COMPUTATIONAL STUDY OF STRUCTURE AND MECHANISM OF *TRYPANOSOMA*
*cruci* TRANS-SIALIDASE

By

ÖZLEM DEMİR

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008
To my parents, Asiye and Mehmet, and my sister, Gülçin.
ACKNOWLEDGMENTS

I first would like to thank my supervisor Adrian Roitberg for all his continuous support, kind guidance, strong encouragement and positive feedback throughout my graduate studies. His trust, patience and motivation accompanied me all the way, even at the most challenging times. I also wish to thank Nicole Horenstein for valuable discussions and endless support. I extend my thanks to all my committee members Nicole Horenstein, Arthur Edison, Jeffrey Krause, Alexander Angerhofer and former member Nigel Richards for accompanying me in the important stages of my graduate career and all their contributions.

Special thanks go to former and present group members Christina Crecca, Seonah Kim, Gustavo Seabra, Georgios Leonis, Hui Xiong, Yilin Meng, Lena Dolghih, Jason Swails and Dan Sindhikara for all their support. I also want to acknowledge Erik Deumens for his excellent technical help whenever needed and all past and present QTP members and staff for providing a nice and warm environment.

This dissertation, like many other nice things in my life, would not be possible without the continuous support of my dear family. I have always felt their encouragement and trust on me and I want to thank each of them for all the joy and happiness they brought to my life. It is a place where words are insufficient to express my feelings.

It is also hard to acknowledge enough my dear friends in Gainesville who have converted the period I have been here to precious times to remember. I would like to thank Evrim Ataş and Avni Argun for their welcoming for all times, continuous support and friendship. Next but not least, I want to thank Asuman and Fatih Gördü for being like a second family to me here in Gainesville. It is hard to forget all the good times we shared with dear friends Emine Demir, Esra Büyükahtakın, Aysun Altan, Ufuk Koca, Tezcan Özrazgat, Emine Dedeoğlu, Tuba Yavuz-Kahveci, Yasemin Alptekinoğlu, Dilber and Dilek Dağdelen, Fatma Aslan-Tutak, Zeynep Çopur,
Zeynep İrem Atalay, Semra Ağralı, Sevnur Kömür thụ, Ece Ünür and Pamela Monterola. Special thanks go to Uğur Başlantı, Enes Çalışk, Yavuz Yağız, Zafer Demir, Bilge Tutak and Murat Keçeli for all their support and friendship.

The time I spent here would not be as valuable without my long-term friends who have been there for me all the time although we were separated by very long distances. Thus, I appreciate and celebrate sincere friendships of Tuba Gülbağcı, Olga Samarskaya, Özgür Çelik, Ebru Yumak, Ülkü Yaylalı, Ülkü Balcı, Mine and Abdullah Özer.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................................................................................. 4  
LIST OF TABLES .............................................................................................................................................. 9  
LIST OF FIGURES ............................................................................................................................................ 10  
ABSTRACT ................................................................................................................................................... 15  

## CHAPTER

### 1 INTRODUCTION ......................................................................................................................................... 17  
1.1. Prologue ................................................................................................................................................ 17  
1.2. Sialic Acids .......................................................................................................................................... 17  
1.3. Chagas' Disease .................................................................................................................................. 19  
1.4. *Trypanosoma cruzi* ............................................................................................................................ 21  
1.5. *Trypanosoma cruzi* trans-Sialidase (TcTS) and *Trypanosoma rangeli* Sialidase (TrSA) ................ 22  
1.5. Structural Information about *Trypanosoma cruzi* trans-Sialidase and *Trypanosoma rangeli* Sialidase ................................................................. 26  
1.6. Mechanistic Information about *Trypanosoma cruzi* trans-Sialidase and *Trypanosoma rangeli* Sialidase ......................................................................................... 37  

### 2 METHODS .................................................................................................................................................. 44  
2.1. Quantum Mechanics ............................................................................................................................. 44  
2.2. Molecular Mechanics ............................................................................................................................ 47  
2.3. Hybrid Methods ..................................................................................................................................... 49  
2.4. Methods for Exploring the Potential Energy Surface ........................................................................ 51  
2.5. Molecular Dynamics Simulation Methods .......................................................................................... 56  

### 3 QUANTUM MECHANICS/ MOLECULAR MECHANICS STUDY OF THE CATALYTIC MECHANISM OF *TRYPANOSOMA cruzi* TRANS-SIALIDASE AND *TRYPANOSOMA rangeli* SIALIDASE USING ONIOM ......................................................................................... 59  
3.1. Introduction ............................................................................................................................................. 59  
3.2. Methods ................................................................................................................................................ 61  
3.2.1. Relaxation of the TcTS and TrSA Systems ..................................................................................... 61  
3.2.2. Preparation of the TcTS and TrSA Model Systems ....................................................................... 62  
3.2.3. ONIOM Input Setup ......................................................................................................................... 63  
3.2.4. Potential Energy Surface Scan Setup ............................................................................................. 64  
3.2.4.1. Simulating the trans-sialidase reaction of TcTS in a $S_N$2 way ..................................................... 64  
3.2.4.2. Simulating the trans-sialidase reaction of TrSA in a $S_N$2 way ..................................................... 65  
3.2.4.3. Simulating the trans-sialidase reaction of TcTS in a $S_N$1 way ..................................................... 65
3.2.4.4. Simulating the sialidase reaction of TcTS ................................................................. 67
3.3. Results and Discussion .................................................................................................... 67
  3.3.1. The SN2-like trans-Sialidase Reaction of TcTS ........................................................ 67
  3.3.2. The SN2-like trans-Sialidase Reaction of TrSA ......................................................... 71
  3.3.3. The SN1-like trans-Sialidase Reaction of TcTS ......................................................... 73
  3.3.4. The Sialidase Reaction of TcTS .............................................................................. 75
3.4. Conclusion ...................................................................................................................... 79

4 QUANTUM MECHANICS/ MOLECULAR MECHANICS STUDY OF THE
CATALYTIC MECHANISM OF TRYPANOSOMA cruzi TRANS-SIALIDASE USING
AMBER .................................................................................................................................. 80
  4.1. Introduction .................................................................................................................. 80
  4.2. Methods ....................................................................................................................... 80
    4.2.1. Model Preparation ............................................................................................... 80
    4.2.2. Initial Minimization .............................................................................................. 81
    4.2.3. QM/MM Potential Energy Surface Preparation .................................................. 81
  4.3. Results and Discussion ............................................................................................... 82
  4.4. Conclusion .................................................................................................................. 85

5 MOLECULAR DYNAMICS STUDY OF TRYPANOSOMA cruzi TRANS-
SIALIDASE AND TRYPANOSOMA rangeli SIALIDASE ................................................... 88
  5.1. Introduction .................................................................................................................. 88
  5.2. Methods ....................................................................................................................... 90
    5.2.1. System Setup and Simulation Details .................................................................... 90
    5.2.2. Analysis Methods ................................................................................................. 92
  5.3. Analysis ....................................................................................................................... 95
    5.3.1. Stability of the Molecular Dynamics Simulations ................................................ 95
    5.3.2. Root Mean Square Fluctuation (RMSF) Analysis of MD Simulations of
          TcTS and TrSA ......................................................................................................... 96
      5.3.2.1. Comparison of RMSF changes due to covalent intermediate formation
              in TcTS and TrSA .............................................................................................. 96
      5.3.2.2. Comparison of RMSF between DANA-bound TcTS and TrSA
              simulations ........................................................................................................ 98
      5.3.2.3. Comparison of RMSF between wild-type TcTS, wild-type TrSA and
              TrSA5mut simulations ......................................................................................... 98
      5.3.2.4. Comparison of RMSF between sialyllactose-bound TcTS and
              sialyllactose-bound TrSA .................................................................................. 98
      5.3.2.5. Comparison of experimental B-factors of free and covalent
              intermediate forms of TcTS and TrSA ............................................................. 99
      5.3.2.6. Comparison of experimental B-factors between DANA-bound TcTS
              and TrSA X-ray crystal structures .................................................................... 99
      5.3.2.7. Comparison of experimental B-factors between wild-type TcTS, wild-
              type TrSA and TrSA5mut X-ray crystal structures ......................................... 99
      5.3.2.8. Comparison of RMSF values from MD simulations and B-factors from
              X-ray crystal structures .......................................................................................103
5.3.3. Comparison of Average Structures from MD Simulations of TcTS in Ligated and Unligated Forms

5.3.3.1. Comparison of unligated TcTS with TcTS covalent intermediate average structures
5.3.3.2. Comparison of unligated TcTS with DANA-bound TcTS average structures
5.3.3.3. Comparison of unligated TcTS with sialyllactose-bound TcTS average structures

5.3.4. Comparison of Average Structures from MD Simulations of TrSA in Ligated and Unligated Forms

5.3.4.1. Comparison of unligated TrSA with TrSA covalent intermediate average structures
5.3.4.2. Comparison of unligated TrSA with DANA-bound TrSA average structures
5.3.4.3. Comparison of unligated TrSA with sialyllactose-bound TrSA average structures

5.3.5. Comparison of Sialyllactose-Bound TcTS and Sialyllactose-Bound TrSA Average Structures

5.3.6. Comparison of DANA-Bound TcTS with DANA-Bound TrSA Average Structures

5.3.7. Behavior of Key Residues of TcTS and TrSA in MD

5.3.7.1. Behavior of Trp312
5.3.7.2. Behavior of Tyr119
5.3.7.3. Behavior of Phe58
5.3.7.4. Behavior of Asp59
5.3.7.5. Behavior of Tyr342
5.3.7.6. Behavior of Glu230
5.3.7.7. Behavior of Asp96
5.3.7.8. Behavior of Leu36

5.3.8. Ligand-Enzyme Interactions in Ligated TcTS and TrSA MD Simulations

5.3.9. Interactions Observed Between the Enzyme Residues in the Active Site

5.4. Discussion

5.4.1. Sialyllactose-Bound Form of TrSA
5.4.2. Effects of the 5 Point Mutations on TrSA that Promote trans-Sialidase Catalysis
5.4.3. Comparison of DANA-Bound Forms of TcTS and TrSA
5.4.4. Comparison of Unligated and DANA-Bound Forms of TcTS
5.4.5. Comparison of the Effect of Covalent Intermediate Formation in TcTS and TrSA

CONCLUSIONS AND PERSPECTIVE

LIST OF REFERENCES

BIOGRAPHICAL SKETCH
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Sialidase and trans-sialidase activities of wild-type and mutant TcTS and TrSA</td>
<td>23</td>
</tr>
<tr>
<td>1-2</td>
<td>Inhibitor (DANA) binding properties of TcTS, TrSA and several TrSA mutants</td>
<td>31</td>
</tr>
<tr>
<td>1-3</td>
<td>Trans-Sialidase activity of various TrSA mutants</td>
<td>35</td>
</tr>
<tr>
<td>1-4</td>
<td>Kinetic constants for the sialidase activity of TrSA mutants</td>
<td>37</td>
</tr>
<tr>
<td>3-1</td>
<td>Results of RC3 scan for TcTS that show stable oxocarbenium ion formation</td>
<td>75</td>
</tr>
<tr>
<td>3-2</td>
<td>Results of RC1 scan for TcTS that show conversion of the oxocarbenium ion to the covalent intermediate</td>
<td>77</td>
</tr>
<tr>
<td>3-3</td>
<td>Results of RC3 scan that show a water molecule attacking the covalent intermediate in TcTS</td>
<td>78</td>
</tr>
<tr>
<td>4-1</td>
<td>Properties of stationary points on potential energy surfaces</td>
<td>83</td>
</tr>
<tr>
<td>5-1</td>
<td>Properties of simulated structures</td>
<td>93</td>
</tr>
<tr>
<td>5-2</td>
<td>Hydrogen-bond interactions of the ligands in covalent intermediate, DANA-bound and sialyllactose-bound TcTS</td>
<td>130</td>
</tr>
<tr>
<td>5-3</td>
<td>H-bond interactions of the ligands in covalent intermediate, DANA-bound and sialyllactose-bound TrSA</td>
<td>132</td>
</tr>
<tr>
<td>5-4</td>
<td>Hydrogen bonding interactions between the residues in the active site in MD simulations of TcTS species</td>
<td>134</td>
</tr>
<tr>
<td>5-5</td>
<td>Hydrogen bonding interactions between the residues that lie within 10 Å of Tyr342.OH atom in MD simulations of free and ligated TrSA species</td>
<td>135</td>
</tr>
<tr>
<td>5-6</td>
<td>Sequence information for different sialidases and trans-sialidases of <em>Trypanosomal</em> species</td>
<td>141</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Chemical structures of neuraminic acid, N-acetyl neuraminic acid and N-glycolyl neuraminic acid</td>
<td>18</td>
</tr>
<tr>
<td>1-2</td>
<td>A) Geographical distribution of Chagas’ disease. B) The insect that transmits Chagas’ disease: Triatominae C) Romana’s sign on the eye, which is a sign of Chagas’ disease</td>
<td>19</td>
</tr>
<tr>
<td>1-3</td>
<td>Different forms of <em>T.cruzi</em> in its life cycle</td>
<td>22</td>
</tr>
<tr>
<td>1-4</td>
<td>Transfer of sialic acids onto the <em>Trypanosoma cruzi</em> cell surface</td>
<td>23</td>
</tr>
<tr>
<td>1-5</td>
<td>Sequence alignment of TcTS and TrSA with amino acid differences highlighted</td>
<td>24</td>
</tr>
<tr>
<td>1-6</td>
<td>Structures of sialic acid oxocarbenium ion, DANA, oseltamivir, zanamivir and sialyllactose</td>
<td>27</td>
</tr>
<tr>
<td>1-7</td>
<td>Two different views from superimposed sialyllactose-bound TcTS and unligated TrSA</td>
<td>29</td>
</tr>
<tr>
<td>1-8</td>
<td>Active site of TcTS when sialyllactose is bound</td>
<td>29</td>
</tr>
<tr>
<td>1-9</td>
<td>Dual conformations of Tyr119</td>
<td>30</td>
</tr>
<tr>
<td>1-10</td>
<td>The two residues that affect Trp312 in TcTS: Pro283 and Tyr248</td>
<td>33</td>
</tr>
<tr>
<td>1-11</td>
<td>Sialidase and trans-sialidase activities of chimeric proteins of TcTS and TrSA</td>
<td>33</td>
</tr>
<tr>
<td>1-12</td>
<td>Different types of bisubstrate enzyme mechanisms</td>
<td>38</td>
</tr>
<tr>
<td>1-13</td>
<td>Structure of 2,3-difluoro sialic acid used for chemical trapping</td>
<td>41</td>
</tr>
<tr>
<td>1-14</td>
<td>The proposed mechanism for trans-sialidase catalysis reaction of TcTS</td>
<td>42</td>
</tr>
<tr>
<td>3-1</td>
<td>The covalent intermediate formation step</td>
<td>65</td>
</tr>
<tr>
<td>3-2</td>
<td>Reaction coordinates used to generate potential energy surfaces</td>
<td>66</td>
</tr>
<tr>
<td>3-3</td>
<td>Potential energy surface of TcTS for $S_N2$-like covalent intermediate formation step</td>
<td>69</td>
</tr>
<tr>
<td>3-4</td>
<td>Changes in geometric properties of sialic acid during $S_N2$-like covalent intermediate formation of TcTS</td>
<td>70</td>
</tr>
<tr>
<td>3-5</td>
<td>The potential energy surface of TrSA for $S_N2$-like covalent intermediate formation step</td>
<td>72</td>
</tr>
</tbody>
</table>
3-6 Changes in geometric properties of sialic acid during S_N2-like covalent intermediate formation of TrSA...........................................................................................................................................74
3-7 The mechanism proposed for trans-sialidase catalysis reaction of TcTS. ..................................76
3-8 Potential energy diagram calculated for trans-sialidase catalysis reaction of TcTS...........78
3-9 Potential energy diagram calculated for sialidase catalysis reaction of TcTS. .....................79
4-1 Reaction coordinates used to generate potential energy surfaces..............................................83
4-2 Potential energy surface for S_N2-like mechanism of TcTS covalent intermediate formation ..................................................................................................................................................84
4-3 Potential energy surface for S_N1-like mechanism of TcTS that shows a stable oxocarbenium ion formation..............................................................................................................................85
4-4 Potential energy surface for S_N1-like mechanism of TcTS that shows covalent intermediate formation from the oxocarbenium ion. .........................................................................................................................86
5-1 Corresponding active site residues of TcTS and TrSA that have at least one atom within 10 Å radius of Tyr342 hydroxyl O atom. ..............................................................................................................90
5-2 The regions of the TcTS and TrSA that are used for MD simulation and analysis............94
5-3 Description of ϕ, ψ, χ1 and χ2 dihedral angles of a residue in a protein........................94
5-4 Root mean square deviation (RMSD) of MD simulations with respect to the initial structure of the production runs. A) TcTS species B) TrSA species ........................................97
5-5 Cα RMSF comparison of free TcTS with TcTS covalent intermediate. ..............................100
5-6 Cα RMSF comparison of free TrSA with TrSA covalent intermediate. ..............................100
5-7 Cα RMSF comparison of free and DANA-bound forms of TcTS........................................101
5-8 Cα RMSF comparison of free and DANA-bound forms of TrSA........................................101
5-9 Cα RMSF comparison of TrSA, TcTS and TrSA5mut..........................................................102
5-10 Cα RMSF comparison of sialyllactose-bound forms of TcTS and TrSA.......................102
5-11 Comparison of crystal structure Cα B-factors of TcTS covalent intermediate with its free form......................................................................................................................................................104
5-12 Comparison of crystal structure Cα B-factors of TrSA covalent intermediate with its free form......................................................................................................................................................105
5-13  \( C_\alpha \) B-factor comparison of free and DANA-bound forms of TcTS X-ray crystal structures. .........................................................................................................................105

5-14  \( C_\alpha \) B-factor comparison of free and DANA-bound forms of TrSA X-ray crystal structures..........................................................................................................................106

5-15  \( C_\alpha \) B-factor comparison of TcTS, TrSA and TrSA\(_{5\text{mut}}\) X-ray crystal structures. ........106

5-16  Crystal contacts of TrSA around the loops bearing Trp312, Gln283 and Gly248 ..........107

5-17  Histograms of Trp312 \( \chi_1-\chi_2 \) populations in MD simulations for TcTS species.........112

5-18  Histograms of Trp312 \( \chi_1-\chi_2 \) populations in MD simulations for TrSA species.........113

5-19  RMSD of the loop that bears Trp312 in MD simulations without fitting to the initial frame. ........................................................................................................................................115

5-20  Correlation of RMSD of the backbone of Trp312’s loop without fitting to the initial conformation and the \( \chi_1 \) dihedral angle of Tyr119 during MD simulations. ..............116

5-21  Histograms of Tyr119 \( \chi_1-\chi_2 \) populations in MD simulations.................................118

5-22  Backbone RMSD for the loop bearing Tyr119/ Ser119 in MD simulations. ...............119

5-23  Phe58 \( \chi_1 \) dihedral angle change in MD simulations. ..................................................120

5-24  Asp59 \( \chi_1 \) change during MD simulations. .................................................................123

5-25  RMSD of the loop bearing Asp59 in MD simulations without fitting ........................124

5-26  Histograms of \( \chi_1-\chi_2 \) populations of Tyr342 in TcTS MD simulations .........................125

5-27  Histograms of \( \chi_1-\chi_2 \) populations of Tyr342 in TrSA MD simulations .........................126

5-28  \( \Phi \) dihedral angle of Ala341/Gly341 in MD simulations..............................................127

5-29  Histograms of \( \chi_1-\chi_2 \) populations of Leu36 in TcTS MD simulations. .......................128

5-30  Histograms of \( \chi_1-\chi_2 \) populations of Ile36 in TrSA MD simulations. .........................129

5-31  Hydrogen bonding and steric interactions in sialyllactose-bound TcTS. ....................131

5-32  Hydrogen bonding and steric interactions in sialyllactose-bound TrSA. ......................133

5-33  Different conformations of Trp312 seen in TcTS and TrSA. ......................................137

5-34  Behavior of Trp312 and its loop backbone in MD simulations of A) modeled sialyllactose-bound TrSA and B) sialyllactose-bound TcTS.........................................................138
5-35 Close crystal contacts in the unit cell of unligated TrSA.................................139
5-36 Binding mode of sialyllactose in A) TcTS and B) TrSA......................................141
5-37 A view from superimposed DANA-bound forms of TrSA and TcTS average structures of the MD simulations.............................................................................................................145
5-38 A view from superimposed DANA-bound TrSA and TcTS average structures of the simulations.............................................................................................................146
LIST OF ABBREVIATIONS

DANA 2, 3-dehydro-3-deoxy-N-acetylneuraminic acid
PES Potential energy surface
RMSD Root mean square deviation
RMSF Root mean square fluctuation
TcTS Trypanosoma cruzi trans-sialidase
TrSA Trypanosoma rangeli sialidase
Trans-sialidase is a vital enzyme for the lifecycle of Trypanosoma cruzi, the protozoa responsible for Chagas’ disease, which is lethal and drastically affects large human populations in Central and South America—widening its epidemic area to North America in recent years. T. cruzi trans-sialidase (TcTS) catalyzes transfer of sialic acids from host glycoconjugates to the parasite’s glycoconjugates, which facilitates the parasite the means to escape from the host immune system and to invade the host cells. Thus, TcTS stands as a potential and appealing therapeutic target for Chagas’ disease.

Experimental evidence suggests that a relatively long-lived covalent intermediate forms in the mechanism of TcTS. If this scenario is correct, sialic acid is scavenged from the host’s glycoconjugates and stays bound to the enzyme until the parasite’s glycoconjugate enters the active site. However, it is unclear whether the covalent intermediate formation occurs through an $S_N1$ or $S_N2$ mechanism. It is crucial to elucidate the mechanism and the transition structure for future inhibitor design studies of TcTS.

Additionally, the common inhibitors for sialidases, which catalyze hydrolysis of sialic acids, do not work for TcTS. The reason for this is unclear since both enzyme families share the first step of the mechanism. Trypanosoma rangeli sialidase (TrSA) stands out among sialidases
to perform a comparative study with TcTS due to their distinct structural similarity (%70 sequence identity and Ca RMSD of 0.59 Å) and yet, different catalytic function. There is experimental evidence about formation of a covalent intermediate in TrSA as well, but only for an activated ligand. Thus, there is a possibility that the mechanism of TrSA is artificially biased towards covalent intermediate formation due to the effect of substituents on the natural ligand. Elucidating the difference in mechanisms of TcTS and TrSA could also pave the way to tailor sialidases into trans-sialidases (and glycosidases into trans-glycosidases) to use for efficient synthesis of molecules that currently require long and low-yield chemical processes.

In this study, the mechanisms of both enzymes are investigated using two different QM/MM methods in Chapters 3 and 4. Potential energy surfaces are constructed for each enzyme by performing constrained minimizations. Based on the potential energy surfaces, the difference in the mechanisms of the two enzymes is discussed.

Furthermore, 50-ns long molecular dynamics simulations are performed for the two enzymes in free, ligand-bound and inhibitor-bound forms and these simulations are analyzed thoroughly in Chapter 5 to distinguish any structural or dynamical differences between the two enzymes and to shed light on the reason of difference in their inhibitor binding ability.
CHAPTER 1
INTRODUCTION

1.1. Prologue

In this chapter, sialic acids, their properties and importance will be introduced first. Then, Chagas’ disease, one of the many serious diseases that involve sialic acids and the key enzymes responsible for sialic acid catabolism, will be described. Brief information about Trypanosoma cruzi, which is the causative agent of Chagas’ disease, will follow. Subsequently, the main focus of this study, the trans-sialidase enzyme of Trypanosoma cruzi will be introduced and contrasted to a structurally very similar enzyme, Trypanosoma rangeli sialidase. In the following two sections, current information about structure and mechanism of these two enzymes will be provided in detail. And the last section will introduce our approach to elucidate the present structural and mechanistic problems of these two enzymes.

1.2. Sialic Acids

Sialic acid is the common name for O- and N-substituted derivatives of a nine-carbon monosaccharide called neuraminic acid (5-amino-3, 5-dideoxy-D-glycero-D-galacto-2-nonulosonic acid) (Figure 1-1). Derivatives of 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid (KDN) which have a hydroxyl group at C5 position instead of the amino group of neuraminic acid are also included in this group recently. Sialic acids were discovered in 1930s and both names, sialic acid and neuraminic acid, reflect the origins of their first isolation; one was isolated from submaxillary mucin (sialos=saliva in Greek) and the other from brain glycolipids (neuro + amine + acid).^1^ More than 50 different sialic acids found in nature indicate the remarkable structural diversity—provided by substituting the amino group at C5 position with acetyl or glycolyl groups and/or substituting the hydroxyl groups at C4, C7, C8 and C9 positions with acetate, lactate, sulfate or phosphate esters, or by methyl ethers.
Sialic acids are found in vertebrate animals, a few invertebrates and microorganisms. They lie at the terminal positions of polysaccharide chains, especially in cell membranes, and are used for recognition purposes by the immune system. Since sialic acids are strongly acidic with pKa values in the range 2.0-2.6, they are negatively charged under physiological conditions. Due to their terminal position, sialic acids can both act as masking recognition sites and, conversely, represent specific recognition sites themselves. Additionally, due to their negative charge, sialic acids help in exerting attractive/repulsive forces to cells and molecules, and in binding and transport of positively-charged molecules. Thus, sialic acids have a significant role in various cell functions, such as host-cell interactions like cell adhesion and invasion, and resistance to nonspecific complement attack.

![Chemical structures of neuraminic acid (Neu, R=H), N-acetyl neuraminic acid (Neu5Ac, R=CH3-CO—) and N-glycolyl neuraminic acid (Neu5Gc, R=HOCH2-CO—).](image)

Due to the importance of sialic acids, the sialidase superfamily, that catalyzes hydrolysis of sialo-sugars releasing sialic acids, includes enzymes that are implicated as virulence factors in pathogenesis of many different diseases. For example, the sialidases of avian influenza virus which causes bird flu, influenza virus which causes flu, paramyxovirus which causes respiratory disease, Vibrio cholerae which causes cholera, Trypanosoma cruzi which causes...
Chagas’ disease,\textsuperscript{5} \textit{Trypanosoma brucei gambiense} and \textit{Trypanosoma brucei rhodesiense}\textsuperscript{13} which cause sleeping sickness in humans, \textit{Trypanosoma brucei brucei}\textsuperscript{14-17} and \textit{Trypanosoma congoense}\textsuperscript{13,18} which cause nagana in cattle, all play important roles for the mentioned diseases.

1.3. Chagas’ Disease

Chagas’ disease, also known as American trypanosomiasis, is a lethal human disease mostly prevalent in Central and South America and recently shifting its endemic area to North America.\textsuperscript{19,20} Dating back to \(~9,000\) years according to the recent evidence,\textsuperscript{21} Chagas’ disease is estimated by WHO to have 17 million cases in 21 endemic countries with a death toll of 45,000 people per year due to the cardiac form of the disease (Figure 1-2.A) and listed as one of the fourteen neglected tropical diseases (www.who.int/TDR).

![Figure 1-2. A) Geographical distribution of Chagas’ disease. B) The insect that transmits Chagas’ disease: Triatominae. C) Romana’s sign on the eye, which is a sign of Chagas’ disease. (Adapted from www.who.int/TDR)](image)

The disease was first described by Dr. Carlos Chagas in 1909 and the causative agent of the disease was identified to be \textit{Trypanosoma cruzi}.\textsuperscript{22} Although first observed mostly in rural areas of South America, the disease reached the cities and even to other continents (North America) due to large numbers of people emigrating from the endemic countries over the years.\textsuperscript{19} Unfortunately, there is still no vaccine or effective cure after a century and the disease remains as a major medical problem in South America threatening a very large population. The most important precautions for this disease are improving the conditions of house—rendering them unavailable for colonization of the Triatominae—and performing blood-screening before any blood transfusions.
The current drugs, nifurtimox and benznidazole, are widely used in the acute phase of the disease but not in the chronic phase.\textsuperscript{23} In addition to having severe side effects including anorexia, loss of weight, psychic alterations, excitability, sleepiness, nausea, skin manifestations like hypersensitivity, depression of bone marrow, these drugs harm the adrenal glands, oesophagus, colon and reproductory and mammary tissues.\textsuperscript{23} They also show neurotoxicity, mutagenetic and carcinogenic effects.\textsuperscript{23} All these side effects generally cause discontinuation of their usage by the patients. Thus, effective drugs for Chagas’ disease are urgently needed.

Chagas’ disease can be transmitted to humans by several known ways:

1. Being bitten by bloodfeeding \textit{Triatominae} vectors (Figure 1-2.B)
2. Transfusion of infected blood\textsuperscript{24}
3. Transplant of an infected organ\textsuperscript{25,26}
4. Congenitally, from mother to the baby\textsuperscript{27}
5. Breastfeeding\textsuperscript{28}
6. Using food or drinks contaminated with crushed \textit{Triatominae} vectors\textsuperscript{29}

The first one is the most common way of transmission responsible for 80-90\% of all cases.

