Computational Insights into the High-Fidelily Catalysis of Aminoacyl-tRNA Synthetases

Mohamed M. Aboelnga

University of Windsor

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

Recommended Citation

https://scholar.uwindsor.ca/etd/6594

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

By

Mohamed M. Aboelnga

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

© 2017 Mohamed Aboelnga
Computational Insights into The High-Fidelity Catalysis of Aminoacyl-tRNA Synthetases

By

Mohamed Aboelnga

APPROVED BY:

_____________________________
P. Fernandes, External Examiner
University of Porto

_____________________________
S. Goodwin
School of Computer Science

_____________________________
S. Ananvoranich
Department of Chemistry & Biochemistry

_____________________________
J. Rawson
Department of Chemistry & Biochemistry

_____________________________
J. Gauld, Advisor
Department of Chemistry & Biochemistry

June 9, 2017
Declaration of Co-Authorship/Previous Publication

I. Co-Authorship

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

**Chapter 4:** The primary contribution and data analysis were performed by myself and the written manuscript was edited in collaboration with Dr. John J. Hayward under the supervision of Prof. James W. Gauld.

**Chapter 5:** The primary contribution and data analysis were performed by myself and the written manuscript was edited in collaboration with Dr. John J. Hayward under the supervision of Prof. James W. Gauld.

**Chapter 7:** The primary contribution and data analysis were performed in collaboration with Sarah Henshaw and Zaid Kaloti under the supervision of Prof. James W. Gauld.

**Chapter 9:** The primary contribution and data analysis were performed in collaboration with Jacqueline Gemus under the supervision of Prof. James W. Gauld.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above material in my thesis.

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

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

<table>
<thead>
<tr>
<th>Thesis chapter</th>
<th>Citation</th>
<th>Publication status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 3</td>
<td><em>J. Phys. Chem. B</em> 2017</td>
<td>Accepted for publication</td>
</tr>
<tr>
<td>Chapter 4</td>
<td><em>Phys. Chem. Chem. Phys.</em> 2017</td>
<td>Submitted</td>
</tr>
<tr>
<td>Chapter 5</td>
<td><em>ACS Catalysis</em> 2017</td>
<td>Accepted for publication</td>
</tr>
</tbody>
</table>

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

III. General

I declare that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

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

Obtaining insights into the catalytic function of enzymes is an important area of research due to their widespread applications in the biotechnology and pharmaceutical industries. Among these enzymes, the aminoacyl-tRNA synthetases (aaRSs) are known for their remarkable fidelity in catalyzing the aminoacylation reactions of tRNA in protein biosynthesis. Despite the exceptional execution of this critical function, mechanistic details of the reactions catalyzed by aminoacyl-tRNA synthetases remain elusive demonstrating the obvious need to explore their remarkable chemistry. During the PhD studies reported in this thesis the mechanism of aminoacylation, pre-transfer editing and post-transfer editing catalyzed by different aaRS have been established using multi-scale computational enzymology.

In the first two chapters a detailed information about aaRS and the addressed questions was given in addition to an overview of the used computational methodology currently used to investigate the enzymatic mechanisms. The aminoacylation mechanism of threonine by Threonyl-tRNA synthetases, glutamine by Glutaminyl-tRNA synthetases and glutamate by Glutamyl-tRNA synthetases have been clearly unveiled in chapter 3 and 4. Also, valuable information regarding the role of cofactors and active site residues has been obtained. While investigating the post-transfer editing mechanisms, which proceed in a remote and distinct active site, two different scenarios were experimentally suggested for two types of threonyl-tRNA synthetase species to correct the misacylation of the structurally related serine. We explored these two mechanisms as in chapters 5 and 6. Moreover, the synthetic site in which the aminoacylation reaction is catalyzed, is also responsible for a second type of proofreading reaction called pre-transfer editing mechanism. In chapter 7, this latter mechanism has been
elucidated for both Seryl-tRNA synthetases and Isoleucyl-tRNA synthetases against their non-cognate substrates cysteine and valine, respectively. In chapter 8, an assessment QM/MM study using a variety of DFT functionals to represent the chemically active layer in aminoacylation mechanism of the unnatural amino acid β-Hydroxynorvaline as catalyzed by Threonyl-tRNA synthetase has been carried out. Overall, it was found that substrate-assisted mechanisms are a common pathway for these enzymes.

One important application of such information is to establish the criteria required for any candidate to inhibit the catalytic functions of aaRS, which was applied in chapter 9 to screen potential competitive inhibitors able to efficiently block the bacterial Threonyl-tRNA synthetases.

The investigations reported herein should provide atomistic details into the fundamental catalytic mechanisms of the ubiquitous and ancient aaRS enzymes. Consequently, they will also help enable a much-needed deeper understanding of the underlying chemical principles of catalysis in general.
Dedication

I dedicate this work to my family.
Acknowledgment

I would like to express my special thanks of gratitude to my advisor Prof. James W. Gauld who gifted me a great opportunity to get my PhD under his supervision, as well as for his constant guidance and help over the course of my PhD study.

Also, I would like to thank my PhD committee, Prof. Scott Goodwin, Prof. Sirinart Ananvoranich and Prof. Jeremy Rawson for taking their valuable time in reading my thesis, and for the valuable discussions we had during the regular meetings. Also, special thanks to Prof. Pedro Fernandes for accepting to be my external examiner.

I would also acknowledge the great help provided by the talented Gauld unit whom I have had the honor to work with; the previous members: Dr. Eric Bushnell, Dr. Hisham Dokainish, Dr. Bogdan Ion, Rami Gherib, Grant Fortowsky, Daniel Simard, Philip Stamov, and Wanlei Wei in addition to the wonderful current members: Travis DeWolfe, Paul Meister, Sahar Nikoo and Taqred Alnakhli. All of them have contributed to provide a beneficial and knowledgeable yet entertaining environment that helped me be productive in the short time I spent in my PhD degree. Also, special thanks to Dr. John Hayward who edited my work and Jacqueline Gemus, Sarah Henshaw and Zaid Kaloti for their involvement in my research.

Furthermore, my sincere appreciation for the endless motivation of my dear parents which made me so productive. Importantly, I am also deeply grateful to all the fabulous efforts exerted by my lovely wife Eman Moussa; her supportive accompany and continuous encouragement are always the principle reasons for my success.

Lastly, I thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Egyptian Cultural Affairs & Mission Sector and Damietta University for financial support and Compute Calcul Canada and SHARCNET for computational resources.
Table of Contents

Declaration of Co-Authorship ----------------------------------------------- III

Abstract ---------------------------------------------------------------- V

Dedication---------------------------------------------------------------- VII

Acknowledgment------------------------------------------------------------- VIII

List of Tables------------------------------------------------------------- XV

List of Figures------------------------------------------------------------ XVI

List of Schemes------------------------------------------------------------ XXIII

List of Abbreviation------------------------------------------------------ XXVI

Chapter 1.

Introduction

1.1 Introduction------------------------------------------------------------ 2

1.2 References------------------------------------------------------------- 8

Chapter 2.

Computational Enzymology Methods

2.1 Molecular Modeling of Enzymes ---------------------------------------- 13

2.2 Molecular Dynamics Simulations --------------------------------------- 14
  2.2.1. Molecular Mechanics: --------------------------------------------- 17

2.3 Quantum Mechanics and Quantum Mechanics/Molecular Mechanics.-------- 17
  2.3.1 QM-only Approach ------------------------------------------------ 18
  2.3.2 QM/MM Modeling --------------------------------------------------- 20
  2.3.2.1 Additive and Subtractive QM/MM schemes: ------------------------ 21
  3.3.2.2 Technical aspects to be considered when constructing a QM/MM model.------- 25

2.4 Density Functional Theory (DFT)-------------------------------------- 27

2.5 References------------------------------------------------------------- 32
Chapter 3.

The Roles of Zn(II) And Active Site Residues in Substrate Discrimination by Threonyl-tRNA Synthetase.

3.1. Introduction: ---------------------------------------------------------------38

3.2. Computational Methods: -----------------------------------------------------41
  3.2.1 Molecular Dynamic (MD) Simulations. ---------------------------------------41
  3.2.2 QM/MM Investigations: ----------------------------------------------------43

3.3. Results and Discussion: ------------------------------------------------------44
  3.3.1 ThrRS⋯L-Thr/Ser-AMP/tRNA\textsuperscript{Thr} Michaelis Complexes:----------------44
  3.3.2 ThrRS⋯L-Thr/Ser-AMP/tRNA\textsuperscript{Thr} Reactive Complexes. -------------48
  3.3.3 Aminoacylation (Threonylation) Mechanism ----------------------------------50
  3.3.4 Aminoacylation (Serinylation) Mechanism. ----------------------------------54
  3.3.5 Discrimination against Val. -----------------------------------------------57
  3.3.6 Chiral Discrimination against D-Threonine: ----------------------------------60

3.4. Conclusion: -----------------------------------------------------------------64

3.5. References: -----------------------------------------------------------------66

Chapter 4.

A Water-Mediated and Substrate-Assisted Aminoacylation Mechanism in Aminoacyl-tRNA synthetases.

4.1 Introduction: -----------------------------------------------------------------72

4.2 Computational Methods: -------------------------------------------------------75
  4.2.1 Molecular Dynamic Simulation: ----------------------------------------------75
  4.2.2 QM-Cluster and QM/MM calculation: -----------------------------------------77

4.3 Results and Discussion: --------------------------------------------------------79
  4.3.1 MD simulation of wildtype/mutant GlnRS and ND-GluRS------------------------79
  4.3.2 QM-Only and QM/MM calculation ---------------------------------------------82
  4.3.2.1 Wildtype GlnRS Catalysed Aminoacylation with Glutamine: ------------------82
    4.3.2.1.1 QM-Cluster: the substrate's phosphate acting as the base. -------------82
    4.3.2.1.2 QM/MM: the substrate's phosphate acting as the base. ----------------83
4.3.2.1.3 QM/MM: the substrate's \(\alpha\)-NH\(_2\) acting as the base:------------------------87

4.3.2.2 Aminoacylation by the Glu34Gln mutant of GlnRS.------------------------89

4.3.2.3 ND-GluRS Catalyzed Misacylation of tRNA\(^{Gln}\) by Glutamate: ---------------91

4.4 Conclusion:---------------------------------------------------------------96

4.5 Referencies:---------------------------------------------------------------97

Chapter 5.
The Enzymatic Post-Transfer Editing Mechanism of \textit{E. coli} Threonyl-tRNA Synthetase.

5.1 Introduction:-------------------------------------------------------------103

5.2 Computational Methods:---------------------------------------------------107
  5.2.1 Molecular Dynamics Simulations.-----------------------------------------107
  5.2.2 QM/MM Investigations.-------------------------------------------------109

5.3 Results and discussion:---------------------------------------------------111
  5.3.1 His73-promoted Editing Mechanisms---------------------------------------111
    5.3.1.1 His73-promoted hydrolytic editing-----------------------------------111
      5.3.1.1.1 Direct water mediated protonation of \(\text{Ado76}^3\text{O}\) by Lys156-------116
      5.3.1.1.2 Indirect water-mediated protonation of \(\text{Ado76}^3\text{O}\) by Lys156:--------116
      5.3.1.1.3 Protonation of \(\text{Ado76}^3\text{O}\) by His73-H\(^+\).----------------------117
      5.3.1.2 His73-promoted editing through self-cyclization/lactone formation:----118
    5.3.2 Cys182-promoted editing Mechanisms:----------------------------------120
      5.3.2.1 Deprotonated Cys182 with protonated His73 and His186--------------121
        5.3.2.1.1 Direct protonation of \(\text{Ado76}^3\text{O}\) by Lys156.-------------------123
        5.3.2.1.2 Protonation of \(\text{Ado76}^3\text{O}\) involving His73-H\(^+\).-------------124
        5.3.2.1.3 Protonation of \(\text{Ado76}^3\text{O}\) involving Cys182-SH.-------------127
      5.3.2.2 Deprotonated Cys182 with neutral His73 and His186------------------127
        5.3.2.2.1. Protonation of \(\text{Ado76}^3\text{O}\) involving Lys156.------------------130
        5.3.2.2.2. Protonation of \(\text{Ado76}^3\text{O}\) involving Cys182.------------------132

5.4 Conclusion---------------------------------------------------------------135
Chapter 6.

Unraveling the Critical Role Played by 2'OH\textsubscript{Ado76} in The Post-Transfer Editing by \textit{Archaeal} Threonyl-tRNA Synthetase.

6.1 Introduction

6.2 Computational Methods

6.2.1 Molecular Dynamics (MD) Simulations:

6.2.2 QM/MM calculations:

6.3 Results and Discussion:

6.3.1 Concerted mechanism:

6.3.1.1 Anchoring Mechanism:

6.3.1.2 Proton-Shuttle Mechanism:

6.3.1.3 Double-Proton shuttle Mechanism:

6.3.2 Two-step Mechanism:

6.3.2.1 Anchoring Mechanism

6.3.2.2 Proton Shuttle Mechanism

6.3.3 D-amino acid deacylase function

6.4. Conclusion

6.5 References

Chapter 7.

QM/QM Investigation on The Pre-transfer Editing Mechanism in Seryl-tRNA and Isoleucyl-tRNA Synthetases

7.1 Introduction

7.2 Computational Methods

7.2.1 Molecular Dynamics Simulations:

7.2.2 QM cluster and QM/QM calculations:

7.3 Results and discussion

7.3.1 MD results

7.3.2 QM/SE Investigations
List of Tables

Table 3.1. Average values of key distances (Angstroms) and angles (degrees) obtained from 10 ns production MD simulations of the ThrRS···aa-AMP/tRNA<sup>Thr</sup> complex (aa=L-Thr, L-Ser, L-Val). ........................................................................................................................................................................ 42

Table 4.1. Free energy values (in kcal/mol) obtained for the two different mechanisms considered in case of the two enzymes using QM-only approach........................................ 82

Table 6.1. Calculated energy barriers in kcal/mol generated using different DFT functionals. ........................................................................................................................................................................ 152

Table 8.1. The calculated bond lengths, in Å, for Zn(II) binding environment using different DFT functionals. Note, the last row replaces Zn(II) with Cd(II)........................................ 206
List of Figures

**Figure 1.1.** The cellular synthesis of an aminoacyl-tRNA and its role protein biosynthesis.

........................................................................................................................................2

**Figure 1.2.** Surface representation of the double discrimination model in threonyl-tRNA synthetase to ensure overall fidelity mechanism; the catalytic domain in blue where the editing one is in grey colour.........................................................................................................................5

**Figure 2.1.** Flow diagram for a common strategy used to obtain a suitable chemical model for subsequent QM/MM or QM-cluster calculations. The PDB ID for the starting X-ray structure in this example is 1TKY. 10 ............................................................................................................................... 15

**Figure 2.2.** Schematic illustration of how the additive (top scheme) and subtractive coupling schemes (bottom scheme) compute the energy of the chemical model.......................................................... 22

**Figure 2.3.** Representative image for the different generations of DFT functionals according to J. Perdew 6d with an example corresponding functional......................................................................................... 29

**Figure 3.1.** Overlay of the substrates in the optimized structures of the bound Thr-AMP substrates in the a) pre-reactive complexes and b) reactive complexes when the Thr β-hydroxyl is either neutral (tube) or deprotonated (ball and stick) when ligated to the Zn(II) ion. ...... 49

**Figure 3.2.** The calculated PES surfaces for aminoacylation (threonylation) as catalyzed by ThrRS when the threonyl moieties β-hydroxyl group ligated to the Zn(II) is neutral (blue) or deprotonated (black) throughout the mechanism. ................................................................. 52

**Figure 3.3.** Overlay of the substrates in the optimized structures of the bound Thr-AMP substrate ( bPRC Thr ) and the studied substrates a) bPRC Ser , b) PRC Val and c) PRC DThr ....... 55
Figure 3.4. The calculated PES surfaces for aminoacylation (serinylation) as catalyzed by ThrRS for when the serinyl moieties β-hydroxyl group ligated to the Zn(II) is neutral (blue) or deprotonated (black) throughout the mechanism. ................................................................. 56

Figure 4.1. Schematic drawing of the QM models used for the QM-cluster and the QM-region within the QM/MM models for (a) GlnRS (PDB ID: 1EUQ) and (b) ND-GluRS (PDB ID: 3AKZ). The substrate's aminoacyl moiety in each is highlighted in red......................... 77

Figure 4.2. Plots of the variation in some key distances in Ångström along 10 ns simulations: a) \(\text{Ado}^76^2'-\text{O}^--\text{O}_1\beta\) distance for wildtype (blue lines) and mutant (red lines) GlnRS; b) \(\text{Ado}^76^2' -\text{O}^--\text{C}_\text{carb}\) distance for wildtype (blue lines) and mutant (red lines) GlnRS; c) \(\text{Ado}^76^2'-\text{O}^--\text{O}_1\beta\) distance for ND-GlnRS; d) \(\text{Ado}^76^2'-\text{O}^--\text{C}_\text{carb}\) distance for ND-GluRS......................................................... 80

Figure 4.3. The first solvation shell directly hydrogen bonded to the phosphate groups in the substrates; the water oxygen density is in blue and the water hydrogen density is in grey. 81

Figure 4.4. Schematic representation of the obtained complexes along the studied aminoacylation mechanism together with selected key distances in Ångstrom; black color distances for the native enzymes, blue are for the Glu34Gln mutant. ............................................. 84

Figure 4.5. The free energy surface for the aminoacylation mechanism of glutamine in the case of the wildtype GlnRS (black color surface) and the Glu34Gln mutant GlnRS (blue color (dashed) surface). \(\text{Gln}^{\text{TS2}}\) and \(\text{Gln}^{\text{TS2'}}\) were obtained through scan calculations. ..... 86

Figure 4.6. Optimized structures for the aminoacylation mechanism in the wildtype GlnRS where the substrate’s α-NH₂ acts as the mechanistic base, together with energy values (in
kcal/mol, relative to the reactive complex) in parentheses. For clarity, the atoms that are involved in the reaction are highlighted. 

**Figure 4.7.** Schematic drawing of the complexes obtained during the O$_{ip}$ substrate-assisted aminoacylation mechanism of ND-GluRS, together with selected key distances in Ångstrom.

**Figure 4.8.** Free energy surface for the aminoacylation mechanism of glutamate by ND-GluRS.

**Figure 5.1.** The template QM-region used to study possible catalytic mechanisms of the ThrRS editing domain in which either (a) His73 or (b) Cys182 acts as the required base.

**Figure 5.2.** Schematic illustration of the optimized structures obtained (see Computational Methods), with select distances shown in Angstroms, for the His73 promoted mechanisms obtained with Lys156 protonating either the (1) $\text{Ado}_{76}^3$O or (2) $\text{O}_{\text{carb}}$ centers of the Ser-tRNA$^{\text{Thr}}$ substrate.

**Figure 5.3.** Free energy surfaces (kcal/mol) calculated (see Computational Methods) for the ThrRS editing domain mechanisms in which His73 acts as a base with Lys156 acting as an acid and protonating either the $\text{O}_{\text{carb}}$ center (blue line) or $\text{Ado}_{76}^3$-O (red line).

**Figure 5.4.** Schematic representation of the optimized complexes, with select distances shown in Angstroms, obtained for a His73-promoted self-cyclization editing mechanism. The relative free energies (kcal/mol) are also given in brackets.
**Figure 5.5.** Schematic illustration of the optimized structures obtained, with select bond distances shown in Angstroms, for the mechanism in which Cys182 acts as a base to activate an active water, and with Lys156 acting as acid and directly protonating the Ado76$^{3'}$-oxygen.

Figure 5.6. FES obtained (kcal/mol) for the three possible editing mechanisms elucidated in which Cys182 acts as a base and His73 and His186 are protonated and the required acid for cleavage of the Ado76$^{3'}$ O—C$_{\text{carb}}$ bond is: Lys156 (black line and labels); His73 (red line and labels); or Cys182 (blue line and labels).

Figure 5.7. Schematic illustration of the optimized structures, with selected bond lengths shown in Angstroms, obtained for the Ila editing mechanism (i.e., initial active site contains Cys182S—/His73-H+/His186-H+) in which cleavage of the Ado76$^{3'}$ O—C$_{\text{carb}}$ bond involves proton transfer from His73 (2: red labels) or Cys182 (3: blue labels).

Figure 5.8. Schematic illustration of the optimized structures (with select bond lengths shown in Angstroms) obtained for the editing pathways in which Cys182 acts as the base in the presence of neutral His73 and His186, and where either Lys156 (1: black labels) or Cys182 (2: red labels) acts as the required acid for the second proton transfer.

Figure 5.9. FES obtained (kcal/mol) for the two possible editing mechanisms elucidated in which Cys182 acts as a base and His73 and His186 are neutral and the required acid for cleavage of the Ado76$^{3'}$ O—C$_{\text{carb}}$ bond is: Lys156 (blue line and labels); or Cys182 (red line and labels).
Figure 5.10. FES obtained (kcal/mol), together with optimized structures with select bond distances shown (Angstroms), for the ThrRS editing domain's mechanism IIc in which Cys182 acts as a base, but Cys182, His73 and His186 are all initially neutral. ........................................... 134

Figure 6.1. Illustration of the high layer (QM region) of the QM/MM model used in this study ................................................................................................................................. 146

Figure 6.2. The optimized molecular structures for the obtained TSs with selected bond lengths in Å ........................................................................................................................................... 150

Figure 6.3. The calculated free energy surface in kcal/mol for the concerted mechanisms using M06HF functional .................................................................................................................................. 151

Figure 6.4. The obtained free energy surface in kcal/mol for the most favorable step-wise mechanism occurring through a 6-membered cycle ........................................................................................................... 157

Figure 6.5. The obtained free energy surface for the deaminoacylation of D-threonine. 159

Figure 7.1. The selected residues treated in the QM layer in SerRS (left model) and in IleRS (right model). ............................................................................................................................................................................. 172

Figure 7.2. The conformation adopted by the substrate as well as a plot indicating the change in selected key distances resulted from the MD simulation; A) for the Cys-AMP in SerRS and B) for Hcy-AMP in IleRS. ............................................................................................................................................................................. 176

Figure 7.3. Optimized molecular structures during the studied substrate-assisted mechanism with selected key distances as well as free energy differences for IleRS. The values in blue color obtained using QM-cluster calculations. ............................................................................................................. 178
Figure 7.4. Optimized molecular structures along the studied substrate assisted mechanism with selected key distances when Cys-AMP is the substrate in SerRS. ................................. 181

Figure 7.5. The relative free energy surface of the self-cyclization mechanism in SerRS; the black, red and blue colored surface is for Cys-AMP substrate, Thr-AMP and Ser-AMP substrates, respectively. ................................................................. 183

Figure 7.6. Optimized molecular structures of the studied substrate-assisted mechanism with selected key distances in Å when Thr-AMP is the substrate. ............................................. 185

Figure 7.7. Optimized molecular structures along the studied substrate-assisted mechanism with selected key distances when Ser-AMP is the substrate. ............................................. 188

Figure 7.8. The relative free energy surface of the self-cyclization mechanism in SerRS utilizing QM-only approach; the black, red and blue colored surface is for Cys-AMP substrate, Thr-AMP and Ser-AMP substrates, respectively. ............................................. 189

Figure 8.1. Optimized structures of the different βHNV isomers with different orientations of the terminal methyl using the B3LYP/6-31++G(d,p) level of theory. ......................... 198

Figure 8.2. Overlay of the representative active site structures for the two substrate isomers (H atoms are omitted for clarity). ................................................................. 199

Figure 8.3. The calculated free energy surface obtained at the ONIOM(B3LYP/6-311G+(2df,p):AMBER96)+ΔG_{corr} level of theory for the overall acylation mechanism for the neutral form (red surface) and deprotonated form (black surface) of βHNV-AMP substrate. ........................................................................................................ 204

Figure 8.4. Overlay of the two QM layers of the reactive complexes for the two different conformations (a and b) of βHNV-AMP (H atoms are omitted for clarity). ...................... 206
Figure 1.1. The cellular synthesis of an aminoacyl-tRNA and its role in protein biosynthesis.

Figure 1.2. Surface representation of the double discrimination model in threonyl-tRNA synthetase to ensure overall fidelity mechanism; the catalytic domain in blue where the editing one is in grey colour.

Figure 2.1. Flow diagram for a common strategy used to obtain a suitable chemical model for subsequent QM/MM or QM-cluster calculations. The PDB ID for the starting X-ray structure in this example is 1TKY.  

Figure 2.2. Schematic illustration of how the additive (top scheme) and subtractive coupling schemes (bottom scheme) compute the energy of the chemical model.

Figure 2.3. Representative image for the different generations of DFT functionals according to J. Perdew with an example corresponding functional.

Figure 3.1. Overlay of the substrates in the optimized structures of the bound Thr-AMP substrates in the a) pre-reactive complexes and b) reactive complexes when the Thrβ-hydroxyl is either neutral (tube) or deprotonated (ball and stick) when ligated to the Zn(II) ion.

Figure 3.2. The calculated PES surfaces for aminoacylation (threonylation) as catalyzed by ThrRS when the threonyl moieties β–hydroxyl group ligated to the Zn(II) is neutral (blue) or deprotonated (black) throughout the mechanism.

Figure 3.3. Overlay of the substrates in the optimized structures of the bound Thr-AMP substrate (bPRCThr) and the studied substrates a) bPRC^Ser, b) PRC^Val and c) PRC^DThr.
List of Schemes

**Scheme 1.1.** General two-step aminoacylation mechanism of aaRS. ........................................ 4

**Scheme 1.2.** Schematic drawing for the general editing mechanisms utilized by ThrRS..... 6

**Scheme 3.1.** The mechanism of ThrRS catalyzed threonylation of the corresponding tRNA$^\text{Thr}$ where the substrate’s $\alpha$-NH$_2$ serves as the required base (based on ref. 22). .................. 40

**Scheme 3.2.** Optimized structures, with selected bond lengths (Angstroms) shown, of the fully-bound active sites of the pre-reactive and reactive ThrRS-aa-AMP/tRNA$^\text{Thr}$ complexes for when the aminoacyl moieties $\beta$-hydroxyl is (a) neutral ($^a$PRC$^\text{aa}$, $^a$RC$^\text{aa}$) or (b) deprotonated ($^b$PRC$^\text{aa}$, $^b$RC$^\text{aa}$). (aa = Thr, R = CH$_3$ (black); aa = Ser, R = H (blue))....... 46

**Scheme 3.3.** Optimized structures, with selected bond lengths (Angstroms), of the transition structures (TS$^\text{aa}$) and product complexes (PC$^\text{aa}$) for threonylation of the cognate tRNA$^\text{Thr}$ as catalyzed by ThrRS. (aa=Thr, R=-CH$_3$; aa=Ser, R=H)........... 53

**Scheme 3.4.** Optimized structures, with selected bond lengths (Angstroms) and relative energies (in parentheses; kJ mol$^{-1}$) shown, of the pre-reactive and reactive complexes of ThrRS-Val-AMP/tRNA$^\text{Thr}$, and the transition structure and product complex for aminoacylation.................................................................................................................................................. 60

**Scheme 3.5.** Optimized structures, with select bond lengths (Angstroms) and relative free energy (values in brackets) shown, of the fully-bound active sites of the pre-reactive (PRC), reactive (RC), transition structure (TS), and product (PC) complexes for aminoacylation by D-threonine as potentially catalyzed by ThrRS. .......................................................... 63
**Scheme 4.1.** The general substrate-assisted acylation mechanism in Gln-/GluRS, (X=NH₂ or O), two different non-bridged oxygens (O₁₆ and O₂₆) in addition to the bridged oxygen (Ob) are available to participate. ................................................................. 74

**Scheme 5.1.** Schematic illustration of the two proposed pathways for post-transfer editing as catalysed by bacterial E. coli ThrRS in which the mechanistic water is activated by either (A) His73 or (B) Cys182 acting as the required base ................................................................. 105

**Scheme 6.1.** Schematic representations for the aminoacylation and editing mechanisms employed by ThrRS. ........................................................................................................... 141

**Scheme 6.2.** Chemical schematic representation of the studied concerted mechanisms, a) for the anchoring mechanism (4-membered cycle), b) for proton Shuttle (6-membered cycle) and c) for double-proton shuttle (8-membered cycle). ................................................................. 149

**Scheme 6.3.** Schematic chemical representation for the studied two-step mechanism, a) proton shuttle 4-membered, b) 6-membered cycles ................................................................. 155

**Scheme 7.1.** The generally proposed self-cyclization mechanism in the pre-transfer editing against Hcy.²⁴ ........................................................................................................... 170

**Scheme 8.1.** Schematic representation of the general two-step aminoacylation mechanism catalyzed by aaRS ........................................................................................................... 195

**Scheme 8.2.** The optimized molecular structures with selected bond lengths for the obtained complexes in case of two different isomers. Blue colored values are for isomer (B) and the black text for isomer (A) ........................................................................................................... 203
Scheme 8.3. Optimized molecular structures with selected bond lengths in Angstroms and relative energies (in parentheses; kJ mol$^{-1}$) for the obtained complexes in case of the Cd(II)-mutant complexes.

Scheme 9.1. The equation used to estimate the B.E, X and Y represent the chelating groups while K and M represent the Push-Pull substituents.

Scheme 9.2. The most promising ligands considered in this study with their corresponding estimated binding energies in kJ/mol (values between brackets). The coordinated atoms are colored in blue.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>aaRS</td>
<td>Aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>AlaRS</td>
<td>Alanyl-tRNA synthetase</td>
</tr>
<tr>
<td>AMBER</td>
<td>Assisted Model Building with Energy Refinement</td>
</tr>
<tr>
<td>AspRS</td>
<td>Aspartyl-tRNA synthetases</td>
</tr>
<tr>
<td>AsnRS</td>
<td>Asparaginyl-tRNA synthetases</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>B3</td>
<td>Becke's 3-parameter exchange functional</td>
</tr>
<tr>
<td>CHARMM</td>
<td>Chemistry at Harvard Macromolecular Mechanics</td>
</tr>
<tr>
<td>CysRS</td>
<td>Cysteiny1-tRNA Synthetases.</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>EE</td>
<td>Electronic embedding</td>
</tr>
<tr>
<td>GlnRS</td>
<td>Glutaminyl-tRNA Synthetase</td>
</tr>
<tr>
<td>GluRS</td>
<td>Glutamyl-tRNA Synthetase</td>
</tr>
<tr>
<td>HisRS</td>
<td>Histidyl-tRNA Synthetases.</td>
</tr>
<tr>
<td>IC</td>
<td>Intermediate complex</td>
</tr>
<tr>
<td>IleRS</td>
<td>Isoleucine-tRNA Synthetases</td>
</tr>
<tr>
<td>IEFPCM</td>
<td>Integral equation formalism-polarizable continuum model</td>
</tr>
<tr>
<td>LYP</td>
<td>Lee-Yang-Parr correlation functional</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>ME</td>
<td>Mechanical embedding</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MetRS</td>
<td>Methionyl-tRNA synthetases.</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular mechanics</td>
</tr>
<tr>
<td>MOE</td>
<td>Molecular Operating Environment</td>
</tr>
<tr>
<td>ONIOM</td>
<td>Our own n-layered integrated molecular orbital and molecular mechanics</td>
</tr>
<tr>
<td>PC</td>
<td>Product complex</td>
</tr>
<tr>
<td>PES</td>
<td>Potential energy surface</td>
</tr>
<tr>
<td>PP&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum mechanics</td>
</tr>
<tr>
<td>QM/MM</td>
<td>Quantum mechanics/molecular mechanics</td>
</tr>
<tr>
<td>RC</td>
<td>Reactant complex</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>SP</td>
<td>Single-point calculation</td>
</tr>
<tr>
<td>SerRS</td>
<td>Serinyl-tRNA Synthetase</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer ribonucleic acid</td>
</tr>
<tr>
<td>TrpRS</td>
<td>Tryptophenyl-tRNA synthetase.</td>
</tr>
<tr>
<td>TS</td>
<td>Transition state</td>
</tr>
<tr>
<td>vdw</td>
<td>van der Waals</td>
</tr>
</tbody>
</table>
Chapter 1.

Introduction
1.1 Introduction

Numerous biochemical reactions take place in every cell in all living organisms. The feasibility of such biochemical reactions might be kinetically challenging but enzymes are catalyze these reactions through tremendous rate enhancements. While many strategies are employed by enzymes, providing stabilization to the transition state lowering the energy barrier is central to enzymatic catalysis. For instance, the phosphodiester bond, the backbone linkage of DNA and RNA molecules, is exceedingly resistant to spontaneous hydrolysis. The half-life scission time of its hydrolysis is estimated to be 30 million years, but in the presence of metallonucleases the hydrolysis rate is accelerated dramatically by a factor of $10^{17}$. Understanding the incredible catalytic power of enzymes has been a goal of chemical and biochemical research for the past century. Considerable progress has been achieved; however, deeper insights into the source of the catalytic power of enzymes and the roles of enzymes associated in different biological processes and diseases are still much in demand. This, in turn, will open many novel routes into the development of many biological and biotechnological applications.

Figure 1.1. The cellular synthesis of an aminoacyl-tRNA and its role protein biosynthesis.
The aminoacylation process catalyzed by the ubiquitous aminoacyl-tRNA synthetases (aaRS) is a critical step in faithful translation of genetic information into proteins. High accuracy in translation is essential for preserving cellular function. Additionally, aaRSs have been associated with viral assembly, chlorophyll biosynthesis, oxidative stress response and antibacterial therapy; however, it is most known for its pivotal role in protein biosynthesis. The family is split into two main classes according to distinct structural and mechanistic elements.

Specifically, aaRSs catalyze the attachment of an amino acid to its corresponding transfer RNA (tRNA), the molecule responsible for providing amino acids to the elongating polypeptide chain, in the form of aminoacyl-tRNA as shown in Figure 1. The enzyme performs this role with outstanding specificity with an error rate of less than 1 in every 10,000 reactions. This accuracy is vitally important, as errors in aminoacylation can lead to a variety of physiological and pathological problems including misfolded proteins, non-functional enzymes, cancer, and possibly cell death.

AaRSs catalyze the aminoacylation reaction via a two-step process: activation and acylation. In the activation step, the aaRS aligns its cognate amino acid together with adenosine triphosphate (ATP) in an appropriate orientation for the carboxylate of the amino acid to attack the α-phosphate of the ATP, leading to the formation of aminoacyl-AMP accompanied by the release of the inorganic pyrophosphate. Subsequently, the acylation reaction takes place where the aaRS catalyzes the aminoacyl transfer onto its cognate tRNA, Scheme 1.1.
As a result of the large sizes of the tRNA molecules, it has been found to be simple for each aaRS to identify its cognate tRNA. However, due to significant structural similarity between the amino acids, it is much more complicated for aaRSs to select their cognate amino acids from a pool of structurally and chemically similar molecules. As a result, aaRSs have evolved numerous strategies to ensure high fidelity in catalysis and accordingly faithful translation of the genetic code. Through a size-base discrimination mechanism, the aminoacylation site of an aaRS recognizes its cognate substrate and rejects non-cognate ones. For example, cysteinyl-tRNA synthetase shows impressive fidelity in substrate recognition and consequently, aminoacylation with $10^8$ fold selectivity.

Additionally, the majority of aaRSs possess an ability to edit mis-aminoacylation, either in the aminoacylation site through pre-transfer editing or by using a separate catalytic editing site through post-transfer editing. In the pre-transfer editing mechanism, the aaRS hydrolyzes aminoacyl adenylate intermediate to regenerate the free non-cognate amino acid which can then be cleared from the active site. In addition, a self-cyclization mechanism has been suggested for the pre-transfer editing against homocysteine and ornithine by Methionyl-tRNA synthetases and Lysyl-tRNA synthetases. Once the wrong amino acid is misacylated to the tRNA a post-transfer editing process can operate, during which the misacylated-tRNA is shuttled to a distal editing active site where the ester

---

**Scheme 1.1.** General two-step aminoacylation mechanism of aaRS.
bond between the misacylated amino acid and tRNA is cleaved, Scheme 1.2.\textsuperscript{29}

For instance, the aminoacylation site of threonyl-tRNA synthetases (ThrRS) accommodates a unique Zn(II) metal ion which assists in the removal of the non-cognate valine.\textsuperscript{30} However, this synthetic site is unable to fully distinguish between its cognate substrate, threonine, and the structurally related noncognate serine.\textsuperscript{31} Rather, it acts under a double-sieve model and employs a range of editing approaches to clear misacylated Seryl-tRNA\textsuperscript{Thr}, Figure 1.2.\textsuperscript{32}

\textbf{Figure 1.2.} Surface representation of the double discrimination model in threonyl-tRNA synthetase to ensure overall fidelity mechanism; the catalytic domain in blue where the editing one is in grey colour.

Only \textit{yeast mitochondrial} ThrRS lacks a separate editing site and thus appears to exploit only pre-transfer editing within its aminoacylation site.\textsuperscript{33} However, ThrRS from any other species is known to follow a double discrimination approach; that is, they exhibit both pre- and post-transfer editing.\textsuperscript{34} Notably, the post-transfer editing mechanism of ThrRS has been
observed to be somewhat species specific; in the case of the N-terminus motif (archaeal) ThrRS, the post-transfer editing process is thought to proceed via an RNA-mediated hydrolysis pathway. Specifically, it is thought to be facilitated by the tRNA co-substrate’s 2′/3′-OH \text{Ado}\text{76} group. Interestingly, sequence analysis has demonstrated a substantial sequence similarity between this motif in ThrRS and D-amino acid deacylases (DTD). The latter is utilized by aaRSs to ensure the homochirality of the synthesized protein by removing the misformed D-aa-tRNA\text{aa} but, unfortunately, the precise mechanistic role is unknown.

Meanwhile, in \textit{E. coli} species, the hydrolysis mechanism is a matter of much more debate. It has been proposed that an editing site cysteine or histidine residue acts as a mechanistic base to activate the nucleophilic water molecule. Interestingly, even by employing a double discrimination mechanism, ThrRS can still have difficulties since the unnatural amino acid β-hydroxynorvaline cannot readily be discriminated against and can be incorporated into proteins. Other aaRS are thought to go even further and employ a triple-sieve approach; for instance, distinguishing between serine and alanine by AlaRS is one of the greatest molecular-recognition challenges in nature.

\textbf{Scheme 1.2.} Schematic drawing for the general editing mechanisms utilized by ThrRS.
However, it has been recently suggested that this inaccuracy in translation is a “double-edged sword”.\textsuperscript{45} Absolute accuracy is not always essential and indeed, several aaRSs have been observed to tolerate moderate levels of misaminoacylation.\textsuperscript{46} In fact, this can be used as a physiological stress response or allow some amino acids to be incorporated into proteins even if their specific aaRS is lacking in certain species.\textsuperscript{47-48} For instance, it has been found that only \textit{eukaryota} and some \textit{bacteria} species possess twenty aaRS, each responsible for a specific, individual amino acid.\textsuperscript{49} Accordingly, indirect acylation pathways are required for the transfer of amino acids in those species that do not possess all 20 aaRS.\textsuperscript{49} For example, most \textit{prokaryota} lack AsnRS and GlnRS enzymes; instead, non-discriminating glutamyl-RS and aspartyl-RS can mis-aminoacylate tRNA\textsuperscript{Gln} and tRNA\textsuperscript{Asn} with glutamate and aspartate amino acids, respectively.\textsuperscript{50} These mischarged tRNAs are then converted to the desired Gln-tRNA\textsuperscript{Gln} and Asn-tRNA\textsuperscript{Asn} by Glu-tRNA\textsuperscript{Gln} and Asp-tRNA\textsuperscript{Asn} amidotransferases.\textsuperscript{51}

The exact mechanisms of the aforementioned reactions remain quite poorly understood. This is highly unfortunate given their tremendous potential impact on molecular recognition, physiological and pathological processes, and catalysis. The outgrowing computer and software capabilities made it possible for the state-of-the-art multiscale computational chemistry to provide significant insights into these critical and far-reaching problems.\textsuperscript{52-53} Specifically, it can shed light on the exact contribution of each active site residue to catalysis and thus precisely identify the catalytically active ones. This can provides the foundation required for the development of targeted and effective therapeutic agents for the treatment of the many disease states associated with these enzymes.\textsuperscript{54}
1.1 References


Chapter 2.

Computational Enzymology Methods
2.1 Molecular Modeling of Enzymes

Elucidating the complete catalytic mechanism of an enzyme means establishing and understanding all the factors that contribute to their remarkable ability to enhance the rate of reactions and to act with great specificity. This often requires, for instance, identifying and characterizing the structures, properties, and energy of all the intermediates and transition states along a reaction cycle. Experimental investigations have made numerous great strides and progress in the field enzymatic catalysis across many different aspects. Unfortunately, however, many aspects of enzymatic catalysis, particularly at the atomistic or electronic level, remain challenging or impossible to elucidate by current experimental techniques.1-2

Fortunately, the exponential growth of computational power has enabled the development and application of computational chemistry methods to, for instance, ever increasingly large and complex chemical problems. Indeed, computational enzymology, the use of such methods to study enzymes, is now able to provide a tremendous wealth of accurate and reliable insights into their structures, properties, energies, and catalytic mechanisms.3 Furthermore, such methods can be used to examine chemical systems from the atomistic to macromolecular level. Thus, computational enzymology has established itself as both a complementary and stand-alone approach to traditional experimental methods, and has become a booming area of active research and development in the field of enzymology.

In this dissertation a range of computational methods have been applied to study a range of enzymes. There have several great reviews in the literature on the field. Hence, within this chapter only a brief introduction and summary of such methods and their application, as relevant to the work herein, is provided.
2.2 Molecular Dynamics Simulations

Computational enzymology studies often start with obtaining or deriving a suitable chemical model from an experimentally derived high-resolution X-ray crystal structure. However, such experimental structures are themselves not necessarily an exact representation of the enzyme or complex in vitro being impacted by, for example, they are static, non-solvated, may contain mutations, and are potentially subject to crystal packing effects.

For example, it has long been recognized that, for instance, within cells enzymes are in dynamic motion and can adopt a range of conformations over time that exist from a few femtoseconds to several seconds. In contrast, the observed time scale for a catalytic reaction can be on the order of a millisecond to microsecond scale.\(^4\) This inherent flexibility has been found to be central to enzymatic catalysis, including their efficient substrate binding and product release, and is represented through a multidimensional free-energy landscape.\(^5\) Indeed, for any enzyme to function properly, it is essential to balance between its inherent flexibility and suitable stability to maintain its original coordinates.

