12-2016

Effect from Substrate Binding and Selected Mutations for Allosteric Communication in Atm1 Type ABC Exporter

Yinling Liu
Clemson University, yinlinl@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation
Liu, Yinling, "Effect from Substrate Binding and Selected Mutations for Allosteric Communication in Atm1 Type ABC Exporter" (2016). All Dissertations. 1832. https://tigerprints.clemson.edu/all_dissertations/1832

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
EFFECT FROM SUBSTRATE BINDING AND SELECTED MUTATIONS FOR ALLOSTERIC COMMUNICATION IN ATM1 TYPE ABC EXPORTER

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry

by
Yinling Liu
December 2016

Accepted by:
Brian Dominy, Committee Chair
Steven Stuart
Julia Brumaghim
Weiguo Cao
ABSTRACT

This thesis employed molecular dynamics (MD) simulations to investigate the effects of substrate binding and site-direct mutation of proteins.

Chapter 2 performed MD simulations to examine the substrate effects on the behavior of Atm1 type ABC exporter. We found that thiol-containing glutathione (GSH) bound Atm1 type ABC exporter undergoes the largest conformational changes characterizing the nucleotide binding domain (NBD) coming closer together. The intracellular loops ICL1 and ICL2 between NBDs and transmembrane domains (TMDs) associated with the biggest fluctuation in the first principle component from PCA analysis in GSH bound system. Also, unlocking and slide along each other of two C-helices were observed in GSH system. The orientation of GSSG and GSH are not sterically constrained by the membrane protein due to its rotation within the large transmembrane hole. Therefore, it is impossible to assign specific binding sites in the transmembrane hole to either substrate.

The following chapter investigated the allosteric communication of both the wild-type and the double mutation Y195F/Q272A of the Atm1 type ABC transporter by MD simulations and compared to two single mutant Y195F and Q272A of Atm1 type ABC exporter in detail. In the 200 ns MD simulations, we observed that both double mutation Y195F/Q272A and single mutation Y195F exhibited large conformational flexibility using RMSD analysis. A ‘semi-closed’ geometry detected in Y195F/Q272A MD trajectory that is closer than the starting crystal structure suggests that the two NBDs approached each other. The single mutation Q272A fail to the link the TM4 and TM6, which in turn exhibits
insensitive ATPase hyperactivity due to the loose contact with the nearby residues with no bulky sidechain. Finally, further PCA indicated that the two NBDs rotate in opposite directions, which results in the asymmetric NBD dimerization. This in turn, leads to the allosteric communication between TMDs and NBDs.

In chapter 4, questions including the physicochemical mechanisms motivating the evolution of Tth UDGa are tackled through density function theory (DFT) methods, MD simulations, and a successive analysis of thermodynamic properties associated with the enzyme activity using MD trajectories. Transition states have been successfully located from DFT frequency and IRC calculations based on the B3LYP/6-31+G(d) level of theory. TthUDGa bound reactant in a typical 3'-exo sugar ring conformation that favors oxacarbenium ion through nucleotide backbone distortion. An encouraged Spearman correlation ($r_s=0.786$, $r^2=0.554$) was achieved by comparison to the experimental data. These results suggest that the electrostatic interaction changes appear to be a major role in enzyme efficiency resulting from mutation. A detailed analysis of the transition state conformation of the glycosidic bond scission, binding affinity between TthUDGa protein and flipping out uridine, and catalytic activity of TthUDGa/DNA complexes suggests that the transition state stabilization might be contributing factors for the evolutionary optimization of TthUDGa/DNA complexes.
DEDICATION

To

my beloved parents,

my husband,

and my daughter, of course
First of all, I would like to express my sincere gratitude to my advisor Prof. Brian N. Dominy, for his expert guidance, full support, and encouragement through my study and research to pursue my PhD. He has taught me much more than science.

Besides my advisor, I would like to express my special appreciation and thanks to Prof. Steve Stuart, for his various courses and insightful discussions in our joint group meeting. I would also like to thank Prof. Weiguo Cao and Prof. Julia Brumaghim for including me to their projects and their insightful comments and encouragement.

I also thank my group members Zhe Jia, Tingting Han, Vibhor Agrawal, and Richard Overstreet, for all the help and fun we have in the past four years. This journey would have been much more difficult without their constant support and encouragement.

Yinling Liu

December, 2016
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>1</td>
</tr>
</tbody>
</table>

1. INTRODUCTION.................................................................................... 1
   IMPLICIT SOLVENT MODELS................................................................... 1
   ATP BINGING CASSETTE TRANSPORTER.................................................. 3
   SUMMARY.............................................................................................. 8

2. EFFECT FROM SUBSTRATE BINDING FOR ALLOSTERIC COMMUNICATION IN ATM1 TYPE ABC EXPORTER....................... 10
   ABSTRACT............................................................................................ 10
   INTRODUCTION...................................................................................... 10
   METHODS............................................................................................. 13
   RESULTS AND DISCUSSION..................................................................... 16
   CONCLUSION......................................................................................... 35

3. EFFECT FROM SELECTED MUTATIONS IN THE LIGAND BINDING SITE FOR ALLOSTERIC COMMUNICATION IN ATM1 TYPE ABC EXPORTER.......................................................................................................................... 37
LIST OF TABLES

Table 2.1. Secondary structure assignment for Atm1 ABC exporter. ......................... 17

Table 2.2. Calculated MM/GBSA interaction energy (kcal/mol) for Atm1 ABC exporter substrates binding system. The standard error was included in parenthesis. Evdw is the Van Der Waals energy, Ecoul is the Coulombic energy, Egb is the polar solvation free energy. GSSG-a, GSSG-b represents the major and the second binding site for substrate GSSG.31

Table 4.1. Summary of Charmm partial atomic charges assigned over atoms in three different charge scaling method. (GS: ground state; TS: transition state; NC: no change, the same Charmm partial atomic charge assigned as in the ground state). ...................... 72

Table 4.2. Summary of distances (Å) between heavy atoms and angles (°, in parentheses) involved in H-bond interaction between uracil and enzymes of wild-type and mutants. 82

Table 4.3. Summary of the MM/PBSA binding free energy of ground state and transition state using ESP scaled charges for Tth UDGa/DNA complexes in native protein and its mutants. The standard errors connected to the averages are given in parentheses. ........ 82
LIST OF FIGURES

Figure | Page
--- | ---
Figure 1.1. Visualization of binding sites for GSH of ScAtm1 (A, PDB access code: 4MYH) and NaAtm1 (B, PDB access code: 4MRP). The approximate positioning of the phospholipid bilayer is indicated by a dotted line. The orange and purple cartoon diagram illustrates two subunits emphasizing the secondary structure. GSH is colored by atom in the VDW representation. ................................................................. 6
Figure 2.1. Relevant molecular models built in the Atm1 ABC exporter systems. (A) Modeled Atm1 ABC exporter with an inward-facing open conformation was inserted into a POPC bilayer. The structure of Atm1 ABC exporter is represented as new-cartoon. The POPC bilayer and TIP3P water are represented as a line. 0.15M NaCl was represented as VDW. (B) Zoom in of ATP binding site interaction, ATP molecule is shown in licorice. Each subunit of the protein is represented by new-cartoon. (C) Zoom in of GSH binding site interaction, GSH molecule is shown in licorice. ......................................................... 18
Figure 2.2. RMSDs for the backbone atoms of whole protein(A), TMDs(B), and NBDs(C), whole protein of subunit 1(D), TMDs of subunit 1(E) and NBDs of subunit 1(F) and whole protein of subunit 2(G), TMDs of subunit 2(H) and NBDs of subunit 2(I). ....................... 19
Figure 2.3. Per residue plot of RMSF of Ca atoms derived from PC1s of GSH (top) and GSSG (bottom). .................................................................................................................. 21
Figure 2.4. Contour plot of the cross-correlation motion against resid for the apo (A), GSH (B) and GSSG bound systems(C), and zoom in the cross-correlation of the GSH bound
state between subunit 1 and subunit2 (D). Representation of subunit 1 emphasizing the secondary structure arrangement.

Figure 2.5. Movement of NBDs. (A) Superimpose of the final snapshot of NBDs over the course of 200 ns MD simulation in the apo (red), GSSG (yellow) and GSH (green) system, respectively. (B) Movement of catalytically important residues E523 pairs in apo (red), GSSG (yellow) and GSH (green) system, respectively.

Figure 2.6. Time evolution of unlocking and sliding of two C-Helices along each other observed in the GSH system.

Figure 2.7. The per residue decomposition of MM/GBSA in transmembrane domain of different substrate bound systems.

Figure 2.8. Comparison of the binding site trajectory of GSSG (A-D) and GSH (E-F) in Atm1 ABC exporter. Two GSSG molecules are colored green and yellow for the major and minor binding site respectively. All figures are drawn in VMD program. (A) Detailed interactions between amino acids from protein and GSSG in two binding sites in the crystal structure (PDB access code 4MRS). Amino acids with key interactions were drawn in CPK by atom name. (B) Top view of the trajectory of the center of mass for two GSSG molecules during 200 ns MD simulation. The permeation pathways of the center of mass of primary and secondary binding site GSSG molecules is colored green and yellow, respectively. TMDs of Atm1 ABC exporter are colored by segment in the NewCartoon. (C) Zoom in front view of permeation pathways of the center of mass trajectory of two GSSGs. (D) Zoom in top view of permeation pathways of the center of mass trajectory of two GSSGs.
(E) Zoom in front view of permeation pathways of the center of mass trajectory of GSH.

(F) Zoom in top view of permeation pathways of the center of mass trajectory of GSH.

Figure 3.1. RMSDs for the backbone atoms of whole protein (blue), TMDs (red), and NBDs (blue) in wild-type (A), double mutation Y195F/Q272A (B), single mutation Y195F (C) and single mutation Q272A (D).

Figure 3.2. The sequence similarities and secondary structure information analysis from aligned sequence of Sav 1866, NaAtm1, and ScAtm1 type ABC exporter. Figure are generated using online program Espript3\textsuperscript{129-130}. Coils, arrows and character TT represent $\alpha$ helical, $\beta$ sheet and three turn secondary structures, respectively. The conserved sequence is labeled red and blocked, while identical residues are highlighted in red. Specifically, secondary structure assignment in TMD with TM1 to TM6 and functional motifs in NBDs of NaAtm1 are also highlighted in yellow and blue, respectively. Residues for double mutational analysis are shown by black arrows.

Figure 3.3. The structure of NBDs in apo state of NaAtm1 type ABC exporter. (A) Three key distances are identified in the NBDs movement. KD1 represents the $\alpha$-C distances between K400 in subunit 1 and S499 in subunit 2. KD2 represents the $\alpha$-C distances between K400 in subunit 2 and S499 in subunit 1. KD3 represents the $\alpha$-C distances between two E523 residues in two subunits. Functional motifs of NBDs are labeled by segment and colored in NewCartoon representation. The figure is generated in VMD program\textsuperscript{118}. (B) Histogram plots of the three key distance movements from three independent 200 ns molecular dynamic simulations. Three dashed lines demonstrate the original distance of KD1, KD2, and KD3 in the crystal structure 4MRN.
Figure 3.4. The final MD structures of NBDs regions and KD3 distance in the NaAtm1 ABC transporter obtained from the apo state (A), Y195F/Q272A (B), Y195F (C), and Q272A (D) system. The NBDs are colored by segment in NewRibbons representation. The figure is generated in VMD program \(^{118}\) ................................................................. 50

Figure 3.5. Asymmetric movement of KD1 and KD2 in MD snapshots in the NaAtm1 ABC transporter obtained from the apo state (A), Y195F/Q272A (B), Y195F (C), and Q272A (D) system. .................................................................................................................. 51

Figure 3.6. Side view (A) and bottom view (B) of the vector representation of PC1. The two NBDs rotates in clockwise and counter-clockwise directions. ................................. 55

Figure 4.1. Hydrolysis of damaged DNA under the catalysis of UDG. A general sketch of the division of the systems into QM and MM atoms was also shown (highlight in the blue curve is QM region). Essentially, the QM region consists of the base uracil, the ribose and two linked H atoms attached to O5’ and O3’ terminal. Amino acids of enzyme were assigned to MM region. .................................................................................................................. 68

Figure 4.2. Properties of transition state and reactants along IRC at B3LYP/6-31+G(d) level drawn by GaussView. A. Total energy; B. RMS gradient norm of total energy; C. C1’-N1 bond distance; D. O4’-C1’ distance; E. Dihedral of C5’-C4’-C3’-O3’; F: Dihedral of O4’-C1’-N1-C2. .................................................................................................................. 69

Figure 4.3. Electrostatic potential maps of optimized reactant (A) and transition state (B) drawn by GaussView for the cleavage of glycosidic bond at B3LYP/6-31+G(d) level. Top: stick diagram; bottom: electrostatic potential surface maps, blue and red regions represent positive and negative potentials, respectively. This result suggested that the transition state
is highly active and need to be stabilized via delocalization of the negative charge on N1 through partial glycosidic bond cleavage. (C): A partial empty 2p_z orbital which hyperconjugates with the lone pair of O4’ and σ (C2’-H2’) electrons in the LUMO of transition state shown in Gaussview. This result showed that the hybridization of the anomeric carbon had been changed from sp3 to sp2, which thus stabilized the transition state in enzyme reaction.........................................................73

Figure 4.4. Visualization of active site interaction between family 4 Tth UDGa and double chain DNA in VMD program. A: Model built: wild-type family 4 Tth UDGa structure (colored by secondary structure in NewCartoon) with aligned-uracil-containing double strand DNA. Uracil is colored in licorice and DNA is colored in NewRibbons; B: Close-up view of wild-type uracil and catalytic amino acids in active site pocket of minimized structure. Hydrogen bond interactions between amino acids and uracil are shown as black dashed lines. Uracil is using the same coloring scheme described in (A). E41, G42, F54, N80, and H155 are colored by name in CPK model represents. ........................................75

Figure 4.5. Per-residue decomposition of the MM/PBSA binding affinity of wild-type Tth UDGa/DNA complexes for substrate uridine in the ground state (A) and transition state (B) and three different binding free energy change (C, D and E) from ground state to transition state in Charmm, where negative values indicate residues that contribute favorable interaction, positive values indicate residues that contribute unfavorable interactions, zero values indicate residues not involved in enzyme catalysis. Four groups of amino acid groups demonstrate obvious favorable interactions with substrate uridine, motif 1 (residue from 40 to 44), motif 2 (residue from 155 to 162), turn 1 (residue 53 to 58) and turn 2

..........................................................
(residue 80 to 93). The greatest contribution of catalytic affinity was from His155. Three
MM/PBSA binding free energy of transition state use all the same parameters besides the
charges which were rescaled different in Charmm. A: Ground state binding free energy; B:
transition state binding free energy of ESP charge scaling method. C: Binding free energy
change from ESP charge scalar methods (method 1). D: Binding free energy change from
Mulliken charge scaling methods (method 2). E: Binding free energy change from Charmm
partial charge initial guess charge scaling (method 3).

Figure 4.6. Interactions and two-dimensional scatter plots of the wild-type and mutant Tth
UDGa proteins with O2 of uracil in the active site. Modeled interactions with O2 of uracil
in the active site of Tth UDGa-WT (A), Tth UDGa-E41Q (B), Tth UDGa-G42D (C) and
Tth UDGa-E41Q-G42D (D). dUMP is colored by atom type. Amino acid residues in the
active site of Tth UDGa are shown in licorice in program VMD. Two-dimensional scatter
plots of heavy atom distances between E41, H155 and uridine in Tth UDGa-WT (E), Tth
UDGa-E41Q (F), Tth UDGa-G42D (G) and Tth UDGa-E41Q-G42D (H).

Figure 4.7. Spearman rank correlation analysis between calculated enzyme efficiency and
experimental $k_{cat}/K_M$ data. These results suggest that the electrostatic interaction changes
appears to be a major role in enzyme efficiency resulting from mutation.

Figure 4.8. Contour profile of backbone RMSDs of individual amino acids relative to the
minimized structure along MD trajectory.