Six to ten days after infection, the clinical signs of immediate acute stage of the disease start to appear and these can continue up to 2 months.\textsuperscript{30} The acute stage typically causes oedema at the infected location—called Romana’s sign if at the eyelid (Figure 1-2.C)—due to rapid reproduction of causative agent of the disease and the lack of the immune response. It is also possible that no noticeable sign of this stage is seen. The acute stage is over once a balance between the host and the parasite is reached due to the immune system reducing the number of parasites in the circulatory system. The subsequent chronic stage continues life-long and can seriously damage the heart, oesophagus, colon and the nervous system.\textsuperscript{30} Megaoesophagus, megacolon, heart failure and sudden death are often seen among the patients. The infected people and vertebrates serve as a potential reservoir since \textit{Triatominae} biting them can transmit the parasites to their future victims.
1.4. *Trypanosoma cruzi*

The causative agent of Chagas’ disease, *Trypanosoma cruzi*, is a flagellatae protozoan. The lifecycle of this parasite involves vertebrate hosts and bloodfeeding insects of Triatominae (a.k.a. kissing bugs), which is a subfamily of *Reduviidae*. The most important of the species that transmit Chagas’ disease are *Triatoma infestans*, *Rhodnius prolixus*, *Triatoma dimidiata*, *Triatoma brasiliensis* and *Panstrongylus megistus*. Most species of *Triatominae* live on wild nesting vertebrates and only a few species are associated with humans and domestic animals. These insects feed at night when the hosts are asleep and they leave their feces that include *T. cruzi* parasites around the bite wound during or right after getting the bloodmeal from the host. The parasites are carried into the body through the wound when the host scratches the wound due to intensive itching.

*T. cruzi* is in epimastigote form when it first enters the insect vector by way of a bloodmeal from an infected host (Figure 1-3). In this stage, *T. cruzi* does not express trans-sialidase and has no sialic acids on its cell surface. After several changes in the gut of the insect, the parasite moves to the rectal gland of the insect where it multiplies and produces infective metacyclic trypomastigote forms of *T. cruzi* ready to leave the insect with defecation. The metacyclic forms express trans-sialidase and once transmitted to the human (or any vertebrate host) by the insect vector by way of a bloodmeal, they acquire sialic acids from the host sialoglycoconjugates and are able to invade cells in the host body. Soon after invasion by the parasite, the membrane of the invaded host cell is ruptured and the parasites are exposed to the cytoplasm where they transform into amastigote form that can multiply very quickly. Amastigotes grow and transform into trypomastigote form. Trans-sialidase expression is found to be stage-specific and turned on only during metacyclic stages.
As with many other parasites, the presence, quantity and localization of sialylated glycoconjugates on the cell surface is very important for the survival and infectivity of *T. cruzi*. Since it is unable to synthesize sialic acids, *Trypanosoma cruzi* depends on its trans-sialidase enzyme to acquire sialic acids from host sialoglycoconjugates (Figure 1-4). Sialic acids transferred to its surface glycoconjugates—called mucins—provide *T. cruzi* the ability to evade the immune system of the host and to adhere to and invade the host cells. The parasite invasion is found to be reduced significantly when sialylated epitopes are neutralized with antibodies, when trans-sialidase is neutralized by various other protocols or when sialic acids are present in neither the host cells nor the external medium. The direct involvement of trans-sialidase in animal pathogenesis and thymocyte apoptosis is also shown.

Figure 1-3. Different forms of *T. cruzi* in its life cycle (Adapted from Figure 2 in Tyler *et al.*

**1.5. *Trypanosoma cruzi* trans-Sialidase (TcTS) and *Trypanosoma rangeli* Sialidase (TrSA)**

Sialidase superfamily catalyzes hydrolysis of sialo-sugars releasing sialic acids and includes enzymes that act as virulence factors in pathogenesis of many different diseases. *Trypanosoma cruzi* trans-sialidase (TcTS), a member of this family, is much more efficient in catalyzing transfer of sialic acids in the presence of suitable sugar acceptors (glycoconjugates with terminal β-galactose) than hydrolysis (Table 1-1). *T. cruzi*, the parasite responsible for Chagas’ disease, is unable to synthesize sialic acids *de novo* but expresses TcTS to evade
immune response of the host and to invade cells. The role of TcTS in survival and infectivity of *T. cruzi* besides the lack of this enzyme in mammals makes it a potential and appealing therapeutic target.

Figure 1-4. Transfer of sialic acids onto the *Trypanosoma cruzi* cell surface.

Table 1-1. Sialidase and trans-sialidase activities of wild-type and mutant TcTS and TrSA.

<table>
<thead>
<tr>
<th>Substrate: α-2,3-sialyllactose</th>
<th>Sialidase activity</th>
<th>Trans-sialidase activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcTS wild type</td>
<td>306.0 ± 4.4</td>
<td>1412.98 ± 21.95</td>
</tr>
<tr>
<td>TcTS Trp312Ala mutant</td>
<td>185.1 ± 7.1</td>
<td>0</td>
</tr>
<tr>
<td>TrSA wild type</td>
<td>4298.7 ± 34.4</td>
<td>0</td>
</tr>
<tr>
<td>TrSA Trp312Ala mutant</td>
<td>3567.8 ± 56.9</td>
<td>0</td>
</tr>
<tr>
<td>TrSA Gln283Pro mutant</td>
<td>11655 ± 95</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Substrate: α-2,6-sialyllactose</th>
<th>Sialidase activity</th>
<th>Trans-sialidase activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcTS wild type</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TcTS Trp312Ala mutant</td>
<td>109.9 ± 8.3</td>
<td>0</td>
</tr>
<tr>
<td>TrSA wild type</td>
<td>0.7 ± 0.2</td>
<td>0</td>
</tr>
<tr>
<td>TrSA Trp312Ala mutant</td>
<td>72.8 ± 11.1</td>
<td>0</td>
</tr>
</tbody>
</table>

Activities are expressed as nmol sialic acid (free sialic acid for sialidase activity, sialic acid transferred to lactose for trans-sialidase activity) min⁻¹mg⁻¹.

TcTS specifically catalyzes transfer of α-2, 3-linked N-acetylneuraminic acids (which will be referred to as sialic acid from now on throughout this study) to acceptors with terminal β-galactosyl residues with retention of configuration. Thus, TcTS differs from sialyltransferases which cause inversion of configuration and require sugar nucleotides like cytidine.
| TcTS | 1   | ---LAPGSSRVELFKRQSKVPEK-DGKVTERVVHSLPVLNVDGV |
| TrSA | 1   | AASLAPGSSRVELFKRKNSTVPFEESNGTIRERVVHSFRIPIVTNVDGV |
| TcTS | 46  | MVIAADARYTESSNDNLIIDTVAKYSVDDGGETWETQIAAICNSRASSSVRV |
| TrSA | 50  | MVIAADARYETSFDNSTEMTAVKYSVDGATNTQIAAICNSRASSSVRV |
| TcTS | 95  | VDPTIVKLNKLYLVLSYNSSRSYWTSHGDARDWIILLAVGVTKSTA |
| TrSA | 99  | MDATIVKLNKLYLVLSYNSSRSYWTQHRGDWPEPLLTVGVTKSA |
| TcTS | 144 | GGTKTASIKWSVPVLKKEFFPAEMGHTNQFLGAGVAIAVSNGLNYV |
| TrSA | 149 | NGKTATASIKWSVPVLKKEFFPAEMGHTNQFLGAGVAIAVSNGLNYV |
| TcTS | 192 | PVQVTNKQVFSDKFSEDEGTWKFVKGRSAAMGCVFPAVELEWEGKLI |
| TrSA | 196 | PVQIADMGCVFPTVIMMYEDEGTWKFVKGRSAAMGCVFPAVELEWEGKLI |
| TcTS | 242 | INTRVDDYRRRLYETWDMQNTWLEAVGTSLRVWGSPPSQSNSQPQSSPT |
| TrSA | 246 | INNVYDDNRLYETWDMQNTWLEAVGTSLRVWGSPPSQSNSQPQSSPT |
| TcTS | 291 | AVTIEKMDVMLFTHLFNKGKRDRLNYLWLDNQRIYTVGQVSTIGDEN |
| TrSA | 295 | AVTIEKMDVMLFTHLFNKGKRDRLNYLWLDNQRIYTVGQVSTIGDEN |
| TcTS | 340 | SAYSSVLYKDDKLYCLHEINSNVYSVLFARLVGELRIKSVQLCWSWKNW |
| TrSA | 344 | GYYSSVLYKDDKLYCLHEINSNVYSVLFARLVGELRIKSVQLCWSWKNW |
| TcTS | 389 | DSHLSICTPAPAAASSERGCGPAVTVGLVGFLSHSATKTEWEAPYR |
| TrSA | 393 | DHNLASICTPAPAAASSERGCGPAVTVGLVGFLSHSATKTEWEAPYR |
| TcTS | 438 | CVNASTANAERVPNLGKPAVGVGGALFWVSQQQGONRYHFAHSAFLVA |
| TrSA | 442 | CDANVANAERVPNLGKPAVGVGGALFWVSQQQGONRYHFAHSAFLVA |
| TcTS | 487 | SVTIEFVPKGASPLGASLDSSGKCOLLSYDKNHQPPIYGSTPVT |
| TrSA | 491 | STVIDELPKGASTPLLGALEGPGDACKLSYDNRQWPYLAGAAPAF |
| TcTS | 537 | TGSWEMKRYHVVLMTNARKISVVIDEPLEGSGQTVVPDERTDPISHF |
| TrSA | 541 | TGSWELHKKYHVVLMTNARKISVVIDEPLEGSGQTVVPDERTDPISHF |
| TcTS | 585 | YVGGYCRSMPDSTRTVVNVLLYRNQMNAAEIRTLFLSQQDLIGTE |
| TrSA | 589 | YIGPSSKAPDSTRTVVNVLLYRNQMNAAEIRTLFLSQQDLIGTE |

Figure 1-5. Sequence alignment of TcTS and TrSA with amino acid differences highlighted.
monophosphate (CMP)-N-acetylneuraminate as sialic acid donors.\textsuperscript{5,50-54} Recently, it was shown that TcTS can also catalyze the transfer of $\alpha$-2, 3-linked N-glycolylneuraminic acid to acceptors with terminal $\beta$-galactosyl residues, even more efficiently than the transfer of sialic acids.\textsuperscript{55} The enzyme is not sensitive to the monosaccharide linked to $\beta$-galactose; open chain derivatives of lactose, like lactitol and lactobionic acid, are also found to be good acceptors of sialic acid.\textsuperscript{56-58}

TcTS is a glycosylphosphatidilinositol(GPI)-anchored surface enzyme that consists of a 70 kDa globular core that accommodates the catalytic site and a variable number of highly immunogenic repeats termed SAPA—shed acute phase antigen.\textsuperscript{43} The enzyme has an optimum activity around pH 7.9.\textsuperscript{46} About 140 genes in the \textit{T. cruzi} genome encodes for TcTS protein family of which most members lack catalytic activity due to a key mutation (Tyr342His).\textsuperscript{59-61} TcTS belongs to the family of exo-$\alpha$-sialidases (EC 3.2.1.18)\textsuperscript{62,63} and is listed as a member of glycoside hydrolase family GH 33 according to the Henrissat classification (\url{www.cazy.org}).\textsuperscript{63}

A closely related parasite, \textit{Trypanosoma rangeli}, expresses a surface sialidase which lacks SAPA tail but has unusual similarity to 631-residue-long globular core of TcTS. \textit{T. rangeli} coexists with \textit{T. cruzi} sharing the same vector, host species and epidemic area and thus, makes it harder to diagnose Chagas’ disease and to distinguish from \textit{T. rangeli} infections.\textsuperscript{64} \textit{Trypanosoma rangeli} sialidase (TrSA) has an optimum activity around pH 5.5\textsuperscript{65} and is also encoded by a multi-gene family some members of which encode for catalytically inactive enzymes.\textsuperscript{66-68} There are 5 glycosylation sites in the natural TrSA enzyme at asparagine residues 14, 23, 114, 428 and 614 which are proven to have no role in catalysis since recombinant TrSA lacking these sugar residues also shows the same enzymatic activity.\textsuperscript{69}

The sequence identity is 70\% between TrSA and TcTS with only four amino acid insertions in TrSA close to the N-terminus (Figure 1-5). One can evaluate the extent of similarity
comparing with the fact that any two proteins having more than 30% sequence identity are known to adopt essentially the same structure.\(^7^0\) When TcTS and TrSA are superimposed using corresponding \(C_\alpha\) atoms, the root mean square deviation (RMSD) is only 0.59 Å. The active sites of the two enzymes are also very similar as will be explained in detail in the next section. Despite this structural similarity, TrSA is found to lack any trans-sialidase activity\(^4^8,6^6,7^1\) which makes these two enzymes an excellent case study to elucidate the reasons for the two different catalytic activities (Table 1-1). Another important difference of TcTS from TrSA is in their inhibitor binding properties. 2,3-dehydro-3-deoxy-N-acetylneuraminic acid (DANA) (3, Figure 1-6), which is a structural analog of sialic acid oxocarbenium ion (2, Figure 1-6) and an efficient inhibitor for TrSA and sialidases in general, does not inhibit TcTS\(^4^4,7^2\) (Table 1-2). Other sialidase inhibitors like oseltamivir (4, Figure 1-6) and zanamivir (5, Figure 1-6) do not inhibit TcTS, either. With such similar active sites, the reasons of this inhibition difference await deciphering and can guide inhibitor design studies for TcTS.

Unraveling the minute but vital differences between TcTS and TrSA could also pave the way to tailor sialidases into trans-sialidases (and glycosidases into trans-glycosidases) to use for efficient synthesis of oligosaccharides with terminal sialic acids.\(^7^2-7^4\) Using modified glycosidases instead of long and low-yield chemical processes to synthesize desired polysaccharides will be a great improvement if the large number of various glycosidases can be tailored into efficient trans-glycosidases that are not naturally available.

1.5. Structural Information about *Trypanosoma cruzi* trans-Sialidase and *Trypanosoma rangeli* Sialidase

Due to its importance for Chagas’ disease, TcTS is one of the most extensively studied enzymes among sialidases. The X-ray crystal structures of free and DANA-bound forms of TcTS were elucidated in 2002\(^7^5\) just after those of TrSA were obtained.\(^6^9,7^6\) The structures of Michaelis
Figure 1-6. Structures of sialic acid oxocarbenium ion (2), DANA (3), oseltamivir (4), zanamivir (5) and sialyllactose (6). The monosaccharide units of sialyllactose are also shown.
complex of TcTS<sub>Asp59Ala</sub> mutant with sialyllactose and the covalent intermediate forms of both
TcTS and TrSA using activated sialic acid substrates were also obtained in subsequent
crystallographic studies. Both enzymes consist of two globular domains that are attached by
a long α-helical segment (Figure 1-7). Equivalent positions in the two enzymes differ by 4 amino
acids because of four residue insertions present in TrSA in the first 25 residues of the N-
terminus. The residue numbering in TcTS will be referred for equivalent positions of both
enzymes to keep consistency throughout this work unless otherwise stated. The N-terminal
domain (residues 1-371) has a six-bladed β-propeller topology similar to bacterial and plant
sialidases and is where the active site resides. An α-helical region (residues 372-395) connects
the N-terminal domain to the C-terminal domain (residues 396-640) which has a β-barrel
topology similar to plant lectins and a short α-helical region at the end. The narrow catalytic cleft
seen in the crystal structures argues against the formerly suggested sequential mechanism which
requires concomitant binding of the donor and the acceptor substrates to the enzyme.

The catalytic cleft of TcTS consists of two sites; the sialic acid binding site which
accommodates the sialic acid part of the donor ligand and is buried deep in the enzyme and the
aglycon binding site which accommodates the lactose part of the donor ligand and subsequently
the acceptor ligand (lactose, 6 in Figure 1-6, or any sugar with a terminal β-galactose) (Figure 1-
8). The aromatic side chains of Trp312 and Tyr119 lie parallel to each other and form the two
lateral walls of the aglycon binding site at the periphery of TcTS. Although both of these two
residues lie on flexible loops, both loops adopt the same pose in all X-ray crystal structures of
TcTS and TrSA. Another interesting property of Tyr119 depicted in Figure 1-9 is that two
different conformations of Tyr119 are observed unless the aglycon binding site is filled with a
ligand. In one conformation (the “down” conformation), Tyr119 side chain swings into the
Figure 1-7. Two different views from superimposed sialyllactose-bound TcTS (red) and unligated TrSA (blue). Sialyllactose is shown in orange.

Figure 1-8. Active site of TcTS when sialyllactose (in CPK form) is bound. A) Trp312 and Tyr119 that lie at the periphery of the active site and surround the lactose part of the ligand. The arginine triad that interacts with the carboxylate group of the sialic acid part of the ligand is also shown. B) The catalytic nucleophile, Tyr342/Glu230 pair, lie at the bottom of the catalytic cleft while Asp59 interacts with the glycosidic O atom that connects sialic acid and lactose parts of the sialyllactose.
catalytic cleft interacting with Asp96, Gln195 and Glu230 and in the presence of sialic acid also with the glycerol side chain of sialic acid, while in the other conformation (the “up” conformation) Tyr119 side chain completely leaves the catalytic cleft extending into the bulk solvent. Emphasizing the role of hydrophobic interactions with Tyr119 and Trp312, the aglycon binding site does not have any direct hydrogen bonding interactions with the lactose moiety of the ligand, but water-mediated interactions instead, except the interaction of Asp59 carboxylic acid with the glycosidic oxygen (Figure 1-8).

Figure 1-9. Dual conformations of Tyr119.

The sialic acid binding site of TcTS catalytic cleft shares several features with microbial sialidases (Figure 1-8). The arginine triad—Arg35, Arg245 and Arg314— that forms strong salt bridges with the carboxylate group of sialic acid, glutamate (Glu357) that stabilizes Arg35, Asp59 that acts as an acid/base catalyst and a tyrosine and a glutamic acid residue (Tyr342 and Glu230) that lie at the bottom of the catalytic cleft and act as a nucleophile couple in microbial sialidases are all conserved in TcTS.
The active site of TrSA is very similar and has the same residues mentioned so far in TcTS active site—in very similar positions—except having a serine residue instead of Tyr119 residue. Due to this difference, TrSA lacks the second lateral wall seen in TcTS that embraces the lactose part of the ligand and thus, the active site of TrSA is more solvent-exposed. However, a simple Ser119Tyr mutation of TrSA does not provide it with any trans-sialidase activity, so the story is clearly more complex. Taking advantage of the very few amino acid differences close to the active sites of TcTS and TrSA, a number of mutagenesis studies have been performed to unravel the critical amino acids necessary for trans-sialidase activity.48,69,81

### Table 1-2. Inhibitor (DANA) binding properties of TcTS, TrSA and several TrSA mutants

<table>
<thead>
<tr>
<th></th>
<th>$K_i$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcTS wild type</td>
<td>12.29</td>
</tr>
<tr>
<td>TrSA wild type</td>
<td>0.0015</td>
</tr>
<tr>
<td>TrSA$_{5mut}$</td>
<td>1.54</td>
</tr>
<tr>
<td>TrSA$_{5mut}$ Ile37Leu</td>
<td>1.01</td>
</tr>
<tr>
<td>TrSA$_{5mut}$ Gly342Ala</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Mutagenesis studies showed the importance of Tyr119 and Trp312 for trans-sialidase catalysis in TcTS. Tyr119 in TcTS is a serine residue in TrSA and thus, could very well be a reason for the trans-sialidase ability of TcTS.48,69,81 However, TrSA$_{Ser119Tyr}$ mutant failed to catalyze sialic acid transfer while the inverse mutation (Tyr119Ser) caused TcTS to lose almost all its trans-sialidase ability.48 These results show that presence of Tyr119 residue is necessary but insufficient by itself for trans-sialidase ability. It is also shown that Trp312Ala mutation completely abolished trans-sialidase activity of TcTS while the same mutation only decreased sialidase activities of TcTS and TrSA indicating the importance of Trp312 for trans-sialidase activity48 (Table 1-1). Trp312 is also found to be the reason of high substrate specificity of TcTS and TrSA for sialyl-$\alpha$-2,3-linked-oligosaccharides since Trp312Ala mutants of both TcTS and TrSA can additionally hydrolyze sialyl-$\alpha$-2,6-linked-oligosaccharides which can not be
hydrolyzed by the wild type enzymes at all\textsuperscript{48} (Table 1-1). Sialidases of \textit{Vibrio cholerae}, \textit{Micromonospora viridifaciens} and the influenza virus which have dual $\alpha$-(2, 3) - $\alpha$-(2, 6) substrate specificity all have a shorter loop in the corresponding region—constituting a completely solvent-exposed catalytic cleft—and no tryptophan residue like Trp312 confirming the role of Trp312 in substrate specificity.\textsuperscript{69} Even the two residues Pro283 and Tyr248 that lie side by side on two loops neighboring the loop of Trp312 in TcTS prove to be important for trans-sialidase activity since both residues sterically help Trp312 side chain to adopt a suitable conformation for binding (Figure 1-10). These two residues are among the critical amino acid differences between TcTS and TrSA; Pro283 and Tyr248 in TcTS correspond to a glutamine and a glycine residue, respectively, in TrSA. However, TrSA$_{\text{Gln283Pro}}$ and even TrSA$_{\text{Gln283Pro/Gly248Tyr}}$ mutants did not show any trans-sialidase activity, but they both showed significantly increased sialidase activities.\textsuperscript{48} Curiously, both trans-sialidase and sialidase activities are abolished in TcTS$_{\text{Pro283Gln}}$ mutants.\textsuperscript{48} Even the double mutant TcTS$_{\text{Tyr248Gly/Pro283Gln}}$ and the triple mutant TcTS$_{\text{Tyr119Ser/Tyr248Gly/Pro283Gln}}$ lack both trans-sialidase and sialidase activities. All these mutagenesis results point to the necessity of the precise location of Trp312 side chain to lock the lactose part of the ligand (especially the acceptor ligand which binds the enzyme only through this interaction) into a proper conformation for trans-sialidase catalysis in TcTS.

A later study further confirmed the importance of Pro283 residue for trans-sialidase activity. Since most of the residues related to catalysis reside among the first 200 residues, chimeric enzymes were constructed exchanging the 200-residue long segment of N terminus of TcTS and TrSA and their sialidase and trans-sialidase activities were measured (Figure 1-11).
Figure 1-10. The two residues that affect Trp312 in TcTS: Pro283 and Tyr248.

Figure 1-11. Sialidase and trans-sialidase activities of chimeric proteins of TcTS and TrSA. The blue and white bars represent TcTS and TrSA-derived sequences, respectively. The amino acid identity at position 283 is also indicated.
The chimera with the N-terminus of TcTS and the C-terminus of TrSA was found to be totally inactive, while the other chimera with the N-terminus of TrSA and the C-terminus of TcTS had a strict sialidase activity. An additional Gln283Pro mutation was found to confer trans-sialidase activity to the second chimera with strict sialidase activity (Figure 1-11).

Unable to confer trans-sialidase activity to TrSA with single, double or triple mutations, other possible mutations in TrSA active site to achieve this goal were investigated. For this purpose, a sequence alignment of TcTS, TrSA and *Trypanosoma brucei* trans-sialidase (TbTS) which has only about 30% sequence similarity to TcTS and TrSA was also referred as a guide.\(^72\)

The three essential mutations for trans-sialidase activity identified in previous studies (Ser119Tyr, Gly248Tyr, Gln283Pro) were supplemented with two mutations in TrSA: Met95Val and Ala97Pro. These two mutations were chosen for several reasons. First, they are both neighbors of a conserved residue in the active site (Asp96). Second, Val95 and Pro97 are conserved both in TcTS and TbTS while these two residues are substituted with Met95 and Ala97 in TrSA. And third, different interactions of Asp96 with the inhibitor DANA are observed in crystal structures of DANA-bound forms of TcTS and TrSA probably related to the identity of neighboring residues. The TrSA quintet mutant (TrSA\(_{5\text{mut}}\)) prepared using this strategy—with Ser119Tyr, Gly248Tyr, Gln283Pro, Met95Val and Ala97Pro mutations—achieved to confer trans-sialidase activity in TrSA scaffold for the first time although it is only 1% of the trans-sialidase activity of wild type TcTS\(^72\) (Table 1-3). In order to increase the trans-sialidase activity of TrSA\(_{5\text{mut}}\), additional mutations were introduced as shown in Table 1-3. Among these Gly341Ala and Ile36Leu mutations prove to be effective significantly obtaining 10% of the trans-sialidase activity of TcTS. Both of these mutations were chosen due to their possible effect on the nucleophile, Tyr342. Additionally, Gly341 (Ala341 in TcTS) probably affects the
flexibility of Tyr342 while Ile36 (Leu in TcTS) is in contact with the side chain of Tyr342 and has alternate conformations. The sialidase activities of all the mentioned mutants were also determined in the same study. As shown in Table 1-4, single Gln283Pro mutation promoted the sialidase activity of TrSA significantly, which further confirmed the importance of Pro283. TrSA5mut and its further mutated forms which have acquired trans-sialidase activity all showed low sialidase activities (Table 1-3, Table 1-4). Also, the insensitivity of the enzymes to inhibition by DANA was observed to increase as they acquire trans-sialidase activity for TrSA5mut and its further mutated forms, however, it still did not reach the level of TcTS (Table 1-2).

DANA-bound forms of TcTS and TrSA are found to differ in the enzyme-ligand interactions seen in X-ray crystal structures. The hydroxyl groups of the glycerol moiety of DANA form several direct hydrogen bonds to TcTS residues while only water-mediated bonds between this region of DANA and TrSA residues are observed. The O4 hydroxyl group of DANA interacts with Asp59 in TcTS while it interacts with Asp96 in TrSA.

Table 1-3. Trans-sialidase activity of various TrSA mutants

<table>
<thead>
<tr>
<th>Protein</th>
<th>TS activity (pmol min⁻¹μg protein⁻¹)a</th>
<th>App Kₘ (mM)b</th>
<th>App Vₘₐₓ (pmol min⁻¹μg protein⁻¹)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>TrSA wt/TrSA Ser119Tyr/TrSA Gln283Pro/TrSA Met95Val, Ala97Pro/TrSA Gly248Tyr, Gln283Pro/TrSA Ser119Tyr, Gly248Tyr, Gln283Pro</td>
<td>Undetectable</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TrSA5mut</td>
<td>3.84 ± 0.18 (0.9)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TrSA5mut Ile36Leu</td>
<td>47.20 ± 5.50 (11.3)</td>
<td>9.44 ± 2.00</td>
<td>14.70 ± 1.11</td>
</tr>
<tr>
<td>TrSA5mut Gly341Ala</td>
<td>45.28 ± 3.30 (10.9)</td>
<td>6.90 ± 2.57</td>
<td>6.39 ± 1.30</td>
</tr>
<tr>
<td>TrSA5mut Val179Ala</td>
<td>10.15 ± 0.21 (2.4)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TrSA5mut Phe113Tyr</td>
<td>11.11 ± 0.57 (2.7)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TrSA5mut Thr38Ala</td>
<td>26.62 ± 0.72 (6.4)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TrSA5mut Asp284Gly</td>
<td>12.13 ± 0.78 (2.9)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TcTS wt</td>
<td>416.71 ± 47.43 (100)</td>
<td>0.24 ± 0.03</td>
<td>3.83 ± 0.42</td>
</tr>
</tbody>
</table>

a The activity is measured using sialyllactose and lactose as donor and acceptor, respectively. The percentage of trans-sialidase activity compared to wild type TcTS is given in parenthesis. b Apparent kinetic constants for lactose are determined keeping sialyllactose at 2 mM concentration and varying the lactose concentration.
The process of trying to obtain X-ray crystal structure of TcTS gave additional structural clues. They found that the crystals of TcTS become instable and crack readily when DANA or sialic acid is bound to it. In addition, TcTS tends to precipitate when it is incubated with DANA or sialic acid. Both of these observations suggest that conformational changes occur upon ligand binding, which ultimately disrupt the molecular packing of crystals. In the same study, they also found that soaking unligated TcTS crystals with lactose failed to produce any lactose-bound TcTS crystals. Since lactose is a natural acceptor ligand for the transfer reaction of TcTS, it definitely can bind TcTS. To elucidate this issue, surface plasmon resonance experiments were performed on the immobilized TcTS—in an enzymatically inactive form due to Asp59Asn mutation—using sialyllactose and lactose as mobile specific ligands separately. No lactose binding to TcTS could be detected even at very high concentrations like 100 mM while sialyllactose binding to TcTS was detected even in the concentration range of 0.1-2 mM. However, if the same experiment was performed after TcTS was preequilibrated in a buffer that contains sialic acid or sialyllactose, lactose binding to TcTS was clearly detected at 10 mM concentration. These results clearly demonstrate that the affinity for the acceptor substrate is modulated by donor substrate binding, probably due to a structural change.