Thus, unraveling enzymatic mechanisms at the atomistic level can require the use of molecular dynamics (MD) simulation, which enable one to model the time-dependent change of the atomic coordinates of the system under study.\(^6\) Specifically, the position of any atom in the protein at any time can be monitored. Furthermore, they can allow one to model the behaviour of a solvated system, as well as include the effects of temperature on conformation. MD simulations use classic Newtonian mechanics to determine the movement behaviour of a system under the pressure of the forces acting on it. Parametrized forcefields such as Assisted Model Building with Energy Refinement (AMBER)\(^7\)\(^-\)\(^8\) or Chemistry at Harvard
Macromolecular Mechanics (CHARMM)\(^9\) are used to compute the applied force on each atom of the complex systems and then integrate over time.

Figure 2.1. Flow diagram for a common strategy used to obtain a suitable chemical model for subsequent QM/MM or QM-cluster calculations. The PDB ID for the starting X-ray structure in this example is 1TKY.\(^{10}\)

In response to the enzyme flexibility, atoms are in a continuous movement during time and thus the applied forces are changing, thereby the latter need to be recalculated at each time step. Eventually, MD simulations sample the generation of all plausible conformations the enzyme can adopt in a certain time. The outcome of the MD simulations is a trajectory which reveals the fluctuations of all atoms in the specified time frame over thousands of time steps. Complete atomistic details of the atoms motion in the system under study can be obtained through the analysis of the generated trajectories. It is also crucial to mention that,
the length of the MD simulation should be adequate to allow the structural fluctuations to reach a plateau indicating the equilibrium of the generated conformations from which a snapshot can be chosen for the following QM/MM calculations, **Figure 2.1.** All the conformations obtained before this equilibrium state are usually discarded and a representative model is chosen from only the equilibrated ones. In complex biomolecular systems such as lipids, membrane proteins and related systems where longer timescales (≥ µs) are needed for proper simulations; a coarse-grained model can be applied.11 This model allows for a significant increase in the timescale of MD simulations by treating small groups of atoms as single particles.12

In this thesis, the Molecular Operating Environment13 (MOE) was used to prepare all the chemical models and the simulations themselves were performed using the NAMD program.14 It is important to bear in mind that the X-ray crystal structures do not always contain the cognate substrates; either an analogue or even no substrate (apoenzyme). In the latter case, molecular docking is the tool we use to insert the substrate in the most favorable binding mode inside the active site. Many docking protocols have been developed based on the idea of inserting the substrate in all the possible orientations using molecular mechanics (MM) method. Each generated complex is then energy minimized and the complex with the lowest energy is the one predicted to exhibit the substrate bound in its most favorable position.

It is also well established that there is a central influence of the conformational fluctuations of the enzyme on the calculated energy barriers for the studied mechanism.15-16 The active site residues change their positions regularly during the simulation and thereby their catalytic contributions are varied from one conformation to another. Inaccurate selection of the starting structure for any enzymatic study can lead to a proposal of the wrong
reaction mechanism. Accordingly, it has been suggested to select multiple initial points from
the generated conformations of MD simulations for subsequent enzymatic study.\textsuperscript{16}

2.2 Molecular Mechanics:

As we mentioned above, MD sampling is usually performed using molecular mechanics
(MM). In this empirical-based approach, the calculation ignores the electronic motion and
the energy is then calculated based on the nuclear positions only. Therefore, the chemical
system is described as spheres joined by springs. Specifically, the energy of this system is
derived as a function of its conformation and expressed as the sum of basic classic equations
that describe the valence terms (such as bond stretching, bond bending, bond torsion), van
der Waals energy, electrostatic interaction and cross terms (which describe the effects of the
motion of one molecule on the other). The constants used in the equation are parametrized
either from experimental data or through \textit{ab initio} calculations. A set of equations with their
respective constants is called a force field. Among many available forcefields, the most
widely used for proteins and nucleic acids are the AMBER and CHARMM forcefields.\textsuperscript{8-9}
This simplicity of the calculations allowed MM minimizations to be applicable to very big
chemical systems.

2.3 Quantum Mechanics and Quantum Mechanics/Molecular Mechanics.

A diverse range of computational tools are available and widely used in computational
enzymology. Two of the most common are the quantum mechanical (QM)-only, also known
as the QM-cluster method, and the Quantum Mechanics/Molecular Mechanics (QM/MM)
methodologies.\textsuperscript{17} Indeed, these approaches have become essential for exploring the structure
and the functions of biomolecules from the quite small to large multi-molecular species such
as enzyme-substrate complexes. However, each approach has its challenges and strengths and in the following sections we highlight at least some of these.

2.3.1 QM-only Approach:

Modeling enzymatic reactions using the QM-only method has been successfully applied to the study of a range of chemical reactions and, in particular metalloenzymes.\textsuperscript{18-20} At the moment, the QM-cluster method can only be used to treat chemical systems of 200 atoms or less, thus it can not treat an entire enzyme. This is due to the fact that a single quantum mechanical method is used to describe the entire chemical model. According to numerous investigations on different metalloenzymes, the error of the QM-only approach is estimated to be less than 5 kcal/mol.\textsuperscript{21}

More specifically, a central foundational idea behind this approach is that for an enzymatic system the active site residues that are catalytically essential in the mechanism or responsible for productive binding of the substrate are excised out of the protein and treated with a high level of theory.\textsuperscript{22} As a result of discarding the protein environment, two main essential factors that influence the excised model are ignored: the steric effect employed by the protein to maintain the geometry of the active site, which can lead to artificial and non-physically relevant geometry changes, as well as long-range interactions and polarization effects induced by the surrounding protein.\textsuperscript{23}

Two common solutions are introduced to overcome these limitations. It is typical to freeze one atom from each residue at its crystallographic or possibly MD minimized coordinate to help maintain the structural integrity of the model in place.\textsuperscript{22} Also, it is presently common to use a continuum medium with a dielectric constant to approximately represent the electrostatic effects due to the surrounding protein. The latter can be included via the use of
single point energy calculations on the optimized geometries. Different values for the dielectric constant can be used based on the polarizability of the protein, but a value of 4 is generally held to be the most representative of common protein environments.\textsuperscript{24}

The number of atoms included in the QM-only chemical model and that are ideal to accurately represent the real active site has been a matter of increasing debate.\textsuperscript{18-19, 25-26} But it is now generally established that the accuracy of the QM-only protocol increases when a large chemical model of 100-200 atoms is used.\textsuperscript{23} This has been concluded based on the finding that the impact of changing the value of the dielectric constant is insignificant when the model used is derived by truncation of the real system at some distance away from the active site.\textsuperscript{27}

Moreover, increasing the size of the QM model provides the system with two essential properties; (i) suitable flexibility to adapt to any required geometry changes along the chemical reaction, and (ii) a suitable compromise for treating the short- and long-range interactions. Many assessment studies have demonstrated that a convergence in the obtained potential energy surfaces of different chemical reactions can be achieved when chemical models in the range of 150-200 atoms are used.\textsuperscript{25-26} However, before expanding the size of the chemical model, special care should be taken to consider the consequences. In particular, the calculation time will increase dramatically and, the larger the model, the higher the chance to obtain various local minima during the optimization of the chemical pathway. In addition, larger chemical models do not allow for the use of the higher and thus more accurate and reliable computational methods.

It is also important to mention that several QM-only investigations have been performed on different metalloenzymes to examine the choice of X-ray structure used to derive a
Chapter 2

chemical model on the reliability of the obtained energy barriers.\textsuperscript{28} Applying different displacements on the fixed atoms was found to result in negligible changes in the kinetic energies as long as the system is composed of more than 100 atoms.\textsuperscript{29} Overall, if the resolution of the considered X-ray structures is reasonable, better than 2.0 Å, the starting structure is not critical to the computed energies by the QM-only approach.

In practicality, the QM-only approach is found to be helpful to provide preliminary insights into the catalytic mechanism. Because of the less expensive time cost, it can be used to determine from a variety of proposed or potential mechanisms the one that is likely most feasible. It can also be useful for benchmark investigations; that is, to determine the most reliable QM method and basis set for reliably and accurately treating the chemical system under study. However, one must also be aware of the inherent limitations associated with the use of such models.

2.3.2 QM/MM Modeling

The QM/MM multilevel methodology inspired by the pioneering work of Warshel and Levitt\textsuperscript{30}, became the method of choice in studying enzymatic catalysis.\textsuperscript{17, 31-35} In this approach, the entire enzyme is split into two main subsystems according to their importance in the chemical reaction. The active site where the bond forming and bond breaking takes place is the electronically active region and is treated by the QM level of theory (inner layer). The criteria for choosing the size and the number of atoms in this region is the same as the QM-only approach. In fact, it is generally accepted that different sizes leads to different results and thus different models should be assessed to establish the consistency of the obtained energy.\textsuperscript{36}
The remaining region of the protein is represented by an empirical MM method which treated by a forcefield such as AMBER or CHARMM. The entire protein can be represented but for the proteins that have many chains in its X-ray structure, the size of the model depends on the location of the active site. In most of the cases, one monomer containing the active site is a good representation for the entire protein. A large size of the chemical model makes it computationally more expensive in addition to the complexity of the generated conformations of the MM layer. The resulting error from the latter can be diminished by either fixing the alpha carbon atom of the residues 8 Å away from the substrate to its coordinates at the MM minimization or even fixing the remaining part of the protein entirely.

Also, the interaction between these two layers (QM–MM coupling) cannot be estimated by just combining the energy between them. Special precautions should be considered when selecting the boundary region. Two general approaches, additive and subtractive, are currently considered for treating the coupling between these two layers.

2.3.2.1 Additive and Subtractive QM/MM schemes

Additive Scheme. In this scheme the QM/MM energies are calculated by adding the MM energy of the low layer (\(E_{\text{MM}}\)), the QM energy of the the active site region or high layer (\(E_{\text{QM}}\)), and the QM energy of the electrostatic interaction between the high (QM) and low (MM) layers (\(E_{\text{QM/MM, coupling}}\)), Figure 2.2.
Figure 2.2. Schematic illustration of how the additive (top scheme) and subtractive coupling schemes (bottom scheme) compute the energy of the chemical model.

Subtractive Scheme. In this scheme, separate QM and MM calculations are performed. That is, the QM region is treated independently and the QM/MM coupling interaction is not counted explicitly but considered within the MM calculation. The energy expression of the subtractive QM/MM scheme is the sum of the MM energy of the entire system ($E_{\text{MM, entire system}}$) and the QM energy of the active site region ($E_{\text{QM, inner layer}}$), minus the MM energy of the inner layer. The latter is done to avoid inclusion of the energy of the inner layer through both the QM and MM calculations, Figure 2.2.

Because of this simplicity, the subtractive approach can be expanded to $n$ number of layers and the ONIOM approach developed by Morokuma and coworkers is an example of this scheme.\textsuperscript{38-39} This example has been successfully applied to combine two QM methods to represent the entire system or covering three layers. In the case of three layers there will be an additional medium layer which can be treated, for example, using a less computationally
expensive QM method (e.g., semiempirical) than used in the high layer. The resulting QM/QM/MM energy expression can be represented by:

\[ E_{\text{QM/MM}} = E_{\text{MM, entire system}} + E_{\text{QM, inner layer}} + E_{\text{QM, medium layer}} - E_{\text{MM, inner layer}} - E_{\text{MM, medium layer}} \]

At the heart of a QM/MM calculation is the treatment of the interaction between the QM and the MM layers and should be described accurately. Two types of interactions between these two regions are normally represented: bonded (bond stretching, bond bending and dihedral rotation) and non-bonded (electrostatic and van der Waals) interactions. According to the treatment of electrostatic coupling between the QM and MM regions, QM/MM calculations are said to use either mechanical or electrostatic (or electronic) embedding.

In the case of the mechanical embedding the interaction between the two layers is treated by a MM calculation, which is reasonable for van der Waals but not electrostatic interactions. In particular, as a result of being treated using an MM forcefield, the charge distribution of the outer MM layer does not directly interact with the inner QM layer, which can be a drawback for the mechanical embedding treatment. Increasing the size of the inner layer to account for the charge distribution is suggested to diminish the influence of this shortcoming.\(^4^0\) The integrated ONIOM method is by default a mechanical embedding method.\(^3^7\)-\(^3^8\) It should be noted that the ONIOM approach can be expanded to include multiple layers. For example, the three layer ONIOM method has been suggested as one possible approach to enhance modeling of the polarity of the protein environment around an enzyme’s active site.\(^3^9\)

In the case of electrostatic embedding the electrostatic interaction between the layers is treated using a more advanced level of computation. While this makes it more accurate it is computationally more expensive. In particular, the change of the charge on the inner region,
resulting from the electrostatic influence of the outer layer is represented by incorporating an additional one-electron term in the QM Hamiltonian to represent the MM point charges. The treatment of the QM-MM van der Waals interaction in electronic embedding is identical to the mechanical embedding treatment.

Interestingly, the mechanical embedding approach, which avoids the overpolarization of the QM layer caused by inclusion of the additional one-electron term in the QM Hamiltonian, was found to be more reliable than electrostatic embedding in several enzymatic investigations. However, electrostatic embedding led to faster convergence of the size of the QM layer relative to the mechanical embedding scheme.

Importantly, the explicit treatment of the junction area between the high and the low layer is found to be significant only in the cases where the outer region of the protein has an electronic influence on the inner layer. In the case of the additive QM/MM approach, however, there is no requirement for parameters to represent the link atoms which is not described by the force field.

In addition, the treatment of the boundary between the QM and MM layers can be largely classified into the link atom (or parametrized atom) and a localized orbitals approach. In the link atom approach, a link atom, usually hydrogen, is used to saturate the chemical valence of the truncated atom in the QM layer. But this hydrogen atom is not treated by either the QM or MM methods. However, the use of a hydrogen atom results in limitations in the accuracy of the obtained QM/MM energy and electron density. This is because it is in a sense artificial, as it does not exist in the original connection, and the distance between this artificial hydrogen atom and the QM/MM boundary is only 0.5 Å, which is shorter than typical of an X-H bond. By comparison, in the localized orbital approach a QM calculation is used to
describe the connection between the inner and outer layers and thus, reduce errors due to overpolarization. However, the link atom approach is more straightforward and widely used.

2.3.2.2 Technical aspects to consider when constructing a QM/MM model.

When truncating the QM-MM conjunction, the coupling boundary should be at an adequate distance from the active site to be feasible computationally. More specifically, the boundary should be at least three bonds away from the location of bond formation and breakage to guarantee enough flexibility to afford dihedral rotation. It is also important to avoid placing the boundary at any polarized or conjugated bond (either a linear bond or in a ring system). Globally, the most appropriate place to cut at is at a single bond between non-polarized atoms such as aliphatic C–C single bonds.

2.3.3 Comparison between QM-cluster and QM/MM approaches:

Studying the enzymatic mechanism using QM-only was found to result in a significant change in the energy barrier compared to studies done using the QM/MM approach. However, absence of the protein environment does not induce significant changes in the obtained relative energies for the chemical reaction if it is not associated with a large degree of charge transfer. To examine the effect of including the protein environment on the energy of the inner layer, \( E_{QM} \), the difference in QM energy of the isolated chemical model in the gas phase \( (E_{QM^*}) \) and the energy of inner layer inside the QM/MM model \( (E_{QM, QM/MM}) \) is calculated, \( E_{MM} = E_{QM, QM/MM} - E_{QM^*} \).

Convergence of energy using QM/MM models relative to the QM-cluster approach with respect to the chosen size of the QM model has been the subject of numerous studies.
general, the QM/MM approach was found to converge faster with increasing system size than the QM-cluster approach. For instance, benchmarking QM/MM studies on the proton transfer from cysteine to histidine in [Ni,Fe] hydrogenase demonstrated the convergence of the obtained energy barrier once the QM/MM junction is moved away from the active site.\textsuperscript{41} Following the same protocol, it was demonstrated that addition of more residues to the QM-cluster system should not occur according to chemical intuition or to the residues’ proximity to the active site, but rather to their contribution to the QM/MM energy.\textsuperscript{48-50} In these studies, buried charged groups up to 20 Å away from the active site were found to significantly contribute to the obtained energy barrier. Moreover, including additional neutral residues in the QM region is significant only if they are located less than 4.5 Å away from the active site. Importantly, it is also suggested that all the polarized residues located up to 10-15 Å away from the active site should be treated at the QM level for accurate modeling of the enzyme.\textsuperscript{41, 49} To illustrate, a recent QM/MM study declared the convergence of the QM region when the size reached 200-300 atoms,\textsuperscript{51} while in another investigation, convergence in the energy was achieved beyond 300 atoms.\textsuperscript{52}

Another comparison between QM-cluster and QM/MM models in the study of enzymatic reactions has been performed on the mechanism of tungsten-dependant acetylene hydratase.\textsuperscript{36, 53} In this comparison, combining the two approaches to complementarily inform each other was found to be a necessary step towards successful elucidation of the enzymatic reaction. However, unlike the previous suggestions,\textsuperscript{49} the authors do not support adopting the energy resulted from expanding the size of an optimized QM region in the QM/MM model through the use of single point energy calculations.\textsuperscript{53} Significant differences in the
obtained energies relative to the fully optimized QM/MM models have been observed and were found to be independent of the chosen models.\textsuperscript{36}

2.4 Density Functional Theory (DFT)

Electronic structure calculations on the QM region can be performed using a variety of methods including \textit{ab initio}, density functional theory (DFT), and semi-empirical calculations. The central goal of these treatments is to provide an approximate solution to the time-independent Schrödinger equation; the fundamental equation of quantum chemistry. Solving the time-independent Hamiltonian operator for multi-electron systems is incredibly complex, making it impossible to exactly solve the Schrödinger equation. A remarkable step toward expanding its applicability has been achieved upon the simplification made by the Born-Oppenheimer approximation. This allows us to assume that the atomic nuclei are stationary relative to the electrons due to their larger mass. As a result, the Schrödinger equation is separated into a nuclear and electronic part, simplifying the molecular Hamiltonian to the electronic Hamiltonian. Thus, many physical and chemical properties can be determined based on the ground state electronic structure.

The Hartree-Fock calculation (HF) is the most basic type of \textit{ab initio} calculation. It is a variational method, meaning the calculated energy is always equal to or greater than the exact one. It solves a series of one electron equations describing how each electron moves in a field of surrounding electrons and is built from a linear combination of basis functions (atomic orbitals). One of the main limitations of HF calculations is that it does not account for explicit electron correlation (repulsion), but just an average effect. Many types of correlated calculation such as Møller-Plesset perturbation Theory (MPn), and coupled cluster theory have been developed to correct for the lack of HF electron correlation.
Similarly, semi-empirical (SE) methods are another approach that was developed on the basis of the wave function to solve for the Hamiltonian operator with the inclusion of approximations obtained from empirical data. Thus, the SE method is much faster than any \textit{ab initio} calculation, but the obtained results are not always accurate. In this approach, most of the core electrons of the system are omitted and not treated in the calculation, extending its application toward much larger systems. SE methods are parametrized to either experimental data or \textit{ab initio} calculations to overcome the errors of ignoring the core electrons. This inaccuracy is a result of the fact that the molecule under study should necessarily have structural similarities to the database used to derive the parameters. The most commonly used SE methods are Austin Model 1 (AM1) and the parametrized method 3 (PM3).

Due to the limitations of electron correlation, the wave function theory that was globally applied to quantum mechanics has been avoided for large systems.\textsuperscript{54} Consequently, Kohn-Sham Density Functional Theory (DFT) is presently the most popular and powerful tool for many applications in computational chemistry.\textsuperscript{54-55} It indicates that all the ground state properties of any system can be derived from the electron density distribution over space.\textsuperscript{56} The electron density is independent of the number of electrons, a substantial benefit when examining larger chemical models (cost scales with N\textsuperscript{4}).\textsuperscript{57} With comparable efficiency to HF, DFT has an outstanding performance-to-cost ratio simultaneously being applicable to larger chemical systems.

Because the exact functional is unknown, DFT relies upon approximating the unknown exchange-correlation (XC) functionals of the electron density\textsuperscript{58}. This approximation is known to be the main shortcoming of DFT calculations. In the scope of this approximation,
significant progress has been achieved in approximating the XC. The first generation approximations is the local spin density approximation (LSDA) in which the XC functionals depend only upon local spin densities. Despite its simplicity, LSDA is known to not be useful in studying chemical compounds as it overestimates bond lengths and underestimates chemical barriers. The second generation includes the generalized gradient approximation (GGA) methods which consider the gradient of the electron density in the XC functionals, improving the accuracy of results. Thereby, introducing this generation of functionals was the cornerstone that enabled computational chemists to use DFT. For instance, the exchange functionals of Becke86 (B86) and the popular Lee-Yang-Parr (LYP) correlation functionals are pioneers of this generation. However, accurately describing the energy barriers is still not fully solved.

Figure 2.3. Representative image for the different generations of DFT functionals according to J. Perdew with an example corresponding functional.

As a result, developing the hybrid density functionals (H-GGA) that mix the XC from the GGA method with a percentage of Hartree-Fock exchange is one reason for the progress of DFT functionals. This remarkable step upon introducing H-GGA, particularly after the introduction of the common B3LYP functional, allowed DFT to become the most widely used method in quantum chemistry. In fact, the B3LYP functional became the functional
of choice in studying numerous chemical properties, together with the hybrid meta density functional (M-GGA), they led to a significant improvement in the accuracy of determining different chemical properties of molecules, Figure 2.3.\textsuperscript{64}

Despite the unparalleled success in the application of DFT functionals, it is unfortunate to know that they suffer from four major challenges.\textsuperscript{65} The largest error is the self-interaction error which arises from the electron interacting with itself in the columnic term described by the DFT Hamiltonian.\textsuperscript{66} This error directly influences the underestimation of barrier heights, but can be somewhat addressed by including additional contribution from the exact HF exchange.\textsuperscript{67} Also, another major limitation is the inability of these functionals to describe non-covalent long-range (van der Waals) interactions.\textsuperscript{68} Presently, there is ongoing progress towards eliminating this shortcoming through introducing empirical dispersion corrections such as Grimme’s empirical formula.\textsuperscript{69} Interestingly, either geometry optimization of the chemical models or single point energy calculations using B3LYP-D3 were found to be equally successful in minimizing this error in different systems.\textsuperscript{70} One more critical drawback of DFT functionals is that they all are ground-state methods and are unable to provide reasonable results for excited state applications. Finally, error arises from the inaccurate description of chemical systems containing transition metals, increasing with increasing the %XC included contrary to a solution to self-interaction error.\textsuperscript{71} To illustrate, M06L (with no HF exchange) led to accurate predictions in the excitation energies in some systems.\textsuperscript{72}

Quite noticeably, there is a clear conflict between the self-interaction error and inclusion of transition metals. To overcome the latter problem, it is suggested to utilize a DFT functional with low HF exchange; however this will lead to inappropriate estimation of the
barrier heights. Currently, the meta hybrid M05, M06, and the improved M08 suites of DFT functionals were found to be superior to other functionals in terms of diminishing the previously mentioned errors. In general, however, it is well known that there is not a single DFT functional that is suitable for all systems; the choice of the functional relies mainly on the property under evaluation and the nature of the chemical system under study. For instance, M062X outperformed the robust B3LYP functional in treating long range as well as providing a better description of barrier heights. Meanwhile, the local M06-L functional, is the most suitable for transition metal chemistry. The M06HF functional with full HF exchange takes advantage of this XC to avoid the self-interaction error, yet it is not suitable for transition metal chemistry.

It is also important to mention that M062X is parametrized to approximately describe long range interactions; however it fails to provide the same level of accuracy as dispersion corrected DFT functionals. In contrast, other studies demonstrated the success of the Minnesota functionals in describing long range interactions at a similar level to dispersion-corrected ones. Generally, other than the description of noncovalent interactions, M062X was recently recommended to be the most accurate among the 14 Minnesota density functionals.

In conclusion, to determine which is the best DFT function to evaluate a chemical system, one of two main tasks should be considered. The literature should be consulted as to which functional best describes the system or perform a benchmarking study of several DFT functionals against a known standard method such as CCSD to represent your chemical model.
2.5 References


Chapter 2

13. (MOE), M. O. E. *Chemical Computing Group Inc*, 2013.08; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.08.


Chapter 3.
The Roles of Zn(II) And Active Site Residues in Substrate Discrimination by Threonyl-tRNA Synthetase.
3.1 Introduction

Proteins have a diverse array of critical roles in cells and organisms. Their proper functioning depends on the accuracy of their synthesis and thus, the process of transcription and translation. Aminoacyl-tRNA synthetases (aaRS) are central enzymes in the translation process as they charge transfer-RNA (tRNA\(^{\text{aa}}\)) with their cognate amino acid (aa). This is achieved via two half-reactions in which they activate their cognate aa through reaction with adenosine triphosphate (ATP) to form an aminoacyl-adenylate (aa-AMP). Then, within the same site, they catalyze transfer of the aminoacyl moiety onto the Ado76 residue of their tRNA (tRNA\(^{\text{aa}}\)). Using a lock-and-key specificity for their substrate, aaRS are able to achieve an outstanding overall reaction fidelity with an error rate of \(\sim 10^{-4}\). Thus, in addition to their central role in protein biosynthesis and potential as drug targets, they are also exemplars of molecular recognition. As a result, they have increasingly been the subject of experimental and computational studies.

Structural and chemical similarities between the amino acids presents a significant challenge to aaRS in distinguishing between cognate and non-cognate substrates. For instance, it has been stated that amino acids that differ by only one methyl group provide no more than 1 kcal mol\(^{-1}\) difference in binding energy. As a result, the active site of aaRS may also bind non-cognate amino acids potentially resulting in mischarging of their tRNA\(^{\text{aa}}\). For example, AlaRS may misacylate its cognate tRNA\(^{\text{Ala}}\) with the sterically larger serine as well as the smaller glycine. Meanwhile, IleRS, PheRS, and ProRS may mischarge their corresponding tRNA\(^{\text{aa}}\)'s with non-cognate valine, tyrosine, and alanine or cysteine, respectively. In order to overcome these errors, some aaRS use a double-sieve model wherein both the above aminoacylation site and a second editing site perform pre- and/or
post-transfer editing.\textsuperscript{11} In the latter site, removal of the mischarged aminoacyl moiety from aa-tRNA is achieved via a tRNA-dependent mechanism.\textsuperscript{12}

In particular, the class II aminoacyl-tRNA synthetase ThrRS must discern its cognate substrate L-threonine from multiple non-cognate amino acids including valine and serine.\textsuperscript{13-14} It has also recently been the target of potential new antimicrobial drugs.\textsuperscript{15} Experimentally, several X-ray crystal structures of ThrRS with and without various ligands bound within its aminoacylation site have been obtained.\textsuperscript{13-14, 16} Notably, based in part on these structures the aminoacylation active site was found to contain an essential Zn(II) ion. Indeed, mutation of any of the three enzyme residues (a cysteinyl and two histidyl's) ligated to the Zn(II) inactivated or inhibited the enzyme.\textsuperscript{14, 17}

In general, aminoacylation as catalyzed by aaRS proceeds via a conserved substrate-assisted mechanism.\textsuperscript{18} More specifically, a non-bridging phosphate oxygen of the aa-AMP substrate acts as the required base to abstract a proton from either the Ado\textsuperscript{76}-2'- (class I) or Ado\textsuperscript{76}-3'OH (class II) group of the tRNA\textsubscript{aa}.\textsuperscript{4} However, in the case of ThrRS, experimental mutation studies concluded that the substrates non-bridging phosphate oxygens do not act as the base.\textsuperscript{19} Recently, using molecular dynamics (MD) and DFT-cluster computational methods we suggested that the experimentally observed bidentate ligation of the threonyl moiety of L-Thr-AMP to the Zn(II) ensures that its $\alpha$-NH$_2$ group is neutral, and enhances the lability of the Zn(II)···N$_{\text{Thr}}$ bond.\textsuperscript{20-21} As a result, the substrates $\alpha$-NH$_2$ is able to act as the required mechanistic base as shown in Scheme 3.1.
However, in addition to aiding substrate recognition, it has also been suggested that a key role of the Zn(II) may be steric or chemical discrimination against non-cognate amino acids.\textsuperscript{11} For example, it may facilitate discrimination against the isosteric but non-cognate valine due to steric hindrance with the latter's side chain $\beta$-methyl group.\textsuperscript{22} Meanwhile, serine, which differs from threonine by a single -CH$_2$- in its side chain, is able to bind in the same manner as threonine within the aminoacylation site of ThrRS. Thus, it has been suggested that the Zn(II) may help ensure that only amino acids possessing a $\beta$-hydroxyl group may bind within the aminoacylation active site.\textsuperscript{13-14} Experimentally, it has been indicated that ThrRS catalyzed aminoacylation by the non-cognate serine occurs at a rate 1000-fold less than for the cognate threonine.\textsuperscript{13-14} This difference in the reaction rate does mean that in addition to pre-transfer editing, post-transfer editing of mischarged Ser-tRNA$^{\text{Thr}}$ is required.\textsuperscript{23-24} It has also been suggested that aaRS play a crucial role in ensuring the homochirality of amino acid residues in proteins, which is essential to the proper folding and thus function.\textsuperscript{25-26} The role of the Zn(II) and active site of ThrRS in discriminating against D-Threonine is, however, as yet unknown.

In this present study, we have used a multi-scale computational approach to investigate substrate binding in the aminoacylation active site of threonyl-tRNA synthetase, and the
subsequent mechanism of aminoacylation. More specifically, the role of the Zn(II) ion in the binding and catalytic mechanism for the cognate substrate L-Thr-AMP and potential non-cognate substrates L-Ser-AMP, L-Val-AMP, and D-Thr-AMP, was examined using molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) methods.

3.2 Computational Methods

3.2.1. Molecular Dynamic (MD) Simulations

The Molecular Operating Environment (MOE) program was used to prepare all Molecular Dynamics (MD) simulations,\(^2^7\) which were then performed using the NAMD program.\(^2^8\) We have previously performed an MD study on the fully bound ThrRS complex.\(^2^0\) As then, the Michaelis complex was constructed using several X-ray crystal structures as templates (PDB ID: 1QF6, 1EVL, and 4EO4)\(^1^3^-^1^4,^2^9\) with the threonyl substrate being manually docked such that both its \(\alpha\)–NH\(_2\) and \(\beta\)-OH were ligated to the Zn(II) center as suggested by the experimental structures. Furthermore, our previous MD protocol was also used herein. First, the generated complete enzyme-substrate (ThrRS⋯L-Thr-AMP/tRNA\(^\text{Thr}\)) complex was minimized using the AMBER99 force field until the root mean square gradient fell below 0.05 kcal/mol Å\(^{-1}\). Using this structure, the bound substrate was then manually modified to generate starting structures for the corresponding ThrRS⋯L-Ser-/Val-AMP/tRNA\(^\text{Thr}\) complexes.

All three models were then solvated by adding a 2 Å layer of water (1361 water molecules) to generate fully solvated complexes of approximately 10971 atoms. Each resulting solvated complex was then minimized using the AMBER99 molecular mechanics force field until the root mean square gradient fell below 0.01 kcal/mol Å\(^{-1}\). The minimized
structures were then annealed over 100 ps from 150 to 300 K at constant pressure. This was followed by a 10 ns MD production run with a time step of 2 fs under constant pressure and temperature and without applying any restraints. In these simulations, the default settings implemented in the MOE software were used, including a cutoff of 10 Å for non-bonded interactions and tether ranges from 0-100 Å applied to the heavy atoms. For each simulation, the conformations obtained were analyzed and clustered based on the root-mean-square deviations (RMSD) of their active site heavy atoms and substrate. An average structure of the most dominant conformation for each ThrRS•••L-aa-AMP/tRNA\textsuperscript{Thr} (aa = Thr, Ser, Val) complex was then selected for analysis and to generate suitable starting models for the QM/MM studies (see below).

**Table 3.1** Average values of key distances (Angstroms) and angles (degrees) obtained from 10 ns production MD simulations of the ThrRS•••aa-AMP/tRNA\textsuperscript{Thr} complex (aa=L-Thr, L-Ser, L-Val).

<table>
<thead>
<tr>
<th>Geometrical Parameter</th>
<th>Aminoacyl substrate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Thr</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>3'O•••N\alpha (Å)</td>
<td>3.01</td>
<td>2.99</td>
<td>2.98</td>
</tr>
<tr>
<td>3'O•••C\text{carb} (Å)</td>
<td>3.17</td>
<td>3.22</td>
<td>3.20</td>
</tr>
<tr>
<td>\alpha N•••Zn (Å)</td>
<td>1.92</td>
<td>1.90</td>
<td>2.03</td>
</tr>
<tr>
<td>\beta O•••Zn (Å)</td>
<td>2.14</td>
<td>2.11</td>
<td>NA</td>
</tr>
<tr>
<td>\angle 3'O-C-O (º)</td>
<td>84.0</td>
<td>81.1</td>
<td>77.4</td>
</tr>
</tbody>
</table>
An alignment analysis was performed on the three average structures obtained. No significant differences in their bound active sites was observed as indicated by the RMSD value obtained of just 0.26 Å. Furthermore, the average values of mechanistically important distances and angles of all three complexes are quite similar as seen in Table 3.1. For the aminoacylation mechanism involving D-Thr, the active site-bound L-Thr-AMP was manually changed to its D enantiomer to generate ThrRS-D-Thr-AMP/tRNA\textsuperscript{Thr}.

3.2.2. QM/MM Investigations:

All QM/MM calculations were performed within the ONIOM formalism\textsuperscript{30-31} as implemented in the Gaussian 09 program.\textsuperscript{32} This approach has been successfully applied in the field of computational enzymology to investigate the catalytic mechanisms of other aminoacyl-tRNA synthetases\textsuperscript{33} and related enzymes.\textsuperscript{34}

A suitable chemical model was obtained by minimizing the average structure obtained from the corresponding MD simulation using the AMBER99 forcefield. The system was truncated to include the aa-AMP substrate and all residues and waters within ~20 Å of the substrate. Using a two-layer ONIOM(QM/MM) approach the entire chemical system was divided into two subsystems based on their relevance to the catalytic mechanism. The high-layer (QM-region) consisted of the aa-AMP substrate, Zn(II) ion, the side-chains of the three residues ligated to the Zn(II) ion (His385, His511 and Cys334), the side-chains of residues directly hydrogen bonded with the aa-AMP moiety (Gln484, Gln381, Arg363, Asp383, and Lys465), the Ado76 ribose of the tRNA\textsuperscript{Thr}, and two active site water molecules. It should be noted that based on available X-ray crystal structures the Gln381, Arg363 and Lys465 residues are thought to help stabilize the negative charge on the substrate's carbonyl oxygen.
Chapter 3

(O<sub>carb</sub>) and the non-bridging phosphate oxygens during the aminoacyl transfer reaction.\textsuperscript{13-14} In total, the QM region consisted of 129 atoms. The low layer (MM region) contained all remaining residues and waters for a total number of 2283 atoms.

For the QM-region, the density functional theory method B3LYP\textsuperscript{35-36} was chosen as it has been shown to perform the best of several common DFT functionals for studying Zn metalloenzymes.\textsuperscript{37} This method was used in conjunction with the 6-31G(d,p) basis set, while the AMBER96 force field was used to describe the MM-region. All QM/MM calculations were performed within the mechanical embedding (ME) formalism. Frequency analyses were also obtained at the same level of theory, i.e., ONIOM(B3LYP/6-31G(d,p):AMBER96) in order to characterize the nature of the stationary points (i.e., minima or transition structure). Relative energies were calculated by performing single point energy calculations on the above optimized structures at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory. To help ensure the integrity of the enzyme model was maintained, the α-carbon's of all low-layer amino acid residues were held fixed at their initial optimized MM positions.

3.3 Results and Discussion

3.3.1 ThrRS⋯L-Thr/Ser-AMP/tRNA<sup>Thr</sup> Michaelis Complexes:

The QM/MM optimized structures of the L-Thr and L-Ser containing pre-reactive Michaelis complexes (PRC), with selected distances (Angstroms), are shown in Scheme 3.2. As the two complexes share many similarities, only the ThrRS⋯L-Thr-AMP/tRNA<sup>Thr PRC</sup> (↑<sup>Thr</sup>PRC) is discussed unless otherwise indicated.

As part of an earlier QM-cluster based computational study on ThrRS we examined the Michaelis complex and aminoacylation mechanism of ThrRS.\textsuperscript{21} It was concluded that in the pre-reactive Michaelis complex, the threonyl moiety of the Thr-AMP substrate was
bidentately ligated to the Zn(II) ion via both its neutral $\alpha$-NH$_2$ and side chain $\beta$-OH groups, giving a pentacoordinate Zn(II) center. In the corresponding reactive complex (RC), however, the substrate's threonyl moiety was only monodentately ligated to the Zn(II) via its side chain $\beta$-hydroxy oxygen. Notably, in this slightly higher energy complex the threonyl side chain hydroxyl was anionic having transferred its proton to the nearby side chain carboxylate of the Asp383 residue. Furthermore, this reactive complex lay just 7.5 kJ mol$^{-1}$ higher in energy than the preceding initial PRC.

In contrast, in this present study, which uses a considerably more extensive chemical model in conjunction with a QM/MM-based method, two pre-reactive complexes are obtained (Scheme 3.2). In the lower energy complex, $^{a}$PRC$^{\text{Thr}}$, the threonyl moiety is singly ligated to the Zn(II) ion via its neutral $\alpha$-NH$_2$ group with a Zn(II)$\cdots$N$_{\text{Thr}}$ bond length of 2.11 Å. Meanwhile, its $\beta$-OH group has shifted away from the Zn(II), as indicated by the large Zn(II)$\cdots$O$_{\beta}$ distance of 3.31 Å, and is now strongly hydrogen bonded to the nearby side chain carboxylate of Asp383, $r$(Asp383COO$^-\cdots$HO$_{\beta\text{Thr}}$) = 1.58 Å. Consequently, the Zn(II) ion is tetracoordinate in its ligation. In addition, in $^{a}$PRC$^{\text{Thr}}$ the side chain amide of the active site Gln381 forms a weak hydrogen bond with the carbonyl oxygen (O$_{\text{carb}}$) of the Thr-AMP substrate, $r$(Gln381NH$_2^+\cdots$O$_{\text{carb}}$) = 2.34 Å. Meanwhile, the side chain guanidinium of Arg363 forms a moderately strong hydrogen bond of length 1.95 Å with a non-bridging phosphate oxygen of the Thr-AMP moiety. The same non-bridging phosphate oxygen also strongly hydrogen bonds with the protonated side chain amine of Lys465 via an active site water with $r$(OH$_2^+\cdots$OP) and $r$(H$_2$O$\cdots$HN$_{\text{Lys465}}$) distances of 1.60 Å and 1.70 Å, respectively (Scheme 3.2).
Scheme 3.2. Optimized structures, with selected bond lengths (Angstroms) shown, of the fully-bound active sites of the pre-reactive and reactive ThrRS•aa-AMP/tRNA\textsuperscript{Thr} complexes for when the aminoacyl moieties β-hydroxyl is (a) neutral (\textsuperscript{a}PRC\textsuperscript{aa}, \textsuperscript{a}RC\textsuperscript{aa}) or (b) deprotonated (\textsuperscript{b}PRC\textsuperscript{aa}, \textsuperscript{b}RC\textsuperscript{aa}). (aa = Thr, R = CH\textsubscript{3} (black); aa = Ser, R = H (blue)).

It has been observed that many Zn(II) metalloenzymes have a basic residue adjacent, but not ligated to the Zn.\textsuperscript{38} Often, the role of this residue is to aid the activation of an R-OH group bound to the Zn(II) ion by accepting its proton. A recent analysis determined that the β-OH groups of threonine and Serine do have acidic character.\textsuperscript{39} This can be enhanced by the Lewis acidity of the Zn(II) ion which often induces a decrease in the $pK_a$ value of the ligated R-OH group.\textsuperscript{40} ThrRS exhibits a similar feature with the positioning of the Asp383
residue adjacent, but not ligated, to the Zn(II) ion (see Scheme 3.2). As a result, an alternate pre-reactive complex (bPRC_{Thr}; Scheme 3.2b) was obtained in which the substrate's threonyl moiety is bidentately ligated to the Zn(II) via both its neutral α-NH$_2$ group and its β-hydroxy oxygen with bond lengths of 2.18 and 2.09 Å, respectively.

Importantly, the threonyl's α-hydroxy group has now transferred its proton to the side chain carboxylate of Asp383 with which it forms a strong Asp383COOH···O$_a$ hydrogen bond of just 1.67 Å. Notably, bPRC$_{Thr}$, which contains a pentacoordinate Zn(II), lies only 14.0 kJ mol$^{-1}$ higher in energy than aPRC$_{Thr}$ in which the Zn(II) is tetracoordinate. In addition, the barrier for interconversion of aPRC$_{Thr}$ to bPRC$_{Thr}$ is just 21.3 kJ mol$^{-1}$. This suggests the possible occurrence of either bound-active site structure in vivo and highlights the flexible coordination of Zn(II) with tetracoordinate being the most common.$^{41}$

In bPRC$_{Thr}$ significant differences are observed in the hydrogen bonding interactions of O$_{carb}$ and the substrates phosphate with the key active site residues (see Scheme 3.2). In particular, the Gln381NH$_2$···O$_{carb}$ distance has shortened from 2.34 to 1.98 Å. Furthermore, the guanidinium of Arg363 now forms quite short and strong hydrogen bonds with non-bridging and bridging (O$_b$) phosphate oxygens of the substrate's AMP moiety with lengths of 1.83 and 1.89 Å, respectively. During aminoacylation of tRNA$_{Thr}$ the O$_b$ centre gains significant negative charge due to the breaking of the O$_b$–C$_{carb}$ bond. It should also be noted that the side chain amine of Lys465 is now more strongly hydrogen bonded via a water bridge with the same non-bridging phosphate oxygen as Arg363 as indicated by $r$(OH$_2$···OP) and $r$(H$_2$O···HN$_{Lys465}$) distances of 1.46 and 1.54 Å, respectively (Scheme 3.2).