Figure 4.9. Close-up views of the Tth UDGb-DNA active site interactions in the energy
minimized structures. Tube trace of the protein is colored in tan. Uridine (U), inosine (I),
and xanthosine (X), as well as amino acids Asp75, Asn120, and His190, are colored by
atom type. Water molecules are labeled in purple. Dashed lines indicate inferred hydrogen bonds or water association. (A), modeled Tth UDGb-U interactions. (B), modeled Tth UDGb-I interactions. (C), modeled Tth UDGb-X interactions. (D), chemical illustration of Tth UDGb-U interactions. (E), chemical illustration of Tth UDGb-I interactions. (F), chemical illustration of Tth UDGb-X interactions. 

Figure 4.10. Comparison of interactions between Tth UDGb-N120 with water and E. coli MUG-N18 with water. A, superimposition of UDGb-AP structure (Protein Data Bank code 2DEM; green) with E. coli MUG structure (Protein Data Bank code 1MUG; silver). The AP site is colored by atom type. The two structures were superimposed using the program VMD. B, close-up view of Tth UDGb-N120-water and E. coli MUG-N18-water interactions using the same coloring scheme described in A. Asn120 and the interacting water in Tth UDGb structure are shown in purple. Asn18 and the interacting water in the MUG structure are shown in orange.

Figure 4.11. Modeling of interactions between glycosylase and uracil. The hydrogen bonds are shown in blue. (A) Interactions between MUG-WT and uracil based on a model built from the solved crystal structure (PDB 1MUG). K68 and uracil are colored by atom type. (B) Interactions between MUG-K68N and uracil. MUG-K68N structure was modeled from PDB MUG-K68N. K68N and uracil are colored by atom type. (C) Interactions between UNG-WT and uracil based on the solved crystal structure (PDB 2EUG). N123 and uracil are colored by atom type. (D) Interactions between UNG-N123A and uracil. UNG-N123A structure was modeled from PDB 2EUG. N123A and uracil are colored by atom type. (E) Interactions between Tth UDGb-WT and uracil based on the solved crystal structure (PDB
2DEM). A111 and uracil are colored by atom type. (F) Interactions between UDGb-A111N and uracil. UDGb-A111N structure was modeled from PDB 2DEM. A111N and uracil are colored by atom type.
CHAPTER ONE

INTRODUCTION

IMPLIED SOLVENT MODELS

Molecular dynamics (MD) examines the dynamic evolution of a system of interacting particles by solving Newton’s equations of motion numerically. The force \( F_i \) on an atom due to a potential \( V \) can be described using the negative gradient of that potential.

\[
F_i = -\nabla V(r_i)
\] (1.1)

The total potential \( V \) of the CHARMM force field is a sum of the bonded potential including bond stretching (Equation 1.2), bond bending (Equation 1.3), bond twisting (Equation 1.4), improper potentials (Equation 1.5) plus the non-bonded potential consisting of Van Der Walls interactions (Equation 1.6) and electrostatic interactions given by Coulomb’s law (Equation 1.7).

\[
V_{\text{bond}} = K_b(b - b_0)^2
\] (1.2)

\[
V_{\text{angle}} = K_\theta(\theta - \theta_0)^2
\] (1.3)

\[
V_{\text{dihedral}} = K_\chi(1 + \cos(n\chi - \delta))
\] (1.4)

\[
V_{\text{improper}} = K_\Psi(\Psi - \Psi_0)^2
\] (1.5)

\[
V_{\text{Lennard–Jones}} = \varepsilon_{i,j}\left[\left(\frac{R_{\text{min},i,j}}{r}\right)^{12} - 2\left(\frac{R_{\text{min},i,j}}{r}\right)^6\right]
\] (1.6)

\[
V_{\text{elec}} = \frac{q_i q_j}{4\pi\varepsilon r}
\] (1.7)

Where \( q_i \) and \( q_j \) are the partial atomic charges, \( r \) represents their separation distance, and \( \varepsilon \) is the dielectric constant of the surrounding environment. In principle, Van Der Walls
interaction and electrostatic interactions act over some distance between all pairs of atoms, which would demand a troublesomely large number of calculations. To solve this issue, a nonbonded cutoff distance is specified to periodically update the nonbonded list during a simulation.

Solvent, usually water, which has been incorporated in MD simulations, has a screening effect on the electrostatic interactions. The simplest treatment of solvent effects is to include a distance dependent dielectric screening constant $\epsilon_{eff} = r_{ij}\epsilon$, where $\epsilon$ ranges from 4 to 20, in the electrostatic term in Equation 1.7. Other implicit screening solvent models based on the solvent exposed surface area of the protein and continuum electrostatic theory have been developed recently.

Among all implicit solvent models, Molecular Mechanics/ Poisson-Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA) approaches have been extensively used in recent years. Excellent review articles about these two methods have also been reported. Both of these two approaches are more computationally efficient than thermodynamic integration (TI) and free energy perturbation (FEP) and more rigorous than docking. They give satisfactory prediction results in the prediction of binding free energies.

If we consider the Poisson equation for the polar solvation energy:

$$\nabla[\epsilon(r)\nabla\phi(r)] = -4\pi\rho(r)$$ \hspace{1cm} (1.8)

Where $\epsilon$ is the dielectric constant, $\phi$ is the electrostatic potential, and $\rho$ is the charge distribution. The Equation 1.8 becomes the Poisson-Boltzmann (PB) equation if an extra Boltzmann factor was added then the ionic strength is considered.
Solving Equation 1.8 under two conditions, one is the gas phase condition where \( \varepsilon = 1 \) and the other is in solvent condition where \( \varepsilon = \varepsilon_w \), the dielectric constant of the solvent, give you \( \varphi_{\text{react}} \), which is the difference between these two potentials.

\[
\Delta G_{\text{pol}} = \frac{1}{2} \sum_i q_i \varphi_{\text{react}}(\vec{r}_i) \tag{1.9}
\]

The potential can be reduced in the case of a simple ion of radius \( a \), which leads to the well-known Born formula

\[
\Delta G_{\text{Born}} = -\frac{q^2}{2a} \left( 1 - \frac{1}{\varepsilon_w} \right) \tag{1.10}
\]

Equation 1.9 can be given as follows by a summation of individual Born terms with pairwise electrostatic potential terms

\[
\Delta G_{\text{pol}} = \sum_{i}^{N} \frac{q_i^2}{2\varepsilon_w} \left( \frac{1}{\varepsilon_w} - 1 \right) + \frac{1}{2} \sum_i^{N} \sum_{j \neq i}^{N} \frac{q_i q_j}{r_{ij}} \left( \frac{1}{\varepsilon_w} - 1 \right) \tag{1.11}
\]

The purpose of GB theory is trying to acquire a fairly straightforward analytical formula which will capture the physics of the PB equation as much as possible for the real molecular geometries. In this thesis, I will use both MM/PBSA and MM/GBSA methods to investigate the binding free energy of ATP-binding cassette (ABC) transporters in Chapter 2 and Uracil DNA Glycosylase -Tth UDGa/DNA complexes in Chapter 4.

**ATP BINDING CASSETTE TRANSPORTER**

Membrane proteins are crucial players in fundamental cellular processes such as surface recognition, cell-cell contact, signaling, energy conversion, and transporting substances across the plasma membrane. Understanding how a biological membrane protein works has both fundamental and practical significance to a human being since life processes are
cellular processes. Considerable experimental efforts have been made in better understanding their biological structure and function in a phospholipid bilayer environment\textsuperscript{19-27}. However, the functions of membrane proteins like transporters and channels encompass dynamic processes which still cannot be captured by any experimental method today\textsuperscript{28}. MD simulations of membrane proteins can fill this gap serving as a “computational microscope” to provide deeper insights into their architectures, functions and interactions with surrounding environment at the atomic level\textsuperscript{29-37}. Considering that high-resolution X-ray crystallographic structures of membrane proteins are particularly difficult to obtain, and theoretical and computational simulation of membrane protein systems are normally about more than 100,000 atoms on a time scale of hundreds of nanoseconds, MD simulation of a realistic membrane protein complex is still challenging and thus a major goal in modern structural biology.

ATP-binding cassette (ABC) transporters establish one of the largest transporter superfamilies and are ubiquitously found in all kingdoms of life\textsuperscript{38-46}. They are principal active transporters that couple the energy provided by ATP binding and hydrolysis to the directional transport diverse substrates across cellular membranes. In eukaryotic cell, ABC transporters are associated with hereditary diseases such as cystic fibrosis\textsuperscript{47} and their role in multidrug resistance (MDR) of cancer cells for extruding therapeutic drugs\textsuperscript{48-51}. Several examples include bacteria catalyzing the intake of elemental nutrients or the excretion of toxic substances, which leads to drug and antibiotic resistance of microbial pathogens\textsuperscript{52-53}. The molecular architecture of ABC transporters includes a nucleotide binding domain (NBD) responsible for ATP binding and hydrolysis, which drives conformational changes
of a transmembrane domain (TMD) in a phospholipid bilayer environment. ABC transporters can be strictly allocated to families of exporters or importers\cite{54-56} based on the sequence homology of their core TMDs with known crystal structures. Multidrug transport Sav1866 from \textit{S. aureus} is the first illuminated ABC exporter by crystallographic analysis\cite{57}. Protein functions mainly depend on their structures. To date, the crystallized ABC exporters comprise 2*6 transmembrane helices (TMs) and is structurally distinct from the importers, which are characterized by a minimal core of 2*5 TMs or 2*10 TMs\cite{58-63}. The NBDs are the motor domains of ABC exporters and several conserved sequence motifs are contributing to ATP binding, hydrolysis and trigger the TMDs\cite{64}. These motifs include the ‘Walker A’ motif that binds the nucleotide; the ‘Walker B’ motif that orients the nucleophilic attack on ATP via a water molecule; the ‘ABC signature motif’-LSGGQ that contacts the nucleotide in the ATP-bound site; the ‘switch motif’ contributes to the catalytic reaction\cite{65}; the ‘A-loop’ critical for ATP binding\cite{66-68}, and the ‘D-loop’ and ‘Q-loop’ that are involved in the contact interface with two NBDs and TMDs, respectively. Domain motions of a heterodimeric exporter TM287/288 suggests that ATP remains bound at the degenerate site, while ATP binding and hydrolysis at the consensus site and drive the transport cycles\cite{69}. Enzyme assays and MD simulation on MsbA demonstrates that the coupling helices CH2 is significantly important for the occurrence of the node-like motion\cite{70}. Cross-linking findings pinpoints that neighboring TM helices in the multidrug ABCB1 are coherent with the helix arrangement detected in multidrug transport Sav1866\cite{71}, indicating that Sav1866 probably acts as a decent model for the center architecture of all ABC exporters\cite{72}. Several molecular dynamic studies about catalytic conformation
transitions on Sav1866, BtuCD, P-glycoprotein and other ABC exporters using homology modeling based on Sav1866 have been reported. The “alternating access” model was used to describe the interconversion between outward-facing and inward-facing conformations of ABC exporters associated with ATP hydrolysis. This simple scheme can explain nutrient import by ABC importers and drug extrusion by ABC exporters theoretically. ABC exporters can facilely undergo conformational rearrangement at the interfaces between TM helices coupled to ATP hydrolysis.

Figure 1.1. Visualization of binding sites for GSH of ScAtm1 (A, PDB access code: 4MYH) and NaAtm1 (B, PDB access code: 4MRP). The approximate positioning of the phospholipid bilayer is indicated by a dotted line. The orange and purple cartoon diagram illustrates two subunits emphasizing the secondary structure. GSH is colored by atom in the VDW representation.
The Atm1 family ABC exporters use ATP as an energy source in the export of substrates across the membrane in transition metal homeostasis and detoxification processes. Two high-resolution X-ray crystallographic structures of Atm1 family ABC exporters from Novosphingobium aromaticivorans (NaAtm1) and Saccharomyces cerevisiae (ScAtm1) have been solved recently\textsuperscript{58,86}. Both NaAtm1 and ScAtm1 have ~1200 amino acids, adopt an inward-facing, open conformation (Figure 1) and share ~45\% sequence identity. The cellular antioxidant thiol-containing glutathione (GSH) binding site located close to the inner membrane surface in a large cavity emphasizes the articulated design of Atm1 type ABC exporters. NaAtm1 and ScAtm1 both form a dimer, with each subunit containing 6 transmembrane helices in the transmembrane domains coupled to the nucleotide binding domains. The strong interaction of the two C-terminal α-helices prevents wider opening of the Atm1-type ABC exporter dimers. The crystallographic analysis and functional searches show that glutathione derivatives can function as substrates for the Atm1 family ABC exporters protect against silver and mercury toxicity in cellular detoxification processes\textsuperscript{58}. Furthermore, both free α-amino and α-carboxyl groups in the transported ligand has been reported to be required but not sufficient for effective stimulation of ATPase activity. Two binding sites of glutathione derivatives GSSG of NaAtm1 had been reported. The kinetic constants $k_{\text{cat}}/K_M$ for the NaAtm1 ATPase activity of S-Ag GSH and S-Hg GSH are 2400 and 200 min$^{-1}$ mM$^{-1}$ which were the highest activity compared to 29 and 3.8 min$^{-1}$ mM$^{-1}$ for GSSG and GSH, respectively. The kinetic constants $K_M$ for the substrates GSSG and GSH in NaAtm1 ATPase activity are 0.97±0.12 and 15±1 mM, respectively. The
proficiency of NaAtm1 to transport GSSG instead of GSH differentiates it from the *E.coli* CydDC ABC transporter which is selective for GSH\(^8\).

**SUMMARY**

Although extensive progress has been made in the structural inquiry of ABC exporters, how Atm1-type ABC exporters selectively distinguish substrates and how substrate binding is coupled to conformational rearrangement of the TMDs remains to be identified. When combined with computationally sophisticated and rigorous free energy techniques, molecular dynamics simulations of Atm1-type ABC exporters can elucidate the microscopic origins of substrate specificity in the primary binding site and the second binding site mentioned above and the driving forces of biological function during the substrate translocation cycle. The effect of the substrate including GSH and GSSG for allosteric communication will be investigated in Chapter 2.

The particular roles of catalytic amino acids in enzymatic reactions is a major source of the massive catalytic power of enzymes. Enzyme reactions work by lowering activation energy (\(\Delta G^\circ\)). Many excellent reviews about the modeling of electrostatic effects in enzymes have been reported\(^89-94\). However, since many factors could have an impact on the alteration in binding energy when moving from the reactant state to the transition state, the origin of the catalytic power of enzymes is still poorly understood\(^95\). Among all the techniques developed to help interpret the role of catalytic amino acids in enzyme reactions, site-directed mutation experiments are widely used. The catalytic groups may perfectly change the orientation in mutant variants and the entire enzyme activity can thus be enhanced or
weakened after mutation. The site-directed mutagenesis studies show that mutations N269A, Q272A, and G319L all effectively minimize GSH stimulation but result in varying degrees of the ATPase activity. Both mutant Y195F and double mutant Y195F/Q272A with a high basal ATPase activity without substrate GSH simulation. In Chapter 3, I will make double mutations and single mutations in the ligand binding site to examine the different binding architectures, functions and interactions with surrounding TMDs through long trajectory of MD simulations at the atomic level.

Finally, in Chapter 4, I will use classical MD simulation and MM/PBSA analysis including per residue decomposition of the free energy to investigate the impact of single amino acid mutations on the binding affinity and catalytic activity toward the family 4 Uracil DNA Glycosylase -Tth UDGa/DNA complexes. Hopefully, my present thesis of UDG dynamics can improve scientific understanding toward the early stage in base excision repair pathway about DNA repair.
CHAPTER TWO

EFFECT FROM SUBSTRATE BINDING FOR ALLOSTERIC COMMUNICATION IN ATM1 TYPE ABC EXPORTER

ABSTRACT

All ATP binding cassette (ABC) transporters couple the energy provided by ATP binding and hydrolysis to the directional transport diverse substrates across cellular membranes. We have performed Molecular dynamics (MD) simulations to examine the substrates effects on the behavior of Atm1 type ABC exporter in detail. We found that thiol-containing glutathione (GSH) bound Atm1 type ABC exporter undergoes the largest conformational changes characterizing the nucleotide binding domain (NBD) coming closer together. The intracellular loops ICL1 and ICL2 between NBDs and transmembrane domains (TMDs) associated with the biggest fluctuation in first PCA component in the GSH bound system. Also, unlocking and sliding of the two C-Helices along each other were observed in the GSH system. The orientation of GSSG and GSH are not sterically constrained by the membrane protein, and it rotates within the large transmembrane hole. Therefore, it is impossible to assign specific binding sites in the transmembrane hole to either substrate. The findings of this study provide essential visions into the structure-function relationship of ABC exporter.