Another result obtained from the X-ray crystallography studies is that TcTS can produce DANA. All TcTS crystals soaked with sialic acid ended up having DANA in their active sites. This was also observed previously in a different study by Todeschini et al. The structural and dynamical properties of TcTS and TrSA will be thoroughly inquired in Chapter 5 by the analysis of molecular dynamics simulations of the two enzymes.
Table 1-4. Kinetic constants for the sialidase activity of TrSA mutants

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_m$ (mM)</th>
<th>$V_m$ (nmol min$^{-1}$ μg protein$^{-1}$)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$10^{-5}k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TrSA wt</td>
<td>0.273 ±0.010</td>
<td>273.11 ±8.82</td>
<td>151.42</td>
<td>5.55</td>
</tr>
<tr>
<td>TrSA Ser119Tyr</td>
<td>0.068 ±0.002</td>
<td>45.48 ±1.17</td>
<td>27.21</td>
<td>3.98</td>
</tr>
<tr>
<td>TrSA Gln283Pro</td>
<td>0.082 ±0.02</td>
<td>1175.14 ±27.60</td>
<td>703.03</td>
<td>86.08</td>
</tr>
<tr>
<td>TrSA Met95Val,Ala97Pro</td>
<td>1.079 ±0.048</td>
<td>285.98 ±11.40</td>
<td>171.09</td>
<td>1.59</td>
</tr>
<tr>
<td>TrSA Gly248Tyr,Gln283Pro</td>
<td>0.070 ±0.001</td>
<td>225.76 ±4.57</td>
<td>135.06</td>
<td>19.2</td>
</tr>
<tr>
<td>TrSA Ser119Tyr, Gly248Tyr, Gln283Pro</td>
<td>0.122 ±0.008</td>
<td>159.32 ±10.33</td>
<td>95.31</td>
<td>7.83</td>
</tr>
<tr>
<td>TrSA5mut Ile36Leu</td>
<td>0.039 ±0.001</td>
<td>71.54 ±1.32</td>
<td>42.80</td>
<td>10.88</td>
</tr>
<tr>
<td>TrSA5mut Gly341Ala</td>
<td>0.200 ±0.009</td>
<td>106.04 ±4.43</td>
<td>63.44</td>
<td>3.17</td>
</tr>
<tr>
<td>TrSA5mut Val179Ala</td>
<td>0.107 ±0.003</td>
<td>38.29 ±0.87</td>
<td>22.91</td>
<td>2.13</td>
</tr>
<tr>
<td>TrSA5mut Phe113Tyr</td>
<td>0.172 ±0.029</td>
<td>49.48 ±8.18</td>
<td>29.60</td>
<td>1.73</td>
</tr>
<tr>
<td>TrSA5mut Thr38Ala</td>
<td>0.038 ±0.001</td>
<td>74.91 ±1.53</td>
<td>44.82</td>
<td>11.79</td>
</tr>
<tr>
<td>TrSA5mut Asp284Gly</td>
<td>0.029 ±0.001</td>
<td>56.03 ±1.17</td>
<td>33.52</td>
<td>11.65</td>
</tr>
<tr>
<td>TcTS wt</td>
<td>0.291 ±0.022</td>
<td>0.30 ±0.022</td>
<td>0.18</td>
<td>0.0062</td>
</tr>
</tbody>
</table>

Molecular mass of TcTS and TrSA are 71.2 kDa and 76.1 kDa, respectively. Sialidase activities are determined using 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid as substrate and fluorescence measurements of released 4-methylumbelliferone. $a$ $K_m$ and $V_m$ values are estimated using a Lineweaver-Burk plot. $b$ $k_{cat}$ and $k_{cat}/K_m$ values are obtained using the $K_m$ and $V_m$ values in the previous columns.

1.6. Mechanistic Information about *Trypanosoma cruzi* trans-Sialidase and *Trypanosoma rangeli* Sialidase

The mechanism of *Trypanosoma cruzi* trans-sialidase (TcTS) is extensively studied due to its role in Chagas’ disease. Both sialidase and trans-sialidase reactions of TcTS have been explored. Kinetic isotope effect studies, initial velocity studies, chemical trapping and chemical rescue experiments are examples of experimental methods utilized.

There are two possible mechanisms for bisubstrate enzymatic reactions; sequential mechanism and ping-pong mechanism. The sequential mechanism requires binding of both substrates to the enzyme before any chemical reaction happens, while the ping-pong mechanism requires chemical reaction to occur after the binding of the first substrate that results in a modified form of the enzyme, which subsequently binds the second substrate and completes the reaction (Figure 1-12). The sequential mechanism has two kinds; ordered sequential mechanism and random sequential mechanism. Ordered sequential mechanism requires binding of one of the
substrates particularly before the other one, while in a random sequential mechanism the order of binding of the substrates does not matter.

Figure 1-12. Different types of bisubstrate enzyme mechanisms. A and B represent the substrates while P and Q represent the products. E’ is modified form of enzyme E due to covalent bonding to a ligand.

To get detailed information about TcTS mechanism, primary and $\beta^{-2}H$ kinetic isotope effects (KIE) were determined both in TcTS and contrasted with the ones in acidic solution.\textsuperscript{82} Primary $^{14}C$ (or $^{13}C$) KIEs give a measure of the associative/dissociative character of the reaction while $\beta^{-2}H$ KIEs inform us about the charge development at the anomeric carbon (Figure 1-1) in the transition state. The data for the acid-catalyzed solvolysis reaction of the substrate were indicative of a transition state with a dissociative character (with very little, if any, nucleophilic participation) and a significant charge development at the anomeric carbon. However, large primary $^{14}C$ KIEs and small $\beta^{-2}H$ KIEs measured show that the transition state in the trans-sialidase reaction of TcTS has an associative character (i.e. has significant nucleophilic participation) with little charge development at the anomeric carbon. The information obtained about the transition state for trans-sialidase reaction of TcTS pointed towards a subsequent formation of a covalent intermediate. Additionally, very similar $\beta^{-2}H$ KIEs obtained in the same
study at different concentrations of the acceptor substrate ruled out the possibility of sequential ordered mechanism for TcTS that was suggested based on previous data.46,79

Initial velocity studies of the trans-sialidase reaction of TcTS were also performed by the same researchers to complement the KIE studies.83 The initial velocity results can distinguish sequential and ping-pong mechanisms; the double reciprocal plot (1/V vs. 1/[Substrate₁] while [Substrate₂] is held constant) consists of intersecting lines for a sequential mechanism—both random and ordered ones—, while parallel lines are typical for a ping-pong mechanism. The pattern for a ping-pong mechanism can change if any branch reactions exist, like the hydrolysis reaction in TcTS, but one can still distinguish a ping-pong mechanism from a random sequential mechanism by measuring initial velocities of both the first and second product formation. The double reciprocal plot obtained monitoring the first product formation of TcTS reaction showed parallel lines. However, the double reciprocal plot obtained by monitoring the second product formation resulted in intersecting lines when performed in low substrate concentrations—at which the hydrolysis reaction becomes significant. These data indicated the mistake of the previous steady-state kinetic studies46,79 that were interpreted as implying a sequential mechanism for TcTS without taking into account the hydrolytic branch reaction of TcTS. Thus, a branched ping-pong mechanism was suggested by this study for TcTS. To further the analysis, a chemical trapping experiment with radioisotope-labeled sialyllactose as the substrate was performed, which successfully captured a labeled intermediate with sialic acid covalently bound to the enzyme.52,83 This study was the first to obtain evidence for a ping-pong mechanism including covalent intermediate formation, much before any structural data from X-ray crystallography became available. Although it was not possible to identify the residue of TcTS acting as the nucleophile, they correctly predicted it to be Tyr342.
The X-ray crystal structures of TcTS were obtained in 2002 showing that Tyr342 is well-positioned to act as a nucleophile. The significance of Tyr342 was already revealed by that TcTS with either natural or site-directed mutation at this position (Tyr342His, Tyr342Ala, Tyr342Gly, Tyr342Phe, Tyr342Thr) completely lacks trans-sialidase and sialidase activity. However, this idea of Tyr342 acting as a nucleophile was still not adopted due to tyrosine’s high pKa until a chemical trapping study was able to identify Tyr342 as the nucleophile in TcTS reaction using an activated substrate. 2, 3-difluoro sialic acid is an activated substrate since the fluorine atom at C3, adjacent to the anomeric center, inductively destabilizes the positively charged oxocarbenium ion-like transition state reducing both the formation and turnover rates while the fluoride leaving group at the anomeric center helps the covalent intermediate to be kinetically accessible. When TcTS was incubated with 2, 3-difluoro sialic acid, it was possible to trap and accumulate the covalent intermediate. Subsequent LC/MS analysis of the peptide digests revealed that the substrate formed a covalent bond to Tyr342, which was later confirmed by X-ray crystal structure of this covalent intermediate. Thus, in TcTS, Tyr342 indeed acts as a nucleophile couple with Glu230. It is suggested that there is a charge relay from Glu230 to Tyr342 in order to prevent the Coulombic repulsion of approaching the carboxylate group of the sialic acid directly with a negatively charged nucleophile. The same procedure was also used to identify Tyr342 as the nucleophile in TrSA mechanism with subsequent X-ray crystal structure determination of the covalent intermediate.

Comparison of the crystal structure of sialyl-lactose bound TcTS and the covalent intermediate shows that the sialic acid position is fixed from two sides by the strong interactions of its carboxylate group and N-acetyl side chain. However, the anomeric carbon goes through an electrophilic migration changing the sialic acid ring from the initial distorted skewed-boat
conformation (B\textsubscript{2,5}) of the Michaelis complex to a more relaxed chair conformation (\textsuperscript{2}C\textsubscript{3}) of the covalent intermediate\textsuperscript{77} (Figure 1-14). This relaxation might also stabilize the covalent intermediate with respect to the Michaelis complex and lengthen its lifetime.

The structural studies clearly indicate Asp\textsubscript{59} as the most suitable residue to act as an acid/base catalyst—to protonate/deprotonate the glycosidic oxygen. Asp\textsubscript{59} acting as an acid catalyst for influenza virus sialidase was debated previously since the pKa value of a solvent-exposed aspartic acid residue may be too low to act as an acid catalyst.\textsuperscript{88,89} However, the presence of negatively-charged sialic acid in the catalytic cleft, and the reduced solvent exposure due to both hydrophobicity of the catalytic cleft and the aglycone binding were suggested to raise the pKa significantly for Asp\textsubscript{59}. A recent chemical rescue study that prove using azide ions can restore the trans-sialidase activity of inactive TcTS\textsubscript{Asp\textsubscript{59}Ala} mutant forming a sialyl azide product confirmed Asp\textsubscript{59} as the acid/base catalyst.\textsuperscript{90}

Our understanding of the mechanism of TcTS has grown over the years as explained in detail in this section and a mechanism shown in Figure 1-14 is now accepted as the mechanism of TcTS. The relative timing of bond cleavages/formations and their extent in the transition structure is not clearly depicted in Figure 1-14, leaving it as an open question. The failure of DANA, a molecule structurally similar to an oxocarbenium ion (which is a possible transition
Figure 1-14. The proposed mechanism for trans-sialidase catalysis reaction of TcTS. The donor and acceptor sugar moieties are colored in red and orange, respectively. In sialidase catalysis reaction of TcTS, a water molecule takes the role of the acceptor lactose.
state), to inhibit TcTS and the KIE results that indicate an associative character with little charge development at the anomeric C for the transition structure in TcTS provide some information about the transition structure of TcTS mechanism. These experimental findings will be inquired with computational methods and the details of the mechanism of TcTS and how it differs from the mechanism of TrSA will be investigated in Chapters 3 and 4.
CHAPTER 2
METHODS

2.1. Quantum Mechanics

Quantum mechanics calculations take the electrons into account explicitly and thus, are able to compute the properties of a system that depend on its electronic distribution. It is also possible to investigate the reactions which involve bond breaking or formation with quantum mechanics. The \textit{ab initio} and semiempirical quantum mechanics methods as well as the density functional theory will be covered here shortly.

The Schrödinger equation lies at the basis of quantum mechanics. The electrostatic potential, $V$, is taken to be time-independent and thus, the time-independent Schrödinger equation (Equation 2-1) will be discussed in this section. The wavefunction of a particle, $\psi(r)$, defines the state of the particle. In the Schrödinger equation, the Hamiltonian operator, $H$,—that is given in Equation 2-2—acts on $\psi(r)$ and returns the wavefunction multiplied by a scalar value—called the eigenvalue—as the result if $\psi(r)$ is an eigenstate. The eigenvalue corresponds to the energy of the particular eigenstate.

$$H\psi(r) = E\psi(r)$$ \hfill (2-1)

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V$$ \hfill (2-2)

The Schrödinger equation for a one-electron atom can be solved resulting in functions that correspond to the atomic orbitals. However, for polyelectronic systems, no exact solution exists for the Schrödinger equation even for He atom since it becomes a three-body problem. Additionally, electron spins have to be accounted for in polyelectronic systems complicating the solution more. Thus, different approaches are taken to obtain approximate solutions for polyelectronic systems. In one approach called the perturbation theory, a similar problem that is
easier to solve than the real one is found and how the difference between this problem and the
real one will be reflected in the solutions is considered.

Born-Oppenheimer approximation which considers the nuclei of the system to be fixed is
used to reduce the complexity of the problem of polyelectronic systems. Once the contribution of
the nuclei are separated using the Born-Oppenheimer approximation, the remaining electronic
Hamiltonian solely depends on the coordinates of the electrons in the system and its
eigenfunctions are called spatial orbitals, \( \Psi_i(\mathbf{r}) \). For each spatial orbital \( \Psi_i(\mathbf{r}) \), two different
spin orbitals—which contains information about both the spatial distribution and the electronic
spin—can be constructed as \( \chi_1 = \Psi(\mathbf{r})\alpha(w) \) and \( \chi_2 = \Psi(\mathbf{r})\beta(w) \) in which \( \alpha(w) \) and \( \beta(w) \)
represent the spin up and spin down functions. To include the effect of electron spins, the exact
wave function of the polyelectronic system is required to satisfy the antisymmetry principle,
which states that the wave function must be antisymmetric with respect to the interchange of the
coordinates of any two electrons, as well as being a solution to the Schrödinger equation.\(^91\) To
satisfy the antisymmetry principle, Slater determinants, an example of which is shown in
Equation 2-3 for an N-electron system, are used.

\[
\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \sqrt{\frac{1}{N!}} \begin{vmatrix}
\chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \cdots & \chi_k(\mathbf{x}_1) \\
\chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \cdots & \chi_k(\mathbf{x}_2) \\
\vdots & \vdots & \ddots & \vdots \\
\chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \cdots & \chi_k(\mathbf{x}_N)
\end{vmatrix}
\]

The Hartree-Fock approximation—equivalent to molecular orbital approximation—is one
of the most ways to find approximate solutions to the electronic Schrödinger equation and
constitutes a cornerstone for understanding modern chemistry. An antisymmetric wave function
for a polyelectronic system can be represented as in Equation 2-3, however, there are no certain
forms of spin orbitals to use. The variation principle states that the best wave function in such a
functional form is the one which gives the lowest energy, $E_0$, when the full electronic Hamiltonian operator acts on it. Using a procedure to minimize $E_0$ with respect to the spin orbitals chosen, Hartree-Fock equation shown in Equation 2-4 can be derived. In this equation, $f(i)$ is an effective one-electron operator, called the Fock operator, as shown in Equation 2-5 in which $v^{\text{HF}}(i)$ is the average potential effective on electron $i$ due to the presence of all other electrons.\(^9\) Thus, the Hartree-Fock approximation reduces the problem of the polyelectronic systems to one-electron problems in which each electron is assumed to experience an average electronic potential. Since the average potential effective on electron $i$, $v^{\text{HF}}(i)$, is dependent on the spin orbitals of all other electrons, the Hartree-Fock equation has to be solved iteratively starting from the spin orbitals initially guessed until self-consistency is reached and this procedure is called the self consistent field (SCF) method. The spin orbitals of initial guess are constructed using a finite set of spatial basis functions.

\[
f(i)\chi(x_i) = \varepsilon\chi(x_i)
\]  

\[
f(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + v^{\text{HF}}(i)
\]

Semiempirical quantum mechanical methods are based on the Hartree-Fock formalism, however, they use a simpler Hamiltonian than the correct one (e.g. by approximating or totally excluding the two-electron integrals) and use parameters obtained by fitting to experimental data or the results of \textit{ab initio} calculations in order to compensate for using an approximate Hamiltonian.

Another approach to the electronic structure of polyelectronic systems is density functional theory (DFT). DFT also uses single-electron functions like Hartree-Fock approach, however, DFT does not calculate the full $N$-electron wave function explicitly but only aims to calculate the total electronic energy. DFT is based on the relationship between the overall
electron density and the total electronic energy, which can be expressed as that the total
electronic energy can be written as a functional form of the overall electron density as in
Equation 2-6. In this equation, \( V_{\text{ext}}(\mathbf{r}) \) is the external potential due to Coulomb interactions with
the nuclei and \( F[\rho(\mathbf{r})] \) is the sum of interelectronic interactions and the kinetic energy of the
electrons. The variational theory is used to minimize the total electronic energy. Finding the
functional forms used for \( F[\rho(\mathbf{r})] \) constitute one of the difficulties in this approach. Kohn-Sham
equation (Equation 2-7) serves for this purpose and is a sum of the kinetic energy of a system of
non-interacting \( N \) electrons (with the same electron density as the real system), electron-electron
Coulombic interaction and exchange-correlation contributions.\(^9\) Self consistent field method is
used to solve Kohn-Sham equations for the electrons of the system.

\[
E[\rho(\mathbf{r})] = \int V_{\text{ext}}(\mathbf{r})\,\rho(\mathbf{r})\,d\mathbf{r} + F[\rho(\mathbf{r})] \tag{2-6}
\]

\[
F[\rho(\mathbf{r})] = \sum_{i=1}^{N} \int \psi_i^*(\mathbf{r})(-\frac{\nabla^2}{2})\psi_i(\mathbf{r})\,d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\,d\mathbf{r}_1\,d\mathbf{r}_2 + E_{\text{XC}}[\rho(\mathbf{r})] \tag{2-7}
\]

2.2. Molecular Mechanics

Most of the biological problems are too large to be treated with quantum mechanics. To
be able to obtain energies and other properties of such large systems, molecular mechanics that
computes the energy of a system using only the nuclear coordinates (and excluding any
information about the electrons) can be used. However, it should be noted that molecular
mechanical methods solely can not deal with processes that include bond formation and/or
cleavage.

To calculate the energy of a system in molecular mechanics, force fields which are
mathematical functions that empirically describe the contributions of the interactions within a
system, like stretching of a bond, changing in a bond angle and dihedral angle, electrostatic
interactions, etc. are used. Current force fields like AMBER,\textsuperscript{93} CHARMM,\textsuperscript{94} GROMACS,\textsuperscript{95} OPLS\textsuperscript{96} that differ in their functional forms and/or parametrisation have proved to produce reasonable results in conformational energetics and charge distributions.

The mathematical functional form of AMBER force field, which is used throughout this study, is given in Equation 2-8. The first two terms in Equation 2-8 account for energetic penalties associated with the deviation of bonds and angles, respectively, from their equilibrium values. The third term describes how the energy changes as the dihedral angles change, while the last two terms introduce non-bonded (or ‘through-space’) interactions describing the van der Waals interactions via a Lennard-Jones potential and the electrostatic interactions via Coulomb’s law.

\[
U(R) = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{\epsilon}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i \neq j} (\frac{A_y}{R_{ij}}^2 - \frac{B_y}{R_{ij}}^6) + \sum_{i \neq j} \frac{q_i q_j}{\epsilon R_{ij}}
\]

As well as the functional form of the force fields, the parameters (the various constants in the equation) are also important. The parameters of force fields are designed empirically to reproduce various experimental properties. Transferability, which is the validity of the same set of parameters for a large group of relevant molecules, is an important and useful feature of a force field.

Most force fields—including AMBER—use “atom types” to distinguish an atom with a specific atomic number, hybridization and local environment. Just like quantum mechanical calculations that require the atomic numbers for each atom to be specified, molecular mechanical calculations require the atom type for each atom to be specified. Each parameter in the force field is expressed in terms of atom types, and each parameter necessary for a particular bond or
dihedral angle in a system is supplied from the parameter data files according to the atom types of the constituent atoms.

When explicit solvent molecules are used to simulate a system, periodic boundary conditions are generally used to exclude the effects of interactions of the atoms with the container walls.\textsuperscript{97} Thus, information on the behavior of the system in bulk solvent, rather than its behavior in a solvent close to a solid surface, can be obtained. Using periodic boundary conditions means that a ‘primary cell’ of volume $V$ which confines all $N$ atoms in the system of interest is imagined to act as a small portion of a bulk material. Basically, the bulk material is imagined to be built by the ‘primary cell’ surrounded by its exact replicas—called ‘image cells’—in all directions. All cells have open boundaries, so atoms can move from one cell to another freely. However, the atom number in each cell is kept constant ($N$) since for each atom leaving the primary cell, an image of that atom simultaneously enters the primary cell from the opposite face of the cell. During a simulation, only the coordinates of the atoms in the primary cell is stored, and the coordinates of the atoms in the image cells are computed using the appropriate coordinate transformation methods when necessary.

\textbf{2.3. Hybrid Methods}

It is currently not feasible to investigate the reactions of very large systems with quantum mechanics (QM), and molecular mechanics (MM) is insufficient for this purpose since it does not let bonds to be formed or broken. As a solution, hybrid methods which combine two or more computational methods in one single calculation are used—first realized by Warshel and Levitt\textsuperscript{98}—to investigate the reactions of large systems with high precision.

QM/MM methods, which are the most common of hybrid methods, treat the region where the chemistry takes place with quantum mechanics, while the rest of the system is treated with molecular mechanics that has a much less computational cost. The ONIOM method (our Own N-
layer Integrated molecular Orbital molecular Mechanics) implemented in Gaussian03, on the other hand, uses an extrapolation approach and is able to combine different methods without having a limitation of two layers at most, and thus is more general.

Despite sharing many common properties, QM/MM methods follow different paths in two details. The first detail is how they treat the covalent interaction between the quantum mechanics (QM) and molecular mechanics (MM) regions. Once the large system is partitioned into two regions, both regions will have dangling bonds and in order to make a quantum mechanical calculation, such dangling bonds have to be saturated. There are several methods to resolve this problem; using link atoms, frozen orbitals, or pseudopotentials. Link atoms are used in most QM/MM methods—including the one implemented in AMBER—as well as in the ONIOM method due to their generality. The link atom positions in both methods are obtained using a scale factor and the position vectors of the atom that is bonded to the link atom and the atom that is replaced by the link atom.

The second detail of QM/MM calculations that differ between various implementations is how the electrostatic interaction between the two regions is treated. In one approach called classical or mechanical embedding, these electrostatic interactions are calculated between partial point charges of the MM region and partial point charges of the QM region as would be in a pure molecular mechanics calculation. For this purpose, partial point charges for each atom in the QM region have to be assigned. In the other approach called electronic embedding, the interaction between charge distribution of the MM region and the actual charge distribution of the QM region is calculated. This method requires inclusion of MM region partial point charges in the QM Hamiltonian and thus, allows the QM region to get polarized according to the MM region charge distribution.
The energy expressions of generic QM/MM and ONIOM methods differ from each other. QM/MM methods calculate the total energy of the systems as a summation seen in Equation 2-9 (The total energy expressions here are written for mechanical embedding approach—which is the approach used for the calculations in this study—as will be obvious from that the interactions at the boundary of the two regions are treated with molecular mechanics.). The model system including the link atoms as well is treated with QM methods, and the MM-only system is treated with MM methods. The QM/MM boundary interactions are calculated including all MM terms that have at least one center in the MM-only region and at least one center in the model-only region.

\[
E_{\text{QM/MM}} = E_{\text{MM-only}_\text{system},\text{MM}} + E_{\text{model}_\text{system},\text{QM}} + E_{\text{model-only}_\text{MM-only}_\text{boundary},\text{MM}}
\]

\[
E_{\text{ONIOM}} = E_{\text{real}_\text{system},\text{MM}} + E_{\text{model}_\text{system},\text{QM}} - E_{\text{model}_\text{system},\text{MM}}
\]

On the other hand, the energy expression of ONIOM method given with Equation 2-10 is in the form of an extrapolation and requires three independent calculations.\(^{107}\) The energy of the real system that consists of all atoms is calculated at the MM level, and the energy of the model system, which is the focus of interest, is calculated at both QM and MM levels separately. One can infer from Equation 2-10 that most of the MM terms in the model system exist also in the real system and thus, will cancel totally in the full expression. To prevent a discontinuous potential and to make the result of the energy expression independent of the chosen partitioning of the system, it is suggested to have at least three bonds between the MM region and any bond that will be formed or broken in the reaction.

2.4. Methods for Exploring the Potential Energy Surface

The potential energy is a complicated function of the coordinates of all atoms in the system. The variation of energy with the coordinates is referred to as the ‘potential energy surface’ or ‘hypersurface’. Visualization of a potential energy surface is generally not possible.
since the potential energy depends on too many variables (3N coordinates in the case of an N-atom-system).

Apart from the global minimum of the potential energy surface (PES), many other local minima exist on the PES. Each local minimum can be thought as the lowest point in a local energy valley surrounded by higher energy hills. Identification of minimum points on PES can be achieved using minimization algorithms. The path that the system follows going from one minimum to another also draws a lot of attention. The highest point on such a pathway between two minima is called a ‘saddle point’ and the geometry of the system at that point is called a ‘transition structure’. Locating the minimum points on the PES can be used to obtain thermodynamic data, while additional knowledge of the saddle points is necessary to have an insight on kinetics of the system.

The minimization problem can mathematically be stated as finding the points on the PES with the first derivatives with respect to each variable being equal to zero and the second derivatives with respect to each variable being positive. Similarly, the saddle points are the points on the PES at which all first derivatives are equal to zero and only one of the second derivatives are negative.

Although analytical methods exist to locate the minima of a PES, they are not generally used due to the complexity of the form of the potential energy. Instead, numerical methods which gradually lower the energy of the system by changing the coordinates slowly are used. Most of the minimization algorithms go downhill on the PES from the starting point until it reaches the nearest minimum. So far, no algorithm known can locate the global minimum in a PES starting from an arbitrary point on the PES. However, it is also known that the ‘active’
structure (for example the active conformation of a drug molecule) is not necessarily the global minimum on the PES or the most highly populated conformation.92

Most minimization methods use derivatives while searching for the minimum. There are different minimization methods that use first derivatives only—like steepest descents method and conjugate gradients method—or both first and second derivatives—like the Newton-Raphson method. The first derivative of energy is called the gradient, and its direction gives an idea about the relative position of the nearest minimum while its magnitude indicates the steepness of the slope. The gradient is in fact equal to the force, but has an opposite sign. While minimizing the energy of a system, each atom is moved in the opposite direction to the force acting on it. On the other hand, second derivatives—called the Hessian matrix when represented in the form of a matrix—inform us about the curvature of the system and thus, are used to locate the stationary points of the PES. Generally, minimization methods go downhill in the PES from the starting point by gradually changing the coordinates.

The steepest descents method uses a ‘line search’ or an arbitrary step approach in a direction parallel to the net force in the system.92 A line search is finding three points along the line which will satisfy the requirement that the middle point is lower in energy than the other two points. At the minimum point obtained from a line search, the gradient will be orthogonal to the previous direction. Subsequently, following an iterative procedure or using a fitted function decreases the distance between the three points—which will confine the region that includes the minimum to a much smaller region. Arbitrary step approach, on the other hand, takes a step of an arbitrary step size in a direction parallel to the net force by changing the coordinates of the system, and iteratively adjusts the step size checking whether the energy is lowered in each step. When a step causes an increase in energy in the arbitrary step approach, the step size is decreased
(e.g. 0.5 of the last step) and thus, leaving the region of the minimum is prevented. The steepest descents method is especially useful at the initial part of the minimization process, since it can robustly approach the region of the minimum—even if it is far from the starting structure—following the direction of the gradient which relieves the most unfavorable features in the starting structure. However, this method follows a path that oscillates and continually overcorrects itself, which is generally not an efficient method in the case of a long narrow potential energy well.

The conjugate gradients method also has the gradients of all successive steps orthogonal but differs from the steepest descents method in having the direction of successive steps conjugate, rather than orthogonal. The direction of a new step is determined using Equation 2-11 in which $v_{k-1}$ and $v_k$ are successive directions, $g_k$ is the gradient and $\gamma_k$ is a scalar constant. The scalar constant can be calculated by different approaches like Fletcher-Reeves method (Equation 2-12) or Polak-Ribiere method (Equation 2-13). The line search algorithm or an arbitrary step approach is also used in the conjugate gradients method. In general, the conjugate gradients method proves to be a much better method after the initial strain is removed. Thus, the common practice is to use the steepest descents method at the beginning of a minimization procedure which is followed by a minimization using the conjugate gradients method.

$$v_k = -g_k + \gamma_k v_{k-1} \quad (2-11)$$

$$\gamma_k = \frac{g_k \cdot g_k}{g_{k-1} \cdot g_{k-1}} \quad (2-12)$$

$$\gamma_k = \frac{(g_k - g_{k-1}) \cdot g_k}{g_{k-1} \cdot g_{k-1}} \quad (2-13)$$

Apart from the two most popular first derivative methods mentioned in the last two paragraphs, the most popular second derivative method, the Newton-Raphson method, will be
described here shortly. For the multidimensional potential energy function, the minimum, \( x^* \), can be found using Equation 2-14 in which \( V'(x_k) \) is the force matrix at the point \( x_k \). Since the PES is not an ideal quadratic function, an iterative process is necessary to locate the minimum. As seen, Equation 2-14 requires finding the inverse of the Hessian matrix, \( V''(x_k) \), and this can be computationally too expensive. Thus, Newton-Raphson method or its variants are generally used for systems of small sizes. Another drawback of Newton-Raphson method is that the minimization can become unstable if the initial point lies in a region for which it is not appropriate to use the harmonic approximation. So, another method is generally used to approach the minimum before using the Newton-Raphson method. Many commercial program packages like Gaussian03\(^99\) generally use quasi-Newton methods which make use of the vectors representing the current and previous points to update the inverse Hessian matrix instead of constructing it for each iteration.

\[
x^* = x_k - V'(x_k)V''^{-1}(x_k)
\] (2-14)

Among the most popular first and second derivative minimization methods, first derivative methods are preferred for molecular mechanics calculations of large systems due to the high cost of calculating and storing the inverse of the Hessian matrix for such systems. Second derivative methods or their variants are generally used for quantum mechanics calculations which can be performed currently for small systems only.

While exploring the PES for a system, locating the saddle points as well as distinguishing them from the minima is necessary to understand the kinetics of the system. At a minimum point, all eigenvalues are positive, while there is one negative (i.e. imaginary) eigenvalue at a first-order saddle point—which is the saddle point type of interest for our purposes. Most of the times, the conversion of a minimum to another minimum occurs through the change of just one or two
coordinates. Thus, gradually changing the relevant coordinates or an appropriate combination of them (called reaction coordinates) and letting the system relax at each step while the reaction coordinates are kept fixed is used to approximately monitor the reaction pathway. Alternatively, the algorithms used for locating the minima can be modified to locate the saddle points. However, the initial point should be very close to the transition structure for such a method to succeed.