We also examined the corresponding pre-reactive complexes (aPRC$_{Ser}$ and bPRC$_{Ser}$) in which the non-cognate substrate Ser-AMP was bound within the aminoacylation active site
of ThrRS. Overall, the binding of Ser-AMP and the hydrogen bond network is similar to that seen in the case of the cognate substrate Thr-AMP (Scheme 3.2). In addition, key substrate-enzyme distances were also generally in close agreement with differences of 0.2 Å or less, especially for $^{b}$PRC$^{\text{Ser}}$ (i.e., when the substrate being bidentately ligated to a pentacoordinate Zn(II)). The only exceptions occur for $^{a}$PRC$^{\text{Ser}}$ (i.e., when the substrate is monodentately ligated to a tetracoordinate Zn(II)) and in the Glu381NH$_2$···O$_{\text{carb}}$ and Arg363···O$_b$ hydrogen bonds which decrease by 0.29 and 1.40 Å to 2.05 and 1.97 Å, respectively. Energetically, the same trends are observed as for when Thr-AMP is bound; $^{a}$PRC$^{\text{Ser}}$ lies lower in energy than $^{b}$PRC$^{\text{Ser}}$ by 9.3 kJ mol$^{-1}$ and can interconvert via TS$^{\text{Ser}}_1$ at a cost of just 18.0 kJ mol$^{-1}$ (see below).

These results are also in good agreement with recent X-ray crystal structures that observed very similar Zn···N$_{\text{substrate}}$ and Zn···O$_{\text{substrate}}$ distances for threonine and serine bound in the active site of ThrRS.$^{16,42}$ It also supports the suggestion that size-based discrimination by the aminoacylation active site of ThrRS is insufficient to wholly discriminate between L-SerAMP and L-ThrAMP.$^{11}$

### 3.3.2 ThrRS···L-Thr/Ser-AMP/tRNA$^{\text{Thr}}$ Reactive Complexes

In our previous QM-cluster based computational study on the aminoacylation mechanism of ThrRS,$^{21}$ the first step of the mechanism was determined to be cleavage of the labile Zn(II)···N$_{\text{Thr}}$ bond so that the resulting unligated neutral α-NH$_2$ group can act as the required base. Notably, the threonyl moiety remained ligated to the Zn(II) via its β-hydroxy oxygen. Based on the above pre-reactive complexes obtained, two reactive complexes are possible ($^{a}$RC$^{\text{Thr/Ser}}$ and $^{b}$RC$^{\text{Thr/Ser}}$) and are shown Scheme 3.2. In general, structurally, the same
features and trends are observed in $^{a/b}\text{RC}^\text{Ser}$ as for $^{a/b}\text{RC}^\text{Thr}$. Thus, unless otherwise indicated only those involving bound L-ThrAMP are discussed.

The complex $^{a}\text{RC}^\text{Thr}$ can be thought of as arising from $^{a}\text{PRC}^\text{Thr}$ by a substitution. More specifically, the threonyl's side chain hydroxyl (Thr$\beta$-OH) shifts and ligates to the Zn(II) with concomitant breaking of the Zn(II)$\cdots$N$\text{Thr}$ bond. Notably, the Thr$\beta$-OH group remains neutral (i.e., not deprotonated). Alternatively, it can be thought as arising from $^{b}\text{PRC}^\text{Thr}$ by breaking of the Zn(II)$\cdots$N$\text{Thr}$ bond with concomitant proton transfer from the side chain carboxylic group of Asp383 onto the threonyl substrates Thr$\beta$-oxygen center.

In $^{a}\text{RC}^\text{Thr}$ the Zn(II)$\cdots$N$\text{Thr}$ distance has increased from 2.11 Å ($^{a}\text{PRC}^\text{Thr}$) to 3.16 Å. Importantly, the -NH$_2$ group now forms a relatively short hydrogen bond of length 1.86 Å with the $\text{A76}^{3'}$-OH group of the tRNA$^\text{Thr}$ (Scheme 3.2). Concomitantly, the Thr$\beta$-OH group is ligated to the Zn(II) ion via its oxygen at a distance of 2.29 Å and retains its strong hydrogen bond with the adjacent Asp383 COO$^-$, though it has lengthened slightly to 1.64 Å. Meanwhile, the mechanistically relevant $\text{A76}^{3'}$$\cdots$C$_{\text{carb}}$ distance, which corresponds to the bond to be formed, is 3.14 Å. In general, the rest of the substrate-enzyme hydrogen bond network remains little altered from $^{a}\text{PRC}^\text{Thr}$ (i.e., distances differ by $\leq$0.12 Å). The largest change occurs for the Gln381 NH$_2$$\cdots$O$_{\text{carb}}$ interaction which shortens by 0.51 Å to 1.83 Å. It is also important to note that the tetratacordinate Zn(II) complex $^{a}\text{RC}^\text{Thr}$ lies 35.3 kJ mol$^{-1}$ higher in energy than the tetracoordinate complex $^{a}\text{PRC}^\text{Thr}$. 
Figure 3.1. Overlay of the substrates in the optimized structures of the bound Thr-AMP substrates in the a) pre-reactive complexes and b) reactive complexes when the Thr β-hydroxyl is either neutral (tube) or deprotonated (ball and stick) when ligated to the Zn(II) ion.

The alternate reactive complex \( b^{RC} \text{Thr} \) also contains a tetracoordinate Zn(II) center and lies 7.7 kJ mol\(^{-1}\) lower in energy than \( a^{RC} \text{Thr} \). It can be considered as being formed from \( b^{PRC} \text{Thr} \) by simple cleavage of the Zn(II)⋯N\(_{\text{Thr}}\) bond. Indeed, the latter has lengthened from 2.18 Å (\( b^{PRC} \text{Thr} \)) to 3.17 Å with the α-NH\(_2\) group again making a strong hydrogen bond with the Ado763′-OH moiety, though now with a shorter length of 1.78 Å (Scheme 3.2).

Meanwhile, the A763′-O⋯C\(_\text{carb}\) and Thr\(_{\beta}\)-O⋯Zn(II) distances in \( b^{RC} \text{Thr} \) are 2.77 and 1.96 Å, respectively. Notably, these latter distances are both markedly shorter than observed in \( a^{RC} \text{Thr} \). The latter shorter bond reflects the fact that the Thr \(_\beta\)-oxygen remains anionic in \( b^{RC} \text{Thr} \), and forms only a moderately strong hydrogen bond (2.05 Å) with the neutral side chain carboxylic of Asp383 (Scheme 3.2). Overall, the substrate-enzyme hydrogen bond network involving Gln381, Arg363, and Lys465 is similar to that observed in \( a^{RC} \text{Thr} \) with distances in agreement within 0.1 Å. An exception, however, occurs with the guanidinium
of Arg363 which is now more strongly hydrogen bonded to the substrates O$_b$ center at a distance of just 1.92 Å compared to 3.34 Å in $^a$RC$^{\text{Thr}}$.

The key cause of these differences in structure, and energy, is the protonation state of the Thr$\beta$-OH group when ligated to the Zn(II) ion. The above observations suggest that when it is deprotonated the threonyl moiety is positioned closer to the Ado76 residue of the cognate tRNA$^{\text{Thr}}$. Meanwhile, the substrates phosphate O$_b$ center is shifted towards the side chain of Arg363. These differences in position of the substrate between the two reactive complexes is illustrated in Figure 3.1. As can be seen, the comparatively modest differences between $^a$RC$^{\text{Thr}}$ and $^b$RC$^{\text{Thr}}$ in the position of the substrate's threonyl in the vicinity of the Zn(II) ion induce larger shifts in the position of its C$_{\text{carb}}$ and O$_b$ centers. Importantly, the $\angle_{A_76'3'O_{\text{carb}}\cdots C_{\text{carb}}}$ angle in $^b$RC$^{\text{Thr}}$ and $^a$RC$^{\text{Thr}}$ is 85.5° and 58.9°, respectively. This further suggests that the A$_{76}'$O center in $^b$RC$^{\text{Thr}}$ is better positioned for its subsequent nucleophilic attack at C$_{\text{carb}}$ than in $^a$RC$^{\text{Thr}}$. The same trends are observed for the corresponding L-SerAMP-bound complexes. The only difference occurs for the $\text{Ado76}3'-O\cdots C_{\text{carb}}$ distance, where it is negligibly shorter in $^a$RC$^{\text{Ser}}$ than $^b$RC$^{\text{Ser}}$ by 0.02 Å.

3.3.3 Aminoacylation (Threonylation) Mechanism.

To further understand the effects of the structural and energetic differences of $^a$RC$^{\text{Thr}}$ and $^b$RC$^{\text{Thr}}$, we examined the aminoacylation (threonylation) mechanism as catalyzed by ThrRS beginning from both reactive complexes. The resulting potential energy surfaces are shown in Figure 3.2, while the corresponding optimized structures, with selected bond lengths, are shown in Scheme 3.3.
In ThrRS catalyzed aminoacyl transfer the substrates $\alpha$-NH$_2$ accepts a proton from the A$_{76}$3'-OH group. In addition, the A$_{76}$3'-oxygen nucleophilically attacks the substrates C$_{\text{carb}}$ center to form the required A$_{76}$3'O—C$_{\text{carb}}$ bond, while the scissile PO—C$_{\text{carb}}$ bond must be broken. In agreement with our previous QM-cluster based computational study$^{21}$, threo-nylation occurs via a concerted mechanism regardless of whether it start from $^{a}\text{RC}^{\text{Thr}}$ or $^{b}\text{RC}^{\text{Thr}}$.

The threo-nylation mechanism involving the reactive complex $^{a}\text{RC}^{\text{Thr}}$ (i.e., the threonyl substrates $^{b}\text{OH}$ remains neutral; pathway A), the reaction proceeds via $^{a}\text{TS}^{\text{Thr}}$ at a cost of 115.6 kJ mol$^{-1}$ with respect to $^{a}\text{RC}^{\text{Thr}}$, or 150.9 kJ mol$^{-1}$ relative to $^{a}\text{PRC}^{\text{Thr}}$ (Figure 3.2). The corresponding product complex $^{a}\text{PC}^{\text{Thr}}$ lies 88.6 kJ mol$^{-1}$ higher in energy than $^{a}\text{RC}^{\text{Thr}}$.
or 123.9 kJ mol\(^{-1}\) higher in energy than \(^{a}\text{PRC}^{\text{Thr}}\). That is, threonylation is markedly endothermic.

**Scheme 3.3.** Optimized structures, with selected bond lengths in Angstroms (Thr (black); Ser (blue)), of the transition structures (TS\(^{aa}\)) and product complexes (PC\(^{aa}\)) for threonylation of the cognate tRNA\(^{\text{Thr}}\) as catalyzed by ThrRS. (aa=Thr, R=\(-\text{CH}_3\); aa=Ser, R=H)

In contrast, when threonylation follows from the alternate reactive complex \(^{b}\text{RC}^{\text{Thr}}\) (i.e., the substrates \(^{a}\)-hydroxy remains deprotonated **pathway B**), the reaction proceeds via \(^{b}\text{TS}^{\text{Thr}}\) at a markedly lower cost of 81.3 kJ mol\(^{-1}\) or 108.9 kJ mol\(^{-1}\) relative to \(^{b}\text{RC}^{\text{Thr}}\) or \(^{b}\text{PRC}^{\text{Thr}}\), respectively. Furthermore, the corresponding product complex \(^{b}\text{PC}^{\text{Thr}}\) lies 37.4 kJ
mol\(^{-1}\) lower in energy than \(^{bRC_{Thr}}\), and 9.8 kJ mol\(^{-1}\) lower in energy than \(^{aPRC_{Thr}}\). That is, threonylation via pathway B is exothermic.

In both pathways, the hydrogen bond networks between the leaving AMP moiety and active site residues are essentially the same and with similar interaction distances. Namely, the side chains of Arg363 and Lys465, the latter indirectly via a water, stabilize the negative charge of the phosphate as well as the increasing negative charge on its bridging oxygen \(O_b\) (Scheme 3.3). The key differences between the pathways instead involve the ligands and coordination of the Zn(II) center.

For pathway A, as can be seen in Scheme 3.3, in the transition structure \(^{aTS_{Thr}}\) an active site water has ligated to the Zn(II) center with a Zn(II)–OH\(_2\) distance of 2.33 Å. Simultaneously, this water hydrogen bonds with both the side chain carboxylate of Asp383 and the substrates carbonyl oxygen \(O_{carb}\) with distances of 2.12 and 1.72 Å, respectively. It is noted that \(O_{carb}\) also retains its hydrogen bond with the side chain amide of Gln381; \(r(O_{carb} \cdots H2N_{Gln381}) = 1.96\) Å. In the resulting product complex \(^{aPC_{Thr}}\) the water remains bound to the Zn(II) center, though now at a slightly longer distance of 2.58 Å. This is likely due in part to the fact that the \(\delta^+\)-OH group remains both neutral throughout the mechanism and hydrogen bonded to the side chain carboxylate of Asp383 in both \(^{aTS_{Thr}}\) (2.29 Å) and \(^{aPC_{Thr}}\) (2.29 Å). Consequently, the cationic Zn(II) center becomes and remains pentacoordinate.

In contrast, in the alternate pathway B in which the substrates \(\beta\)-hydroxy group is deprotonated, the Zn(II) ion remains tetracoordinate throughout. More specifically, as seen in Scheme 3.3, in \(^{bTS_{Thr}}\) an active site water forms a strong hydrogen bond bridge between the \(\beta\)-O and \(O_{carb}\) centers with distances of 1.66 and 1.94 Å, respectively, Scheme 3.3. The
$O_{\text{carb}}$ center again retains its hydrogen bond to the amide of Gln381 with length $r(O_{\text{carb}}\cdots\text{H}_2\text{N}_{\text{Gln381}}) = 1.92$ Å. These shorter hydrogen bonding interactions compared to those observed in $^a\text{TS}^{\text{Thr}}$ suggest that in $^b\text{TS}^{\text{Thr}}$ there is greater stabilization of the negative charge buildup on $O_{\text{carb}}$ that occurs during aminoacyl transfer. It should also be noted that despite maintaining a hydrogen bond with the neutral side chain carboxylate of Asp383 throughout the mechanism, $\text{Thr}_{\beta}$-O also remains strongly ligated to the Zn(II) ion in both $^b\text{TS}^{\text{Thr}}$ and $^b\text{PC}^{\text{Thr}}$ with distances of 2.02 and 2.00 Å, respectively (Scheme 3.3). The lower energies of $^b\text{TS}^{\text{Thr}}$ and $^b\text{PC}^{\text{Thr}}$ compared to $^a\text{TS}^{\text{Thr}}$ and $^a\text{PC}^{\text{Thr}}$ is due to several factors including that for pathway B the Zn(II) ion is tetracoordinate and forms a neutral complex, being ligated to two histidyl's and a cysteiny1 thiolate residue as well as the $\text{Thr}_{\beta}$-O$^-$ centre. It has been previously noted that neutral charged Zn complexes are more stable than corresponding complexes with either positive or negative charges.\(^{43}\)

![Diagram of tRNA and Zn(II) ion with ligands](image)

**Figure 3.3.** Overlay of the substrates in the optimized structures of the bound Thr-AMP substrate ($^b\text{PRC}^{\text{Thr}}$) and the studied substrates a) $^b\text{PRC}^{\text{Ser}}$, b) PRC$^{\text{Val}}$ and c) PRC$^{\text{DThr}}$.

### 3.3.4 Aminoacylation (Serinylation) Mechanism.

Aminoacylation using the non-cognate but viable substrate SerAMP exhibited, in general, similar trends to that observed for ThrAMP (Figure 3.4). However, some key differences in the structures and mechanisms were seen (see also Scheme 3.2).
For instance, aminoacyl transfer in which the serinyl's $\text{Ser}_{\beta}$-hydroxyl group ligated to the Zn(II) is deprotonated (i.e., pathway B) occurs with a significantly lower barrier than the alternate mechanism for when it is neutral (pathway A; Scheme 3.2). More specifically, $^a\text{PRC}_{\text{Ser}}$ can readily interconvert with $^b\text{PRC}_{\text{Ser}}$ via $\text{TS}_{1\text{Ser}}$ with a barrier of just 18.0 kJ mol$^{-1}$. This step is simply proton transfer from the $\text{Ser}_{\beta}$OH group to Asp383. The pre-reactive complex $^b\text{PRC}_{\text{Ser}}$ can then readily form $^b\text{RC}_{\text{Ser}}$ by breaking of the labile Zn(II)$\cdots\text{NH}_2$ substrate bond. Notably, $^b\text{RC}_{\text{Ser}}$ lies lower in energy than $^a\text{PRC}_{\text{Ser}}$ by 14.5 kJ mol$^{-1}$. The subsequent aminoacyl transfer occurs via $^b\text{TS}_{\text{Ser}}$ with a barrier of 128.9 kJ mol$^{-1}$ with respect to $^b\text{RC}_{\text{Ser}}$ (or 114.4 kJ mol$^{-1}$ relative to $^a\text{PRC}_{\text{Ser}}$). The final product complex lies slightly higher in energy than $^a\text{PRC}_{\text{Ser}}$ by 7.2 kJ mol$^{-1}$. Thus, the overall serinylation mechanism is slightly endothermic.

**Figure 3.4.** The calculated PES surfaces for aminoacylation (serinylation) as catalyzed by ThrRS for when the serinyl moieties $\beta$-hydroxyl group ligated to the Zn(II) is neutral (blue) or deprotonated (black) throughout the mechanism.
Chapter 3

The alternate mechanism in which the $\text{Ser}_\text{OH}$ group remains neutral while ligated to the Zn(II) ion (pathway A), proceeds via $^a\text{TS}^\text{Ser}$ with a significantly higher barrier of 153.5 kJ mol$^{-1}$ relative to $^a\text{PRC}^\text{Ser}$ (Figure 3.4). Unlike that observed for threonylation, both possible serinylation pathways give the same final product complex $^b\text{PC}^\text{Ser}$.

That is, the product obtained is that in which the serinyl's $\beta$-hydroxy is deprotonated while ligated to the Zn(II) ion (Scheme 3.2). Thus, unlike that observed for ThrAMP, serinylation via either potential pathway is endothermic and hence disfavored. It is noted that the barriers for serinylation via $^a\text{TS}^\text{Ser}$ (153.5 kJ mol$^{-1}$) and $^b\text{TS}^\text{Ser}$ (128.9 kJ mol$^{-1}$) both lie higher in the relative energy than those for the analogous pathways for Thr-AMP which had barriers of 150.9 kJ mol$^{-1}$ ($^a\text{TS}^\text{Thr}$) and 108.9 kJ mol$^{-1}$ ($^b\text{TS}^\text{Thr}$), respectively (Figure 3.2). Experimentally, both cognate Thr-AMP and non-cognate Ser-AMP are observed to be viable substrates for aminoacylation by ThrRS, with Thr-AMP preferred.$^{13-14, 23}$ The present results suggest that in both cases aminoacylation proceeds via pathway B in which the substrates $\beta$-hydroxy group is deprotonated when ligated to the Zn(II) center.

3.3.5 Discrimination against Val.

As noted in the Introduction, ThrRS must necessarily also discriminate against valine due to its similar shape and volume to threonine. However, unlike threonine, and serine, it lacks a side chain $\beta$-OH to interact with the Zn(II) ion. In addition, it does not undergo ThrRS catalyzed aminoacylation.$^{13}$ However, in order to obtain a more complete understanding of discrimination by ThrRS we also examined binding of Val-AMP within the active site of ThrRS and the subsequent possible mechanisms of aminoacylation. Optimized structures obtained for the pre-reactive, reactive, transition structure and product complexes
are shown in Scheme 3.4 along with select bond distances (Angstroms) and relative energies (kJ/mol).

Two possible pre-reactive complexes, \( {^a\text{PRC}}^{\text{Val}} \) and \( {^b\text{PRC}}^{\text{Val}} \), were obtained (Scheme 3.4). In both, the valinyl moiety of the substrate is bound to the Zn(II) ion by only its neutral \( \alpha \)-amine group. In the lower energy pre-reactive complex \( {^a\text{PRC}}^{\text{Val}} \), the Zn(II)−⋅⋅⋅N_{\text{Val}} \) distance is 2.19 Å. More importantly, the Zn(II) is tetracoordinate as it is also still ligated to the enzyme via a cysteinyl thiolate and two histidyl imidazoles. In addition, the active site water which is ligated to the Zn(II) in the apoenzyme but presumably displaced upon substrate binding,\(^{13}\) forms a hydrogen bond bridge between the carboxylate of Asp383 and the substrates carbonyl oxygen with \( \text{Asp383COO}^− \cdot⋅⋅\text{H}_2\text{O} \) and \( \text{O}_{\text{carb}}\cdot⋅⋅\text{H}_2\text{O} \) distances of 1.66 Å and 2.19 Å, respectively. Meanwhile, the Zn(II)⋅⋅⋅OH\(_2\) distance is 3.69 Å.

The alternate pre-reactive complex \( {^b\text{PRC}}^{\text{Val}} \) lies markedly higher in energy by 63.5 kJ mol\(^{-1}\). Notably, the active site water is now also ligated to the Zn(II) ion with a distance of 2.19 Å and as a consequence the Zn(II) ion is pentacoordinate, Scheme 3.4. The most significant difference in the substrate−enzyme hydrogen bonding network between the two pre-reactive complexes involves the substrates \( \text{O}_b \) center. In \( {^a\text{PRC}}^{\text{Val}} \) it is hydrogen bonded to the guanidinium of Arg363, whereas in \( {^b\text{PRC}}^{\text{Val}} \) the latter instead hydrogen bonds to one of the substrate's phosphate oxygens, \( r(\text{Arg363NH}_2\cdot⋅⋅\text{O}_{\text{pro-R}}) = 2.32 \) Å. It should also be noted that unlike the orientation of the substrate in the corresponding ThrAMP analogue, \( {^b\text{PRC}}^{\text{Thr}} \), the position of the substrate’s \( \text{O}_b \) center in \( {^b\text{PRC}}^{\text{Val}} \) is also 1.24 Å further away from the Arg363’s guanidinium group with \( r(\text{Arg363NH}_2\cdot⋅⋅\text{O}_b) = 3.13 \) Å. This difference is also seen in an overlay of the active site-bound substrate structures of \( {^b\text{PRC}}^{\text{Val}} \) and \( {^b\text{PRC}}^{\text{Thr}} \) (Figure 3.3).
We then examined possible mechanisms by which aminoacylation may proceed from either pre-reactive complex (i.e., beginning from a tetra- or pentacoordinate Zn(II) complex). However, a pathway could only be elucidated for the latter with a water ligated to the Zn(II) throughout. In the reactive complex (RC\textsuperscript{V}), which lies 80.1 kJ mol\textsuperscript{-1} higher in energy than a\textsuperscript{PRC}\textsubscript{Val}, the Zn(II)⋯N\textsubscript{Val} bond has broken. As for the other reactive complexes considered herein the nitrogen of the valinyl's α-NH\textsubscript{2} now forms a strong hydrogen bond with the A\textsubscript{76}3′-OH group of the tRNA\textsuperscript{Thr} with a distance of 1.76 Å. Concomitantly, the \( r(\text{Ado76}3′-\text{O}⋯\text{C carb}) \) distance has shortened to 2.82 Å, which is 0.05 Å longer than observed in b\textsuperscript{RC}\textsuperscript{Thr} (cf. Scheme 3.2).

The Zn(II) center is now tetracoordinate and the ligated water has been deprotonated, transferring one of its protons onto the carboxylate of Asp383. Consequently, the Zn⋯O\textsubscript{water} distance has shortened markedly to 1.94 Å (Scheme 3.4). Aminoacyl transfer then occurs via TS\textsubscript{2}\textsuperscript{Val} with an energy barrier of 161.2 kJ mol\textsuperscript{-1} relative to a\textsuperscript{PRC}\textsuperscript{Val}. This barrier is the highest of all aminoacylation pathways elucidated in this study and is not enzymatically feasible. The product complex PC\textsuperscript{Val}, in which the valinyl moiety had been transferred onto the A\textsubscript{76}3′-oxygen, lies 91.6 kJ mol\textsuperscript{-1} higher in energy than a\textsuperscript{PRC}\textsuperscript{Val} indicating that overall mechanism is also thermodynamically highly unfavorable.
Scheme 3.4. Optimized structures, with selected bond lengths (Angstroms) and relative energies (in parentheses; kJ mol\(^{-1}\)) shown, of the pre-reactive and reactive complexes of ThrRS...Val-AMP/tRNA\(^{Thr}\), and the transition structure and product complex for aminoacylation.

3.3.6 Chiral Discrimination against D-Threonine

It has been stated that for correct protein folding it is imperative that their constituent amino acids be homochiral.\(^44\) Indeed, amino acids in proteins occur in their L-isomer form. It is known that D-amino acid deacylases (DTD), which show similarities to the editing site of ThrRS, edit D-aminoacylated-tRNA formed by aaRS.\(^{45-46}\) However, ThrRS may also help enforce the homochirality of the aminoacylation process by discriminating, at least partially, between D- and L-threonine. Indeed, for AspRS and HisRS differences in the enzyme-substrate recognition network between the non-cognate D- and cognate L-substrates in their aminoacylation sites has been suggested as the main reason for their discrimination and
hence, stereospecificity of the aminoacylation process. Hence, the ability of the aminoacylation site of ThrRS to discriminate against the non-cognate enantiomer D-threonine was investigated. Key optimized structures obtained along the aminoacylation pathway, with selected bond lengths (Angstroms) and relative energies (kJ mol\(^{-1}\)), are shown in Scheme 3.5.

As observed for its corresponding enantiomeric cognate substrate L-Thr-AMP, two different pre-reactive complexes were obtained, \(^{a}\text{PRC}^{\text{DT}}\text{Thr}\) and \(^{b}\text{PRC}^{\text{DT}}\text{Thr}\). In the former complex the substrates threonyl moiety is ligated to the Zn(II) center via only its \(\alpha\)-amine nitrogen with a distance of 2.14 Å (Scheme 3.5). The Zn(II)\(\cdot\cdot\cdot\)(H)O\(\beta\)DThr distance is considerably longer at 2.92 Å, though this is 0.39 Å shorter than observed in \(^{a}\text{PRC}^{\text{Thr}}\). In contrast, in \(^{b}\text{PRC}^{\text{DT}}\text{Thr}\), which lies 33.0 kJ mol\(^{-1}\) higher in energy, the substrates threonyl component is ligated to the Zn(II) center via both its side-chain \(\beta\)-oxygen and \(\alpha\)-amine nitrogen with distances of 2.09 and 2.34 Å, respectively. The latter Zn(II)\(\cdots\)N\(\text{DT}\) interaction is 0.16 Å longer than observed in the analogous complex \(^{b}\text{PRC}^{\text{Thr}}\). It is important to note that all other Zn(II)\(\cdots\)N\(_{\text{substrate}}\) and Zn(II)\(\cdots\)O\(_{\beta\text{substrate}}\) interaction distances obtained in the stationary points along the D-threonylation pathway are within 0.05 Å of those obtained for their corresponding L-Threonyl containing complexes (cf. Scheme 3.2 and Scheme 3.3). Furthermore, and as also observed in \(^{b}\text{PRC}^{\text{Thr}}\) (see Scheme 3.2), the threonyl's \(\beta\)-hydroxy group has transferred its proton to the carboxylate of Asp383. But, in contrast to that seen in \(^{b}\text{PRC}^{\text{Thr}}\), the side chain guanidinium of Arg363 does not form a hydrogen bond with the mechanistically key bridging phosphate oxygen O\(_{b}\) as indicated by their separation of \(r(O_{b}\cdots\text{H}_{2}\text{NArg363}) = 3.25\) Å. Instead, as in \(^{a}\text{PRC}^{\text{DT}}\text{Thr}\) it remains hydrogen bonded to one of the substrate's non-bridging phosphate oxygens and carbonyl oxygens with distances of 2.23
and 1.78 Å, respectively. Comparison of the $^b\text{PRC}^{\text{Thr}}$ and $^b\text{PRC}^{\text{DThr}}$ complexes (Figure 3.3) suggests that the shift in position of the latter's $O_b$ center relative to a terminal amine of the adjacent Arg363 is one of the main differences between these two complexes.

In contrast to that observed for threonylation using L-Thr-AMP (cf. Figure 3.2), the corresponding reactive complex $^b\text{RC}^{\text{DThr}}$ lies 6.0 kJ mol$^{-1}$ lower in energy than the preceding $^b\text{PRC}^{\text{DThr}}$. However, despite this, its energy relative to $^a\text{PRC}^{\text{DThr}}$ (27.0 kJ mol$^{-1}$) is in close agreement with that obtained when L-Thr-AMP is the substrate (Figure 3.2; 27.6 kJ mol$^{-1}$). Structurally, as observed in the reactive complex for L-Thr-AMP, the Zn(II)⋯$N_{D\text{Thr}}$ bond in $^b\text{RC}^{\text{DThr}}$ has cleaved as indicated by its distance of 3.39 Å. Again, the nitrogen of the $\alpha$-NH$_2$ group now forms a relatively short hydrogen bond to the Ado76-3'OH group of the tRNA$^\text{Thr}$ with a length of 1.86 Å. Meanwhile, the mechanistically relevant $\text{Ado76-3'O}⋯C_{\text{carb}}$ distance is now 2.73 Å, which is 0.04 Å shorter than observed in $^b\text{RC}^{\text{Thr}}$ (Scheme 3.2). It is also notable that the $O_{\text{carb}}$ center also now forms strong hydrogen bonds with the nearby Gln381, $r(O_{\text{carb}}⋯H_2N_{Q381})= 1.91$ Å, and Arg363, $r(C_{\text{carb}}O⋯H_2N_{R363}= 1.79$ Å). However, the $O_b$ center still has not formed a hydrogen bond with Arg363, the shortest $bO⋯HN_{R363}$ distance being 3.14 Å (Scheme 3.5).

Aminoacylation of the tRNA then proceeds via $^b\text{TS}^{\text{DThr}}$ at a cost of 116.9 kJ mol$^{-1}$ relative to $^a\text{PRC}^{\text{DThr}}$. This is 8.0 kJ mol$^{-1}$ higher in energy than obtained for the analogous mechanism involving L-Thr-AMP (cf. Figure 2). This may in part be due to the lack of charge stabilization on the $O_b$ center via hydrogen bonds with any enzyme residues, in particular Arg363 as observed in $^b\text{TS}^{\text{Thr}}$ (cf. Scheme 3.5). It is also noted that the increasing negative charge on the $O_{\text{carb}}$ center is stabilized by hydrogen bonds with the side chains of Gln385 and
Arg363 with $r(O_{\text{carb}} \cdots \text{HN}_{363})$ and $r(O_{\text{carb}} \cdots \text{HN}_{\text{Arg363}})$ distances of 2.0 Å and 1.69 Å, respectively (Scheme 3.5).

**Scheme 3.5.** Optimized structures, with select bond lengths (Angstroms) and relative free energy (values in brackets) shown, of the fully-bound active sites of the pre-reactive (PRC), reactive (RC), transition structure (TS), and product (PC) complexes for aminoacylation by D-threonine as potentially catalyzed by ThrRS.

In the final product complex $\text{PC}^{\text{DThr}}$ the $\text{A76}^3'-\text{O} \cdots \text{C}_{\text{carb}}$ bond has formed with a length of 1.36 Å. Aminoacylation is accompanied by concomitant release of the substrates adenosine monophosphate moiety. However, the $\text{C}_{\text{carb}} \cdots \text{O}_b$ bond in $\text{PC}^{\text{DThr}}$ has only elongated to 1.53 Å whereas in the analogous L-threonyl containing complex $\text{bPC}^{\text{Thr}}$ the $\text{C}_{\text{carb}} \cdots \text{O}_b$ distance is 2.98 Å. This likely reflects a lack of stabilization of the negative charge on the AMP’s $\text{O}_b$ center in $\text{bPC}^{\text{DThr}}$ by hydrogen bonding with the guanidinium of Arg363, unlike that
observed in $^bPC^{\text{Thr}}$. In addition, distinct from that observed in all other product complexes examined herein, the AMP's O$_{\text{pro-S}}$ center remains unprotonated. In other words, the side chain protonated amine of Lys465 and active site water have not transferred a proton onto the O$_{\text{pro-S}}$ center in $^bPC^{\text{DThr}}$ though it does retain a quite strong hydrogen bond with the bridging water, $r$(O$_{\text{pro-S}}$····H$_{\text{H2O}}$) = 1.49 Å. Importantly, PC$^{\text{DThr}}$ lies significantly higher in energy relative to $^aPRC^{\text{DThr}}$ by 100.9 kJ mol$^{-1}$, indicating that aminoacylation involving D-Thr-AMP is highly unfavorable (Scheme 3.5) Furthermore, the reverse reaction has a barrier to reaction of just 16 kJ mol$^{-1}$ suggesting that it is likely to occur quite readily for any product formed.

3.4 Conclusion

The roles of the active site Zn(II) and residues in the aminoacylation mechanism catalyzed by threonyl-tRNA Synthetase (ThrRS) has been computationally investigated. Specifically, their role in ThrRS's ability to discriminate between its cognate substrate $L$-threonine and non-cognate $L$-serine, $L$-valine, and D-threonine has been examined using molecular dynamics (MD) and ONIOM QM/MM methods.

For both $L$-Thr- and $L$-Ser-AMP two pre-reactive complexes (PRC) were obtained. In the lower energy complex ($^aPRC^{\text{Thr/Ser}}$) the ligand binds to the Zn(II) only via its aminoacyl $\alpha$-NH$_2$ nitrogen; i.e., the Zn(II) is tetracoordinate. In the alternate PRC complexes lying higher in energy by 14.0 ($^bPRC^{\text{Thr}}$) and 9.3 ($^bPRC^{\text{Ser}}$) kJ mol$^{-1}$, the aminoacyl is bidentately ligated to the Zn(II) via its $\alpha$-NH$_2$ nitrogen and $\beta$-hydroxyl oxygen; i.e., the Zn is pentacoordinate.

Similarly, for both $L$-Thr- and $L$-Ser-AMP two possible reactive complexes (RC), in which the Zn(II)···N-$\alpha_{\text{substrate}}$ bond has been broken, were obtained. In the higher energy
complexes, $^a\text{RC}^{\text{Thr/Ser}}$, the Zn(II)-ligated $\beta$-OH group of the substrate's aminoacyl moiety remains neutral. In contrast, in the lower energy reactive complexes $^b\text{RC}^{\text{Thr/Ser}}$ the aminoacyl's Zn(II)-ligated $\beta$-hydroxyl has transferred its proton to the nearby carboxylate of Asp383. This transfer and the substrate's $\beta$-OH group acidity is facilitated by the Lewis acidity of the Zn(II) and the presence of the Asp383 residue. Consequently, in $^b\text{RC}^{\text{Thr/Ser}}$ a neutral tetracoordinate Zn(II)-ligand complex is formed. Deprotonation of the $\beta$-hydroxyl results in shorter Zn⋯O-$\beta_{\text{substrate}}$ bond which in turn helps better position the substrate relative to the tRNA$^{\text{aa}}$ and important active site residues for the subsequent aminoacyl transfer. Notably, the substrates $\alpha$-NH$_2$ forms shorter hydrogen bonds with the $A_{76}'$-OH moiety and, in general, the mechanistically central $A_{76}'$-O⋯C$_{\text{carb}}$ distance is greatly reduced in $^b\text{RC}^{\text{Thr/Ser}}$ compared to $^a\text{RC}^{\text{Thr/Ser}}$. The substrate's non-bridging phosphate oxygen O$_b$ forms shorter, stronger hydrogen bonds with Arg363. The subsequent ThrRS catalyzed aminoacyl transfer occurs in one step with barriers of 108.9 and 128.9 kJ mol$^{-1}$ for L-Thr- and L-Ser-AMP, respectively. From both kinetic and thermodynamic perspectives, aminoacylation of threonine is preferred. When the substrates Zn(II)-ligated $\beta$-hydroxyl group remains neutral, aminoacylation occurs with significantly higher barriers of 150.9 and 153.5 kJ mol$^{-1}$ for L-Thr- and L-Ser-AMP, respectively. In addition, in their respective transition structures and product complexes the Zn(II) ion is pentacoordinate as they also bind an active site water and thus, is also not charge balanced by its ligands. Also, the increasing negative charge on O$_b$ during the reaction is not stabilized by hydrogen bonding with Arg363.

L-Val-AMP, lacking a $\beta$-hydroxyl, is unable to form a Zn⋯O-$\beta$ bond and thus help form a neutral Zn(II) complex and facilitate cleavage of the Zn⋯N-$\alpha$ bond. Furthermore, it also
does not form a stabilizing $O_b \cdots H_2N_{\text{Arg363}}$ hydrogen bond. Hence, aminoacylation by L-valine is kinetically and thermodynamically disfavoured.

The enantiomeric D-Thr-AMP interacts with the Zn(II) center throughout the aminoacylation process in a manner similar to that observed for L-Thr-AMP. However, unlike L-Thr-AMP, its key non-bridging phosphate $O_b$ center does not form and thus is not stabilized by a hydrogen bond with Arg363. Hence, the barrier for aminoacylation (116.9 kJ mol$^{-1}$) is higher than that of its cognate L-enantiomer. Furthermore, the final product complex is 100.9 kJ mol$^{-1}$ higher in energy than the initial complex $^{a\text{PRC}}_{\text{DThr}}$ and hence is not thermodynamically favoured. This may indicate a possible role for the aminoacylation site of ThrRS in chiral discrimination. The present results illustrate how enzymes are able to modify the properties of their substrates and in doing so, can cause subtle but critical positional shifts in the bound substrate that facilitate the reaction. Furthermore, it also shows how this can be exploited by enzymes to discriminate against potential non-cognate substrates including chiral enantiomers. The present results also provide insights for possible future experimental studies including mutation of identified key active site residues such as Arg363, Asp383, and Gln381.

3.5 References


Chapter 4.

A Water-Mediated and Substrate-Assisted Aminoacylation Mechanism in the Discriminating Aminoacyl-tRNA Synthetase GlnRS and Non-Discriminating GluRS
4.1 Introduction

Aminoacyl-tRNA synthetases (aaRSs) are essential enzymes with extensive roles in a wide range of biological functions; however, protein biosynthesis remains their central purpose.\(^1\)\(^-\)\(^3\) AaRSs are a family of enzymes that are well known to ligate an amino acid to its corresponding tRNA with exceptional specificity.\(^4\) Many encoded amino acids are structurally or chemically quite similar and consequently it can be challenging to discriminate between them.\(^5\) Thus, to avoid any deleterious mistranslation errors, these enzymes possess different editing functions and almost half of them incorporate a separate proofreading domain.\(^4\)\(^,\)\(^6\)\(^-\)\(^7\)

Glutaminyl-tRNA synthetase, GlnRS, is a class I aaRS which specifically attaches glutamine to the cognate tRNA\(^{\text{Gln}}\) yielding Gln-tRNA\(^{\text{Gln}}\). It is able to discriminate against the isosteric non-cognate amino acid glutamate by \(10^7\)-fold.\(^8\) This high accuracy is achieved without any free-standing editing domain.\(^9\) It has been suggested that unlike the noncognate glutamate, glutamine adopts a catalytically favourable binding mode in the active site of GlnRS; an arginine residue acts as a negative selectivity determinant and positions glutamate in a less productive orientation.\(^10\)

Although it is commonly thought that translation requires remarkable fidelity, it has recently been proposed that inaccuracies in the process should be considered a “double-edged sword”.\(^11\) Particularly, many aaRSs are found to tolerate moderate levels of mistranslations,\(^12\) and these can be better described as adaptive translation.\(^13\) Only eukaryote and some bacteria species have the aaRSs for all twenty canonical amino acids, and accordingly tRNA-dependent indirect (noncanonical) pathways are needed for the biosynthesis of different amino acids in the remaining species.\(^14\) The aaRSs involved in
these pathways, in addition to identifying their cognate tRNAs, can also accommodate non-cognate tRNA for which the corresponding aaRS is missing.

For example, most prokaryotic species lack GlnRS. Instead, they use an alternative indirect route that involves a distinct type of non-discriminating (ND) GluRS enzyme, ND-GluRS, to achieve glutamine aminoacylation. Specifically, the ND-GluRS covalently links glutamate to the noncanonical tRNA\textsuperscript{Gln} to yield the misacylated Glu-tRNA\textsuperscript{Gln}. A similar situation exists for the aminoacylation of tRNA\textsuperscript{Asn} with aspartate, which is also catalysed by a non-discriminating enzyme, ND-AspRS. These mischarged tRNAs (Glu-tRNA\textsuperscript{Gln} and Asp-tRNA\textsuperscript{Asn}) are subsequently converted to Gln-tRNA\textsuperscript{Gln} and Asn-tRNA\textsuperscript{Asn} by Glu-tRNA\textsuperscript{Gln} and Asp-tRNA\textsuperscript{Asn} amidotransferase, respectively.

The main difference between discriminating and non-discriminating GluRS is the presence of an arginine residue in the latter; its mutation to glutamine results in a ND-GluRS. Aside from this structural difference, both GlnRS and ND-GluRS can recognize tRNA\textsuperscript{Gln} and participate in glutamine aminoacylation through direct and indirect routes.

AaRS-catalysed aminoacylation mechanism involving glutamine occurs within the aminoacylation site through two main steps, namely activation and acylation. Specifically, the glutamine amino acid is first reacted with ATP through an inline displacement mechanism to yield glutaminyl-AMP; in this form, the glutamine moiety is favourably oriented for the subsequent acylation. This is initiated when a base abstracts the proton from the Ado\textsuperscript{76′}-OH group of the tRNA sugar terminal which attacks the carbonyl (C\textsubscript{carb}) of the substrate, producing the glutaminylated-tRNA\textsuperscript{Gln} (Gln-tRNA\textsuperscript{Gln}), Scheme 4.1. It is generally accepted that the catalytic domain of aaRSs lack a mechanistic base residue and, accordingly, a general substrate-assisted catalytic mechanism is believed to operate.
particular, the aminoacylation mechanism of GlnRS has been proposed to exploit the same general mechanism as in HisRS, TrpRS, MetRS and PheRS: one of the non-bridged phosphate oxygens (O1_p or O2_p) of the glutaminyl-AMP acts as the requisite base.26-28

Scheme 4.1. The general substrate-assisted acylation mechanism in Gln-/GluRS, (X=NH₂ or O), two different non-bridged oxygens (O1_p and O2_p) in addition to the bridged oxygen (Ob) are available to participate.

In the case of HisRS, experimental substitution of one of the substrate's non-bridging oxygens for sulphur diminished the rate of reaction 10000-fold which is strongly indicative of the involvement of these oxygens in this substrate-assisted mechanism.24 More recently our group performed a DFT-cluster study on the aminoacylation mechanism in HisRS.29 In particular the potential for each of the substrate's phosphate oxygens to act as the catalytic base was examined, i.e. the bridging (Ob) or non-bridging oxygens (pro-R and pro-S). Importantly, the pro-S non-bridging oxygen was found to be the most favourable base and lead to the only feasible activation energy.