INTRODUCTION
Anticancer drug resistance is mostly attributable to the overexpression of efflux drug transporters in cancer cells\textsuperscript{96}. Even the mechanisms of multidrug resistance (MDR) appear to be complex, one of the most important mechanisms underlying MDR is the overexpression of ATP binding cassette (ABC) transporters\textsuperscript{97-103}. ABC transporters form one of the largest families of transmembrane proteins. ABC exporter uses ATP as an energy source in the export of substrates. Structurally, all ABC exporters have two transmembrane domains (TMD) and two nucleotide binding domains (NBD). The NBDs are the motor domains of ABC exporters and several highly conserved sequence motifs are contributing to ATP binding and hydrolysis\textsuperscript{70, 80, 104-105}. These motifs include the ‘Walker A’ motif that interacts with the phosphate group of ATP; the ‘Walker B’ motif that coordinates Mg\textsuperscript{2+} and orients the nucleophilic attack on ATP molecule via a water molecule; the ‘ABC signature motif’-LSGGQ that binds the nucleotide in the ATP-bound site; the ‘A loop’ that pi-pi stacks with the adenine ring of ATP; and the ‘D-loop’ and the ‘Q-loop’ that is involved in the communication between NBDs and TMDs inter-domain. The hydrophobic TMDs which is composed of six TM helices per subunit are involved in substrates recognition and efflux. The TMDs are believed to undergo substantial conformation changes during the entail transport process.

Despite the availability of more resolved crystal structures during recent years\textsuperscript{57, 62, 106-109}, no substrates-binding ABC exporter crystal structure have been solved until the Atm1 type ABC exporter with bound antioxidant thiol-containing glutathione (GSH) in 2014\textsuperscript{58}. In Atm1 type ABC exporter, binding site of the cellular antioxidant thiol-containing glutathione (GSH), located near to the inner membrane surface in a big cavity features the
articulated design of Atm1 type ABC exporters (Figure 2.1). Two binding sites of glutathione derivatives GSSG of Atm1 type ABC exporter had been reported. The mechanistic questions of how binding of substrates at TMDs is interconnected to the NBDs to direct the transport cycle in ABC exporter remain unresolved. The “alternating access” model was used to describe the interconversion between outward-facing and inward-facing conformations of ABC exporters associated with ATP binding and hydrolysis. Questions remain, that although residues in the TMD-NBD interface involved in the communication pathway of TMD-NBD allosteric mechanism have been identified, the exact signaling mechanism remains unclear. The reciprocating model proposes that substrates binding to TMD induce an isomerization of the transporter which will lead them towards the reciprocal conformations. Although the reciprocating model provides a mechanistic path by which substrates binding is connected to the NBDs, it appears apparently difficult to reconcile with the available crystallographic information of ABC transporter families45. Although considerable advancement has been made in the crystal structural analysis of ABC exporters, several inquiries remain. Specifically, how Atm1-type ABC exporters selectively distinguish substrate GSSG and GSH, how substrate binding is linked to the conformational rearrangement of TMDs, and how the allosteric effects between the TMD and NBD induced by GSSG and GSH binding remain to be identified. MD simulations of Atm1-type ABC exporters can elucidate the microscopic origins of substrate specificity in the TMDs and the driving forces of biological function during the substrate translocation cycle. In order to decipher how the mechanical force produced by substrates binding is transmitted in the ABC transporter, we use an Atm1 type ABC exporter as a model. We
have used molecular modeling and graph theoretic analysis techniques of the six possible ligand-bound states of the molecule: no ligand (apo), GSH bound, GSSH bound, ATP-bound, GSH/ATP-bound, and GSSG/ATP-bound. The simulations revealed that GSH bound undergoes distinct structural changes. We found that thiol-containing glutathione (GSH) bound Atm1 type ABC exporter undergoes the largest conformational changes characterizing the nucleotide binding domain (NBD) coming closer together. In GSH bound systems, the intracellular loops ICL1 and ICL2 between NBDs and the transmembrane domains (TMDs) associated with the largest fluctuation in the first PCA component. Also, unlocking and sliding of the two C-Helices along each other were observed in the GSH system. The MD simulations were successfully interpreted for several physical and dynamic processes from the empirical analysis.

METHODS

Molecular Dynamic simulations

Pre-oriented protein coordinates with reference to the membrane normal of apo, GSH bound and GSSG bound Atm1 type ABC exporter (PDB code 4MRN, 4MRP, and 4MRS respectively) were taken from OPM database\textsuperscript{110} and used as the starting structures for all simulations. Topology and parameters for the GSH and GSSG molecules were derived from CHARMM General Force Field (ParamChem)\textsuperscript{111}. Engineered mutation residues selenomethionine (MSE) was reversed to methionine (MET). The homodimer protein was embedded into 1-Palmitoyl-2-oleoylphosphatidylcholine (POPC) lipid bilayer using CHARMM-GUI server\textsuperscript{112-113}. The lipid-protein systems were solvated via TIP3P water
model and neutralized by adding 0.15 M NaCl counter ions. The ultimate system size of individually membrane embedded models is 111*111*169 Å³ and consisting of 189,656 total number of atoms (Figure 2.1A).

Molecular dynamic simulations were performed using NAMD 2.9\textsuperscript{114} with the CHARMM36\textsuperscript{115-116} force field. The SHAKE algorithm was applied to constrain the bonds containing hydrogens. A cutoff of 12 Å was used for van der Waals interactions. Particle-mesh Ewald summation\textsuperscript{117} was used in the periodic boundaries condition for the efficient calculation of long-range electrostatic interaction. Langevin dynamics were employed to maintain a constant temperature of 310K in each simulation. Constant pressure simulations were run at 1 atm using a Langevin piston technique with a decay period of 250 fs and a damping timescale of 125 fs. An integration time step of 1fs was used in order to avoid any substantial structural deformation during equilibration and production run. Coordinates were saved every 2ps. The equilibrated structure was used for the production run in which no restraints were applied. A total time of 200 ns was implemented for the production simulation and subsequent structural analysis. VMD 1.9.1\textsuperscript{118} had been used for visualization purposes. With the purpose of regulating whether the interactions between substrates GSH or GSSG and the TMDs of Atm1 ABC exporter observed in the MD simulation were biased by the initial velocity of the protein, three independent MD simulations were performed.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) of the C-alpha atom coordinates in Atm1 ABC exporter was performed to recognize and characterize large-scale conformational changes.
PCA analysis is created on a diagonalization of the variance-covariance matrix, obtained from the atomic motions of the MD simulation trajectories. The eigenvectors correspond to the dominant or independent modes of atomic fluctuations, and the eigenvalues correspond to the contribution of each eigenvector to the global movements of the system. On the whole, the principal eigenvalues and the corresponding eigenvectors represent the most meaningful global motions of the system. PCA of the C-alpha atom trajectories was achieved using the VMD program. N-terminal amino acids 8–20 and C-terminal amino acids 600–607 were not included in the PCA analysis. Fluctuations were referenced to the first frame of the trajectory.

To illustrate the correlation of the fluctuations between residues, the covariance of the C-alpha pairs displacement of a dynamic trajectory were calculated using the “coor cova” module in Charmm. The covariance values are within the scope of (-1,1), where negative numbers reveal the pairs of residues moving in the opposite direction along the eigenvector mode, positive numbers suggest pairs of residues moving in the same direction. The cross-correlation function is normalized where the +1 value means a fully correlated motion between a residue pair. A cross-correlation of -1 suggests fully anti-correlated motion, and zero means uncorrelated fluctuations.

**MM/GBSA interaction energy calculation**

Free energy calculation method, Molecular Mechanics Generalized Born Surface Area (MM/GBSA) approach, \( G = \langle G_{\text{MM}} \rangle + \langle G_{\text{GB}} \rangle + \langle G_{\text{SA}} \rangle \), was used to predict Atm1 ABC exporter–substrates binding affinity. The molecular mechanics energy, \( G_{\text{MM}} \), used an infinite cutoff to evaluate the non-bonded interactions. The bulk solvent dielectric
constant and protein interior dielectric constant were set to 80.0 and 4.0, respectively. Snapshots for MM/GBSA analysis were extracted every 10 ps from the 200 ns production phase MD trajectories. The non-entropic contribution of MM/GBSA interaction energies change between each amino acid of the enzyme and the flipping out of a nucleotide after binding, \( \Delta G = G_{\text{complex}} - G_{\text{enzyme}} - G_{\text{sub}} \), were also evaluated.

RESULTS AND DISCUSSION

In this study, the homodimer protein with an inward-facing open conformation was embedded into a POPC lipid bilayer (Figure 2.1A). The secondary structure assignment for Atm1 ABC transporter was list in Table 2.1. First, we performed MD simulations of the three possible states of Atm1 ABC exporter with no ATP-bound: no ligand (apo), GSH bound, and GSSG bound. As a comparison, we performed three additional MD simulations with ATP bound corresponding to the three systems mentioned above. To simulate the ATP-bound state, two ATP molecules were bound to the NBDs of Atm1 ABC exporter using the following procedure. First, according to the E-value (3e-97) and identity score (56%) using Basic Local Alignment Search Tool (BLAST) program, nucleotide binding domain of human ABCB6 (PDB access code 3NH6) was selected as the template. Then, the ATP-bound Atm1 ABC exporter was obtained by manually adding the ATP molecule which was extracted from the crystal structure of the ATP-bound form of the nucleotide binding domain of human ABCB6 (PDB access code 3NH9) to the superimposed structure of human ABCB6 and Atm1 ABC exporter. To summarize, there are in total six systems investigated in this study: apo, GSH-apo, GSSG-apo, ATP, GSH/ATP, and GSSG/ATP
systems. The specific ATP binding site interaction and substrates GSH or GSSG binding site interaction were illustrated in the Figure 2.1B and C. As shown in the Figure 2.1B, ICL1 and ICL2 are located near the adenine part of ATP, which directly interacts with the adenine ring in a hydrophobic manner (ND1@H132--N1@ATP). We can also see that the substrate GSH maintain multiple interactions with surrounding amino acids in the ligand binding site.

Table 2.1. Secondary structure assignment for Atm1 ABC exporter.

<table>
<thead>
<tr>
<th>Name</th>
<th>Resid TMD</th>
<th>Resid NBD</th>
<th>Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Helix</td>
<td>16-27 TM1</td>
<td>34-68 Walker A</td>
<td>394-402</td>
</tr>
<tr>
<td>ECL1</td>
<td>69-75 TM2</td>
<td>76-126 Q loop</td>
<td>441-445</td>
</tr>
<tr>
<td>ICL1</td>
<td>129-136 TM3</td>
<td>137-179 Signature</td>
<td>498-504</td>
</tr>
<tr>
<td>ICL2</td>
<td>231-236 TM4</td>
<td>183-229 Walker B</td>
<td>518-523</td>
</tr>
<tr>
<td>ECL2</td>
<td>294-297 TM5</td>
<td>240-292 D loop</td>
<td>525-529</td>
</tr>
<tr>
<td>C-Helix</td>
<td>587-607 TM6</td>
<td>299-340 H loop</td>
<td>553-559</td>
</tr>
</tbody>
</table>
Figure 2.1. Relevant molecular models built in the Atm1 ABC exporter systems. (A) Modeled Atm1 ABC exporter with an inward-facing open conformation was inserted into a POPC bilayer. The structure of Atm1 ABC exporter is represented as new-cartoon. The POPC bilayer and TIP3P water are represented as a line. 0.15M NaCl was represented as VDW. (B) Zoom in of ATP binding site interaction, ATP molecule is shown in licorice. Each subunit of the protein is represented by new-cartoon. (C) Zoom in of GSH binding site interaction, GSH molecule is shown in licorice.
Figure 2.2. RMSDs for the backbone atoms of whole protein (A), TMDs (B), and NBDs (C), whole protein of subunit 1 (D), TMDs of subunit 1 (E) and NBDs of subunit 1 (F) and whole protein of subunit 2 (G), TMDs of subunit 2 (H) and NBDs of subunit 2 (I).
GSH system exhibited the largest conformational changes from RMSD analysis. The root mean square deviations (RMSDs) of the backbone atoms from the minimized structure were separately calculated for the TMDs, NBDs, and the whole protein among the three systems studied (Figure 2.2). As shown in the RMSD profiles, our simulations are stable during the 200 ns MD simulations. The reasonable RMSDs are in the range of 1-5 Å during the 200 ns MD simulation for the whole protein. The final whole-protein RMSD value of the GSH system exhibited a value of 5.0 Å, which is greater than that obtained for the GSSG system (approximately 3.5 Å) and apo system (approximately 2.6 Å). As shown in Figure 2.2, the RMSD values of the GSH system showed a larger value than that of the GSSG and apo systems in both TMDs and NBDs calculations (Figure 2.2 B-C). With a focus on the individual subunits (Figure 2.2 D-I), asymmetric changes in the RMSD of the backbone atoms have been observed among all three systems. In the subunit 1 state, the RMSD values of GSH system was the largest with a final value around 4.1 Å. However, the difference between these three systems is not significant for the subunit 2 state. RMSD values of the GSSG system were larger than the corresponding values in the GSH system during a 200 ns simulation. Interestingly, we found that the two subunits in the starting crystal structure are asymmetric. The RMSD value of two subunits is 1.25 Å and the RMSD value of two TMDs and two NBDs is 1.0 and 0.4 Å in the apo crystal structure, respectively. These findings imply that NBD dimerization can be initiated by substrate binding. However, considering the structural equivalence of the two subunits and their stochastic nature, similar events would occur at the alternate subunit if a greater number or longer MD simulations were conducted.
Figure 2.3. Per residue plot of RMSF of Ca atoms derived from PC1s of GSH (top) and GSSG (bottom).
Key groups contributed to the fluctuation identified using Principal Component Analysis (PCA)

To understand how the NBD of the GSH bound system can undergo such a large conformation change, we conducted a principal component analysis (PCA) of 1198 C-alpha atoms in three systems. PCA outlines a set of eigenvectors derived from the variance-covariance matrix, obtained from the atomic motions of MD simulation trajectories. Eigenvectors are ranked corresponding to the amplitude of the global motions they describe. In our case, the relative contribution of the first PCA component (PC1), was found to be 51.8%, 41.1% and 29.4% in GSH, GSSG, and apo systems, respectively. It showed that the PC1 explains the majority of the protein motion in Atm1 ABC exporter systems and thus we will mostly focus on the PC1 component. Figure 2.3 shows the root mean square fluctuation (RMSF) of per residue plot from the dominated first PCs. It is worthwhile to mention that both of the intracellular loop ICL1 and extracellular loop ECL1 of apo protein drastically changes the dynamics after binding of substrates GSH or GSSG. The intracellular loops ICL1 and ICL2 between NBDs and TMDs associated with the largest fluctuation in the first PCA component in the GSH bound system. To our interest, the ICL2 caused the most significant RMSF change when compared GSH to the apo system. PCA analysis elucidates that ICL1 and ICL2 are responsible for large-scale conformational changes. As shown in the structure of the ATP-binding site in Atm1 ABC exporter (Figure 2.1B), ICL1 and ICL2 are located near the adenine part of ATP, which directly interacts with the adenine ring in a hydrophobic manner (ND1@H132--N1@ATP). The structural and functional roles of the intracellular and extracellular loop in the ABC transporter are
revealed by other experimental researchers recently. The mutation of ICL2 decreased the ATPase activity to 32% and significant reductions in the ATP binding affinities of that of wild-type of ABC transporter MsbA\textsuperscript{70}. ICLs are essential for the maintenance of the native 3D arrangement between the TMDs and NBDs. Therefore, the different interactions of ICLs with NBDs are likely to cause the increased ATPase activity because of the substrates binding.
Figure 2.4. Contour plot of the cross-correlation motion against resid for the apo (A), GSH (B) and GSSG bound systems (C), and zoom in the cross-correlation of the GSH bound state between subunit 1 and subunit 2 (D). Representation of subunit 1 emphasizing the secondary structure arrangement.
To illustrate the correlated motion between residues of Atm1 ABC exporter, the covariance of the C-alpha pairs displacement of a dynamic trajectory from three systems was calculated (Figure 2.4). In contrast to the apo state that showed the weakest correlation and anti-correlated motion (Figure 2.4A), a significant correlation and anti-correlated motion involved in several areas occurred in the GSH bound state (Figure 2.4B, D). The most notable of this correlation involved an alteration in the correlation motion between TM1&2 in the transmembrane domain. As shown in the zone a (Figure 2.4D), TM1 and TM2 had a very strong correlated motion within itself but very strong anti-correlated motion between the two subunits. The analysis showed that in all studied three systems, very large TM4&5 correlated motion movements with TM1&2 in the other subunit (zone b1 and b2 in Figure 2.4D). The TM4&5 movement of subunit 1 is also correlated with motions in the NBD of subunit 2. In the GSH bound state, the ICL1 in subunit 1 involves a strong anti-correlated motion with the NBD of subunit 2. Also, ICL2 in subunit 1 undergoes concerted motions with the NBD of subunit 2 (zone c and d in Figure 2.4D). Functional motions of a protein may be represented by only a few low-frequency modes. Both the motions between ICL1, ICL2, and the NBD can be characterized quantitatively in the cross-correlation of C-alpha atoms. Generally, the alternating access transport mechanism is used to describe the interconversion between outward- and inward-facing conformations of the ABC transporter. As seen from the cross-correlation of C-alpha atoms from MD trajectories above, we found that the interconversion of Atm1 ABC transporter illustrated by the departure of either TM1&2 or TM4&5 from the remaining four transmembrane helices within the same subunits. More specifically, the allosteric communication between
GSH/GSSG binding and conformation changes in NBDs cannot be explained using rigid body movement of individual subunits, but rather the accommodating conformational rearrangement through TM1&2, TM4&5 linked by ICL2 and the interaction association with a covalent association from TM6 in TMD to the NBDs. The transported substrates such as GSH or GSSG span between these TM helices in the transmembrane domain to create adaptable interfaces associated with the alteration between outward- and inward-facing conformations during the transport cycle.

