi- ($2A-D_{K^+}$) and one tridentate ($3A_{K^+}$) complexes were obtained in which the K⁺ is bound to the ribose at the same positions as previously observed for Na⁺ (see **Figure 7.2**). In addition, in $3A_{K^+}$ the ribose sugar again adopts an O4'-*exo* conformation. Furthermore, a lengthening of the C—O and O—H bonds for those oxygens ligated to the K⁺ is again observed. For

example, in $1A_{K^+}$ the 5'-O—C and 5'-O—H bonds lengthen by 0.032 and 0.012 Å to 1.450 and 0.977 Å respectively (Table E.1 in Appendix E). Energetically, some similar trends to those of the Na^+ –ribose complexes were also observed. In particular, the monodentate species again have the lowest binding energies, which all now lie in the range of 67.5–91.9 kJ mol⁻¹ (see Table 7.2). The most stable bidentate complexes are again found to be **2A**, **2B** and **2C_{K⁺}** with binding energies of 123.2, 124.8 and 119.6 kJ mol⁻¹ respectively. That is, they all within just 5.2 kJ mol⁻¹ of each other. It is noted that these are 57.1, 53.0 and 54.7 kJ mol⁻¹ lower than those of the corresponding Na^+ –ribose complexes. Furthermore, the bidentate complex **2D_{K⁺}** in which the K^+ ion is coordinated to the ribose via its 2'- and ring 4'-oxygen is again found to have a distinctly lower ΔE (108.3 kJ mol⁻¹). Finally, the tridentate complex (**3A_{K⁺}**) again has a binding energy (110.4 kJ mol⁻¹) only marginally higher by 2.1 kJ mol⁻¹ of the weakest bidentate complex **2D_{K⁺}** (see Table 7.2).

However, the K^+ –ribose complexes also exhibit some key differences from their analogous Na^+ –ribose complexes. First, the $K^+ \cdots O$ interaction distances are markedly larger than those observed for the $Na^+ \cdots O$ interactions. For example, the $K^+ \cdots O$ -5' distance in $1A_{K^+}$ is 0.379 Å longer than that of $Na^+ \cdots O$ in $1A_{Na^+}$. This is due in large part to the greater ionic radius of K^+ compared to that of Na^+ ($r_{K^+}=1.33$ Å, $r_{Na^+}=0.97$ Å).⁵² This may also help to explain the fact that the calculated binding energies are much lower than those of the corresponding Na^+ –ribose complexes by 37.1–57.1 kJ mol⁻¹. Perhaps most importantly, however, the most stable K^+ –ribose complex is found to be **2B_{K⁺}** in which the K^+ ion binds via both the 4'- and 5'-ribose oxygens rather than **2A_{K⁺}**. However, it should be noted that $\Delta E(2B_{K^+})$ is only marginally higher than $\Delta E(2A_{K^+})$ by 1.6 kJ mol⁻¹.

7.3.2.3 Mg^{2+} -Ribose Complexes

The Mg^{2+} ion has twice the positive charge of Na^+ or K^+ , and has a markedly smaller ionic radius ($r_{Mg^{2+}}=0.66 \text{ \AA}$).⁵² Thus, it shows a number of important differences in its binding with ribose. First, the corresponding $1A_{Mg^{2+}}$ complex in which Mg^{2+} is bound to the ribose solely via its 5'-oxygen was not a minimum on the B3LYP/6-311+G(2df,p) potential energy surface. Rather, attempts to optimize $1A_{Mg^{2+}}$ resulted in the proton of 5'-OH transferring to the nearby 3'-OH group. Similarly, the corresponding $2D_{Mg^{2+}}$ complex in which the Mg^{2+} ion binds via both the 2'- and ring 4'-oxygens also was not found to be a minimum at this level of theory but instead upon optimization collapses to the tridentate complex $3A_{Mg^{2+}}$. In addition, unlike Na^+ and K^+ , $3A_{Mg^{2+}}$ has the largest binding energy ($\Delta E=735.6 \text{ kJ mol}^{-1}$) of all the Mg^{2+} -ribose complexes. This is likely due in part to the smaller ionic radius of Mg^{2+} , thus enabling the ion to better position itself between all three oxygens. The larger charge and smaller radius also result in shorter $Mg^{2+}\cdots O$ interaction distances and larger binding energies compared with their Na^+ and K^+ counterparts. Furthermore, it is also noted that the difference in binding energies between $2A_{Mg^{2+}}$, $2B_{Mg^{2+}}$ and $2C_{Mg^{2+}}$ is now decidedly larger at 19.6 kJ mol^{-1} compared to that observed for both Na^+ and K^+ (see above).

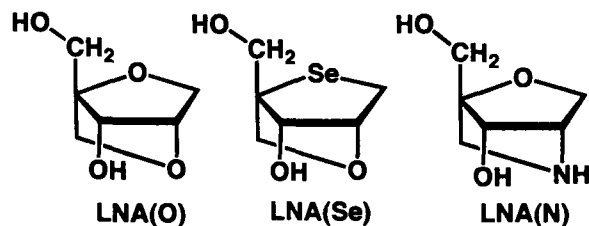
7.3.2.4 Cd^{2+} -Ribose Complexes

Like Mg^{2+} , Cd^{2+} is also a divalent cation. However, it has a significantly larger ionic radius ($r_{Cd^{2+}}=0.97 \text{ \AA}$)⁵⁴ that is in fact quite close to that of Na^+ . Thus, the complexation of Cd^{2+} with ribose shows a number of important differences not only to the Mg^{2+} -ribose complexes but also with the Na^+/K^+ -ribose species. For example, Cd^{2+} was found to bind to ribose in almost all the same positions as observed for Na^+ . The only exception is that no equivalent $2D_{Cd^{2+}}$ -type complex, in which the ion is bound via the 2'- and ring 4'-oxygens, was obtained. It should be noted that the $Cd^{2+}\cdots O$ distances in the monodentate species ($1A_{Cd^{2+}}$, $1B_{Cd^{2+}}$, and $1C_{Cd^{2+}}$) are *longer* than those in the bidentate

complexes ($2A_{Cd^{2+}}$, $2B_{Cd^{2+}}$, and $2C_{Cd^{2+}}$). However, the largest $Cd^{2+}\cdots O$ interaction distances are observed in the tridentate complex $3A_{Cd^{2+}}$. In addition, the $Cd^{2+}\cdots O$ distances are longer than in their corresponding Mg^{2+} -ribose complexes due to the larger radius of Cd^{2+} (see above). Comparing the metal-oxygen interaction distances in all complexes it is found that the distances in the monodentate complexes follow the sequence: $r(Mg^{2+}\cdots O) < r(Na^+\cdots O) < r(Cd^{2+}\cdots O) < r(K^+\cdots O)$ while those in the bidentate and tridentate complexes follow the sequence: $r(Mg^{2+}\cdots O) < r(Cd^{2+}\cdots O) < r(Na^+\cdots O) < r(K^+\cdots O)$. Thus, notably, the order of the $Na^+\cdots O$ versus $Cd^{2+}\cdots O$ distances depends on the binding mode of the ion.

Energetically, while the binding energies of $1B_{Cd^{2+}}$ and $1C_{Cd^{2+}}$ differ by just 0.2 kJ mol^{-1} , they are both markedly higher than that of $1A_{Cd^{2+}}$ by approximately 17 kJ mol^{-1} . In contrast, for Mg^{2+} -ribose the binding energies of $1B_{Mg^{2+}}$ and $1C_{Mg^{2+}}$ differ by over 10 kJ mol^{-1} and no $1A_{Mg^{2+}}$ species was obtained (see above). Furthermore, the binding energies of the three bidentate complexes $2A_{Cd^{2+}}$ ($\Delta E=704.5\text{ kJ mol}^{-1}$), $2B_{Cd^{2+}}$ ($\Delta E=682.6\text{ kJ mol}^{-1}$) and $2C_{Cd^{2+}}$ ($\Delta E=679.2\text{ kJ mol}^{-1}$) are found to differ from each other by as much as 25.3 kJ mol^{-1} ($\Delta E[2A_{Cd^{2+}}-2C_{Cd^{2+}}]$). This is the largest spread of bidentate-binding energies observed for any of the four metal ions considered herein. Interestingly, similar to that found for Na^+ and K^+ but in contrast to that observed for Mg^{2+} , the binding energy of $3A_{Cd^{2+}}$ ($\Delta E=697.2\text{ kJ mol}^{-1}$) is less than that of the most stable bidentate complex $2A_{Cd^{2+}}$.

Scheme 7.3 Schematic illustration of locked ribose considered in this chapter.