In the case of GlnRS the catalytic mechanism is contested, being suggested to be promoted either by the nearby Glu34 active site residue or one of the phosphate oxygens.27,30-31 According to an early crystallographic study, Glu34 may be able to act as the required mechanistic base, via a bridging water molecule, that deprotonates the tRNA’s Ado76-2’-OH group;30 the authors disputed the previously proposed mechanism where one of the
phosphate oxygens is the base. Furthermore, the $pK_a$ of the phosphate oxygens were calculated to be 1.5-2.0, which suggests an inability to abstract a proton. However, more recent mutagenesis analysis countered this hypothesis and emphasized that Glu34 plays no role in initiating the acylation mechanism. When Glu34 was replaced by glutamine the rate of reaction was decreased by a factor of $10^3$-$10^4$ fold, but not completely inhibited.\textsuperscript{26} This significant impact on the reaction rate was found to be a result of a structural change in the conformation of the active site binding pocket upon mutation. This argument was supported by the observation that this glutamate residue is not conserved among the other aaRS, even in the structurally similar GluRS.\textsuperscript{26} In addition, they found that the $pK_a$ of phosphate group rises during the reaction, meaning the basicity of the oxygen atoms is enhanced.

Our group has a long-standing interest in the mechanisms operating in aaRS enzymes;\textsuperscript{29, 32} the principal aim of this study is to provide atomistic details regarding the direct and indirect pathways towards the formation of Gln-tRNA$^{\text{Gln}}$. We have sought to clarify the impact of the Glu34 residue within the synthetic site of GlnRS and to identify the base in the mechanism. To further validate these findings, we have expanded our investigation to the non-discriminating enzyme, ND-GluRS. Our results clarify the suggested substrate-assisted scenario for this aminoacylation mechanism, which may be common to other class I aaRSs.

4.2 Computational Methods

4.2.1. Molecular Dynamic Simulation

Suitable experimental X-ray crystal structures for GlnRS (PDB ID: 1EUQ\textsuperscript{33}) and ND-GluRS (PDB ID: 3AKZ\textsuperscript{34}) with substrate analogues bound were used as templates for the chemical models. Using the Molecular Operating Environment (MOE) software\textsuperscript{35} we
prepared the two X-ray structures for Molecular Dynamics (MD) simulations. This preparation included mutation of the substrates to the native forms, addition of missing hydrogen atoms considering protonation states as predicted by MOE, and solvation by adding a 2 Å layer of water. This generated solvated GlnRS•••Gln-AMP/tRNA\textsuperscript{Gln} and ND-GluRS•••Glu-AMP/tRNA\textsuperscript{Gln} complexes, with total atom counts of 22158 and 20906 respectively. Also, the Glu34Gln mutant of GlnRS, an isostructural mutation designed to investigate the structural role of Glu34 in the active site, was generated in silico. A recent experimental site-directed mutagenesis study on the Glu34Gln mutant was observed to have comparable overall kinetics to the wild-type enzyme.\textsuperscript{26}

Finally, unconstrained molecular mechanics (MM) minimizations using the AMBER12 force field until the root mean square gradient fell below 0.01 kcal/mol·Å were performed on all the studied systems. It should be noted that both GlnRS and ND-GluRS enzymes are considered ribonucleoprotein enzymes where the tRNA is essential for the catalytic mechanism to proceed properly.\textsuperscript{36} Thus all the MD simulations have been performed with the tRNA bound. The minimized structures were then submitted for 100 ps annealing equilibration from 150 to 300 K at constant pressure. The resultant structures were set up for a 10 ns simulation run with a time step of 2 fs as per the default settings of the MOE software. These settings include a cutoff at 10 Å for non-bonded interactions and tether ranges from 0–100 Å applied to the heavy atoms. All MD simulations were performed using the NAMD program.\textsuperscript{37} All sampled conformations in each MD simulation were analysed and clustered based on their root mean square deviations (RMSD) relative to the first structure. Notably, the equilibrium state was achieved after the first 5 ns where the system reached a stable conformation state. A representative structure of the most
Chapter 4

prominent conformation was then truncated to include all atoms within 20 Å of the substrate to generate the chemical model for the following QM/MM calculation.

4.2.2 QM-Cluster and QM/MM calculation

The Gaussian09 program package was used for all QM and ONIOM(QM/MM) calculations. In order to compare between the two different possible mechanisms, we started our investigation by performing QM-cluster-based calculations. This approach has been successfully used to examine many catalytic mechanisms. The chemical cluster models used herein for GlnRS and ND-GluRS, included the Gln-/Glu-AMP substrates and the 3′-terminal adenosine (Ado76) nucleotide of the tRNA moiety (Figure 4.1). In addition, they also included all residues that may participate in initiating the mechanism, stabilizing the transition state (TS), and/or neutralizing the phosphate-leaving group. The rest of the protein has been omitted after capping the carbon atoms.

Figure 4.1. Schematic drawing of the QM models used for the QM-cluster and the QM-region within the QM/MM models for (a) GlnRS (PDB ID: 1EUQ) and (b) ND-GluRS (PDB ID: 3AKZ). The substrate's aminoacyl moiety in each is highlighted in red.
More specifically, for the GlnRS model (total of 146 atoms) Asp66 has been included due to its role in positioning the Gln-AMP substrate through salt bridge formation with the $\alpha$-NH$_2$ group in the optimum orientation for the reactions. His43 and Lys270 are also included since they stabilize the phosphate groups and the carbonyl oxygen (O$_{\text{carb}}$) of the substrate. Glu34 was also included in order to study the impact of its mutation on the mechanism. Finally, Lys72 was included due to its direct hydrogen bond interaction with Glu34. In the mutated Glu34Gln GlnRS model, the same residues have been represented except the Glu34Gln mutation.

For the ND-GluRS QM-cluster model (total of 138 atoms), Arg28 and Arg216 were included due to their direct hydrogen bond interactions with the carboxylate groups of the Glu34 residue and the substrate (i.e. GluCOO$^-$). In addition, Lys257 and Pro31 were included as they stabilize the developing negative charge on the phosphate group. Asp64 was also incorporated because of the salt bridge formation with the substrate's $\alpha$-NH$_3^+$ group. His220 was included to stabilize the developing negative charge on the substrate’s O$_{\text{carb}}$ centre.

The M06-2X density functional method was used throughout this study as the QM method of choice as it is stated to provide a good description for non-covalent long-range interactions and has previously outperformed B3LYP in representing the kinetic energy barriers of different systems.$^{40}$

For all QM-cluster studies, and in order to approximately represent the protein’s polar environment, optimizations were obtained at the M06-2X/6-31G(d,p) level of theory. Vibrational frequencies were also obtained at this level of theory so as to identify the nature of the stationary points and to estimate the thermal energy corrected to 298.15 K ($\Delta E_{298}$).
Afterwards, single point energy calculations were performed on the above structures using the IEFPCM solvation model with a dielectric constant ($\varepsilon$) of 4.0 at M06-2X/6-311+G(2df,p) level of theory with inclusion of the appropriate $\Delta E_{298}$.

Subsequently, the ONIOM(QM/MM) method was employed to more extensively explore the catalytic mechanism. In this approach each enzyme-complex system was divided into two subsystems. The QM layer (reactive region) of each included the key active site residues (Figure 4.1) most relevant to the mechanism and was effectively identical to the corresponding QM-cluster model (except for the capped hydrogen atoms). Meanwhile, the surrounding protein (low-layer) was modelled using the AMBER96 forcefield. Optimized geometries, frequencies, and approximate thermal energy corrected to 298.15 K ($\Delta E_{298}$) were obtained at the ONIOM(M06-2X/6-31G(d,p):AMBER96) level of theory. Relative energies were determined by single point calculations on the above optimized structures at the ONIOM(M06-2X/6-311+G(2df,p):AMBER96) level of theory with inclusion of the corresponding $\Delta E_{298}$.

## 4.3 Results and Discussion

### 4.3.1. MD simulation of wildtype/mutant GlnRS and ND-GluRS

As noted in the Introduction, in the substrate-assisted aminoacylation reaction it is necessary for the Ado76′′-OH proton to be abstracted by one of the oxygen atoms of the substrate’s phosphate. For both the solvated wildtype GlnRS···Gln-AMP/tRNA<sup>Gln</sup> and mutant Glu<sup>34</sup>GlnRS···Gln-AMP/tRNA<sup>Gln</sup> complexes, a plot of Ado76′′O···O<sub>p</sub> distances obtained during the course of the MD simulations is shown in Figure 4.2a. Importantly, their average distances were found to be 6.75 Å and 6.04 Å, respectively. These distances are too great for there to be a direct interaction between the Ado76′′O and O<sub>p</sub> centres. This
suggests that if the phosphate is to act as the base there needs to be a bridging water molecule to facilitate the proton transfer process. Indeed, in the MD simulations of both complexes, it was observed that the non-bridging phosphate oxygens (O1_p and O2_p) of the substrate were typically engaged in hydrogen bonding interactions with one or more water molecules. For GlnRS···Gln-AMP/tRNA^{Gln} and ND-GluRS···Glu-AMP/tRNA^{Glu} the 3D-RISM solvent analysis tool, as implemented in MOE, was used to further characterize the water distribution (i.e. average positions) around the phosphate (Figure 4.3). Notably, there is a significant localization of the water molecules around the substrates O1_p and O2_p centres in both complexes. This highly hydrated environment suggests water may play a role as a mediator in catalysis.

Figure 4.2. Plots of the variation in some key distances in Ångström along 10 ns simulations: a) \( \text{Ado76}^2\text{-O}\cdots\text{O1}_p \) distance for wildtype (blue lines) and mutant (red lines) GlnRS; b) \( \text{Ado76}^2\text{-O}\cdots\text{C}_{\text{carb}} \) distance for wildtype (blue lines) and mutant (red lines) GlnRS; c) \( \text{Ado76}^2\text{-O}\cdots\text{O1}_p \) distance for ND-GlnRS; d) \( \text{Ado76}^2\text{-O}\cdots\text{C}_{\text{carb}} \) distance for ND-GluRS.
For both the wildtype GlnRS⋯Glu/tRNA\textsuperscript{Gln} and mutated Glu\textsuperscript{34}GlnGlnRS⋯Glu/tRNA\textsuperscript{Gln} complexes the mechanistically relevant \textit{Ado76}O\textsubscript{2′}⋯C\textsubscript{carb} distance was monitored over the course of the 10 ns MD simulation and are shown in Figure 4.2b. Importantly, the average distances are calculated to be reasonably similar at 3.98 Å and 3.81 Å, respectively. This suggests that they are positioned reasonably close to each other for the subsequent reaction.

For the ND-GlnRS⋯Glu/tRNA\textsuperscript{Gln} complex, analogous plots of the \textit{Ado76}O\textsubscript{2′}⋯O\textsubscript{1\textsuperscript{p}} (Figure 4.2c) and \textit{Ado76}O\textsubscript{2′}⋯C\textsubscript{carb} (Figure 4.2d) distances obtained over the course of the MD simulations give average values of 4.54 Å and 3.25 Å, respectively. Similar to the corresponding discriminating GluRS complexes (see above), while the average \textit{Ado76}O\textsubscript{2′}⋯O\textsubscript{1\textsuperscript{p}} distance may seem too large for a proton transfer to occur, water may be able to facilitate this process by acting as a bridge between the \textit{Ado76}O\textsubscript{2′}OH and O\textsubscript{1\textsuperscript{p}} centre (see Figure 4.3). Thus, both GluRS and ND-GlnRS may share a common substrate-assisted mechanism (c.f. Figure 4.2a and c).

Figure 4.3. The first solvation shell directly hydrogen bonded to the phosphate groups in the substrates; the water oxygen density is in blue and the water hydrogen density is in grey.
4.3.2 QM-Only and QM/MM calculation

4.3.2.1 Wildtype GlnRS Catalysed Aminoacylation with Glutamine

4.3.2.1.1 QM-Cluster: the substrate's phosphate acting as the base

If the substrate's phosphate group acts as the base there is the potential for one of its non-bridging oxygens (O1_p) or bridging (O_b) oxygen to act as the proton acceptor. Hence, using a QM-cluster approach we examined the feasibility of either to act as the required base in the aminoacylation process. It should be noted that given that the MD results suggest that the phosphate may play such a role via a water bridge, an appropriate model was used for these studies. The results obtained are given in Table 4.1.

As can be seen, the QM-cluster calculations (see Computational Methods) indicated a clear kinetic and, in general, thermodynamic preference for O1_p to act as the base through the assistance of a bridging water molecule, rather than the O_b atom. An energy barrier of 23.5 kcal/mol is obtained when the reaction is initiated by the O1_p atom compared to a significantly higher barrier of 60.8 kcal/mol in the case of the O_b atom. The phosphate leaving groups in the obtained intermediate complexes (IC) are in their mono-protonated forms, since there is no participation from the nearby lysine residues. Accordingly, in the subsequent hybrid ONIOM(QM/MM) investigations on the aminoacylation mechanism in which the phosphate is the base, only the O1_p atom was considered.

Table 4.1. Free energy values (in kcal/mol) obtained for the two different mechanisms considered in case of the two enzymes using QM-only approach.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>RC</th>
<th>TS-O1_p</th>
<th>IC-O1_p</th>
<th>TS-O_b</th>
<th>IC-O_b</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlnRS</td>
<td>0.0</td>
<td>25.6</td>
<td>26.2</td>
<td>59.7</td>
<td>14.0</td>
</tr>
<tr>
<td>ND-GluRS</td>
<td>0.0</td>
<td>25.0</td>
<td>42.0</td>
<td>51.1</td>
<td>−17.3</td>
</tr>
</tbody>
</table>
4.3.2.1.2 QM/MM: the substrate's phosphate acting as the base

The QM/MM optimized (see Computational Methods) structures of the stationary points (reaction, intermediate, product complexes, and transition structures) obtained for the GlnRS catalysed aminoacylation of tRNA\textsuperscript{Gln} by Gln in which one of the substrates non-bridging phosphate oxygens acts as the base are schematically shown in Figure 4.4. The corresponding potential energy surface obtained is shown in Figure 4.5.

In the reactive complex (GlnRC) moderately strong hydrogen bond interactions are observed between \textit{Ado}762′OH of the tRNA and a water (W1) molecule, \(r(\text{Ado}762′\text{OH} \cdots \text{O}_{W1}) = 1.82\ \text{Å}\), and between the same W1 and the substrate's non-bridging phosphate oxygen O1\textsubscript{p}, \(r(\text{W1H} \cdots \text{O}_{1\text{p}}) = 1.80\ \text{Å}\). The other non-bridging oxygen of the phosphate, O2\textsubscript{p}, is stabilized through hydrogen bonds with the side chains of two different residues, the cationic Lys270 residue and the nearby His43 residue at distances of 1.92 and 1.86 Å, respectively (Figure 4.4). Notably, this hydrogen bond network around O2\textsubscript{p} indicates that the O1\textsubscript{p} atom will be the more basic one and supports the hypothesis that it is the mechanistic base. Notably, there is no direct interaction between the Glu34 residue and the \textit{Ado}762′-OH group. Instead, the former is strongly hydrogen bonded to the nearby Lys72 residue at a distance of 1.67 Å. This strong salt bridge interaction suggests that Glu34 is also unlikely to act as the mechanistic base.

In addition, the tRNA adenosine terminal is concomitantly appropriately situated for nucleophilic attack upon the C\textsubscript{carb} centre of the Gln-AMP substrate with \(r(\text{Ado}762′\text{O} \cdots \text{C}_{\text{carb}}) = 3.10\ \text{Å}\). This geometry of the substrates in the GlnRC thus appears to support the feasibility of a water-mediated substrate-assisted mechanism.
Figure 4.4. Schematic representation of the obtained complexes along the studied aminoacylation mechanism together with selected key distances in Ångstrom; black color distances for the native enzymes, blue are for the Glu34Gln mutant.

Aminoacyl transfer proceeds through a concerted transition state, $^{\text{Gln}}$TS (Figure 4.4). More specifically, O1$_p$ has deprotonated the $^{\text{Ado}76}2'-\text{OH}$ group through the assistance of the bridging W1 molecule. Concomitantly, the $^{\text{Ado}76}2'-\text{oxygen}$ has approached the C$_{\text{carb}}$ centre of the Glu-AMP substrate such that the $^{\text{Ado}76}2'-\text{O} \cdots \text{C}_{\text{carb}}$ distance is now just 1.63 Å with concurrent elongation of the C$_{\text{carb}}$—O$_b$ bond from 1.34 to 1.52 Å. This is found to be a late transition state, since the proton of the W1 molecule is completely transferred to the non-
bridged oxygen $O_{1p}$ of the phosphate group, $r(H—O_{1p}) = 1.01 \text{ Å}$ (Figure 4.4). Concomitantly, a typical single bond has been generated between the former $\text{Ado}_7\text{H}^2'$-H proton and the oxygen of W1 ($O_{\text{w1}}$) with a bond length of 1.00 Å. The accumulated negative charge on the $O_{\text{carb}}$ atom is stabilized through hydrogen bond formation to the nearby backbone –NH– of the Glu34 residue, $r(N(H)\cdots O_{\text{carb}}) = 1.81 \text{ Å}$, which is 0.11 Å shorter than the corresponding distance in the reactive complex $\text{GlnRC}$. It is noted that the hydrogen bonding interactions between the Lys270 and His43 residues and $O_{2p}$ are maintained with distances of 1.72 and 2.00 Å, respectively. The energy barrier for $\text{GlnTS}$ is 25.0 kcal/mol relative to the corresponding $\text{GlnRC}$, indicating an enzymatically permissible step (Figure 4.5).

Collapse of $\text{GlnTS}$ results in formation of the intermediate complex $\text{GlnIC1}$ featuring a tetrahedral $C_{\text{carb}}$ (Figure 4.4). In this intermediate the $\text{Ado}_7\text{O}^2—C_{\text{carb}}$ single bond has been formed with a length of 1.53 Å, while the $C_{\text{carb}}\cdots O_{\text{PO4}}$ distance has further elongated to 1.57 Å. Similar to $\text{GlnRC}$, the developed negative charge on the $O_{\text{carb}}$ atom in $\text{GlnIC1}$ is stabilized through hydrogen bond interaction with the backbone amide –N(H)– group of the nearby Glu34 residue with a $\text{Glu34N(H)}\cdots O_{\text{carb}}$ distance of 1.81 Å. Thermodynamically $\text{GlnIC1}$ was calculated to be 4.0 kcal/mol higher in energy than the initial reactant complex $\text{GlnRC}$ (Figure 4.5).
Figure 4.5. The free energy surface for the aminoacylation mechanism of glutamine in the case of the wildtype GlnRS (black color surface) and the Glu34Gln mutant GlnRS (blue color (dashed) surface). Gln\textsuperscript{TS2} and Gln\textsuperscript{TS2}' were obtained through scan calculations.

It should be noted that in Gln\textsuperscript{IC1} the distance between the side chain amine of the Lys270 residue and the substrates O2\textsubscript{P} oxygen is shortened to 1.66 Å, which is 0.05 Å shorter than the corresponding distance in the preceding transition state (Gln\textsuperscript{TS}; Figure 4.4). This decrease in the distance better stabilizes the AMP group and facilitates the subsequent proton transfer step from the Lys270 residue to the adjacent O2\textsubscript{P} atom via Gln\textsuperscript{TS2} (Figure 4.4). As shown in the calculated potential energy surface this second proton transfer occurs without a barrier (−12.0 kcal/mol relative to the Gln\textsuperscript{RC}, Figure 4.4). That is to say, once formed Gln\textsuperscript{IC1} can react without a further barrier to give the final product complex, Gln\textsuperscript{PC}. In Gln\textsuperscript{TS2} the AMP moiety, cleaved from the initial Gln-AMP substrate, has moved 2.87 Å
away from the glutamine moiety, which has itself now been transferred onto its cognate tRNA$^{\text{Gln}}$ as indicated by the typical single-bond character of the $\text{Ado}^\text{162O}—\text{C}_{\text{carb}}$ bond which has length of 1.38 Å.

As noted, the product complex $\text{GlnPC}$, is obtained once the second proton transfer is complete. Notably, the O2$^\text{P}$—H bond has now formed with a length of 0.98 Å while the glutamine and AMP moieties now lay 3.15 Å apart, Figure 4.4. Importantly, the obtained product complex $\text{GlnPC}$ lies $-29.9$ kcal/mol lower in energy relative to $\text{GlnRC}$ and hence is thermodynamically favourable, Figure 4.5. Taken together, the obtained potential energy surface of this water mediated substrate-assisted mechanism implies a kinetically and thermodynamically feasible pathway.

4.3.2.1.3 QM/MM: the substrate's $\alpha$-$\text{NH}_2$ acting as the base

It has been suggested that the substrate-assisted aminoacylation mechanism in some aaRS enzymes exploit the aminoacyl substrate’s $\alpha$-$\text{NH}_2$ group as the base. This was initially proposed by our group during a study on the aminoacylation of ThrRS,$^{31}$ and has also recently been proposed for the case of leucyl-tRNA synthetases (LeuRS).$^{45}$ In ThrRS the $\alpha$-$\text{NH}_2$ group is activated (made neutral and thus able to act as a base) by the presence of a Zn(II) ion, while in LeuRS the protonated form of the substrate's $\alpha$-$\text{NH}_3^+$ group is deprotonated by a conserved aspartate residue.$^{45}$ Due to the presence of an aspartate residue (Asp66) proximal to the substrate $\alpha$-$\text{NH}_2$ in GlnRS, we expanded our study to explore the possibility of the latter group being the mechanistic base. The optimized structures obtained, with select bond distances and corresponding free energies, are summarised in Figure 4.6.
As can be seen, the relative energies of each of the stationary points (intermediate and product complexes, and transition states) along the pathway are significantly higher in energy compared to the values obtained for the previous case where O1\textsubscript{p} acted as the base. For example, the corresponding first intermediate, \( \text{GlnIC1}'' \), is 26.8 kcal/mol higher in energy than the initial reactive complex \( \text{GlnRC} \), which is the common complex to both possible pathways. Notably, the \( \alpha \)-NH\textsubscript{2} group no longer interacts with the Asp66 residue but instead forms a moderately strong hydrogen bond to the \( \text{Ado76}2'\text{OH} \) group with a distance of \( r(\text{N}\text{NH}_2\cdots\text{HO-2'}\text{Ado76}) = 1.74 \text{ Å} \). The subsequent step is transfer of a proton from the \( \text{Ado76}2'\text{OH} \) group to the \( \alpha\)-NH\textsubscript{2} moiety with concomitant formation of the \( \text{Ado76}2'\text{O}--\text{C}_{\text{carb}} \) bond. Now, however, this occurs via the transition state \( \text{GlnTS}'' \) at a cost of 57.8 kcal/mol relative to \( \text{GlnRC} \) (Figure 4.6). This is 32.8 kcal/mol more than for aminoacyl transfer via \( \text{GlnTS1} \) (c.f. Figure 4.5). Additionally, the product complex \( \text{GlnPC}'' \) is thermodynamically unfavourable lying 44.4 kcal/mol higher in energy relative to \( \text{GlnRC} \). Accordingly, these findings discount the possibility of the substrate’s \( \alpha\)-NH\textsubscript{2} group acting as the catalytic base in the GlnRS enzyme from both kinetic and thermodynamic perspectives.

**Figure 4.6.** Optimized structures for the aminoacylation mechanism in the wildtype GlnRS where the substrate’s \( \alpha\)-NH\textsubscript{2} acts as the mechanistic base, together with energy values (in kcal/mol, relative to the reactive complex) in parentheses. For clarity, the atoms that are involved in the reaction are highlighted.
4.3.2.2 Aminoacylation by the Glu34Gln mutant of GlnRS.

The results reported in the above sections indicate that the Glu34 residue is not directly involved in the aminoacylation mechanism. Instead, it participates in a salt bridge with the protonated side chain of a nearby lysyl residue, \( r(\text{Glu34COO}^\ldots\text{+H3N}_{\text{Lys72}}) = 1.67 \text{ Å} \). However, as noted in the Introduction, experimental studies on the mutation of Glu34 to glutamine observed a rate decrease in the enzyme but not inhibition. Thus, to better understand the role of the Glu34 residue in the mechanism \textit{in silico} mutagenesis was performed in which it was substituted by the isoelectronic and isostructural residue glutamine. Except for the Glu34Gln mutation all other active site residues are the same as in the wildtype. The potential energy surface obtained is shown in \textbf{Figure 4.5}.

Significant changes are observed in the reactive complex obtained for the Glu34Gln mutant, \textit{GlnRC}\textit{'}\textit{C}, compared to that obtained for the wildtype (\textit{GlnRC}). In particular, the mutation causes the substrate to shift position relative to the surrounding residues. For instance, unlike in the case of the wildtype enzyme, there no longer is a hydrogen bond between the \textit{Ado762'OH} group and the W1 molecule. The latter W1 does retain its hydrogen bond with a non-bridging oxygen of the phosphate group though it is slightly longer at \( r(\text{W1OH}\ldots\text{O1}_p) = 1.85 \text{ Å} \) (\textbf{Figure 4.4}). Moreover, the \( \text{O}_\text{carb}\ldots\text{N(H)}_{\text{Gln34}} \) hydrogen bond distance has increased by 0.35 Å to 2.28 Å and accordingly weakened relative to the wildtype enzyme. Similarly, the \( \text{O2}_p\ldots\text{HN}_{\text{His43}} \) hydrogen bond has now been extended by 0.14 Å to 2.00 Å. These two interactions are intended to stabilize the negative charges on the \( \text{O}_\text{carb} \) and \( \text{O2}_p \), respectively, and the longer distances indicate less charge stabilization in the Glu34Gln mutant. However, it is noted that the weakening of the \( \text{O2}_p\ldots\text{HN}_{\text{His43}} \)
hydrogen bond is at least partially compensated for by the large decrease of 0.33 Å in the Lys270 NH$_3^+$···O$_2^p$ hydrogen bond to 1.59 Å (Figure 4.4).

The aminoacylation pathway obtained for the Glu34Gln mutant was in general the same as the wildtype. For instance, the first step of the mechanism proceeds through the tetrahedral oxyanion transition state $^{Gln}TS1'$. Notably, for the latter no significant changes from the analogous wildtype $^{Gln}TS1$ (c.f. Figure 4.4) are observed in the interactions between the substrate and the surrounding residues. Thermodynamically, however, $^{Gln}TS1'$ is 39.8 kcal/mol higher in energy with respect to its corresponding reactive complex, Figure 4.4. That is, for the Glu34Gln mutant the barrier for the first step is 14.8 kcal/mol higher in the energy. Since there is a negligible difference in the geometries of $^{Gln}TS1$ and $^{Gln}TS1'$, this energetic difference may be due to the structural influences of the Glu34Gln mutation on the reactive complex $^{Gln}RC'$. More specifically, in the latter the mutation causes the $^\text{Ado76}2'\text{OH}$ to be positioned in a less productive orientation for the subsequent substrate-assisted mechanism and thus, larger structural changes are required for the reaction to proceed.

Collapse of the transition state $^{Gln}TS1'$ leads to the generation of intermediate complex $^{Gln}IC'$, which is 37.2 kcal/mol higher in energy than $^{Gln}RC'$, Figure 4.5. It is noted that this is 33.2 kcal/mol higher in relative energy to the corresponding initial reactant complex than $^{Gln}IC$. As shown in Figure 4.5 the wildtype and Glu34Gln intermediate complexes $^{Gln}IC$ and $^{Gln}IC'$ respectively, are geometrically analogous; the two complexes possessing the same interactions, with similar distances, between the ligand and active site residues, Figure 4.4. Indeed, the largest difference is just 0.04 Å in the His43NH···O$_2^p$ hydrogen bond.
Similar to the wildtype catalysed aminoacylation, the subsequent proton transfer from the side-chain protonated amine of Lys270 to the O$_2p$ atom occurred without a barrier via Gln$^\text{TS2'}$. As shown in Figure 4.4 there are no significant differences in the distances between Gln$^\text{TS2}$ and Gln$^\text{TS2'}$; although the latter is 21.7 kcal/mol higher in energy. In the resulting product complex formed, Gln$^\text{PC'}$, the AMP and glutamine are 3.09 Å apart from each other. The product Gln$^\text{PC'}$ is thermodynamically favourable by 9.3 kcal/mol compared to Gln$^\text{RC'}$. Notably, however, it is less stable than the corresponding wildtype product Gln$^\text{PC}$ by 20.6 kcal/mol.

4.3.2.3 ND-GluRS Catalyzed Misacylation of tRNA$^\text{Gln}$ by Glutamate

As noted in the Introduction, in most prokaryotes the aminoacylation of tRNA$^\text{Gln}$ is catalysed by the non-discriminating aaRS ND-GluRS. In particular, it binds and charges the non-cognate tRNA$^\text{Gln}$ with glutamate, yielding Glu-tRNA$^\text{Gln}$, which is then converted to Gln-tRNA$^\text{Gln}$ by a tRNA-dependent amidotransferase.\textsuperscript{18} To investigate the generality of this water-mediated substrate-assisted aminoacylation mechanism within this class of enzymes, we also investigated its feasibility in the ND-GluRS enzyme.

As for discriminating GlnRS initially, QM-only calculations were performed on the two possible substrate-dependent mechanisms in which either the phosphates O$_b$ or O$_1p$ oxygen centres act as the base, Table 4.1. Similar to the previous findings in the case of GlnRS, of these two when the O$_1p$ is acting as the required base leads to the most kinetically feasible mechanism with an activation energy of 24.2 kcal/mol, compared to 50.3 kcal/mol in the case of O$_b$ atom. It is worth noting that for the O$_b$ mechanism, the intermediate complex IC-O$_b$ is lower in energy than IC-O$_1p$ due to a proton transfer from the Lys270 residue to the O$_2p$ atom, resulting in a neutral phosphate leaving group.
Chapter 4

However, due to the high barrier in the case of the $O_b$ mechanism, all subsequent hybrid QM/MM calculations used $O1_p$ as the base.

Figure 4.7. Schematic drawing of the complexes obtained during the $O1_p$ substrate-assisted aminoacylation mechanism of ND-GluRS, together with selected key distances in Ångstrom.

The QM/MM optimized structure of the $GluRC$, derived from the previous MD simulations, is schematically shown in Figure 4.7. As for GluRS the Glu-AMP and the tRNA adenosine terminal ($Ado76'OH$) appear to be suitably positioned with respect to each other and the active site residues for the substrate-assisted mechanism to take place. For instance, again an active site water, W1, is positioned as a bridge between the 2$'OH_{Ado76}$ of
the tRNA and the O₁₉ of the Glu-AMP substrate with O(H)ₚ₋₋₋O₁₉ and W₁O₋₋₋Ado76²⁻(H)O distances of 1.71 Å and 1.65 Å, respectively (Figure 4.4).

Notably, the O₂₉ atom is more stabilized, and thus less basic than O₁₉, due to the hydrogen bonding network it forms with the backbone –N(H)– group of the nearby Ser32 residue (r(O₂₉⋅⋅⋅HN₃₉₂) = 2.16 Å) and two distinct water molecules with O₂₉⋅⋅⋅O₅₉ distances of 1.70 and 1.90 Å. The His220 residue, which will help stabilize the negative charge that develops on the substrate’s O_carb centre, directly hydrogen bonds to the latter with r(H₉₂₆₂₀⋅⋅⋅O_carb) = 2.18 Å. Meanwhile, the α-NH₃⁺ group of the Glu-AMP substrate is stabilized through salt-bridge formation with the nearby Asp64 residue at distance of 1.77 Å. In addition, the substrate’s carboxylate group is stabilized by hydrogen bonding to the guanidinium side chains of two arginine residues, Arg216 and Arg28. Specifically, Arg28 forms two moderately strong Arg28(NH)⋅⋅⋅O Glu hydrogen bonds at distances of 1.77 and 1.62 Å while other Arg216’s side chain forms a single hydrogen bond at 1.82 Å.

The orientation of the substrate in Glu-RC appears to favour a W₁-mediated deprotonation of the Ado76²OH group by O₁₉. This facilitates the first step of the aminoacylation mechanism; attack of the Ado76²OH on the substrate’s C_carb centre which occurs via Glu_TS1, Figure 4.7. It was noted that in Glu_TS1 the O₁₉ centre has effectively fully accepted a proton from the bridging water W1. It does, however, retain a hydrogen bond with the O₉₁ centre at a distance of 1.51 Å. Concomitantly, the Ado76²O⋅⋅⋅C_carb distance has significantly shortened to 1.54 Å while the O₉₁⋅⋅⋅C_carb bond has elongated slightly to 1.62 Å. Furthermore, the H₉₂₂₂₀⋅⋅⋅O_carb hydrogen bond distance has shortened by 0.39 Å to 1.89 Å, which is 0.2 Å shorter than the analogous interaction in GlnRS, Figure 4.4. Importantly, the energy required for this step, i.e., the energy of Glu_TS1 relative to
Glu\textsubscript{RC}, is calculated to be 25.4 kcal/mol, indicating an enzymatically feasible barrier (Figure 4.8).

The required energy for this step is 3.7 kcal/mol less than the obtained activation barrier for the same step in GlnRS, Figure 4.8. This may be due to the presence of an extra hydrogen bond in Glu\textsubscript{TS} between the transiently charged Ado\textsubscript{76}2′O and the nearby side chain hydroxyl of Ser32, $r(\text{Ado76}2′\text{O} \cdots \text{H}_{\text{Ser32}}) = 1.91$ Å, which would provide more charge stabilization to the accumulated negative charge on the Ado\textsubscript{76}2′-oxygen.

![Figure 4.8](image.png)

**Figure 4.8.** Free energy surface for the aminoacylation mechanism of glutamate by ND-GluRS.

This transition state leads to the formation of the intermediate Glu\textsubscript{IC}, in which the substrates glutamyl moiety is now bound to the tRNA adenosine terminal by a covalent bond with the Ado\textsubscript{76}2′-oxygen. Concomitantly, the covalent bond between the substrates
glutamyl and AMP components has been cleaved as indicated by a \( C_{\text{carb} \cdots O_b} \) distance of 2.93 Å, Figure 4.7. Notably, \( \text{Glu}^\text{IC} \) is almost thermoneutral with \( \text{Glu}^\text{RC} \), lying 0.9 kcal/mol lower in energy, Figure 4.8.

It should be noted that there are significant structural differences between \( \text{Glu}^\text{IC} \) and \( \text{Gln}^\text{IC} \). Most significantly, the latter is a tetrahedral intermediate, the Gln\cdots AMP (\( C_{\text{carb} \cdots O_b} \)) bond not yet being cleaved. Consequently, it lies higher in energy than the corresponding reactant complex, \( \text{Gln}^\text{RC} \). In contrast, \( \text{Glu}^\text{IC} \) more closely resembles the product complex wherein the Glu\cdots AMP (\( C_{\text{carb} \cdots O_b} \)) bond has been cleaved and the \( \text{Ado}^7 \text{O-Glu} \) covalent bond has formed. The thermodynamic stability of \( \text{Glu}^\text{IC} \) is also due in part to the extra charge stabilization provided to the \( O_2^p \) and \( O_b \) centres through the formation of strong hydrogen bonds to the nearby water and side chain hydroxyl of Ser32 at distances of 1.52 and 1.54 Å, respectively.

Unlike the intermediate complex obtained for GlnRS (\( \text{Gln}^\text{IC} \)) where a lysyl is directly hydrogen bonded to the ligand, in \( \text{Gln}^\text{IC} \) the lysyl residue (Lys257) is hydrogen bonded to a second active site water molecule W2, \( r_{(\text{Lys257}H \cdots O_W)} = 1.52 \) Å. Consequently, the subsequent proton transfer from Lys257 to a non-bridging oxygen of the AMP's phosphate to stabilise the charge on the leaving group takes place through W2 via \( \text{Glu}^\text{TS2} \), Figure 4.7. In this transition state, the transferring proton is approximately mid-way between the side chain amine nitrogen of lysyl257 (\( N_{\text{Lys257}} \)) and the W2 oxygen with distances of 1.28 Å and 1.23 Å, respectively. Meanwhile, the proton being concomitantly donated by W2 to the phosphate is 1.07 and 1.41 Å from the \( O_W \) and \( O_2^p \) centres, respectively, Figure 4.7. As was the case for GlnRS, at the present level of theory this second step of the mechanism is calculated to occur via \( \text{Glu}^\text{TS2} \) without a barrier, Figure 4.8.
Moreover, the $O_b$ atom is now involved in hydrogen bonding interactions with the $-N(H)-$ and side chain hydroxyl of Ser32 with distances of 1.86 and 1.62 Å, respectively. In the final product complex $\text{GluPC}$, Lys257 is now neutral having donated a proton via W2 onto the $O_2p$ atom, $r(W_2O\cdots H) = 1.51$ Å. Again, the overall aminoacylation process is predicted to be thermodynamically favorable as $\text{GluPC}$ lies 3.7 kcal/mol lower in energy than the initial reactant complex $\text{GluRC}$.

These results are in accord with our observations on GlnRS. The water-mediated, substrate-assisted mechanism is kinetically and thermodynamically feasible in case of the ND-GluRS. The first step, deprotonation of the $\text{Ado}^{2'}\text{OH}$ group by $O_1^p$ and concomitant nucleophilic attack upon the substrate's $C_{\text{carb}}$ is rate-limiting.

4.4 Conclusion

The aminoacylation mechanism of glutamine and glutamate amino acids to tRNA$_{\text{Gln}}$ by Glutaminyl- and ND-Glutamyl-tRNA synthetases was investigated computationally. Initial MD simulations indicated that the tRNA $\text{Ado}^{2'}\text{OH}$ terminus and the Gln-/Glu-AMP substrates are correctly orientated for a substrate-assisted mechanism with consistent $\text{Ado}^{2'}\text{O}\cdots C_{\text{carb}}$ and $\text{Ado}^{2'}\text{O}\cdots O_1^p$ distances over the course of the simulations. Moreover, our preliminary QM–only calculations supported the preference for the $O_1^p$ atom to act as the base, over the $O_b$ atom. Further investigations also supported the preference of this $O_1^p$ base over the substrate’s $\alpha$-NH$_2$.

The ONIOM calculations established that this mechanism proceeds through two steps. The first step is a water mediated substrate assisted deprotonation of the $\text{Ado}^{2'}\text{OH}$ group with concomitant formation of a $C_{\text{carb}}$–$O-2'\text{Ado}$ bond. This step is also the rate-limiting
with energy barriers of 25.0 and 25.4 kcal/mol in the cases of GlnRS and ND-GluRS, respectively.

For both GlnRS and ND-GluRS, a lysine residue was found to be positioned near the substrate such that it is able to neutralize the phosphate leaving group through barrierless proton transfers upon formation of a tetrahedral intermediate. The latter helps stabilize the anionic charge on the AMP leaving group and thus contributes to the overall thermodynamic favourability of the mechanism.

The Glu34Gln mutation of GlnRS leads to an increase in the activation energy of the mechanism to 39.8 kcal/mol. This is likely due at least in part to the substrates not being as favourably positioned for reaction, demonstrating at least an important structural role for the Glu34 residue, although no direct catalytic role was discovered. Importantly, these findings suggest the substrate-assisted mechanism elucidated for GlnRS and ND-GluRS may occur within other class I aminoacyl-tRNA synthetases.

4.5 References


Chapter 4

35. Molecular Operating Environment (MOE), *Chemical Computing Group Inc.*, 2013.08; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.08.


Chapter 5.
The Enzymatic Post-Transfer Editing Mechanism of *E. coli* Threonyl-tRNA Synthetase
5.1 Introduction

The remarkable accuracy in the translation of the genetic code of all organisms into proteins is closely tied to the outstanding specificity of the ubiquitous aminoacyl-tRNA synthetase (aaRS) family of enzymes.\textsuperscript{1-2} These enzymes are responsible for catalyzing the activation of their cognate amino acid via its reaction with ATP to form aminoacyl-AMP. Subsequently, and within the same synthetic (aminoacylation) active site they catalyze the transfer of the activated aminoacyl group onto its corresponding tRNA\textsuperscript{aa}, yielding the desired aminoacyl-tRNA\textsuperscript{aa}.\textsuperscript{3} Importantly, they are stated to be able to perform this overall process with an exemplary fidelity of $10^{-4}$, i.e., 1 aminoacylation error per 10000 reactions.\textsuperscript{2} Consequently, this crucial class of enzymes has been the subject of many experimental and theoretical studies seeking to gain deeper insight into their catalytic accuracy.\textsuperscript{2,4}

On the basis of size or chemical discrimination the active size pocket of an aaRS typically accommodates the corresponding cognate substrate amino acid while excluding non-cognate amino acids.\textsuperscript{5} However, due to electronic and/or structural similarities between some amino acids (e.g., threonine, serine, and valine), it is challenging for almost half of the aaRSs to correctly distinguish between cognate and non-cognate amino acids.\textsuperscript{6-7} Except for specific cases where moderate levels of mistranslation errors can be tolerated, or may even be necessary,\textsuperscript{8-9} any defects in the accuracy of translation can lead to protein misfolding, which is responsible for many fatal diseases.\textsuperscript{10}

In order to ensure accurate translation most aaRSs possess proofreading mechanisms. Specifically, in the synthetic site they may use pre-transfer editing mechanisms for which the most common is hydrolysis.\textsuperscript{11} For instance, a water molecule nucleophilically cleaves, the labile phosphate bond of a misactivated amino acid. This type of editing is observed in many enzymes; e.g. LeuRS hydrolyzes misactivated Val-AMP\textsuperscript{12} and ThrRS hydrolyzes
misactivated Ser-AMP. However, many aaRS may also exploit a distinct second editing active site – the editing site – which catalyzes post-transfer editing of misacylated tRNA. For example, AlaRS, ThrRS, and LeuRS enzymes utilize this type of correcting pathway to reject mischarged Gly-tRNAAla, Ser-tRNAThr, and Ile-tRNALeu, respectively. It has also been noted that some enzymes have a second, free-standing editing domain; for example, AlaRS has the AlaXp active site to correct any mislinked serine.