In fact, none of the numerical minimization methods can exactly locate the minima or the saddle points, but approximate it instead. Thus, the algorithms using these methods have to determine a convergence criterion to stop the minimization process. For this purpose, either the energy or the coordinates of the atoms of the system can be monitored and the calculation can be stopped once there is a smaller difference between the successive steps than a predetermined threshold value. Also the root mean square gradient for a system of N atoms can be calculated as shown in Equation 2-15 and be used as a convergence criterion.

\[
RMS = \sqrt{\frac{g^T g}{3N}}
\]  

(2-15)

2.5. Molecular Dynamics Simulation Methods

Molecular dynamics—which was first performed as early as 1957 for a condensed phase system\textsuperscript{108} and 1971 for liquid water\textsuperscript{109}—uses Newton’s laws of motion to generate successive configurations of a system. Solving the differential equations that represent Newton’s 2\textsuperscript{nd} law as shown in Equation 2-16, the trajectory of particles can be obtained. It is possible to solve these equations if the force acting on a particle is either zero or constant everywhere independent of the particle’s position. However, when the force acting on each particle depends on the positions of all other particles in the system, the equations become too impossible to solve analytically since the motion of particles are coupled. Thus, finite difference method, that breaks down the
integration into small stages each covering an infinitesimal $\delta t$ time window and assumes the force is constant during this time window, is used to integrate the equations of motion. Verlet algorithm\textsuperscript{110} is the most widely used method to integrate the equations of motion in molecular dynamics simulations and is summarized in Equations 2-17 as an example for such integration algorithms.

\[
\begin{align*}
\frac{d^2 x_i}{dt^2} &= \frac{F_{x_i}}{m_i} \\
\frac{d^2 y_i}{dt^2} &= \frac{F_{y_i}}{m_i} \\
\frac{d^2 z_i}{dt^2} &= \frac{F_{z_i}}{m_i}
\end{align*}
\]

(2-16)

\[
\begin{align*}
\mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \delta \mathbf{v}(t) + \frac{1}{2} \delta^2 \mathbf{a}(t) + ... \\
\mathbf{r}(t - \delta t) &= \mathbf{r}(t) - \delta \mathbf{v}(t) + \frac{1}{2} \delta^2 \mathbf{a}(t) - ... \\
\mathbf{v}(t + \frac{1}{2} \delta t) &= \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t)}{\delta t}
\end{align*}
\]

(2-17)

An important point before starting a molecular dynamics simulation is to decide the time step. A large time step can cause instabilities in the simulation, while a small time step significantly limits the ability to cover the phase space due to its high cost of computer time. Thus, a compromise between these two factors is necessary while deciding the time step. In simulations of flexible molecules, the time step is suggested to be about $1/10^{\text{th}}$ of the repeat period of the fastest motion of the molecule—which is C-H bond stretching in most biological systems with a period of 10 fs.\textsuperscript{92} This is why most molecular dynamics simulations either use a time step of 1 fs or freeze the hydrogen stretchings by constraining those bonds to their equilibrium bond lengths (e.g. SHAKE algorithm\textsuperscript{111}).
The initial coordinates of a system can be obtained from experimental or theoretical modeling studies. In either case, the initial velocities have to be assigned for each atom which can be done by randomly choosing a velocity for each atom from a uniform distribution, a simple Gaussian distribution or a Maxwell-Boltzmann distribution at a particular temperature. Maxwell-Boltzmann distribution is represented in Equation 2-18 which gives the probabilities for each atom $i$ with a mass of $m_i$ to have a velocity $v_i$ in the x direction at the simulation temperature $T$.\(^2\)

The initial velocities are then generally adjusted to provide an overall momentum of zero for the system.

$$P(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left(-\frac{m_i v_{ix}^2}{2k_B T}\right)$$  \hspace{1cm} \text{(2-18)}$$

Prior to the production phase molecular dynamics simulation, the system is relaxed to reach a stable configuration. Various properties of the system, especially its energy, are monitored during this phase until they reach stable values. For the simulations of proteins, it is necessary to constrain the protein backbone—or the entire protein in the case of explicit solvent MD simulations—at the beginning of the relaxation phase to conserve its secondary and tertiary structures. These constraints are then gradually reduced and finally released during the relaxation phase.
3.1. Introduction

*Trypanosoma cruzi* trans-sialidase (TcTS) specifically catalyzes the transfer of α-(2-3) sialic acid from sialoglycoconjugates to β-galactosyl glycoconjugates, retaining the anomeric configuration, while *Trypanosoma rangeli* sialidase (TrSA) is responsible only for the hydrolysis of α-(2-3) glycoside bonds. TcTS is found to act through a ping-pong mechanism (Figure 1-14) and shares with TrSA the first step in which the sialic acid is scavenged from the donor glycoconjugate with some nucleophilic participation of the enzyme. Whether this nucleophilic participation ends up collapsing into a covalent intermediate or remains as an oxocarbenium ion intermediate (2 in Figure 1-6) in each mechanism has been long discussed for enzymes of the similar families, even for the lysozyme.\(^{112-114}\) Strong nucleophilic participation in the transition structure of TcTS revealed by KIE studies pointed towards subsequent covalent intermediate formation.\(^{83,115}\) Quenching the catalytic reaction of a radioisotope-labeled ligand clearly indicated that TcTS mechanism involved a covalent bond formation between the ligand and the enzyme.\(^{83,115}\) A separate effort later achieved trapping and identifying the covalent intermediate of TcTS and TrSA using 2-deoxy-2,3-difluorosialic acid as the ligand (7 in Figure 1-13).\(^{77,78,86}\)

Despite elucidating the identity of the unusual nucleophile of a Tyr/Glu couple, there is still a possibility that activated fluoro-ligands follow a different reaction path than the natural ligand. This possibility should especially be investigated for TrSA for which there is no direct evidence for the natural substrate forming a covalent intermediate and the covalent intermediate formation might be the result of a modified mechanism forced by the activated substrate. The short lifetime of the covalent intermediate of TrSA revealed by the kinetic studies of the covalent
may also be an indicator of a finely-tuned mechanism which has just shifted from a strong nucleophilic participation to a covalent intermediate formation due to the effect of fluorine substituents of the substrate.

Since no difference in the mechanisms of TcTS and TrSA could be elucidated with the available experimental studies so far and DANA, an oxocarbenium-like molecule and a very efficient inhibitor for TrSA, is unable to inhibit TcTS\(^{44,72}\) possibly pointing to different transition states, we decided to investigate the mechanisms of the two enzymes computationally in detail. Is there a covalent intermediate formation for both TcTS and TrSA reactions with the natural substrates actually? Is the catalysis an S\(_N\)1-like or an S\(_N\)2-like mechanism in each case? How do the energetic barriers compare to each other? Can the energetic barriers explain why TcTS prefers transferring the sialic acid to the glycoconjugate rather than to a water molecule? Why can TrSA only catalyze hydrolysis? The answers to these questions can shed light to our understanding of the mechanism and can serve as a guide for future inhibitor design studies.

First, we investigated the covalent intermediate formation step (Figure 3-1) in TcTS and TrSA. For this purpose, we prepared potential energy surfaces (PES) using possible reaction coordinates—bond lengths—for both TcTS and TrSA mechanisms. Our study is presented here in a chronological order. At the time this study began, the KIE studies had just shown covalent intermediate formation in TcTS and had anticipated Tyr342 to be the nucleophile,\(^{52,83}\) as well as the first X-ray crystal structures of TcTS were just obtained.\(^{75}\) However, there was still a lot being debated about the mechanism by several experimental groups, even about whether Tyr342 could behave as a nucleophile or not.\(^{75}\) At the beginning, we assumed an S\(_N\)2-like reaction and used RC1 and RC2 (Figure 3-2) as our reaction coordinates. We were also limited by Gaussian03\(^{99}\) not allowing the use of combination of two or more bonds as a reaction coordinate.
However, the results from these 2-dimensional $S_n2$-like PESs as well as the results from hybrid calculations performed in AMBER10 that will be presented in the next chapter indicated the use of RC3 and RC4 would be a better choice. Thus, we performed 1-dimensional potential energy scans using RC3.

To compare with the trans-sialidase reaction, we have also investigated the energetics of the sialidase reaction of TcTS using an RC3 scan. Since the sialidase reaction differs from the trans-sialidase reaction only in the step after the covalent intermediate formation, only the differing step is studied for the sialidase reaction.

### 3.2. Methods

#### 3.2.1. Relaxation of the TcTS and TrSA Systems

The X-ray crystal structure of TcTS in complex with sialyllactose ([www.rcsb.org](http://www.rcsb.org), pdbID: 1S0I) is used as a starting point to investigate Tyr342 attack to the sialyllactose in TcTS. Everything except for the first 371 residues from chain A and sialyllactose is deleted. The mutated residue Ala59 is mutated back to an aspartic acid. The resulting pdb is fed into the Xleap module of AMBER9 molecular dynamics package$^{93,116,117}$ where the necessary H atoms and any other missing atoms are added as well as being solvated in a truncated octahedral cell of TIP3P$^{118}$ explicit water molecules. FF99SB$^{119,120}$ and Glycam04$^{121-123}$ force fields are used to construct the topology files. Sander module of AMBER9 is used first to minimize the energy of the system and then to relax it under periodic boundary conditions with a procedure explained in detail as follows:

1. 1000 steps of minimization with the protein and the ligand fixed,
2. 2500 steps of minimization,
3. 20 ps of molecular dynamics at constant volume with the protein and the ligand weakly restrained while increasing temperature from 0 K to 300 K with Langevin dynamics,$^{124}$
4. 100 ps of molecular dynamics at constant pressure while keeping temperature at 300 K with Langevin dynamics,
Throughout all minimizations and molecular dynamics, the cutoff used for nonbonded distances is 10 Å. All molecular dynamics is performed using SHAKE algorithm for H-involving bonds.

Since there is no X-ray crystal structure available for sialyllactose-bound form of TrSA, the X-ray crystal structure of TrSA in complex with 2-deoxy-2, 3-dehydro-N-acetyl neuraminic acid (DANA) (www.rcsb.org, pdbID: 1N1T) is used as a starting point to investigate Tyr342 attack to the sialyllactose in TrSA. Only the first 375 residues and DANA are conserved. The sialic acid part of the sialyllactose in 1S0I.pdb is superimposed with DANA in 1N1T.pdb and the appropriate coordinates for sialyllactose for TrSA are constructed in this way. The resulting pdb is completed for the missing H atoms and solvated as stated above for TcTS by the Xleap module. Minimization and relaxation of the system with the same procedure as in the TcTS case followed.

The X-ray crystal structure of TcTS covalent intermediate (www.rcsb.org, pdbID: 2AH2) is used as a starting point to investigate water attack to TcTS covalent intermediate. Everything except for the first 371 residues and sialic acid is deleted. The resulting pdb is completed for the missing H atoms and solvated as stated above for sialyllactose-bound TcTS by the Leap module of AMBER9. Minimization and relaxation of the system with the same procedure as in the sialyllactose-bound TcTS case followed.

### 3.2.2. Preparation of the TcTS and TrSA Model Systems

The relaxed geometry of TcTS with sialyllactose is used for model system preparation of TcTS to use in trans-sialidase reaction simulation. Sialyllactose, Trp312 and all residues that are within 3 Å of sialic acid—the ligand, 17 protein residues and 4 water molecules—are chosen to
constitute the model system. A hydrogen atom at the N-end and an OH group at the C-end are added for each residue for completeness. Total number of atoms in the model system is 464.

The relaxed geometry of TrSA with sialyllactose is used for the model system preparation of TrSA. Sialyllactose and all residues within 3 Å of sialic acid—the ligand, 14 protein residues and 13 water molecules—are chosen to constitute the model system. The N-end and C-end of included residues are completed with an H atom and an OH group as well. Total number of atoms in the model system is 420.

The relaxed geometry of TcTS covalent intermediate is used for model system preparation of TcTS to use in sialidase reaction simulation. Sialic acid, Trp312, Tyr119 and all residues that are within 3 Å of sialic acid—the ligand, 16 protein residues and 8 water molecules—are chosen to constitute the model system. A hydrogen atom at the N-end and an OH group at the C-end are added for each residue for completeness. Total number of atoms in the model system is 404.

3.2.3. ONIOM Input Setup

For ONIOM calculations of sialyllactose-bound TcTS using Gaussian03,99 the quantum region is chosen to consist of Asp59, Glu230, Tyr342 and the sialyllactose, a total of 139 atoms. All backbone atoms except for Cα atoms of Tyr342, Glu230 and Asp59 are frozen. All other atoms including the entire sialyllactose and water molecules are free. Only one link atom is necessary at the N-end of Glu230. The charge of the whole system and the charge of the quantum region are +2 and -2, respectively. The initial ONIOM calculations used Hartree-Fock level of theory with a 3-21G basis set as the high level of theory and AMBER force field for the low level of theory.

The quantum region for sialyllactose-bound TrSA is chosen to consist of Asp59, Glu230, Tyr342 and sialyllactose, a total of 139 atoms. All backbone atoms except for Cα atoms of Asp59, Glu230 and Tyr342 are frozen. The charge of the whole system and the charge of the
quantum region are +1 and -2, respectively. The initial ONIOM calculations for TrSA are performed with the same levels of theory as those used for TcTS.

For ONIOM calculations of TcTS covalent intermediate using Gaussian03, the quantum region is chosen to consist of Asp59, Glu230, Tyr342 and the sialic acid, a total of 139 atoms. All backbone atoms except for C$_a$ atoms of Tyr342, Glu230 and Asp59 are frozen. All other atoms including the entire sialic acid and water molecules are free. Only one link atom is necessary at the N-end of Glu230. The charge of the whole system and the charge of the quantum region are +1 and -2, respectively. The initial ONIOM calculations used Hartree-Fock level of theory with a 3-21G basis set as the high level of theory and AMBER force field for the low level of theory.

3.2.4. Potential Energy Surface Scan Setup

3.2.4.1. Simulating the trans-sialidase reaction of TcTS in a $S_n$2 way

A 2-dimensional potential energy surface (PES) for the covalent intermediate formation step (Figure 3-1) is prepared for TcTS by scanning the two distances:

(i) The distance between the anomeric C of sialic acid and the O atom of Tyr342 hydroxyl group (RC1 in Figure 3-2),
(ii) The distance between the O atom of Glu230 carboxylate group and the H atom of Tyr342 hydroxyl group (RC2 in Figure 3-2).

The attack distance is scanned between values of 1.46 Å and 3.46 Å with increments of 0.25 Å and the H transfer distance is scanned between values of 0.96 Å and 1.76 Å with increments of 0.2 Å. In order to get more precise values, the initial PES is reconstructed by gradually increasing the high level of theory in ONIOM calculations first to HF/6-31G* and then to B3LYP/6-31G* using the output of each level of theory as an input for the subsequent higher level of theory. A final PES is constructed by enlarging the quantum region to include the
arginine triad (Arg35, Arg245 and Arg314) that results in 220 atoms in the quantum region.

Figure 3-1. The covalent intermediate formation step

3.2.4.2. Simulating the trans-sialidase reaction of TrSA in a \( S_n2 \) way

For TrSA, a similar 2-dimensional PES as in the TcTS case is prepared by scanning the same two distances—RC1 and RC2. The attack distance is scanned between values of 1.42 Å and 3.42 Å with increments of 0.25 Å and the H transfer distance is scanned between values of 0.96 Å and 1.96 Å with increments of 0.2 Å. The same procedure of consecutive PES calculations by increasing the level of theory for TcTS is followed for TrSA. The final quantum region consists of Asp59, Glu230, Tyr342 and sialyllactose and the Arg triad (Arg35, Arg245 and Arg314).

3.2.4.3. Simulating the trans-sialidase reaction of TcTS in a \( S_n1 \) way

A 1-dimensional potential energy surface (PES) is prepared for TcTS by scanning the distance between the anomeric C of the sialic acid and the glycosidic O of lactose (RC3 in Figure 3-2). The starting geometry is the same as the one used for investigating the \( S_n2 \) mechanism. The attack distance RC3 is scanned between values of 1.45 Å and 3.85 Å with increments of 0.20 Å. The PES is constructed at the level of B3LYP with 6-31G* basis set and using a large quantum
region including the Arg triad (only the guanidinium regions of Arg35, Arg245 and Arg314 this time) as well as Asp59, Glu230, Tyr342 and sialyllactose (the 2nd and 3rd C atoms on the glycerol side chain of sialic acid and their substituents are excluded from the quantum region as well as the entire glucose and upper half of the galactose).

Figure 3-2. Reaction coordinates used to generate potential energy surfaces. The sialic acid and lactose parts of sialyllactose are colored in black and orange, respectively.
Since the RC3 scans attempting to simulate SN1 reaction above ended up in oxocarbenium ion formation, a subsequent RC1 scan is performed to see the energetic profile of covalent intermediate formation from the oxocarbenium ion. Starting from the optimized structure of RC3=3.05 Å in the very first RC3 scan above, we prepared a 1-dimensional PES by scanning the RC1 distance from 2.76 Å to 1.46 Å with increments of 0.10 Å.

3.2.4.4. Simulating the sialidase reaction of TcTS

The trans-sialidase catalysis and sialidase/hydrolysis catalysis only differ in the steps after the covalent intermediate formation. The attack of a water molecule to the covalent intermediate is also simulated by gradually pushing the nearest water molecule towards the anomeric C in the TcTS covalent intermediate until the sialic acid forms—from 3.46 Å to 1.46 Å. The two minima in the potential energy scans are optimized releasing all constraints and the resulting geometries of these unconstrained optimizations are confirmed to be minima by the all-positive frequencies computed for them. The PES is constructed at the level of B3LYP with 6-31G* basis set and using a quantum region including Asp59, Glu230, Tyr342 and the sialic acid—bound to Tyr342.

3.3. Results and Discussion

3.3.1. The SN2-like trans-Sialidase Reaction of TcTS

For TcTS and TrSA, the first step of the reaction, which is scavenging the sialic acid from the lactose, has been the subject of discussions about whether the nucleophilic participation from Tyr342 ends up collapsing into a covalent intermediate or remains as a stable oxocarbenium ion intermediate. Since computational methods allow us to work with the natural ligand by easily changing the F atom used to obtain X-ray crystals to an H atom, we investigated the detailed mechanism of TcTS as well as of TrSA. For this purpose, we obtained the potential energy surfaces (PES) of an SN2-like reaction by constraining two critical bond distances for TcTS and TrSA:
(i) RC1, the distance between Tyr342 hydroxyl O and the anomeric C of sialic acid, (ii) RC2, the distance between Tyr342 hydroxyl H and Glu230 carboxylate O.

The PES of TcTS (Figure 3-3) shows two minima, one that corresponds to the Michaelis complex and another to the covalent intermediate. The covalent intermediate minimum (RC1=1.46 Å, RC2=0.96 Å) is only 0.21 kcal/mol higher in energy than the Michaelis complex minimum (RC1= 3.21 Å, RC2= 1.16 Å), showing that the covalent intermediate is energetically accessible. Ideally, one would do a frequency calculation to characterize a stationary state, however, it was not possible to do it with such a large system in ONIOM. Thus, unconstrained optimizations starting from the best local minimum candidate are used to reach the closest minimum and to show that we are indeed at a stable local minimum. The covalent bond formation is confirmed by totally unconstrained optimization of the covalent intermediate structure in the TcTS PES, which resulted in a structure with RC1=1.50 Å and RC2=0.99 Å and an energy that is 0.56 kcal/mol lower than the TcTS Michaelis complex. Being 1.50 Å long, the new covalent bond formed is a little longer than that of the X-ray crystal structure (1.42 Å). The calculations of Bottoni et al. also showed a similar C-O bond distance of 1.522 Å for the covalent bond in a glycosyl-enzyme intermediate of lysozyme.

Additionally, the structure that corresponds to the TcTS Michaelis complex (RC1= 3.21 Å, RC2= 1.16 Å) shows a strong hydrogen bond between Tyr342 hydroxyl group and Glu230 carboxylate group. We also observe that the hydrogen transfer between Tyr342 and Glu230 can occur easily in TcTS with a low energy cost of 2.47 kcal/mol despite the high pKa difference of tyrosine and glutamic acid residues—10 and 4, respectively. To show the effect of the enzyme environment, we have calculated the energetic difference between (Tyr.O− + Glu.H) and (Tyr.OH + Glu−)—which just differ in the H atom position—using the exact same coordinates for these two residues from the QM/MM calculations including the entire active site. The difference is
found to be 9.97 kcal/mol if only the tyrosine and glutamic acid residues are included in the calculation, and it rises to 16.18 kcal/mol if the sialic acid is included as well as the two residues. These calculations show that the enzyme environment indeed significantly lowers the energetic barrier of a rather unfavorable hydrogen transfer reaction.

Analyzing the structures that correspond to the individual points on TcTS PES presents clues about the detailed mechanism of the reaction. The top left plot in Figure 3-4 indicates that lactose conserves its bond to the sialic acid (a ~1.45 Å long C-O bond) in the Michaelis complex until the bond breaks at RC1=2.21 Å. As RC1 is forced to become shorter, lactose keeps moving further away from the sialic acid. In short, lactose breaks its bond to the sialic acid much before that Tyr342 forms a covalent bond to the sialic acid, which happens at RC1~1.46 Å.

![Figure 3-3. Potential energy surface of TcTS for S_n2-like covalent intermediate formation step. Energies are in kcal/mol.](image-url)
While breaking its bond to the sialic acid, the lactose molecule simultaneously transfers a H atom from the Asp59 carboxylic acid group. This is shown clearly in the top right plot in Figure 3-4 as the distance of the glycosidic O to H atom in Asp59 carboxylic acid decreases abruptly from 2.32 Å at RC1=2.46 Å to 0.98 Å at RC1=2.21 Å.

Figure 3-4. Changes in geometric properties of sialic acid during S_N2-like covalent intermediate formation of TcTS. Pyramidalization is monitored by the dihedral angle of C3-C1-O5-C2 in sialic acid which becomes 0° for a completely planar oxocarbenium ion. Distances are in Å and dihedral angles are in degrees.
The pyramidalization around the anomeric C of the sialic acid is another important parameter to monitor since the planarity around the anomeric C is a sign of oxocarbenium ion formation. The bottom right plot in Figure 3-4 shows that the pyramidalization values are all negative for the region where RC1 is 2.46 Å or larger and all positive for the region where RC1 is 2.21 Å or smaller; that is, a change in the direction of the apex of the pyramid with the anomeric C as the apex has occurred. Thus, the planar oxocarbenium form of sialic acid must be forming in the region of 2.21 Å < RC1 < 2.46 Å. Accordingly, in this region, the bond between the anomeric C and its neighboring O5 atom in the sialic acid ring shortens as a result of the electron donation from the lone pair of O5 atom to the empty pz orbital of the positively charged anomeric C (bottom left plot in Figure 3-4). This C-O bond becomes shortest (1.27 Å) at the point of RC1 = 2.21 Å and RC2 = 1.76 Å.

In summary, this PES shows that the covalent intermediate formation is energetically plausible with a covalent intermediate almost equienergetic with the Michaelis complex. The observation of that the leaving group—lactose—breaks its bond to the sialic acid long before Tyr342 forms a covalent bond to the anomeric carbon and there is a metastable oxocarbenium-like intermediate forming before the covalent intermediate (in the valley right next to the covalent intermediate in Figure 3-3) led us to think that an S₅1-like mechanism is more likely for this step in TcTS. Thus, we performed 1-dimensional potential energy scans of RC3 (instead of RC1/RC2) to investigate the covalent intermediate formation mechanism further.

3.3.2. The S₅2-like trans-Sialidase Reaction of TrSA

The PES of TrSA (Figure 3-5) also has two minima that correspond to the Michaelis complex and the covalent intermediate. The covalent intermediate minimum (RC1 = 1.67 Å, RC2 = 0.96 Å) is 4.44 kcal/mol higher in energy than the Michaelis complex minimum (RC1 = 3.42 Å, RC2 = 1.76 Å) and is energetically accessible. The covalent bond formation is
confirmed by totally unconstrained optimization starting from the covalent intermediate structure at RC1=1.46 Å and RC2=0.96 Å, which resulted in a structure with RC1=1.50 Å and RC2=0.97 Å and an energy that is 1.09 kcal/mol higher than the TrSA Michaelis complex.

The structure that corresponds to the TrSA Michaelis complex (RC1=3.42 Å, RC2=1.76 Å) also has H-bonding between Tyr342 hydroxyl group and Glu230 carboxylate group, but not as strong as in the TcTS case. The hydrogen transfer between Tyr342 and Glu230 is much more difficult in TrSA with an energy barrier of 16 kcal/mol. When the H is located on Glu230 instead of Tyr342, the structure is about 10 kcal/mol higher in energy compared to the case in which H is retained on Tyr342. Favoring H to reside on Tyr342, but not on Glu230 for the Michaelis complex, TrSA differs from TcTS.

Figure 3-5. The potential energy surface of TrSA for SN2-like covalent intermediate formation step. Energies are in kcal/mol.
The top left plot in Figure 3-6 shows that lactose to sialic acid bond is conserved in the TrSA Michaelis complex until the bond is broken at RC1=1.92 Å. Similar to the TcTS case, lactose slowly moves away from the sialic acid as RC1 becomes shorter. However, TrSA differs from TcTS in requiring shorter RC1 values compared to TcTS to break the lactose to sialic acid bond—the bond was broken at RC1=2.21 Å in TcTS. This can clue a more $S_N2$-like mechanism for TrSA. The simultaneous hydrogen transfer from Asp59 carboxylic acid group to lactose is realized here, too, as seen in the top right plot in Figure 3-6.

The pyramidalization values are all negative for RC1 of 2.17 Å and larger values and all positive for RC1 of 1.92 Å and shorter values (bottom right plot in Figure 3-6). These data point that the planar oxocarbenium form of sialic acid forms in the region 1.92 Å < RC1 < 2.17 Å. This is also supported by the shortening of the bond between anomeric C and neighbouring O5 atom in this region, as seen in Figure 3-6.

3.3.3. The $S_N1$-like trans-Sialidase Reaction of TcTS

To investigate the $S_N1$-like mechanism, the bond distance between the anomeric C and the glycosidic O (RC3 in Figure 3-2) is scanned. This potential energy scan (Table 3-1) showed a stable oxocarbenium ion formation with an energy barrier of about 31 kcal/ mol (and an energy barrier of 22 kcal/ mol for the reverse reaction). Even when lactose is pulled further away from the oxocarbenium ion form of the sialic acid, no simultaneous attack of Tyr342 to the anomeric carbon is observed. The two minima are also confirmed by totally unconstrained optimizations which gave almost exactly the same geometries as those of the starting structures and with all-positive frequencies. As the oxocarbenium ion forms, the hydroxyl O atom of Tyr342 approaches the anomeric C and gets more polarized. This nucleophilic assistance is also reflected in the decrease of anomeric C charge as Tyr342 approaches.
Figure 3-6. Changes in geometric properties of sialic acid during S$_\nu$2-like covalent intermediate formation of TrSA. Pyramidalization is monitored by the dihedral angle of C3-C1-O5-C2 in sialic acid which becomes 0° for a completely planar oxocarbenium ion. Distances are in Å and dihedral angles are in degrees.

Following the RC3 scan that facilitated the formation of the oxocarbenium ion form of sialic acid, an RC1 scan is performed to simulate the conversion of the oxocarbenium ion to the covalent intermediate. For this purpose, the optimized geometry at RC3=3.05 Å (in which RC1=2.76 Å) is used as a starting geometry and Tyr342 hydroxyl group is pushed first to RC1=2.36 Å and then gradually closer to the anomeric C with increments of 0.10 Å. Table 3-2
shows that the energy barrier for oxocarbenium ion to form the covalent intermediate is very low (5 kcal/mol). Also, the covalent intermediate obtained from this RC1 scan is optimized without any constraints showing its stability and the final covalent intermediate structure is proved to be a minimum by its all-positive frequencies calculated.