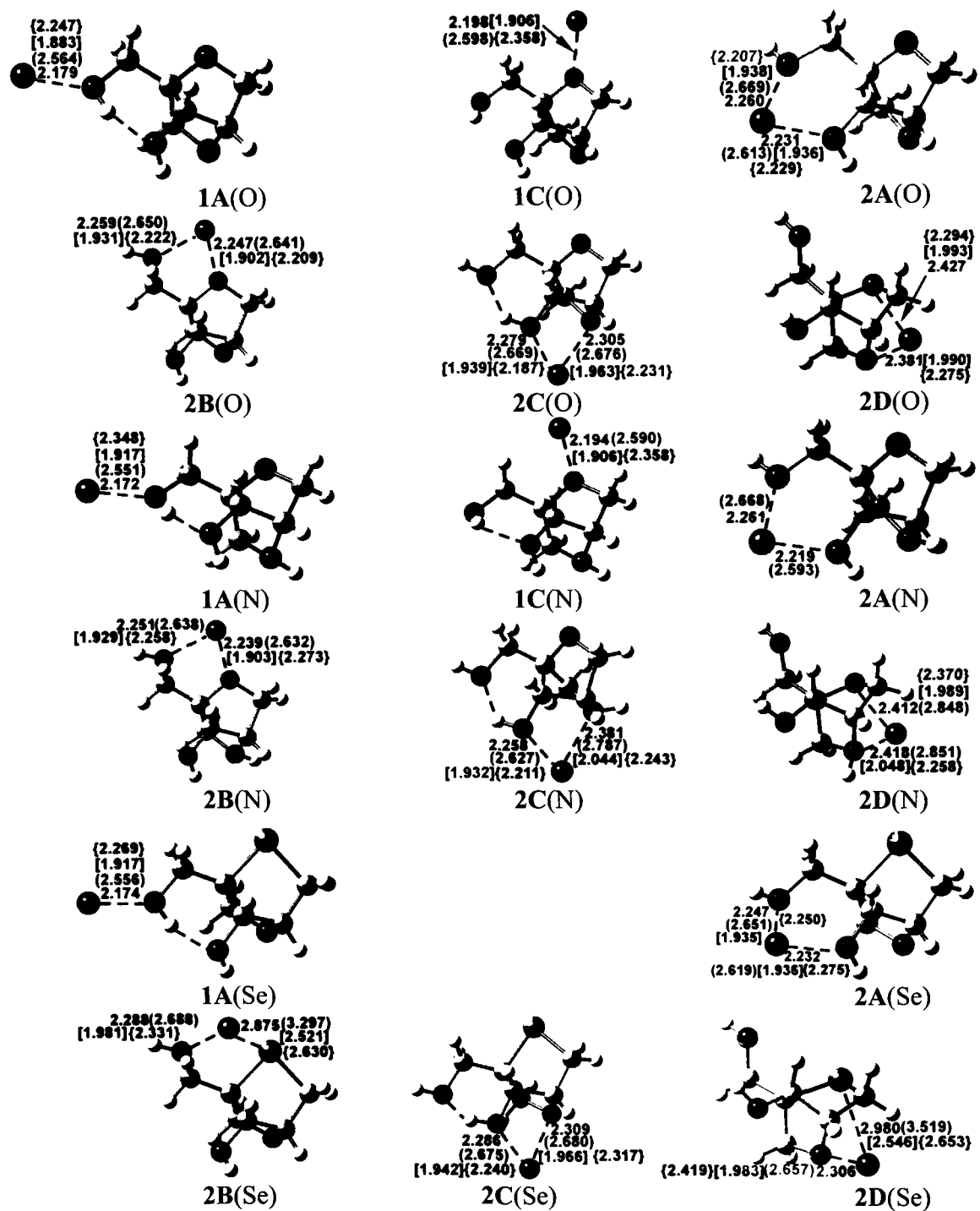


Figure 7.3 Optimized structures [B3LYP/6-311+G(2df,p)] of the complexes of locked ribose with metal ions (Na⁺, K⁺, Mg²⁺, Cd²⁺). Selected distances (angstroms) are shown without brackets for Na⁺ and in round, square and curly brackets for K⁺, Mg²⁺ and Cd²⁺, respectively.

Table 7.3 Calculated^a Binding Energies (ΔE) in kJ mol^{-1} of the Complexes of Locked Ribose with Metal Ions ($M^{n+} = \text{Na}^+, \text{K}^+, \text{Mg}^{2+}, \text{Cd}^{2+}$) Obtained in This Chapter.

Structure	ΔE	Structure	ΔE	Structure	ΔE
1A(O) _{Na⁺}	127.8	1C(O) _{Na⁺}	106.4	2A(O) _{Na⁺}	168.5
2B(O) _{Na⁺}	172.6	2C(O) _{Na⁺}	172.1	2D(O) _{Na⁺}	103.3
1A(N) _{Na⁺}	146.9	1C(N) _{Na⁺}	123.3	2A(N) _{Na⁺}	191.7
2B(N) _{Na⁺}	188.9	2C(N) _{Na⁺}	197.4	2D(N) _{Na⁺}	139.3
1A(Se) _{Na⁺}	129.5	1C(Se) _{Na⁺}	—	2A(Se) _{Na⁺}	164.6
2B(Se) _{Na⁺}	148.5	2C(Se) _{Na⁺}	170.2	2D(Se) _{Na⁺}	127.2
1A(O) _{K⁺}	91.3	1C(O) _{K⁺}	70.2	2A(O) _{K⁺}	114.1
2B(O) _{K⁺}	120.6	2C(O) _{K⁺}	123.1	2D(O) _{K⁺}	—
1A(N) _{K⁺}	109.1	1C(N) _{K⁺}	85.6	2A(N) _{K⁺}	135.6
2B(N) _{K⁺}	135.2	2C(N) _{K⁺}	140.9	2D(N) _{K⁺}	92.1
1A(Se) _{K⁺}	91.9	1C(Se) _{K⁺}	—	2A(Se) _{K⁺}	109.9
2B(Se) _{K⁺}	100.0	2C(Se) _{K⁺}	121.5	2D(Se) _{K⁺}	82.2
1A(O) _{Mg²⁺}	527.3	1C(O) _{Mg²⁺}	531.6	2A(O) _{Mg²⁺}	710.4
2B(O) _{Mg²⁺}	700.9	2C(O) _{Mg²⁺}	683.4	2D(O) _{Mg²⁺}	582.5
1A(N) _{Mg²⁺}	581.9	1C(N) _{Mg²⁺}	571.2	2A(N) _{Mg²⁺}	—
2B(N) _{Mg²⁺}	726.5	2C(N) _{Mg²⁺}	757.7	2D(N) _{Mg²⁺}	667.5
1A(Se) _{Mg²⁺}	576.7	1C(Se) _{Mg²⁺}	—	2A(Se) _{Mg²⁺}	738.7
2B(Se) _{Mg²⁺}	699.1	2C(Se) _{Mg²⁺}	701.6	2D(Se) _{Mg²⁺}	658.2
1A(O) _{Cd²⁺}	606.9	1C(O) _{Cd²⁺}	605.8	2A(O) _{Cd²⁺}	689.0
2B(O) _{Cd²⁺}	684.0	2C(O) _{Cd²⁺}	668.2	2D(O) _{Cd²⁺}	566.4
1A(N) _{Cd²⁺}	693.5	1C(N) _{Cd²⁺}	667.7	2A(N) _{Cd²⁺}	—
2B(N) _{Cd²⁺}	731.6	2C(N) _{Cd²⁺}	768.1	2D(N) _{Cd²⁺}	674.0
1A(Se) _{Cd²⁺}	669.2	1C(Se) _{Cd²⁺}	—	2A(Se) _{Cd²⁺}	733.0
2B(Se) _{Cd²⁺}	755.2	2C(Se) _{Cd²⁺}	707.2	2D(Se) _{Cd²⁺}	723.0

^a B3LYP/6-311+G(2df,p) + ZPVE

7.3.3 Complexes of Locked Ribose with Metal Ions

In this chapter, three types of locked ribose were considered: 2'-O,4'-C-methylene ribose, 2'-N,4'-C-methylene ribose and 2'-O,4'-C-methylene 4'-selenoribose. For

simplicity, they are abbreviated as LNA(O), LNA(N) and LNA(Se) respectively (**Scheme 7.3**). In addition, the metal–LNA complexes are named as, for example, **1A(O)**, **1A(N)** and **1A(Se)**. All the optimized structures of metal–LNA complexes obtained with selected bond distances are shown in **Figure 7.3** while the corresponding binding energies are given in **Table 7.3**.

7.3.3.1 Metal–LNA(O) Complexes

The binding of all four metal ions with LNA(O) shows some similar trends to those detailed above for when they are complexed with native ribose. For instance, for those oxygens ligated to the metal ion, their C—O and O—H (if appropriate) bonds are generally lengthened upon complexation. For example, the 5'-O—C and 5'-O—H distances in **1A(O)_{Na⁺}** increase from 1.431 to 1.453 Å and from 0.961 to 0.982 Å respectively upon binding of Na⁺ to the 5'-oxygen.

Importantly, however, there are a number of significant differences due to the inclusion of the —CH₂— bridge. Overall it is generally found that the metal–oxygen distances are now longer and binding energies smaller than those of the corresponding metal–ribose complexes. More specifically, only two monodentate complexes were now found in which the metal ions bind with either 5'- [**1A(X)_{Mⁿ⁺}**] or ring 4'-oxygen [**1C(X)_{Mⁿ⁺}**]. That is, no complexes were obtained in which the metal ions were *only* coordinated via the 2'-oxygen. Indeed, any attempts to optimize such structures led to either **2C(O)** or **2D(O)**. This likely reflects the fact that once the 2'-O is involved in the cross-link its lone pairs are directionally constrained to lie on the same sides as the 3'- and 4'-oxygens (see **Figure 7.3**). Another significant difference is that a complex in which the Mg²⁺ ion singly coordinated to the LNA(O) via the 5'-oxygen [**1A(O)_{Mg²⁺}**] was found to in fact be a stable minimum. In contrast, as detailed above, an analogous complex was not stable for the native ribose due to a coordination induced proton transfer from the 5'-OH

to 3-OH group (see above). However, the 5'-O—H bond in $2A(O)_{Mg^{2+}}$ is lengthened considerably to 1.063 Å while concomitantly the 5'-OH...O-3' distance is shortened significantly to 1.444 Å. Thus, the 5'-OH group could be thought of as partially deprotonated. Finally, for Na^+ , K^+ and Cd^{2+} the most stable monodentate complexes are $2A(O)$ with binding energies of 127.8, 91.3 and 606.9 $kJ\ mol^{-1}$ respectively. For Mg^{2+} the binding energy of the $1A(O)$ complex is 4.3 $kJ\ mol^{-1}$ lower than that of the alternate $1C(O)_{Mg^{2+}}$ complex (Table 7.3).