Figure 2.5. Movement of NBDs. (A) Superimpose of the final snapshot of NBDs over the course of 200 ns MD simulation in the apo (red), GSSG (yellow) and GSH (green) system, respectively. (B) Movement of catalytically important residues E523 pairs in apo (red), GSSG (yellow) and GSH (green) system, respectively.

Movement of NBDs and intracellular loop

As a measure of the structure changes of NBDs, we chose the distance between Glu523 pairs. These residues are located at the end of the Walker B motif in the catalytic sites of two NBDs and assumed to activate the catalytic water molecule of hydrolysis at the active
site\textsuperscript{49}. A change in its distance corresponds to the opening or closing of the two NBDs. Figure 2.5B shows the departure distances of the catalytically significant glutamic acid E523 pairs over the simulated trajectory. The initial distance of the E523 pairs is 30 Å in the crystal structure. Over time, the MD simulation showed that the change in separation of the catalytic glutamic acid was significantly different among the three systems. The GSH system exhibited the largest change in its E523 pair distances with a final value of 22 Å, suggesting that the substrate GSH binding to the TMD domain causes significant instability in the NBD structure. In the apo system, although the E523 pair does move throughout the trajectory, they remain relatively constant at approximately 30 Å. The separation distance of E523 pairs was reduced fairly smoothly from 29 to 26 Å in the GSSG system. Taken together, the data showed that both the substrates GSH and GSSG induce a closing of the two NBDs during simulation (Figure 2.5). A closing of the two NBDs induced by substrates binding, consistent with the previous report\textsuperscript{53, 80, 119-120}. Some assumptions were made of the necessary sequence of events during ATP hydrolysis in the NBD: ATP binds to an open site subunit, two NBDs close, ATP hydrolysis, site opening and product release\textsuperscript{45}. The Glutamic acid E523 at the Walker B motif acts as a general base to active a water molecule for nucleophilic attack of the gamma phosphate ion, and thus plays a critical role in the ATPase activity. Therefore, it can safely be concluded that the decreases of the E523 pair distances of the GSH system are greater than that of an apo and GSSG system, which agrees well with experimental \(k_{\text{cat}}\) data.

As shown in Figure 2.1B, the ATP molecule is assumed to bind the NBD through mainly three attractive interactions in ABC transporters. The first is a pi-pi stacking contact
between the adenine ring of ATP and its nearby aromatic residue (A-loop). The second is an electrostatic interaction between the negative charges of the r-phosphate group of ATP and nearby positively charged residues (Walker A and H-loop). The third is through coordinating a Mg$^{2+}$ ion and nearby negatively charged residues (Q-loop and Walker B)$^{70}$. In the case of Atm1 ABC exporters, the aromatic residue near the ATP adenine ring is Y370 of the A loop. The distance between the center of the adenine ring and the side chain ring of Y370 is about 3.2Å. The G397, A398, G399, K400, S401, and T402 are located at Walker A and all have electrostatic interactions with the O atom of the phosphate group on the ATP molecule. It is noteworthy that the RMSF of Walker A from PC1 of GSH system is considerably greater than that for the apo system.

One crucial attribute of Atm1 family ABC exporter configurations was the accomplished resolution of a 24 amino acid residues C-terminus which had not been discerned before in any other ABC exporter crystal structure. In the crystal structure of NaAtm1 ABC transporter, the two long C-terminus helices tightly interact with each other to form a linking bridge involving the two subunits, thus stabilizing the NaAtm1 homodimer in an open, inward-facing conformation. Figure 2.6 showed the time evolution of unlocking of two C-terminus helices observed in the GSH system. It showed that the C-terminus helix in one subunit is parallel to its counterpart helix from the other subunit with a distance around 5.2 Å at 0ns. During the GSH simulation, the two α-helices unlock and slide along each other to allow the tight interaction between the Walker A motif and signature motif on the opposing NBDs. It is marked that during the Atm1 ABC exporter transport cycle, the two C-terminus α-helices either become more tilted or have to unlock and slide along
each other to allow the strong contact of the two NBDs\textsuperscript{86}. Considering that the ATP fragment was stuck between the Walker A and the signature motifs, the observation of unlocking of two C-helices in the GSH system but none in the apo or GSSG system is in fairly good agreement with experimental data.

Figure 2.6. Time evolution of unlocking and sliding of two C-Helices along each other observed in the GSH system.

GSSG results in higher binding affinity than GSH

To investigate the origin of the allosteric communication between substrate binding in TMDs and large conformation change in NBDs, we conducted binding site interaction analysis after GSH or GSSG binding. The per residue decomposition of MM/GBSA in the transmembrane domain of different substrate bound systems in the crystal structure showed that GSH and GSSG binds at different but overlapping sites within the TMs. Key interactions are formed between the amino acids Y156, F157, Y195, L265, L268, Q272, M317 and M320 and the free carboxyl and amino groups of GSH or GSSG. The main
binding site of GSH or GSSG substrates is located about 5 Å into the TMD toward the cytoplasmic surface, characterized by interactions from TM5-TM6 and their dimeric equivalent TM11-12. A discontinuity in the TM 6 covers residue 314 to 318 was found in the crystal structure. As shown in Figure 2.1C and Figure 2.8 A, the binding site of GSH in the crystal structure overlaps with the major binding site of GSSG. Those interactions including following: side chain of Q272 in TM5 make bidentate hydrogen bonds to the terminal N and O of GSH or GSSG, backbone N from G319 and M320 in TM6 forms strong hydrogen interactions with the free carboxyl groups of GSH or GSSG, side chain of Y156 forms a hydrogen bond to the a-carboxylate of the Gly at the other end of the GSH (GSSG). It is worth mentioning that another less frequently occupied GSSG binding site, is located about 5 Å from the major binding site from the cytoplasmic side (Figure 2.8A). Three arginine residues, R206, R210, and R323, are positioned toward GSSG to form hydrogen bonds in this minor binding site of GSSG in the crystal structure.
Figure 2.7. The per residue decomposition of MM/GBSA in transmembrane domain of different substrate bound systems

Table 2.2. Calculated MM/GBSA interaction energy (kcal/mol) for Atm1 ABC exporter substrates binding system. The standard error was included in parenthesis. Evdw is the Van Der Waals energy, Ecoul is the Coulombic energy, Egb is the polar solvation free energy. GSSG-a, GSSG-b represents the major and the second binding site for substrate GSSG.

<table>
<thead>
<tr>
<th>Binding site</th>
<th>Total IE</th>
<th>Evdw</th>
<th>Ecoul</th>
<th>Egb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSH</td>
<td>-32.12 (2.90)</td>
<td>-27.56 (2.86)</td>
<td>-8.73 (7.12)</td>
<td>4.16 (6.70)</td>
</tr>
<tr>
<td>GSSG-1</td>
<td>-47.85 (5.15)</td>
<td>-42.92 (4.24)</td>
<td>-6.32 (9.87)</td>
<td>1.39 (9.52)</td>
</tr>
<tr>
<td>GSSG-2</td>
<td>-34.53 (3.63)</td>
<td>-30.41 (3.40)</td>
<td>-9.19 (10.88)</td>
<td>5.07 (10.33)</td>
</tr>
</tbody>
</table>

Table 2.2 showed the calculated MM/GBSA interaction energy during the 200 ns MD simulation. The average interaction energy of the major and secondary binding site for
GSSG is -47.85 and -34.53 kcal/mol, which is stronger than the binding affinity of the GSH about -32.12 kcal/mol. To our surprise, the table showed that the Coulombic energy for the major GSSG, minor GSSG, and GSH binding system is -6.32, -9.19, and -8.73 kcal/mol with the standard error about 81.6%, 156.2% and 118.4% of itself. As shown in Table 2.2, the polar solvation energy Egb for the major GSSG, minor GSSG, and GSH binding system showed a larger standard error as well. To identify the possible contributing factors of such great variance of the Coulombic interaction and polar solvation energy, we explored the possible binding sites spread for GSSG and GSH on Atm1 ABC exporter during simulation. The trajectory of the center of mass for substrate GSSG and GSH during the 200 ns MD simulation were depicted in Figure 2.8B-F. As shown in the figures, the substrates GSSG have a limited conformational flexibility due to the rigid structure compared to the GSH. The GSSG in the primary binding site generates 139 different H-bond interactions with the protein while the minor binding site only owns 111 interactions through the simulation, which should make apparent that the binding affinity in the primary site is stronger compared to later. The orientation of GSSG and GSH are not sterically constrained by the system, and it rotates within the large transmembrane hole. Therefore, it is impossible to assign specific binding sites in the transmembrane hole to either substrate. However, they do adopt certain orientations with regard to the transmembrane pore once it can form two or more hydrogen bonds with different amino acids of the protein at the same time. These substrates can form a stable interaction and thus lead to a high occupancy. More detailed hydrogen bonding interactions between Atm1 ABC exporter and substrates are provided in the Appendix. Note that the hydrogen bond interactions to be considered here are those
in which the distances between heavy atoms are within 3.5 Å and within a 30° degree cutoff. We found that although GSSG and GSH primarily interact within the same general region of Atm1 ABC exporter, they interact with residues in different TMs within the transmembrane hole. The permeation pathways for GSSG and GSH were separate but implicated overlapping sets of key amino acids as the trajectory of the center of mass spreads. The center of mass for GSSG and GSH trajectories provide significant implications for us to understand the substrates uptake and transportation by Atm1 ABC exporter. We can safely conclude that the GSH and GSSG substrate with free carboxyl and amino groups were able to make extensive contacts with all kinds of residues of the TMDs because of the large hydrophilic cavity in the TMDs and the comparative twisting of the NBDs while they move toward each other. In the future, a range of crystallographic conformations with several substrate binding states, such as the presence of GSH or GSSG, should be obtained experimentally, that would elucidate the physical location of the substrate binding sites within the TMD of ABC exporter. The binding site interaction observed here could elucidate why GSH has greater conformational changes but with weaker binding affinity in comparison to the GSSG and thus provide significant insights into the structure-function relationship of ABC exporter.
Figure 2.8. Comparison of the binding site trajectory of GSSG (A-D) and GSH (E-F) in Atm1 ABC exporter. Two GSSG molecules are colored green and yellow for the major and minor binding site respectively. All figures are drawn in VMD program. (A) Detailed interactions between amino acids from protein and GSSG in two binding sites in the crystal structure (PDB access code 4MRS). Amino acids with key interactions were drawn in CPK by atom name. (B) Top view of the trajectory of the center of mass for two GSSG molecules
during 200 ns MD simulation. The permeation pathways of the center of mass of primary and secondary binding site GSSG molecules is colored green and yellow, respectively. TMDs of Atm1 ABC exporter are colored by segment in the NewCartoon. (C) Zoom in front view of permeation pathways of the center of mass trajectory of two GSSGs. (D) Zoom in top view of permeation pathways of the center of mass trajectory of two GSSGs. (E) Zoom in front view of permeation pathways of the center of mass trajectory of GSH. (F) Zoom in top view of permeation pathways of the center of mass trajectory of GSH.

CONCLUSION

Atm1 type ABC exporter with bound antioxidant GSH or GSSG is the first solved substrate-binding ABC transporter crystal structure. In this study, molecular dynamic simulations have been performed to understand substrates effects on the behavior of Atm1 type ABC exporter. The simulations showed that the GSH bound system undergoes the largest conformational changes. In a GSH bound system, the intracellular loops ICL1 and ICL2 between NBDs and transmembrane domains as well as a functional motif Walker A in the nucleotide binding domain associated with the biggest fluctuation in the first PCA component. The substrate GSH and GSSG with free carboxyl and amino groups were able to make extensive contacts with all kinds of residues of the TMDs due to the large hydrophilic cavity in TMDs and the comparative twisting of the NBDs while they move toward each. Therefore, both GSSG and GSH adopted a range of different conformations. The MD simulations successfully interpreted for several physical and dynamic processes from the experimental examinations. Several of the residues identified throughout
simulation involved in the hydrogen bond interaction with GSH/GSSG were assessed by site-directed mutagenesis on the ATPase activity. Simulations on the substitution of these residues will be discussed in next chapter.
CHAPTER THREE

EFFECT FROM SELECTED MUTATIONS IN THE LIGAND BINDING SITE FOR
ALLOSTERIC COMMUNICATION IN ATM1 TYPE ABC EXPORTER

ABSTRACT
The Y195F/Q272A double mutation is considered as the key residue mutation of the Atm1 type ABC transporter because it can significantly enhance ATPase activity ($k_{\text{cat}}$) in the ligand-binding site. In the present study, the allosteric communication to both of the wild-type and the Y195F/Q272A double mutation of the Atm1 type ABC transporter were investigated by molecular dynamics (MD) simulations and was compared to those of two single mutant Y195F and Q272A of Atm1 type ABC exporter in detail. In the 200 ns MD simulations, we observed that both double mutation Y195F/Q272A and single mutation Y195F exhibited large conformational flexibility using the RMSD analysis. A ‘semi-closed’ geometry detected in Y195F/Q272A MD trajectory that is closer than starting crystal structure suggested that the two NBDs approached each other. The single mutation Q272A failed to the link the TM4 and TM6, which in turn exhibits insensitive ATPase hyperactivity due to the loose contact with the nearby residues with no bulky sidechain. Finally, further PCA indicated that the two NBDs rotates in opposite directions, which resulted in the asymmetric dimerization of the NBD, this in turn leads to the allosteric communication between TMDs and NBDs.
INTRODUCTION

Any species that enters or exits a cell, whether an ion, small molecules or macromolecules, must penetrate one or more biological plasma membranes. To accommodate the variety of objects, numerous transporter proteins are encoded within the genomes of organisms. One of the largest classes of transporter proteins is the ATP binding cassette (ABC) transporter superfamily. ABC transporters use ATP as an energy source in the transport of substrates across cellular membranes. Many different ABC transporters are reported in individuals from all kingdoms of life; whereas exporters which excrите drugs, toxins, and lipids across membranes are found in both eukaryotes and prokaryotes. The overexpression of Atm1/ABCB7/HMT1/ABCB6 ortholog from *Novosphingobium aromaticivorans* DSM 12444, NaAtm1, in *Escherichia coli* confers protection against mercury and silver toxicity and presents with a physiological role in cellular detoxification processes.