Our group has a long-standing interest in the mechanisms operating in aaRS enzymes, both regarding their synthetic activity and the pre-transfer editing functions of the synthetic site. The principal aim of this study is to expand the understanding of the post-transfer editing functions of threonyl-tRNA synthetase (ThrRS) enzymes. While it is thought that yeast mitochondrial ThrRS relies only upon pre-transfer editing, the ThrRSs found in all other species are suggested to also an editing site. This latter site is remote from the synthetic site and is responsible for the hydrolysis of misacylated Ser-tRNAThr. Structurally, the editing domain has been observed to be species-specific, with the two main types being bacterial and archaeal, and accordingly there are two different editing mechanisms.

In archaeal ThrRS, for example from Pyrococcus abyssi, the editing region is located in the N-terminal domain and the editing mechanism proceeds by a universal tRNA-mediated hydrolysis mechanism. In contrast, in bacterial ThrRS, for instance from Escherichia coli, the editing region has a quite distinct structure and different hydrolytic editing mechanisms have been proposed. Unfortunately, despite experimental studies on various possible editing pathways, the exact mechanism remains an open question.

Based on mutational and kinetic analyses it has been suggested that the binding domain’s His73, Lys156, His186 and Cys182 residues are the most mechanistically important. Notably, His73Ala mutation was observed to cause the most dramatic decrease in the rate of
hydrolytic editing, followed by Cys182Ala mutation.\textsuperscript{24,25} In addition, based on experimental X-ray crystal structures it was concluded that the His73 residue is within hydrogen bonding distance of a water molecule (W1) thought to be positioned near the substrate.\textsuperscript{23} Subsequently, the Schimmel and co-workers also concluded that the His73 and Lys156 residues are well-positioned to partake in catalysis.\textsuperscript{25} Consequently, it was proposed that, as shown in Scheme 5.1a the His73 residue is neutral and acts as the required mechanistic base. Specifically, it promotes activation of W1 through proton abstraction thus enabling the oxygen of W1 to nucleophilically attack the misacylated substrate's (Ser-tRNA\textsuperscript{Thr}) C\textsubscript{carb} centre. This ultimately results in cleavage of the labile C\textsubscript{Carb}—O3′ ester bond. The regenerated free Ado\textsubscript{76}3′O-tRNA oxyanion is neutralized \textit{via} a second proton transfer from the nearby Lys156 residue through a bridging water molecule, W2. The enzymatic use of a neutral histidyl as a mechanistically important base, such as in serine proteases and flavocytochrome b2, is widely documented.\textsuperscript{26-27}

\textbf{Scheme 5.1.} Schematic illustration of the two proposed pathways for post-transfer editing as catalysed by bacterial \textit{E. coli} ThrRS in which the mechanistic water nucleophile is activated by either (A) His73 or (B) Cys182 acting as the required base.
Recently, however, based on experimental oxidative stress studies on the editing domain of ThrRS, it has alternatively been suggested that the active site Cys182 residue may instead be the base, as shown in Scheme 5.1b.\(^{28-29}\) In particular, it was observed that treatment of *E. coli* ThrRS by H\(_2\)O\(_2\) resulted in formation of a Cys182-derived sulfenic acid (Cys\(_{182}\)SOH). For such a species to be formed the Cys182 must be in its thiolate form. Notably, this oxidative modification significantly increased misincorporation mistakes, and it was concluded that this was due to inhibition of the post-transfer editing of Ser-tRNA\(^{Thr}\) by ThrRS. Furthermore, and in contrast to the previously proposed mechanism above, it was suggested that of the three active site histidyl residues (His73, His77, and His186), His73 and His186 are both protonated and help activate the thiol of Cys182 and stabilize its thiolate form.\(^{28}\) Hence, an alternate post-transfer editing mechanism was proposed and is shown in Scheme 5.1b. These studies\(^{29}\) also noted that the editing site of ThrRS is similar to the active site of cysteine proteases, which are known to form sulfenic acids, and that oxidation of cysteine residues to sulfenic acids often occurs in ROS signaling proteins.\(^{30}\) Moreover, it has been noted that the editing sites of AlaRS and ThrRS share considerable sequence similarity and that their active sites contain the same conserved residues; namely, two histidyl residues, and a cysteinyl.\(^{37}\) Indeed, in the editing site of AlaRS the cysteinyl (Cys666) is in a nearly identical orientation to that of Cys182 in the editing site of ThrRS. Furthermore, mutation of Cys666 to alanine severely inhibited the ability of the editing site of AlaRS to deacetyl Ser-tRNA\(^{Ala}\). Hence, it has been suggested that these two aaRSs may share a common post-transfer editing pathway.\(^{31-33}\)

In the current study we have conducted a detailed, systematic multi-scale computational investigation using MD simulations and ONIOM(QM/MM) methodology to gain insight into the post-transfer editing mechanism in *E. coli* ThrRS. In particular, we have examined both
the previously proposed hydrolytic editing mechanisms of ThrRS in which either His73 or Cys182 acts as the mechanistic base, as well as possible alternate mechanisms and for a range of different potential active site protonation states.

5.2 Computational Methods

5.2.1 Molecular Dynamics Simulations

The high-resolution experimental X-ray crystal structure of the editing domain of threonyl-tRNA synthetase isolated from *E. coli* with the substrate analogue seryl-3'-aminoadenosine (SerAA) bound within the active site (PDB ID: 1TKY)\textsuperscript{24} was selected as the template structure for all chemical models used in this study. The bound analogue was mutated back to the desired substrate (Ser-AMP) by replacing the relevant bridging –NH– group in SerAA with an oxygen atom. Hydrogen atoms were added as appropriate and according to their protonation states as determined from the \( pK_a \) values calculated by both PROPKA 3.1\textsuperscript{34} and the default protonation tool in the Molecular Operating Environment (MOE) software.\textsuperscript{35} Unconstrained Molecular Mechanics (MM) energy minimization using the AMBER12 forcefield was then performed on the generated model. The model was then solvated by adding a 2 Å layer of water around the entire complex (total 725 water molecules) followed by a second MM energy minimization which terminated once the root mean square gradient fell below 0.01 kJ/mol·Å\(^2\). The resulting complex was initially submitted for 100 ps equilibration from 0 to 300 K at constant pressure using the default settings of MOE which includes cut-offs for long range interactions of 10 Å.

The mechanisms examined in this present study can be grouped into two broad categories; either His73- or Cys182-promoted editing. Hence, two template enzyme-substrate complexes were generated. For the former scenarios, i.e. those in which His73 acts
as the base, the complex contained neutral His73, His186, and Cys182. These protonation states are consistent with their predicted $pK_a$ values (see above). For the alternate scenarios in which Cys182 acts as the base, consistent with the experimentally suggested ionization states the enzyme-substrate complex instead contained protonated His73 and His186 residues while Cys182 was deprotonated. Under constant pressure and temperature, the generated structures were then subjected to a second simulation for 500 ps, under the same conditions and 2 fs time steps, using the NAMD engine. The final conformation obtained in each 500 ps simulation was then checked to ensure it was a suitable representative structure, and then MM minimized as above. Suitable QM/MM chemical models were then generated by truncating the resulting complex to include all residues and waters within 10 Å of the substrate.

5.2.2 QM/MM Investigations

A two-layer ONIOM(QM/MM)\textsuperscript{37-38} approach was used and consequently the above derived chemical models were each divided into QM- (high-level) and MM-subsystems (low-level) as shown in Figure 5.1. In particular, the substrate, mechanistically most relevant residues, as well as select active site water molecules were included in the QM-region. The surrounding protein environment, and remaining waters, were placed in the low-layer and described using an appropriate MM method. This methodology has previously been successfully applied to explore different catalytic mechanisms.\textsuperscript{39} All ONIOM(QM/MM) calculations were performed using the Gaussian09 software package.\textsuperscript{40}

For those chemical models used to examine mechanisms in which His73 acts as the base the QM region contained: the substrate (Ser-Ado); the charge stabilizing residues Tyr104 and Gly95; the proposed catalytic residues His73 and Lys156; and Asp180 due to its role in
orienting the substrate appropriately by formation of a salt bridge with the substrate’s α-NH₂. In addition, in the appropriate MD simulations an active site water (W1) molecule was observed to be hydrogen bonded to the His73 residue and concomitantly in close proximity to the C_carb of the substrate. Meanwhile, a second water (W2) molecule was hydrogen bonded to the side-chain amine of Lys156. Hence, both of these bridging water molecules (W1 and W2) were also included. The QM-region consisted of a total of 111 atoms and is shown in Figure 5.1a.

**Figure 5.1.** The template QM-region models used to study possible catalytic mechanisms of the ThrRS editing domain in which either (a) His73 or (b) Cys182 acts as the required base.

For the chemical models used to examine mechanisms in which Cys182 acts as the base the QM-region was expanded to include the Cys182 and His186 residues. These QM-regions contained a total of 121 atoms and are shown in Figure 5.1b. Moreover, for all chemical models the α-carbon atoms of the amino acids in the low layer (at least one residue away from the QM layer) were held fixed at their initial positions in order to ensure the structural integrity of the model.

To obtain optimized structures, each QM layer was treated using the hybrid DFT B3LYP method, widely applied in the study of catalytic and enzymatic mechanisms, in
conjunction with the 6-31G(d,p) basis set. In order to enhance the modelling of non-covalent and dispersion interactions, the empirical D3 dispersion correction by Grimme was applied.\(^{45}\) This has previously been shown to significantly improve the accuracy of kinetic barriers.\(^{46}\) Meanwhile, the AMBER96 force field was used for the MM region.\(^{47}\) Frequency calculations were also performed at this level of theory, ONIOM(B3LYP-D3/6-31G(d,p):AMBER96), to determine the nature of the stationary points and to calculate Gibbs free energy correction values (\(\Delta G_{\text{corr}}\)). Relative energies were obtained via single point energy calculations on the above structures at the ONIOM(B3LYP-D3/6-311+G(2df,p):AMBER96) level of theory with inclusion of the corresponding \(\Delta G_{\text{corr}}\). It is noted that all ONIOM calculations were performed within a mechanical embedding formalism.

### 5.3 Results and discussion

As noted above it has been suggested that either His73 or Cys182 acts as the base to facilitate and promote editing through activation of an active site water. However, this is complicated by the existence of several acid/base residues within the editing site. In addition, once a base has gained a proton it could, in some scenarios, potentially then act as an acid. Consequently, in this present study the mechanisms studied can be categorized as falling into either His73- or Cys182-promoted editing mechanisms. However, for each of these we have examined several possible variations that differ by, for example, the protonation states of other editing site residues. In total, 11 different possible mechanisms have been elucidated and compared in terms of their chemistry and thermochemical feasibility.
5.3.1 His73-promoted Editing Mechanisms

5.3.1.1 His73-promoted hydrolytic editing

The editing mechanism of archaeal ThrRS proceeds through a substrate-assisted mechanism in which the \( \text{Ado76}^2'\text{OH} \) group of the \( \text{tRNA}^\text{Thr} \) moiety facilitates hydrolysis of the \( \text{C}_{\text{carb}}-\text{O}_3'\text{Ado76} \) bond.\(^{23}\) In contrast, in the presently obtained optimized structure of the reactive complex \((\text{RC})\) of the editing site of bacterial ThrRS, shown in Figure 5.2, the substrate is positioned such that no hydrogen bonding interaction is observed between the substrate’s \( \text{Ado76}^2'\text{OH} \) group and the nucleophilic active site water (W1). Instead, W1 is in close proximity to the catalytic His73 residue with which it forms a moderately strong hydrogen bond, \( r(\text{His73}^\text{N}^\text{…H} \text{W1}) = 1.88 \text{ Å} \). In addition, W1 is positioned in good proximity to the substrate’s \( \text{C}_{\text{carb}} \) center with a distance of \( r(\text{W1}^\text{O}^\text{…C}_{\text{carb}}) = 3.17 \text{ Å} \), Figure 5.2. That is, W1 seems well positioned for the subsequently required proton transfer onto His73 and nucleophilic attack at \( \text{C}_{\text{carb}} \). The Lys156 residue, specifically its side chain ammonium, forms a strong hydrogen bond with a second water (W2), \( r(\text{Lys156}^\text{H}^\text{…O} \text{W2}) = 1.81 \text{ Å} \), which also positions it near to the substrate. Meanwhile, the Asp180 residue helps hold the substrate in more catalytically productive conformation through formation of an \( \text{Asp180}^\text{COO}^\text{…H}_3\text{N}^+ \text{SerAA} \) salt bridge, \( r = 1.64 \text{ Å} \). The substrate’s \( \text{O}_{\text{carb}} \) atom also forms a hydrogen bond with the backbone \(-\text{N(H)}-\) of the Gly95 residue at a distance of 1.89 Å. Optimized structures of stationary points (energy minima and transition states) along the mechanism, together with selected key distances, are shown schematically in Figure 5.2. The corresponding free energy surface (FES) for this mechanism is shown in Figure 5.3.
**Figure 5.2.** Schematic illustration of the optimized structures obtained (see Computational Methods), with select distances shown in Angstroms, for the His73 promoted mechanisms obtained with Lys156 protonating either the (1) \( \text{Ado}^{76}3'\text{O} \) or (2) \( \text{O}_{\text{carb}} \) centers of the Ser-tRNA\(^{Thr} \) substrate.

The reaction is initiated by the imidazole of His73 abstracting a proton from W1 while concomitantly, the water's \( \text{O}_{W1} \) center nucleophilically attacks the \( \text{C}_{\text{carb}} \) center of the substrate. This step proceeds via \( ^{1a}\text{TS1} \) at a cost of 27.6 kcal/mol relative to \( ^1\text{RC} \), suggesting an enzymatically high-energy barrier. In \( ^{1a}\text{TS1} \) the proton liberated from W1 has significantly
shifted toward the His73 residue as indicated by the distances $r(W_1 O^{-} H W_1) = 1.41 \, \text{Å}$ and $r(W_1 H^+ N_{H73}) = 1.16 \, \text{Å}$. Furthermore, the oxygen of the developing $W_1 OH^-$ moiety is now only $1.75 \, \text{Å}$ away from the substrate’s $C_{\text{carb}}$ center having formed a weak partial bond.

**Figure 5.3.** Free energy surfaces (kcal/mol) calculated (see Computational Methods) for the ThrRS editing domain mechanisms in which His73 acts as a base with Lys156 acting as an acid and protonating either the $O_{\text{carb}}$ center (blue line) or $\text{Ado76}^{3'}-O$ (red line).

Collapse of $^{1a}TS1$ results in generation of the energetically stable tetrahedral intermediate complex $^{1a}IC1$, which lies 21.6 kcal/mol higher in energy than $^{1RC}$, **Figure 5.3.** In this complex, the His73 residue is now protonated, $r(H_{73} N-H) = 1.06 \, \text{Å}$ while the partial $C_{\text{carb}}-O_{W1}$ single bond has shortened markedly to $1.54 \, \text{Å}$. Concomitantly, the $C_{\text{carb}}-O_{\text{carb}}$ and $C_{\text{carb}}-O3'_{\text{Ado76}}$ bonds have lengthened by 0.06 and 0.14 Å, respectively, **Figure 5.2.** These
latter increments are due to the change in the hybridization of the $C_{\text{carb}}$ toward pseudo $sp^3$-hybridization and stabilization of the increased negative charge on $O_{\text{carb}}$ by the single moderately strong hydrogen bond of length 1.76 Å it forms with the $-N(H)-$ of the Gly95.

The subsequent and final stage of the overall mechanism is cleavage of the $C_{\text{carb}}-O^{3'}_{\text{Ado76}}$ bond and formation of the neutral $O_{\text{Ado76}}^{3'}\text{H}$ group. An examination of $^{1a}\text{IC1}$ suggests that there are potentially three process by which this could occur. Namely, it may either occur by proton transfer from the side chain ammonium of Lys156 through the active site water W2 directly onto $O_{\text{Ado76}}^{3'}$ or indirectly via $O_{\text{carb}}$. Alternatively, the now protonated imidazole of His73 may act as an acid and directly protonate $O_{\text{Ado76}}^{3'}$-oxygen. Thus, each of these mechanisms were examined with the two most feasible being shown in Figures 5.2 and 5.3.

5.3.1.1.1 Direct water mediated protonation of $O_{\text{Ado76}}^{3'}$ by Lys156

The side chain ammonium of Lys156 is able to essentially directly protonate the $O_{\text{Ado76}}^{3'}$ center of $^{1a}\text{IC1}$ via a mediating water W2. This step proceeds via $^{1a}\text{TS2}$ with a free energy barrier of 33.4 kcal/mol, Figure 5.3. In this possible overall mechanism, the proton transfer represents the rate-limiting step. In addition, the high barrier suggests that the mechanism is unlikely to be enzymatically feasible. It is noted that $^{1a}\text{TS2}$ appears to be a dissociative transition state since the proton has essentially been transferred onto the $O_{\text{Ado76}}^{3'}$-oxygen while the $O_{\text{Ado76}}^{3'}\cdots C_{\text{carb}}$ bond has concomitantly significantly elongated to 2.16 Å, Figure 5.2. The resultant product complex $\text{I}^\text{PC}$ lies 4.3 kcal/mol higher in energy than $\text{I}^\text{RC}$; that is, this mechanism is also thermodynamically unfavorable. In $\text{I}^\text{PC}$ the Lys156 side chain amine is now neutral while the neutral $O_{\text{Ado76}}^{3'}\text{OH}$ group has been fully formed. In addition, the $O_{\text{Ado76}}^{3'}\cdots \text{H}$ distance is that expected of a typical $\text{O}--\text{H}$ single bond with length 0.96 Å, while
the carbC−3′O_{Ado76} ester bond has been fully cleaved as indicted by its distance of 2.8 Å, Figure 5.3

5.3.1.1.2 Indirect water-mediated protonation of Ado76-3′O by Lys156

Alternatively, Lys156 may first transfer its ammonium proton via W2 onto the oxyanionic O_{carb} center in IaIC1. Subsequently, an intramolecular proton transfer can occur from -O_{carb}H onto the 3′O_{Ado76} center of the leaving tRNA moiety, Figure 5.2. The first step in this process proceeds via IaTS2 at a cost of 32.6 kcal/mol relative to I1RC, Figure 5.3. It is noted that while IaTS2 lies 0.8 kcal/mol lower in energy than IaTS1, it is still too high to likely be enzymatically feasible. Furthermore, structurally, W2 has substantially transferred its proton onto the substrates O_{carb} center as indicated by the relevant distances of r(W2H⋯O_{W2}) = 1.37 Å and r(W2H⋯O_{carb}) = 1.11 Å, Figure 5.2. The resultant subsequent tetrahedral 1,1-diol intermediate complex (IaIC2) lies 1.1 kcal/mol higher in energy than IaIC, being 22.7 kcal/mol higher in energy relative to I1RC, Figure 5.3.

The subsequent proton transfer from the newly formed -O_{carb}H group onto the Ado76-3′-oxygen is found to take place via the 4-membered-ring transition state ITS3, Figure 5.2. As noted earlier, the Ado76-2′OH group in the optimized chemical model herein is oriented in such a way that it is unable to readily participate in this step to help form a 6-membered-ring transition state. The energy required for this second step is extremely high with a barrier of 69.6 kcal/mol relative to I1RC, Figure 5.3, clearly indicating that editing via this possible mechanism is enzymatically unfeasible. It is noted that even if an alternative pathway for this second proton transfer reaction was possible, the barrier for proton transfer via IaTS2 in itself is higher than likely to be enzymatically feasible. The final product complex formed,
\( ^1\text{PC} \), is the same as that obtained for the above described direct water meditated protonation mechanism. Hence, this possible alternate mechanism is also thermodynamically unfavorable.

5.3.1.1.3 Protonation of \( \text{Ado}^{76}3'\text{O} \) by His73-H^+.

The His73 residue has also been proposed to act as an acid-base catalyst and thereby the proton transferred to the \( \text{Ado}^{76}3'\text{-oxygen} \), cleaving the \( C_{\text{carb}}—O3'_{\text{Ado}^{76}} \) bond, may possibly originate from the protonated His73 residue formed in the first step of the mechanism, Figure 5.2.\textsuperscript{25} However, the relative energy of the required transition state is 45.8 kcal/mol. Moreover, the product complex generated from this step is 24.1 kcal/mol higher in energy than \( ^1\text{RC} \) and is thus significantly more thermodynamically unfavorled than \( ^1\text{PC} \).

5.3.1.2 His73-promoted editing through self-cyclization/lactone formation:

In the reactive complex \( ^1\text{RC} \) it was observed that the substrate's serinyl side-chain, and in particular its \( \beta\)-OH group, could potentially undergo a conformational change. In particular, it could rotate such that it no longer interacts with the side-chain carboxylate of Asp180 as is common,\textsuperscript{5, 22} but instead could potentially interact with His73. Such a geometry and interaction may facilitate an alternate editing pathway in which cleavage of the aminoacyl-tRNA's ester bond proceeds through formation of a 4-membered lactone ring, Figure 5.4.
Figure 5.4 Schematic representation of the optimized complexes, with select distances shown in Angstroms, obtained for a His73-promoted self-cyclization editing mechanism. The relative free energies (kcal/mol) are also given in brackets.

Similar cyclizations have been suggested to play a role in editing of other misacylated substrate's including the pre-transfer editing mechanism of MetRS previously reported by our group.18

A scan of the potential energy surface for rotation around the substrate's serinyl side chain C–C bond leads to formation of the intermediate complex $^{1b}IC_1$ lying just 5.3 kcal/mol higher in energy than $^{1RC}$. It is noted that the scan was performed as the energy surface was very flat; that is, the rotational barrier was quite small. Indeed, after empirical free energy...
corrections were included the barrier estimated to be slightly lower than 0.0 kcal/mol, reinforcing that in vivo this rotation can occur essentially without a barrier. In $^{118}$IC1 the substrate’s seriny1 β-OH is strongly hydrogen bonded to the neutral imidazole of His73 whilst also simultaneously in close proximity to the substrate's $C_{\text{carb}}$ center with distances of $r(H73N \cdots H\text{OH}) = 1.76 \text{ Å}$ and $r(OH^- \cdots C_{\text{carb}}) = 2.79 \text{ Å}$.

The oxygen of the β-OH is then able to nucleophilically attack the $C_{\text{carb}}$ center with concomitant transfer of the hydroxyl's proton onto His73. This step proceeds through $^{118}$TS2 at a quite high cost of 31.4 kcal/mol relative to the initial reactant complex $^1$RC. This high barrier may again be due in part to the increasing anionic charge on the $O_{\text{carb}}$ center during this step being stabilized by only its single hydrogen bond with the backbone -NH- of Gly95. The resulting tetrahedral intermediate $^{118}$IC2 lies slightly lower in free energy than $^1$RC by $-3.4 \text{ kcal/mol}$. Importantly, it must be noted that intermediate ligand has shifted position slightly so that the $O_{\text{carb}}$ oxyanion center is now hydrogen bonded to the protonated imidazole of His73 (i.e., His73-H$^+$). This suggests that the latter residue may be able to transfer its proton onto the $O_{\text{carb}}$ oxyanion center in order to generate a neutral lactol. And indeed, such a proton transfer can occur essentially without a barrier via $^{118}$TS3, as indicated by the fact that its calculated free energy after empirical corrections is lower than that of $^{118}$IC2. Importantly, this results in formation of the low energy lactol $^{118}$IC3 which lies markedly lower in energy than $^1$RC by 42.5 kcal/mol and is in fact the lowest energy complex along this pathway.

Several possible reactions (not shown) were considered for the subsequent required cleavage of the $C_{\text{carb}}-3'\text{O}_{\text{Ado76}}$ bond. The lowest energy pathway was found to involve an intramolecular proton transfer from the newly formed $-O_{\text{carb}}$H group onto the $\text{Ado76-3'}$-oxygen.
This step occurs via the four-membered ring transition state $^{1b}TS_4$ with a barrier of 33.3 kcal/mol relative to $^{1b}IC_3$. This is the rate-limiting step of the reaction and its height suggests that at least in the editing domain of ThrRS this pathway is enzymatically unfeasible. The final product complex formed, $^{1b}PC$, while thermodynamically favoured compared to the initial reactive complex $^1RC$ by 13.3 kcal/mol, is much higher in energy than $^{1b}IC_3$ by 29.2 kcal/mol, Figure 5.3.

5.3.2. Cys182-promoted editing Mechanisms:

As noted above, it has also been suggested that the active site cysteiny| Cys182 may instead be able to act as a base to promote/catalyse the editing mechanism. Thus we systematically examined possible mechanisms through which this may occur, with a key difference between them being the protonation states of the conserved active site His73, Cys182 and His186 residues. However, as for the above mechanisms, in some cases several scenarios were considered based on for instance the source of the proton in the second stage and/or the site of substrate to which it is initially transferred. It is noted that as for the above His73-promoted editing mechanisms, the side chain of Lys156 was considered to be in its protonated form.

5.3.2.1. Deprotonated Cys182 with protonated His73 and His186.

In the QM/MM optimized structure (see Computational Methods) of the reactant complex $^{1a}RC$, the anionic thiolate is stabilized by hydrogen bonds with the protonated imidazoles of His73 and His186 with distances of $r(\text{HisN} \cdots \text{Cys}) = 2.26 \text{ Å}$ and $1.90 \text{ Å}$, respectively, Figure 5.5. In addition, it forms a moderately strong hydrogen bond with the
nucleophilic water W1 with \( r(W_1 \cdot H^- \cdot S_{Cys182}) = 1.89 \, \text{Å} \). Similar to that observed in \(^{1a}\text{RC}\) the Asp180 anchors the substrate via salt bridge formation with the substrate’s \( \alpha\)-NH\(^3\) group at a distance of 1.51 Å. Meanwhile, the W2 molecule is positioned close to the other side of the substrate through hydrogen bonding to the backbone carbonyl oxygen of the nearby Ile94 residue with \( r(W_2 \cdot H^- \cdot O_{\text{Ile94}}) = 1.65 \, \text{Å} \), Figure 5.5.

Deprotonation of W1 by the Cys182 thiolate facilitates nucleophilic attack of the oxygen of W1 on the substrate’s C\(_{\text{carb}}\) center. This reaction proceeds through the transition state \(^{1a}\text{TS1}\) at a cost of 17.4 kcal/mol relative to \(^{1a}\text{RC}\), which is 10.2 kcal/mol less than the corresponding step in the above described His73-promoted mechanisms, Figures 5.3 and 5.6. It is noted that it represents a late transition state as the W1 proton is substantially shifted toward the sulfur atom and the W1 oxygen has moved quite close to the C\(_{\text{carb}}\) as indicated by their distances of \( r(W_1 \cdot H^- \cdot S_{Cys182}) = 1.49 \, \text{Å} \) and \( r(W_1 \cdot O^- \cdot C_{\text{carb}}) = 1.73 \, \text{Å} \), Figure 5.5. Similar to \(^1\text{TS1}\) in the His73-promoted mechanisms, the substrate’s O\(_{\text{carb}}\) center is stabilized by formation of a strong hydrogen bond with the backbone of Gly95, \( r(\text{Gly95-N\( \cdot \cdot \cdot \)O_{\text{carb}}}) = 1.78 \, \text{Å} \) which, due to the increased negative charge on O\(_{\text{carb}}\), is 0.26 Å shorter than the same interaction in \(^{1a}\text{RC}\), Figure 5.5. In addition, however, it also forms a strong hydrogen bond with a nearby active site water, \( r(O_{\text{carb}} \cdot H^- \cdot W) = 1.81 \, \text{Å} \), which is simultaneously hydrogen bonded to the protonated imidazole of His73. The latter is made possible due to the Cys182 now being neutral and remains hydrogen bonded to His186-H\(^{\text{\+}}\). This additional stabilizing hydrogen bond of O\(_{\text{carb}}\) may help explain the markedly lower reaction barrier obtained for this step.
Figure 5.5. Schematic illustration of the optimized structures obtained, with select bond distances shown in Angstroms, for the mechanism in which Cys182 acts as a base to activate an active water, and with Lys156 acting as acid and directly protonating the Ado763′-oxygen.

Collapse of IIaTS1 in leads to the formation of intermediate complex IIaIC1, which possesses typical \( \text{carb} \text{− OH} \) and \( \text{RS} \text{− H} \) bonds \( (r = 1.49 \text{ Å} \) and \( 1.36 \text{ Å} \), respectively). Notably, IIaIC1 lies 13.7 kcal/mol higher in energy than IIaRC, which is markedly lower than the relative energy (21.6 kcal/mol) than the corresponding intermediate IIC1 in the His73-promoted mechanisms, Figure 5.3.

As for the above mechanisms in which His73 acted as the base, there are three possible proton donors (Lys156, His73, and Cys182) present for the continuation of the reaction; cleavage of the \( \text{Ado763′O − C}_{\text{carb}} \) bond. Thus, each of the possibilities was then investigated, the most feasible mechanism obtained for each is discussed herein.
5.3.2.1.1. Direct protonation of $\text{Ado}763'O$ by Lys156

In $\text{IIaIC1}$ the $\text{Ado}763'O$ center is strongly hydrogen bonded to a proton of W2 at a distance of 1.79 Å. Notably, W2 is simultaneously also positioned near the side chain ammonium of Lys156. For the above His73 catalyzed process, proton transfer from Lys156 onto the $\text{Ado}763'$-oxygen gave the lowest barriers. For this present scenario it is found that such proton transfer from Lys156 onto the $\text{Ado}763'O$ center can occur with the involvement of the bridging water molecule, but with a high free energy barrier, via $\text{IIaTS2}1$ of 32.5 kcal/mol relative to $\text{IIaRC}$ (Figure 5.6). This step also appears to share structural similarities to that described above in the analogous His73-promoted mechanism that proceeds via $\text{IaTS2}1$. Namely, $\text{IIaTS2}1$ also appears to be a dissociative transition state as the $\text{H} \cdots \text{N}^+_{\text{Lys156}}$ distance has elongated significantly to 1.85 Å (i.e., has been transferred onto W2) while the carbC–O3'$\text{Ado}76$ bond has also been essentially wholly cleaved as indicated by its distance of 1.71 Å, Figure 5.5.

Collapse of $\text{IIaTS2}1$ leads directly to formation of the product complex $\text{IIaPC1}$ in which the tRNA's Ado76 ribose sugar and the noncognate serine amino acid are separated by a distance of 2.84 Å. The $\text{Ado}763'O$–H single bond has been formed with a length of 0.99 Å, Figure 5.4. Notably, this overall pathway was found to be thermodynamically favorable as $\text{IIaPC1}$ is 33.6 kcal/mol lower in energy than $\text{IIaRC}$, Figure 5.6.

5.3.2.1.2. Protonation of $\text{Ado}763'O$ involving His73-$\text{H}^+$.

In $\text{IIaIC}$ it was observed that with neutralization of the Cys182 thiolate group, the protonated imidazole of His73 (His73-$\text{H}^+$) has switched position and now is hydrogen bonded to a third editing site water molecule (W3). Thus, His73-$\text{H}^+$ may in fact be able to act as an acid and transfer its proton on the intermediate. It was found that in fact His73-$\text{H}^+$
is able to effectively transfer its proton via W3 onto the oxa
cyanion O_{carb} center of $^{IIa}{\text{IC}}$. Furthermore, this effectively occurs without a barrier via $^{IIa}{\text{TS}^2}$ as suggested by its slightly lower free energy relative to $^{IIa}{\text{IC}}$ (Figure 6). This low barrier also reflects the fact that $^{IIa}{\text{TS}^2}$ occurs early along this step; the His73N—H distance has lengthened only slightly to 1.15 Å and the distance to the O_{W3} center is still quite long at 1.36 Å. Meanwhile, the W3O$^-$—H distance for the proton W3 concurrently donates to the substrate’s O_{carb} center has increased significantly to 1.21 Å and the corresponding HW$^-$O_{carb} distance is relatively short at 1.25 Å, Figure 5.7.

**Figure 5.6.** FES obtained (kcal/mol) for the three possible editing mechanisms elucidated in which Cys182 acts as a base and His73 and His186 are protonated and the required acid for cleavage of the Ado763’O—C_{carb} bond is: Lys156 (black line and labels); His73 (red line and labels); or Cys182 (blue line and labels).
Indeed, the subsequently formed 1,1 diol intermediate complex $^{11}\text{IC}2$, which contains a now tetrahedral sp$^3$ hybridized $\text{C}_{\text{carb}}$ center, lies 11.0 kcal/mol lower in energy than the initial reactive complex $^{11}\text{aRC}$, Figure 5.6. The next and final step is transfer of the proton from the newly formed $\text{C}_{\text{carb}}$–OH groups onto the $3'\text{O}_{\text{Ado76}}$ center. The lowest energy pathway for such a transfer was found to occur involve the –OH group formed from W1 and proceeds via the 6-membered ring transition state $^{11}\text{aTS}3$. However, energetically this step has a high-energy barrier of 33.3 kcal/mol with respect to $^{11}\text{IC}2$ and is in fact the overall rate-limiting step for this possible pathway, Figure 5.5 and Figure 5.6. This is likely at least partly a reflection of the high pKa (i.e. low acidity) of the $\text{Ado76}3'\text{OH}$ group which is integrally involved in this step (Figure 5.7). However, the resulting product complex ($^{11}\text{aPC}2$) formed, in which the $\text{Ado76}3'\text{O}–\text{C}_{\text{carb}}$ bond is fully broken, is thermodynamically very favorable having an energy relative to $^{11}\text{aRC}$ of −50.0 kcal/mol (Figure 6).

**Figure 5.7.** Schematic illustration of the optimized structures, with selected bond lengths shown in Angstroms, obtained for the Ila editing mechanism (i.e., initial active site contains
Cys182SH–/His73-H+/His186-H+) in which cleavage of the Ado763′O–C_carb bond involves proton transfer from His73 (2: red labels) or Cys182 (3: blue labels).

5.3.2.1.3. Protonation of Ado763′O involving Cys182SH.

Alternatively, the neutral thiol of Cys182 formed during the first step could potentially act as an acid to facilitate cleavage of the Ado763′O–C_carb bond, Figure 5.7. More specifically, Cys182SH is able to transfer its proton onto the nearby Ado762′OH moiety while the latter concomitantly transfers its proton onto the Ado763′-oxygen, resulting in cleavage of the Ado763′O–C_carb bond. However, this step proceeds through IIa_TS2 with an energy barrier of 27.9 kcal/mol relative to IIa_RC (14.2 kcal/mol relative to IIa_IC1). Again this is the overall rate-limiting step for this possible pathway. But, though it remains high, it is the lowest barrier amongst the three possibilities elucidated for when the editing site initially contains a deprotonated Cys182, and protonated His73 and His186, Figure 5.6. The resulting product complex IIa_PC3 is the thermodynamically most favored of these three related pathways with an energy 51.0 kcal/mol lower than that of IIa_RC.

5.3.2.2. Deprotonated Cys182 with neutral His73 and His186 (Cys182S–/His73/His186).

It is noted that all related AlaXps enzymes there are three residues (His9, His13 and Cys116) that are positioned similarly to those of His73, His186 and Cys182 in the editing site of ThrRS. In addition, the Zn(II) binding-site in the aminoacylation site of ThrRS contains a similarly arranged deprotonated cysteiny1 and two neutral histidyl residues. Furthermore, the pKₐs predicted using PROPKA for these three residues in the editing site of ThrRS suggested that Cys182 may be deprotonated with His73 and His186 being neutral. Hence, we considered mechanisms in which these where their initial protonation states in the
editing site with the optimized structures and thermochemistry of the two possible pathways elucidated given in Figures 5.8 and 5.9.

Figure 5.8. Schematic illustration of the optimized structures (with select bond lengths shown in Angstroms) obtained for the editing pathways in which Cys182 acts as the base in the presence of neutral His73 and His186, and where either Lys156 (1: black labels) or Cys182 (2: red labels) acts as the required acid for the second proton transfer.
In the QM/MM optimized structure of the corresponding reactive complex $^{11b}RC$, the thiolate of Cys182 forms a hydrogen bond with an active site water molecule (W1) with $r(\text{Cys182}S^-\cdot\text{H}_{\text{W1}}) = 2.18$ Å (Figure 5.8). Notably, this orients W1 such that its oxygen (O$_{\text{W1}}$) is positioned just 2.64 Å from the substrate's C$_{\text{carb}}$ center, a distance which is significantly shorter by 1.40 Å than observed in $^{11a}RC$. This suggests that W1 is well placed for the subsequent required nucleophilic addition, Figure 5.8. This seemingly better positioning may also be due to the fact that unlike in $^{11a}RC$, the thiolate of Cys182 does not form a hydrogen bonding network with His73 and His186 due to their being neutral. Instead, and in addition to its hydrogen bond with W1, Cys182S$^-$ also forms a moderately strong hydrogen bond of length 2.16 Å directly with the Ado76$^{'2'}$OH group (Figure 5.8). This appears to help bind and orient the substrate in a more conducive conformation for subsequent reaction.

The analogous pathway for activation and nucleophilic attack of W1 on the substrate, as previously obtained for $^{11a}RC$ was elucidated. Again, the thiolate of Cys182 is able to abstract a proton from W1 while the oxygen (O$_{\text{W1}}$) of the latter simultaneously nucleophilically attacks the substrate's C$_{\text{carb}}$ center, Figure 5.8. In contrast, however, this reaction proceeds via $^{11b}TS1$ at an cost of just 9.9 kcal/mol, 8 kcal/mol lower than for the analogous reaction via $^{11a}TS1$. (Figure 5.9). Indeed, of all the systems considered herein, this is the lowest barrier for activation and subsequent nucleophilic attack by the active site water.

It is noted that in $^{11b}TS1$, the Cys182S$^-$·HO$^{2'}_{\text{Ado76}}$ hydrogen bond has been maintained though it has lengthened slightly to 2.28 Å. Meanwhile, and the developing negative charge on the O$_{\text{carb}}$ center is again stabilized through its quite strong hydrogen bond with the Gly95N·H group; $r(\text{Gly95N(H)}\cdot\cdot\text{O}_{\text{carb}}) = 1.67$ Å. This latter distance is 0.12 Å shorter, and thus stronger, than observed in the $^{11a}TS1$. As before, this step results in formation of an oxyanionic tetrahedral intermediate complex ($^{11b}IC$), in which the $r(\text{Gly95N(H)}\cdot\cdot\text{O}_{\text{carb}})$
interaction has shortened markedly to 1.58 Å (Figure 5.8). Significantly, as shown in Figure 5.9, \( {\text{IIbIC}} \) is just 4.4 kcal/mol higher in energy than the initial reactive complex \( {\text{IIbRC}} \). Hence, in terms of relative free energy, it is the lowest in energy of all such intermediates \( {\text{IIC}}, {\text{IIaIC}}, \) and \( {\text{IIbIC}} \) obtained herein.

As for the other systems examined, the subsequent cleavage of the \( {\text{Ado763'O}}{-}{\text{C}}_{\text{carb}} \) bond requires an acid to ultimately transfer its proton, directly or indirectly, onto the leaving \( {\text{tRNA}}_{\text{Thr}}{\text{Ado763'-oxygen}} \). In this present scenario, the most probable proton sources are the ammonium of Lys156 or the now neutral of Cys182. Hence, both possible pathways were considered.

5.3.2.2.1. Protonation of \( {\text{Ado763'O}} \) involving Lys156.

In both \( {\text{IIbRC}} \) and, in particular, \( {\text{IIbIC}} \), a second active water (W2) is positioned near both the ammonium of Lys156 and the \( {\text{Ado763'-oxygen}} \). Indeed, in \( {\text{IIbIC}} \) the water W2 simultaneously forms strong hydrogens bonds with both of these groups; \( r({\text{Ado763'O}}{\cdots}{\text{HOW2}} = 1.78 \, \text{Å}) \) and \( r({\text{Lys156NH}}{\cdots}{\text{O}}_{\text{W2}} = 1.70 \, \text{Å}) \), Figure 5.8. Water-mediated transfer of a proton from the ammonium of Lys156 onto the \( {\text{Ado763'-oxygen}} \) is able to proceed through \( {\text{IIbTS2}} \) at an enzymatically feasible cost of 20.8 kcal/mol relative to the initial reactive complex \( {\text{IIbRC}} \), Figure 5.9. As for all other mechanisms elucidated this barrier is the rate-limiting step of the overall mechanism. Importantly, however, it is the lowest barrier obtained for this step of all systems considered in this investigation.
Figure 5.9. FES obtained (kcal/mol) for the two possible editing mechanisms elucidated in which Cys182 acts as a base and His73 and His186 are neutral and the required acid for cleavage of the $\text{Ado}^3\text{O} - \text{C}_{\text{carb}}$ bond is: Lys156 (blue line and labels); or Cys182 (red line and labels).

In the product complex, $\text{I}^\text{mPC}^1$, the noncognate serine and tRNA$^{\text{Thr}}$ formed are now significantly separated as illustrated by $r(\text{C}_{\text{carb}} - \text{O}^{3'}\text{Ado}^76 = 3.19 \text{ Å})$. Notably this product complex was also thermodynamically the most favorable of all obtained being 53.6 kcal/mol lower in energy relative to $\text{I}^\text{mRC}$, Figure 5.9. Collectively, the free energy values obtained for this pathway, both kinetically and thermodynamically, support the likelihood of the editing mechanism at least involving Cys182 as a water activating base.
5.3.2.2. Protonation of $\text{Ado}763'O$ involving Cys182.

The close proximity of the now neutral thiol of Cys182 to the $\text{Ado}762'O$ group suggests the possibility that Cys182 may be able to acts as the acid, with the assistance of the $\text{Ado}762'O$ group to protonated its adjacent $\text{Ado}763'O$ center, Figure 5.8. However, such a process occurs via $^{\text{IIb}TS2}2$ and requires an energy of 26.7 kcal/mol relative to $^{\text{IIb}RC}$, which is higher than obtained above for when Lys156 acted as the required acid. Furthermore, the resulting product complex $^{\text{IIbPC}}2$ is less favoured than $^{\text{IIbPC}}1$ being only 16.7 kcal/mol lower in energy than $^{\text{IIbRC}}$. Thus, the Lys156 residue is kinetically and thermodynamically preferred as the required acid.

5.3.2.3. Activation of Cys182 by His73, in the presence of a protonated or neutral His186.

It has been suggested that at least part of the role His73 may be to activate, deprotonate, the thiol of Cys182. In this scenario both Cys182 and His73 would necessarily initially be neutral, while His186 maybe either protonated or neutral. For completeness, both possible scenarios were examined; that is, we examined possible mechanisms which being with activation of Cys182. For the case in which initially both His73 and Cys182 are neutral but His186 is protonated the optimized structure of the pre-reactive complex lies 10.3 kcal/mol lower in energy than $^{\text{IIaRC}}$; the reactive complex formed after activation. Furthermore, proton transfer from the thiol of Cys182 onto His73 occurred with a barrier of 21.6 kcal/mol. While this barrier is feasible, the previously described subsequent editing mechanisms that proceed via $^{\text{IIaRC}}$ (Figure 5.6) all have markedly high barriers that are likely enzymatically unfeasible. This proposal was thus discounted.