Four bidentately coordinated complexes were found for all metal ions except K^+ . Specifically, in $2A(O)$ the metals bind via the 3'- and 5'-oxygens, while in $2B(O)$, $2C(O)$ and $2D(O)$ the ions bind via the ring 4'- and 5'-, 2'- and 3'-, and 2'- and ring 4'-oxygens, respectively. Remarkably, unlike that observed for the bidentate complexes with ribose, complexes were obtained in which the Mg^{2+} [$2D(O)_{Mg^{2+}}$] and Cd^{2+} [$2D(O)_{Cd^{2+}}$] ions were bound only via the 2'- and ring 4'-oxygens. This is due to the fact that introduction of the $-CH_2-$ bridge prevents the formation of tridentate complexes. In addition, unlike for ribose, no corresponding $2D(O)_{K^+}$ species could be obtained. This may be due to the combined effects of constraining the position of the 2'-oxygen in LNA(O) coupled with the fact that K^+ has the largest atomic radii in all ions presently considered. That is, it cannot 'fit'. Indeed, the distance between 2'- and 4'-oxygens in LNA(O) is 2.784 Å while in contrast, in the ribose it is considerably larger at 3.152 Å. Interestingly, Na^+ is the only metal ion that is able to bind in this position with both ribose and LNA(O). However, it should be noted that the binding energy of $2D(O)_{Na^+}$ (103.3 $kJ\ mol^{-1}$) is considerably smaller than that of the analogous Na^+ -ribose complex ($2D_{Na^+}$; 159.8 $kJ\ mol^{-1}$) by 56.5 $kJ\ mol^{-1}$ (cf. Table 7.2). This is in fact the largest difference in binding energies observed between complexes in which the metal ion is bound to the ribose and LNA(O) via the *same* oxygens.

Similar to that observed for the metal–ribose complexes, the binding energies obtained for those bidentate complexes in which the metal is coordinated via the 3′-/5′-, 4′-/5′ and 2′-/3′-oxygens are all generally close to each other. For instance for Na⁺, K⁺, Mg²⁺ and Cd²⁺ they differ by at most 4.1, 9.0, 27.0 and 20.8 kJ mol⁻¹ respectively. Again, the largest variation is observed for the more positively charged cations. Now, however, the binding energy for the 2′-/4′-oxygen bidentately coordinated complexes [**2D(O)**], if they exist at all, are all drastically smaller. In fact, for Na⁺, Mg²⁺ and Cd²⁺ they are smaller than those of the most stable bidentate complex by 69.3, 127.9 and 122.6 kJ mol⁻¹ respectively.

Finally, for Na⁺ and K⁺ the lowest energy complexes are **2B(O)**_{Na⁺} and **2C(O)**_{K⁺}. Furthermore, both dications now prefer to bind bidentately via 3′-O and 5′-O [**2A(O)**_{M²⁺}]. These results differ from the preferred binding modes observed for ribose (cf. **Figure 7.2** and **Table 7.2**). Thus, the introduction of a –CH₂– bridge has clearly influenced the binding and positioning of the metal ions.

7.3.3.2 Metal–LNA(N) Complexes

In the LNA(N) sugar the bridging 2′-oxygen of LNA(O) has now been replaced by the isoelectronic –NH– group. Thus, while the nitrogen centre is more basic it only possesses one lone pair of electrons.

In general, the metal ions bind with LNA(N) in a similar manner to that described above for LNA(O). There are, however, two important exceptions. First, in contrast to that observed for LNA(O), a stable complex was obtained in which the K⁺ ion is bound via both the 2′-heteroatom (N) and the 4′-ring oxygen [**2F(N)**_{K⁺}]. This may be due at least in part to the fact that in LNA(N) the distance between 4′-oxygen and 2′-heteroatom (2.818 Å) is slightly larger than in LNA(O) (2.784 Å), thus enabling the largest metal ion

of those considered (K^+) to 'fit'. Second, when the Mg^{2+} and Cd^{2+} ions bind bidentately via the 5'- and 3'-oxygens they cause the 3'-OH proton to transfer to the $-NH-$ moiety thus forming a bridging protonated $-NH_2^+$ group.

Energetically, the calculated binding energies for all metal-LNA(N) complexes are higher than those of their corresponding metal-LNA(O) complexes (see **Table 7.3**). For the monodentate complexes only a slight increase in the difference in binding energies between the **1A(N)** and **1C(N)** complexes is generally observed compared to the corresponding **1A(O)** and **1C(O)** complexes. The exception to this trend, however, is observed for Cd^{2+} for which they now differ by 25.8 kJ mol^{-1} compared to just 1.1 kJ mol^{-1} for LNA(O). In addition, unlike that observed for LNA(O), the preferred complex for all metal ions entails them binding via both the 2'-nitrogen and 3'-oxygen [**2C(N)_Mⁿ⁺**]. The binding energies (ΔE) for each of these complexes for Na^+ , K^+ , Mg^{2+} and Cd^{2+} are 197.4, 140.9, 757.7 and 768.1 kJ mol^{-1} respectively.

7.3.3.3 Metal-LNA(Se) Complexes

Unlike all the above complexes, only one monodentate binding manner is found for the metal-LNA(Se) complexes, in which metal ions bind via the 5'-oxygen. We were unable to find the complex in which the metal ions only bind via the 4'-selenium. Hence, there are in total five possible complexes, in which the metal ions do not directly bind with selenium in **1A(Se)**, **2A(Se)** and **2C(Se)**, while they bind with both selenium and an oxygen in **2B(Se)** and **2D(Se)** respectively

As the 4'-oxygen is replaced by the less electronegative selenium, the other oxygens are expected to become more negative. Indeed, the Mulliken charges for the 2', 3' and 5' oxygens in LNA(O) are -0.156, -0.392 and -0.399 respectively while those in LNA(Se) increase to -0.183, -0.415 and -0.414 respectively. Thus, ΔE 's of **1A(Se)**, **2A(Se)** and

2C(Se) are expected to be larger than the metal–LNA(O) counterpart **1A(O)**, **2A(O)** and **2C(O)**. This is indeed found for the divalent ions Mg^{2+} and Cd^{2+} , in which ΔE 's increase by 18.2—62.3 kJ mol^{-1} . In contrast, such effect is marginal for the monovalent ions Na^+ and K^+ in which ΔE 's of **1A(Se)**, **2B(Se)** and **2C(Se)** only change by -4.2—1.7 kJ mol^{-1} . For **2B(Se)** in which the metal ions directly interacts with 4'-selenium and 5'-oxygen, $\Delta E[\mathbf{2B(Se)}_{\text{Na}^+}]$, $\Delta E[\mathbf{2B(Se)}_{\text{K}^+}]$ and $\Delta E[\mathbf{2B(Se)}_{\text{Mg}^{2+}}]$ decrease by 24.1, 20.6 and 1.8 kJ mol^{-1} compared with $\Delta E[\mathbf{2B(O)}_{\text{Na}^+}]$, $\Delta E[\mathbf{2B(O)}_{\text{K}^+}]$ and $\Delta E[\mathbf{2B(O)}_{\text{Mg}^{2+}}]$. However, an increase of 71.2 kJ mol^{-1} is found for **2B(Se)** $_{\text{Cd}^{2+}}$. This can be explained in terms of Na^+ , K^+ and Mg^{2+} being hard Lewis acids. In contrast, Cd^{2+} is a soft Lewis acid. Therefore, it is more favorable for Cd^{2+} to interact with the soft Lewis base Se. Indeed, **2B(Se)** $_{\text{Cd}^{2+}}$ is found to be the most preferred position for Cd^{2+} . We note that this is different from Cd^{2+} –LNA(O) complexes. Thus, the substitution of oxygen by other heteroatoms is able to change metal ion binding properties. For **2D(Se)** in which the metal ions directly interacts with 4'-selenium and 2'-oxygen, increases of ΔE by 23.9 and 75.7 are found even for the hard acids Na^+ and Mg^{2+} (note that **2D(O)** $_{\text{K}^+}$ cannot be found). This may be due to the larger atomic size of selenium, which extends the room for metal ions binding, with the distance between 4'-selenium and 3'-oxygen at 3.155 Å compared with 2.784 Å in LNA(O). Thus, the increases of ΔE may come from a better interaction between metal ions and 2'-oxygen. Indeed, such distances are 0.075 and 0.007 Å shorter in **2D(Se)** than **2D(O)** for Na^+ and Mg^{2+} respectively. Furthermore, the largest increase of 156.6 kJ mol^{-1} in **2D(Se)** compared with **2D(O)** is found for Cd^{2+} . This is due to both the enlarged space and better interaction between the soft acid Cd^{2+} and soft base Se.

7.4 Conclusions

In this chapter, the interactions of metal ions (Na^+ , K^+ , Mg^{2+} and Cd^{2+}) with ribose and selected locked ribose: 2'-*O*,4'-*C*-methylene ribose, 2'-*N*,4'-*C*-methylene ribose and 2'-*O*,4'-*C*-methylene 4'-selenoribose were studied using the B3LYP method.

First, a typical bidentate complex in which the metal ions (Na^+ , K^+ , Mg^{2+}) bind via the 2'- and 3'-oxygens was chosen to perform a systematic basis set assessment in order to find an appropriate basis set to obtain both the reliable geometries and binding energies. As for the geometries, we found that the monovalent ions Na^+ and K^+ are more sensitive to the inclusion of diffuse functions while the divalent metal ion Mg^{2+} is more sensitive to addition of a second set of *d*-functions on heavy atoms. In addition, diffuse functions are also important to obtain the accurate binding energies for all these three metal ions. Therefore, our calculations were performed using the 6-311+G(2df,p) basis set for Na^+ , K^+ , Mg^{2+} complexes and one that includes an ECP for core electrons and further augmented with similar diffuse and polarization functions for valence electrons was used for Cd^{2+} within its complexes.