Table 3-1. Results of RC3 scan for TcTS that show stable oxocarbenium ion formation

<table>
<thead>
<tr>
<th>RC3 (Å)</th>
<th>Energy (Hartree)</th>
<th>Relative energy (kcal/mol)</th>
<th>RC4 (Å)</th>
<th>RC1 (Å)</th>
<th>RC2 (Å)</th>
<th>Pyramidalization around anomeric C (°)</th>
<th>Charge on the anomeric C</th>
<th>Charge on Tyr342 hydroxyl O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.45</td>
<td>-3631.020164</td>
<td>0.00</td>
<td>2.49</td>
<td>3.80</td>
<td>1.68</td>
<td>-33.0</td>
<td>0.665</td>
<td>-0.863</td>
</tr>
<tr>
<td>1.65</td>
<td>-3631.010161</td>
<td>6.28</td>
<td>2.48</td>
<td>3.60</td>
<td>1.65</td>
<td>-28.4</td>
<td>0.628</td>
<td>-0.875</td>
</tr>
<tr>
<td>1.85</td>
<td>-3630.994516</td>
<td>16.09</td>
<td>2.50</td>
<td>3.42</td>
<td>1.63</td>
<td>-23.6</td>
<td>0.618</td>
<td>-0.885</td>
</tr>
<tr>
<td>2.05</td>
<td>-3630.980832</td>
<td>24.68</td>
<td>2.53</td>
<td>3.25</td>
<td>1.61</td>
<td>-18.2</td>
<td>0.619</td>
<td>-0.896</td>
</tr>
<tr>
<td>2.25</td>
<td>-3630.970787</td>
<td>30.98</td>
<td>2.57</td>
<td>3.11</td>
<td>1.59</td>
<td>-13.4</td>
<td>0.619</td>
<td>-0.906</td>
</tr>
<tr>
<td>2.45</td>
<td>-3630.974628</td>
<td>28.57</td>
<td>2.32</td>
<td>2.85</td>
<td>1.57</td>
<td>-8.2</td>
<td>0.634</td>
<td>-0.908</td>
</tr>
<tr>
<td>2.65</td>
<td>-3630.995837</td>
<td>15.27</td>
<td>1.03</td>
<td>2.84</td>
<td>1.61</td>
<td>-5.2</td>
<td>0.607</td>
<td>-0.906</td>
</tr>
<tr>
<td>2.85</td>
<td>-3630.994196</td>
<td>7.55</td>
<td>1.03</td>
<td>2.79</td>
<td>1.62</td>
<td>-3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.05</td>
<td>-3631.006762</td>
<td>8.41</td>
<td>1.03</td>
<td>2.76</td>
<td>1.63</td>
<td>-1.4</td>
<td>0.592</td>
<td>-0.907</td>
</tr>
<tr>
<td>3.25</td>
<td>-3631.006251</td>
<td>8.73</td>
<td>1.03</td>
<td>2.74</td>
<td>1.65</td>
<td>0.5</td>
<td>0.591</td>
<td>-0.906</td>
</tr>
<tr>
<td>3.45</td>
<td>-3631.006098</td>
<td>8.83</td>
<td>1.03</td>
<td>2.70</td>
<td>1.65</td>
<td>1.9</td>
<td>0.596</td>
<td>-0.907</td>
</tr>
<tr>
<td>3.65</td>
<td>-3631.005888</td>
<td>8.96</td>
<td>1.03</td>
<td>2.83</td>
<td>1.68</td>
<td>1.0</td>
<td>0.585</td>
<td>-0.900</td>
</tr>
<tr>
<td>3.85</td>
<td>-3631.005800</td>
<td>9.01</td>
<td>1.00</td>
<td>3.00</td>
<td>1.70</td>
<td>2.0</td>
<td>0.579</td>
<td>-0.888</td>
</tr>
<tr>
<td>1.45 nor</td>
<td>-3631.020164</td>
<td>0.00</td>
<td>2.49</td>
<td>3.80</td>
<td>1.68</td>
<td>-33.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.85 nor</td>
<td>-3631.005798</td>
<td>9.01</td>
<td>1.00</td>
<td>3.00</td>
<td>1.70</td>
<td>1.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

nor = optimization with restraints released. Charge values are Mulliken charges.

One can easily imagine that for the sialic acid transfer reaction to be completed the exact reverse reactions have to occur as seen in Figure 3-7 if lactose acts both as the sialic acid acceptor and donor. Thus, with the potential energy profiles obtained, the potential energy diagram for the complete trans-sialidase reaction can be built as seen in Figure 3-8.

3.3.4. The Sialidase Reaction of TcTS

The trans-sialidase catalysis and sialidase/hydrolysis catalysis only differ in the steps after the covalent intermediate formation. To be able to compare the trans-sialidase catalysis with sialidase catalysis energetically, the attack of a water molecule to the covalent intermediate is also simulated. For this purpose, the nearest water molecule is gradually pushed to the anomeric
Figure 3-7. The mechanism proposed for trans-sialidase catalysis reaction of TcTS. The donor and acceptor sugar moieties are colored in red and orange, respectively. In sialidase catalysis reaction of TcTS, a water molecule will take the role of the acceptor lactose.
Table 3-2. Results of RC1 scan for TcTS that show conversion of the oxocarbenium ion to the covalent intermediate.

<table>
<thead>
<tr>
<th>RC1 (Å)</th>
<th>Energy (Hartree)</th>
<th>Relative energy (kcal/mol)</th>
<th>RC2 (Å)</th>
<th>RC3 (Å)</th>
<th>RC4 (Å)</th>
<th>Pyramidalization around An.C (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.36</td>
<td>-3631.003782</td>
<td>10.28</td>
<td>1.61</td>
<td>3.13</td>
<td>1.03</td>
<td>6.0</td>
</tr>
<tr>
<td>2.26</td>
<td>-3631.002332</td>
<td>11.19</td>
<td>1.61</td>
<td>3.38</td>
<td>1.03</td>
<td>10.2</td>
</tr>
<tr>
<td>2.16</td>
<td>-3631.000186</td>
<td>12.54</td>
<td>1.58</td>
<td>3.40</td>
<td>1.03</td>
<td>12.5</td>
</tr>
<tr>
<td>2.06</td>
<td>-3630.997624</td>
<td>14.14</td>
<td>1.55</td>
<td>3.42</td>
<td>1.03</td>
<td>15.2</td>
</tr>
<tr>
<td>1.96</td>
<td>-3630.997617</td>
<td>14.15</td>
<td>1.47</td>
<td>3.44</td>
<td>1.03</td>
<td>18.0</td>
</tr>
<tr>
<td>1.86</td>
<td>-3631.002469</td>
<td>11.10</td>
<td>1.01</td>
<td>3.49</td>
<td>1.03</td>
<td>22.8</td>
</tr>
<tr>
<td>1.76</td>
<td>-3631.004610</td>
<td>9.76</td>
<td>1.00</td>
<td>3.56</td>
<td>1.03</td>
<td>25.8</td>
</tr>
<tr>
<td>1.66</td>
<td>-3631.007052</td>
<td>8.23</td>
<td>0.99</td>
<td>3.62</td>
<td>1.03</td>
<td>28.4</td>
</tr>
<tr>
<td>1.56</td>
<td>-3631.009020</td>
<td>6.99</td>
<td>0.99</td>
<td>3.68</td>
<td>1.03</td>
<td>31.0</td>
</tr>
<tr>
<td>1.46</td>
<td>-3631.015184</td>
<td>3.12</td>
<td>0.99</td>
<td>4.07</td>
<td>1.03</td>
<td>34.3</td>
</tr>
<tr>
<td>1.46 nor</td>
<td>-3631.015929</td>
<td>2.65</td>
<td>0.99</td>
<td>4.05</td>
<td>1.03</td>
<td>33.0</td>
</tr>
</tbody>
</table>

nor = optimization with restraints released.

C of the TcTS covalent intermediate until the sialic acid forms (Table 3-3). The three minima in this potential energy profile are optimized releasing all constraints and the resulting geometries of these unconstrained optimizations are confirmed to be the minima of this reaction by frequency calculations (Table 3-3).

Combining the potential energy profiles obtained, the complete potential energy diagram of the sialidase reaction of TcTS is constructed as seen in Figure 3-9. Comparing Figure 3-8 and Figure 3-9 shows us the importance of covalent intermediate formation. The potential energy barrier for water attacking the covalent intermediate is higher than the barrier for the acceptor lactose attacking. If the covalent intermediate did not form and the reaction was proceeding directly through the oxocarbenium ion form of sialic acid instead, the potential energy barriers would favor the water attack rather than the lactose attack. So the potential energy diagrams obtained in this study are in line with the experimental results showing that trans-sialidase catalysis is more efficient in the presence of acceptor sugar molecules and only sialidase catalysis is observed in the absence of acceptor sugar molecules (Table 1-1).
Figure 3-8. Potential energy diagram calculated for trans-sialidase catalysis reaction of TcTS.

Table 3-3. Results of RC3 scan that show a water molecule attacking the covalent intermediate in TcTS.

<table>
<thead>
<tr>
<th>RC3 (Å)</th>
<th>Energy (Hartree)</th>
<th>Relative energy (kcal/mol)</th>
<th>RC4 (Å)</th>
<th>RC1 (Å)</th>
<th>RC2 (Å)</th>
<th>Pyramidalization around anomeric C (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.46</td>
<td>-2857.099947</td>
<td>0.00</td>
<td>1.88</td>
<td>1.51</td>
<td>0.99</td>
<td>30.0</td>
</tr>
<tr>
<td>3.21</td>
<td>-2857.096624</td>
<td>2.09</td>
<td>1.85</td>
<td>1.52</td>
<td>0.99</td>
<td>28.8</td>
</tr>
<tr>
<td>2.96</td>
<td>-2857.089697</td>
<td>6.43</td>
<td>1.84</td>
<td>1.54</td>
<td>0.99</td>
<td>27.0</td>
</tr>
<tr>
<td>2.71</td>
<td>-2857.080102</td>
<td>12.45</td>
<td>1.84</td>
<td>1.57</td>
<td>0.99</td>
<td>24.4</td>
</tr>
<tr>
<td>2.56</td>
<td>-2857.071826</td>
<td>17.65</td>
<td>1.83</td>
<td>1.60</td>
<td>0.99</td>
<td>22.3</td>
</tr>
<tr>
<td>2.21</td>
<td>-2857.076973</td>
<td>14.42</td>
<td>1.63</td>
<td>2.70</td>
<td>1.57</td>
<td>-7.9</td>
</tr>
<tr>
<td>1.96</td>
<td>-2857.067301</td>
<td>20.49</td>
<td>1.56</td>
<td>2.90</td>
<td>1.54</td>
<td>-16.6</td>
</tr>
<tr>
<td>1.71</td>
<td>-2857.071060</td>
<td>18.13</td>
<td>1.40</td>
<td>3.06</td>
<td>1.53</td>
<td>-23.3</td>
</tr>
<tr>
<td>1.46</td>
<td>-2857.091428</td>
<td>5.35</td>
<td>0.99</td>
<td>3.22</td>
<td>1.53</td>
<td>-31.1</td>
</tr>
<tr>
<td>3.46 nor.</td>
<td>-2857.100446</td>
<td>-0.31</td>
<td>1.89</td>
<td>1.50</td>
<td>0.99</td>
<td>30.5</td>
</tr>
<tr>
<td>2.21 nor.</td>
<td>-2857.085089</td>
<td>9.32</td>
<td>1.77</td>
<td>2.49</td>
<td>1.52</td>
<td>5.4</td>
</tr>
<tr>
<td>1.46 nor.</td>
<td>-2857.091438</td>
<td>5.34</td>
<td>0.99</td>
<td>3.22</td>
<td>1.54</td>
<td>-31.2</td>
</tr>
</tbody>
</table>

nor = optimization with restraints released.
Figure 3-9. Potential energy diagram calculated for sialidase catalysis reaction of TcTS.

3.4. Conclusion

The results obtained by the calculations with ONIOM method confirm the covalent intermediate formation for both TcTS and TrSA. In both $S_N2$-like and $S_N1$-like approaches, the covalent intermediate is found to be almost equienergetic with the Michaelis complex. For TcTS, the trans-sialidase and sialidase catalysis reactions are also compared preparing their potential energy diagrams which pointed out the importance of covalent intermediate formation for trans-sialidase catalysis. The potential energy diagrams also showed that if the reaction proceeded directly through oxocarbenium ion, sialidase catalysis would be favored rather than trans-sialidase catalysis in TcTS even in the presence of an acceptor molecule. In the next chapter, a similar study using QM/MM method in AMBER10 will be presented.
CHAPTER 4
QUANTUM MECHANICS/MOLECULAR MECHANICS STUDY OF THE CATALYTIC MECHANISM OF TRYPANOSOMA cruzi TRANS-SIALIDASE USING AMBER

4.1. Introduction

In order to check whether the results presented in Chapter 3 are dependent on the QM/MM method used, a similar study on the mechanism of Trypanosoma cruzi trans-sialidase is done using a different QM/MM method—the QM/MM method in AMBER10. Constrained minimizations using the same reaction coordinates are performed for this purpose.

4.2. Methods

4.2.1. Model Preparation

The starting structures for simulations are prepared from the sialyl-lactose (sialic acid in TrSA case) bound high-resolution crystal structures of TcTS and TrSA (www.rcsb.org, pdbID: 1S0I & 1N1T). Only the catalytic domains of the proteins—the first 371 residues in TcTS and the first 375 residues in TrSA—are included for the models. All the crystallographic water molecules are deleted and lactose is replaced by a methoxy group bound at the anomeric C atom of the sialic acid using the LEAP module which also adds the hydrogen atoms at pH 7. However, the protonation states of the three active site residues, Tyr342, Glu230 and Asp59 (numbered as in TcTS), are adjusted in accord to the proposed mechanisms and will be mentioned for each case. Then, the systems are solvated in a truncated-octahedron water box of TIP3P water molecules. Parameters from FF99SB120,126 and Glycam04121-123 force fields are used. The systems are relaxed using the following procedure in AMBER9 suite:

1. Minimization for 1000 steps—the first 10 steps using the steepest descents method and the rest using the conjugate gradients method—fixing the positions of all the main chain atoms except for H and carboxylic O atoms. The cutoff used to truncate nonbonded pairs is 8.0 Å.
2. Molecular dynamics simulation for 1.5 picoseconds at 300 K at constant pressure with SHAKE used for fixing bonds that involve H atoms.
3. Minimization for 1000 more steps just as in step 1.
(4) Minimization for 1000 more steps as in step 1 except with much weaker constraints on the main chain atoms.
(5) Minimization for 1000 more steps as in step 1 with all constraints on atoms released.
(6) Molecular dynamics simulation of 5 picoseconds as in step 2 except that the constraints on the atoms are 1/5th of the initial value.
(7) Molecular dynamics simulation of 15 picoseconds as in step 2 except that the constraints on the atoms are 1/10th of the initial value.
(8) Molecular dynamics simulation of 5 picoseconds as in step 2 except that the constraints on the atoms are all released.
(9) Molecular dynamics simulation of 0.5 nanoseconds as in step 8.

4.2.2. Initial Minimization

All water molecules in the relaxed model system are deleted and a nonperiodic, vacuum model of TcTS system is simulated. 5000 steps of pure MM minimization for the entire system are followed by 5000 steps of QM/MM minimization. The initial pure MM minimization is performed by 10 cycles of steepest descent method followed by conjugate gradients method for the rest of the minimization with a 1000 Å cutoff for nonbonded interactions at 100 K—provided by a Langevin thermostat with 2 ps\(^{-1}\) collision frequency. The QM/MM minimization is performed with the same procedure as the pure MM minimization. PDDG/PM3,\(^{127}\) a semiempirical model, is used as the QM method and the QM region consists of Tyr342, Glu230, Asp59 and the ligand. The convergence criterion used for the SCF calculation is 10\(^{-4}\) kcal/mol and for the root mean square of the energy gradient is 10\(^{-4}\) kcal/mol Å.

4.2.3. QM/MM Potential Energy Surface Preparation

The potential energy surfaces (PES) of sialyl-transfer reaction of TcTS is prepared by constrained minimizations using QM/MM calculations of Amber10 suit. PDDG/PM3 method is used as the QM method and the QM region consists of Tyr342, Glu230, Asp59 and the ligand. The same procedure as the initial QM/MM minimization in vacuum is followed for 1000 steps except that the convergence criterion for the root mean square of the energy gradient is 10\(^{-2}\) kcal/mol.
mol Å this time and a harmonic biasing potential with a force constant of 500 kcal/mol Å² is used.

First, an S₈2-like mechanism is simulated using RC1 and RC2 as the reaction coordinates
and preparing a potential energy surface (PES). RC1 values are scanned from 3.55 Å to 1.35 Å
and RC2 values are scanned from 2.00 Å to 0.80 Å, using 0.1 Å increments for each coordinate.

An independent two dimensional PES is prepared using RC3 and RC4 as in Chapter 3. The
PES is constructed by scanning RC3 between values of 2.7 Å and 0.8 Å and RC4 between values
of 1.2 Å and 5.0 Å, using 0.1 Å increments for both coordinates. Using the most stable structure
in the energetic valley of products of this RC3-RC4 PES, a subsequent two dimensional scan is
done using RC1 and RC2 as the reaction coordinates. RC1 is scanned between values of 4.5 Å
and 1.0 Å and RC2 is scanned between values of 2.0 Å and 0.8 Å.

4.3. Results and Discussion

The PES shown in Figure 4-2 is prepared by scanning RC1 and RC2 and shows that Asp59
carboxylic acid transfers a proton to the glycosidic oxygen atom of the ligand before the
nucleophilic attack of Tyr342 to the anomeric carbon atom. However, the energy barrier for this
reaction is very high (78.8 kcal/mol) indicating that this reaction is not likely to happen. Also,
the sudden change seen on the PES close to the covalent intermediate might indicate RC1 and
RC2 are not proper reaction coordinates to monitor this reaction. One should still note that in the
PES the covalent intermediate lies in an energy valley and is 11 kcal/mol lower in energy than
the Michaelis complex, indicating the stability of the covalent intermediate.

The PES prepared scanning RC3 and RC4 instead, shown in Figure 4-3, gave a more
reasonable picture in terms of energetics with a barrier of 40.2 kcal/mol. However, this scan
does not form a covalent intermediate directly at the end but a stable oxocarbenium ion (i.e. a
Figure 4-1. Reaction coordinates used to generate potential energy surfaces. The sialic acid and lactose parts of sialyllactose are colored in black and orange, respectively.

Table 4-1. Properties of stationary points on potential energy surfaces

<table>
<thead>
<tr>
<th></th>
<th>Energy (kcal/mol)</th>
<th>RC1 (Å)</th>
<th>RC2 (Å)</th>
<th>RC3 (Å)</th>
<th>RC4 (Å)</th>
<th>Anomeric C to ring O distance (Å)</th>
<th>Pyramidalization around anomeric C (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RC1-RC2 scan:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michaelis complex</td>
<td>10.9</td>
<td>3.45</td>
<td>1.70</td>
<td>1.42</td>
<td>2.59</td>
<td>1.40</td>
<td>-31.3</td>
</tr>
<tr>
<td>Transition state</td>
<td>89.8</td>
<td>1.85</td>
<td>0.81</td>
<td>1.52</td>
<td>2.58</td>
<td>1.40</td>
<td>-16.9</td>
</tr>
<tr>
<td>Covalent intermediate</td>
<td>0.0</td>
<td>1.45</td>
<td>0.95</td>
<td>3.79</td>
<td>0.98</td>
<td>1.40</td>
<td>29.0</td>
</tr>
<tr>
<td><strong>RC3-RC4 scan:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michaelis complex</td>
<td>13.1</td>
<td>3.62</td>
<td>1.65</td>
<td>1.40</td>
<td>1.70</td>
<td>1.40</td>
<td>-32.5</td>
</tr>
<tr>
<td>Transition state</td>
<td>53.3</td>
<td>3.36</td>
<td>1.65</td>
<td>1.80</td>
<td>1.10</td>
<td>1.34</td>
<td>-9.0</td>
</tr>
<tr>
<td>Oxocarbenium ion</td>
<td>34.2</td>
<td>2.68</td>
<td>1.64</td>
<td>3.06</td>
<td>0.98</td>
<td>1.28</td>
<td>-0.4</td>
</tr>
<tr>
<td>Subsequent RC1-RC2 scan:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition state</td>
<td>33.7</td>
<td>2.25</td>
<td>1.00</td>
<td>4.87</td>
<td>0.95</td>
<td>1.31</td>
<td>7.8</td>
</tr>
<tr>
<td>Covalent intermediate</td>
<td>0.0</td>
<td>1.45</td>
<td>1.00</td>
<td>4.47</td>
<td>0.95</td>
<td>1.40</td>
<td>28.8</td>
</tr>
</tbody>
</table>
Figure 4-2. Potential energy surface for S$_{N}$2-like mechanism of TcTS covalent intermediate formation.

carbocation) is located on the PES. The oxocarbenium ion formation is also confirmed by the planarity around the anomeric carbon atom and shortening of the bond between the anomeric carbon and its neighboring oxygen atom as seen in Table 4-1. The PES also shows that the hydrogen transfer from Asp59 carboxylic acid to the oxygen of the methoxy group—representing the glycosidic oxygen—occurs prior to the bond breaking between the methoxy group and the sialic acid.

To complete the reaction, a subsequent two dimensional scan is performed using RC1 and RC2 as the reaction coordinates. The PES in Figure 4-4 shows that the oxocarbenium ion readily forms the covalent intermediate with a small energy barrier of 1.9 kcal/ mol. The covalent intermediate lies in an energy valley and is 3.2 kcal/ mol lower in energy than the Michaelis complex.
A potential energy diagram for the entire trans-sialidase reaction is constructed bringing all the information obtained from AMBER QM/MM calculations together and shown in Figure 4-5. We are able to reproduce the energetical plausibility of the covalent intermediate formation and confirm the oxocarbenium ion as a stable intermediate—shown in Chapter 3 with ONIOM method—using a different QM/MM method.

4.4. Conclusion

The results of the two different QM/MM methodologies described in Chapters 3 and 4 agree on the energetic plausibility of the covalent intermediate formation—showing it is almost equienergetic with the Michaelis complex—for both TcTS and TrSA. The trans-sialidase reaction of TcTS is found to proceed with an S_N1-like mechanism. Here, the term “S_N1-like mechanism” is used to describe that the cleavage of the bond between the glycosidic O and the anomeric C occurs before the bond formation between the anomeric C and Tyr342. However, the
Figure 4-4. Potential energy surface for S_N1-like mechanism of TcTS that shows covalent intermediate formation from the oxocarbenium ion.

Figure 4-5. Potential energy diagram of trans-sialidase reaction of TcTS using QM/MM.
mechanism is not a typical $S_N1$ mechanism in which the cationic intermediate formation occurs further away from the nucleophile without any nucleophilic assistance. Instead, the nucleophile, Tyr342, stands in such a position in the active site to assist the oxocarbenium ion formation. As the reaction proceeds from the Michaelis complex to the oxocarbenium ion, the anomeric C migrates away from the lactose and gets closer to Tyr342, which results in stabilization of the positive charge developed on the anomeric C by the polarized negative charge on the hydroxyl group of Tyr342. The spatial approach of Tyr342 to the anomeric C, the increasing negative charge on Tyr342’s hydroxyl O atom and the decreasing positive charge on the anomeric C during this process (shown in Table 3-1) support this view altogether. Overall, the active site modifies an otherwise $S_N1$ mechanism with nucleophilic participation and the new mechanism is a complex mechanism which has both $S_N1$-like and $S_N2$-like properties. This reaction is an example of the large spectrum of reaction mechanisms that lie between a strict $S_N1$ mechanism and a strict $S_N2$ mechanism.\textsuperscript{128,129}

Both of our methodologies point to a stable oxocarbenium ion intermediate, although with a small energetic barrier for collapsing into a covalent intermediate. Our calculations also show that the rate determining step in the TcTS mechanism is the cleavage of the bond between the lactose and the sialic acid. For the rate determining step, a late transition structure similar to an oxocarbenium ion is anticipated by the potential energy diagrams obtained. These results agree with the KIE results that were interpreted to be indicative of a transition structure with significant nucleophilic participation and little charge development.
CHAPTER 5
MOLECULAR DYNAMICS STUDY OF *Trypanosoma cruzi* TRANS-SIALIDASE AND *Trypanosoma rangeli* SIALIDASE

5.1. Introduction

Enzymes are far from being static structures and their dynamics can play significant roles in catalysis. The range of dynamics can extend from small conformational changes of a single side chain to correlated motion of one or more regions of the enzyme with time scales in the nanosecond to millisecond region and can be responsible for activation/inactivation of an enzyme.

Difference in catalytic functionality despite the high structural similarity point to minute differences between *Trypanosoma cruzi* trans-sialidase (TcTS) and *Trypanosoma rangeli* sialidase (TrSA) which may be related to differences in the dynamics of the two enzymes. Thus, the dynamical properties of the two active sites will be thoroughly investigated in this chapter by monitoring measurable properties for the all-atom simulated systems of free and ligated forms of TcTS and TrSA.

There are a number of questions which we focus on. The first and main question is what structural or dynamical differences between the two enzymes can cause the difference in catalytic activities. The second one arises from the recent study that succeeded to transform TrSA which has no trans-sialidase activity into a form that shows 10% of the trans-sialidase activity of TcTS using only 5 point mutations (TrSA\textsubscript{5mut}: Met95Val, Ala97Pro, Ser119Tyr, Gly248Tyr, Gln283Pro). Since none of these five residues are directly implicated in the chemical catalysis step, the change of conformational dynamics and/or structure due to these mutations should be responsible for the acquired trans-sialidase activity. Using the available X-ray crystal structures of all three enzymes—wild type TcTS, wild type TrSA and TrSA\textsubscript{5mut}—as
starting structures, a comparative study of all-atom molecular dynamics (MD) simulations of these systems can shed some light into their differences in detail.

The third question arises from the difference in DANA binding properties of TcTS and TrSA. DANA, a common inhibitor for sialidases, is significantly less efficient for TcTS ($K_i=12.29$ mM) compared to TrSA ($K_i=1.5 \mu$M). Detailed analysis of MD simulations of DANA-bound TcTS and TrSA can shed some light on the reasons of this difference.

The fourth question is related with the experimental results showing that lactose which acts as a sialic acid acceptor can not bind TcTS in the absence of sialic acid in the medium. This observation brings to mind the possibility of a structural change or a change in the dynamics in the active site due to occupation of the sialic acid binding site which will favor subsequent lactose binding. A thorough comparison of free TcTS and TcTS covalent intermediate MD simulations will be made searching for any structural or dynamical changes due to covalent intermediate formation or DANA binding that can modify the active site to a more proper binding site for lactose.

We also investigate how TrSA binds sialyllactose since there is no X-ray crystal structure available.

The active site residues are listed for TcTS and TrSA in Figure 5-1. Only 12 mutations are found out of 43 residues in the active site—when defined as the residues that have at least one atom in 10 Å of Tyr342’s hydroxyl O atom. From now on, the residue numbering of TcTS will be used for corresponding residues of both enzymes—unless otherwise stated—to prevent confusion and to provide better means to compare the two enzymatic systems.
<table>
<thead>
<tr>
<th>TcTS</th>
<th>Arg35</th>
<th>Leu36</th>
<th>Pro37</th>
<th>Asp51</th>
<th>Arg53</th>
<th>Phe58</th>
<th>Asp59</th>
<th>Asn60</th>
<th>Val95</th>
<th>Asp96</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg39</td>
<td>Ile40</td>
<td>Pro41</td>
<td>Asp55</td>
<td>Arg57</td>
<td>Phe62</td>
<td>Asp63</td>
<td>Asn64</td>
<td>Met99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TcTS</td>
<td>Pro97</td>
<td>Tyr119</td>
<td>Trp120</td>
<td>Leu176</td>
<td>Gly177</td>
<td>Gly178</td>
<td>Ala179</td>
<td>Gly180</td>
<td>Gln195</td>
<td>Ser229</td>
</tr>
<tr>
<td>TrSA</td>
<td>Ala101</td>
<td>Ser123</td>
<td>Trp124</td>
<td>Val180</td>
<td>Gly181</td>
<td>Gly182</td>
<td>Val183</td>
<td>Gly184</td>
<td>Gln199</td>
<td>Ser233</td>
</tr>
<tr>
<td>TcTS</td>
<td>Glu230</td>
<td>Pro231</td>
<td>Val232</td>
<td>Arg245</td>
<td>Asp247</td>
<td>Tyr248</td>
<td>Arg251</td>
<td>Pro283</td>
<td>Gly284</td>
<td>Ser285</td>
</tr>
<tr>
<td>TrSA</td>
<td>Glu234</td>
<td>Pro235</td>
<td>Ala236</td>
<td>Arg249</td>
<td>Asp251</td>
<td>Gly252</td>
<td>Arg255</td>
<td>Gln287</td>
<td>Asp288</td>
<td>Ser289</td>
</tr>
<tr>
<td>TcTS</td>
<td>Glu286</td>
<td>Ser287</td>
<td>His304</td>
<td>Pro305</td>
<td>Trp312</td>
<td>Arg314</td>
<td>Ser340</td>
<td>Ala341</td>
<td>Tyr342</td>
<td>Ser343</td>
</tr>
<tr>
<td>TrSA</td>
<td>Glu290</td>
<td>Ser291</td>
<td>His308</td>
<td>Pro309</td>
<td>Trp316</td>
<td>Arg318</td>
<td>Ser344</td>
<td>Gly345</td>
<td>Tyr346</td>
<td>Ser347</td>
</tr>
<tr>
<td>TcTS</td>
<td>Glu357</td>
<td>Glu362</td>
<td>Tyr364</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TrSA</td>
<td>Glu361</td>
<td>Asp366</td>
<td>Tyr368</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5-1. Corresponding active site residues of TcTS and TrSA that have at least one atom within 10 Å radius of Tyr342 hydroxyl O atom. The sequences are aligned using Basic Local Alignment Search Tool (BLAST). The residues that are different between the two enzymes are underlined and original numbering in each enzyme is used.

5.2. Methods

5.2.1. System Setup and Simulation Details

The initial structures of TcTS and TrSA are obtained from the Protein Databank (www.rcsb.org, PdbID: 1MS3, 2AH2, 1MS1, 1S0I, 1N1S, 2A75, 1N1T, 1WCS). All these 8 structures correspond to different points in the reaction coordinate (Table 5-2). Only the catalytic domain—the first 371 residues for TcTS, the first 375 residues for free and DANA-bound TrSA, the first 373 residues for TrSA covalent intermediate—and the ligand, if any, are kept and the rest of the proteins and any solvent molecules are excluded while preparing the systems for simulation (Figure 5-2). In addition to the above 8 structures available in the Protein Databank, the missing sialyllactose-bound form of TrSA is modelled in silico superimposing DANA-bound TrSA (1N1T.pdb) and the sialyllactose-bound TcTS in VMD program and imitating the sialyllactose pose seen in TcTS.

While preparing the input files for MD simulations several changes have been performed on the original X-ray crystal structure files to imitate the natural catalysis event better. In the systems that have the Asp59Ala mutation, the alanine59 is mutated back to an aspartic acid to
obtain the wild type enzyme. If the X-ray crystal structure is obtained as a homodimer, only chain A is preserved and chain B is deleted. And for the covalent intermediates, the F atom inserted into sialic acid to be able to obtain a stable crystal of the covalent intermediate is changed to a H atom as in the natural catalysis event.