Alternatively, His73 may activate Cys182 in the presence of a neutral His186, i.e., all three conserved residues (His73, His186 and Cys182) are initially neutral. The resulting free
energy surface for the mechanism elucidated, along with schematic illustration of the corresponding QM/MM optimized structures obtained, is shown in Figure 5.10. As before, the first step is deprotonation of the catalytic Cys182 thiol by His73; that is, the conversion of the pre-reactive complex $^\text{IIcPRC}$ into the required reactive complex $^\text{IIcRC}$. This proton transfer occurs via $^\text{IIcTS1}$ with a required energy of 21.7 kcal/mol. Notably, this barrier is only 0.1 kcal/mol higher than obtained for the analogous step in the presence of a protonated His186 suggesting that it does not exert significant influence on the activation step. In the reactive complex $^\text{IIcRC}$ the Cys182 is now deprotonated while His73 protonated and His186 remains neutral. Importantly, it lies just 0.4 kcal/mol higher in energy than the pre-reactive complex $^\text{IIcPRC}$, Figure 5.10

Figure 5.10. FES obtained (kcal/mol), together with optimized structures with select bond distances shown (Angstroms), for the ThrRS editing domain's mechanism IIc in which Cys182 acts as a base, but Cys182, His73 and His186 are all initially neutral.
In $^{11}\text{RC}$ the Cys182 thiolate is hydrogen bonded to an active site water (W1), $r_{(\text{Cys182}S^−H_{W1})} = 2.04$ Å. The latter moiety W1, as observed in the other mechanisms examined herein, is also positioned in reasonable proximity to the substrate's $C_{\text{carb}}$ center with a $w_{1}O^−C_{\text{carb}}$ distance of 3.22 Å. As in the previous scenarios, the thiolate of Cys182 facilitates nucleophilic attack of W1 at the substrate's $C_{\text{carb}}$ center by simultaneously abstracting a proton from W1. In this present chemical system, this step proceeds through $^{11}\text{TS2}$ at a cost of 27.9 kcal/mol (Figure 5.10). This is in fact the highest barrier obtained of all mechanisms considered herein for the formation of a tetrahedral intermediate complex. Notably, it is 18.0 kcal/mol higher than that obtained for the corresponding step in the most preferred model $^{11}\text{b}$ (i.e., $^{11b}\text{TS1}$; Figure 5.9).

The oxyanionic tetrahedral intermediate formed, $^{11}\text{IC}$, is 16.7 kcal/mol higher in energy than the initial pre-reactive complex $^{11}\text{PRC}$. The second and final step is cleavage of the $\text{Ado76}^{3'}\text{O}−C_{\text{carb}}$ bond. This is again found to preferentially occur simultaneously by a water-mediated proton transfer from Lys156 onto the $\text{Ado76}^{3'}$-oxygen. This reaction proceeds via $^{11}\text{TS3}$ and requires just 8.4 kcal/mol with respect to $^{11}\text{IC}$; 25.1 kcal/mol relative to $^{11}\text{PRC}$. In the product complex $^{11}\text{PC}$ the now hydrolysed non-cognate serine and tRNA$^{\text{Thr}}$ are 2.84 Å apart. In addition, it's formation is thermodynamically favoured having a relative energy of -43.2 kcal/mol. However, while this is therefore clearly an exergonic mechanism, and more favourable than some of the alternative pathways considered, its barrier for the rate-limiting step in this case is thermodynamically less favorable than that of the preferred mechanism, $^{11}\text{b1}$ (see Figure 5.9).
5.4 Conclusion

In this study we have performed an extensive systematic MD and QM/MM investigation on the post-transfer editing mechanism catalyzed by the editing domain of ThrRS from *Escherichia coli* for deaminoacylation of mischarged Ser-tRNA\(^{\text{Thr}}\). In particular, a range of possible scenarios have been examined in which either of the potential mechanistic bases His73 or Cys182 deprotonate an active site water to facilitate the latter’s nucleophilic attack on the substrate’s carbonyl center in a hydrolytic editing process. In addition, we examined the potential of a His73-promoted pathway involving a non-hydrolytic self-cyclization/lactone formation. In total, the chemistry and feasibility of 11 different possible mechanisms, within different possible active site protonation states, were examined and compared.

When His73 acts as the initial base five different mechanisms were considered in which it deprotonates either a nucleophilic water molecule (W1, resulting in hydrolytic editing) or the Ser-AMP substrate’s serinyl \(\beta\)-OH, resulting in lactone formation. For those pathways in which His73 activates W1 the mechanism essentially occurred in two stages: (i) nucleophilic attack of W1 at the Ser-AMP's \(C_{\text{carb}}\) center to form a tetrahedral intermediate; and (ii) cleavage of the Ado\(76^3\text{O}—C_{\text{carb}}\) bond with protonation of the Ado\(76^3\text{-oxygen}. The first stage is a common one-step reaction with a free energy barrier of 27.6 kcal/mol. For the second stage, several possible proton sources were considered. The most favorable pathway proceeded *via* a one-step reaction in which Lys156 protonates the Ado\(76^3\text{-oxygen*} through a second active site water molecule (W2) with a rate-limiting barrier of 33.4 kcal/mol. For the His73-promoted lactone formation pathway, deprotonation of the \(\beta\)-OH group with formation of the cyclic tetrahedral intermediate occurred with a barrier of 31.4 kcal/mol. The
rate-limiting step of this pathway, however, was cleavage of the $\text{Ado76}^{3'}\text{O}—\text{C}_{\text{carb}}$ bond at a cost of 32.8 kcal/mol.

For the Cys182-base scenarios, three chemical models with varying combinations of protonation states for the conserved His73, His186 and Cys182 residues were used to investigate the feasibility of six possible mechanisms. As with the alternate His73-promoted hydrolytic editing mechanism, these occurred in two stages as above. Of all mechanisms considered herein, the most kinetically and thermodynamically feasible pathway was obtained when both His73 and His186 are in their neutral states, and Cys182 is in its thiolate form. The first stage of the mechanism, Cys182 activation of W1 and formation of the tetrahedral intermediate, occurs in one-step with a barrier of 9.9 kcal/mol. For the second stage, cleavage of the $\text{Ado76}^{3'}\text{O}—\text{C}_{\text{carb}}$ bond, the most favorable pathway also occurs in one step with Lys156 protonating the $\text{Ado76}^{3'}$-oxygen via W2 with a free energy barrier of 20.8 kcal/mol. The resulting product complex, with a relative free energy of $-53.6$ kcal/mol with respect to its corresponding initial reactive complex, was also the most favored of all of the mechanisms. Hence, the present results support the conserved Cys182 residue as being in its thiolate form and acting as the required mechanistic base that initiates editing within the editing domain of \textit{E. coli} ThRS.

Importantly, given the noted similarities to the editing site of AlaRS and related enzymes that share the same catalytic motif, this mechanism may be more generally applicable.
5.5 References


35. Molecular Operating Environment (MOE) Chemical Computing Group Inc., 2013.08; 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.08.


Chapter 6.

Unraveling the Critical Role Played by $\text{Ado}^{\text{76}}\text{2'}-\text{OH}$ in The Post-Transfer Editing by *Archaeal* Threonyl-tRNA Synthetase.
6.1 Introduction

The central role of the aminoacyl-tRNA synthetase (aaRS) family of enzymes is to catalyze the linkage of amino acids to their corresponding tRNA.\(^1\) For each existing amino acid there is an aaRS that catalyzes its coupling onto the cognate tRNA via two half-steps: activation and acylation. First, the amino acid is activated via reaction with adenosine triphosphate (ATP) to form an aminoacyl-adenylate intermediate (aaAMP), followed by transfer of the aminoacyl (aa) moiety onto its cognate tRNA\(^{\text{aa}}\). Impressively, this loading process occurs with a misacylation error of \(~1\) in every 10,000 reactions.\(^3\) Thus, aaRSs play a key role in the accurate translation of an organism's genetic code into proteins.\(^2\) Defects in the aminoacylation process can result in misfolded and thus incorrectly functioning proteins, which can eventually lead to disease states such as neurodegeneration.\(^4\)

Due to structural and chemical similarities between some amino acids it can be challenging for the synthetic site of aaRSs to achieve proper discrimination. Consequently, many exploit proof-reading (editing) mechanisms that selectively act against incorrectly activated amino acids or aminoacylated tRNA\(^{\text{aa}}\) to degrade them, often back to their constituent components (Scheme 6.1).\(^7\) For example, several aaRS use a tRNA-independent pre-transfer editing mechanism (Scheme 6.1; reaction 3) whereby the aminoacyl-adenylate is hydrolyzed within the synthetic site.\(^8\)

However, pre-transfer editing is not always solely sufficient to ensure the necessary fidelity of aminoacylation, such as between isoelectric amino acids. Hence, an additional post-transfer correction (editing) mechanism is often employed using a distal active site.\(^9\) Indeed, almost half of the aaRSs utilize post-transfer editing and thus behave as double sieve
models.\textsuperscript{10} In this proofreading mechanism, misacylated tRNA\textsuperscript{aa} is shuttled to the editing site where the ester bond between the incorrect aminoacyl moiety and tRNA\textsuperscript{aa} is cleaved (Scheme 6.1). Interestingly, a triple-sieve editing mechanism is used by alanyl-tRNA and prolyl-tRNA synthetases to ensure accurate aminoacylation.\textsuperscript{11}

\begin{center}
\textbf{Scheme 6.1.} Schematic representations for the aminoacylation and editing mechanisms employed by ThrRS.
\end{center}

For instance, threonyl-tRNA synthetase (ThrRS) is a class II synthetase that must necessarily discriminate between its cognate substrate threonine from the non-cognate substrate serine. It is known to utilize a variety of editing mechanisms including pre-transfer editing against serinyln-adenylate.\textsuperscript{7,12-13,14} Unfortunately, such editing is not sufficient to achieve the necessary required high fidelity. With the exception of mitochondrial ThrRS, bacterial, eukaryotic and archaeal ThrRS all possess a remote active site for post-transfer editing.\textsuperscript{13} It is generally accepted that the editing domain sequence of ThrRS is not evolutionarily conserved.\textsuperscript{3} Indeed, distinct from bacterial and eukaryotic versions that
possess a universal editing domain found in both Thr- and AlaRS, *archaeal* ThrRS employs a unique N-terminal post-transfer editing region. Accordingly, two different catalytic scenarios have been suggested for the respective editing mechanisms in *bacterial/eukaryote* and *archaeal* ThrRS. In the post-transfer editing mechanism of *E. coli* (bacterial) ThrRS, an active site cysteiny1 or histidyl residue is thought to act as the base that deprotonates the nucleophilic water and initiates the reaction.

In contrast, the editing mechanism of *archaeal* ThrRS is thought to be a paradigm for most of the editing domains in the other aaRSs. Moreover, previous sequence analysis demonstrated a substantial sequence similarity between the archaeon *Pyrococcus abyssi* ThrRS (Pab-NTD) and D-amino acid deacylases (DTD). The latter domain is used by aaRSs responsible for hydrolyzing misacylated D-aa-tRNA, thus preserving the homochirality of proteins. The activity of this enzyme is extended to all the tRNAs misacylated by a D-amino acid. Given their similarities, Pab-NTD ThrRS has been shown to be capable of accommodating several D-amino acids, and thus is proposed to complement the role of DTD in ensuring protein homochirality.

For Pab-NTD, due to the apparent lack of direct involvement of the enzyme residues in a way that could facilitate catalysis, post-transfer editing has been suggested to take place through a substrate-assisted mechanism. In particular, based on experimentally obtained structures, the free hydroxyl group of the adenosine ribose sugar of the tRNA (Ado76^2'- or 3'-OH for class I and II respectively), appears to be the only potential base in close proximity to the substrate's scissile ester bond. Furthermore, a significant inhibition in editing has been observed upon its removal. Thus, it has been suggested that the Ado76^2'-OH or Ado76^3'-OH group of the substrate (aa-tRNA) triggers the reaction by orienting a nucleophilic H2O
molecule in close proximity to the aa-tRNA ester bond, i.e., the $\text{Ado}^{2'-3'}\text{OH}$ group plays a structural or anchoring role.\textsuperscript{10} However, the precise role played by this hydroxyl group is still debated.

The post transfer editing mechanism for different aaRSs has been the subject of several computational studies. For example, Tateno and coworkers\textsuperscript{22} performed a Molecular Dynamics (MD) study in order to identify essential residues in the editing mechanism of leucinyl-tRNA synthetases for removing the noncognate valine.\textsuperscript{22} A nucleophilic H$_2$O molecule was observed to be consistently hydrogen bonded with the $\text{Ado}^{3'}\text{OH}$ group and in close proximity to the carbonyl carbon ($\text{C}_{\text{carb}}$) of the substrate (Val-tRNA$^{\text{leu}}$). Consequently, quantum mechanics/molecular mechanics (QM/MM) free energy simulations were performed to explore the reaction. It was concluded that it involves a self-cleaving hydrolytic mechanism assisted by the substrate's own $\text{Ado}^{3'}\text{OH}$ group.\textsuperscript{7,16} The hydrolytic editing mechanism of the mislinked alanine in the freestanding editing domain (INS) of prolyl-tRNA synthetases has also been explored using QM/MM calculations.\textsuperscript{23} Similarly, it was concluded that the $\text{Ado}^{2'}\text{OH}$ group is crucial for positioning the nucleophilic H$_2$O for subsequent nucleophilic attack on the substrate's $\text{C}_{\text{carb}}$ center.

In this current study, we have complementarily used both MD simulations and QM/MM methods to gain insights into the role played by the substrate's $\text{Ado}^{2'}\text{OH}$ group in the post-transfer editing mechanism of Pab-NTD ThrRS. Similarities between the fully-substrate bound editing site and biocatalysts involving ribozymal catalysis were noted.\textsuperscript{24,25} In particular, the function of the free RNA hydroxyl group is either to appropriately position and anchor the nucleophilic H$_2$O\textsuperscript{26} or to directly participate in the catalytic mechanism in the ribosome.\textsuperscript{27} Hence, the investigations were broadened to examine the applicability and
feasibility of analogous tRNA substrate-mediated pathways in the editing site of ThrRS including concerted and step-wise anchoring, and single or double proton shuttle mechanisms (Schemes 6.2 and 6.3).

6.2 Computational Methods

6.2.1 Molecular Dynamics Simulations

The Molecular Operating Environment (MOE) program\textsuperscript{28} was used to prepare all chemical models for the MD simulations with the X-ray crystal structure of the editing domain of ThrRS from \textit{Pyrococcus abyssi} with bound seryl-3\textquotesingle-aminoadenosine (PDB ID: 2HL1)\textsuperscript{3} used as the initial template. The link nitrogen atom in the ester bond was replaced with an oxygen atom and the protonation states of all the residues were assigned according to the PropKa protonation tool implemented in MOE. All crystallographic water molecules were removed except for two positioned near the substrate. The model was then minimized using the molecular mechanics (MM) forcefield AMBER12. The complex was then solvated by adding a layer of water to 6 Å around the enzyme-ligand system, resulting in a system with total number of 11000 atoms. The generated chemical model was then submitted for a second MM minimization using AMBER12.

The final complex was then submitted for an unconstrained 10 ns MD simulation using the NAMD program,\textsuperscript{29} with a time step of 2 fs under constant pressure and temperature until the system reached an equilibrium state. The generated conformations from this MD simulation were analyzed based on their root mean square deviations (RMSD) of the heavy atoms of their active site residues. The obtained RMSD values were then clustered and the most representative structure (with the most prominent conformation) was chosen for the
subsequent QM/MM calculations. This structure was then minimized using the AMBER12 forcefield. Finally, a suitable chemical model for the QM/MM calculations was then derived by truncating the system to only include all residues and waters within 20 Å of the active site’s substrate (2000 atoms in total).

6.2.2 QM/MM calculations

To elucidate the proofreading mechanism, we utilized the hybrid ONIOM QM/MM approach\textsuperscript{30-31} as implemented in the Gaussian 09 suite of programs.\textsuperscript{32} This approach has been shown to be a powerful tool for examining many related catalytic mechanisms.\textsuperscript{33-34} The entire chemical model was divided into two subsystems based on their level of contribution to the reaction, Figure 6.2. The active region, high layer, was described using a quantum mechanical (QM) method while the remaining protein environment is treated using a MM method.

Figure 6.1. Illustration of the high layer (QM region) of the QM/MM model used in this study.
The QM region, consisting of 88 atoms, included the substrate 3'-seryl-adenosine (SerAA), a model of serine bound to the A76 residue of tRNA\textsuperscript{Thr}, two H\textsubscript{2}O molecules and the backbone chain of Pro116 and Ala82 as they are thought to stabilize the accumulated negative charge on the oxygen atom (O\textsubscript{carb}) in the transition state. In addition, the R-group of Lys121 was included as it is thought to be important in orienting and positioning the nucleophilic H\textsubscript{2}O molecule in close proximity to the substrate’s ester group.\textsuperscript{7} In addition, Lys121 is also conserved in the editing sites of other aaRSs where it is thought to play a similar role.\textsuperscript{35} Glu134 was also included in the QM region due to its role in positioning the substrate through a salt bridge formation with the substrate’s α-NH\textsubscript{2}.\textsuperscript{3} Indeed, mutation of either Glu134 or Lys121 dramatically diminishes post-transfer editing activity.\textsuperscript{3} To describe the QM region the density functional theory methods B3LYP,\textsuperscript{36} M062X and M06HF\textsuperscript{37} in conjunction with the 6-31G(d,p) basis set were used, while the AMBER96 forcefield\textsuperscript{38} was used to describe the surrounding protein environment, i.e., the low (MM) layer.

Hence, optimized geometries and frequencies were obtained at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory, as were the corresponding Gibbs free energy corrections (ΔG\textsubscript{corr}). Relative energies were determined by performing single point energy calculations on the above optimized structures at the ONIOM(B3LYP/6-311+G(2df,p):AMBER96) level of theory. Only the Cα centers in the low layer were held fixed, all other atoms being free to move during optimizations.

It is important to note that all mechanisms were studied using three different DFT functionals; B3LYP, M062X and M06HF. In particular, we have evaluated the ability of these functionals to reliably and accurately describe the studied mechanisms and their thermochemistry, with the results summarized in Table 6.1. It was observed that the
mechanism was sensitive to the % of HF exchange-correlation (XC) included. The M06HF functional, that has the highest XC% contribution, provided a better kinetic description of the mechanism. It is also clear from this table that, M062X gave more reasonable energy values relative to B3LYP, in agreement with previous theoretical studies on ribozymal catalytic mechanisms.\textsuperscript{39-40} Thus, in the following discussion we will focus only on the data obtained using the M06HF functional to describe the QM region.

6.3 Results and Discussion

6.3.1. Concerted mechanism

Initially, the substrate Ser-AA in the optimized reactive complex (RC) was found to be positioned in the optimal orientation required for the subsequent nucleophilic attack through salt-bridge formation between the Glu134COO\(^{-}\) group and the serine’s side chains (\(\alpha\)-NH\(_2\) and the \(\beta\)-OH) at \(r_{\text{Glu134COO}^{-}\cdots\text{HO}_{\beta\text{ser}}} = 1.89\, \text{Å}\) and \(r_{\text{Glu134COO}^{-}\cdots\text{H}_{2\alpha\text{ser}}} = 1.57\, \text{Å}\). More importantly, the nucleophilic H\(_2\)O molecule (W) is also placed nearby the substrate in optimal position to interact with the ester bond, \(r_{\text{O}_{\text{w}}\cdots\text{C}_{\text{car}}b} = 2.93\, \text{Å}\), and the angle of nucleophilic attack (\(\angle_{\text{O}_{\text{w}}\cdots\text{C}_{\text{car}}b\cdots\text{O}_{\text{car}}b\) is 87.6\(°\). The latter water molecule is held in this position through forming a strong hydrogen bond network with the main chain O\(_{\text{car}}b\) of Pro116 residue, \(r_{\text{H}_{\text{w}}\cdots\text{O}_{\text{Pro116}}} = 1.43\, \text{Å}\), the side chain of Lys121 residue, \(r_{\text{H}_{\text{w}}\cdots\text{N}_{\text{Lys121}}} = 1.41\, \text{Å}\), as well as the Ado\(_{76}2'\text{OH}\) group, \(r_{\text{H}_{\text{w}}\cdots\text{O}_{\text{Ado76}}} = 1.91\, \text{Å}\). Moreover, the O\(_{\text{car}}b\) atom forms a moderately strong hydrogen bond interaction with the main chain \(\cdots\text{NH\cdots}\) of the Ala82 residue at a distance of 2.15 \(\text{Å}\). The hydrolytic mechanism is initiated by a nucleophilic attack of the oxygen atom of the water molecule (O\(_{\text{w}}\)) on the C\(_{\text{car}}b\) of the substrate, leading to
the formation of a new C$_{\text{carb}}$−O$_w$ bond. This is followed by cleavage of the ester bond, where the bridged oxygen (O$_b$) abstracts proton from the dissociated H$_2$O molecule.

6.3.1.1 Anchoring Mechanism:

In this mechanism, the nucleophilic H$_2$O molecule attacks the C$_{\text{carb}}$ atom and its proton is instantly shifted toward O$_b$ atom of the ester bond through a 4-membered ring transition state (I$^\text{TS^4}$), Scheme 6.1. The obtained free energy barrier for this step is found to be 32.7 kcal/mol, which is quite a high barrier enzymatically, as a result of the geometrically constrained 4-membered ring. In this transition state, the substrate is positioned in a less productive hydrogen bond orientation relative to the main chain $\text{−NH−}$ of Ala82 residue, which is required to stabilize the generated negative charge on the oxyanion. In this position, the O$_{\text{carb}}$⋯HN$_{\text{Ala82}}$ interaction is now weaken at distances of 2.30 Å with 0.15 Å further relative to RC, Table 6.2
Scheme 6.2. Schematic illustration of the concerted mechanisms studied for: (a) the anchoring mechanism (4-membered ring); (b) proton Shuttle (6-membered ring); and (c) double-proton shuttle (8-membered ring).

$^{1}$TS$^4$ is followed by the formation of the product complex (PC1) where the $C_{\text{carb}} \cdots O_b$ distance is drastically elongated to 3.07 Å. Obviously, the $\text{Ado7}^2 O$ of the adenosine-leaving group is stabilized by forming a new (O$_b$—H) single bond with a length of 0.97 Å, whereas the departure of serine amino acid is facilitated by forming a new $C_{\text{carb}} \cdots \text{OH}_{w1}$ bond with length of 1.32 Å. Relative to the initial reactive complex RC, the product complex PC1 lies 5.6 kcal/mol lower in energy suggesting a thermodynamically favorable product complex. The free energy surface for this concerted scenario is shown in Figure 6.1.
Figure 6.2. The optimized molecular structures for the obtained transition states (TSs) with selected bond lengths in Angstroms (Å).

Moreover, to better clarify the precise role for Ado762′OH group following the current anchoring pathway, we investigated the mechanism in the presence of the deoxy substrate Ado762′H group instead of Ado762′OH. Notably, the missing hydrogen bond between W molecule and the substrate’s Ado762′OH led to a negligible change in the position of the W molecule to the C_carb by an increase of just 0.07 Å relative to the distance obtained in the wildtype RC. Identical to the RC, the W1 molecule is tightly held in this place by forming a strong hydrogen bond with the main chain O_carb of P116 with a distance of 1.40 Å in addition to the presence of the conserved Lys121 residue (O_w1···H_Lys121) at a distance of 1.41 Å. Also, the angle of the nucleophilic attack (∠O_w···C_carb—O_carb) is now 90.14°. Thereby,
the previously proposed role of the $\text{Ado76}2'O\text{H}$ group to align the nucleophilic W in appropriate orientation might be possible inaccurate. More interestingly, the obtained energy barrier for the 4-membered concerted mechanism was found to be even more favorable than the one obtained in the presence of the $\text{Ado76}2'O\text{H}$ group with energy value of 25.4 kcal/mole (data not shown).

**Figure 6.3.** The calculated free energy surface in kcal/mol for the concerted mechanisms to cleave the ester bond of the mischarged Ser-tRNA$^{\text{Thr}}$ using M06HF functional.

In order to fill this gap of understanding, an alternative mechanism that might provide a complete picture regarding the actual role of the $\text{Ado76}2'O\text{H}$ group in the mechanism has been explored.

6.3.1.2 Proton-Shuttle Mechanism

In this pathway, the proton transfer to the leaving $O_b$ atom is shuttled through the assistance of the bridging $\text{Ado76}2'O\text{H}$ group, see Scheme 6.1, and subsequently we rather
obtained a 6-membered cycle transition state, **Scheme 6.1B**. In fact, the 6-membered cycle formation/breaking mechanism in the ribosome studies is found to proceed with a lower energy barrier relative to 4-membered one.\(^{37}\) This preference is due in part to the presence of more hydrogen bonds and more productive angles for the protons transfer process. Importantly, the driving force for this kind of mechanisms is the development of low barrier hydrogen bond (LBHB), which is known to significantly contribute in facilitating many enzymatic catalysis.\(^{38}\) This type of LBHB interaction takes place when two atoms or more with similar \(pK_a\) values form strong hydrogen bonds and thus can share a proton matching the geometry of our system. The optimized transition states with the important bond lengths for this mechanism (\(^{1}\text{TS}^6\)) are shown in **Figure 6.2**. The energy barrier for this step was found to be 26.1 kcal/mol, approximately 6.7 kcal lower than the value obtained in the 4-membered one (\(^{1}\text{TS}^4\)) and hence enzymatically more feasible (**Figure 6.3**). In this \(^{1}\text{TS}^6\), the developed negative charge on the oxyanion atom is stabilized through hydrogen bond formation with the amide linkage \(-\text{NH}−\) of the Ala82 residue at a distance of 2.05 Å which is 0.10 Å shorter and thus stronger interaction than the corresponding one in \(^{1}\text{TS}^4\).

**Table 6.1.** Calculated energy barriers in kcal/mol for the various mechanisms (represented by their transition state label) obtained at the ONIOM(DFT method/(6-31G(d,p):AMBER96).

<table>
<thead>
<tr>
<th>DFT Method</th>
<th>Transition State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(^{1}\text{TS}^4)</td>
</tr>
<tr>
<td><strong>B3LYP</strong></td>
<td>31.1</td>
</tr>
<tr>
<td><strong>M062X</strong></td>
<td>34.9</td>
</tr>
<tr>
<td><strong>M06HF</strong></td>
<td>32.7</td>
</tr>
</tbody>
</table>
In an effort to identify a lower energy barrier via better stabilization of the oxyanion in $^{1}\text{TS}^6$, we attempted to provide extra hydrogen bond interaction as suggested before. Specifically, we expanded our chemical model in the QM layer to comprise more residues (main side chain of His83 residue and an extra bridged H$_2$O molecule), offering more hydrogen bond to the O$_{\text{carb}}$ atom, $r(\text{O}_{\text{carb}}\cdots\text{H}_2\text{O}) = 1.80$ Å. Reoptimization of the new complexes along the same concerted proton shuttling pathway did not lead to a considerable change in the energy barrier relative to the smaller QM model (data not shown). Thus, charge stabilization on the O$_{\text{carb}}$ atom does not contribute significantly in the obtained energy barrier in the system studied herein.

6.3.1.3 Double–Proton shuttle Mechanism

Notably, there are two water molecules trapped near the critical Ado76$^{2'}$OH group in our chemical model. Thereby, an 8-membered ring concerted mechanism is also explored in our investigation. Specifically, the proton transfer process from the nucleophilic W molecule to the Ado76$^{2'}$OH group takes place through a second bridged water molecule (double-proton shuttle mechanism), $^{1}\text{TS}^8$ in Scheme 6.2c.

In contrast to $^{1}\text{TS}^6$, $^{1}\text{TS}^8$ costs a higher energy barrier with a value of 31.47 kcal/mol, Figure 6.3 and Table 6.1 However, regardless of the increased number of hydrogen bonds involved in the proton transfer process in $^{1}\text{TS}^8$, it is less energetically favorable due in part to its geometric distortion, suffers from steric hindrance, Figure 6.2 Moreover, $^{1}\text{TS}^6$ has a quite strong hydrogen bond with more planar $\angle\text{O}_{\text{carb}}\cdots\text{O}_{\text{carb}}\cdots\text{NH}_{\text{Ala82}}$ bond angle of 156.2° for the oxyanion stabilization; while the same angle in the case of $^{1}\text{TS}^8$ is found to be less productive
with value of 116.2°, (although $r(O_{\text{carb}} \cdot \cdot \cdot NH_{\text{Ala82}}) = 1.87$ Å), indicating a deviation of the substrate from the most productive orientation in the active site.

The optimized molecular structures with selected bond lengths for places where bond formation and bond breaking take place are shown in Figure 6.2. Comparing between the key bond distances in these three transition states, ($^1\text{TS}^4$, $^1\text{TS}^6$ and $^1\text{TS}^8$) it was noticed that, the most significant change in the geometry of the models corresponds to locations where bond forming and breaking takes place. Specifically, the $C_{\text{carb}} - O_b$ bond distance ($r_l$) is mostly cleaved in case of $^1\text{TS}^4$ and $^1\text{TS}^6$ complexes with values of 1.79 and 1.65 Å demonstrating a late transition state while in the $^1\text{TS}^8$ the same bond is slightly cleaved at a distance of 1.48 Å. In agreement with this observation, we noticed the newly forming bond $O_W - C_{\text{carb}}$ ($r_2$) is more advanced in the case of $^1\text{TS}^4$ and $^1\text{TS}^6$ with distances of 1.53 and 1.48 Å than its value of 1.61 Å in $^1\text{TS}^8$, Figure 6.2.

6.3.2. Two-step Mechanisms

6.3.2.1. Anchoring Mechanism

The accumulated negative charge on the oxyanion group in the transition states was thought to be the main reason for the overestimated barriers. As a result, an alternative step-wise mechanism was proposed through two subsequent steps with two 4-membered ring transition states along the pathway, Scheme 6.3, pathway a.$^{35,40-41}$

In the first step of this mechanism, a 4-membered ring ($^{II}\text{TS}^4$) is formed through a concurrent step of a nucleophilic attack on $C_{\text{carb}}$ and the proton of the W molecule is instead transferred to $O_{\text{carb}}$ atom. The latter proton transfer might provide more charge neutralization to the developed negative charge on the $O_{\text{carb}}$ atom, Scheme 6.3. This transition state leads
to a formation of diol intermediate complex (IC) where $C_{\text{carb}}$ atom forms two single bonds with two different OH groups (OH$_W$ or OH$_{\text{carb}}$). This intermediate complex is followed by a second proton transfer from one of the new hydroxyl groups to the $O_b$, stabilizing the leaving group in the product complex (PC1 or PC2) through another 4-membered transition states ($^{\text{II}T_2^4}$ or $^{\text{II}T_2^4'}$), Scheme 6.3.

Scheme 6.3. Schematic representation for the two-step mechanisms studied involving a: (a) 4-membered ring proton shuttle; and (b) 6-membered ring proton shuttles.
After exploring all these possibilities, the obtained barriers are found to be quite high, Table 6.1. Accordingly, the two-step anchoring mechanism is also kinetically inappropriate. Overall, the generated oxyanion is not the only factor that influences the obtained energy barriers; the constrained 4-membered ring transition state has a significant impact on the energy costs as noticed previously in similar chemical systems.\textsuperscript{36, 42} Subsequently, another two-step proton shuttle mechanism through a 6-membered cycle transition state has been considered, Scheme 6.3.

6.3.2.2 Proton-Shuttle Mechanism

This proton shuttle mechanism is initiated by a proton shift from the $\text{Ado}_7\text{OH}$ group to the $O_{\text{carb}}$ atom. Though $\text{II}TS1$, an advanced $O_w\cdots C_{\text{carb}}$ distance ($r_2$) at 1.31 Å is observed to represent a new bond formation. Importantly, the negatively charged $O_{\text{carb}}$ atom is further counterbalanced via the shuttled proton from the $\text{Ado}_7\text{OH}$ group at $O_{\text{carb}}\cdots \text{H}$ bond length of 1.42 Å ($r_2$), Figure 6.2. However, the energy barrier for this step is found to be 25.9 kcal/mol, which is just 0.17 kcal lower in energy than the concerted one ($^1\text{TS}_6$). Indeed, the latter transition state is considered the most feasible step among the other transition states. The next step is the formation of the diol IC, which is identical to the previous one mentioned in the 4-membered step-wise mechanism. Afterwards, a second proton transfer process occurs from one of the newly generated hydroxyls of the IC to the $O_b$ atom and then formation of product complex where the labile ester bond is cleaved and the serine is released, Scheme 6.3, pathway b. As we discussed earlier in the step-wise anchoring mechanism, there are two different possibilities for the reactions to move from the
intermediate IC into one of the product complexes PC1 or PC2. In the first possibility, the second proton shuttle step will take place from OH\textsubscript{w} hydroxyl group to the O\textsubscript{b} atom though a cyclic 6-membered transition state (\textsuperscript{II}TS\textsubscript{6}). A higher energy barrier than the first one with value of 30.5 kcal/mol is needed for this \textsuperscript{II}TS\textsubscript{6} to proceed. Alternatively, the second step proton transfer (\textsuperscript{II}TS\textsubscript{6}′) occurred through a proton shuttle from the OH\textsubscript{carb} atom to the O\textsubscript{b} atom at a lower energy barrier with value of 25.9 kcal/mol, Figure 6.4. The obtained key distances for these two transition states are shown in Figure 6.2. Notably, the facile O\textsubscript{carb}–O\textsubscript{b} bond (\textit{r}5) is slightly cleaved to a distance of 1.58 Å in case of \textsuperscript{II}TS\textsubscript{6}. Meanwhile the same bond is observed to be 1.87 Å in case of \textsuperscript{II}TS\textsubscript{6}′, indicating a late transition state. This significant difference in the bond lengths induces a geometric change in the surrounding residues; the negatively charged O\textsubscript{b} center in \textsuperscript{II}TS\textsubscript{6}′ is counterbalanced by forming a quite stronger hydrogen bond with the nearby Lys121 residue with the assistance of a bridged water molecule, \textit{r}(O\textsubscript{b}…H\textsubscript{W1}) = 1.46 Å. This hydrogen bond is not observed in the \textsuperscript{II}TS\textsubscript{6}, which might be the reason for its higher energy barrier with respect to \textsuperscript{II}TS\textsubscript{6}′. In addition, the accumulated negative charge on O\textsubscript{carb} atom, that is neutral in \textsuperscript{II}TS\textsubscript{6}, is neutralized by the amide chain –NH– of Ala82 at distance of 2.42 Å in \textsuperscript{II}TS\textsubscript{6}′.

The following step will be the generation of different product complexes according to the source of the offered proton to the leaving group as well as the position of the OH group in the releasing serine, Scheme 6.3. Both product complexes are thermodynamically favorable and lie at -5.8 and -2.3 kcal/mol in energy for PC1 and PC2 respectively. The obtained free energy surface for the optimized complexes along these pathways is shown in Figure 6.4.
Collectively, the post transfer editing pathway is believed to be more favored through a step-wise proton shuttle mechanism from a kinetic perspective. Additionally, the two-step mechanism gave a slightly more stable transition state ($^{II}\text{TS}^6$) than the concerted one ($^{I}\text{TS}^6$), Table 6.1.

6.3.3 D-amino acid deacylase function

AaRSs play a central role in preserving the overall homochirality of the protein by discriminating between the enantiomeric species of the cognate L-amino acid against the noncognate D-amino acid. This homochirality is maintained by using another freestanding checkpoint called D-amino acid deacylases (DTD) responsible for hydrolyzing the
misacylated D-aa-tRNA.\textsuperscript{44,45} The activity of this enzyme is not restricted for specific tRNA, but is extended to all the tRNAs misacylated by a D-amino acid. Interestingly, previous sequence analysis demonstrated a substantial sequence similarity between the N-terminus motif in Pab-NTD ThrRS and DTD.\textsuperscript{44,45} As a result, Pab-NTD ThrRS was found to be capable of accommodating several D-amino acids, and thus is proposed to complement the role of DTD in perpetuating the homochirality.\textsuperscript{3,46} To study the ability of the current editing model to hydrolyze the D-Thr–tRNA\textsubscript{Thr} and function as a deacylase, we investigated the proton shuttle editing mechanism with D-Thr-AA as the substrate.

\textbf{Figure 6.5.} The obtained free energy surface for the deaminoacylation of D-threonine by the editing site of ThrRS.

Notably, the optimized \textsuperscript{III}RC for this new complex shows that the nucleophilic W molecule is held in the optimal position required for nucleophilic attack, $r(O_w\cdots C_{\text{carb}}) = 2.92$
Å. Similar to the obtained geometry in RC in which serine is the substrate, W is oriented productively as a result of forming strong hydrogen bonds with Ado76′OH, the side chain of Lys121 and the main chain O_{carb} of Pro116 with distances of 1.93, 1.41 and 1.43 Å, respectively. Following the concerted proton-shuttle mechanism where the rate limiting step is a 6-membered cycle transition state (^{III}TS^6), we observed a kinetically favorable barrier with 29.2 kcal/mol with respect to ^{III}RC. The obtained bond distances for this transition state indicates a late transition state where the {carb}O−O_w bond is more advanced toward a typical single bond (1.48 Å) and the C_{carb}⋯O_b bond is mostly cleaved at a distance of 1.65 Å. The substrate’s oxyanion group is stabilized through forming a moderately strong hydrogen bond with the main chain of Ala82 residue, r(Ala82NH⋯O_{carb}) = 2.02 Å. Later, ^{III}TS^6 is followed by the formation of ^{III}PC where the ester bond of D-threonine-AA substrate is entirely cleaved, r(C_{carb}⋯O_{carb}) = 2.45 Å, and a new single OH bond is formed in the leaving group. The free energy surface for the correcting mechanism of D-threonine is shown in Figure 6.5

Accordingly, the elucidated editing mechanism takes place through a kinetically feasible barrier if D-threonine is the substrate. This barrier is a strong precursor suggesting the possibility of the Pab-NTD editing model to employ the extra DTD function. Importantly, this editing pathway for eliminating the attached D-amino acid might be universal for the other DTD enzymes, which has substantial similarity to the one studied here.

6.4 Conclusion

To explore the post-transfer proofreading mechanism of the *archaeal* ThrRS editing site, we applied molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) methodologies. Assessing three different DFT functionals; B3LYP, M062X and M06HF indicated the sensitivity of the computational models to the amount of XC%
included and the M06HF functional with a full-HF exchange functional resulted in the lowest energy barriers.

After exploring different editing mechanisms, the obtained energy barriers of either the concerted or the stepwise mechanisms via the geometrically constrained 4-membered ring transition state are kinetically less favorable compared to the 6-membered ones. Interestingly, the role played by the $A_{do762}'OH$ group to align the nucleophilic $H_2O$ molecule in close vicinity to the ester group is suggested experimentally, removing it during the 4-membered ring mechanisms showed negligible differences in the calculated energy barriers. Significantly, the contribution of the $A_{do762}'OH$ group in the proton shuttle from the nucleophilic water molecule to either the $O_b$ (concerted) or $O_{carb}$ (two-step) dramatically decreases the energy barrier to 26.1 and 25.9 kcal/mol, respectively. Importantly, in a good match with the experimental results, the $A_{do762}'OH$ group is found to necessarily trigger the correction mechanism. Additionally, using this elucidated mechanism, an extra deacylase activity of the editing site of *archaeal* ThrRS has been confirmed.

### 6.5 References


Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09* Gaussian, Inc.: Wallingford, CT, 2009.


Chapter 7.

QM/QM Investigation on The Pre-transfer Editing Mechanism in Seryl-tRNA and Isoleucyl-tRNA Synthetases
7.1 Introduction

The accuracy of the genetic translation of RNA is of central importance for the proper functioning of the coded proteins and cell survival. This process is carried out by the fundamental aminoacyl-tRNA synthetases (aaRSs), enzymes that are noted for their remarkable precision.\(^1\) Importantly, any limitation in the fidelity of translation can result in fatal diseases including cancer.\(^2\)-\(^3\) The entire family of enzymes is split into two main subclasses, class I and class II, according to structural differences in their catalytic sites.\(^5\) Each of the twenty aaRSs is only responsible for the transfer of its native amino acid to the cognate tRNA. Therefore, a deeper understanding of the accuracy of the catalytic function of the aaRSs is a matter of growing interest.\(^4\)

The active site of the aaRS catalyzes a two-step process, namely amino acid activation and tRNA acylation. Initially, it juxtaposes both the amino acids and the ATP in a productive orientation for the activation step to occur. In this step, the carboxylate group’s oxygen atom of the amino acid nucleophilically attacks the phosphorous atom of the ATP molecule leading to the formation of aminoacyl-AMP and releases the inorganic diphosphate group (Scheme 7.1).\(^6\) The acylation step then takes place, whereby the amino acid is covalently attached to the terminal adenosine of the cognate tRNA, resulting in aminoacyl-tRNA.

In fact, aaRSs perform their functions with a translation error rate of 1:10000, which is remarkable enough for aaRS to be known as the paradigm of molecular specificity.\(^7\) In order for the aaRS to achieve this outstanding specificity, their active sites recognize the corresponding amino acids and reject any larger or similar ones. However, it is a complex task for some of these enzymes to correctly distinguish their native amino acids and reject
the isosteric and isoelectronic noncognate ones.\textsuperscript{8} To ensure the high fidelity of the translation process, aaRS utilize numerous proofreading mechanisms including pre-transfer (removing the misactivated adenylate) or post-transfer (excluding the misacylated-tRNA) editing mechanisms.\textsuperscript{9-10}

The location for post-transfer editing to take place is a separate active site known as the editing domain.\textsuperscript{7} Meanwhile, the hydrolysis of the undesired aa-AMP intermediate can occur in several different places.\textsuperscript{12} Firstly, the synthetic site might eject the misactivated aa-AMP substrate into the cytoplasm which then undergoes uncatalyzed hydrolysis.\textsuperscript{13} Alternatively, this hydrolysis might take place in the synthetic site by a tRNA-independent pre-transfer editing process. However, within the synthetic site is the most common location for pre-transfer editing pathway in aaRSs.