Like other ABC exporters, NaAtm1 type ABC exporter has a characteristic architecture that contains two nucleotide binding domains (NBDs) that are located in the cytoplasm, and two transmembrane domains (TMDs) that are embedded in the membrane bilayer. The functional importance of NaAtm1 ABC exporter was approached by evaluating the effects of mutation on the ATPase activity. Among site-direct mutagenesis, the Y195F/Q272A double mutation is considered as the key residue mutation of the Atm1 type ABC transporter due to its nine-fold enhanced ATPase activity ($k_{cat}$) in the ligand binding site compared to the apo state. Single mutation Y195F also increased the catalytic efficiency by eight-fold, however, the catalytic efficiency is significantly decreased to one-eighth over
the single mutation Q272A. The crystallographic analysis of NaAtm1 type ABC exporter was solved at 2.4 Å resolution in 2014\(^{58}\). Crystallographic analysis showed that the sidechain of residue Q272 form bidentate hydrogen bonds to the backbone N and C atom of the γ-Glu group of Glutathione, with a distance of about 2.80 Å and 3.23 Å, respectively. The hydroxyl group of sidechain Y195 does not interact with any group of Glutathione directly but is about 5.3 Å away and pointed to the terminal carboxylate groups of Glutathione. Although considerable progress has been accomplished in the structural analysis of exporters from many families of ABC transporters\(^{57-60, 63, 86, 107-108, 122-123}\), much less is known about how the site-directed mutagenesis at these two particular positions altered the ATPase activity.

In the current study, the allosteric communication to both of the wild-type and the Y195F/Q272A double mutation of the Atm1 type ABC transporter were investigated by molecular dynamics (MD) simulations and was compared to those of two single mutations Y195F and Q272A of Atm1 type ABC exporter. We have performed MD simulations to examine the mutation effects on the behavior of Atm1 type ABC exporter in detail. First, both double mutation Y195F/Q272A and single mutation Y195F exhibited large conformational flexibility in RMSD analysis. Then, a ‘semi-closed’ geometry observed in Y195F/Q272A MD trajectory that is closer than starting crystal structure suggested that the two NBDs approached each other such that ATP molecules could be sandwiched between the walker A and the ABC signature motifs. Single mutation Q272A pre-shaped and distorted the geometry of the binding site and thus weaken the interactions between TMs and substrates. Asymmetric movement (\(\Delta K_{D1} > \Delta K_{D2}\)) showed that there is
allosteric communication between TMDs and NBDs. PCA analysis suggested that the two NBDs rotate in opposite directions, which resulted in the occurring of the asymmetric fashion of the NBD dimerization.

METHODS

Molecular Dynamic simulations

Pre-oriented protein coordinates with reference to the membrane normal of apo Atm1 type ABC exporter (PDB access code 4MRN) was taken from OPM database\textsuperscript{110} and used as the starting structures for all simulations. Engineered mutation residues MSE was reversed to MET. The homodimer protein was embedded into POPC lipid bilayer using CHARMM-GUI server\textsuperscript{112-113}. Mutants Y195F/Q272A, Y195F, and Q272A of Atm1 type ABC exporter were also made using the mutation tool in the CHARMM-GUI server. The lipid-protein systems were solvated via TIP3P water model and neutralized by adding 0.15 M NaCl counter ions. The ultimate system size of individually membrane embedded model is 111*111*169 Å\textsuperscript{3} and constituting 189,656 total number of atoms.

Molecular dynamic simulations were performed using NAMD 2.9\textsuperscript{114} with the CHARMM36\textsuperscript{115-116} force field. The SHAKE algorithm was applied to constrain the bonds containing hydrogens. A cutoff of 12 Å was used for van der Waals interactions. Particle-mesh Ewald summation\textsuperscript{117} was used in the periodic boundaries condition for the efficient calculation of long-range electrostatic interaction. Langevin dynamics was employed to maintain a constant temperature of 310K in each simulation. Constant pressure simulations were run at 1 atm using a Langevin piston technique with a decay period of 250 fs and a
damping timescale of 125 fs. An integration time step of 1 fs was used in order to avoid any substantial structural deformation during equilibration and production run. Coordinates were saved every 2 ps. The equilibrated structure was used for the production run in which no restraints were applied. A total of 200 ns production simulation were implemented for each structural analysis. VMD 1.9.1 has been used for visualization purposes. With the purpose of regulating whether the interactions between NBDs and the TMDs of Atm1 ABC exporter observed in the MD simulation were biased by the initial velocity of protein, three independent MD simulations were performed for each mutant systems.

Multiple Sequence Alignment

The first step was to obtain the fasta file for Sav 1866, NaAtm1, and ScAtm1 type ABC exporter sequences. These were acquired from the Protein Data Bank, with the access code 2HYD, 4MRN, and 4MYC, respectively. Then, chain A of the above three sequences was extracted and combined to a single dataset file. The sequences in the dataset files were then aligned using the multiple sequence alignment in Cluster Omega. In order to reduce the number of gaps within the multiple sequence alignment, parameters including the substitution matrix BLOSUM62 with a gap opening penalty of 20 and gap extension penalty of 6 were used. The sequence similarities and secondary structure information from the aligned sequence was then analyzed using the Espript 3 webserver.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) of the C-alpha atom coordinates in Atm1 ABC exporter was performed to recognize and characterize large-scale conformational changes.
PCA analysis is created on diagonalization of the variance-covariance matrix, obtained from the atomic motions of MD simulation trajectories. The eigenvectors correspond to the dominant or independent modes of atomic fluctuations, and the eigenvalues correspond to the contribution of each eigenvector to the global movements of the system. On the whole, the principal eigenvalues and the corresponding eigenvectors arrest the most meaningful global motions of the system. PCA of the C-alpha atom trajectories was achieved using the VMD program. N-terminal amino acids 8–20 and C-terminal amino acids 600–607 were not included in the PCA analysis. Fluctuations were referenced to the first frame of the trajectory.

To illustrate the correlation of the fluctuations between residues, the covariance of the C-alpha pairs displacement of a dynamic trajectory were calculated using the “coor cova” module in Charmm. The covariance values are within the scope of (-1,1), where negative numbers reveal the pairs of residues moving in the opposite direction along the eigenvector mode, positive numbers suggest the pair of residues moving in the same direction. The cross-correlation function is normalized so that the +1 value means a fully correlated motion between a residue pair. A cross-correlation of -1 suggests fully anti-correlated motion, and zero means uncorrelated fluctuations.

RESULTS AND DISCUSSION

Double mutation Y195F/Q272A and single mutation Y195F exhibited large conformational change in RMSD analysis.
The root mean square deviations (RMSDs) of the backbone atoms from the minimized structure were separately calculated for the whole protein, TMDs, and NBDs among the four systems studied. The results are shown in Figure 3.1. As shown in the RMSD profiles, our simulations are stable during the 200 ns MD simulations. The RMSDs of the whole protein are in the range of 1.0 - 4.8 Å during the 200 ns MD simulation. Both of the final whole-protein RMSD value of the double mutation Y195F/Q272A and single mutation Y195F exhibited a value of 3.8 Å, which are greater than that obtained for the apo system (approximately 2.7 Å) and single mutation Q272A system (approximately 2.8 Å). With a focus on the individual TMDs (red line in each plot), the RMSD values of Y195F system showed the largest value about 4.0 Å than that of a double mutation Y195F/Q272A (3.0 Å) and single mutation Q272A systems (2.9 Å). The lowest conformation flexibility of the NBDs backbone atoms RMSD has been observed in the apo system, with a constant around 2.0 Å. The rank of RMSD values of TMDs agrees well with the RMSD values of NBDs in the four simulated systems. As shown in the blue line of each plot, the RMSD values of Y195F system showed the largest value of about 3.9 Å than that of a double mutation Y195F/Q272A (3.6 Å) and single mutation Q272A systems (2.6 Å) in TMDs calculations. The apo system showed the lowest conformation flexibility of the TMDs backbone atoms RMSD, which is around 2.1 Å. In summary, RMSD values of double mutation Y195F/Q272A and single mutation Y195F exhibited the largest conformational flexibility during 200 ns simulations. The RMSD analysis showed a strong correlation with the experimental ATPase activity of selected mutations in the ligand binding site, where the
catalytic efficiency increased by nine-fold and eight-fold over the double mutation Y195F/Q272A and single mutation Y195F, respectively.

Figure 3.1. RMSDs for the backbone atoms of whole protein (blue), TMDs (red), and NBDs (blue) in wild-type (A), double mutation Y195F/Q272A (B), single mutation Y195F (C) and single mutation Q272A (D).
The sequence similarities and secondary structure information from aligned sequence
A structure-based search using BLAST for a homologous protein of NaAtm1 ABC exporter was initiated. The results showed that there is 45% sequence identity of NaAtm1 to the yeast mitochondrial ABC transporter Atm1 (ScAtm1, PDB access code 4MYC), which is much higher than that of the 32% sequence identity to Sav1866 (PDB access code 2HYD). Sav1866 is the first and well-characterized bacterial ABC exporter. The sequence similarities and secondary structural information from aligned sequence of those three ABC transporters were analyzed (Figure 3.2). When aligned with sequences corresponding to these three ABC transporters, substantial conservation of secondary structure elements in TMDs and significant functional motifs in NBDs was spotted. NaAtm1 type ABC exporter is characterized by a set of highly conserved functional motifs in the NBDs at the sequence level. On the contrary, the sequences and architectures of the TMDs of NaAtm1 type ABC exporter are variable, reflecting the chemical diversity of the transported substrates.
Figure 3.2. The sequence similarities and secondary structure information analysis from aligned sequence of Sav 1866, NaAtm1, and ScAtm1 type ABC exporter. Figure are generated using online program Espript3\textsuperscript{129-130}. Coils, arrows and character TT represent \( \alpha \) helical, \( \beta \) sheet and three turn secondary structures, respectively. The conserved sequence is labeled red and blocked, while identical residues are highlighted in red. Specifically, secondary structure assignment in TMD with TM1 to TM6 and functional motifs in NBDs of NaAtm1 are also highlighted in yellow and blue, respectively. Residues for double mutational analysis are shown by black arrows.
The conformation changes of the NBDs in the apo state

To characterize the conformations of NaAtm1 ABC exporter in the apo state, three key distances of NBDs were measured in the NBDs movement from three independent 200ns MD simulations. The first distance we chose (represented KD3 in Figure3A) is the \( \alpha-C \) distances between two E523 residues in two subunits of NaAtm1 ABC exporter. These residues are located at the end of the Walker B motif in the catalytic sites of two NBDs and assumed to active the catalytic water molecule of hydrolysis at the active site\(^{49}\). A change in E523 pair distance corresponds to the opening or closing of the two NBDs. Another two key distances are the \( \alpha-C \) distances between a pair of K400 in one subunit and S499 from opposite NBD (represented KD1 and KD2 in Figure3A). K400 is located at the tail of the Walker A motif and coordinates with the \( \gamma \)-phosphate group of ATP due to the positively charged sidechain \( \varepsilon \)-amino group. By contrary, S499 is positioned at the head of the ABC signature motif and binds ATP with the side chain hydroxyl group. Since the ATP fragment was stuck between the Walker A and the ABC signature motifs, the KD2 and KD3 form a functionally important catalytic cavity. Open, the inward-facing conformation of NaAtm1 ABC exporter was observed in the crystal structure. Three dashed lines demonstrate the original distance of KD1, KD2, and KD3 in the crystal structure 4MRN, which is 26.7 Å, 25.0 Å, and 31.3 Å, respectively. Figure 3.3B showed the histogram plots of the KD1, KD2, and KD3 movements from three independent 200 ns molecular dynamic simulations of the apo state of the NaAtm1 ABC exporter. It was revealed that the departure distances of the catalytically significant glutamic acid E523 pairs KD3 decreased slightly over the simulated trajectory with an observed value of 30.30 ±0.77 Å. However, a much larger
deviation from the crystal structure of KD1 and KD2, with an observed value of 24.65 ±1.44 Å and 24.73 ±1.10 Å, were found in the three independent simulations. Therefore, the two NBDs of NaAtm1 type ABC exporter would maintain full separation in a plasma membrane environment in the apo state. A ‘semi-closed’ geometry that is closer than starting crystal structure observed in MD simulation suggested that the two NBDs approached each other such that ATP molecules could be sandwiched between the Walker A and the ABC signature motifs. Interestingly, asymmetric changes of KD1 and KD2 movement have been found in the apo state. The asymmetric movement (ΔKD1 > ΔKD2) showed that there is allosteric communication between TMDs and NBDs such as the intracellular loops located on the TMD-NBD interfaces in the apo state NaAtm1 ABC exporter. In summary, we can safely conclude that symmetric movement events would occur at the alternate subunit if a large number or longer MD runs were conducted due to the structural equivalence of the two subunits and their stochastic nature in the APT hydrolysis process.
Figure 3.3. The structure of NBDs in apo state of NaAtm1 type ABC exporter. (A) Three key distances are identified in the NBDs movement. KD1 represents the $\alpha$-C distances between K400 in subunit1 and S499 in subunit 2. KD2 represents the $\alpha$-C distances between K400 in subunit2 and S499 in subunit 1. KD3 represents the $\alpha$-C distances between two E523 residues in two subunits. Functional motifs of NBDs are labeled by segment and colored in NewCartoon representation. The figure is generated in VMD
program. (B) Histogram plots of the three key distance movements from three independent 200 ns molecular dynamic simulations. Three dashed lines demonstrate the original distance of KD1, KD2, and KD3 in the crystal structure 4MRN.

Monitor large conformation changes of the NBDs in the mutation systems

Figure 3.4. The final MD structures of NBDs regions and KD3 distance in the NaAtm1 ABC transporter obtained from the apo state (A), Y195F/Q272A (B), Y195F (C), and Q272A (D) system. The NBDs are colored by segment in NewRibbons representation. The figure is generated in VMD program.
Figure 3.5. Asymmetric movement of KD1 and KD2 in MD snapshots in the NaAtm1 ABC transporter obtained from the apo state (A), Y195F/Q272A (B), Y195F (C), and Q272A (D) system.

We first investigated the structure and dynamics of the NBD regions using KD3 distance in NaAtm1 ABC transporter systems (Figure 3.4). Considering that the initial distance of the KD3 is 30 Å in the crystal structure, the change of separation of the catalytic glutamic acid was significantly different among mutation systems compared to the apo state over 200 ns simulation. In the apo system (Figure 3.4A), although KD3 does move throughout the trajectory, they remain relatively constant at approximately 30 Å. On the contrary, double mutation Y195F/Q272A system exhibited a large change in its E523 pair distances with a final value of 26.1 Å (Figure 3.4B), suggesting that the double mutation
Y195F/Q272A on the ligand binding site causes significant instability in the NBD structure. Single mutation Y195F system exhibited the largest change and deviation from its average 25.5 Å in KD3 distances with a final value of 26.1 Å (Figure 3.4C). The separation distance of E523 pairs was increased slightly to 31.6 Å in single mutation Q272A system (Figure 3.4D). Similar events have been found in both KD1 and KD2 values during the 200ns MD simulation in NaAtm1 ABC transporter systems (Figure 3.5). In the double mutation Y195F/Q272A system, KD1 and KD2 both shortened to 23.38±1.27 Å and 20.80±1.41 Å, respectively, which deviated from the distance of 26.7 Å and 25.0 Å observed in the crystal state. On the contrary, in the single mutation Q272A system, KD1 is observed to 25.16±1.27 Å and KD2 to 26.15±1.33 Å, which is slightly greater than the distances 25.0 Å observed in the crystal state. Taken together, the data showed that both the double mutation Y195F/Q272A and single mutation Y195F induce a closing of the two NBDs during simulation. A ‘semi-closed’ geometry that is closer than the starting crystal structure observed in both the double mutation Y195F/Q272A and single mutation Y195F MD simulation suggested that the two NBDs approached each other such that ATP molecules could be sandwiched between the Walker A and the ABC signature motifs. This semi-closed geometry which is similar to the canonical head-to-tail type of NBD dimer, could partially explain the cause of previously observed large RMSD fluctuations during MD simulations.