For the metal–ribose complexes, all the four metal ions are able to bind with the ribose monodentately, bidentately and tridentately. The binding energies of the K^+ –ribose complexes are smaller than the Na^+ –ribose complexes since K^+ is larger than Na^+ . Furthermore, the two divalent metal ions Mg^{2+} and Cd^{2+} have larger binding energies than Na^+ and K^+ due to the larger formal charges.

Modification of the ribose is able to change its binding with metal ions. For example, replacing the 2'-oxygen with nitrogen increases the binding energies of the complexes for all the metal ions. In addition, replacing the 4'-oxygen to a softer Lewis

base, such as selenium, is also able to increase the binding energy with some soft metal ions, such as Cd^{2+} .

References

- (1) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 4th ed.; W H. Freeman and Company: New York, 2004.
- (2) Rodgers, M. T.; Armentrout, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 8548.
- (3) Russo, N.; Toscano, M.; Grand, A. *J. Am. Chem. Soc.* **2001**, *123*, 10272.
- (4) Russo, N.; Toscano, M.; Grand, A. *J. Phys. Chem. B* **2001**, *105*, 4735.
- (5) Zhu, W.; Luo, X.; Pua, C. M.; Tan, X.; Shen, J.; Gu, J.; Chen, K.; Jiang, H. *J. Phys. Chem. A* **2004**, *108*, 4008.
- (6) Kabeláč, M.; Hobza, P. *J. Phys. Chem. B* **2006**, *110*, 14515.
- (7) Šponer, J. E.; Sychrovsky, V.; Hobza, P.; Šponer, J. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2772.
- (8) Munoz, J.; Sponer, J.; Hobza, P.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2001**, *105*, 6051.
- (9) Russo, N.; Toscano, M.; Grand, A. *J. Phys. Chem. A* **2003**, *107*, 11533.
- (10) Petrov, A. S.; Pack, G. R.; Lamm, G. *J. Phys. Chem. B* **2004**, *108*, 6072.
- (11) Rulisek, L.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 1913.
- (12) Gresh, N.; Sponer, J. E.; Spacková, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669.
- (13) Kosenkov, D.; Gorb, L.; Shishkin, O. V.; Šponer, J.; Leszczynski, J. *J. Phys. Chem. B* **2008**, *112*, 150.
- (14) Lamsabhi, A. M.; Mó, O.; Yáñez, M.; Boyd, R. J. *J. Chem. Theory Comput.* **2008**, *4*, 1002.

- (15) Moussatova, A.; Vázquez, M. V.; Martínez, A.; Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V.; Pedersen, D. B.; Simard, B. *J. Phys. Chem. A* **2003**, *107*, 9415.
- (16) Pedersen, D. B.; Simard, B.; Martinez, A.; Moussatova, A. *J. Phys. Chem. A* **2003**, *107*, 6464.
- (17) Noguera, M.; Bertran, J.; Sodupe, M. *J. Phys. Chem. A* **2004**, *108*, 333.
- (18) Pavelka, M.; Shukla, M. K.; Leszczynski, J.; Burda, J. V. *J. Phys. Chem. A* **2008**, *112*, 256.
- (19) Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 166.
- (20) Leclerc, F.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 3395.
- (21) Stahley, M. R.; Strobel, S. A. *Science* **2005**, *309*, 1587.
- (22) Stahley, M. R.; Strobel, S. A. *Curr. Opin. Chem. Biol.* **2006**, *16*, 319.
- (23) Honcharenko, D.; Varghese, O. P.; Plashkevych, O. P.; Barman, J.; Chattopadhyaya, J. *J. Org. Chem.* **2006**, *71*, 299.
- (24) Pradeepkumar, P. I.; Cheruku, P.; Plashkevych, O.; Acharya, P.; Gohil, S.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2004**, *126*, 11484.
- (25) Koshkin, A. A.; Singh, S. K.; Nielson, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Wengel, J. *Tetrahedron* **1998**, *54*, 3607.
- (26) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K., I.; In, Y.; Ishida, T.; Imanishi, T. *Tetrahedron Lett.* **1997**, *38*, 8735.
- (27) Singh, S. K.; Kumar, R.; Wengel, J. *J. Org. Chem.* **1998**, *63*, 10035.
- (28) Singh, S. K.; Kumar, R.; Wengel, J. *J. Org. Chem.* **1998**, *63*, 6078.
- (29) Kaur, H.; Babu, B. R.; Maiti, S. *Chem. Rev.* **2007**, *107*, 4672.
- (30) Kumar, N.; Patowary, A.; Sivasubbu, S.; Petersen, M.; Maiti, S. *Biochemistry* **2008**, *47*, 13179.
- (31) Gunaga, P.; Moon, H. R.; Choi, W. J.; Shin, D. H.; Park, J. G.; Jeong, L. S. *Curr. Med. Chem.* **2004**, *11*, 2585.

- (32) Jeong, L. S.; Tosh, D. K.; Kim, H. O.; Wang, T.; Hou, X.; Yun, H. S.; Kwon, Y.; Lee, S. K.; Choi, J.; Zhao, L. X. *Org. Lett.* **2008**, *10*, 209.
- (33) Watts, J. K.; Johnston, B. D.; Jayakanthan, K.; Wahba, A. S.; Pinto, B. M.; Damha, M. J. *J. Am. Chem. Soc.* **2008**, *130*, 8578.
- (34) Jayakanthan, K.; Johnston, B. D.; Pinto, B. M. *Carbohydrate Res.* **2008**, *343*, 1790.
- (35) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.
- (36) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (38) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (39) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.
- (40) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

-
- (41) Corral, I.; M6, O.; Y6fnev, M.; Scott, A. P.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 10456
- (42) Hartmann, M.; Radom, L. *J. Phys. Chem. A* **2000**, *104*, 968.
- (43) Crespo-Otero, R.; Montero, L. A.; Stohrer, W. D.; de la Vega, J. M. G. *J. Chem. Phys.* **2005**, *123*, 134107.
- (44) Cheng, J.; Kang, C.; Zhu, W.; Luo, X.; Puah, C. M.; Chen, K.; Shen, J.; Jiang, H. *J. Org. Chem.* **2003**, *68*, 7490.
- (45) Masamura, M. *Theor. Chem. Acc.* **2001**, *106*, 301.
- (46) Kar, T.; Ponec, R.; Sannigrahi, A. B. *J. Phys. Chem. A* **2001**, *105*, 7737.
- (47) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.
- (48) Rulíšek, L.; Havlas, Z. *J. Am. Chem. Soc.* **2000**, *122*, 10428
- (49) Rulíšek, L.; Havlas, Z. *J. Phys. Chem. A* **2002**, *106*, 3855.
- (50) Rulíšek, L.; Havlas, Z. *J. Phys. Chem. B* **2003**, *107*, 2376.
- (51) Sun, G.; Voigt, J. H.; Filippov, I. V.; Marquez, V. E.; Nicklaus, M. C. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1752.
- (52) *CRC Handbook of Chemistry and Physics*, 54th ed.; CRC Press: Cleveland, Ohio.

Chapter 8

Influences of Computational Methods and Chemical Models on Amide Bond Formation

8.1 Introduction

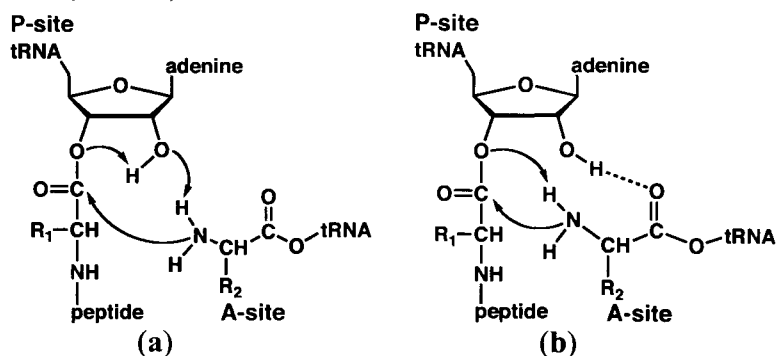
Recently, quantum mechanical methods have been widely used to investigate the mechanisms of enzymatic reactions.¹⁻¹⁴ A commonly used approach^{13,14} is to truncate an active site model from the crystal structure. Geometry optimizations on these models are typically performed using the density functional theory methods, in particular B3LYP, with a basis set of medium size (6-31G or larger with polarization functions). In order for the chemical model of the enzyme to be computationally tractable, the substrate and selected catalytic groups from the active site are included for simulation while the rest of the enzyme is modeled as a continuum.

With the current development of computer power, it is now routine to treat up to 100 atoms using the B3LYP method with a medium size basis set. However, a 100-atom model is still relatively small compared to the enzyme active site. If the number of atoms of the active site increases, normally one has to use a lower level method, or a smaller basis set, or both. As a result, the accuracy of the calculations may be sacrificed.