The resulting pdb file is fed into the LEAP module of AMBER9 molecular dynamics package where the necessary H atoms and any other missing atoms are added as well as being solvated with $\approx 23,000$ TIP3P explicit water molecules and truncated octahedral boundary conditions.

Since the Asp59 is responsible for acid/base catalysis, its protonation state is changed to the aspartic acid form instead of the aspartate form which is found normally under physiological conditions. The protonation states of Asp59, Glu230 and Tyr342 change over the course of the reaction and are adjusted in the covalent intermediates accordingly (Table 5-1): Asp59 is in the aspartate form while Glu230 is in the glutamic acid form instead of the glutamate form due to the H transfer from Tyr342 to Glu230 in the reaction. In all other systems—free enzymes and Michaelis complexes—Asp59 and Tyr342 are protonated while Glu230 is unprotonated.

Amber.FF99SB and Glycam04 force fields are used to construct the topology files. Since the inhibitor DANA is not a standard residue in Glycam04, the parameters for this molecule had to be constructed. First the geometry for DANA is obtained modifying the standard terminal sialic acid geometry from Glycam04 using HYPERCHEM program and minimizing it to get planarity around its anomeric C atom. HF/6-31G* in GAUSSIAN03 package is used to produce RESP charges for DANA and these charges are then fed into the ANTECHAMBER module of AMBER9 package to produce the parameter and topology files using GAFF force field.
The PMEMD module of AMBER9 package is used first to relax the systems under periodic boundary conditions with a procedure explained in detail as follows:

(1) 1000 steps of minimization using steepest descent method using a cutoff of 8.0 Å for truncation of nonbonded pairs while restraining all backbone atoms of the enzyme with a restraint of the form $k \times (\Delta x)^2$ where the force constant is 5.0 kcal/mol Å$^2$.

(2) 15 picoseconds of molecular dynamics with a timestep of 1 femtosecond at 300 K and constant pressure using periodic boundary conditions and Berendsen weak-coupling algorithm$^{126}$ with a coupling time constant of 0.5 picoseconds and a pressure relaxation time of 0.5 picoseconds to provide constant temperature. SHAKE algorithm is also used to constrain bonds involving hydrogen atoms. The constraints for the backbone atoms and the nonbonded cutoff value used in the 1st step are kept.

(3) Three subsequent minimizations with the same procedure as in step 1 except decreasing the force constant for the backbone atoms to 2.0 kcal/mol Å$^2$, 0.1 kcal/mol Å$^2$ and finally to 0, respectively.

(4) 5 picoseconds of molecular dynamics with the same procedure as in step 2 except decreasing the restraint force constant to 1.0 kcal/mol Å$^2$.

(5) 15 picoseconds of molecular dynamics with the same procedure as in step 2 except decreasing the restraint force constant to 0.5 kcal/mol Å$^2$.

(6) 1 nanosecond of molecular dynamics with the same procedure as in step 2 except releasing the restraints and increasing the timestep to 2 femtoseconds.

Following the relaxation, 50 ns of production run is performed following the same procedure as in the last step of the relaxation.

5.2.2. Analysis Methods

The residue numbering for the analysis is adjusted for all systems simulated to prevent any confusion and to provide consistency between different systems as described in Figure 5-2. The residue numbering in TcTS is used for all systems throughout this chapter and Chapter 6 except otherwise stated.

While comparing the average structures of MD simulations, all residues that have at least one atom in 10 Å of Tyr342’s hydroxyl O atom are included.

The conformations of the residues in the protein are monitored through $\phi$, $\psi$, $\chi_1$ and $\chi_2$ dihedral angles which are defined in Figure 5-3. The histograms of $\chi_1$-$\chi_2$ populations of Trp312
and Tyr119 are prepared using 10° X 10° grids. The histograms of $\chi_1$-$\chi_2$ populations of Tyr342 and Leu36 are prepared using 5° X 5° grids.

Table 5-1. Properties of simulated structures

<table>
<thead>
<tr>
<th>PdbID</th>
<th>1MS3</th>
<th>2AH2</th>
<th>1MS1</th>
<th>1S0I</th>
<th>1N1S</th>
<th>2A75</th>
<th>1N1T</th>
<th>-</th>
<th>1WCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>TcTS</td>
<td>TcTS</td>
<td>TcTS-DANA</td>
<td>TcTS-SLT</td>
<td>TrSA</td>
<td>TrSA-DANA</td>
<td>TrSA-SLT</td>
<td>TrSA_Smut</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.65</td>
<td>1.60</td>
<td>1.80</td>
<td>1.60</td>
<td>1.64</td>
<td>1.95</td>
<td>1.60</td>
<td>-</td>
<td>2.80</td>
</tr>
<tr>
<td>Extra molecules in active site in X-ray crystal structure</td>
<td>-</td>
<td>Glycerol</td>
<td>Glycerol</td>
<td>-</td>
<td>SO$_4$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mutations to get X-ray crystal structure</td>
<td>N58F</td>
<td>N58F</td>
<td>N58F</td>
<td>N58F, D59A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Double conformations in X-ray structure in the catalytic domain</td>
<td>L36, S115, S116, S118, Y119, S122, H123, S257, Y342</td>
<td>E9, K12, I49, F58, D59, Y119, M170, T262, L264, M297</td>
<td>L36, D96, Y119, Y342</td>
<td>-</td>
<td>-</td>
<td>K11, F58, N60, E167</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prot. State of Asp59 in MD</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Prot. State of Tyr342 in MD</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Prot. State of Glu230 in MD</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The extra molecules found in the active site in X-ray crystal structures which are part of the solvent are also designated. The + and – represent aspartic acid and aspartate forms of Asp59 (and likewise for other residues), respectively. Only the mutations in the catalytic domain are mentioned in this table.

The salt bridges which reside in 10 Å radius of Tyr342’s hydroxyl oxygen are determined with the Saltbridge option of VMD. In this option, a salt bridge is considered to be formed if the distance between the oxygen atoms of acidic residues (aspartic acid and glutamic acid) and the nitrogen atoms of basic residues (arginine, histidine and lysine) are within the cutoff distance—
3.2 Å used here—in at least one frame. The results are used as a starting point to monitor the important hydrogen bonding interactions around the active site and additionally glutamine, asparagine and tyrosine residues closeby are included in the hydrogen-bond analysis. Hydrogen bonds among the residues are monitored with ptraj module of AMBER9. The criteria used for hydrogen-bonding are that heavy atom to heavy atom distance is 3.5 Å or less and 1st heavy atom-hydrogen-2nd heavy atom angle is 140° or larger.

Figure 5-2. The regions of the TcTS and TrSA that are used for MD simulation and analysis. The red regions are the lectin-like domains excluded in the MD simulations. The blue regions are the amino acid inserts that differ between TcTS and TrSA and so are excluded in residue numbering in the MD analysis part to provide consistency between the two enzymes and prevent confusion.

Figure 5-3. Description of $\phi$, $\psi$, $\chi_1$ and $\chi_2$ dihedral angles of a residue in a protein. $\phi$ is the dihedral angle of $C_{\text{precedent residue}}$-$C_\alpha$-$C$, $\psi$ is the dihedral angle of $N$-$C_\alpha$-$C$ $N_{\text{subsequent residue}}$, $\chi_1$ is the dihedral angle of $N$-$C_\alpha$-$C_\beta$-$C_\gamma$, $\chi_2$ is the dihedral angle of $C_\alpha$-$C_\beta$-$C_\gamma$-$C_\delta$. The dihedral angles being 0° means that the terminal atoms are oriented cis to each other.
The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) values are also calculated with ptraj module in AMBER9. RMSD is a custom method to measure the overall similarity of 3-dimensional structures. RMSD is defined with Eq. 5-1 in which $\delta_i$ is the distance between a pair of equivalent atoms in the two structures being compared and $N$ is the number of particles included in the calculation. The minimal RMSD value is found trying different superpositions. In our study, all $C_\alpha$ atoms of the enzyme are included in this calculation giving a measurement of the deviation of the enzyme over time from the starting structure. On the other hand, RMSF is a measurement of deviation of a particle from a reference structure—average structure of production run in our case—over time. RMSF is defined with Eq. 5-2 where $t$ is time and $\bar{x}$ is the time-averaged position of the same particle. RMSF can be used as a measure of the mobility of a particle in a simulation and is related to experimental B-factors. One can relate RMSF from simulations to experimental B-factors by the relationship given in Equation 5-3.

$$B\text{-factor} = 8\pi^2 (\text{RMSF})^2$$

Ptraj module of AMBER9 and VMD are both used to monitor the distances, angles and dihedral angles in the simulations.

### 5.3. Analysis

#### 5.3.1. Stability of the Molecular Dynamics Simulations

The root mean square deviation (RMSD) with respect to the starting geometries is generally used to show the stability of the simulations. Figure 5-4 shows that the RMSD value of $C_\alpha$ atoms of the protein with respect to the starting structure of the production MD simulation...
does not exceed 2.5 Å in any simulation as well as indicating that the systems in our simulations are stable.

5.3.2. Root Mean Square Fluctuation (RMSF) Analysis of MD Simulations of TcTS and TrSA

In this section, the root mean square fluctuations (RMSF) of C\textalpha atoms of the enzymatic systems will be thoroughly investigated. Since the loops extending from residue 142 to residue 147 and from residue 123 to 128 are more mobile for all TrSA species compared to all TcTS species, these loops will not be mentioned in each comparison repeatedly. Also, the mobilities of the loops extending in the range 22-25, 142-147, 163-169, 257-263 and 295-296, which are all far from the active site, will not be paid much attention due to their distance and irrelevance to the active site although their mobilities vary in each simulation.

5.3.2.1. Comparison of RMSF changes due to covalent intermediate formation in TcTS and TrSA

The main difference between TcTS and TrSA in RMSF changes due to covalent intermediate formation is seen for the RMSF of the loop bearing Pro283 in TcTS—which is equivalent to Gln283 in TrSA (Figure 5-1). Comparing Figures 5-5 and 5-6 shows that there is significant RMSF increase for this loop in TcTS due to the covalent intermediate formation while a significant RMSF decrease occurs in the case of TrSA. A similar trend is observed for the loop bearing Tyr248 in TcTS—which is equivalent to Gly248 in TrSA. The loop bearing Tyr248/Gly248 lies right next to the loop bearing Pro283/Gln283.

Another difference is observed for the loop bearing Trp312. In TcTS, lower RMSF values of this loop are seen in the covalent intermediate compared to the free enzyme, while the reverse trend is seen in TrSA. This difference can be related to the fact that the loop bearing Trp312 moves outwards into the exterior region forming a more open active site in TrSA covalent intermediate and unligated TcTS MD simulations, while this loop remains mainly intact in TcTS
Figure 5-4. Root mean square deviation (RMSD) of MD simulations with respect to the initial structure of the production runs. A) TcTS species B) TrSA species
covalent intermediate and unligated TrSA MD simulations. This active site opening will be
discussed later in this chapter. For the β-strand bearing Asp96, TrSA covalent intermediate has
higher RMSF values compared to free TrSA, while almost no difference is seen between free
TcTS and TcTS covalent intermediate.

5.3.2.2. Comparison of RMSF between DANA-bound TcTS and TrSA simulations

When the RMSF values of DANA-bound forms of TcTS and TrSA catalytic regions are
compared, the only difference is seen in the three loops bearing Tyr248, Trp312 and Asp96,
being higher in DANA-bound TrSA (Figure 5-7 and 5-8).

For both TcTS and TrSA, DANA binding promotes a decrease in the mobilities of the
three loops bearing Tyr248, Pro283 and Trp312 (Figure 5-7 and 5-8).

5.3.2.3. Comparison of RMSF between wild-type TcTS, wild-type TrSA and TrSA5mut
simulations

The RMSF values of TrSA5mut for the three loops bearing Tyr248, Pro283 and Tyr119 are
lower compared to wild type TrSA, and are close to the values observed for wild-type TcTS
(Figure 5-9). TrSA5mut has lower RMSF values for the loop bearing Trp312 compared to TrSA,
too, however, corresponding RMSF values for wild-type TcTS are much higher than both TrSA
forms. The locations of three out of the five point mutations (Met95Val, Ala97Pro, Ser119Tyr,
Gly248Tyr, Gln283Pro) performed on TrSA are observed to change their local mobility and
approach to that of TcTS. No effect on the local flexibility of the two mutations surrounding
Asp96 is observed.

5.3.2.4. Comparison of RMSF between sialyllactose-bound TcTS and sialyllactose-bound
TrSA

There are two main differences between the RMSF values of the active sites of
sialyllactose-bound forms of TcTS and TrSA (Figure 5-10). The loop bearing Pro283 has higher
RMSF in sialyllactose-bound TcTS than sialyllactose-bound TrSA, while the loop bearing
Trp312 has higher RMSF in sialyllactose-bound TrSA as seen in Figure 5-10. The high RMSF values for the loop bearing Trp312 is related to reorientation of Trp312 to adopt the binding conformation and will be explained in detail in the discussion of sialyllactose binding of TrSA.

5.3.2.5. Comparison of experimental B-factors of free and covalent intermediate forms of TcTS and TrSA

The comparison in Figure 5-11 shows that B-factors of free TcTS and TcTS covalent intermediate are almost identical except from that the loop bearing Asp59 is more flexible in the covalent intermediate. The comparison of B-factors of free TrSA and TrSA covalent intermediate (Figure 5-12) does not really inform us about the flexibility changes due to covalent intermediate formation since the resolution of them are different—1.64 Å and 1.95 Å, respectively.

5.3.2.6. Comparison of experimental B-factors between DANA-bound TcTS and TrSA X-ray crystal structures

No noticeable effect of DANA binding is seen on the B-factors of the active site in TcTS (Figure 5-13). The same trend is seen in Figure 5-16 for TrSA.

The comparison of Figure 5-13 and Figure 5-16 shows that DANA-bound TcTS has higher B-factors than DANA-bound TrSA for the three loops bearing Tyr248, Pro283 and Trp312. Additionally, the loop bearing Asp59 has higher B-factors in DANA-bound TrSA than in DANA-bound TcTS.

5.3.2.7. Comparison of experimental B-factors between wild-type TcTS, wild-type TrSA and TrSA5mut X-ray crystal structures

The trace of B-factors of TrSA5mut is very flat compared to TcTS and TrSA due to its lower resolution (Figure 5-15). Since the resolution of wild-type TcTS, wild-type TrSA and TrSA5mut are 1.65 Å, 1.64 Å and 1.95 Å, respectively, we can only compare the B-factors of TcTS and TrSA. The three loops bearing Tyr248, Pro283 and Trp312 have higher RMSF values in TcTS.
Figure 5-5. $C_\alpha$ RMSF comparison of free TcTS with TcTS covalent intermediate. The arrows show the RMSF differences in the active site.

Figure 5-6. $C_\alpha$ RMSF comparison of free TrSA with TrSA covalent intermediate. The arrows show the RMSF differences in the active site.
Figure 5-7. $\mathrm{C}_\alpha$ RMSF comparison of free and DANA-bound forms of TcTS. The arrows show the RMSF differences in the active site.

Figure 5-8. $\mathrm{C}_\alpha$ RMSF comparison of free and DANA-bound forms of TrSA. The arrows show the RMSF differences in the active site.
Figure 5-9. $C_\alpha$ RMSF comparison of TrSA, TcTS and TrSA$_{5\text{mut}}$. The differences in the active site are indicated with arrows.

Figure 5-10. $C_\alpha$ RMSF comparison of sialyllactose-bound forms of TcTS and TrSA. The differences in the active site are indicated with arrows.
while the loop bearing Tyr119 has higher RMSF values for TrSA.

5.3.2.8. Comparison of RMSF values from MD simulations and B-factors from X-ray crystal structures

The analysis of RMSF values of MD simulations showed higher mobility in TrSA than TcTS for the loops bearing Tyr248, Pro283 and Trp312, while the experimental B-factors showed higher mobility for these loops in TcTS. These 3 loops reside back to back at the outer surface of the enzymes. The reason of this inconsistency is most likely the crystal contacts observed for these three loops in TrSA. When the crystal contacts are investigated using the symmetry operations on X-ray crystal structure with SwissPdbViewer, strong interactions from a neighboring enzyme is observed for all these 3 loops constraining their motion and resulting in unnatural and low B-factors for TrSA (Figure 5-14). No such contacts exist for TcTS X-ray crystal structure in these three loops. Thus, the RMSF results obtained from MD simulations—which are free from the crystal contact problem—should be more reliable for these regions. Based on the RMSF analysis results, we observe that these three loops are more rigid in TcTS compared to TrSA due to the nonidentical residues, namely Tyr248/Gly248 and Pro283/Gln283, in these particular loops. This rigidity is important since Trp312 forms the aglycon binding site of the catalytic cleft. The two loops bearing Tyr248 and Pro283 are found to stay intact even in free TcTS MD simulation, in which active site opening due to outwards motion of Trp312 loop is observed.

5.3.3. Comparison of Average Structures from MD Simulations of TcTS in Ligated and Unligated Forms

In this section, the average structures of 50-ns MD simulations of different forms of TcTS are compared. Overall, they look very similar with RMSD values less than 1.1 Å when their backbone atoms are superimposed. Only the residues that adopt different conformations in the
Figure 5-11. Comparison of crystal structure Cα B-factors of TcTS covalent intermediate with its free form. The difference in active site B-factors is shown with an arrow.

two structures being compared are mentioned and all others can be assumed to be very similar.

**5.3.3.1. Comparison of unligated TcTS with TcTS covalent intermediate average structures**

Tyr342, Glu230, Asp59, Leu36 and Gln286 conformations differ due to change of $\chi_1$ dihedral angles. Ser287 conformation is also different due to a slight backbone motion. Trp312 is found in the binding conformation in the covalent intermediate while the loop opening observed in unligated TcTS results in a more open active site and different positioning of Trp312. In both structures Tyr119 is in the “down conformation” (Figure 1-9). It should also be noted that covalent intermediate formation induced a slight outwards motion in the loop structure causing Tyr119 backbone to lie higher in the active site.

**5.3.3.2. Comparison of unligated TcTS with DANA-bound TcTS average structures**

Tyr119 and Leu36 conformations differ in unligated and DANA-bound forms of TcTS due
Figure 5-12. Comparison of crystal structure $C_\alpha$ B-factors of TrSA covalent intermediate with its free form.

Figure 5-13. $C_\alpha$ B-factor comparison of free and DANA-bound forms of TcTS X-ray crystal structures. The arrows indicate the loops whose mobility differ between DANA-bound forms of TcTS and TrSA.
Figure 5-14. $C_\alpha$ B-factor comparison of free and DANA-bound forms of TrSA X-ray crystal structures.

Figure 5-15. $C_\alpha$ B-factor comparison of TcTS, TrSA and TrSA$_{5\text{mut}}$ X-ray crystal structures.
Figure 5-16. Crystal contacts of TrSA around the loops bearing Trp312, Gln283 and Gly248 (shown in licorice form). Tyr342 is also shown in licorice form to clarify relative orientation of the active site. A neighboring enzyme (its backbone shown in red) limits the mobility of these three loops.

to changing $\chi_1$ dihedral angles. Tyr119 is in the “up conformation” (Figure 1-9) in DANA-bound TcTS and DANA-binding induces a small change in the loop structure of Tyr119 causing its backbone to lie higher in the active site. Tyr342 has moved up to interact with DANA in the DANA-bound enzyme, which also causes the loop bearing Tyr342 to move up and the Ala341 conformation to change. Trp312 is found in the binding conformation in the DANA-bound TcTS. Overall, we observe that DANA-binding locks the arginine triad, Tyr342, Glu230 and Trp312 into position to embrace the ligand tightly.

5.3.3.3. Comparison of unligated TcTS with sialyllactose-bound TcTS average structures

When the average structures of unligated and sialyllactose-bound forms of TcTS are compared, Glu230, Tyr119, Asp59, Leu36 and Gln286 conformations differ due to changing $\chi_1$ dihedral angles. Ser287 and Ser288 conformations are also different due to a slight loop motion. In the sialyllactose-bound TcTS, Tyr119 is in the “up conformation” (Figure 1-9) and the loop bearing Tyr119 approaches a little to the ligand. Tyr342 $\chi_1$ is different in the complex to interact with the ligand, the loop bearing Tyr342 changes its conformation slightly and Ala341
conformation is different. Glu230 has moved to a conformation unfavorable for a hydrogen-bonding interaction with Tyr342. Asp59 position is also different in the complex due to loop motion on top of conformational change. Trp312 is found in the binding conformation in the complex. Overall, sialyllactose binding also locks the arginine triad, Tyr342, Trp312 and Tyr119 into position to embrace the ligand tightly. However, Asp59 is not found in a conformation to interact with O3 atom of the galactose part of the ligand (glycosidic O in Figure 1-8) and Glu230 is not hydrogen-bonded to Tyr342.

5.3.4. Comparison of Average Structures from MD Simulations of TrSA in Ligated and Unligated Forms

In this section, the average structures of 50-ns MD simulations of different forms of TrSA are compared. Overall, they look very similar with RMSD values less than X Å when their backbone atoms are superimposed. Only the residues that adopt different conformations in the two structures being compared are mentioned and all others can be assumed to be very similar.

5.3.4.1. Comparison of unligated TrSA with TrSA covalent intermediate average structures

Ile36, Gln283 and Tyr342 conformations differ between the unligated TrSA and the covalent intermediate due to changing $\chi_2$ dihedral angle. Asp59 conformation is also different due to $\chi_1$ difference. Trp312 is in the binding conformation in unligated TrSA while its conformation has changed on top of a slight outwards movement of its loop in the covalent intermediate. The loop bearing Tyr342 stays intact even after covalent intermediate formation. The arginine triad rearranges in the covalent intermediate to embrace the carboxylate of the sialic acid tightly. Also, the loop bearing Asp247 approaches more to the sialic acid and Asp247 conformation changes to interact with Arg245 in the covalent intermediate.
5.3.4.2. Comparison of unligated TrSA with DANA-bound TrSA average structures

Ile36 and Gln283 conformations differ between the unligated TrSA and the DANA-bound TrSA due to changing $\chi_2$ dihedral angle. Asp59 conformation is also different due to $\chi_1$ difference. The arginine triad rearranges in the DANA-bound TrSA to embrace the carboxylate of the sialic acid tightly. Also, Trp312 conformation resembles more to the binding conformation with only a slight change in its loop, and the loop bearing Asp247 lies closer to the sialic acid in DANA-bound TrSA.

5.3.4.3. Comparison of unligated TrSA with sialyllactose-bound TrSA average structures

Ile36 and Gln283 conformations differ between the unligated TrSA and the sialyllactose-bound TrSA due to changing $\chi_2$ dihedral angle. Phe58 and Asp59 conformations are also different due to $\chi_1$ difference. The Phe58 conformation in sialyllactose-bound TrSA provides Phe58 to embrace the lactose part of the ligand. The arginine triad rearranges in sialyllactose-bound TrSA to embrace the carboxylate of the sialic acid tightly. Sialyllactose binding also promotes Trp312 conformation to change to the binding conformation and movement of its loop to make stacking interactions with the lactose part of the ligand. The loop bearing Asp247 and the loop bearing Ser119 are also closer to the active site in sialyllactose-bound form.

5.3.5. Comparison of Sialyllactose-Bound TcTS and Sialyllactose-bound TrSA Average Structures

The sialyllactose pose is very similar in the two enzymes. In the active sites, the conformations of Phe58 and Asp59 differ for the two enzymes. There is also a slight difference in the loop bearing Tyr342 which might be related to presence of Ala341 in TcTS instead of Gly341 in TrSA. There are also several mutations that are constraining the active sites differently:

(i) Val95 in TcTS is Met95 in TrSA which results in a more restricted active site in TrSA around O4 atom of sialic acid.
(ii) Leu176 in TcTS is Val176 in TrSA which results in a more restricted active site in TcTS around N-acetyl group of sialic acid.
(iii) Pro283 in TcTS is Gln283 in TrSA which can affect the Trp312 conformation.

5.3.6. Comparison of DANA-bound TcTS with DANA-bound TrSA Average Structures

The main difference between the two average structures is that the glycerol branch of DANA (3 in Figure 1-6) adopts different conformations in the two enzymes. Asp59 conformation is also different due to $\chi_1$ change. Asp59 conformation differs due to $\chi_1$ change. There is also a slight difference in the loop bearing Tyr342 which might be related to the presence of Ala341 in TcTS instead of Gly341 in TrSA. Trp312 is seen in the binding conformation in DANA-bound TcTS.

5.3.7. Behavior of Key Residues of TcTS and TrSA in MD

Among the key residues in the active site, Trp312 and Asp59 are on flexible loops, Tyr119 resides on a very short 3-10 helix which is on a flexible loop, Glu230 and Asp96 lie on $\beta$-strands which are parts of the $\beta$-propeller architecture that is characteristic for neuraminidases while Tyr342 lies on a small loop surrounded by $\beta$-strands. The conformations and overall motion of these key residues will be investigated in this section.

5.3.7.1. Behavior of Trp312

The conformation of Trp312 is found very different in TcTS and TrSA X-ray crystal structures which can be distinguished by both $\chi_1$ and $\chi_2$ dihedral angles. The behavior of Trp312 side chain is monitored preparing a 2-dimensional histogram of its $\chi_1$ and $\chi_2$ dihedral angles during the simulations. The histograms in Figure 5-17 show that there is a dominant binding mode of Trp312 which can be described around the region of $\chi_1 = -50^\circ$ and $\chi_2 = -75^\circ$. Among TcTS species, the binding mode is highly populated by Trp312 in DANA or sialyllactose-bound cases while in the free enzyme or the covalent intermediate the binding mode is visited much less although it is still the dominant conformation.
Interestingly, Trp312 also visits the binding mode in the free and sialyllactose-bound TrSA although populating it to a lower extent compared to the TcTS simulations (Figure 5-18). The reasons of this population difference are found to be related to the starting conformation ($\chi_1 = 50^\circ$ and $\chi_2 = 75^\circ$) obtained from the X-ray crystal structures as well as the flexibility difference in this region between TcTS and TrSA and will be discussed later in this chapter. In MD simulations of the covalent intermediate and the DANA-bound forms of TrSA as well as the TrSA$_5$mut, Trp312 is seen to sample different conformations which are different in $\chi_1$ or $\chi_2$ dihedral angle from the binding mode seen in TcTS but closer to their initial conformation.

Since Trp312 resides on a flexible loop, the conformation and the motion of the backbone is also monitored through the RMSD of the backbone atoms of the loop—residues 307 to 317 (Figure 5-19). The high RMSD values for free TcTS and TrSA covalent intermediate indicate significant change in the architecture of the loop. A similar motion is observed for TcTS covalent intermediate and TcTS-DANA although to a much smaller extent. The visualization of the all-atom simulation shows that the high RMSD values are an indicator of a loop motion which forms a more open active site. However, when all the available X-ray crystal structures of empty and ligated TcTS and TrSA are superimposed, the backbone of the loop bearing Trp312 is seen to adopt the same pose in all. A closer look shows that both ends of the loop are attached at Lys309 and Arg314 strongly to neighboring residues with two strong H-bonding interactions. Especially, the salt bridges strongly connecting several loops together to finally attach to Lys309 is interesting. Lys309 side chain amino group extends to form H-bonds to both Asp337 and Asn339, while Glu338 in between forms a salt bridge to Arg7. This chain of strongly jointed residues can induce motion of the loop or its immobilization under different circumstances. Since
Arg7 lies close to the N-terminal of TcTS, the position of N-terminal is important for the loop bearing Trp312.

Figure 5-17. Histograms of Trp312 $\chi_1$-$\chi_2$ populations in MD simulations for TcTS species. Dihedral angles are in degrees.

The mobility of the loop bearing Trp312 seen in MD simulations, which is reflected neither in multiple conformations nor in larger B-factors for the loop in the X-ray crystal structure, is of curiosity. The reason of this difference can be related to the close crystal contacts observed. For example, in the X-ray crystal structure of empty TcTS, the two monomers forming
the homodimer lie very close to each other around this N-terminal region as well as some other regions. Direct H-bond distances less than 3.5 Å and water mediated H-bond distances less than
3 Å are observed at the contact surface around the N-terminal region and around residues 336 and 338. These close contacts can serve to lock the N-terminal to a certain position which will also implicate a certain position for the loop of Trp312 in the X-ray crystal structures.

Another point to mention is that the opening of Trp312 loop is only seen after the Tyr119 $\chi_1$ angle takes the value of ~180° (the “down” conformation of Tyr119, Figure 1-9). (Figure 5-20) It becomes easier for Trp312 and its loop first to move backwards and then move randomly forming a more open active site only after the stacking interaction of Trp312 with Tyr119 is lost due to Tyr119 side chain swinging down into the active site.

5.3.7.2. Behavior of Tyr119

Dual conformation is observed for Tyr119 residue which lies at the periphery of the active site in X-ray crystal structures of free, DANA-bound and covalent intermediate forms of TcTS. The two conformations observed differ in $\chi_1$ dihedral angle; in one conformation Tyr119 side chain swings down into the active site while in the other conformation it points up extending into the exterior medium (Figure 1-9). Figure 5-21 shows the histograms of Tyr119 $\chi_1-\chi_2$ dihedral angles in MD simulations of TcTS species and TrSA$_{5mut}$. Different $\chi_2$ angles at the same $\chi_1$ dihedral angle represent a flipping of the phenyl ring around itself. The histograms show a very localized “down” conformation for free and covalent intermediate forms of TcTS and TrSA$_{5mut}$ and a very localized “up” conformation for the sialyllactose-bound TcTS. Interestingly, in DANA-bound TcTS simulation, Tyr119 visits 3 different conformations which differ in $\chi_1$. The third conformation also extends out into the exterior medium. Another point to notice is that the “down” conformation—$\chi_1= 180^\circ$—must be very tightly held due to its interactions with the enzyme that even phenyl-ring flipping ($\chi_2$ change) is not observed.