Unusually, IleRS, a class I aaRS, seldom utilizes a tRNA-dependant pre-transfer editing mechanism.\textsuperscript{14-15} This atypical reaction contributes almost one-third to the entire editing mechanism employed by \textit{Escherichia coli} IleRS. In fact, IleRS’s exclusion of the structurally similar valine takes place by employing both pre- and post-transfer hydrolysis processes.\textsuperscript{16-17}

In the pre-transfer tRNA-dependant mechanism, the misactivated Val-AMP is transferred from the synthetic site to the remote editing region and then the pre-transfer proofreading takes place. This suggestion was supported by a recent X-ray crystal structure of the misactivated Val-AMP substrate in the editing region of \textit{E.coli} IleRS.\textsuperscript{18} Accordingly, it is suggested that the editing domain is the dedicated location that hosts all the different types of editing mechanism employed by IleRS in contrast to most of the aaRS.\textsuperscript{19} Furthermore, it has been argued that IleRS is able to accommodate the homocysteine and
Chapter 7

edit it before its incorporation to the tRNA to help prevent, for instance, cardiovascular diseases. In response, we explored the pre-transfer editing mechanism against homocysteine in the editing region of IleRS using a suitable X-ray structure (PDB ID: 1WK8).

The class II aaRS seryl-tRNA synthetase (SerRS), which is responsible for the ligation of serine to the corresponding tRNA\textsuperscript{Ser}, relies only on pre-transfer editing to discriminate against the misactivation of the noncognate and structurally similar cysteine and threonine aminoacids. Despite the lack of a separate editing region, the pretransfer editing in SerRS is sufficiently efficient as to achieve an outstanding level of discrimination. In particular, methanogenic \textit{archaeal} SerRS contains a Zn(II) metal ion in its active site that is essential for substrate recognition. Moreover, SerRS is also responsible for charging the tRNA\textsuperscript{Sec} with serine for the indirect biosynthesis of selenocysteine (Sec).

It is generally accepted that the active site of most aaRSs do not have a catalytic residue that can promote the reactions, and as a result a substrate-assisted mechanism is the most common scenario, although there are suggested exceptions. Importantly, a self-cyclization mechanism has been suggested in the active sites of many enzymes to edit against the toxic homocysteine (Hcy). In this mechanism, one of the nonbridging oxygens of the phosphate group (O1\textsubscript{p}) of the aa-AMP substrate acts as the base that deprotonates the substrate's thiol. The resulting thiolate sulfur then attacks the \textit{sp\textsuperscript{2}} carbonyl carbon of the substrate (C\textsubscript{carb}) and the resultant thiolactone is released, \textbf{Scheme 7.1}. Similarly, LysRS excludes misactivated ornithine-AMP in the form of ornithine-\textit{δ}-lactam, by which the terminal ammonium group of the substrate was deprotonated and the resultant R-NH\textsubscript{2} is the nucleophile. Consequently, the pre-transfer editing mechanism by IleRS against the
unnatural amino acid Hcy, and SerRS against the non-cognates cysteine and threonine might follow the same pathway through self-cyclization mechanism.

Scheme 7.1. The generally proposed self-cyclization mechanism in the pre-transfer editing against Hcy.\textsuperscript{11}

AaRSs are widely recognized as novel drug targets for different infectious diseases.\textsuperscript{32} IleRS in particular has been identified as the target for the antimicrobial compound mupirocin, which is currently being used for treatment of both bacterial skin infections and the fatal parasitic infection African Trypanosomiasis.\textsuperscript{33-34} Essential to the development of new and more potent therapeutic drugs is a more complete understanding of how such enzymes may act against them.\textsuperscript{35-36}

In this study, we have computationally investigated the pre-transfer editing mechanisms of IleRS and SerRS against possible noncognate amino acids. In particular, the applicability of the common substrate-assisted self-cyclization mechanism for such reactions was examined.

7.2 Computational Methods

2.1 Molecular Dynamics Simulations

We started our simulation by considering X-ray structures of each enzyme with bound substrate analogues, PDB ID: 1WK8\textsuperscript{18} and 2CJ9\textsuperscript{23}, respectively. Using the Molecular Operating Environment (MOE) software\textsuperscript{37} we performed the required chemical
modification for each enzyme to mutate the substrate to its native form. The Val-adenosine monosulfate (AMS) analogue substrate in 1WK8 was changed to Hcy-AMP; the sulfur atom in sulphate group was mutated to the native phosphorous atom Hcy-AMP…IlleRS/tRNA\textsuperscript{Ile}. Moreover, in case of 2CJ9, the sulfur atom in the sulfate group of the Cys-AMS was mutated to phosphorous atom and three different models containing three different substrates naming Ser-AMP….SerRS/tRNA\textsuperscript{Ser} and the noncognate Cys-AMP…CysRS/tRNA\textsuperscript{Cys} and Thr-AMP….SerRS/tRNA\textsuperscript{Ser} have been prepared. The ionization states of all the residues have been identified based on the predicted $p$Ka using the available tool in MOE. In each generated model, we have performed solvation using 2 Å layer followed by MM minimization using AMBER12 forcefield. Later, the minimized systems were initially submitted for 100 ps equilibration from 0 to 300 K temperatures at constant pressure where tether ranges from 0 to 100 Å have been applied on heavy atoms.

Figure 7.1. The selected residues treated in the QM layer in SerRS (left model) and in IleRS (right model).
Then, under constant pressure and temperature, each generated model was submitted for 10 ns MD simulation with a time step of 10 ps using the NAMD engine. The conformations generated from the MD simulations were then analyzed based on their root mean square deviation (rmsd) values which then clustered and the average structure with the most prominent conformation has been selected to be the representative model. Starting from this structure we derived our QM/QM chemical models after truncating it to include the first shell of the residues around the substrates. (Total number of atoms is 744 atoms in IleRS and 656 atoms in SerRS).

7.2.2 QM cluster and QM/QM calculations

To investigate the validity of self-cyclization mechanism in the pre-transfer editing pathway in the studied models, the alternative hybrid ONIOM (QM/SE) scheme was used. In this approach, each system has been divided into two main subsystems according to their importance into the catalytic mechanism. The active site residues are considered in the high layer that is represented by DFT using the hybrid functional B3LYP as implemented in Gaussian 09, Figure 7.1. Meanwhile, the surrounding protein environment is treated by semiemperical (SE) PM3 method instead of the standard QM/MM to improve the description of the coupling interaction. Indeed, this methodology has been noticed to be successful in exploring different catalytic mechanisms and Zn-metalloenzymes in particular. In the case of IleRS, the QM layer comprises the Hcy-AMP substrate, Asp328 residue that form a salt-bridge interaction with the amino group of the substrate as well as the backbones of Pro324 and Gly325 residues. In addition to four water molecules were added to stabilize the accumulated negative charges on the O$_{\text{carb}}$ and O$_2p$ during the progress of the reaction, total number of atoms in this layer is 84 atoms.
Meanwhile, the QM layer in SerRS contained the substrate, Arg353, Arg336 and Asn345 residues and two water molecules to help neutralizing the phosphate oxygens as well as $O_{\text{carb}}$ atom. Moreover, the tetracoordinated Zn(II) with its binding residues (Glu355, Cys461 and Cys306) have been added to the high layer due to the direct coordination between the substrate-$\text{NH}_2$ group and the Zn(II) metal ion, total number of atoms in this QM layer is 107 atoms, Figure 7.1. Later, frequency analyses were also computed at ONIOM (B3LYP/6-31G(d,p):SE), which is the optimized level of theory, in order to ensure the nature of the stationary points and also to calculate Gibbs free energy ($\Delta G_{\text{corr}}$). Relative energies were calculated by performing single point energy calculations at the ONIOM (B3LYP/6-311+G(2df,p): PM3) level of theory.

To gain further details regarding the impact of protein environment on the behavior of the catalytic mechanism and the overall geometry of the active site, we have explored the mechanism using Quantum Mechanics cluster calculations. The QM cluster approach is known to be a successful tool for studying enzymatic reactions. In this approach, only the active site’s significant residues (which is identical to high layer of QM/SE system used herein) are considered in the cluster calculation. As described in the methods section, only the active site residues are excised and the remaining protein is omitted. Each residue was capped by adding a hydrogen atom to the respective carbon (143 atoms). Afterwards, single point energy calculations were performed using B3LYP/6-311+G(2df,p). Meanwhile, the frequency calculations were also performed at the optimization level of theory (B3LYP/6-31G(d)) to estimate the Gibbs free energy corrections ($\Delta G_{\text{corr}}$).
7.3 Results and discussion

7.3.1 MD results

According to our MD simulation on both IleRS and SerRS, the substrates were noticed to adopt a unique bent conformation in the active site. The aminoacyl-adenylates functional groups, i.e., the R–SH terminal of Hcy and Cys or the R–OH of Thr and Ser amino acids are in an optimum orientation for a proton transfer reaction to take place with the assistance of O1_p of the AMP group. Figure 7.2 indicates the change in these distances along the 10 ns MD simulation as well as the adopted conformation for both Cys and Hcy in SerRS and IleRS, respectively. During 10 ns simulation run the average S_{Hcy}⋯O1_p distances are found to be 3.58 Å, 5.53 Å and 5.39 Å for Cys-AMP, Thr-AMP and Ser-AMP, respectively in the active site of SerRS, Figure 7.2A. In addition, the intramolecular distance between the R-S(O) group and the C_carb atom is noticed to be 3.27, 2.87 and 2.86 Å for Cys-AMP, Thr-AMP and Ser-AMP, respectively. Similarly, in the active site of IleRS, the average H_{Hcy}⋯O1_p distance during the simulation run was found to be 4.54 Å while the H_{Hcy}⋯C_carb is found to be 4.57 Å, Figure 7.2B. Accordingly, the active site pockets of IleRS and SerRS recognize all the substrates in a specific conformation required for the self-cyclization mechanism.
Figure 7.2. The conformation adopted by the substrate as well as a plot indicating the change in selected key distances resulted from the MD simulation; A) for the Cys-AMP in SerRS and B) for Hcy-AMP in IleRS.

7.3.2 QM/SE Investigations

7.3.2.1 Pre-transfer editing against Hcy-AMP by IleRS

The first optimized model using QM/SE calculations for Hcy-AMP in the active site of IleRS suggested a less favorable orientation of the substrate, $^{1}R^{C}_{HCY}$. In this state, the $\alpha$-H$_{3}$N$^{+}$ is hydrogen bonded to the nearby Asp328 residue with $\text{Asp328C}O^{\cdot}\cdots \text{N}^{+}H_3$ distance of 1.85 Å, Figure 7.3. Interestingly, the R–SH terminal of the Hcy-AMP substrate is not positioned nearby the O1$_p$ atom which is required for initiation of the reaction. Consequently, we performed a dihedral scan calculation around $C_{\beta}$–$C_{\gamma}$ bond to obtain a better conformation of the substrate for the pre-transfer editing mechanism to proceed. This dihedral scan cost a free energy barrier of 24.3 kcal/mol ($^{1}$$\text{TSI}_{HCY}^\circ$) and lead to the formation
of $^{1}\text{IC}^{\text{Hcy}}$ where the R-SH group of the Hcy substrate is tilted down to form hydrogen bond with the O1p atom, Figure 7.3. This hydrogen bond interaction is formed with the assistance of a bridging water molecule (W1) at $\text{HcyS(H)} \cdots \text{O}_{\text{W1}}$ and O1p$\cdots$H$_{\text{W1}}$ distances of 2.15 and 1.69 Å, respectively, Figure 7.3.

In this intermediate complex, the substrate is positioned productively for the subsequent concerted mechanism through a cyclic transition state $^{1}\text{TS}^{2\text{Hcy}}$. The O1p atom deprotonates the R-SH of the Hcy-AMP with the assistance of the mediating W1 and the thiolate group attacks the C$_{\text{carb}}$ atom of the substrate to complete the substrate-assisted mechanism with an energy barrier of 23.2 kcal/mol. Relative to the energy barrier required for $^{1}\text{TS}^{1\text{Hcy}}$, $^{1}\text{TS}^{2\text{Hcy}}$ is more kinetically favorable and therefore the first transition state (of the dihedral scan) is the rate-limiting step. In $^{1}\text{TS}^{2\text{Hcy}}$ the HcyS$\cdots$C$_{\text{carb}}$ distance has shortened to 2.22 Å and concomitantly the C$_{\text{carb}}$$\cdots$O$_{b}$ bond has significantly lengthened to 1.72 Å. Moreover, in this late transition state the O1p atom becomes protonated by forming typical O1p$\cdots$H single bond with a length of 1.01 Å. Also, it should be noted that the transient thiolate ion in this step is stabilized via the formation of moderately strong hydrogen bonds to the bridged W1 and another water molecule with distances of 2.17 and 2.59 Å, respectively. Moreover, the negative charge on the O2p atom is also stabilized through hydrogen bond formation with a nearby water molecule with distance of 1.71 Å.
Figure 7.3. Optimized molecular structures during the studied substrate-assisted self-cyclization mechanism in IleRS with selected key distances as well as free energy differences. The values in blue color obtained using QM-cluster calculations.

In the product complex the homocysteine has formed a thiolactone, $\text{IPCH}_{\text{Hcy}}$, which can then be released from the active site, Figure 7.3. Notably, this step is found to be exergonic as the $\text{IPCH}_{\text{Hcy}}$ lies lower in energy than $\text{IRC}_{\text{Hcy}}$ at $-24.6$ kcal/mol. In this step, the $\text{Hcy}S$–$C_{\text{carb}}$ distance is 1.78 Å, indicating a typical C–S single bond. Meanwhile, the $C_{\text{carb}}$–$O_b$ bond is entirely cleaved at 3.11 Å. The $\text{Asp328COO} \cdots \text{H3N}_{\text{Hcy}}$ salt bridge interaction is maintained during the reaction, demonstrating the importance of the Asp328 residue in positioning the substrate for the reaction to proceed. After studying the same mechanism using QM methodology, the overall scenario is found to be kinetically feasible following the same two-step mechanism obtained in the QM/SE calculations, with
insignificant changes to the bond distances. This reasonable change in the interactions may be a result of the flexible geometry of the models when studied in the gas phase. We also explored the possibility that the bridged oxygen (O_b) could be the base that initiates the reaction. According to our QM calculations, the rate-limiting step is now the second transition state with an energy barrier of 45.9 kcal/mol, indicating an enzymatically unacceptable barrier (data not shown). Thus, the O1_p atom is the initiator of the reaction.

7.3.2.2 Pre-transfer editing against Cys-AMP by SerRS

The model generated from the initial MD simulation of the Cys-AMP···SerRS structure was submitted for QM/SE simulation. In the first optimized reactive complex, \textsuperscript{3}HRC\textsuperscript{Cys}, the Cys-AMP substrate adopts a more linear conformation, unlike its orientation in the previous MD simulation. In this orientation, the O_{carb} atom is hydrogen bonded to the amide NH\textsubscript{2} of Asn345, r(NH\textsubscript{2}···O_{carb}) = 2.41 Å. The O1_p atom is stabilized by the formation of hydrogen bonds to the guanidine groups of two nearby arginine residues, Arg336 and Arg353; Arg363 provides two moderately strong hydrogen bonds with distances of 2.44 Å and 1.85 Å while a single hydrogen bond with distance of 1.75 Å is provided by Arg353, Figure 7.4. Considering the charge stabilization around the O2_p atom, the O1_p is relatively less susceptible to hydrogen bonding interactions except for the formation of two hydrogen bonds with two different water molecules, r =1.85 and 1.75 Å. The O1_p atom is thus more basic and hence more likely to act as the base. The Zn(II) ion adopts a tetracoordinate geometry where the N atom of the substrate’s α-NH\textsubscript{2} occupies the fourth coordinated ligand with distance of 2.15 Å. Notably, this geometry is maintained during the progress of the reaction. Similar to the self-cyclization mechanism of Hcy-AMP in the editing site of IleRS, the overall mechanism occurs through two main steps. Initially,
a dihedral scan was performed around the $C_{\beta} - C_{\gamma}$ bond to obtain a more favorable orientation of the $cys_{SH}$ group for the substrate-assisted mechanism. Then, the deprotonation/nucleophilic attack of thiolate group onto the $C_{carb}$ to form the cysteine thiolactone, **Figure 7.4**.

**Figure 7.4.** Optimized molecular structures along the studied substrate assisted mechanism with selected key distances when Cys-AMP is the substrate in IleRS.

Owing to obtain more favorable position for the nucleophilic attack to take place, the $[^{11}RC]_{Cys}$ complex is then submitted for dihedral scan calculation which noticed to be barrier-lees with an energy of $-8.1 \text{ kcal/mol}$, $[^{11}TS1]_{Cys}$ in **Figure 7.5**. This remarkable low barrier
demonstrated the strong preference of the Cys-AMP to adopt the telted conformation for the pre-transfer editing mechanism to occur.

Later, is the formation of the intermediate complex $^{II}$IC$_1^{Cys}$ where the substrate is in more reactant-like orientation, Figure 7.4. This intermediate complex lies at −5.7 kcal/mol relative to $^{II}$RC$_C^{Cys}$ supporting the feasibility of this intermediate, Figure 7.5.

The binding orientation of the Cys-AMP enables the Cys-SH group to be in direct interaction with the O1$_p$ atom of the substrate, $r_{(Cys-S-H\cdots O1p)} = 2.89$ Å, while maintaining the other hydrogen bond interactions with the substrate, Figure 7.4. This $^{II}$IC$_C^{Cys}$ is followed by a concerted transition state, $^{II}$TS$_2^{Cys}$, where an intramolecular deprotonation of the thiol group by the O1$_p$ atom is accompanied by nucleophilic attack on the C$_{carb}$ atom, Figure 7.4. This step was found to be rate limiting with an energy barrier of 14.6 and 20.4 kcal/mol relative to $^{II}$RC$_C^{Cys}$ and $^{II}$IC$_1^{Cys}$, respectively, Figure 7.5. Accordingly, the proposed pre-transfer editing in SerRS against the noncognate Cys takes place through a kinetically feasible self-cyclization mechanism. In this transition state, the C$_{carb}\cdots$O$_b$ bond is partially cleaved at 1.60 Å, while the S\cdots C$_{carb}$ is partially formed at 2.12 Å, Figure 7.4.

Moreover, the hydrogen atom of the substrate’s R-SH group is entirely shifted towards the O1$_p$ atom, $r = 0.98$ Å, indicating a late transition state. Arg336 and Arg353 are found to play essential roles in stabilizing the developing negative charge on the O$_{carb}$ and O1$_p$ atoms. Arg336 forms two hydrogen bonds with distances of 2.05 and 1.84 Å, while Arg353 provides further charge stabilization to the O$_{carb}$ atom with a hydrogen bond distance of 1.96 Å, Figure 7.4. All these factors may contribute to the obtained enzymatically feasible barrier. From this transition state is formed the product complex $^{II}$PC$_C^{Cys}$ where the Cys
thiolactone is completely established and $C_{\text{carb}}$ has moved 2.97 Å away from the $O_b$ atom. Notably, the Arg336 and Arg353 residues maintain their hydrogen bond interaction with the $C_{\text{carb}}$ and $O_{2p}$ atoms. Arg336 forms two moderately strong hydrogen bonds with distances of 2.13 and 1.79 Å to the $O_{\text{carb}}$ and $O_{1p}$ atoms, respectively. Similarly, the guanidine group of Arg353 forms hydrogen bond interactions at 2.03 and 1.74 Å, Figure 7.4.

![Figure 7.4](image)

**Figure 7.4.** The free energy surface of the self-cyclization mechanism in SerRS; the black, red and blue colored surface is for Cys-AMP substrate, Thr-AMP and Ser-AMP substrates, respectively.

This $\text{II}_{\text{PC}}^\text{Cys}$ is found to be thermodynamically favorable with an energy 3.6 kcal/mol lower than $\text{II}_{\text{RC}}^\text{Cys}$, Figure 7.5. The formation of this 4-membered cysteine thiolactone is thermodynamically less favorable than the formation of the 5-membered homocysteine thiolactone, which is likely to be a result of the geometrically constrained 4-membered ring versus the more favorable 5-membered one obtained in the $\text{I}_{\text{PC}}^\text{Hcy}$. 
7.3.2.3 Pre-transfer editing against Thr-AMP by SerRS

As mentioned in the Section 7.1, the aminoacylation site of SerRS accommodates the native serine amino acid and edits against the structurally similar cysteine (Section 7.3.2.2) and threonine. Accordingly, we expanded our study to examine the validity of a self-cyclization mechanism in the pre-transfer editing against the non-cognate Thr-AMP. Overall, the mechanism is noticed to be identical to the case of Cys-AMP and takes place through two main steps, Figure 7.6. The Thr-AMP is in the linear conformation in the first optimized structure, $^{II} \text{RC}^{\text{Thr}}$, similar to our observation of $^{II} \text{RC}^{\text{Cys}}$. Notably, the type of interaction between Thr-AMP substrate and surrounding residues is almost identical to the analogous Cys-AMP, Figure 7.6. The only significant difference between the two substrates is the absence of a hydrogen bond between substrate’s $O_{\text{carb}}$ with the nearby Asn345. The Zn(II) ion adopts a tetracoordinate, tetrahedral geometry with the substrate being the fourth ligand through its amino group, $r(\alpha-\text{NH}_2\cdots\text{Zn}) = 2.10 \text{ Å}$.

The first step in the mechanism is the dihedral scan which was found to be kinetically accessible with an energy barrier of just 0.7 kcal/mol relative to $^{II} \text{RC}^{\text{Thr}}$, Figure 7.5. This scan process lead to the generation of a thermodynamically favored ($-0.6$ kcal/mol lower in energy than $^{II} \text{RC}^{\text{Thr}}$ intermediate complex $^{II} \text{IC}^{1\text{Thr}}$, where the Thr-AMP substrate adopts a bent conformation, Figure 7.6. In this conformation, the OH group of the substrate is directly hydrogen bonded to the O1p atom with distance of 2.26 Å. The following cyclisation step is initiated by intramolecular proton abstraction from the alcohol group by the substrate’s O1p atom and concomitant nucleophilic attack onto $C_{\text{carb}}$ to form a cyclic tetrahedral intermediate. The energy barrier for this step is 26.6 kcal/mol relative to $^{II} \text{IC}^{1\text{Thr}}$ indicating an enzymatically acceptable rate-limiting step, Figure 7.5.
In the product complex, $^{II}$PC$^{\text{Thr}}$, the $O_b-C_{\text{carb}}$ bond is cleaved entirely at 3.08 Å and a cyclic lactone derivative is formed, Figure 7.6. However, the formation of $^{II}$PC$^{\text{Thr}}$ is thermodynamically unfavorable at 2.6 kcal/mol higher than $^{II}$RC$^{\text{Thr}}$ which is 6.1 kcal/mol more than the $^{II}$PC$^{\text{Cys}}$, Figure 7.5.

Figure 7.6. Optimized molecular structures of the studied substrate-assisted editing mechanism of SerRS with selected key distances in Å when Thr-AMP is the substrate.

7.3.2.4 The validity of pre-transfer editing against the cognate Ser-AMP

Our calculation was further extended to include the native Ser-AMP substrate to investigate the difference between the cognate and the noncognate substrates in the self-cyclization mechanism. Initially, the information obtained from our MD simulation did not
highlight any significant difference between them. Using the QM/SE approach, the first optimized model, $^{II}RC^{\text{Ser}}$, indicates an identical binding geometry of Ser-AMP to the other non-cognate Cys-AMP and Thr-AMP substrates. The Ser-AMP substrate is positioned linearly in the active site and its $\alpha$-NH$_2$ group coordinates to the tetrahedral Zn(II) metal ion with distance of 2.11 Å. Moreover, careful investigation of the type and distances of interaction between Ser-AMP substrate and the surrounding environment indicates a negligible change during the progress of the reaction relative to the Cys-AMP and Thr-AMP cases.

Following the two-step substrate assisted mechanism, we have successfully identified all the structures along the pathway, Figure 7.7. The dihedral scan results in an energy barrier of 4.3 kcal/mol which is 12.5 and 3.6 kcal/mol higher than the required energy for the corresponding step in Cys-AMP and Thr-AMP, respectively, Figure 7.5.

This $^{II}\text{TS}1^{\text{Ser}}$ is followed by the formation of the $^{II}\text{IC}1^{\text{Ser}}$ where the serine’s OH group forms a hydrogen bond with the O1p atom of the substrate with distance of 2.12 Å, Figure 7.7. Unlike the corresponding complex in the case of the non-cognate substrates, the $^{II}\text{IC}^{\text{Ser}}$ is thermodynamically higher in energy than $^{II}RC^{\text{Ser}}$ by 7.4 kcal/mol indicating that this is a less favorable orientation for the Ser-AMP substrate relative to the linear one, Figure 7.5.

Next is the deprotonation of the OH group of the substrate by the adjacent O1p atom and simultaneous nucleophilic attack of the incipient alkoxide on the $sp^2$ C$_{\text{carb}}$ of the substrate, Figure 7.7. This step was observed to have a high-energy barrier at 31.5 and 24.1 kcal/mol relative to its corresponding $^{II}RC^{\text{Ser}}$ and $^{II}\text{IC}1^{\text{Ser}}$, Figure 7.5.

In fact, this is the highest energy barrier obtained among all the studied substrates, indicating the infeasibility of SerRS to edit out the cognate serinyl substrate. We tried to
explore any geometrical change which might be the reason for this significant difference in the energy barrier between $\text{Cys}_{TS}$ and $\text{Ser}_{TS}$. Our careful investigations on the transition states demonstrate the identical binding of the substrates in both cases. Particularly, the exact same types of interaction with negligible change in the distances have been noticed, Figure 7.4 and Figure 7.7. Considering the difference in the $pK_a$ value between the Cys and Ser amino acids with the former being more likely to give its proton, this may be a considerable factor indicating the lower cost in energy to obtain $\text{II}_{TS}^{\text{Cys}}$, Figure 7.5.

In the product complex $\text{II}_{PC}^{\text{Ser}}$ the cyclic serine intermediate is formed 2.97 Å away from the monoprotonated AMP. Interestingly, unlike our finding for the other substrates, the formation of $\text{II}_{PC}^{\text{Ser}}$ is found to be energetically unfavorable and lies 17.7 kcal/mol higher in energy than the corresponding $\text{II}_{RC}^{\text{Ser}}$, Figure 7.5.

Notably, relative to $\text{II}_{PC}^{\text{Cys}}$ there is a significant change in the geometry of $\text{II}_{PC}^{\text{Ser}}$. Specifically, Arg353, which is known to have a stabilizing interaction with the O$_{\text{carb}}$ atom in both $\text{II}_{PC}^{\text{Cys}}$ and $\text{II}_{PC}^{\text{Thr}}$, shows no interaction in $\text{II}_{PC}^{\text{Ser}}$, Figure 7.7. Instead, it forms a single hydrogen bond interaction with O2p atom with a distance of 2.18 Å. Accordingly, the pre-transfer editing against Ser-AMP is not enzymatically feasible from both a kinetic and thermodynamic perspective.
Figure 7.7. Optimized molecular structures along the studied substrate-assisted mechanism with selected key distances when Ser-AMP is the substrate.

7.3.3. Self-cyclization mechanism using QM-only approach

To study the impact of treating the protein environment implicitly, we have utilized QM-only approach to study the same mechanism. Except for Asn345, similar models to the QM layer in the QM/SE model have been prepared with total number of 65 atoms. We successfully characterized all the intermediates and transition state along the pathways for the different substrates Cys-AMP Thr-AMP and Ser-AMP. Notably, due to the gas phase optimization utilized in this case, considerable change in the types of interaction between each substrate and the surrounding amino acids (data not shown).
**Figure 7.8.** The relative free energy surface of the self-cyclization mechanism in SerRS utilizing QM-only approach; the black, red and blue colored surface is for Cys-AMP substrate, Thr-AMP and Ser-AMP substrates, respectively.

Despite these structural changes the obtained free energy surfaces (Figure 7.8) is quite similar to the ones obtained utilizing the more accurate QM/SE approach. The rate limiting step is found to be II$^{TS2'}$ prior to the formation of the final product complex II$^{PC'}$. As for the QM/SE findings, the rate-limiting step for the self-cyclization of Ser-AMP is noticed to have the highest energy cost. Furthermore, the obtained intermediate complex for the Ser-AMP substrate is less thermodynamically stable relative the cases for the other non-cognate substrates. In all the studies cases, the product complex was found to be energetically favorable yet it is the least stable in the case of the cognate Ser-AMP substrate. Overall, the QM-only methodology could provide satisfactory estimations of the free energy surface in less time.
7.4 Conclusion

The pre-transfer editing mechanism in IleRS and SerRS has been extensively explored utilizing Molecular Dynamics, Quantum Mechanics/Semiempirical (QM/SE) and QM approaches. Our findings indicated that both enzymes employ self-cyclization mechanism to prevent the incorporation of the misactivated Hcy, Cys and Thr substrates. Our MD simulation demonstrated the tendency of all the studied substrates to adopt bent conformations in their corresponding active sites. Furthermore, in case of the editing of IleRS against Hcy-AMP, the rate limiting step is found to be the first transition state that involves a dihedral scan and costs relative free energy of 24.3 kcal/mol. However, in the case of the pre-transfer editing in SerRS, the rate limiting step is the second transition state with a concerted step of intramolecular proton shift from the R-SH/-OH group of the substrate to the substrate’s O1p atom and a concomitant nucleophilic attack on the \( sp^2 \) \( C_{\text{carb}} \) atom. This step has energy barriers of 26.0 and 14.6 kcal/mol for the non-cognate Cys-AMP and Thr-AMP substrates, respectively.

Meanwhile the same step is found to enzymatically infeasible if Ser-AMP is the substrate and a kinetic barrier of 31.5 kcal/mol was obtained. Additionally, the following intermediate complex in the studied substrate-assisted mechanism for IleRS (with the Hcy-AMP substrate) and SerRS showed a relatively stable complex where the substrates adopt tented conformations suitable for the editing mechanism to proceed through self-cyclization. Lastly, the generation of the cyclic product complex where the non-cognate amino acids and AMP are moved apart are found to energetically favorable except the case for the native Ser-AMP. Interestingly, omitting the impact of the protein environment during the QM only approach lead to a successful treatment of the mechanism and the
obtained free energy surface shows significant homology with the corresponding ones from the more accurate QM/SE methodology.

7.5 References


Chapter 8.
Comparative QM/MM Study on the Aminoacylation Mechanism of β-Hydroxynorvaline by Threonyl-tRNA Synthetase.
8.1 Introduction

Due to increasing resistance against current antibacterial drugs, an urgent need to develop new antimicrobial agents for novel targets has arisen.\textsuperscript{1} Among the various enzymes used as antibiotic targets, aminoacyl tRNA synthetases (aaRS) have been widely validated as novel antimicrobial agents.\textsuperscript{1-3} AaRS are a ubiquitous family of enzymes that play a crucial role in protein biosynthesis. Based on some differences in structural characteristics, this family of enzymes is divided into two main classes: class I and class II.\textsuperscript{4} AaRS catalyze the attachment of the cognate amino acid onto its respective tRNA through two main half-steps. Initially, the amino acid is activated by reacting with adenosine triphosphate (ATP), forming an aminoacyl-adenylate (aa-AMP), Scheme 8.1. Then, the activated amino acid is covalently attached to either the 2′- or 3′-OH group (based on the class) of the 3′-terminal adenosine (Ado76) of the corresponding tRNA\textsuperscript{aa}.

Insertion of a noncognate amino acid into a protein sequence will lead to abnormally folded proteins\textsuperscript{6} that directly trigger serious diseases such as neurodegeneration, tumorigenesis, and eventually cell death.\textsuperscript{7-9} Hence, for each amino acid there is a specific aaRS which must be able to discriminate between the structurally similar amino acids in the cellular pool to guarantee faithful translation; employing different proofreading functions in either their catalytic sites (pre-transfer editing) and/or in a separate editing region (post-transfer editing).\textsuperscript{5,10} Consequently, they are able to perform their task of aminoacylating their corresponding tRNA with outstanding fidelity; the mistranslation error is on the order of $10^{-4}$.\textsuperscript{5} Since there are known structural differences between bacterial and human aaRS, inhibition of aaRS function in bacteria can be exploited to cause selective prevention of their growth and thus elimination of the infection.\textsuperscript{11} In particular, bacterial threonyl-tRNA
synthetase (ThrRS) is a primary antibiotic target that has been considered recently through the uncompetitive inhibitor Borriledin which inhibits the catalytic function of ThrRS through binding to a hydrophobic region near the active site. More specifically, upon binding it induces a conformational change in the adjacent active site and, as a result, obstructs the optimal binding of the substrate to the cofactor. In addition, ThrRS has been proven to have a significant function related to angiogenesis as antiangiogenesis action is observed upon its inhibition by Borriledin. However, the binding site of Borriledin is conserved in both bacterial and human ThrRS, and thus it lacks selectivity for the bacterial enzyme.

Scheme 8.1. Schematic representation of the general two-step aminoacylation mechanism catalyzed by aaRS.

In contrast, the aminoacylation (catalytic) site does exhibit structural differences between the bacterial and human species. Hence, the alternative use of competitive inhibitors that bind in the catalytic site may display better selectivity for the bacterial enzyme. In such cases, binding of the inhibitor would render the cognate threonine unable to bind and hence the rate of aminoacylation would be significantly diminished. Most of these types of
inhibitors are naturally occurring compounds; for instance, the most clinically widely used aaRS inhibitor is mupirocin, a bacterial isoleucyl-tRNA (IleRS) synthetase inhibitor.\textsuperscript{16} Other naturally occurring inhibitors include indolmycin\textsuperscript{17} that acts against tryptophanyl-tRNA synthetase, cispentacin\textsuperscript{18} that functions against IleRS and prolyl-tRNA synthetase, ascamycin\textsuperscript{2} that inhibits phenylalanyl-tRNA synthetase, and albomycin\textsuperscript{19} which represses serinyl-tRNA synthetase function. Another competitive inhibitor is β-Hydroxynorvaline (βHNV), an unnatural amino acid that differs by a single –CH\textsubscript{2}– group from threonine, has been experimentally shown to readily bind to the catalytic site of ThrRS.\textsuperscript{20-21}

A number of experimental studies have been performed to understand the catalytic activity of ThrRS as well as the main features of its active site.\textsuperscript{20,22-25} It is well-known that its catalytic site contains an essential Zn(II) that adopts a penta-coordinated structure:\textsuperscript{20,23} binding to three enzyme residues (His385, His351 and Cys334) in addition to the threonine substrate's α-NH\textsubscript{2} and β-OH groups. The noncognate substrate serine, also having a β-OH group, can bind in a similar manner to the Zn(II) metal atom and thus can be activated by ThrRS while valine cannot.\textsuperscript{26} Importantly, ThrRS appears to only recognize and activate amino acids that can form a Zn(II)-O-β interaction. The non-natural amino acid βHNV contains just such a substituent OH group and does bind within the active site Zn(II) in a manner similar to that of the threonine. As a result, ThrRS is unable to discriminate against it with high fidelity, and instead catalyses its activation and subsequent aminoacylation at a high rate; only 30-fold less than the cognate threonine.\textsuperscript{20} Moreover, ThrRS appears to possess no ability to edit against βHNV in either its aminoacylation site nor editing sites. As a result it is incorporated into the protein sequence, acting as inhibitor for its growth.\textsuperscript{21}
Herein, we have performed detailed computational studies to gain deeper insights into the aminoacylation of \( \beta \text{HNV} \) within the synthetic site of ThrRS in order to more fully understand the required chemical characteristics to achieve proper inhibition. We have followed the same substrate-assisted concerted scenario elucidated by our group previously and shown in Scheme 8.1,\(^{27}\) using MD simulations in combination with detailed ONIOM(QM/MM) calculations.\(^{28}\) Furthermore, we investigated the active site-bound \( \beta \text{HNV} \) which may be helpful for designing potentially more potent analogues. To determine the most effective chemical conformation, we have considered the acylation mechanism for a systematic isomeric series of substrates, based on the position of the substituted ethyl relative to the substrate’s \( \alpha \)-NH\(_2\) group. Moreover, since protein biosynthesis is a homochiral process, we have also investigated enantioselective translation of \( \beta \text{HNV} \).\(^{29}\) Additionally, we have also performed an assessment on the performance of a variety of GGA density functionals to obtained reliable results. Finally, we performed further calculations to verify the critical role played by the Zn(II) in the active site by mutation with the structurally similar but larger and toxic Cd(II) which also has a closed shell. We then recalculated the relative free energy of the aminoacylation pathway in the presence of Cd(II).

### 8.2 Computational Methods

To prepare the \( \beta \text{HNV} \) chemical model with the right conformation, we compared between the chemical stability of three different possibility according to the orientation of the \(-\text{C}_2\text{H}_5\) group as in Figure 8.1. Notably, isomer b was found to have the lowest energy accordingly, the most stable amongst the other isomers followed by conformer c., Figure 8.1. Therefore, we mainly considered these two conformers for subsequent MD and QM/MM calculations.
8.2.1 Molecular Dynamics simulation:

Based on recent X-ray crystal structures (PDB ID: 1QF6) for the cognate threonine in ThrRS active site\textsuperscript{22} and previous studies done by our group, we adjusted the βHNV ligand orientation mimicking the Thr-AMP….ThrRS/tRNA\textsuperscript{Thr} template to obtain the βHNV-AMP….ThrRS/tRNA\textsuperscript{Thr} Michaelis complex for the two conformers, a and b, Figure 8.1, using the Molecular Operating Environment (MOE)\textsuperscript{30} software package. The protonation states of the residues were assigned according to the protonation tool implemented in MOE. The two obtained models were solvated by adding a 2 Å layer of water molecules followed by molecular mechanics (MM) minimization using the AMBER12 forcefield until the root mean square gradient fell below 0.01 kcal/mol·Å. Under constant pressure, the generated structures were then submitted for 100 ps annealing equilibration from 150 to 300 K. Finally, the generated complexes were then submitted for 10 ns MD simulation with a time step of 2 fs using NAMD engine.\textsuperscript{31} All simulations were carried out under unconstrained pressure and temperature and at a cutoff of 10 Å for non-bonded interactions and tether ranges from 0–100 Å applied to the heavy atoms. Latter, we analyzed the generated trajectories based on their root mean square deviation (RMSD) differences. Clustering analyses were then carried

\textbf{Figure 8.1.} Optimized structures of the different βHNV isomers with different orientations of the terminal methyl using the B3LYP/6-31++G(d,p) level of theory.
out and a representative structure with the most dominant conformation for each complex was chosen for the subsequent analysis. Notably, after performing alignment analysis between these two representative models, we did not observe any considerable changes in the average distances for the most important interactions between the ligand and the surrounding active site residues. After superimposing the two active sites, an RMSD value of 0.31 Å was noticed, Figure 8.2. Thereafter, we minimized the representative models using the AMBER12 forcefield and then truncated them to generate our chemical models for the subsequent QM/MM analyses (see below).

![Figure 8.2. Overlay of the representative active site structures for the two substrate isomers (H atoms are omitted for clarity).](image)

**Figure 8.2.** Overlay of the representative active site structures for the two substrate isomers (H atoms are omitted for clarity).

### 8.2.2 QM/MM Investigations

The overall chemical structure is divided into two main subgroups within the active site based on their importance to the catalytic mechanism applying the ONIOM formalism\(^{32-34}\) as implemented in Gaussian 09.\(^{35}\) The first subsystem is the chemically reactive region consisting of 129 atoms and described at a QM level. This region include the βHNV-AMP substrate, the Zn\(^{2+}\) atom and its three ligating residues, His385, His511, and Cys334, two glutamine residues (Gln484, Gln381), three positively charged residues, Arg363, Arg383,
Lys465, the adenosine76 of the cognate tRNA\textsuperscript{Thr}, and two water molecules. The second subsystem is the remaining environment and was modeled using MM (total number of atoms is 2283). The density functional theory B3LYP method\textsuperscript{36-38} that has been previously demonstrated to give the best performance in the description of Zn metalloenzymes\textsuperscript{39} with the 6-31G(d,p) basis set was used to treat the QM region within the mechanical embedding (ME) formalism, while the AMBER96 force field\textsuperscript{40} was used to describe the MM layer. Furthermore, to estimate the sensitivity of the QM region to the amount of Hartree-Fock (HF) exchange-correlation included, we performed an assessment on the kinetic performance of a variety of GGA density functional (B3LYP* (15\% HF contribution), B3LYP (10\% HF), BP86 (0\% HF), and M06L (0\% HF)). Moreover, in order to address the major limitation of the B3LYP functional in the description of vdW effects,\textsuperscript{41} we included Grimme’s DFT-D3 empirical dispersion corrections.\textsuperscript{42-43} It has been found that including dispersion corrections enhances the reliability of the calculated energy barriers to better replicate experimental results.\textsuperscript{44} Thus, we reoptimized our complexes after including such dispersion corrections.

For all the ONIOM calculations studied herein, relative energies were calculated by performing single point energy calculations at the ONIOM (X/6-311+G (2df,p):Amber96) level of theory where X is the different functionals used for the optimization. To characterize the nature of the stationary point as well as to calculate Gibb’s free energy ($\Delta G_{\text{corr}}$) and zero-point vibrational energy (ZPVE) correction values, frequency analyses were computed at the optimization level of theory, i.e., ONIOM (X/6-31G(d,p):Amber96)-ME.

8.3 Results and Discussion

8.3.1 Exploring the most enzymatically feasible isomer

In agreement with our previous calculations for the most dominant coordination state of
Zn(II), the first structure we obtained from our QM/MM optimization on the first substrate (case b in Figure 8.1) contains a tetrahedral coordinated Zn(II) metal ion where the βHNV-AMP ligand is monoligated to it only through its α-NH₂ group (⁰PRC). A high-energy barrier of 183.8 kJ mol⁻¹ was obtained for the rate-limiting step; too high to be considered enzymatically feasible (data not shown). Alternatively, the Lewis acidity character of Zn(II) with the assistance of the unbounded Asp383 residue allows the substrate to be in its deprotonated form, a common behavior in many Zn-metalloenzymes. More specifically, we obtained alternative pre-reactive complex inside this active site which differs in the protonation state of βHNV-AMP’s hydroxyl group, Scheme 8.1. This enables the βHNV-AMP substrate to be bidentately ligated to the Zn with both α-NH₂ and the β-O⁻ groups. The latter observation is facilitated by the known flexibility of Zn(II) which can switch between coordination states. Notably, the difference in energy between these two prereactive complexes (⁰PRC and PRC) is obtained to be just 0.74 kJ/mol and with a barrierless proton transfer process at -22.6 kJ/mol, Figure 8.3. In this starting from the structure of βHNV-AMP with deprotonated hydroxyl group, we followed the same suggested scenario, where the subsequent step is the cleavage of the Zn(II)···NβHNV bond to form the reactive complex, RC. In this RC, the NβHNV atom forms a moderately strong hydrogen bond with the hydrogen atom of the Ado7ε3'-OH group, r(βHNVN···H3'ε) = 1.74 Å, and the Zn(II) atom is then switched from pentacoordinated to tetracoordinate geometry whereas the βHNV substrate is monoligated to the Z(II) through only its deprotonated β-OH group, r(βHNVOZn)= 1.99 Å, Scheme 8.2. The optimized structures for the obtained transition states and product complexes with selected bond lengths are shown in Scheme 8.2. The free energy surface for the aminoacylation mechanism of βHNV-AMP is shown in Scheme 8.3.
Figure 8.3. The calculated free energy surface obtained at the ONIOM(B3LYP/6-311G+(2df,p):AMBER96)+ΔG_corr level of theory for the overall acylation mechanism for the neutral form (red surface) and deprotonated form (black surface) of βHNV-AMP substrate.