A reaction studied at such a risk is the amide bond formation as catalyzed by the ribosome, whose mechanism is still under debate. One of the experimentally proposed

mechanism involves a six-membered ring transition structure (TS_6), mediated by the P-site tRNA ribose sugar.¹⁵⁻¹⁹ In TS_6 , the P-site 2'-oxygen abstracts a proton from the A-site amine with the latter nucleophilically attacking the P-site carbonyl group. Simultaneously, the proton of the 2'-OH transfers to 3'-oxygen with the peptide grown by one amino acid (**Scheme 8.1a**). Several computational studies²⁰⁻²² also supported this mechanism. However, a different mechanism, which involves a four-membered ring transition structure (TS_4), was first proposed in a computational study.²³ In TS_4 , the A-site amine directly transfers a proton to the P-site 3'-oxygen (**Scheme 8.1b**). In addition, the hydrogen bond between the 2'-OH and A-site carbonyl group was proposed to perform a catalytic role. A recent computational study²⁴ further compared TS_6 and TS_4 . Based on their 444-atom model optimized at HF/3-21G level of theory, TS_6 was found to be 10.8 kJ mol^{-1} higher in energy than TS_4 , even if TS_6 should be more favored than TS_4 as the latter involves a quite constrained proton transfer process. Indeed, several previous computational studies²⁵⁻²⁷ on enzymatic reactions found that six-membered ring transition structures are preferred to four-membered counterparts. Considering the low level of theory, it is questionable whether HF/3-21G is able to give reasonable results for this reaction.

Scheme 8.1 Schematic illustration of the proposed mechanism for amide bond formation as catalyzed by the ribosome via (a) six-membered¹⁵⁻²² and (b) four-membered^{23,24} ring transition structure (see text).

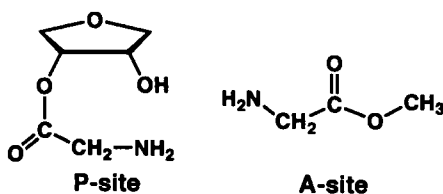


With this question in mind, in this chapter we have examined the influence of computational methods and models on amide bond formation as catalyzed by the ribosome. For simplicity, only the simplest amino acid, glycine, was used.

8.2 Computational Methods

All calculations were performed using the Gaussian 03 suite of programs.²⁸ For the small model, optimized structures were obtained using both the Hartree-Fock (HF) and density functional theory (DFT) method B3LYP²⁹⁻³¹ in conjunction with a variety of basis sets ranging from 3-21G to 6-311+G(2df,p). Frequency calculations were also performed at the corresponding level of theory in order to obtain zero-point vibrational energy (ZPVE) corrections. For the large model, geometry optimizations were performed at the HF/3-21G level of theory. In addition, large single point calculations were also performed at the B3LYP/6-31G(d) and B3LYP/6-31G(d,p) levels based on the HF/3-21G optimized geometries, i.e., B3LYP/6-31G(d)//HF/3-21G and B3LYP/6-31G(d,p)//HF/3-21G.

Scheme 8.2 Schematic illustration of the P-site and A-site used in the small model.



The small model of the P-site is a glycine condensed to a ribose sugar at its 3'-OH position while the A-site is a glycine methyl ester (**Scheme 8.2**). The big model was derived from a previous computational study, in which **TS₄** was found to be lower in energy than **TS₆**.²⁴ Specifically, the active site was taken from the crystal structure (PDB: 1Q86). The side chains of the P-site and A-site phenylalanines were replaced by hydrogens. The nucleobase residues around the active site include C2104, A2485, A2486,

C2487, U2541, U2542, G2618, U2619 and U2620. In addition, the ribose sugars of C2104 and A2486 were also included (see **Figure 8.1**). All the heavy atoms of these nucleobases and ribose sugars were held fixed during optimization. In addition, the 4'-oxygen of the P-site sugar was also fixed.

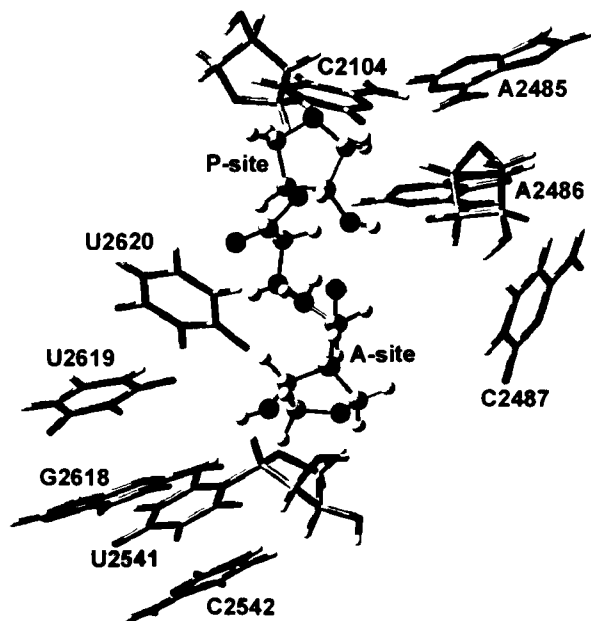


Figure 8.1 The large active model used in this chapter. Color key: C (gray); O (red); N (blue); and H (white).

8.3 Results and Discussion

8.3.1 Small Model Calculations

As noted in **Section 8.2**, a small model was first studied using a variety of methods. **Table 8.1** lists the key interaction distances and reaction barriers of amide bond formation via **TS₆** and **TS₄** using both HF and B3LYP methods in combination with a variety of basis sets ranging from 3-21G to 6-311+G(2df,p). Regarding the transition state geometries, compared with B3LYP, HF generally overestimates the breaking C—O atom bond (r_1 in **TS₆** and r_7 in **TS₄**) and forming O—H bonds (r_2 and r_4 in **TS₆** and r_8 in **TS₄**),

and underestimates the forming C—N bond (r_6 in TS_6 and r_{10} in TS_4) and breaking O—H bonds (r_3 and r_5 in TS_6 and r_9 in TS_4). In addition, for both HF and B3LYP methods, the distances obtained using the 3-21G basis set largely deviate from other basis sets. For example, in TS_6 the differences of r_1 – r_6 between HF/3-21G and HF/6-311+G(2df,p) range from -0.765 to 0.128 Å while such differences at the B3LYP method are between 0.278 and 0.143 Å.

As for the reaction barriers, the HF method in combination with both 3-21G and 6-31G basis sets give poor results. For example, E_6 and E_4 obtained using the 3-21G basis set are smaller than other basis sets by 50.2 — 68.9 kJ mol^{-1} and 40.2 — 80.9 kJ mol^{-1} respectively. With the slightly larger 6-31G basis set, E_6 and E_4 are 50.2 and 40.2 kJ mol^{-1} higher in energy than 3-21G but still 12.0 — 18.7 and 31.2 — 40.7 kJ mol^{-1} lower than those in which the polarization functions on heavy atoms were included. Furthermore, the 3-21G and 6-31G are the only two basis sets that give E_6 marginally lower or even higher than E_4 . In contrast, all other basis sets with polarization functions predict E_6 to be 10.3 — 17.3 kJ mol^{-1} lower in energy than E_4 .

The B3LYP method again shows similar trends. The barriers obtained using the 3-21G and 6-31G basis sets are much smaller than those obtained with other basis sets. In addition, 3-21G predicts ΔE ($E_4 - E_6$) to be 39.9 kJ mol^{-1} . However, all the other basis sets give smaller ΔE of 18.7 — 23.5 kJ mol^{-1} . Importantly, as the basis set is systematically increased to 6-311+G(2df,p), ΔE appears to converge to about 20 kJ mol^{-1} .

Table 8.1 Selected Interaction Distances (Å) and Reaction Barriers (kJ mol⁻¹) of Amide Bond Formation in the Small and Large Active-Site Model via Six- and Four-Membered Ring Transition Structures.

													E_6^a	E_4^b	ΔE^c
	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}					
HF/3-21G	2.018	1.250	1.178	1.432	1.120	1.567	1.684	1.269	1.226	1.569	191.4	3.9			
HF/6-31G	2.777	1.471	1.049	1.446	1.109	1.498	1.928	1.420	1.128	1.532	231.6	-6.1			
HF/6-31G(d)	2.646	1.327	1.101	1.505	1.088	1.501	2.171	1.727	1.053	1.514	266.4	13.6			
HF/6-31G(d,p)	2.634	1.382	1.061	1.471	1.097	1.499	2.062	1.594	1.070	1.518	264.3	14.6			
HF/6-311G(d,p)	2.716	1.388	1.054	1.459	1.099	1.500	2.135	1.742	1.045	1.521	262.8	12.6			
HF/6-31G(2d,p)	2.662	1.376	1.063	1.467	1.098	1.497	2.099	1.651	1.058	1.651	272.3	17.3			
HF/6-31G(df,p)	2.637	1.381	1.061	1.474	1.095	1.500	2.011	1.536	1.080	1.523	266.5	10.3			
HF/6-31+G(d,p)	2.758	1.408	1.051	1.462	1.100	1.497	2.212	1.837	1.043	1.511	265.0	15.3			
HF/6-311+G(2df,p)	2.783	1.403	1.050	1.467	1.097	1.496	2.177	1.800	1.040	1.515	271.0	14.6			
B3LYP/3-21G	1.959	1.220	1.264	1.338	1.209	1.613	1.797	1.328	1.272	1.597	108.4	39.9			
B3LYP/6-31G	2.098	1.267	1.215	1.451	1.134	1.596	1.898	1.367	1.206	1.588	144.0	18.7			
B3LYP/6-31G(d)	2.164	1.281	1.175	1.456	1.132	1.577	1.909	1.390	1.188	1.575	174.9	22.5			
B3LYP/6-31G(d,p)	2.131	1.303	1.146	1.430	1.139	1.578	1.867	1.382	1.182	1.584	171.7	23.5			
B3LYP/6-311G(d,p)	2.210	1.322	1.130	1.448	1.129	1.577	1.928	1.416	1.161	1.585	173.8	20.0			