Since Tyr119 resides on a very short 3-10 helix which is surrounded by flexible loops, the conformation and motion of the loop is also monitored through the RMSD of the backbone.
Figure 5-19. RMSD (in Å) of the loop that bears Trp312 in MD simulations without fitting to the initial frame.
Figure 5-20. Correlation of RMSD of the backbone of Trp312’s loop without fitting to the initial conformation and the $\chi_1$ dihedral angle of Tyr119 during MD simulations. RMSD values more than 3 Å, implicating a loop opening, is only observed in case of Tyr119 conformation in which the hydroxyl group swings down into the active site ($\chi_1 = \sim 180^\circ$).

atoms—residues 114 to 124 (Figure 5-22). The higher RMSD values observed for covalent intermediate and DANA-bound forms of TrSA as well as the TrSA$_{5\text{mut}}$ for the loop of...
Ser119/Tyr119 are indicators of a change of loop architecture although not significant as confirmed by visualization of the MD simulations. No such motion is observed for any of TcTS MD simulations.

5.3.7.3. Behavior of Phe58

Since different Phe58 conformations are observed for sialyllactose-bound TcTS and TrSA average structures of the simulations and dual conformations are observed in X-ray crystal structures of both TcTS and TrSA covalent intermediates, the $\chi_1$ change of Phe58 is also monitored. In fact, this phenylalanine which exists in both TcTS and TrSA X-ray crystal structures is not a natural residue for TcTS, which indeed has an asparagine residue instead. This Phe58 is one of the 7 mutations performed to obtain stable diffraction-quality crystals. All the other 6 mutations are in the lectin-like part which is not included in our simulations. The $\chi_1$ dihedral angles in Figure 5-23 show that Phe58 visits all three possible conformations in all simulations except two. In sialyllactose-bound TrSA simulation, Phe58 adopts one conformation ($\chi_1 = 180^\circ$), different from the initial one, and retains it for the rest of the simulation. And in sialyllactose-bound TcTS simulation, Phe58 very rarely samples the conformation that allows the stacking interaction and strongly prefers the other two possible conformations (Figure 5-23). Visualization of the trajectory file reveals that the conformation of Phe58 seen in sialyllactose-bound TrSA provides a stacking interaction with the lactose part of the sialyllactose. Phe58 conformation will be discussed in more detail later in this chapter.

5.3.7.4. Behavior of Asp59

$\chi_1$ dihedral angle change affects whether Asp59 carboxylic acid group swings towards glycosidic oxygen of lactose—the oxygen that accepts a H atom from Asp59. $\chi_1 = -70^\circ$ provides the proper orientation for Asp59 for its interaction with glycosidic O as well as having a hydrogen-bonding interaction with the guanidinium group of Arg53. Figure 5-24 shows that
Figure 5-21. Histograms of Tyr119 $\chi_1$-$\chi_2$ populations in MD simulations. Dihedral angles are in degrees.

Asp59 adopts the proper orientation for catalysis only in DANA-bound TrSA, sialyllactose-bound TrSA and TrSA$_{5\text{mut}}$ simulations. None of TcTS simulations samples this conformation significantly. Additionally, the RMSD change of the loop bearing Asp59 is investigated. The
Figure 5-22. Backbone RMSD for the loop bearing Tyr119/ Ser119 in MD simulations.
Figure 5-23. Phe58 $\chi_1$ dihedral angle change in MD simulations. Dihedral angles are in degrees.

loop architecture change is insignificant (RMSD < 1 Å, data not shown) in all simulations but there is noticeable overall loop motion in only TrSA$_{5\text{mut}}$ simulation in which the loop temporarily leaves its original position in the second 10 ns of the simulation (Figure 5-25).
5.3.7.5. Behavior of Tyr342

Dual conformations of Tyr342 which differ slightly both in $\chi_1$ and $\chi_2$ dihedral angles were observed in X-ray crystal structures of free TcTS and DANA-bound TcTS. Figure 5-26 and Figure 5-27 shows no sign of dual conformations of Tyr342 in our MD simulations and Tyr342 behaves very similar in all systems. The loop motion is also found to be insignificant since RMSD value of the loop backbone is lower than 1.25 Å in all systems throughout the simulation (data not shown). Since Ala341 that precedes Tyr342 in TcTS is replaced by a glycine in TrSA, we also investigated the effect of backbone $\phi$-$\psi$ angle changes. The $\psi$ angle of Ala341 is found to be -50° in all TcTS simulations while the homologous Gly341 shows a $\psi$ angle of -130° in all TrSA simulations including TrSA5mut (data not shown). The $\phi$ dihedral angle of Ala341/Gly341 is depicted in Figure 5-28 which shows the $\phi$ dihedral angle changes from mostly sampling -150° in free TcTS and TcTS covalent intermediate to mostly sampling -75° in DANA-bound TcTS and sialyllactose-bound TcTS. The same plot shows that the $\phi$ dihedral angle remains the same for all species of TrSA.

5.3.7.6. Behavior of Glu230

Glu230 helps the nucleophile, Tyr342, by capturing its H before attacking the ligand, thus, its conformation is important. But since Tyr342 conformation and position is found to stay intact in the simulations, the relative position of Glu230 is monitored by the hydrogen bond analysis results and no further analysis is done for Glu230.

5.3.7.7. Behavior of Asp96

Asp96 has dual conformations in DANA-bound TcTS X-ray crystal structure which differ in $\chi_1$ dihedral angle. Asp96 stays intact in all MD simulations around -70° (data not
shown). Asp96 lies on a β-strand whose RMSD is found to be lower than 3 Å in the simulations, pointing to a mostly stable backbone around this region.

5.3.7.8. Behavior of Leu36

Leu36 has dual conformations in free TcTS and DANA-bound TcTS X-ray crystal structures. Since Leu36 lies close to Tyr342, the conformation of Leu36 might be affecting Tyr342. Figure 5-29 and Figure 5-30 shows how Leu36 $\chi_1$ and $\chi_2$ dihedral angles change in the MD simulations. We observe multiple conformations of Leu36 in most of the MD simulations.

5.3.8. Ligand-Enzyme Interactions in Ligated TcTS and TrSA MD Simulations

In this section, the interactions between the ligand and the enzyme in different ligated forms of TcTS and TrSA will be compared. Table 5-2 and Table 5-3 gives detailed information about the stability of hydrogen-bonding interactions observed in the catalytic cleft in our MD simulations. Figure 5-31 and Figure 5-32 illustrate the sialyllactose-bound forms of TcTS and TrSA, respectively, as an example of the active site interactions of each enzyme. In all ligated TcTS, the carboxylate O atoms of sialic acid form salt bridges to Arg314, Arg245 and Arg35 (the arginine triad) guanidinium groups. These constitute the most stable interactions of all. Arg53 guanidinium group forms a hydrogen bond to the hydroxyl group (O4) of the sialic acid for all three cases, except being very scarcely observed in the covalent intermediate. Instead, this O4 atom interacts with the carboxylate O atoms of Asp96 in the covalent intermediate.

There are four hydrogen bond interactions seen only in sialyl-lactose-bound TcTS. Two of these are intramolecular interactions of sialyl-lactose—one between the sialic acid O7 atom and the galactose O2 atom and the other between the galactose O5 atom and the glucose O3 atom. It is also observed that Glu362 carboxylate O atoms form hydrogen bonds to O6 atom of the galactose part of the sialyl-lactose.
In all three ligated TcTS, Asp96 carboxylate O atoms are interacting with the N5 atom of the acetyl side chain of the sialic acid.
It is observed that the interactions of the glycerol side chain of the sialic acid differ for each ligated TcTS. The most prominent hydrogen bonds are between Glu230 carboxylate O atoms and O9 atom of the sialic acid in sialyllactose-bound TcTS and between Trp120’s side

Figure 5-25. RMSD (in Å) of the loop bearing Asp59 in MD simulations without fitting.
Figure 5-26. Histograms of $\chi_1-\chi_2$ populations of Tyr342 in TcTS MD simulations. Dihedral angles are in degrees.

chain N atom and O8 atom of the sialic acid, and Gln195’s sidechain O atom and O9 atom of the sialic acid in the TcTS covalent intermediate. In DANA-bound TcTS, Asp96 carboxylate atoms interact with both O8 and O9 atoms of the sialic acid as well as Trp120’s and Gln195’s side chain N atoms interacting with O9 atom.

In all ligated TrSA simulations, the carboxylate group of sialic acid interacts with Arg314, Arg35 and Arg245 (the arginine triad) guanidinium groups. Additionally, Arg53
guanidinium group and Asp96 carboxylate group form persistent H-bonds to the hydroxyl group (O4) of the sialic acid.

Figure 5-27. Histograms of $\chi_1-\chi_2$ populations of Tyr342 in TrSA MD simulations. Dihedral angles are in degrees.
Figure 5-28. $\Phi$ dihedral angle (in degrees) of Ala341/Gly341 in MD simulations.
Figure 5-29. Histograms of $\chi_1$-$\chi_2$ populations of Leu36 in TcTS MD simulations. Dihedral angles are in degrees.

The N-acetyl side chain of the sialic acid has its N atom (N5) interacting with the Asp96 carboxylate O atoms in all TrSA simulations. There is an additional interaction seen only in DANA-bound and SLT-bound TrSA simulations between Asp59’s carboxylate O atoms and O atom (O5N) of the acetyl side chain of the sialic acid. This H-bond forms due to a spatial change in Asp59 by changing the $\chi_1$ angle.
Figure 5-30. Histograms of $\chi_{1}-\chi_{2}$ populations of Ile36 in TrSA MD simulations. Dihedral angles are in degrees.

The glycerol side chain of the sialic acid has different interactions in each simulation of ligated TrSA. The most prominent hydrogen bonds lie from Glu230 carboxylate O atoms to O9 atom of the sialic acid in sialyllactose-bound TrSA, and from Trp120’s side chain N atom to O8
atom of the sialic acid and from Gln195’s sidechain O atom to O9 atom of the sialic acid in the
TrSA covalent intermediate. In DANA-bound TrSA, Arg245 amino N atoms interact with O9
atom of the sialic acid as well as Trp120’s side chain N atom interacting with O8 atom. All these
interactions in TrSA simulations are similar to the ones in TcTS simulations except that Asp96
H-bonds to O8 and O9 in DANA-bound TcTS are lost in DANA-bound TrSA. This is due to the
different conformations that glycerol side chain of the sialic acid adopts in the two average
structures.

Table 5-2. Hydrogen-bond interactions of the ligands in covalent intermediate, DANA-bound
and sialyllactose-bound TcTS.

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>TcTS-SLT</th>
<th>TcTS cov.int.</th>
<th>TcTS-DANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SA.O1A</td>
<td>Arg314.NH1</td>
<td>100 (2900 ps)</td>
<td>96 (260 ps)</td>
<td>92 (121 ps)</td>
</tr>
<tr>
<td>1SA.O1A</td>
<td>Arg314.NH2</td>
<td>9  (11 ps)</td>
<td>45 (22 ps)</td>
<td>48 (22 ps)</td>
</tr>
<tr>
<td>1SA.O1A</td>
<td>Arg245.NH1</td>
<td>90 (110 ps)</td>
<td>79 (56 ps)</td>
<td>22 (23 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg314.NH2</td>
<td>99 (1500 ps)</td>
<td>74 (48 ps)</td>
<td>91 (113 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg35.NH1</td>
<td>90 (107 ps)</td>
<td>93 (149 ps)</td>
<td>73 (40 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg35.NH2</td>
<td>65 (34 ps)</td>
<td>47 (22 ps)</td>
<td>79 (51 ps)</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.O4</td>
<td>42 (45 ps)</td>
<td>39 (37 ps)</td>
<td>-</td>
</tr>
<tr>
<td>Asp96.OD2</td>
<td>1SA.O4</td>
<td>19 (23 ps)</td>
<td>39 (37 ps)</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O4</td>
<td>Arg53.NH2</td>
<td>92 (169 ps)</td>
<td>12 (83 ps)</td>
<td>88 (98 ps)</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.N5</td>
<td>61 (31 ps)</td>
<td>42 (35 ps)</td>
<td>37 (18 ps)</td>
</tr>
<tr>
<td>Asp96.OD2</td>
<td>1SA.N5</td>
<td>38 (19 ps)</td>
<td>15 (77 ps)</td>
<td>89 (108 ps)</td>
</tr>
<tr>
<td>Tyr119.OH</td>
<td>1SA.O8</td>
<td>-</td>
<td>13 (17 ps)</td>
<td>-</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.O8</td>
<td>-</td>
<td>13 (594 ps)</td>
<td>86 (1070 ps)</td>
</tr>
<tr>
<td>1SA.O8</td>
<td>Trp120.NE1</td>
<td>-</td>
<td>45 (76 ps)</td>
<td>-</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.O9</td>
<td>-</td>
<td>-</td>
<td>75 (814 ps)</td>
</tr>
<tr>
<td>Glu230.OE2</td>
<td>1SA.O9</td>
<td>16 (24 ps)</td>
<td>-</td>
<td>9 (46 ps)</td>
</tr>
<tr>
<td>Glu230.OE1</td>
<td>1SA.O9</td>
<td>88 (230 ps)</td>
<td>-</td>
<td>14 (261 ps)</td>
</tr>
<tr>
<td>Gln195.OE1</td>
<td>1SA.O9</td>
<td>-</td>
<td>32 (71 ps)</td>
<td>12 (154 ps)</td>
</tr>
<tr>
<td>1SA.O9</td>
<td>Trp120.NE1</td>
<td>-</td>
<td>7 (21 ps)</td>
<td>71 (61 ps)</td>
</tr>
<tr>
<td>1SA.O9</td>
<td>Gln195.NE2</td>
<td>-</td>
<td>-</td>
<td>39 (29 ps)</td>
</tr>
<tr>
<td>1SA.O7</td>
<td>3LB.O2</td>
<td>78 (79 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3LB.O5</td>
<td>4GA.O3</td>
<td>70 (38 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu362.OE2</td>
<td>3LB.O6</td>
<td>48 (72 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu362.OE1</td>
<td>3LB.O6</td>
<td>19 (46 ps)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Percentage occupancies are reported for each interaction with average lifetimes in parentheses. All
ligand atom namings are according to sialyl-lactose-bound system. All ligand atoms are colored blue
and 4GA, 3LB, and 1SA are the residue numbers of glucose, galactose and sialic acid parts of the
sialyl-lactose, respectively.
Figure 5-31. Hydrogen bonding (green dashed lines) and steric (red dashed lines) interactions in sialyllactose-bound TcTS. The ligand sialyllactose is shadowed and its sialic acid (black), galactose (blue) and glucose (red) parts are colored separately.

There are five hydrogen bond interactions which are seen only in sialyl-lactose-bound TrSA. Three of these are intramolecular interactions of sialyl-lactose—one between the sialic acid O5 atom and the galactose O2 atom, another one between the sialic acid O7 atom and the
galactose O2 atom and the third one between the galactose O5 atom and the glucose O3 atom.

The two other H-bonding interactions are observed from the glucose O6 and sialic acid O7 atoms to Ser119 hydroxyl group. It is also observed that Asp362 carboxylate O atoms form hydrogen bonds to O6 atom of the galactose part of the sialyl-lactose and Gln122 side chain forms hydrogen bonds to O1 atom of the glucose part.

Table 5-3. H-bond interactions of the ligands in covalent intermediate, DANA-bound and sialyllactose-bound TrSA.

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>TrSA-SLT</th>
<th>TrSA cov. int.</th>
<th>TrSA-DANA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SA.O1A</td>
<td>Arg314.NH1</td>
<td>17 (13 ps)</td>
<td>44 (20 ps)</td>
<td>9 (12 ps)</td>
</tr>
<tr>
<td>1SA.O1A</td>
<td>Arg314.NH2</td>
<td>99 (1600 ps)</td>
<td>94 (198 ps)</td>
<td>97 (406 ps)</td>
</tr>
<tr>
<td>1SA.O1A</td>
<td>Arg245.NH1</td>
<td>97 (308 ps)</td>
<td>87 (82 ps)</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O1A</td>
<td>Arg245.NH2</td>
<td>-</td>
<td>-</td>
<td>23 (134 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg314.NH1</td>
<td>98 (522 ps)</td>
<td>84 (71 ps)</td>
<td>98 (474 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg314.NH2</td>
<td>-</td>
<td>-</td>
<td>16 (12 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg35.NH2</td>
<td>84 (67 ps)</td>
<td>92 (134 ps)</td>
<td>85 (68 ps)</td>
</tr>
<tr>
<td>1SA.O1B</td>
<td>Arg35.NH2</td>
<td>-</td>
<td>45 (25 ps)</td>
<td>79 (50 ps)</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.O4</td>
<td>99 (1600 ns)</td>
<td>40 (250 ps)</td>
<td>17 (104 ps)</td>
</tr>
<tr>
<td>Asp96.OD2</td>
<td>1SA.O4</td>
<td>-</td>
<td>-</td>
<td>31 (104 ps)</td>
</tr>
<tr>
<td>1SA.O4</td>
<td>Arg53.NH2</td>
<td>94 (196 ps)</td>
<td>16 (26 ps)</td>
<td>83 (69 ps)</td>
</tr>
<tr>
<td>1SA.O4</td>
<td>Arg53.NH1</td>
<td>-</td>
<td>26 (341 ps)</td>
<td>-</td>
</tr>
<tr>
<td>Asp96.OD2</td>
<td>1SA.N5</td>
<td>94 (168 ps)</td>
<td>46 (101 ps)</td>
<td>53 (32 ps)</td>
</tr>
<tr>
<td>Asp96.OD1</td>
<td>1SA.N5</td>
<td>-</td>
<td>10 (25 ps)</td>
<td>70 (53 ps)</td>
</tr>
<tr>
<td>1SA.O5N</td>
<td>Asp59.OD2</td>
<td>72 (227 ps)</td>
<td>-</td>
<td>10 (445 ps)</td>
</tr>
<tr>
<td>1SA.O8</td>
<td>Trp120.NE1</td>
<td>-</td>
<td>36 (60 ps)</td>
<td>13 (24 ps)</td>
</tr>
<tr>
<td>1SA.O9</td>
<td>Arg245.NH1</td>
<td>-</td>
<td>-</td>
<td>23 (55 ps)</td>
</tr>
<tr>
<td>Glu230.OE1</td>
<td>1SA.O9</td>
<td>92 (1000 ns)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O9</td>
<td>Trp120.NE1</td>
<td>-</td>
<td>11 (23 ps)</td>
<td>-</td>
</tr>
<tr>
<td>Gln195.OE1</td>
<td>1SA.O9</td>
<td>-</td>
<td>35 (116 ps)</td>
<td>-</td>
</tr>
<tr>
<td>3LB.O5</td>
<td>4GA.O3</td>
<td>43 (25 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O5</td>
<td>3LB.O2</td>
<td>10 (13 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O7</td>
<td>3LB.O2</td>
<td>78 (68 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4GA.O6</td>
<td>Ser119.OG</td>
<td>13 (46 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1SA.O7</td>
<td>Ser119.OG</td>
<td>10 (27 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp362.OD1</td>
<td>3LB.O6</td>
<td>12 (33 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp362.OD2</td>
<td>3LB.O6</td>
<td>10 (27 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4GA.O1</td>
<td>Gln122.NE2</td>
<td>24 (30 ps)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gln122.OE1</td>
<td>4GA.O1</td>
<td>9 (30 ps)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Percentage occupancies are reported for each interaction with average lifetimes in parentheses. All ligand atom namings are according to sialyl-lactose-bound system. All ligand atoms are colored blue and 4GA, 3LB, and 1SA are the residue numbers of glucose, galactose and sialic acid parts of the sialyl-lactose, respectively.
Figure 5-32. Hydrogen bonding (green dashed lines) and steric (red dashed lines) interactions in sialyllactose-bound TrSA. The ligand sialyllactose is shadowed and its sialic acid (black), galactose (blue) and glucose (red) parts are colored separately.

5.3.9. Interactions Observed Between the Enzyme Residues in the Active Site

The hydrogen bonding interactions among the residues that lie within 10 Å of Tyr342’s hydroxyl O atom in MD simulations are investigated to detect any differences between TcTS and TrSA active sites (Table 5-4 and 5-5). Persistent hydrogen bonds are observed in both enzymes
Table 5-4. Hydrogen bonding interactions between the residues in the active site in MD simulations of TcTS species.

<table>
<thead>
<tr>
<th>H-bonding atoms</th>
<th>TcTS</th>
<th>TcTS cov.int.</th>
<th>TcTS-DANA</th>
<th>TcTS-SLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp51.Arg53</td>
<td>OD2-NE: 52 (162)</td>
<td>77 (342)</td>
<td>75 (279)</td>
<td>21 (279)</td>
</tr>
<tr>
<td></td>
<td>OD1-NH2: 50 (62)</td>
<td>34 (87)</td>
<td>73 (249)</td>
<td>84 (83)</td>
</tr>
<tr>
<td></td>
<td>OD1-NE: 48 (133)</td>
<td>22 (76)</td>
<td>36 (45)</td>
<td>75 (143)</td>
</tr>
<tr>
<td></td>
<td>OD2-NH2: 49 (59)</td>
<td>51 (47)</td>
<td>27 (82)</td>
<td>-</td>
</tr>
<tr>
<td>Glu230.Arg245</td>
<td>OE1-NH1: 39 (80)</td>
<td>99 (1200)</td>
<td>75 (98)</td>
<td>78 (71)</td>
</tr>
<tr>
<td></td>
<td>OE2-NH1: 60 (121)</td>
<td>-</td>
<td>31 (32)</td>
<td>-</td>
</tr>
<tr>
<td>Glu357.Arg35</td>
<td>OE1-NE: 94 (183)</td>
<td>96 (285)</td>
<td>93 (154)</td>
<td>99 (694)</td>
</tr>
<tr>
<td></td>
<td>OE2-NH1: 96 (267)</td>
<td>70 (37)</td>
<td>80 (55)</td>
<td>66 (32)</td>
</tr>
<tr>
<td></td>
<td>OE1-NH1: -</td>
<td>-</td>
<td>16 (12)</td>
<td>17 (12)</td>
</tr>
<tr>
<td></td>
<td>OE2-NE: -</td>
<td>-</td>
<td>12 (12)</td>
<td>-</td>
</tr>
<tr>
<td>Asp247.Arg245</td>
<td>OD2-NH2: 53 (31)</td>
<td>47 (59)</td>
<td>56 (31)</td>
<td>61 (53)</td>
</tr>
<tr>
<td></td>
<td>OD1-NH2: 60 (35)</td>
<td>45 (57)</td>
<td>55 (32)</td>
<td>42 (34)</td>
</tr>
<tr>
<td></td>
<td>OD2-NE: 40 (53)</td>
<td>44 (41)</td>
<td>50 (52)</td>
<td>62 (69)</td>
</tr>
<tr>
<td></td>
<td>OD1-NE: 51 (58)</td>
<td>45 (38)</td>
<td>49 (49)</td>
<td>36 (44)</td>
</tr>
<tr>
<td>Glu230.Gln195</td>
<td>OE2-NH2: -</td>
<td>-</td>
<td>-</td>
<td>86 (513)</td>
</tr>
<tr>
<td>Gln174.Tyr113</td>
<td>OE1-OH: 80 (4990)</td>
<td>69 (2480)</td>
<td>96 (2650)</td>
<td>96 (463)</td>
</tr>
<tr>
<td>Glu362.Asn359</td>
<td>OE2-ND2: 30 (75)</td>
<td>26 (55)</td>
<td>49 (114)</td>
<td>25 (121)</td>
</tr>
<tr>
<td></td>
<td>OE1-ND2: 20 (60)</td>
<td>47 (83)</td>
<td>45 (104)</td>
<td>65 (143)</td>
</tr>
<tr>
<td>Glu362.Tyr364</td>
<td>OE1-OH: 42 (80)</td>
<td>17 (47)</td>
<td>48 (96)</td>
<td>24 (78)</td>
</tr>
<tr>
<td></td>
<td>OE2-OH: 34 (77)</td>
<td>36 (53)</td>
<td>40 (79)</td>
<td>65 (138)</td>
</tr>
<tr>
<td>Glu230.Tyr342</td>
<td>OE1-OH: 22 (43)</td>
<td>-</td>
<td>56 (74)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>OE2-OH: 17 (30)</td>
<td>99 (1650)</td>
<td>45 (69)</td>
<td>-</td>
</tr>
</tbody>
</table>

Occupancies are given as percentage and average lifetimes are given in picoseconds in parenthesis.

between Asp51 and Arg53, Glu230 and Arg245, Glu357 and Arg35, Asp247 and Arg245, Glu362/Asp362 and Tyr364 side chains. However, TcTS has two additional stable hydrogen bonding interactions between Gln174 and Tyr113, and between Glu362 and Asn359, which are not seen in TrSA due to differences in the corresponding amino acids (Gln174 in TcTS/Glu174 in TrSA, Glu362 in TcTS/Asp362 in TrSA). In DANA-bound TrSA MD simulation, a unique hydrogen bond between Asp59 and Arg53 is seen.

The hydrogen bonding interaction between Tyr342 and Glu230 is also monitored since these two residues act as a nucleophile couple in the catalytic reactions. In our MD simulations, we found stable hydrogen bonding between these two residues only in the covalent intermediate forms of the two enzymes and DANA-bound TcTS. A less stable interaction is observed for
unligated TcTS and TrSA. Interestingly, no such interaction is observed in TrSA<sub>5mut</sub>, either of the sialyllactose-bound enzymes or DANA-bound TrSA MD simulations.

Table 5-5. Hydrogen bonding interactions between the residues that lie within 10 Å of Tyr342.OH atom in MD simulations of free and ligated TrSA species.

<table>
<thead>
<tr>
<th>H-bonding atoms</th>
<th>TrSA</th>
<th>TrSA cov.int.</th>
<th>TrSA-DANA</th>
<th>TrSA-SLT</th>
<th>TrSA&lt;sub&gt;5mut&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp51. OD2-NE</td>
<td>11 (239)</td>
<td>64 (502)</td>
<td>-</td>
<td>90 (108)</td>
<td>-</td>
</tr>
<tr>
<td>Arg53 OD1-NH2</td>
<td>76 (69)</td>
<td>65 (109)</td>
<td>77 (65)</td>
<td>16 (140)</td>
<td>80 (68)</td>
</tr>
<tr>
<td>OD1-NE</td>
<td>86 (269)</td>
<td>34 (151)</td>
<td>86 (150)</td>
<td>-</td>
<td>96 (272)</td>
</tr>
<tr>
<td>OD2-NH2</td>
<td>-</td>
<td>28 (33)</td>
<td>-</td>
<td>78 (99)</td>
<td>-</td>
</tr>
<tr>
<td>Glu230. OE1-NH1</td>
<td>39 (457)</td>
<td>99 (1600)</td>
<td>11 (64)</td>
<td>48 (27)</td>
<td>49 (203)</td>
</tr>
<tr>
<td>Arg245 OE2-NH2</td>
<td>19 (34)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glu357. OE1-NE</td>
<td>99 (747)</td>
<td>99 (1300)</td>
<td>99 (823)</td>
<td>99 (1380)</td>
<td>94 (202)</td>
</tr>
<tr>
<td>Arg35 OE2-NH1</td>
<td>95 (188)</td>
<td>78 (49)</td>
<td>79 (51)</td>
<td>63 (28)</td>
<td>94 (158)</td>
</tr>
<tr>
<td>OE1-NH1</td>
<td>10 (11)</td>
<td>18 (13)</td>
<td>14 (12)</td>
<td>15 (12)</td>
<td>-</td>
</tr>
<tr>
<td>OE2-NE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp247. OD2-NH2</td>
<td>18 (98)</td>
<td>48 (241)</td>
<td>26 (47)</td>
<td>38 (90)</td>
<td>56 (65)</td>
</tr>
<tr>
<td>Arg245 OD1-NH2</td>
<td>-</td>
<td>52 (211)</td>
<td>22 (37)</td>
<td>64 (159)</td>
<td>44 (54)</td>
</tr>
<tr>
<td>OD2-NE</td>
<td>14 (43)</td>
<td>45 (25)</td>
<td>18 (31)</td>
<td>50 (29)</td>
<td>33 (22)</td>
</tr>
<tr>
<td>OD1-NE</td>
<td>-</td>
<td>61 (35)</td>
<td>14 (28)</td>
<td>45 (26)</td>
<td>28 (22)</td>
</tr>
<tr>
<td>Glu230. OE2-NE2</td>
<td>40 (246)</td>
<td>-</td>
<td>10 (451)</td>
<td>99 (1380)</td>
<td>48 (914)</td>
</tr>
<tr>
<td>Gin195 OE1-NE2</td>
<td>34 (337)</td>
<td>-</td>
<td>89 (1200)</td>
<td>-</td>
<td>32 (379)</td>
</tr>
<tr>
<td>Asp362. OD1-OH</td>
<td>92 (192)</td>
<td>49 (82)</td>
<td>48 (161)</td>
<td>44 (167)</td>
<td>26 (174)</td>
</tr>
<tr>
<td>Tyr364 OD2-OH</td>
<td>-</td>
<td>22 (72)</td>
<td>43 (158)</td>
<td>42 (139)</td>
<td>66 (204)</td>
</tr>
<tr>
<td>Glu230. OE1-OH</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyr342 OE2-OH</td>
<td>11 (55)</td>
<td>99 (1100)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp247. OD1-NE2</td>
<td>13 (36)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gin283 OD2-NE2</td>
<td>13 (35)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Asp59. OD1-NH1</td>
<td>-</td>
<td>-</td>
<td>80 (124)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Occupancies are given as percentage and average lifetimes are given in picoseconds in parenthesis.