Following this reactive complex, the reaction takes place through a concerted step where the $\text{Ado76}^{3’}$-OH is deprotonated, $r(\text{Ado76}^{3’}\text{O} \cdots \text{H}) = 1.52 \, \text{Å}$, by the α-NH$_2$ group of the substrate concomitant with nucleophilic attack of the $\text{Ado76}^{3’}$-oxygen on the C$_{\text{carb}}$ center, $r(\text{Ado76}^{3’}\text{O} \cdots \text{C}_{\text{carb}}) = 1.88 \, \text{Å}$, with an enzymatically feasible energy barrier of 90.2 kJ/mol. The next step is formation of the product complex, PC in which a single bond between the C$_{\text{carb}}$ and the now deprotonated $\text{Ado76}^{3’}$-O of the ribose sugar is formed, $r(\text{Ado76}^{3’}\text{O} \cdots \text{C}_{\text{carb}}) = 1.37 \, \text{Å}$. As part of the concerted step, this is synchronous with cleavage of $b\text{O} \cdots \text{C}_{\text{carb}}$ bond as indicated by the 3.15 Å separation.

Indeed, because of the complete proton transfer from the nearby Lys468 residue through a bridging H$_2$O molecule to form a typical pro-R-O-H single bond, the accumulated negative charge on the O$_{\text{pro-R}}$ atom has been partially neutralized, facilitating the release of the phosphate group as in the product complex PC, Scheme 8.2. This product complex lies 28.8 kJ mol$^{-1}$ lower in energy with respect to $b\text{PRC}$. 


Collectively, βHNV can be easily acylated in the catalytic site of ThrRS matching the experimental results.\textsuperscript{21} Moreover, we tested the possibility of acylation of the other conformations of βHNV-AMP (isomer a Figure 8.1). The optimized molecular structures with selected bond lengths for this conformation are shown in Figure 8.2. It is clear from this scheme that the types of interaction of these two isomers with the surrounding residues are almost identical in all the obtained complexes along the aminoacylation pathway. For clarity, an overlay of the active sites of the reactive complexes for the two conformations is shown in Figure 8.4. However, following the same aminoacylation mechanism, the energy barrier for the proton transfer step where the neutral substrate (\textsuperscript{9}PRC) is converted to its
deprotonated form (PRC) proceeded with a higher energy barrier than the first case with a value of 20.8 kJ/mol (TS1). In comparison with conformation b, this value is approximately 45.5 kJ/mol higher in energy. More importantly, the rate limiting-step was obtained to occur through transition state with an enzymatically unfavorable energy barrier of 159.6 kJ/mol (TS2). The product complex (PC) lies 21.6 kJ/mol higher in energy with respect to \(^{a}\text{PRC}\); a relative energy difference of 50.4 kJ/mol, supporting the preference of only the first isomer to be acylated which might guide the experimentalists for developing more potent inhibitors.

![Figure 8.4](image.png)

**Figure 8.4.** Overlay of the two QM layers of the reactive complexes for the two different conformations (a and b) of βHNV-AMP (H atoms are omitted for clarity).

8.3.2. An assessment on the performance of HF-XC:

Furthermore, we performed an assessment of a wide range of functionals to describe the free energy surface of aminoacylation given the b conformation of unnatural amino acid βHNV. Full reoptimizations were carried out using the functionals described in the methods section. Notably, the mechanisms underwent through a similar trend displaying negligible changes in the geometry; the bidentate substrate is more stabilized than the monodentate one.
Specifically, the prereactive complex (PRC) with the deprotonated βHNV-AMP substrate lies lower in energy than the corresponding neutral one by -5.47, -6.77, -6.62 kJ mol⁻¹ for B3LYP*, B3LYP⁺ and BP86, respectively. Continuing the calculation starting from PRC, we obtained the free energy surface for each functional, Figure 8.5.

![Relative Energies kJ mol⁻¹](image)

**Figure 8.5.** Free energy surface for the aminoacylation mechanism of βHNV using variety of DFT functionals.

Notably, there is a good correlation between the amount of HF-XC included in each method and the respective energy barrier. The smaller the portion of %XC, the lower the energy barrier. Specifically, with respect to PRC (deprotonated substrate and thus pentacoordinate Zn(II) ion), the calculated free energy barriers are found to be 87.6, 77.8, 61.8 kJ/mol, for B3LYP*, B3LYP⁺ and BP86, respectively with BP86 to be most suitable functional to describe such systems kinetically. Although varying the functionals shows a dramatic impact on the energy barriers, the thermodynamic aspect is relatively unaffected.
since the calculated energy differences of the product complex for each method are quite similar; -27.5, -26.8 and -25.1 kJ/mol for B3LYP*, B3LYP± and BP86, respectively.

Overall, the coordination geometry of the Zn(II) ion in the all studied functionals changed typically from pentacoordinated in the case of a bidentate ligand (PRC) to tetracoordinated in RC and the obtained bond lengths in PRCs are almost identical, Table 8.1. The latter geometry was found to be consistent until formation of the product complex, PC. The significant improvement in the obtained barrier was not a result of the representation of the Zn(II) binding environment.

Table 8.1. The calculated bond lengths, in Å, for Zn(II) binding environment using different DFT functionals.

<table>
<thead>
<tr>
<th>Functional</th>
<th>Zn-C334</th>
<th>Zn-H351</th>
<th>Zn-H385</th>
<th>Zn-NH2</th>
<th>Zn-OH</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP</td>
<td>2.47</td>
<td>2.02</td>
<td>2.13</td>
<td>2.14</td>
<td>2.14</td>
</tr>
<tr>
<td>B3LYP-D3</td>
<td>2.46</td>
<td>2.10</td>
<td>2.00</td>
<td>2.14</td>
<td>2.15</td>
</tr>
<tr>
<td>B3LYP*</td>
<td>2.46</td>
<td>2.12</td>
<td>2.00</td>
<td>2.14</td>
<td>2.15</td>
</tr>
<tr>
<td>B3LYP±</td>
<td>2.46</td>
<td>2.11</td>
<td>2.00</td>
<td>2.12</td>
<td>2.17</td>
</tr>
<tr>
<td>BP86</td>
<td>2.45</td>
<td>2.09</td>
<td>2.00</td>
<td>2.19</td>
<td>2.19</td>
</tr>
<tr>
<td>M06L</td>
<td>2.37</td>
<td>2.14</td>
<td>2.01</td>
<td>2.19</td>
<td>2.09</td>
</tr>
<tr>
<td>B3LYP/LANL2DZ</td>
<td>2.49(^a)</td>
<td>2.19(^a)</td>
<td>2.09(^a)</td>
<td>2.22(^a)</td>
<td>2.09(^a)</td>
</tr>
</tbody>
</table>

8.3.3. The impact of including dispersion correction on the PES:

Adding dispersion correction to the GGA functionals has been widely shown to reduce their limitations of describing long range interactions.\(^{50-51}\) The geometry of the obtained complexes were reoptimized using Grimme’s dispersion corrected B3LYP-D3 and BP86-
D3. Although there are slight differences in the interaction distances in the overall complexes, the calculated energy barrier for the acylation mechanism treated by B3LYP-D3 functional was found to be 93.3 kJ/mol relative to the PRC, Figure 8.6. Moreover, the energy of the formed PC was thermodynamically favored with value of -51.1 kJ/mol with respect to PRC. Therefore, the B3LYP-D3 method better represents our transition metal complexes and gives results that are more suitable kinetically as well as thermodynamically than B3LYP. Also, we performed the same calculation using BP86-D3. Typically, in comparison with BP86, including long range interaction correction greatly improves the description of the energy barrier to be 54.4 kJ/mol relative to the corresponding PRC, Figure 8.6.

Consequently, our system is found to be very sensitive to both the dispersion correction as well as the amount of XC% contribution in the utilized functional. Thus, we decided to investigate the behavior of the hybrid meta exchange-correlation functional M06-L. As one of Minnesota functionals, it was originally developed to approximately account for long range interactions through empirical parameterization. In addition, it has 0% HF contribution which was found to be reliable for describing systems containing transition metals. Using this functional to describe the mechanism under study, the resulting barrier was not as good as the one obtained using BP86, 128.1 kJ/mol relative to PRC.

**Figure 8.6.** The obtained free energy surface for the acylation mechanism using dispersion corrected functionals.
However, the obtained product complex, PC, lies 25.2 kJ/mol lower in energy relative to PRC, Figure 8.5. The description of this functional for the Zn(II) complex interactions in PRC does not possess any marked difference in the distances relative to the other functionals, Table 8.1. To expand our study, we also assessed the performance of the effective core potential LANL2DZ basis set for the description of Zn(II), while the remaining atoms were represented by 6-31G(d,p). After reoptimizing the complexes, the obtained free energy surface for the acylation of βHNV-AMP is shown in Figure 8.5 and the energy barrier for the rate-limiting step increased slightly by to 101.7 kJ/mol with respect to \(^b\)PRC which was described with a normal basis set. The generated PC is 24.9 kJ/mol lower in energy than \(^b\)PRC. Although this basis set gave slightly less reliable kinetic energy in comparison with the common 6-31G(d,p) basis set, it provided us with a considerable barrier, supporting our mechanism at less expense computational time, certainly an advantage for TM complexes. In addition, by measuring the interaction distances between Zn(II) and the bound ligands, Table 8.1, the obtained values are very similar; more evidence supporting the use of LANL2DZ to describe this system.

8.3.4. Mutation of the active site’s Zn(II) to Cd(II):

To investigate whether the Zn(II) ion is essential, we determined the free energy surface for aminoacylation of βHNV replacing Zn(II) with the chemically similar Cd(II). Replacement with Cd(II) is expected to exhibit a great impact on proceeding with the proposed mechanism. Indeed, we used the GenECP methodology to treat the metal with the LANL2DZ ECP and the remaining atoms with the 6-31G(d,p) basis set, similar to the investigations above. Unlike the Zn(II) complexes, the ligand binds differently to the Cd(II), forming a unique prereactive complex, PRC\(^{Cd}\), in an octahedral coordination geometry. In
this structure, the interaction distances with the other protein residues (His383, His351 and Cys334) are elongated and weakened, Table 8.1. Also, the β-OH group of the substrate remains neutral and forms a quite strong hydrogen bond with the nearby Asp383 residue, \( r(\text{AspCOO}^\text{−}\cdot\cdot\cdot\text{HO}_{\beta\text{HNV}}) = 1.49 \) Å. However, this β-OH group is coordinated to the Cd(II), \( r(\text{Cd}^\text{−}\cdot\cdot\cdot\text{OH}_{\beta\text{HNV}} = 2.21 \) Å), Scheme 8.3. In addition to an active site H₂O molecule that binds to Cd(II), \( r(\text{Cd}^\text{−}\cdot\cdot\cdot\text{OH}_2) = 2.34 \) Å, and becoming the sixth ligand in the octahedral geometry, PRC⁻Cd.

Interestingly, in case of the reactive complex where \( r(\beta\text{HNV}^\text{−}\cdot\cdot\cdot\text{HO}_{\text{Ado76}}) = 1.86 \) Å, the β-OH group is deprotonated and it is now firmly binds to the Cd (II) atom, \( r(\text{Cd}^\text{−}\cdot\cdot\cdot\text{OH}_{\beta\text{HNV}}) = 2.44 \) Å) and the \( \text{Cd}^\text{−}\cdot\cdot\cdot\text{OH}_2 \) bond is lengthened to 2.94 Å. For the rest of the obtained complexes along the reaction pathway, the Cd(II) is in a tetracoordinated geometry and the H₂O molecule no longer binds to Cd(II). Following the same acylation mechanism, the energy barrier for the rate-limiting step of the mutant complex (TS⁻Cd) is observed to be 172.63 kJ/mol, much higher than the one observed for the native Zn(II) complex, Scheme 8.3.

To investigate the reasons for this high barrier, we superimposed the geometry of the two transition states (for the mutant and wild type enzymes) and there are no considerable differences; however, a distinct agreement in the obtained types of interactions was observed. The obtained high-energy barrier may be due in part to the change in the coordination geometry of Cd(II) from the 6-coordinate PRC⁻Cd to the 4-coordinate TS⁻Cd.
Scheme 8.3. Optimized molecular structures with selected bond lengths in Angstroms and relative energies (in parentheses; kJ/mol) for the obtained complexes in case of the Cd(II)-mutant complexes.

Similarly, the resulting \( \text{PC}^{\text{Cd}} \) is thermodynamically unfavorable, lying 59.7 kJ/mol higher in energy than \( \text{PRC}^{\text{Cd}} \). Based on these results, we argue that the single mutation of Zn(II) to Cd(II) results in a drastic kinetic and thermodynamic changes in the free energy surface. Thereby, we have unveiled the pivotal role played by Zn(II) its presence is mandatory for the mechanism to proceed. Relative to Cd(II), the size of Zn(II) is small enough to bind only three residues, Cys334, His351 and His385, in addition to the substrate. Except for the substrate-binding mode, the tetracoordinated geometry obtained in the wildtype ThrRS
between the Zn and the protein residues was consistent during the overall mechanism, Scheme 8.3.

8.4 Conclusion

Using MD simulations followed by ONIOM(QM/MM) calculations, we elucidated the aminoacylation mechanism of βHNV in the active site of ThrRS. In agreement with our previous studies, the substrate should be in its ionized form for the reaction to occur. Although two different conformations of βHNV-AMP bind in a similar manner to Zn(II) and interact similarly with the other residues in the active site of ThrRS, a specific conformation was found to be more kinetically favorable. This information is valuable in terms of understanding the chemical features of the unnatural βHNV that could compete with the cognate threonine for aminoacylation.

Moreover, the accuracy of different GGA density functionals, namely; B3LYP (20% HF), B3LYP* (15%HF), B3LYP\(^+\) (10%HF), BP86 (0%HF), and M06L have been assessed in terms of their reliability in providing reasonable energy barriers for our system. The BP86 functional was found to be the most reliable kinetically and including dispersion correction (BP86-D3) improve its accuracy in description of the pathway kinetically. The less computationally expensive LANL2DZ was successful in representing the chemical reaction kinetically without any considerable changes in the Zn(II) coordination geometry. The importance of Zn(II) to promote the catalytic mechanism has been further verified by performing our calculations in a mutated complex (Zn(II) to Cd(II)). The calculated energies of the transition state as well as the product complex of the mutant are drastically increased, preventing progress of the reaction.
8.5 References


Chapter 9.

Computational Investigations on a Library of Potential Competitive Inhibitors for Bacterial Threonyl-tRNA Synthetase.
9.1 Introduction

The fundamental aminoacyl-tRNA synthetase (aaRS) family of enzymes are well known by their vital role in protein biosynthesis\(^1\) in which they display exceptional catalytic fidelity.\(^2\) For each amino acid, there is a particular enzyme responsible for its aminoacylation to its corresponding tRNA.\(^3\)\(^4\) This family of enzymes is divided into two classes, class I and class II, which are based on the distinct architecture of the protein. Each class is comprised of ten amino acids.\(^5\) AaRSs catalyze aminoacylation processes in two steps, namely, activation and acylation. In the activation step, the amino acid reacts with adenosine triphosphate (ATP) forming an aminoacyl-adenylate intermediate (aa-AMP).\(^6\) This step is followed by an acylation step in which the amino acid is covalently linked to its cognate tRNA through either the 2'- or 3'OH\(_{\text{Ado76}}\) of the tRNA terminal adenosine, depending on the enzyme class, forming the aminoacyl-tRNA product.\(^7\)

In fact, it is well established that the accurate translation of amino acids is crucial for proteins to function properly. A high number of mistranslation errors can lead to misfolded proteins that promote many fatal diseases including neurodegeneration.\(^8\)\(^9\) Importantly, in order to perform this outstanding task, aaRS employ editing functions by pre- and/or post-transfer proofreading pathways either in the synthetic site or in a remote editing domain.\(^10\)\(^11\) Mitochondrial ThrRS harbors a pre-transfer editing mechanism to correct misactivated serine before it is linked to tRNA\(^{\text{Thr}}\)\(^12\) However, \textit{archaeal} and \textit{bacterial} enzymes employ post-transfer correcting mechanisms in the editing domain to hydrolyze misacylated Ser-tRNA\(^{\text{Thr}}\)\(^13\)\(^14\)

On the other hand, due to growing antibiotic resistance towards present antibacterial drugs, it has become essential to expand pharmaceutical research toward discovering novel
antimicrobial targets. The inhibition of the bacterial aaRS activity results in restricting bacterial growth and thus eliminating microbial infection. Accordingly, aaRSs are accepted as targets in antimicrobial therapy. One of the most common pathways exploited to inhibit the function of bacterial aaRS is through competitive inhibitors that are able to bind firmly inside the corresponding active sites. These competitive inhibitors are analogues to the native substrates with similar or even stronger binding affinities and thereby diminish the natural activity of the enzyme.

ThrRS is one of the primary antibiotic targets and a number of competitive inhibitors have been synthesized and experimentally identified. In this study, many of the synthesized inhibitors that exhibited potent binding affinity are able to chelate to the Zn(II) in a bidentate fashion, similar to the native threonine. This Zn(II) ion is a unique feature of the active site of ThrRS and adopts a pentacoordinate geometry with the remaining three sites being occupied by three protein ligands (His351, His385 and Cys334). Intriguingly, the Zn(II) ion has a central role in recognizing threonine and serine as well as rejecting the isosteric valine. Moreover, the unnatural β-hydroxynorvaline (βHNV), differing from the cognate threonine by an ethyl group instead of methyl, was experimentally shown to be a substrate for ThrRS. βHNV shares the same side chain groups as the cognate threonine which enables it to bind identically to the Zn(II) ion, preventing ThrRS from being able to discriminate against it.

Zn(II) is the second most abundant metal in enzymes and Zn(II) metalloenzymes are encoded by approximately 10% of the human genome. Among other characteristics, its ability to adopt a range of coordination geometries contributes to its remarkable ubiquity in biological systems. Most commonly, a tetracoordinate Zn(II) is observed. The role of
Zn(II) as a cofactor can vary from structural, where it stabilizes protein conformation, to a catalytic one where it is directly involved in the mechanism.\textsuperscript{26}

![Scheme 9.1](image)

**Scheme 9.1.** The equation used to estimate the B.E, X and Y represent the chelating groups while K and M represent the Push-Pull substituents.

Our main objective in this study was to suggest competitive binders in the active site of *bacterial* ThrRS by conducting hybrid DFT calculations on numerous proposed ligands. Based on our knowledge of the required criteria for an excellent chelating group to Zn(II), we attempted to identify potential ligands. Initially, we focused on elucidating the specific role of Zn(II) in the discrimination mechanism performed by this active site. This understanding then enabled us to tune a variety of ligand sets looking for ideal chemical reactivity. This protocol is widely used in drug discovery for similar Zn-containing systems.\textsuperscript{27-29} In these studies, the main factor representing potency was the value of the binding energy (B.E) of the proposed ligand which is derived from the previous equation, **Scheme 9.1**.

### 9.2 Computational Methodology

We performed our calculations utilizing the M05-2X functional\textsuperscript{30} and 6-31+G(d) basis set as implemented in Gaussian09\textsuperscript{31} for the optimizations of the complexes, except for Zn(II) where 6-311G(d) is used. Based on a recent benchmark study, M05-2X outperformed a list of other DFT functionals in describing Zn-ligand compounds.\textsuperscript{32} Also,
the overall optimization method was used successfully in similar calculations.\textsuperscript{28} Frequency analysis was conducted at the optimization level of theory to ensure the nature of the stationary points. The considered Zn(II) geometry included the first coordination sphere in which two imidazole rings and methanethiol were used to represent two histidine (His351, His385) and cysteine (Cys334) protein residues Scheme 9.1, respectively. Moreover, a library of ligands with terminal heteroatoms were constructed, Scheme 9.2.

9.2.1 Docking Analysis

For the initial model, the X-ray structure of \textit{Escherichia coli} ThrRS (PDB code: 1QF6)\textsuperscript{20} was docked with the ligand. Docking calculations were performed using the Molecular Operating Environment (MOE) software.\textsuperscript{33} Then the entire complex was prepared for the investigation through protonation and solvation (adding a 2 Å layer of water solvent molecules). The considered ligands were also minimized using molecular mechanics (MM) minimization utilizing the AMBER12 force field available in MOE.

Using the induced fit protocol, the receptor is kept rigid in the native conformation with the studied ligands free to move. The most favorable binding modes were selected. Thereafter, the generated series of poses were rescored using the London dG protocol where the most favorable scores, determined by estimating the corresponding binding free energy values, indicate the conformation that has the most favorable hydrophobic, hydrophilic, and hydrogen bond interactions.

We modified the second rescoring step to retain 100 different poses, the maximum value available in MOE. This is followed by conducting forcefield refinements through energy minimization on the resulting poses from the previous placement stage. During this
energy minimization stage, a cut-off distance of 6 Å in combination with fixation of the side chains of the receptor was applied to speed up the calculation. A dielectric constant of 4 was used to represent the electrostatics of solvation and the final energy is then estimated using the generalized Born solvation model. Lastly, the second rescoring has been performed using the GBV1/WSA dG forcefield-based scoring function which evaluates the binding free energy of the ligand in each pose. Notably, during the two rescoring stages, any duplication in the produced binding (H-bonding or hydrophobic interactions) pattern has been eliminated.

9.3 Results and Discussion

9.3.1 DFT calculations

A small model to represent the ThrRS Zn-binding site was constructed for the subsequent calculations. It has been demonstrated that the presence of a Zn(II) metal ion with its Lewis acidic character in most Zn-metalloenzymes imposes a change in the $pK_a$ values of the binding ligands.\textsuperscript{34-35} Specifically, the $pK_a$ of the ligand decreases followed by a relative increase in the $pK_a$ of neighboring bases, facilitating the proton transfer from the ligand’s β-OH group to this base. Consistent with this hypothesis, the active site of ThrRS has an adjacent base residue (Asp383) not ligated to the Zn(II) and strongly hydrogen bonded to the β-OH group of the substrate. Thus, we initially performed our calculations considering two different analogues, 1- amino-2-propanol and 1-amino-2-propoxide to represent the native L-threonine in the neutral and deprotonated OH group forms, respectively.

Notably, the neutral ligand is observed to bind in a monodentate fashion to the Zn(II) through its amine group, and thereby the overall Zn(II) is tetracoordinate. Meanwhile,
deprotonation of the OH group in the ionized threonine analogue enhances its Lewis basicity and thus enables the substrate to bind in a bidentate fashion to the metal ion to give a pentacoordinated Zn(II). Indeed, the latter coordination is consistent with the native threonine’s binding mode in the ThrRS active site, according to a recent X-ray structure.\textsuperscript{21} Notably, the estimated B.E. for the ionized ligand (1a) is markedly larger (-507.1 kJ/mol) than the corresponding value of the neutral one (1g), only -59.2 kJ/mol, Scheme 9.2.

\textbf{Scheme 9.2.} The most promising ligands considered in this study with their corresponding estimated binding energies in kJ/mol (values between brackets). The coordinated atoms are colored in blue.
Similar to other Zn complexes, ThrRS is energetically favored to have an overall neutral charge for the complexation.\textsuperscript{25} One of the reasons for the preference of the pentacoordinate Zn over the tetracoordinated one is that the overall charge of the complex is neutral only in the pentacoordinate form. Subsequently, the proposed library of ligands in this study contains ionized heteroatoms allowing for the formation of a pentacoordinated metal ion, Scheme 9.2. On the other hand, it is well established that aaRSs play a critical role in preserving the homochirality of the synthesized protein.\textsuperscript{36} It is possible that the active site has editing activity against D-amino acids and accommodates only the L-form.

Moreover, ThrRS maintains this enantioselectivity by harboring a separate editing domain called amino acid deacylases (DTD). Its role is to hydrolyze any misacylated D-threonyl-tRNA\textsuperscript{Thr}.\textsuperscript{37} Accordingly, we considered two different enantiomers of the threonine analogue in the studied model (1a and 1b) and the B.E. for each model has been calculated. Notably, the B.E. values vary according to the studied enantiomer and the most preferred configuration that gave the highest B.E. is (R)-1-amino-2-propoxide (1a) with value of 35.5 kJ/mol higher than the other conformation (1b), Scheme 9.2. In order to better describe this phenomenon, two additional ligands, with chirality \(\alpha\) to the amine rather than the alcohol, have been examined (1c and 1d). Importantly, various B.E. values have been observed with preference of the S-enantiomer, 1c in Scheme 9.2.

As previously discussed, the synthetic active site of ThrRS cannot adequately distinguish between threonine and the structurally similar serine even with the presence of Zn(II) as a cofactor. Consistent with this observation, we also investigated the binding of a serine analogue (2-aminoethanol), 1e, to the Zn(II) complex. The resulting B.E. is found to be slightly lower than the corresponding value of the threonine analogue with a small
difference of 2.4 kJ/mol. In fact, it is documented that the difference in B.E. between the two amino acids is only 4.2 kJ/mol.\textsuperscript{38} Arguably, the subtle added donating ability of the extra methyl group in the threonine analogue enriches the electron density on the chelating oxygen atom and subsequently strengthens its ability to bind to the Zn(II) atom.

Importantly, a comparison between 1c and 1a complexes showed that the methyl group attached to an $\alpha$-NH$_2$ group has less impact on the obtained B.E. unlike when it is positioned next to the O\textsuperscript{-} atom. In agreement with that, inserting an extra methyl group on to 1a close to the $\alpha$-NH$_2$ group, ligand 2a, does not enhance the binding capability, and thus has nearly no influence on the estimated B.E. (-515.6 and -514.3 kJ/mol for 1a and 2a). However, adding an NH$_2$ group to the chiral atom in ligand 2d, with it inductive ability as a donor group, greatly improved the binding efficacy and resulted in a B.E. of -524.1 kJ/mol, \textbf{Scheme 9.2}. Generally, tuning the binding potency has been achieved by changing the electronic properties of the attached substituents and electron donating substituents are energetically favored. Furthermore, the binding of a $\beta$-HNV analogue has been studied, 2b, and the observed B.E is considerable with a value of -513.5 kJ/mol. This B.E. value is 2.1 kJ/mol less than the cognate threonine analogue, reasonably indicating $\beta$-HNV can be recognized in the active site of ThrRS.\textsuperscript{22}

Meanwhile, valine can be rejected easily from this active site due to its inability to bind to the Zn(II) in a similar fashion to the cognate threonine.\textsuperscript{39} Based on our analysis on the binding of valine, it coordinates only through its $\alpha$-NH$_2$ and this results in tetracoordinated Zn(II) ion. Importantly, it exhibits extremely low B.E. with value of just $\sim$44.0 kJ/mol indicating an exceptionally weak binding affinity to the Zn(II) atom, matching experimental observations.\textsuperscript{23, 39} According to a previous related study,\textsuperscript{28} another
set of asymmetric conjugated ligands have been analyzed, 2e, 2f and 2g. The B.E. increased by increasing the electron donating capability of the substituent group, the ligand 2g giving the most favorable B.E. with value of -498.6 kJ/mol. However, all three ligands displayed a weaker binding ability relative to the B.E. of the threonine analogues.

In light of these findings, different sets of ligands have been studied, 3a-3g in Scheme 9.2. Overall, unlike the previous study that demonstrated the remarkable binding ability of aromatic ligands, our work showed aliphatic cyclic rings have a better binding ability. By comparing between 3b and 3e, both 6-membered ring ligands, we obtained a difference of 87.3 kJ/mol in the estimated B.E. with the aliphatic ligand much preferred. The reason for the weak interaction of aromatic ligands is partially due to the involvement of the lone pairs of the bonded heteroatoms in the conjugated system. Also, the binding strength of 3b improved by 16.9 kJ/mol to -431.8 kJ/mol after inserting a N atom close to the bonded oxygen atom, ligand 3a in Scheme 9.2. Importantly, we also considered the two aromatic ligands that gave the highest B.E. to histone deacylase, 28 3c and 3d. According to our calculations, the proposed aliphatic rings (3e, 3f and 3g) exceeded these ligands in terms of their binding capability, Scheme 9.2.
Figure 9.1. Optimized complexes with key bond lengths in Å of the binding geometries of selected ligands to the Zn(II) coordination sphere.

Furthermore, among the three different sizes of the aliphatic cyclic ligands studied, 6-membered cyclic ligands have the most favorable B.E., probably due to their less strained geometry. Globally, aliphatic chains are still energetically preferred with respect to the values of the B.E.. The optimized complexes for selected ligands together with some key distances are shown in Figure 9.1.

Two additional sets of compounds were also examined, 4a-4g and 5a-5f. Our goal was to determine which atom has the strongest binding affinity. In this series of ligands, Scheme 9.2, comparisons between the type of electron donors in the dative bond (amine, alkoxide, sulfoxide) and the electron donating/withdrawing ability of the substituents
(hydrogen, amine, cyano) were performed. These compounds were run as neutral and dianionic ligands. In case of the neutral ligands, 4a, 4b and 4c, ligand 4b with two oxygen atoms showed the highest binding affinity followed by 4c with terminal sulfur atoms. Interestingly, relative to the neutral forms, the B.E. of the corresponding ionized ligands, 4d, 4e and 4g, increased drastically due to having two anionic terminal atoms. The dianionic species in general showed extraordinary binding ability (almost twice the B.E. value of the threonine analogue). Notably, the most favorable binding ability was found in the dianionic species with two oxygen donors, ligand 4g, Scheme 9.2.

In agreement with the former findings, push-pull substituents induce a change in the value of the estimated B.E.. Furthermore, when the strong electron withdrawing cyano group (–CN) was placed on the two main carbons in different combinations, a considerable reduction in the B.E. has been noticed. Placing two CN groups has a greater impact on decreasing the value of the B.E. than a single CN group, as it is explicit in 5d and 5a. Also, a comparison between 5a and 4b demonstrates the influence of single CN group on the chelating ability. In case of 4b, the B.E. is slightly diminished relative to 5a due to the presence of the NH₂ group with the latter energetically favored. It is also interesting to note the higher B.E. values for the ligands with no EWGs, 4d, 4e and 4f, compared to the structures with one or more EWGs 5a-5e, Scheme 9.2.

9.3.2 Docking Results

We performed this investigation considering only the most favored warheads based on DFT calculations. The phosphate group in Thr-AMP is replaced by a sulfamoyl group to link the ligand to the adenosine moiety. The ligand is surrounded by a net of hydrogen bonds with Arg363 residue as well as the adenosine fragment is replaced by indazole. All
four ligands studied, 1a, 4c, 4e and 4d bound properly in the active site. More specifically, the adenosine analogue is sandwiched through different hydrogen bond interactions, the most conserved one with the carboxylic group of Glu365.

Figure 9.2. Cocrystal structures of selected ligands docked in the active site of *E. coli* ThrRS. The estimated values of the B.E. in kJ/mol are shown in blue.
Importantly, all the ligands chelate to the Zn(II) ion in a similar fashion to the cognate threonine, **Figure 9.2.** The calculated binding free energy indicated that the proposed ligands are more potent than the native threonine. Furthermore, in agreement with our DFT calculations, 4d is the most potent ligand having the highest binding energy of -20.7 kJ.

9.3.3 QM/MM calculation

In previous work by our group\(^{40}\), we performed a detailed QM/MM investigation on the aminoacylation mechanism of threonine in ThrRS using the same DFT functional and basis set. Using the same chemical model, we expanded the current work to explore the binding efficiency of three suggested candidates using a QM/MM model. Specifically, we replaced the native substrate threonine with the three ligands that showed the best binding energies, 4c, 4f and 4g. After performing full optimization of the new QM/MM models, we noticed that the three ligands chelate to Zn(II) through the two heteroatoms, **Figure 9.3.** Similar to the native threonine, Zn(II) is pentacordinate and the two systems showed remarkable stability.

**Figure 9.3.** Optimized QM/MM models for the catalytic site of ThrRS with the bound substrates, threonine(1a-2), key residues considered in the QM layer have been hided for clarity.
9.4 Conclusion

Based on our DFT calculations, it was observed that the B.E. is a notable indicator of the discrimination mechanism that takes place in the synthetic site of ThrRS against the structurally related serine, valine, and βHNV amino acids. Moreover, the binding capability is strongly related to the chirality and the basicity of the ligated atoms. Relative to the neutral ligands, the deprotonated ligands display a remarkable binding ability as demonstrated by their corresponding B.E. values. Furthermore, the B.E. of the ligands with two deprotonated heteroatom terminals dramatically increased, with the symmetric one preferred. This understanding guided us to propose a series of potent ligands that could bind more efficiently to Zn(II) relative to the native threonine.

Additionally, to verify the reliability of the proposed potent ligands, we performed molecular docking as well as QM/MM investigations on the selected ligands, further supporting our preliminary results. Overall, our findings inform about the essential characteristics required for competitive inhibition against bacterial ThrRS and will hopefully inspire development of novel inhibitors.

9.5 References


Chapter 10.
Conclusions and Further Work
10.1 Conclusions

Using a variety of computational enzymology protocols, we have successfully clarified the mechanisms of the aminoacylation and editing reactions catalyzed by the fundamental aminoacyl-tRNA synthetases and valuable atomistic details regarding their chemistry have been identified.

In chapter 3, we used MD and QM/MM methods to examine the role played by the Zn(II) ion and active site residues in ThrRS's ability to discriminate between its cognate substrate L-threonine, and the non-cognate L-serine, L-valine and D-threonine. The present results suggest that a role of the Zn(II) ion, with its Lewis acidity, is to facilitate deprotonation of the side chain hydroxyl groups of the aminoacyl moieties of cognate Thr-AMP and non-cognate Ser-AMP substrates. In their deprotonated forms these substrates are able to adopt a conformation preferable for aminoacyl transfer from aa-AMP onto the Ado-3'OH of the tRNA\textsuperscript{Thr} cosubstrate. Relative to the neutral substrates, when the substrates are deprotonated with the assistance of the Zn(II) ion the barrier for the rate-limiting step is decreased significantly by 42.0 and 39.2 kJ/mol for L-Thr-AMP and L-Ser-AMP, respectively. An active site arginyl residue also plays a key role in stabilizing the build-up of negative charge on the substrate's bridging phosphate oxygen during the mechanism. For the enantiomeric substrate, D-Thr-AMP, product formation is highly disfavoured and as a result, the reverse reaction has a very low barrier of 16.0 kJ/mol.

In chapter 4, the aminoacylation mechanisms of glutaminyl-tRNA synthetase (GlnRS) and the non-discriminating glutamyl-tRNA synthetase (ND-GluRS) have been investigated by employing MD simulation, QM-cluster and QM/MM calculations. Our investigations demonstrated the feasibility of a water-mediated, substrate-assisted catalysis pathway with
rate limiting steps occurring at energy barriers of 25.0 and 25.4 kcal/mol for GlnRS and ND-GluRS, respectively. A conserved lysine residue participates in a second proton transfer to facilitate the departure of the adenosine monophosphate (AMP) group. Thermodynamically stable (−29.9 and −9.3 kcal/mol for GlnRS and ND-GluRS) product complexes are obtained only when the AMP group is neutral. This substrate-assisted mechanism may be common to the structurally similar aspartyl-tRNA synthetase (AspRS) and asparaginyl-tRNA synthetase (AsnRS).

In chapter 5, we have explored multiple possible post-transfer editing mechanisms for ThrRS from *Escherichia coli*. The editing site is known to contain two conserved histidyl's (His73 and His186) and a cysteiny1 (Cys182), all of which could act as the required mechanistic base. We have performed detailed molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) studies in which the protonation states of each of these residues was varied. Furthermore, using the various substrate-bound active site models obtained, we have examined previously proposed and alternative possible mechanisms for deaminoacylation of Ser-tRNA<sup>Thr</sup> by ThrRS in which His73 or Cys182 act as the base; 11 mechanisms in total. The present results suggest that the most feasible mechanism is obtained when both His73 and His186 are neutral, while the thiol of Cys182 is deprotonated and acts as a base. The resulting reaction was found to proceed in two steps. First, deprotonation of an active site water by the thiolate of Cys182 with its concomitant nucleophilic attack at the substrate's C<sub>carb</sub> center occurs with a calculated free energy barrier of 9.9 kcal/mol. The subsequent, and overall rate-limiting step, is a water-mediated proton transfer from Lys156 to the Ado<sub>76</sub>3′-oxygen resulting in simultaneous cleavage of the Ado<sub>76</sub>3′O—C<sub>carb</sub> bond with a free energy barrier of 20.8 kcal/mol.
Chapter 10

In contrast to bacterial ThRS’s editing site, discussed in chapter 5, in archaeal ThrRS, it has been suggested that the substrate’s 2’OH_{Ado76} is the only available mechanistic base, which is the focus of the investigations reported in chapter 6. The suggested role of this 2’OH group was to promote the reaction by orienting a nucleophilic water molecule close to the scissile Ser-tRNA^{Thr} ester bond. In this study, we performed an extensive computational investigation, using both Molecular Dynamics (MD) and hybrid ONIOM Quantum Mechanics/Molecular Mechanics (QM/MM) methods, to consider all possible editing mechanisms. The results of our investigation show that the 2’OH_{Ado76} group plays a crucial role; the editing mechanism was found to proceed step-wise via the formation of 6-membered ring transition structure. Moreover, the elucidated mechanism was found to be applicable for the d-amino acid deacylase activity that is exhibited by the same editing domain. Three different functionals were considered in this study and the M06-HF functional was found to give the most enzymatically feasible energy barriers.

In chapter 7, we used MD, QM/SE and QM methodologies to provide atomistic details regarding the pre-transfer editing against the non-cognate homocysteine (Hcys) by isoleucyl-tRNA synthetase (IleRS) as well as cysteine (Cys) and threonine (Thr) by seryl-tRNA synthetase (SerRS). Notably, in the two enzymes considered herein, pre-transfer editing follows a substrate-assisted self-cyclization mechanism. This mechanism is found to take place via two main steps, a dihedral scan around the substrate’s C_{β}–C_{γ} bond followed by a concerted R-S(O)H deprotonation and nucleophilic attack on the substrate’s carbonyl carbon. Notably, the rate limiting step to edit against Hcy-AMP by IleRS is the first step with an activation barrier of 24.3 kcal/mol. Meanwhile, the highest energy barrier obtained during the editing against Cys and Thr by SerRS resulted from the second step
Chapter 10

with values of 20.4 and 26.6 kcal/mol, respectively. Interestingly, following the same pathway to investigate the possibility of pre-transfer editing against the native Ser-AMP by SerRS, an energy barrier of 31.4 kcal/mol was obtained indicating an enzymatically infeasible process. Additionally, an agreement between the relative free energy obtained by QM/SE with the corresponding values obtained by QM-cluster one was observed.

In chapter 8, using MD and QM/MM methodologies we elucidated the escape from the fidelity by the unnatural β-Hydroxynorvaline (βHNV), in the aminoacylation site of ThrRS. Due to the presence of Zn(II) with its Lewis acidity character, only the ionized form of βHNV results in an enzymatically feasible barrier. Furthermore, consistent with the homochiral behavior of this active site, we observed that there is a specific conformation of βHNV that could be aminoacylated. Benchmarking analysis was performed and our system was found to be responsive to the %HF as well as the dispersion correction included in the used density functional theory. Importantly, the BP86-D3 functional was found to be superior to all the studied functionals in representing the rate-limiting step kinetically. The importance of Zn(II) in the aminoacylation mechanism was further emphasized when its substitution with the chemically similar Cd(II) led to a dramatic increase in the energy values and thus, less feasible from both thermodynamic and kinetic perspectives.

In chapter 9, detailed DFT quantum chemical study was conducted to examine the binding ability of various ligands to Zn(II) in ThrRS by comparing the value of the binding energy (B.E) for each ligand relative to the native substrate, threonine. Our screening investigation showed that the native threonine should ligate in a bidentate fashion to this Zn(II) which lead to the highest B.E. Thereby, the synthetic site of ThrRS rejects noncognate amino acids that cannot perform this type of interaction. Moreover, based on
their ligation to the Zn(II) and the obtained B.E values compared to the cognate threonine, many potent ligands have been suggested. Importantly, ligands with deprotonated warheads showed the highest binding ability. Further investigation on the selected ligands using molecular docking and QM/MM confirmed our initial findings of the suggested competitive ligands being able to bind efficiently in the active site of ThrRS. The suggested ligands from this study are potential candidates for competitive inhibitors against bacterial ThrRS.

This research should enable mechanistic investigations on the rest of the crucial family of aminoacyl-tRNA synthetases. Several of the elucidated catalytic mechanisms can be generalized to other members of this family, and thereby future work will be to explore applicability of such mechanisms. Moreover, many other questions need further analysis, including the first step of activation mechanism as well how each enzyme recognizes its cognate tRNA. Eventually, these findings should provide significant insights into the outstanding chemical fidelity of these enzymes. Consequently, this understanding should open many novel routes to approach our long-term goal in establishing collaboration with many experimental leaders who are currently working on the biochemical analysis of aminoacyl-tRNA synthases.
VITA AUCTORIS

Name: Mohamed Mohamed Wahba Aboelnga

Place of birth: Damietta, Egypt

Year of birth: 1986

Education

Ph.D. Chemistry                          October 2017
University of Windsor                  Windsor, Ontario, Canada

M.Sc Chemistry                          Septembre 2011
Damietta University                   Damietta, Egypt

B.Sc Chemistry                          June 2007
University of Mansoura                 Damietta, Egypt