B3LYP/6-31G(2d,p)	2.161	1.313	1.137	1.439	1.133	1.572	148.8	1.900	1.406	1.168	1.577	171.5	22.7
B3LYP/6-31G(df,p)	2.138	1.305	1.144	1.435	1.137	1.579	153.3	1.853	1.372	1.184	1.587	173.8	20.5
B3LYP/6-31+G(d,p)	2.193	1.328	1.132	1.461	1.126	1.574	156.1	1.892	1.396	1.170	1.581	176.4	20.3
B3LYP/6-311+G(2df,p)	2.237	1.340	1.121	1.471	1.120	1.570	166.9	1.928	1.419	1.154	1.583	185.9	19.0
MP2/6-311+G(2df,p)//HF/6-31G(d,p)	2.634	1.382	1.061	1.471	1.097	1.499	148.8	2.062	1.594	1.070	1.518	173.7	24.9
MP2/6-311+G(2df,p)//B3LYP/6-31G(d,p)	2.131	1.303	1.146	1.430	1.139	1.578	136.8	1.867	1.382	1.182	1.584	153.6	16.8
B3LYP/6-31G(d)//HF/3-21G	2.018	1.250	1.178	1.432	1.120	1.567	155.6	1.684	1.269	1.226	1.569	186.3	30.7
B3LYP/6-31G(d,p)//HF/3-21G	2.018	1.250	1.178	1.432	1.120	1.567	151.7	1.684	1.269	1.226	1.569	181.0	29.3
Large model													
HF/3-21G	2.299	1.639	1.009	1.466	1.107	1.515	219.3	1.933	1.359	1.137	1.560	221.1	1.8
B3LYP/6-31G(d)//HF/3-21G	2.299	1.639	1.009	1.466	1.107	1.515	161.7	1.933	1.359	1.137	1.560	172.9	11.2
B3LYP/6-31G(d,p)//HF/3-21G	2.299	1.639	1.009	1.466	1.107	1.515	161.6	1.933	1.359	1.137	1.560	169.7	8.1

^a E_4 : Reaction barrier via four-membered ring transition structure.

^b E_6 : Reaction barrier via six-membered ring transition structure.

^c $\Delta E = E_4 - E_6$

Furthermore, we also performed two single point calculations at the MP2/6-311+G(2df,p) level of theory based on the HF/6-31G(d,p) and B3LYP/6-31G(d,p) optimized geometries: i.e. MP2/6-311+G(2df,p)//HF/6-31G(d,p) and MP2/6-311+G(2df,p)//B3LYP/6-31G(d,p). The values of ΔE were found to be 24.9 and 16.8 kJ mol⁻¹ (see **Table 8.1**). Hence, in agreement with B3LYP, the MP2 method also predicts that **TS₆** is lower in energy than **TS₄**.

In summary, calculations from the small model agree with most chemical evidence that **TS₆** is more favored than **TS₄** when either HF or B3LYP methods with a medium size basis set (6-31G(d) or larger) are employed. Importantly, the previously used²⁴ HF/3-21G level appears not to be able to give accurate results based on the small model. We also performed a single point calculation at the B3LYP/6-31G(d,p) level based on the HF/3-21G optimized geometry, i.e., B3LYP/6-31G(d,p)//HF/3-21G. The values of E_4 and E_6 are 181.0 and 151.7 kJ mol⁻¹ respectively. The latter is significantly lowered compared with that at the HF/3-21 level. As a result, ΔE also increases to 29.3 kJ mol⁻¹ (see **Table 8.1**), which is in good agreement with that at the B3LYP/6-31G(d,p) level. Thus, a single point calculation using a higher level of theory may be needed if the HF/3-21G level has to be used.

8.3.2 Catalytic Hydrogen Bonding Interactions

As noted in **Section 8.1**, the hydrogen bond between the P-site 2'-OH and A-site carbonyl groups (2'-OH...O=C see **Scheme 8.1**) was proposed to be catalytic. In order to obtain further insights into the possible role of this hydrogen bond, we considered three mutations: (i) the P-site 2'-OH group is changed to a hydrogen atom or (ii) a fluorine atom, and (iii) the A-site CH₃COO- group is changed to CH₃-. In all these cases, the original 2'-OH...O=C hydrogen bond no longer exists. The barriers via **TS₄** at B3LYP/6-31G(d,p) level of theory were found to be 163.9, 168.3 and 172.1 kJ mol⁻¹ respectively.

Compared to that of the unmutated model ($171.3 \text{ kJ mol}^{-1}$; see **Table 8.1**), conversion of the 2'-OH to 2'-H even decreases the barrier by 7.4 kJ mol^{-1} while the other two mutations have only marginal effects on the barriers of only -3.0 and 0.9 kJ mol^{-1} . Thus, we concluded that the 2'-OH \cdots O=C hydrogen bond may not be catalytic as originally proposed.^{23,24}

8.3.3 Large Active-Site Model Calculations

The optimized structures of **TS₄** and **TS₆** in the large active-site model are shown in **Figure 8.2** and the key interaction distances and reaction barriers are shown in **Table 8.1**.

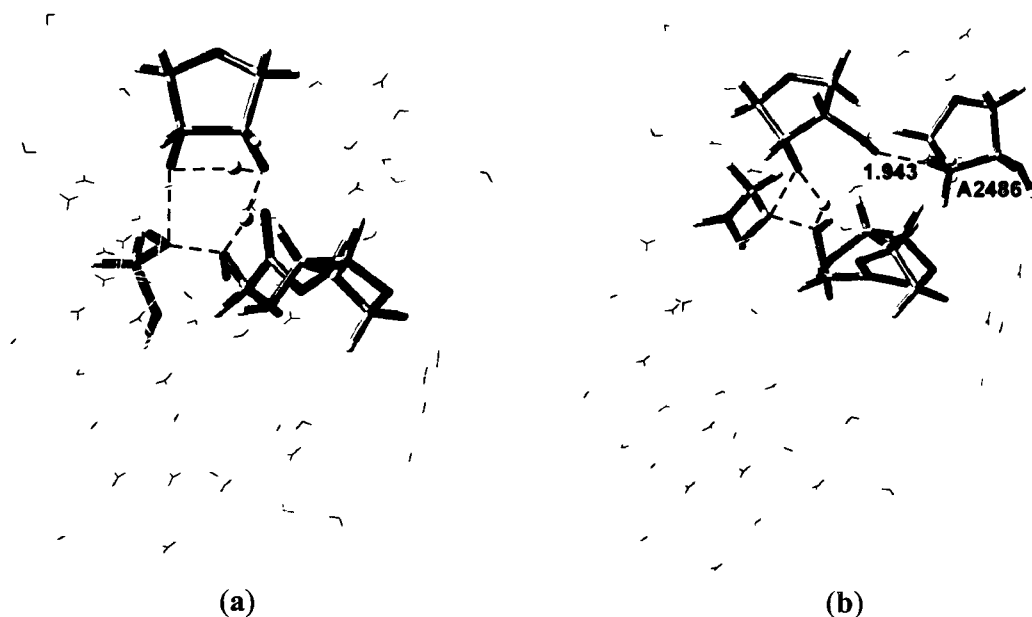


Figure 8.2 The optimized structures of (a) **TS₄** and (b) **TS₆** in the large active-site model at the HF/3-21G level of theory. Selected distances are shown in angstroms.

In the large active-site model, the reaction was found to follow the same mechanisms via either **TS₄** or **TS₆** as in the small model. However, a notable difference is that the P-site 2'-OH group no longer forms a hydrogen bond with the A-site carbonyl oxygen in **TS₄**. Instead, it forms a hydrogen bond with the 2'-OH of A2486 with a

distance of 1.943 Å (see **Figure 8.2**). Thus, in contrast to the previous computational study,²⁴ our large model calculation does not support the existence of the ‘catalytic hydrogen bonding’ if the reaction occurs via **TS₄**.

As for the reaction barriers, at the HF/3-21G level of theory, **TS₄** lies only marginally higher in energy than **TS₆** by 1.8 kJ mol⁻¹ (see **Table 8.1**). This is similar to the results from the small model. However, it is in disagreement with a previous computational study²⁴ at the same level of theory, in which **TS₄** was found to be 10.8 kJ mol⁻¹ lower in energy than **TS₆**. To examine the effects of increasing the level of theory on the barriers, we also performed two single point calculations at the B3LYP/6-31G(d)//HF/3-21G and B3LYP/6-31G(d,p)//HF/3-21G levels and the above energy difference increases to 11.2 and 8.1 kJ mol⁻¹, respectively. Again, this is in agreement with **TS₆** being lower in energy than **TS₄**. Importantly, the conclusion from the previous computational study²⁴ that **TS₄** is more favored than **TS₆** may not be correct most likely due to the fact that the energies were computed using the non-correlated HF method in conjunction with a very small basis set.

8.4 Conclusions

In this chapter, we have examined the influence of computational methods and models on the amide bond formation as catalyzed by the ribosome. In particular, two possible mechanisms via either a four- or six-membered transition structure were considered. First, the HF and B3LYP methods with a variety of basis sets were used to obtain the barriers via **TS₆** and **TS₄**. Except for HF/6-31G, all the other levels of theory predict that **TS₆** is lower in energy than **TS₄**. Importantly, as the basis set is systematically increased, the energy differences between **TS₆** and **TS₄** appear to converge to 20 kJ mol⁻¹ in favor of **TS₆**. In addition, we also made several modifications to our model so that the

previously proposed catalytic hydrogen bonding interaction between the P-site 2'-OH and A-site carbonyl groups no longer exists. However, the barriers via TS_4 do not change dramatically compared to the unmutated model, which suggests that such a hydrogen bond does not have a catalytic role. Furthermore, the large active model calculations also confirmed the above conclusions: (i) TS_6 lies lower in energy than TS_4 , and (ii) the hydrogen bond between the P-site 2'-OH and A-site carbonyl groups is not catalytic.

The study present in this chapter suggests that incorrect results may be obtained when a large model is studied with a low level of theory. Hence, caution must be paid. We suggest that at least large basis set single point calculations should be performed to obtain more accurate energies.