5.4. Discussion

5.4.1. Sialyllactose-Bound Form of TrSA

Trp312 forms one of the lateral walls of the catalytic cleft in TcTS. Experiments proved that Trp312 is crucial for trans-sialidase activity of TcTS. The mutation Trp312-Ala has decreased sialidase activity of TcTS from 306.0 nmol min<sup>-1</sup> mg<sup>-1</sup> to 185.1 nmol min<sup>-1</sup> mg<sup>-1</sup> while totally abolishing its trans-sialidase activity which was 1412.98 nmol min<sup>-1</sup> mg<sup>-1</sup>. Buschiazzo et
also reveals that Pro283-Gln and Tyr248-Gly mutations can abolish both sialidase and trans-sialidase activities of TcTS. These mutations are effective on the conformation of Trp312. Proline 283 in TcTS, which is replaced by a glutamine in TrSA, lies right next to Trp312 in TcTS and backs the tryptophan that adopts a conformation forming a binding site for the sugar donor/acceptor molecules. In the X-ray crystal structure, the glutamine in TrSA—the one that corresponds to Pro283 in TcTS—interferes the space that Trp312 occupies in TcTS, seemingly preventing Trp312 to occupy the same space in TrSA. There is strong experimental evidence pointing to the importance of Pro283 for trans-sialidase activity:

(i) Pro283-Gln mutation abolishing trans-sialidase activity of TcTS

(ii) Chimeric protein that has the first 200 residues from TrSA’s C-terminus and the rest from TcTS’s N-terminus shows strict sialidase activity while a single additional Glutamine to Proline (Pro283) mutation can promote both sialidase and trans-sialidase activities.

In all available TrSA X-ray crystal structures, the conformation of Trp312 is different from the one seen in TcTS (Figure 5-33). The two conformations differ both in $\chi_1$ and $\chi_2$ dihedral angles and the conformation in TcTS puts the indole ring of Trp312 in a position to back the lactose part of the sialyllactose ligand. There is no X-ray crystal structure of sialyllactose-bound TrSA available to show whether the binding mode of Trp312 to the donor/acceptor lactose molecule is the same as in TcTS or not. Thus, we modeled sialyllactose-bound TrSA by manually placing sialyllactose ligand into the active site of TrSA using DANA-bound TrSA as the starting point and mimicking the sialyllactose pose in TcTS. The conformational difference between available TcTS and TrSA X-ray crystal structures in Trp312 conformation as well as the proven importance of Trp312 and Pro283 for sialyl-transfer catalysis draws attention to Trp312.

MD simulations of sialyllactose-bound TrSA model indicated that the loop bearing Trp312 moves away from the sialyllactose resulting in a slightly more open active site during relaxation
Figure 5-33. Different conformations of Trp312 seen in TcTS and TrSA. Sialyllactose is shown in orange. The conformation of Trp312 that backs the lactose part of the sialyllactose is seen in all TcTS while Trp312 lies away from the ligand in the other conformation in all available TrSA X-ray crystal structures.

of the system. However, in the production run, Trp312 slowly changes its conformation to the proper conformation ($\chi_1$ and $\chi_2$) observed in sialyllactose-bound TcTS in about 17 nanoseconds and subsequently re-approaches to the sialyllactose arriving at the exact same binding mode seen in TcTS at the end of 23 ns and conserves this conformation for the rest of the simulation (Figure 5-34). For comparison, the situation in sialyllactose-bound TcTS MD simulation is also shown. The binding mode of Trp312 and the overall loop conformation is conserved throughout the simulation indicating its stability in sialyllactose-bound TcTS.

Similar results can also be deduced from the Trp312 $\chi_1$ and $\chi_2$ histograms in Figure 5-17 and Figure 5-18 which show that Trp312 in TrSA also visits the dominant conformation seen in TcTS cases, even preferring it over the other conformations in some cases. The reason why this conformation is not observed in available X-ray crystal structures of TrSA is in fact found to be related to close crystal contacts. When the unit cell is formed using proper symmetry operations with SwissPdbViewer, the sidechain of Lys495 in the neighboring enzyme swings into the active
site of the enzyme, sterically hindering Trp312 to move freely and adopt the conformation seen in TcTS (Figure 5-35). The conformation seen in TrSA X-ray crystal structures causes hydrogen

Figure 5-34. Behavior of Trp312 and its loop backbone in MD simulations of A) modeled sialyllactose-bound TrSA and B) sialyllactose-bound TcTS. The loop of Trp312 approaches to the sialyllactose ligand and Trp312 adopts the binding mode conformation during simulation in TrSA while the loop of Trp312 and binding mode conformation remains stable during the whole 50-ns simulation.
bonding interactions between Trp312’s indole N atom and Asp362’s carboxylate O atoms and between Lys495’s side chain N atom and Tyr364’s phenol O atom.

All the above results in addition to the MD simulation of TrSA with manually placed sialyllactose point to a unique Trp312 conformation both in TcTS and TrSA forming the binding site for the lactose moiety. This unique conformation is the Trp312 conformation seen in all free and ligated TcTS X-ray structures.

Tyr119 constitutes the second lateral wall of the catalytic cleft facing the lateral wall constituted by Trp312 and is also found to be very significant for the trans-sialidase catalytic ability of TcTS. The mutation Tyr119Ser causes the trans-sialidase activity to decrease to 2.5% of that of the wild type TcTS while the sialidase activity only decreases to 77.8% of that of the wild type TcTS (Table 4 in Paris et al\(^8\)). In place of Tyr119 in TcTS, a serine exists in TrSA. The mutation of this serine in TrSA to a tyrosine is unable to promote any trans-sialidase activity as well as decreasing the sialidase activity to 50% compared to the wild type (Table 3 in Paris et al\(^8\)). These data were initially interpreted that two different binding sites exist for the acceptor

![Figure 5-35. Close crystal contacts in the unit cell of unligated TrSA. Orange and blue ribbons show the two neighboring enzymes. Lys495 side chain lies inside the active site of TrSA hindering Trp312 to move freely. Ser119 is also shown to clarify the relative orientation.](image)
and the donor lactose molecules. However, the X-ray crystal structures elucidated later showed that the active sites of TcTS and TrSA were not large enough to accommodate the donor and acceptor at the same time.

While comparing the average structures of sialyllactose-bound TcTS and TrSA produced by MD simulations, we noticed a structural difference that can shed light on this issue. Phe58 in TcTS, which is also present in TrSA, adopts different conformations in the two enzymes. In TcTS, Phe58 leans back ($\chi_1=60^\circ$ or $-60^\circ$) and the lactose part of the ligand is embraced from the two sides by Trp312 and Tyr119. However in TrSA, Phe58 adopts to a conformation ($\chi_1=180^\circ$)—which is different from the starting conformation—such that it embraces the lactose part of the ligand from the opposite side of Trp312 (Figure 5-36). Thus, the absence of Tyr119 in TrSA is covered by the presence of Phe58, complementing the binding site formed by Trp312 on one side. Still, the data imply that the lactose should be delicately oriented for trans-sialidase catalysis to happen. A shift in orientation of the lactose caused by interacting with Phe58 instead of Tyr119 can significantly decrease the trans-sialidase activity.

Figure 5-23 shows that in sialyllactose-bound TrSA model, Phe58 $\chi_1$ dihedral angle adopts a binding mode different than the initial conformation which was the same as the conformation seen in sialyllactose-bound TcTS and keeps this conformation for the rest of the simulation. This binding conformation is depicted in the average structure in Figure 5-36 for sialyllactose-bound TrSA.

In fact, this phenylalanine which exists in both TcTS and TrSA X-ray crystal structures is not a natural residue for TcTS which indeed has an asparagine residue instead. This Phe58 is one of the 7 mutations performed to obtain stable diffraction-quality crystals (All the other 6 mutations are in the lectin-like part.). Thus, the presence of Phe58 might be playing a role in
sialidase action in TrSA, helping it to bind the lactose part of the ligand—taking the role of Tyr119 in TcTS. Table 5-6 gives the sequence information for sialidases and trans-sialidases from *Trypanosoma cruzi*, *brucei* and *rangeli*. We see a phenyl—either right before or right after—around Asp59 homologs in each sialidase while the sequence differs in trans-sialidases. Thus, presence of the Phe58 homologs next to the acid/base catalyst aspartic acid residue could be a signature for trypanosomal sialidases.

![Figure 5-36. Binding mode of sialyllactose in A) TcTS and B) TrSA. Trp312 and Tyr119 are embracing lactose part of the ligand in TcTS. Trp312 and Phe58 are embracing lactose part of the ligand in TrSA which has a serine in place of Tyr119.](image)

Table 5-6. Sequence information for different sialidases and trans-sialidases of *Trypanosomal* species.

<table>
<thead>
<tr>
<th>Source</th>
<th>Enzyme</th>
<th>PdbID/SequenceID</th>
<th>Sequence around Asp59 analogs</th>
<th>Acid/base catalyst</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. cruzi</em></td>
<td>trans-sialidase</td>
<td>1MS3</td>
<td>NDN</td>
<td>Asp59</td>
</tr>
<tr>
<td><em>T. rangeli</em></td>
<td>sialidase</td>
<td>1N1S</td>
<td>FDN</td>
<td>Asp63</td>
</tr>
<tr>
<td><em>T. brucei</em></td>
<td>sialidase</td>
<td>Q57YT6-1</td>
<td>EDF</td>
<td>Asp108</td>
</tr>
<tr>
<td><em>T. brucei</em></td>
<td>trans-sialidase</td>
<td>Q57XJ2</td>
<td>TDY</td>
<td>Asp157</td>
</tr>
<tr>
<td><em>T. brucei</em></td>
<td>trans-sialidase</td>
<td>Q9GSF0</td>
<td>TDY</td>
<td>Asp157</td>
</tr>
</tbody>
</table>

The average sialyllactose pose is found to be very similar in sialyllactose-bound TcTS and TrSA MD simulations. However, in the active sites, the conformation of Asp59 differs for the two enzymes; Asp59 adopts a proper conformation for catalysis in sialyllactose-bound TrSA in contrast to sialyllactose-bound TcTS. There is also a slight difference in the loop bearing Tyr342
which might be related to the presence of Ala341 in TcTS instead of Gly341 in TrSA. There are also several mutations that are constraining the active sites differently:

(i) Val95 in TcTS is Met95 in TrSA which results in a more restricted active site in TrSA around 1SA.O4.
(ii) Leu176 in TcTS is Val176 in TrSA which results in a more restricted active site in TcTS around N-acetyl group of sialic acid.

Comparing the ligand-enzyme interactions, one notices the stable hydrogen-bonding of 1SA.O4 with Asp96 carboxylate O atoms in TrSA is much less in TcTS and the one between 1SA.O5N and Asp59 carboxylic acid O atoms in TrSA is absent in TcTS due to Asp59 conformational change (Table 5-2, Table 5-3). Additionally, Gln122 in TrSA which corresponds to a serine residue in TcTS interacts with 4GA.O1 atom of the sialyllactose.

The RMSF analysis results depicted in Figure 5-10 show no sign of a big difference in the active sites except for the loop bearing Trp312—which is a sign of this loop rearranging for proper ligand binding in TrSA—and Ala341 which lies right next to catalytic Tyr342. The absence of Tyr342 conformational change as seen in the histograms in Figure 5-26 and Figure 5-27 and the higher conformational flexibility seen in $\phi$ dihedral angle of Ala341 in TcTS compared to Gly341 in TrSA (Figure 5-28) may be a sign of the effect of Ala341 flexibility on catalysis in TcTS.

In summary, the sialyllactose binding to TrSA is found to be very similar to the case in TcTS except for that Phe58 and Gln122 residues help binding on one side of the catalytic cleft in TrSA instead of Tyr119 residue in TcTS.

**5.4.2. Effects of the 5 Point Mutations on TrSA that Promote trans-Sialidase Catalysis**

Since only 5 point mutations (Met95Val, Ala97Pro, Ser119Tyr, Gly248Tyr, Gln283Pro) were enough to modify the strict sialidase TrSA into a trans-sialidase, the effects of these 5 point mutations will be discussed here.
Due to lower resolution, the B-factors of TrSA5mut X-ray crystal structure are found to be incomparable with the wild-type TcTS and TrSA as depicted in Figure 5-15. The RMSF results of MD simulations are thus used to find the differences in flexibilities of the residues of each system. Figure 5-9 showed that the flexibilities in the loops of Ser119, Gly248 and Gln283 in TrSA have decreased due to the mutations, approaching to the flexibilities of homologous loops in TcTS. It is also important that in TcTS, the flexibilities of Tyr248 and Pro283 are still lower compared to TrSA despite the loop opening observed in Trp312 loop of TcTS—which gives these two neighboring loops more degrees of freedom. These 3 loops are important in forming the binding site for the lactose part of sialyllactose ligand in TcTS; Tyr119 forms one lateral wall of the binding site for lactose and Trp312 forms the second lateral wall which is supported firmly by Tyr248 and Pro283 by hydrophobic interactions. RMSF results confirm that the mutations of Gly248Tyr and Gln283Pro in TrSA provided a stable support for Trp312.

However, we observed that Trp312 did not sample the binding conformation (Figure 5-17) in the time scale of our TrSA5mut MD simulation. The initial conformation of Trp312 in TrSA5mut is different from the binding conformation due to crystallographic constraints as explained for TrSA above and affects our ability to sample the binding conformation. However, combining all information obtained from MD simulations of TcTS and TrSA, the loop of Trp312 is able to form a more open active site especially in the absence of the stacking interaction with Tyr119—due to Tyr 119 side chain swinging down in TcTS or a serine replacing Tyr119 in TrSA—and once the loop opens, Trp312 can sample all possible conformations. Especially the free enzymes—wild-type TcTS, wild-type TrSA and TrSA5mut—are observed to move more freely. Introduction of sialyllactose limits this degree of freedom and locks Trp312 into proper conformation as we have seen in MD simulation of modeled sialyllactose-bound TrSA. Figure 5-
17 also indicates the stability of binding conformation of Trp312 in TcTS once the sialic acid binding region is filled while no similar behavior is seen in TrSA. Whether the binding conformation of Trp312 in TrSA is not stable due to lack of hydrophobic interactions or the simulation was not long enough to sample all conformational space is not certain. Either the energetics of this conformational change should be elucidated or more MD simulations with different initial conformations of Trp312 should be performed to give a certain answer to this question.

The other two mutations Met95Val and Ala97Pro did not change the flexibility of the β-strand bearing these or the conformation of Asp96 to a noticeable degree. However, the replacement of methionine with a less bulky residue of valine decreases the constraints around O4 of sialic acid and affects the hydrogen bonding network in that region.

5.4.3. Comparison of DANA-Bound Forms of TcTS and TrSA

DANA, which is a very effective inhibitor for sialidases and specifically for TrSA, was unsuccessful in inhibiting TcTS. We investigated the structure and dynamics of the enzymes to find the reason of this inhibition difference.

The main difference between DANA-bound forms of TcTS and TrSA is that the glycerol branch of DANA adopts different conformations in the two enzymes (Figure 5-37) which is also seen between the X-ray crystal structures. In DANA-bound TcTS the glycerol branch of DANA leans downwards interacting with both Glu230 and Gln195 and forcing Glu230 to move towards Tyr342 and make a hydrogen-bonding interaction that is very stable throughout the simulation. However, in DANA-bound TrSA, the glycerol branch lies straight leaving the lower cavity for Glu230 which is now leaned away from Tyr342 and bended towards Gln195 having a very stable hydrogen-bonding interaction with it. Due to this conformation difference, the hydrogen bonding interactions of the glycerol branch with Trp120 indole ring and Asp96 carboxylate group seen in
DANA-bound TcTS are also lost in DANA-bound TrSA. No such difference in Glu230 conformations is observed in the X-ray crystal structures.

One could argue that different conformations of the glycerol branch of DANA seen in the average structures might be due to the different conformations of them in the X-ray crystal structures that are used as starting structures since these conformations are preserved throughout the simulation mainly. Whether there are at least two conformational minima for the glycerol branch of DANA in both TcTS and TrSA—that we could sample only one in each of our simulations—or there is really one single minimum for each enzyme which constitutes a real difference in DANA binding of TcTS and TrSA is a question. The simulations of TcTS covalent intermediate and sialyllactose-bound TcTS show that the glycerol branch of these ligands adopt the conformation seen in DANA-bound TrSA and do not even sample the conformation seen in DANA-bound TcTS. Since all these ligands occupy the same space in TcTS in the sialic acid binding site, it is more likely that there are at least two different stable conformations for the

![Figure 5-37. A view from superimposed DANA-bound forms of TrSA (orange) and TcTS (gray) average structures of the MD simulations. DANA is shown in licorice form and Tyr342, Glu230 and Gln195 are shown in CPK form. The conformation of glycerol side chain of DANA differs in the two cases also affecting Glu230 conformation.](image)
glycerol branch of DANA (and sialic acid) in TcTS. In fact, dual conformations of the glycerol side chain are observed in the X-ray crystal structure of DANA-bound TcTS. Thus, the conformation of this glycerol branch needs further investigation.

The Asp59 conformations are also different in the average structures of DANA-bound TcTS and TrSA (Figure 5-38) although there is no such difference in the original X-ray crystal structures of these two enzymes. The conformation in the crystal structures is the same as the one seen in the average structure of DANA-bound TrSA in which Asp59 leans towards DANA—as it would do in the presence of the glycosidic O of the sialyllactose that accepts a proton from Asp59 in the proposed mechanism. Investigating the crystal contacts of DANA-bound TcTS and DANA-bound TrSA reveals that there are two different crystallographic constraints forcing the same conformation of Asp59:

(i) The unit cell obtained from X-ray crystal structure of DANA-bound TrSA shows that Asp59 is interacting with the Lys495 of the neighboring enzyme (shown in Figure 5-35) through a water-mediated hydrogen bond.
The X-ray crystal structure of DANA-bound TcTS includes a solvent glycerol molecule that locates itself into the active site occupying the same position that would be occupied by galactose part of the lactose molecule. The hydrogen bonding interaction between one hydroxyl group of this glycerol molecule and Asp59 promotes a certain Asp59 conformation.

Due to these two constraints found, the Asp59 conformations in X-ray crystal structures are not reliable or natural. When the $\chi_1$ dihedral angle of Asp59—which is the source of the conformation difference—is followed over the simulation time, each of the two DANA-bound enzymes is found to adopt mostly one single conformation that is different for each case (Figure 5-24). Different Asp59 conformations can be related to the neighboring bulky residue Met95 in TrSA occupying more space compared to the corresponding residue of Val95 in TcTS. The Asp59 conformation difference is also reflected on that hydrogen bonding between Asp59 and Arg53 is observed 80% of the simulation in DANA-bound TrSA compared to 3% in DANA-bound TcTS.

Another difference between DANA-bound forms of TcTS and TrSA are in hydrogen bonding interactions of Asp96. Asp96 interacts with O4 of DANA in DANA-bound TcTS MD simulation, but no such interaction is observed in DANA-bound TrSA simulation, in which Asp96 interacts with the last hydroxyl group on the glycerol branch of DANA instead. Since the position and conformation of Asp96 is found to stay intact in MD simulations, a slight conformational change of DANA must exist between the two enzymes. In DANA-bound TrSA, DANA is tilted slightly down on the side of O4 hydroxyl group to interact with Asp96 while in DANA-bound TcTS, the glycerol branch of DANA bends downwards to interact with Asp96.

Trp312 did not sample the binding conformation in DANA-bound TrSA MD simulation as seen in the histograms of Figure 5-17 in contrast to DANA-bound TcTS (Figure 5-18). Since Trp312 is found to retain its initial conformation mainly, it is hard to comment on the possible
effect of DANA binding on Trp312 conformation. RMSF analysis shows that the loops of Gly248/Tyr248 and Trp312 are more flexible in DANA-bound TrSA compared to DANA-bound TcTS (Figure 5-7 and 5-8). As mentioned before, the presence of Tyr248 and Pro283 promotes the binding mode of Trp312 and decreases the flexibilities of these loops in TcTS compared to TrSA.

The slight difference in the positioning of DANA and the mentioned differences in several hydrogen bonding interactions between DANA-bound forms of TcTS and TrSA might be responsible for the significant difference in their Kᵢ values. More simulations are necessary to confirm these observations. The dual conformations of the glycerol branch of DANA in DANA-bound TcTS crystal structure also require more attention and follow up. A new simulation of DANA-bound TcTS starting with the alternative conformation of the glycerol branch of DANA can shed some light.

5.4.4. Comparison of Unligated and DANA-Bound Forms of TcTS

Surface plasmon resonance studies\(^7\) mentioned in the Introduction clearly showed that lactose could not bind TcTS in the absence of sialic acid or sialyllactose in the medium and there is no known reason for this observation. Several different studies also showed that TcTS can produce DANA as a by-product.\(^4\),\(^9\),\(^7\) Combining these data, we can hypothesize that TcTS might be producing DANA using sialic acid or sialyllactose in the medium which subsequently causes some modifications in TcTS to promote lactose binding.

In our simulations, we observed that DANA binding causes a decrease in the mobilities of the three loops bearing Tyr248, Pro283 and Trp312 (Figure 5-7 and 5-8). The lower mobilities of these three loops will mean a more rigid aglycon binding site readily available for lactose binding which can explain the observed change of lactose binding ability of TcTS in the presence of sialic acid or sialyllactose in the medium.
5.4.5. Comparison of the Effect of Covalent Intermediate Formation in TcTS and TrSA

We have noticed a difference between TcTS and TrSA in the effect of covalent intermediate formation. As shown in Figure 5-19, the loop bearing Trp312, which resides at the periphery of the catalytic cleft, is observed to move outwards and form a more open active site in unligated TcTS and covalent intermediate of TrSA. However, no such motion is observed in unligated TrSA and covalent intermediate of TcTS.

It is known that TcTS can catalyze both hydrolysis and sialyl-transfer reactions, while TrSA can only catalyze hydrolysis. The ability of TcTS to protect the covalent intermediate from the attack of surrounding water molecules until an acceptor molecule binds and completes the transfer reaction is of curiosity. Since in TrSA, a very similar active site the covalent intermediate readily reacts with the water molecules, it is possible that solvent exposure of the catalytic active sites have a role.

The MD simulation results in Figure 5-19 might be pointing to the change of solvent exposure of the active sites differently in the two enzymes in the course of the reaction. TcTS active site is solvent-exposed in the unligated form but once the covalent intermediate forms, solvent exposure is limited by the closure of the loop bearing Trp312. This will help TcTS to protect the covalent intermediate from the intensive attack of water molecules. However, in TrSA, the active site becomes more solvent-exposed after the covalent intermediate formation which is in line with its very high hydrolysis efficiency and no sialyl-transfer ability.

Although there might be other structural, energetical or dynamical reasons, or a combination of those for the observed catalytical difference of TcTS and TrSA, water accessibility also stands as a good candidate for such an outcome and requires further investigation. More simulations, with different starting conformations of Trp312 to eliminate its
possible effect in the observed loop opening difference, should be performed to test this hypothesis.
CHAPTER 6
CONCLUSIONS AND PERSPECTIVE

Trans-sialidase (TcTS) facilitates Trypanosoma cruzi, the parasite responsible for Chagas’ disease, to evade from the host immune response and to invade the host cells and thus, is identified as an appealing therapeutic target for uncurable Chagas’ disease—which is a drastic threat to large human populations in Central and South America. TcTS catalyzes the transfer of sialic acids—which is used by the host immune system to distinguish host cells from others and can not be synthesized de novo by T. cruzi—from host glycoconjugates to the parasite’s glycoconjugates.

The mechanism of TcTS has been the subject of various research efforts which aim to design potential inhibitors. Trypanosoma rangeli sialidase (TrSA), which acts as a strict sialidase despite its distinct structural similarity to TcTS, pointed out that minute structural modifications are responsible for trans-sialidase catalytic ability. Thus, TrSA is used as a case study together with TcTS to elucidate the requirements for achieving trans-sialidase catalysis. Continuous efforts recently succeeded in transforming TrSA into a trans-sialidase with only 5 point mutations at the active site.

The results from kinetic isotope effect studies together with chemical trapping experiments for TcTS anticipated that an associative mechanism is active in the first step of TcTS catalytic reaction—which is scavenging the sialic acid from sialyllactose—with significant nucleophilic participation which results in collapsing into a covalent intermediate. The ping-pong mechanism implicated by these results was indeed ruled out mistakenly due to earlier initial velocity studies. X-ray crystal structures of different forms of TcTS obtained later also started a debate about the identity of residues that act as nucleophile and acid/base catalyst since the candidates for these roles according to their location were very unusual.
This study started while the mechanism and the identity of the nucleophile and the acid/base catalyst were still being debated. We approached the open-ended questions in this topic with a computational chemist’s point of view. More experimental results became available in the meanwhile shedding light on our path. Evidence for covalent intermediate formation in TrSA using an activated ligand—which might possibly be biased by the artificial substituents of the ligand—directed us to also investigate the possibility of covalent intermediate formation in TrSA.

Our QM/MM studies with two separate methods have confirmed energetic plausibility of covalent intermediate formation in both TcTS and TrSA with natural ligands using Tyr342/Glu230 couple as the nucleophile and Asp59 as the acid/base catalyst. Potential energy surfaces constructed for TcTS are in line with the experimental data showing that TcTS acts as a better trans-sialidase than a sialidase in the presence of acceptor glycoconjugates. Our calculations have identified the oxocarbenium ion form of sialic acid to be an intermediate rather than a transition structure. However, the energy barrier calculated for the oxocarbenium ion to convert into the covalent intermediate is low. Scavenging the sialic acid from sialyllactose (i.e. the formation of oxocarbenium ion) is determined to be the rate-determining step that has a late oxocarbenium-like transition structure. Due to the relative timing of bond cleavages/ formations and the formation of a stable ionic intermediate on the reaction path, we can describe the TcTS mechanism as $S_N1$-like although it includes significant nucleophilic participation and is more complex than a simple $S_N1$ mechanism.

Molecular dynamics simulations we performed, on the other hand, shed some light on the path to be followed to elucidate the inhibitor binding differences between TcTS and TrSA. The glycerol side chain of sialic acid is found to be a good candidate to adopt different conformations
in the two enzymes, thereby causing different inhibitor binding abilities. The simulations also showed that the conformation of Trp312 in the aglycon binding sites were biased by crystal contacts in the X-ray crystal structures of TrSA, and Trp312 indeed adopts the same conformation in TrSA as in TcTS. However, still a difference exists between TcTS and TrSA in that Trp312 is found to be more flexible in TrSA, most probably related to the replacement of Tyr119 in TcTS by a serine residue. Additionally, the flexible loop bearing Trp312 which is found in a specific conformation in all X-ray crystal structures proved to be artificial and the movement of this loop has a possible role in the binding affinity of the enzyme to its ligands. This finding is also an indication of the fact that considering the X-ray crystal structures for direct use in ligand docking studies can mislead the researchers, especially in the case of enzymes such as TcTS and TrSA for which the ligand binding site mostly consists of flexible loops.
LIST OF REFERENCES


34. Previato, J.; Andrade, A. F. B.; Pessolani, M. C. V.; Mendonca-Previato, L. Molecular and Biochemical Parasitology 1985, 16(1), 85-96.


57. Previato, J. O. Glycobiology 2004, 14(10), 25G-a-.

58. de Lederkremer, R. M.; Frasch, A. C. C. Glycobiology 2004, 14(10), 26G-.


D. O.; Seyler, A.; Sharma, R.; Shetty, J.; Simpson, A. J.; Sisk, E.; Tammi, M. T.;
Tarleton, R.; Teixeira, S.; Van Aken, S.; Vogt, C.; Ward, P. N.; Wickstead, B.; Wortman,
J.; White, O.; Fraser, C. M.; Stuart, K. D.; Andersson, B. Science (Washington, DC,
United States) 2005, 309(5733), 409-415.

9(5), 915-921.


64. Guhl, F.; Hudson, L.; Marinkelle, C. J.; Jaramillo, C. A.; Bridge, D. Parasitology 1987,
94, 475-484.

339-348.


67. Buschiazzo, A.; Cremona, M. L.; Campetella, O.; Frasch, A. C. C.; Sanchez, D. O.
Molecular and Biochemical Parasitology 1993, 62(1), 115-116.

Whittington, A. R.; Scicinski, J.; Bethell, R. C.; Taylor, N. Bioorganic & Medicinal
Chemistry Letters 1996, 6(24), 2931-2936.


70. Rost, B. Protein Engineering 1999, 12(2), 85-94.

71. Pontesdecarvalho, L. C.; Tomlinson, S.; Nussenzweig, V. Molecular and Biochemical

72. Paris, G.; Ratier, L.; Amaya, M. F.; Nguyen, T.; Alzari, P. M.; Frasch, A. C. C. Journal of
Molecular Biology 2005, 345(4), 923-934.

G. Chemical Communications 2000(12), 1013-1014.

74. Watt, G. M.; Lowden, P. A. S.; Flitsch, S. L. Current Opinion in Structural Biology 1997,
7(5), 652-660.

75. Buschiazzo, A.; Amaya, M. F.; Cremona, M. L.; Frasch, A. C.; Alzari, P. M. Molecular
Cell 2002, 10(4), 757-768.


90. Damager, I.; Buchini, S.; Amaya, M. F.; Buschiazzo, A.; Alzari, P.; Frasch, A. C.; Watts, A.; Withers, S. G. Biochemistry 2008, 47(11), 3507-3512.


137. Humphrey, W.; Dalke, A.; Schulten, K. Journal of Molecular Graphics 1996, 14(1), 33-


BIOGRAPHICAL SKETCH

Özlem Demir got her BS degree in 2000 and her Master of Science degree in 2002 from the Chemistry Department of Bilkent University in Turkey. She then came to University of Florida to pursue PhD in computational chemistry.