The residue E523 at the Walker B motif plays a critical role in the ATPase activity due to acting as a general base to activate a water molecule for nucleophilic attack the γ phosphate ion, as mentioned before. In ABC transporters, the ATP molecule is assumed to be stuck
between the Walker A and the signature motifs in NBDs and to bind NBD through the following three attractive interactions: (1) a pi-pi stacking contact between the adenine ring of ATP and its nearby aromatic residue (A-loop); (2) an electrostatic interaction between the negative charges of the \( \gamma \)-phosphate group of ATP and nearby positively charged residues (Walker A and H-loop); (3) coordination between the Mg\(^{2+} \) ion and nearby negatively charged residues(Q-loop and Walker B)\(^{70} \). In the case of NaAtm1 ABC exporters, the aromatic residue near the ATP adenine ring is Y370 from A loop. The distance between the center of the adenine ring and the sidechain ring of Y370 is about 3.2Å. The G397, A398, G399, K400, S401, and T402 are located at Walker A and all have electrostatic interactions with the O atom of the phosphate group of ATP molecule.

The unlocking of two C-helices was observed in both the double mutation Y195F/Q272A and single mutation Y195F system but none was observed in the apo or single mutation Q272A system. The double mutation Y195F/Q272A and single mutation Y195F are considered as the key residue mutation of the Atm1 type ABC transporter due to their nine-fold and eight-fold enhanced ATPase activity in the ligand binding site compared to the apo state, respectively. The catalytic efficiency is significantly decreased to one-eighth over the single mutation Q272A. In the crystal state, the sidechain of residue Q272 forms bidentate hydrogen bonds to the backbone N atom and carboxylate groups of the \( \gamma \)-Glutamic acid group of Glutathione, with a distance of about 2.80 Å and 3.23 Å, respectively. At the same time, Q272 is close to the sidechain of Y195 in TM4 that forms a strong hydrogen bond with the backbone of L315 in TM6. The residue Q272 in TM5 located in the ligand-binding site makes strong contact with the surrounding residues of
the TM4 and TM6, which would modify the motion of the TM4 and TM5 to TM6, which in turn lead to the closure of the NBDs through the intracellular loops at the TMD - NBDs interfaces. As a result, the single mutation Q272A which loses the bulky sidechain and exhibits loose contact with the nearby residues, which thus potentially fails to the link the TM4 and TM6, which in turn exhibits insensitive ATPase hyperactivity. Therefore, we concluded that the conformational changes characterized by movement of three key distances KD1, KD2 and KD3 values of both double mutation and single mutation systems is in fairly good agreement with experimental ATPase activity data.

Principal Component Analysis (PCA) of double mutation Y195F/Q272A system

Principal component analysis (PCA) were performed to understand the significant conformation changes of NBDs in double mutation Y195F/Q272A system. PCA outlines a set of eigenvectors derived from the variance-covariance matrix, which is obtained from the atomic motions of MD simulation trajectories. Figure 3.6 demonstrates the PC1s of the double mutation Y195F/Q272A system. It is showed that in the double mutation Y195F/Q272A system, the motion brings the two NBDs close to each other to form the “semi-closed” conformation. The two NBDs rotate in opposite directions, which resulted in the asymmetric dimerization of the NBD.
Figure 3.6. Side view (A) and bottom view (B) of the vector representation of PC1. The two NBDs rotates in clockwise and counter-clockwise directions.

CONCLUSION

In this study, we have performed MD simulations of the double mutation Y195F/Q272A and two single mutation Y195F and Q272A to examine the mutation effects on the behavior of Atm1 type ABC exporter in detail. RMSD analysis showed that both double mutation Y195F/Q272A and single mutation Y195F exhibited large conformational flexibility. A ‘semi-closed’ geometry observed in Y195F/Q272A MD trajectory that is closer than starting crystal structure suggested that two NBDs approached each other such that ATP molecules could be sandwiched between the walker A and the ABC signature
motifs. PCA showed that the two NBDs rotate in opposite directions, which resulted in the asymmetric dimerization of the NBD. This in turn leads to the allostERIC communication between TMDs and NBDs. The current study provides invaluable insights into the early stage of the antioxidant transporter cycle catalyzed by NaAtm1 ABC exporter.
CHAPTER FOUR

BINDING AFFINITY AND CATALYTIC ANALYSIS of UDG/DNA COMPLEXES

Sections of this work have been published as:


ABSTRACT

Questions including the physicochemical mechanisms motivating the evolution of TthUDGa are tackled through density function theory (DFT) methods, molecular dynamics (MD) simulations, and a successive analysis of thermodynamic properties associated with the enzyme activity using MD trajectories. The transition state had been successfully located from DFT frequency and IRC calculations based on the B3LYP/6-31+G(d) level of theory. TthUDGa bound reactant in a typical 3’-exo sugar ring conformation that favors oxacarbenium ion through nucleotide backbone distortion. An encouraged Spearman correlation ($r_s=0.786, r^2=0.554$) was achieved in comparison to the experimental data. These results suggest that the electrostatic interaction changes appear to be a major role in
enzyme efficiency resulting from mutation. A detailed analysis of the transition state conformation of the glycosidic bond scission, binding affinity between TthUDGa protein and flipping out of uridine, and catalytic activity of TthUDGa/DNA complexes suggests that the transition state stabilization might be contributing factors for the evolutionary optimization of TthUDGa/DNA complexes.

INTRODUCTION
Understanding how the enzyme works has a both fundamental and practical significance to human beings since most life processes are controlled by enzymatic reactions. One of the challenges in modern biophysics is how the particular roles of catalytic amino acids play a part in enzymatic reactions with the main origin of the massive catalytic power of enzymes. Enzyme reactions work by lowering activation energy ($\Delta G^\dagger$). However, since many factors could impact in binding affinity during the process when moving from the ground state to its transition state, the source of the enormous catalytic power of enzymes are still poorly understood. Among all the techniques developed to help to interpret the role catalytic amino acids play in enzyme reactions, site-directed mutation experiments are widely used. The catalytic groups may perfectly change the orientation in mutant variants and the entire enzyme activity can thus be enhanced or weaken after mutation. The catalytic efficiency ($k_{cat}/K_M$) of the wild-type and mutants of the enzyme are compared to elucidate the origin of the large effect of the catalytic power of enzymes in an active site environment. To explore the origin of the large effect of the catalytic power of enzymes and its mutants and to distinguish between different catalytic pathways, quantitative methods for
calculating the catalytic efficiency is indispensable. Developing and validating methods which can evaluate the potential energy surface and find the activation free energy between the connected reactant and product accurately are the key factors. Combined quantum mechanics/molecular mechanics (QM/MM) techniques are often used to characterize enzyme activity and changes due to mutation. Quantum mechanics (QM) seeks the approximate solution to the Schrödinger equation, which is limited to small systems since it is an electronic structure method and computationally expensive. Molecular mechanics (MM) methods do not explicitly treat the electrons and thus properties like bond forming and breaking in a chemical reaction that count on the features of the electronic structure could not be modeled. QM/MM methods, a concept introduced by Warshel and Levitt from 1976\textsuperscript{131}, enable the modeling of reactive enzyme systems at reasonable computational detail providing the necessary accuracy as well. The QM/MM approach found extensive recognition beginning with a report by Karplus in 1990\textsuperscript{132}. Over the last two decades, numerous reviews\textsuperscript{133} and articles\textsuperscript{134} of the QM/MM approach have been developed with an emphasis on enzymatic catalysis.

DNA mutation and damage are the causes of many diseases such as cancer, which occur either naturally or induced by contacting toxic materials including organic dyes\textsuperscript{135-140}, DNA binders\textsuperscript{141-142}, and heavy metals, etc. Deamination of cytosine to uracil is one of the major events of DNA damage\textsuperscript{143-145}. Structural damaged DNA molecule can alter or eliminate the cell’s ability to replicate and transcribe the gene and therefore need to be repaired. When DNA repair processes fail, irreparable DNA damage may occur, which eventually lead to malignant tumors or cancer\textsuperscript{146}.
Repair of damaged uracil DNA is achieved in a base excision repair pathway initiated by a Uracil DNA glycosylase (UDG, E.C. 3.2.2.27)\textsuperscript{134, 147-150}. UDG is a hydrolytic deamination repair enzyme (Figure 4.1), of which six families have so far been identified\textsuperscript{134, 144-145, 151-153}. The research object here is a family 4 UDGa from \textit{Thermus thermophiles} HB8 (TthUDGa,) which crystal structure was solved with product uracil base at 1.5 Å resolution in 2003\textsuperscript{153}. To the best of our knowledge, the catalytic and physio-chemical properties of the catalytic domain of TthUDGa have not been reported, issues concerning the binding affinity affect catalysis activity in TthUDGAA/DNA complexes still remain to be further clarified. Our goal in this study is to understand what the actual catalytic amino acids are, how this different binding between the wild-type and its mutants is accomplished, and potentially be able to predict the catalytic activities of mutant variants of TthUDGa from a molecular computational simulation view. In the present study, questions involving the physicochemical mechanisms motivating the evolution of TthUDGa are addressed through density function theory (DFT) methods, molecular dynamics (MD) simulations, and the following analysis of thermodynamic properties associated with the activity of the enzyme using MD trajectories. Especially, first, a DFT calculation based on the B3LYP/6-31+G(d) level was used to evaluate the potential energy surface and locate the correct transition state (TS) conformation between the connected reactant and product. Then, classical MD simulation and molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) analysis were applied to investigate the impact of single amino acid mutations on the binding affinity and catalytic activity toward the TthUDGa/DNA complexes. Last but not the least, the kinetic properties and structural models were characterized to demonstrate
how the different binding between wild-type and its mutants is accomplished and possible pathways behind the N-glycosidic bond cleavage have been proposed.

**METHODS**

A summary of the approach taken in this study is as follows. Questions involving the physicochemical mechanisms underlying the evolution of Tth UDGa are addressed through density function theory (DFT) methods, molecular dynamics (MD) simulations, and a subsequent analysis of thermodynamic properties related to the activity of the enzyme using MD trajectories.

Transition state analysis in Gaussian

TS analysis provides a detailed view of the mechanism of an enzyme reaction. In a hybrid QM/MM model, a small part of the system that undergo glycosidic bond hydrolysis in this study is treated with electronic process at a QM level, while the rest (e.g., protein and solvent) is treated by highly efficient, force field based MM methods. The nucleophilic substitution reaction involving in part QM DFT methods was performed in the Gaussian 09\textsuperscript{154} and Gaussview 5.0\textsuperscript{155} package. All Gaussian 09 calculations including geometry optimizations (including transition state), frequency calculations, electrostatic potential calculations, and intrinsic reaction coordinate (IRC) calculations were performed at the B3LYP/6-31+G(d) level of theory. Optimization to a transition state using the Berny optimization method were performed with increasing the C1’-N1 distance in 0.005 Å increment from 1.4 Å up to 3.0 Å and the dihedral of O4’-C1’-N1-C2 in 5° increment from 150° to -150°. The transition state structure was confirmed by the presence of single
imaginary frequency in vibration analysis and the corresponding mode motion is the stretching of the C-N glycosidic bond associated with bond cleavage. The B3LYP/6-31+G(d) optimized reactant and transition state complex were used to generate electrostatic potential maps, which calculates the interaction between every part of the electron density cloud and a +1 probe charge. The electrostatic potential was mapped with an isodensity surface of 0.0004 electron/Å³ which then colored according to their potential with the regions of positive, neutral and negative electrostatic potential shown in blue, green, and red, respectively. The minimum energy reaction pathway in mass weighted Cartesian coordinates between the TS of the glycosidic bond cleavage and its reactant and product was examined to observe the conformation change during the reaction. The QM/MM linking atoms were replaced with hydrogen atoms for more accurate calculations and the detailed QM/MM atoms distribution are presented later.

Molecular modeling and mutant making

The crystal structure of Tth UDGα and product complex was acquired from the Protein Data Bank (accession code 1UI0), and used as a model for computational analysis later. The structure of substrates flipped-out DNA was extracted from the crystal structure of human UDG-DNA complex (PDB accession code 1EMH) using the Swiss-PDB Viewer (SPDBV) program. Mutants E41A, E41Q, G42D, E41Q/G42D, E47A, F54A, N80A, N89A, and H155A were also made with the mutation tool in Swiss-PDB Viewer program and the “best rotamer” was chosen with the lowest clash score. The activation energy of reaction in wild-type and mutants systems from single trajectory MMPBSA analysis are
compared. Their relative magnitude is reliable and allows to draw qualitative conclusions about the proposed pathways.

Molecular dynamics simulations

After building the initial complex structures, an explicit solvent system using the TIP3P water model was constructed in the CHARMM c35b6 molecular mechanics package\textsuperscript{116} using a suitably sized box. The minimum distance between the box boundary and any of the atoms of the solvated TthUDGa/DNA complex was maintained to at least 9 Å. Sodium chloride ions were added to the system to achieve an electrical neutral system. The CHARMM 27 all hydrogen force field for proteins\textsuperscript{115} and nucleic acids\textsuperscript{157} was used. Particle-mesh Ewald summation\textsuperscript{117} was applied in the periodic boundaries condition for the efficient calculation of long-range electrostatic interaction. Energy minimization was performed by using 4000 steepest descent steps followed by adopted basis Newton-Raphson (ABNR) method with the harmonic constraints from 10 to 1 kcal/(mol•Å\textsuperscript{2}) in decrements of 3 kcal/(mol•Å\textsuperscript{2}) every 1000 steps to remove any unfavorable van der Waals clashes while minimally perturbing the original model x-ray structure. Using a Langevin barostat\textsuperscript{158}, an isothermal-isobaric ensemble (NPT) was constructed in the NAMD program\textsuperscript{114} and the system heated gradually from 100K to 300K over a period of 400ps. An integration time step of 1fs was used so as to avoid any noteworthy structural deformation during equilibration and production run. Coordinates were saved every 2ps. A 2 ns equilibration and 3 ns production simulation were implemented for each structural analysis. VMD 1.9.1\textsuperscript{118} had been used for visualization purposes.

MM-PBSA interaction energy calculation
Molecular mechanics Poisson-Boltzmann solvent accessible surface area (MM-PBSA) approach\textsuperscript{12,159-160}, \( G = < G_{MM} > + < G_{PB} > + < G_{SA} > \), was used to calculate binding free energy difference of each constructed Tth UDGa/DNA complexes. The molecular mechanics energy, \( G_{MM} \), used an infinite cutoff to evaluate the non-bonded interactions. Poisson-Boltzmann polar solvation interaction energies, \( G_{PB} \), were solved with CHARMM using the PBEQ module, which allows the setup of a successive over relaxation method to solve the Poisson-Boltzmann equation on a discretized grid for the complexes. Snapshots for MM/PBSA analysis were extracted every 10 ps from the 3 ns production phase MD trajectories. The dielectric constant for bulk solvent and interior protein were set to 80.0 and 4.0, respectively. The non-polar solvation energy, \( G_{SA} \), was approximated with \( G_{SA} = \gamma(SASA) + \beta \), where \( SASA \) was solvent-accessible surface area, \( \gamma = 0.00542 \text{ kcal/molÅ}^2 \) and \( \beta = 0.92 \text{ kcal/mol} \). Solvent accessible surface with a 1.4 Å solvent probe radius was constructed for the solvent-solute dielectric boundary. Non-entropic contribution of MM-PBSA interaction energies change between each amino acid of the enzyme and flipping out nucleotide after binding, \( \Delta G = G_{complex} - G_{enzyme} - G_{sub} \), were also evaluated.