References

- (1) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.
- (2) Himo, F.; Siegbahn, P. E. M. *Chem. Rev.* **2003**, *103*, 2421.
- (3) Velichkova, P.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 16.
- (4) Hopmann, K. H.; Himo, F. *J. Phys. Chem. B* **2006**, *110*, 21299.
- (5) Hopmann, K. H.; Guo, J. D.; Himo, F. *Inorg. Chem.* **2007**, *46*, 4850.
- (6) Chen, S. L.; Marino, T.; Fang, W. H.; Russo, N.; Himo, F. *J. Phys. Chem. B* **2008**, *112*, 2494.
- (7) Siegbahn, P. E. M.; Borowski, T. *Acc. Chem. Res.* **2006**, *39*, 729.
- (8) Liu, H.; Robinet, J. J.; Ananvoranich, S.; Gaudl, J. W. *J. Phys. Chem. B* **2007**, *111*, 439.
- (9) Cho, K. B.; Gaudl, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 10267.
- (10) Cho, K. B.; Gaudl, J. W. *J. Phys. Chem. B* **2005**, *109*, 23706.
- (11) Robinet, J. J.; Cho, K. B.; Gaudl, J. W. *J. Am. Chem. Soc.* **2008**, *130*, 3328.

- (12) Robinet, J. J.; Gauld, J. W. *J. Phys. Chem. B* **2008**, *112*, 3462.
- (13) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232.
- (14) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515.
- (15) Dorner, S.; Polacek, N.; Schulmeister, U.; Panuschka, C.; Berta, A. *Biochem. Soc. Trans.* **2002**, *30*, 1131.
- (16) Dorner, S.; Panuschka, C.; Schmid, W.; Barta, A. *Nucleic Acids Res.* **2003**, *31*, 6536.
- (17) Chamberlin, S. I.; Merino, E. J.; Weeks, K. M. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14688.
- (18) Beringer, M.; Rodnina, M. V. *Mol. Cell* **2007**, *26*, 311.
- (19) Schmeing, T. M.; Huang, K. S.; Kitchen, D. E.; Strobel, S. A.; Steitz, T. A. *Mol. Cell* **2005**, *20*, 437.
- (20) Das, G. K.; Bhattacharyya, D.; Burma, D. P. *J. Theor. Biol.* **1999**, *200*, 193.
- (21) Sharma, P. K.; Xiang, Y.; Kato, M.; Warshel, A. *Biochemistry* **2005**, *44*, 11307.
- (22) Trobro, S.; Åqvist, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12395.
- (23) Gindulyte, A.; Bashan, A.; Agmon, I.; Massa, L.; Yonath, A.; Karle, J. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13327.
- (24) Thirumoorthy, K.; Nandi, N. *J. Phys. Chem. B* **2008**, *112*, 9187.
- (25) Gorb, L.; Asensio, A.; Tuñón, I.; Ruiz-López, M. F. *Chem. Eur. J.* **2005**, *11*, 6743.
- (26) Qiao, Q.-A.; Yang, C.; Qu, R.; Jin, Y.; Wang, M.; Zhang, Z.; Xu, Q.; Yu, Z. *Biophys. Chem.* **2006**, *122*, 215.
- (27) Rangelov, M. A.; Vayssilov, G. N.; Yomtova, V. M.; Petkov, D. D. *J. Am. Chem. Soc.* **2006**, *128*, 4964.
- (28) Gaussian 03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.;

Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B., Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Wallingford CT, 2004.

(29) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(30) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(31) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

Chapter 9

Conclusions

This thesis details our computational studies on various aspects of nucleic acid-related chemistry. This chapter summarizes the main conclusions and highlights the significance and possible extensions of the current work.

In Chapter 3, the possibility of using a phosphate oxygen as the general base in the mechanism of the histidyl-tRNA synthetase is studied. Three possible mechanisms with the bridging phosphate oxygen, non-bridging pro-*R* and -*S* phosphate oxygens acting as the base to deprotonate the 3'-OH of A76 were considered. It was found that the preferred reaction pathway is an associative stepwise mechanism with the pro-*S* phosphate oxygen acting as the base. This is due to a combination of three reasons: the more linear proton transfer process, the greater basicity of the pro-*S* oxygen and stabilization effects by the active site residues.

It should be noted that at this stage, the exact position of the A76 substrate is still unknown except for some suggestions from experimental studies.¹⁻⁴ Thus, we manually placed A76 in our active site model according to these suggestions. If the position of A76 were known from the crystal structure of the tRNA^{His}-HisRS-histidyl-adenylate complex or its analogue in the future, it may be necessary to reexamine our calculated substrate-bound complex. Although the conclusions obtained from our study are unlikely to change, it is probable that a better understanding of the binding of A76 in the active site will be able to provide new insights into the mechanism of this enzyme.

Another possible extension of the current work is to consider the effects of chirality of the histidyl-adenylate substrate on the reaction. It is well known that all the amino acids synthesized in cells are in the *S* configuration.⁵ However, the reason for such high stereospecificity is still unclear. Previous computational studies^{6,7} have found that the amino acids with *S* configurations are more favorable than *R* configurations in the ribosome. However, it appears to be more efficient for nature to determine such chirality in the initial step of protein synthesis. That is, only *S* aminoacyl-tRNA may be formed in the tRNA-synthetases and then delivered to the ribosome. According to the current experimental suggestions on where A76 is bound,¹⁻⁴ the side group of histidine should be pointing outwards from A76. Thus, this may potentially favor the *S* amino acids due to less steric hindrance. It is apparent that future studies are needed on the roles of chirality of aminoacyl-adenylate on the substrate binding and the reaction barrier.

In Chapters 4 — 6, three aspects of nucleobase chemistry are studied: (i) the repair of a methylated nucleobase, (ii) the use of a nucleobase for catalysis, and (iii) the protonation of a nucleobase quartet and a quartet stack.

In Chapter 4, the catalytic mechanism of the alkylated nucleobase repair enzyme AlkB was studied. This enzyme belongs to the non-heme iron enzyme superfamily and utilizes a unique oxidative dealkylation mechanism to repair the damaged nucleobase. From our calculations, the reaction was found to consist of four stages: (i) the binding of O₂ to the iron center in the active site, (ii) the activation of O₂ to form the highly reactive ferryl-oxo or ferric-oxyl intermediate, (iii) the reorientation of ferryl-oxo group from the axial to equatorial position and (iv) the repair of 1-methyl adenine by the ferryl-oxo intermediate. In the first stage, the lowest energy O₂-bound active site complex was found to be in the spin-septet state, with an electron transferred from Fe(II) to O₂. In the subsequent O₂ activation stage, the spin-quintet and -septet states have similar barriers in

the initial steps. However, the lowest energy ferryl-oxo intermediate is in the spin-quintet state. It should also be noted that the mechanism of this stage on the spin-septet surface is different from the spin-quintet surface due to a distinct electron transfer process. In the third stage, the iron-oxygen group reorients from the axial to equatorial position with the spin-quintet state preferred to the spin-triplet state. However, the spin-septet state loses its catalytic activity by deprotonation of a nearby arginine. In the last stage, the ferryl-oxo moiety abstracts a hydrogen atom from the substrate, followed by a radical rebound process to give an alcohol, which then decomposes to an aldehyde and native adenine. The spin-quintet state was again found to be preferred to the spin-triplet state.

In Chapter 5, the catalytic mechanism of the hepatitis delta virus (HDV) ribozyme is studied. This ribozyme catalyzes the self-cleavage of its own phosphodiester bond. However, it was unclear whether the important nucleobase (C75) acts as a general base or acid. We started our investigations by considering the properties of individual components within the active site, in particular the effects of general environment and explicit hydrogen bonds on the proton affinity of C75 as well as the basicity of the hydrated Mg^{2+} ion in a variety of coordination environments. We found that both the general environment and hydrogen-bonding interactions are able to increase the proton affinity of C75. However, neither of them, individually or combined, makes C75 basic enough to directly deprotonate the 2'-OH of the leaving ribose sugar. Regarding the hydrated Mg^{2+} ion, we have examined the outer- and inner-sphere coordination via a phosphate oxygen of U23. The latter, i.e., $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$, is found to have a larger proton affinity than the former, i.e., $[\text{Mg}(\text{H}_2\text{O})_5\text{OH}]^+$. However, they are still much lower than that of the 2'-O⁻ ribose anion. Based on these results, an appropriate chemical model was built to investigate the catalytic mechanism of the HDV ribozyme. We found that both C75 and $[\text{Mg}(\text{H}_2\text{O})_4(\text{CH}_3\text{PO}_4\text{CH}_3)\text{OH}]$ are involved in the first step. Specifically, the latter acts as a base to deprotonate the 2'-OH of the ribose sugar while

the former donates a proton to a non-bridging phosphate oxygen to form a phosphorane intermediate and is therefore converted to its tautomeric form. In the last step, the cytosine tautomer acts an acid to protonate the leaving 5'-oxygen and simultaneously abstracts a proton from the phosphorane P—OH moiety.

The catalytic mechanisms of ribozymes have attracted considerable interest both experimentally and theoretically. In addition to the HDV ribozyme, several other ribozymes, such as the hammerhead,⁸ hairpin^{9,10} and varkud satellite (VS)^{11,12} ribozymes have also been proposed to use nucleobases as a general base or acid in several experimental studies. In addition, a number of computational studies on the catalytic mechanisms of ribozymes have also been reported.¹³⁻¹⁷ However, the mechanisms of some ribozymes are still under debate. It is clear that further investigations are needed in this area.