**QM/MM charge scaling of transition state model**

The catalytic power of enzymes which lower activation energy (\( \Delta G^i \)) is almost exclusively due to electrostatic effects\textsuperscript{89}. Three different charge scaling approaches were implemented to model the transition state of Tth UDGa/DNA complexes. Specifically, the partial atomic charges of the ground state in Charmm had been scaled in three different ways, then the
resulting partial atomic charges were assigned to original atoms in the MM/PBSA
calculation of the transition state.

\[
\text{Charmm}\_\text{TS1} = \text{Charmm}\_\text{G} \frac{\text{Esp}_{\text{DFT}}\_\text{TS}}{\text{Esp}_{\text{DFT}}\_\text{G}} \quad (4.1)
\]

\[
\text{Charmm}\_\text{TS2} = \text{Charmm}\_\text{G}^* \frac{\text{Mulliken}\_\text{TS}}{\text{Mulliken}\_\text{G}} \quad (4.2)
\]

In method 1, the differences of electrostatic potential (ESP) derived atomic charges from
DFT at B3LYP/6-31+G(d) level of theory between the optimized transition state and the
ground state was scaled according to Equation 4.1. In method 2, Mulliken charges from
the reference\textsuperscript{134} were scaled using Equation 4.2 and applied to the Charmm partial atomic
charge for the purpose of the simulating transition state. In method 3, based on the initial
guess of the electronegative properties of glycosidic bond cleavage, partial charges of four
atoms around the glycosidic bond had been changed while the remaining Charmm partial
charges were kept unchanged in the transition state. The detailed Charmm partial atomic
charges from ground state and transition state were list in Table 4.1. Two binding free
energies of Tth UDGa/DNA complexes (\(\Delta G\_\text{TS}\) and \(\Delta G\_\text{G}\)) based on different partial atomic
charges was compared from single trajectory MMPBSA analysis. Differences between
them were then estimated according to\(\Delta \Delta G = \Delta G\_\text{TS} - \Delta G\_\text{G}\), which can be correlated to enzyme efficiency.

**Molecular modeling of other families**

Molecular modeling of family 5 UDGb, family 2 mismatch-specific UDG (MUG) and
family 7 UDG were conducted as follows: the crystal structure of family 5 Tth UDGb was
acquired from the RCSB Protein Data Bank (accession code 2DEM), and used as a model
for subsequent computational analysis. The initial structures of uridine, inosine, xanthosine, and oxanosine nucleotides bound to the family 5 UDG enzyme were obtained by manually modifying the structure of 2-deoxy-5-phosphono-ribose extracted from the crystal structure of Tth UDGb using the Swiss-PDB Viewer program. Specifically, the 2-deoxy-5-phosphono-ribose was aligned to a cytidine, and the cytosine was changed to uracil, hypoxanthine, xanthine, and oxanine, respectively. The crystal structure of family 2 (E. coli MUG) and family 1 (E. coli UNG) UDG was acquired from the RCSB Protein Data Bank (accession codes 1MUG and 2EUG). The crystal structure of the human UDG–DNA complex (PDB accession code 1EMH) was used as the DNA model to build the flipped out double-stranded DNA complexed with the protein MUG. The single amino acid mutants MUG-K68N, UNGN123A and UDGb-A111N were also made using the mutation tool in the Swiss-Pdb Viewer program. The crystal structure of family 7 UDG was acquired from the RCSB Protein Data Bank (accession code 3IKB). Structure of substrates DNA d(ATGTTGCBTTAGTCC), where the B was the base to be flipped out of the helix, was extracted from the crystal structure of family 5 Tth UDGb-DNA complex (PDB accession code 2DEM). The base complementary to Uridine, Inosine and Xanthosine were systematically modified to guanine, thymine, and cytosine, respectively. Mutants P84N of family 7 UDG was also made using the mutation tool in Swiss-Pdb Viewer program and the “best rotamer” was chosen with the lowest clash score. After building the initial complex structures for family 5, family 2 mismatch-specific UDG (MUG) and family 7 UDG, an explicit solvent system using the TIP3P water model was constructed in the
CHARMM c35b6 molecular mechanics package. A detailed description of the methods of molecular dynamics simulation can be found in the family 4 Tth UDGa/DNA complex.

RESULTS AND DISCUSSION

Transition state conformation

TS conformation analysis provide the detailed knowledge of enzymatic transition state in catalysis. Figure 4.1 demonstrates the hydrolysis process of damaged DNA under the catalysis of UDG. A conventional sketch of the split of the systems into QM and MM atoms was also shown in the substrate (highlight in the blue curve is QM region in Figure 4.1). Essentially, the QM region consists of the base uracil, the ribose and two linked H atoms attached to O5’ and O3’ terminal. Amino acids of enzyme were assigned to MM region. The 5’-phosphate group was replaced by linked H atom since the effect of neglecting the 5’-phosphate in calculation was negligible. Generally, DN*AN and ANDN are two major hydrolytic mechanisms utilized by glycoside hydrolases for recognizing, binding substrates and cleaving a specific glycosidic bond. In DN*AN mechanism, a discrete oxacarbenium ion intermediate before water attack at the anomeric carbon was formed and the formation of oxacarbenium ion intermediate is the rate determining step in enzyme reaction.

From changing conformation properties in frequency calculation using the B3LYP function and 6-31+G(d) basis sets, the transition state during N-glycosidic bond cleavage was confirmed by a single imaginary vibrational frequency of 153.4i cm⁻¹ indicating of C-N glycosidic bond stretching. In addition, the corresponding motion of this unique
imaginary frequency is the C1’-N1 bond breaking and reforming in GaussView (movie not shown).

Figure 4.1. Hydrolysis of damaged DNA under the catalysis of UDG. A general sketch of the division of the systems into QM and MM atoms was also shown (highlight in the blue curve is QM region). Essentially, the QM region consists of the base uracil, the ribose and two linked H atoms attached to O5’ and O3’ terminal. Amino acids of enzyme were assigned to MM region.
Figure 4.2. Properties of transition state and reactants along IRC at B3LYP/6-31+G(d) level drawn by GaussView. A. Total energy; B. RMS gradient norm of total energy; C. C1’-N1 bond distance; D. O4’-C1’ distance; E. Dihedral of C5’-C4’-C3’-O3’; F: Dihedral of O4’-C1’-N1-C2.

IRC results are consistent with the frequency calculation, that is, complete dissociation of the N-glycosidic bond at the transition state of TthUDGa/DNA complexes. The ionization of the leaving group, ribose conformation change and the geometry of the reaction coordinate, are depicted in the IRC (Figure 4.2 A to F). A maximum of the potential energy function was found (Figure 4.2A) and its RMS gradient norm is near 0 (Figure 4.2B) along IRC. The primary C1’-N1 distance was the most useful probe for determining the N-glycosidic bond cleavage. While the reaction moves towards the transition state from the reactant direction, the C1’-N1 bond distance between the ribose and uracil leaving group appears increase from 1.47 Å to 2.78 Å (Figure 4.2C). At the same time, C1’-O4’ bond
distance appears decrease from 1.42 Å to 1.27 Å and reached a local minimum of 1.27 Å at the transition state (Figure 4.2D), which indicated a sp2-sp2 C-O hybridization at the anomeric carbon. Similar to other enzyme systems like human 5’-Methylthioadenosine phosphorylase\textsuperscript{170}, the change in hybridization from sp3 to sp2 on anomeric carbon partially neutralize the positive charge itself, created a partial empty 2p\textsubscript{z} orbital which hyperconjugates with the lone pair of O4’ and \(\sigma(C2’-H2’)\) electrons, and thus stabilized the transition state. The dihedral analysis also support the movement of base ring away from the ribose. Analysis of the dihedral of C5’-C4’-C3’-O3’ (Figure 4.2E) and O4’-C1’-N1-C2 (Figure 4.2F) revealed that ribose ring has been distorted to an oxacarbenium ion-like conformation. In undistorted nucleotide, the \(\delta\) dihedral (C5’-C4’-C3’-O3’) is a usual 2’-endo sugar ring conformation with angle \(\delta=132^\circ\pm8^\circ\). The computational TS had a local maximum \(\delta\) dihedral around 147°, which was a typical 3’-exo sugar ring conformer. Most enzymatic reactions went through transition states in the 3’-exo conformation and the catalytic acceleration initiates from formation of a 3’-exo at the transition state, since only the 3’-exo sugar ring conformer is stable computationally\textsuperscript{143}. Similar to the transition state of another DNA repair enzyme MutY\textsuperscript{161}, UDG enzyme bound substrate in a conformation that favors oxacarbenium ion through nucleotide backbone distortion.

Since the catalytic power of enzymes which lower activation energy (\(\Delta G^i\)) is almost exclusively due to electrostatic effects\textsuperscript{89}, the electrostatic effects of the transition state and ground state had been analyzed. ESP charges on O4’ of transition state increased by +0.17 and negative charges on N1 and O2 of transition state decreased by 0.37 and 0.14
comparative to the reactant, respectively (Table 4.1). This sharing of charges was distinctive from oxacarbenium ion, which cause in enlarged cationic character at the transition state in nucleophilic substitution reaction of UDG. To more clearly visualize the electronic properties of the optimized reactant and TS complex, electrostatic potential maps at the B3LYP/6-31+G(d) level (Figure 4.3) were calculated. The electrostatic potential map of the optimized transition state (Figure 4.3B) showed slightly more electron density on the N1 position than the optimized reactant (Figure 4.3A). This result suggests that the transition state is highly active and needed to be stabilized via delocalization of the negative charge on N1 through partial glycosidic bond cleavage.
Table 4.1. Summary of Charmm partial atomic charges assigned over atoms in three different charge scaling method. (GS: ground state; TS: transition state; NC: no change, the same Charmm partial atomic charge assigned as in the ground state).

<table>
<thead>
<tr>
<th>Atom</th>
<th>Methods 1 Esp_DFT</th>
<th>Methods 1 Esp_DFT</th>
<th>Methods 2 Mulliken</th>
<th>Methods 2 Mulliken</th>
<th>Methods 3 Charmm</th>
<th>Methods 3 Charmm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GS</td>
<td>TS</td>
<td>GS</td>
<td>TS1</td>
<td>GS</td>
<td>TS1</td>
</tr>
<tr>
<td>C4'</td>
<td></td>
<td></td>
<td>0.09</td>
<td>0.11</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>H4'</td>
<td>0.17</td>
<td>0.17</td>
<td>0.09</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O4'</td>
<td>-0.46</td>
<td>-0.29</td>
<td>-0.5</td>
<td>-0.31</td>
<td>-0.31</td>
<td>-0.22</td>
</tr>
<tr>
<td>C1'</td>
<td>0.49</td>
<td>0.43</td>
<td>0.16</td>
<td>0.14</td>
<td>0.21</td>
<td>0.34</td>
</tr>
<tr>
<td>H1'</td>
<td>0.07</td>
<td>0.15</td>
<td>0.09</td>
<td>0.2</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>N1</td>
<td>-0.34</td>
<td>-0.71</td>
<td>-0.34</td>
<td>-0.71</td>
<td>-0.29</td>
<td>-0.41</td>
</tr>
<tr>
<td>C6</td>
<td>0.16</td>
<td>0.29</td>
<td>0.2</td>
<td>0.37</td>
<td>0.1</td>
<td>0.07</td>
</tr>
<tr>
<td>H6</td>
<td>0.15</td>
<td>0.08</td>
<td>0.14</td>
<td>0.08</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>C2</td>
<td>0.79</td>
<td>0.99</td>
<td>0.55</td>
<td>0.69</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>O2</td>
<td>-0.61</td>
<td>-0.75</td>
<td>-0.45</td>
<td>-0.56</td>
<td>-0.46</td>
<td>-0.52</td>
</tr>
<tr>
<td>N3</td>
<td>-0.67</td>
<td>-0.79</td>
<td>-0.46</td>
<td>-0.54</td>
<td>-0.32</td>
<td>-0.32</td>
</tr>
<tr>
<td>H3</td>
<td>0.37</td>
<td>0.38</td>
<td>0.36</td>
<td>0.37</td>
<td>0.3</td>
<td>0.29</td>
</tr>
<tr>
<td>C4</td>
<td>0.86</td>
<td>0.94</td>
<td>0.53</td>
<td>0.58</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td>O4</td>
<td>-0.6</td>
<td>-0.67</td>
<td>-0.48</td>
<td>-0.53</td>
<td>-0.56</td>
<td>-0.6</td>
</tr>
<tr>
<td>C5</td>
<td>-0.58</td>
<td>-0.69</td>
<td>-0.15</td>
<td>-0.18</td>
<td>-0.35</td>
<td>-0.39</td>
</tr>
<tr>
<td>H5</td>
<td>0.22</td>
<td>0.21</td>
<td>0.1</td>
<td>0.1</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>C2'</td>
<td>-0.29</td>
<td>-0.11</td>
<td>-0.18</td>
<td>-0.07</td>
<td>-0.32</td>
<td>-0.34</td>
</tr>
<tr>
<td>H2'</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>0.11</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>H2''</td>
<td>0.1</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.1</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Figure 4.3. Electrostatic potential maps of optimized reactant (A) and transition state (B) drawn by GaussView for the cleavage of glycosidic bond at B3LYP/6-31+G(d) level. Top: stick diagram; bottom: electrostatic potential surface maps, blue and red regions represent positive and negative potentials, respectively. This result suggested that the transition state
is highly active and need to be stabilized via delocalization of the negative charge on N1 through partial glycosidic bond cleavage. (C): A partial empty 2p\textsubscript{z} orbital which hyperconjugates with the lone pair of O4’ and \( \sigma \) (C2’-H2’) electrons in the LUMO of transition state shown in Gaussview. This result showed that the hybridization of the anomeric carbon had been changed from sp\textsubscript{3} to sp\textsubscript{2}, which thus stabilized the transition state in enzyme reaction.

Binding affinity of Tth UDGa/DNA complexes from MM/PBSA calculations

Being a monomer consisting of 192 amino acid residues, monofunctional Tth UDGa has a single domain \( \alpha/\beta \) fold in a common scaffold with compact three \( \alpha/\beta/\alpha \) sandwich (Figure 4.4A). In the crystal structure of Tth UDGa (PDB accession code: 1UI0), seven \( \alpha \)-helices sandwiched a single layer \( \beta \)-sheets of Tth UDGa, which is composed of two antiparallel \( \beta \)-sheets and four parallel \( \beta \)-sheets.

The crystal structure of human UDG-DNA complex was sourced as the DNA model to build the flipping out double DNA strand into the protein Tth UDGa. Single uracil base which bound to the active site groove on the C-terminal side in crystal structure of Tth UDGa was first aligned to the deoxyuridine in double DNA strand extract from the human UDG-DNA complex. Then deoxyuridine in the flipping out DNA strand had been modified to uridine manually. In the built model, the uracil ring was in a way similar to the human UDG-DNA complex\textsuperscript{148}, where the base uracil was rotated \( \sim 90^\circ \) on its C1’-N1 axis to a position almost halfway between syn and anti which causes a steric clash between the deoxyribose O4’ and the uracil C6 hydrogen, and thus could significantly stretched the
glycosidic bond which pulls the uracil deep into the enzyme active-site pocket \(^{149}\) (Figure 4.4A).

Figure 4.4. Visualization of active site interaction between family 4 Tth UDGa and double chain DNA in VMD program. A: Model built: wild-type family 4 Tth UDGa structure (colored by secondary structure in NewCartoon) with aligned-uracil-containing double strand DNA. Uracil is colored in licorice and DNA is colored in NewRibbons; B: Close-up view of wild-type uracil and catalytic amino acids in active site pocket of minimized structure. Hydrogen bond interactions between amino acids and uracil are shown as black dashed lines. Uracil is using the same coloring scheme described in (A). E41, G42, F54, N80, and H155 are colored by name in CPK model represents.