In Chapter 6, the protonation of guanine quartets (**G4**) and quartet stacks (**2G4**) was studied using DFT methods. For **G4**, all multi-protonated complexes favor protonation at the Hoogsteen faces, except the singly protonated complex that favors protonation at the Watson-Crick face of the O6 moieties. In addition, the first proton affinity of **G4** is higher than that of the guanine monomer at both the Watson-Crick and Hoogsteen faces due to the charge delocalization to the adjacent neutral guanine and the presence of a nearby electron rich unprotonated O6. Furthermore, both the first and second proton affinities of **G4** are higher than those of water at the same level of theory. For **2G4**, the most stable complexes in all the protonation states correspond to protonation at the Watson-Crick faces. In addition, each plane is able to accept a maximum of two protons. Furthermore, protons can also sit in the middle of the two planes by forming inter-plane hydrogen bonds.

The findings in this chapter provide initial insights into the binding of the simplest cation H^+ with guanine quartets and quartet stacks. There are several possible avenues to extend this study. First, our studies suggest that the guanine quadruplexes ($G4_n$'s) may act as a base in porphyrin metallation. However, the catalytic mechanism for this reaction still remains unclear. Future studies may focus on the mechanism of $G4_n$'s catalyzed porphyrin metallation. Due to the potential large size of the model, QM/MM methods may be used. In addition, if $G4_n$'s are protonated, they may be able to bind anions. However, most of the currently known $G4_n$'s ligands are cationic. We note that it has been reported that a synthetic Pt-purine quartet is able to bind with SO_4^{2-} anion due to the inversion of the $G4_n$'s charges by Pt^{2+} .¹⁸ Thus, the ability of protonated $G4_n$'s to act as anion binders awaits for further studies both experimentally and theoretically. Furthermore, the protonation of other nucleobase quartets and quadruplexes may also be examined in future studies.

In Chapter 7, the binding of metal ions (Na^+ , K^+ , Mg^{2+} and Cd^{2+}) with ribose sugar and locked ribose sugar is investigated. An assessment was first performed in order to find an appropriate basis set to use with B3LYP to obtain reliable geometries and accurate binding energies. A suitable basis set was found to be 6-311+G(2df,p), and hence used for all the complexes, except Cd^{2+} in which an ECP was used for core electrons. For the complexes of metal ions with ribose, the metal ions are able to bind with the ribose sugar monodentately, bidentately and tridentately. In addition, the Na^+ -ribose complexes were found to have larger binding energies than the K^+ -ribose complexes, due to a smaller radius of Na^+ . Furthermore, the binding energies of two divalent cations Mg^{2+} and Cd^{2+} were found to be much larger than Na^+ and K^+ , due to their larger charges. For complexes of metal ions with locked ribose, the modifications of the structure of the ribose sugar lead to different properties upon binding with metal ions. For example, replacing the 2'-oxygen to nitrogen increases the binding energies for all the metal ions. In addition,

replacing the 4'-oxygen to a soft Lewis base such as selenium is also able to increase the binding energy with some soft metal ion, such as Cd^{2+}

Modified ribose sugars have recently attracted considerable interest. Currently, several possible modifications have been reported in the literature. For example, the 2' and 4', or 1' and 2' positions of the ribose can be locked by a methylene linkage.¹⁹ In addition, the 4'-oxygen can be mutated into other heteroatoms, such as sulfur and selenium.²⁰ These modified nucleic acids show increased thermostabilities when hybridizing with complementary DNA or RNA and may act as new therapeutic tools.¹⁹ However, the properties of the locked ribose interacting with metal ions remain unknown. Our study clearly shows that such locked ribose is able to change the binding abilities with metal ions. Thus, the study performed in this chapter may help discover new metal-ribose complexes and to better understand the roles of metal ions in biological systems.

In Chapter 8, the influences of computational methods and models on the amide bond formation as catalyzed by the ribosome are examined. The motivation for this work comes from a recent computational study⁷ in which a more constrained four-membered ring transition structure (TS_4) was found to be more favored compared to a less constrained six-membered ring transition structure (TS_6) at the HF/3-21G level of theory. We have used a small P- and A-site model as well as a large active site model, and the reaction barriers via either TS_4 or TS_6 were obtained using the HF and B3LYP methods in conjunction with a variety of basis sets. As for the small model, except for HF/6-31G, all other levels of theory predict that TS_6 lies lower in energy than TS_4 . In addition, as the basis set is gradually increased, the energy difference between TS_4 and TS_6 appears to converge to 20 kJ mol^{-1} . However, the small basis set such as 3-21G is unable to provide accurate results for this reaction. For example, TS_6 is only marginal higher in energy than TS_4 by 3.9 kJ mol^{-1} at the HF/3-21G level of theory. This is again observed in the large

active model calculation that shows the energetic difference between TS_4 and TS_6 is only 1.8 kJ mol^{-1} . However, single point calculations at the B3LYP/6-31G(d) and B3LYP/6-31G(d,p) levels based on the optimized geometries at HF/3-21G, i.e., B3LYP/6-31G(d)//HF/3-21G and B3LYP/6-31G(d,p)//HF/3-21G, increase such energy difference to 11.2 and 8.1 kJ mol^{-1} , respectively. Hence, this clearly shows that the previously used⁷ HF/3-21G level may not be appropriate for this reaction.

The materials presented in this chapter also have important implications for future quantum modeling of enzymatic reactions. With the current computer power, it is now routine to treat up to 100 atoms using DFT methods with a medium size basis set. However, as the size of the active site model grows larger, one has to use a lower level of theory, such as HF/3-21G. In this case, caution must be paid when a low level of theory is used, as it may not be accurate enough to describe the chemistry of interest.

In summary, this thesis provides examples of studying the nucleic acid-related chemistry by quantum mechanical methods. Currently, computational methods are widely used to study biomolecules and have shown to be able to provide complementary insights to experimental studies. Further applications of computational methods are deemed to make invaluable contributions in this area.

References

- (1) Guth, E.; Connolly, S. H.; Bovee, M.; Francklyn, C. S. *Biochemistry* **2005**, *44*, 3785.
- (2) Hawko, S. A.; Francklyn, C. S. *Biochemistry* **2001**, *40*, 1930.
- (3) Connolly, S. A.; Rosen, A. E.; Musier-Forsyth, K.; Francklyn, C. S. *Biochemistry* **2004**, *43*, 962.

- (4) Guth, E. C.; Francklyn, C. S. *Mol. Cell* **2007**, *25*, 531.
- (5) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 4th ed.; W. H. Freeman and Company: New York, 2004.
- (6) Thirumoorthy, K.; Nandi, N. *J. Phys. Chem. B* **2007**, *111*, 9999.
- (7) Thirumoorthy, K.; Nandi, N. *J. Phys. Chem. B* **2008**, *112*, 9187.
- (8) Han, J.; Burke, J. M. *Biochemistry* **2005**, *44*, 7864.
- (9) Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2259.
- (10) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R. *Science* **2002**, *298*, 1421.
- (11) Zhao, Z.-Y.; McLeod, A.; Harusawa, S.; Araki, L.; Yamaguchi, M.; Kurihara, T.; Lilley, D. M. J. *J. Am. Chem. Soc.* **2005**, *127*, 5026.
- (12) Wilson, T. J.; McLeod, A. C.; Lilley, D. M. J. *EMBO J.* **2007**, *26*, 2489.
- (13) Lee, T.-S.; Lopez, C. S.; Giambasu, G. M.; Martick, M.; Scott, W. G.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 3053.
- (14) Lee, T.-S.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 7168.
- (15) Nam, K.; Gao, J.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 4680.
- (16) Wei, K.; Liu, L.; Cheng, Y. H.; Fu, Y.; Guo, Q. X. *J. Phys. Chem. B* **2007**, *111*, 1514.
- (17) Banáš, P.; Rulíšek, L.; Hánošová, V.; Svozil, D.; Walter, N. G.; Šponer, J.; Otyepka, M. *J. Phys. Chem. B* **2008**, *112*, 11177.
- (18) Roitzsch, M.; Lippert, B. *Angew. Chem. Int. Ed.* **2006**, *45*, 147.
- (19) Kaur, H.; Babu, B. R.; Maiti, S. *Chem. Rev.* **2007**, *107*, 4672.
- (20) Jeong, L. S.; Tosh, D. K.; Kim, H. O.; Wang, T.; Hou, X.; Yun, H. S.; Kwon, Y.; Lee, S. K.; Choi, J.; Zhao, L. X. *Org. Lett.* **2008**, *10*, 209.

Vita Auctoris

Name: Haining Liu
Place of Birth: Wuhan, China
Year of Birth: 1981

Education

2004 – 2009 Ph.D. in Chemistry
University of Windsor
1999 – 2003 B.Sc. in Chemistry
Wuhan University, China

Publications

1. Liu, H.; Robinet J. J.; Ananvoranich, S.; Gauld, J. W. *J. Phys. Chem. B* **2007**, *111*, 439.
2. Liu, H.; Gauld, J. W. *Phys. Chem. Chem. Phys.* **2009**, *11*, 278.
3. Liu, H.; Gauld, J. W. *J. Phys. Chem. B* **2008**, *112*, 16874.
4. Liu, H.; Llano, J.; Gauld, J. W. *J. Phys. Chem. B* **2008**, submitted.

Conference Presentations

1. Poster presentation
WATOC 2008, Sydney, Australia, 2008.
2. Oral presentation
91st Canadian Chemistry Conference and Exhibition, Edmonton, AB, 2008.

3. Oral presentation

American Chemical Society 234th National Meeting and Exposition, Boston, MA
USA, 2007.

4. Poster presentation

89th Canadian Chemistry Conference and Exhibition, Halifax, NS, 2006.

5. Poster presentation

88th Canadian Chemistry Conference and Exhibition, Saskatoon, SK, 2005.