Figure 4.5A shows the per-residue decomposition of the MM/PBSA binding free energy of wild-type Tth UDGa/DNA complexes in the ground state. Most of the amino acids had
no contribution to the binding affinity except four chunks of amino acid groups concentrated around the active site groove showing obvious favorable interactions with substrate uridine: motif 1 (residue from 40 to 44), motif 2 (residue from 155 to 162), turn 1 (residue 53 to 58) and turn 2 (residue 80 to 93). A conserved histidine residue H155 in motif 2, which is located in an analogous position to that of H268 in human UDGs, had a binding free energy of -1.70 kcal/mol. This favorable free energy change of H155 came from the interaction with the O2 of uracil (Figure 4.4B) and appeared to play an analogous role in substrate binding. An asparagine residue N80 in turn 2 (N204 in human UDGs) had a binding free energy of -1.16 kcal/mol since it made two parallel H bonds to N3 and O4 of uracil, respectively. In turn 1, at least two factors can contribute to the -4.69 kcal/mol binding free energy of F54 (F158 in human UDGs). First, the benzene ring of F54 stacked with the pyrimidine ring of uracil, which helped orienting the uracil ring through the Pi-Pi stacking effect. Furthermore, a strong H-bond was found between backbone N in F54 and O4 of uracil to increase the accuracy of the recognition in glycosidic bond cleavage. In motif 1, two more H bonds were also made between backbone N of E41 (Q144 in human UDGs) and backbone of G42 (Q145 in human UDGs) with the O2 carbonyl group of uracil, respectively. The binding free energy of E41 and G42 in wild-type Tth UDGA/DNA complexes were -2.01 and -2.19 kcal/mol, respectively. K57 and R161 also had large favorable interactions since they were basic amino acids and within 8 Å of VDW radius of the substrate uridine in wild-type Tth UDGA/DNA complexes. These simulation results agreed well with the crystal structure results, which suggested that despite very weak amino acid sequence homology with human UDGs (family 1) and E.coli MUG (family 2), Tth
UDGa specifically recognized uracil prefer a manner as human UDGs (family 1) than E.coli MUG (family 2).
Figure 4.5. Per-residue decomposition of the MM/PBSA binding affinity of wild-type Tth UDGa/DNA complexes for substrate uridine in the ground state (A) and transition state (B) and three different binding free energy change (C, D and E) from ground state to transition state in Charmm, where negative values indicate residues that contribute favorable interaction, positive values indicate residues that contribute unfavorable interactions, zero values indicate residues not involved in enzyme catalysis. Four groups of amino acid groups demonstrate obvious favorable interactions with substrate uridine, motif 1 (residue from 40 to 44), motif 2 (residue from 155 to 162), turn 1 (residue 53 to 58) and turn 2 (residue 80 to 93). The greatest contribution of catalytic affinity was from His155. Three MM/PBSA binding free energy of transition state use all the same parameters besides the charges which were rescaled different in Charmm. A: Ground state binding free energy; B: transition state binding free energy of ESP charge scaling method. C: Binding free energy change from ESP charge scalar methods (method 1). D: Binding free energy change from Mulliken charge scaling methods (method 2). E: Binding free energy change from Charmm partial charge initial guess charge scaling (method 3).

Catalytic analysis of Tth UDGa/DNA complexes

However, the catalytic mechanism of Tth UDGa/DNA complexes is not as clear as the binding process. In order to investigate the contribution of residues to the catalytic affinity around uracil base in the active site, two kinds of techniques were employed. First, three different Charmm partial atomic charge scalar methods were compared to model the transition state binding free energy from single trajectory MMPBSA analysis. The active energy of reaction was estimated in each method according to the differences between
ground state and transition state: \( \Delta \Delta G = \Delta G_{TS} - \Delta G_G \). Second, mutants E41A, E47A, F54A, N80A and H155A which mimicked the secondary structure preference in catalytic had been made from homology modeling on wild-type Tth UDGa/DNA complexes model. Special amino acids had been mutant to Alanine since alanine scanning is a simple and widely used technique to determine whether the side chain of a specific amino acid plays a significant role to the function of given enzyme. How well can the transition state geometric and electrostatic potential from DFT discussed above be used in enzyme reaction? The charge assigned over atoms was a key concept in learning many types of chemical reactions. The MM/PBSA binding free energy of transition state and ground state for Tth UDGa/DNA complexes using all the same parameters except the charges which were recalculated. We assume that the magnitude of electrostatic potential charge for both the transition state and ground state was less important than the percentage changes. We also assume that no great difference for the partial charge percentage changes between transition state and ground state whether the model was based on quantum mechanics or classic mechanics level. Therefore, the partial charges of ground state in Charmm were scaled by the same extent of change in DFT calculation, then the resulted partial charges were assigned to atoms in the MM/PBSA calculation of the transition state (Table 4.1).

Figure 4.5B shows the per-residue decomposition of the MM/PBSA binding affinity of wild-type Tth UDGa/DNA complexes in the transition state from the ESP charge which had been identified from DFT at B3LYP 6-31+G(d) level. The major contribution to the binding affinity of transition state came from the same motifs and turns as in the ground state (Figure 4.5A). Figure 4.5C to E demonstrates binding energy changes from three
different scaling charge methods. The trend of contributions each amino acid played in the catalytic reaction were generally quite similar among these three charge scaling calculations. Four amino acids in the active site were characterized as the major contributors to the catalytic affinity of Tth UDGa/DNA complexes (Figure 4.5C to E), that are, Glu41, Gly42, Glu47, and His155. The greatest contribution of catalytic affinity was from His155, which binding free energy increased by -0.19, -0.19 and -0.91 kcal/mol from the ground state to the transition state in ESP charge scaling, Mulliken charge scaling and initial guess method, respectively. This favorable free energy change of H155 came from a strong H-bond interaction with the O2 of uracil (Table 4.2) and played a dominant role in the base leaving group leaving. As in other family UDGs, the glycosidic bond was hydrolyzed by at least two factors: either by using strained energy acquired by the specific recognition and scission of uracil base or by precise association with a catalytic residue147. The deformation caused by uracil recognition delivers the energy needed for the dissociative reaction of TthUDGa/DNA complexes to hydrolyze the N-glycosidic bond.

To further examine the role of H155 played in the catalysis, we performed atom by atom decomposition of the binding free energy change of wild-type Tth UDGa/DNA complexes. The greatest contribution of catalytic affinity was from atom N1 and O2 of uridine with His155, which confirmed that the interaction between H155 and the O2 of uracil is the leading component in N-glycosidic bond scission. Interestingly, although the side chain of N80 formed two H-bonds with uracil in the ground state, N80 did not contribute to lowering the activation energy because it interacted transition state only -0.01 kcal/mol stronger than ground state. The role of N80 played here in Tth UDGa/DNA complexes
agreed with the N204 of human family 1 UDG by other authors\textsuperscript{134}. Another amino acids F54, which was also expected to be important for specific binding with reactant as discussed above, interacted almost equally well with the ground state and transition state. Therefore, we conclude that the Tth UDGa/DNA complexes hydrolysis the glycosidic bond in a manner analogous to the human family 1 UDG. In the hydrolysis process, both N80 and F54 did not contribute to lowering the activation energy, in addition, the participation of H155 in glycosidic bond scission to form oxacarbenium ion was the rate determining step in $\text{DN}^*\text{A}_\text{N}$ mechanism.

**Mutagenesis studies on the Tth UDGa/DNA complex**

The functional importance of key amino acids was assessed by evaluating the consequences of site-directed mutagenesis on the Tth UDGa activity. Interactions between uracil and Tth UDGa in minimized structures of native protein and alanine scanning point mutants E41A, F54A, N80A, and H155A had been summarized in Table 4.2. Most of the distances between heavy atoms involved in H-bond interaction were maintained well around 2.7-3.4 Å. Conformation of Tth UDGa/DNA complexes did not undergo observable changes since amino acids Glu41, Phe54, Asn80, and His155 that form hydrogen bonds to the uracil substrate were highly conserved among wild-type and mutants.
Table 4.2. Summary of distances (Å) between heavy atoms and angles (°, in parentheses) involved in H-bond interaction between uracil and enzymes of wild-type and mutants

<table>
<thead>
<tr>
<th>H bond</th>
<th>WT</th>
<th>E41A</th>
<th>F54A</th>
<th>N80A</th>
<th>H155A</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>E41N...UraO2</td>
<td>2.94(136.7)</td>
<td>3.15(140.7)</td>
<td>3.17(141.9)</td>
<td>3.21(140.3)</td>
<td>3.18(130.5)</td>
<td>Medium</td>
</tr>
<tr>
<td>G42N...UraO2</td>
<td>3.11(157.2)</td>
<td>3.36(156.9)</td>
<td>3.22(157.6)</td>
<td>3.23(155.1)</td>
<td>3.03(169.4)</td>
<td>Strong</td>
</tr>
<tr>
<td>F54N...UraO4</td>
<td>2.93(156.3)</td>
<td>2.93(162.0)</td>
<td>3.16(159.3)</td>
<td>2.85(154.4)</td>
<td>2.93(154.5)</td>
<td>Strong</td>
</tr>
<tr>
<td>N80ND2...UraO4</td>
<td>2.91(156.2)</td>
<td>2.85(161.2)</td>
<td>2.93(154.7)</td>
<td>none</td>
<td>2.97(157.6)</td>
<td>Strong</td>
</tr>
<tr>
<td>N80OD1...UraN3</td>
<td>2.78(157.8)</td>
<td>2.81(157.3)</td>
<td>2.91(164.1)</td>
<td>none</td>
<td>2.86(162.2)</td>
<td>Strong</td>
</tr>
<tr>
<td>H155NE2...UraO2</td>
<td>2.77(171.2)</td>
<td>2.83(158.8)</td>
<td>2.79(166.9)</td>
<td>2.79(166.5)</td>
<td>none</td>
<td>Strong</td>
</tr>
</tbody>
</table>

Table 4.3. Summary of the MM/PBSA binding free energy of ground state and transition state using ESP scaled charges for Tth UDGa/DNA complexes in native protein and its mutants. The standard errors connected to the averages are given in parentheses.

<table>
<thead>
<tr>
<th>Substrates</th>
<th>ΔG_G (kcal/mol)</th>
<th>ΔG^TS (kcal/mol)</th>
<th>ΔΔG (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>-33.24(0.065)</td>
<td>-33.49(0.067)</td>
<td>-0.25(0.0057)</td>
</tr>
<tr>
<td>E41Q</td>
<td>-32.50(0.109)</td>
<td>-32.30(0.110)</td>
<td>0.20(0.0096)</td>
</tr>
<tr>
<td>G42D</td>
<td>-27.53(0.065)</td>
<td>-27.59(0.065)</td>
<td>-0.06(0.0051)</td>
</tr>
<tr>
<td>E47A</td>
<td>-27.84(0.070)</td>
<td>-27.87(0.070)</td>
<td>-0.03(0.0056)</td>
</tr>
<tr>
<td>F54A</td>
<td>-26.23(0.059)</td>
<td>-26.18(0.059)</td>
<td>0.04(0.0053)</td>
</tr>
<tr>
<td>N80A</td>
<td>-28.84(0.077)</td>
<td>-28.71(0.077)</td>
<td>0.13(0.0049)</td>
</tr>
<tr>
<td>N89A</td>
<td>-30.76(0.064)</td>
<td>-30.94(0.064)</td>
<td>-0.18(0.0057)</td>
</tr>
<tr>
<td>H155A</td>
<td>-30.16(0.084)</td>
<td>-30.02(0.085)</td>
<td>0.14(0.0056)</td>
</tr>
<tr>
<td>E41Q/G42D</td>
<td>-29.8(0.089)</td>
<td>-30.04(0.089)</td>
<td>-0.24(0.0056)</td>
</tr>
</tbody>
</table>

MM/PBSA binding free energy studies of ground state and transition state using ESP scaled charges for Tth UDGa/DNA complexes in native protein and its mutants demonstrated that substitution of residues alter the Tth UDGa/DNA complexes activity at both ground state and transition state (Table 4.3). Although mutations of these residues resulted in varying degree of Tth UDGa/DNA complexes binding affinity, they all effectively weaken the Tth UDGa/DNA complexes interaction in ground state. The order
for ground state binding affinity change ($\Delta G_G$) is F54A < G42D < E47A < N80A < H155A < N89A < E41Q. The greatest variance of binding affinity was from F54A, which had a $\Delta \Delta G_G$ of 6.82 kcal/mol compared to the native protein, this was because F54A lost the aromatic ring which helped orienting the uracil ring through the Pi-Pi stacking effect. However, F54A only lead 0.04 kcal/mol change of the catalytic efficiency $\Delta \Delta G$. Mutant N80A had a 4.21 kcal/mol and 0.13 kcal/mol for binding affinity change $\Delta \Delta G_G$ and catalytic efficiency $\Delta \Delta G$, respectively. Mutant N80A worked in a manner analogous to F54A, that is, both N80 and F54 contribute more in lowering the binding affinity with the reactant than in lowering the activation energy of the glycosidic bond hydrolysis. In contrast to F54A and N80A, H155A had a much smaller change of binding affinity $\Delta \Delta G_G$ about only 2.89 kcal/mol but with the greater change of catalytic efficiency $\Delta \Delta G$ about 0.14 kcal/mol. The mutagenesis studies were agreed with the per-residue decomposition of the MM/PBSA calculations of wild-type Tth UDGa/DNA complexes discussed above. A hybrid QM/MM approach$^{134}$ used to determine the mechanism of catalysis by family 1 UDG showed that an “autocatalysis” derives from the burial and positioning of four phosphate groups that stabilize the rate determine step in the N-glycosidic bond cleavage. The comparison between wild-type and mutagenesis studies proposed that the steric interferences caused by residue substitutions and the distinctive charge distributions in the binding pockets may be major determinants of TthUDGa/DNA complexes.

Co-evolutionary of E41 and G42 positions
To understand the structural and functional correlation between E41 and G42 positions in family 4 Tth UDGa, average hydrogen bond distances were analyzed from MD trajectories. In the wild-type enzyme, the average hydrogen bond distances between the mainchain of E41 and O2 of uracil and between the sidechain of H155 and O2 of uracil are 3.26 Å and 2.86 Å, respectively (Figure 4.6A and E). The short distance between H155-NE2 to the O2 of uracil is suggestive of a strong hydrogen bond. E41Q mutation increased the distances between the O2 of uracil to the mainchain of E41Q and the sidechain of H155 to 3.38 Å and 3.39 Å, respectively (Figure 4.6B and F). This outcome would substantially weaken the hydrogen bonds to O2, resulting in a large loss of UDG activity. The structural effect caused by G42D mutation is more profound for the hydrogen bond distance between the uracil and the E41 than that between the uracil and H155. The increases in average distance are 4.12 Å and 3.03 Å, respectively (Figure 4.6C and G). The concurrent change of E41Q and G42D, however, shortens the hydrogen bond distances between O2 of uracil and the mainchain of E41Q and between O2 of uracil and the sidechain of H155 to 3.27 Å and 2.91 Å, respectively (Figure 4.6D and H, likening what is observed in the wild-type Tth UDGa. The structural alignment of the two important hydrogen bonds brought about by E41Q-G42D doublet is in line with the large recovery of the lost UDG activity in individual amino acid change. These analyses suggest that these two positions are intrinsically correlated and the EG doublet or the QD doublet works in concert to exert its structural and functional impact on family UDGa. The presence of a significant number of co-evolutionary interactions could be used to infer that these residues play a major role in the protein’s
identity and function and thus, the system has evolved in such way that balancing mutations in co-evolving residues can compensate for mutations in these residues.

Figure 4.6. Interactions and two-dimensional scatter plots of the wild-type and mutant Tth UDGa proteins with O2 of uracil in the active site. Modeled interactions with O2 of uracil
in the active site of Tth UDGa-WT (A), Tth UDGa-E41Q (B), Tth UDGa-G42D (C) and Tth UDGa-E41Q-G42D (D). dUMP is colored by atom type. Amino acid residues in the active site of Tth UDGa are shown in licorice in program VMD. Two-dimensional scatter plots of heavy atom distances between E41, H155 and uridine in Tth UDGa-WT (E), Tth UDGa-E41Q (F), Tth UDGa-G42D (G) and Tth UDGa-E41Q-G42D (H).

Figure 4.7. Spearman rank correlation analysis between calculated enzyme efficiency and experimental $k_{cat}/K_M$ data. These results suggest that the electrostatic interaction changes appears to be a major role in enzyme efficiency resulting from mutation.

Spearman rank correlation analysis with experimental data

To further verify the effectiveness of ESP charge scaling model, Spearman rank correlation coefficient was calculated to measure the degree of similarity between calculated enzyme
catalytic efficiency and experimental activity for 8 systems (WT, E41Q, G42D, E47A, F54A, N80A, N89A and E41Q/G42D). The Spearman rank correlation coefficient $r_s$ was chosen here instead of a Pearson correlation for two reasons. First, our emphasis here is to characterize the interaction energy change from ground state to the transition state because of the strong electrostatic changes during the catalytic process. As a result, the qualitative association between the calculated and experimentally determined conducts is sufficient within the scope of the current study. In addition, MM/PBSA binding affinity was noticed to be in exceptional qualitative agreement with experiment in previous studies but the ability to assess absolute binding affinities quantitatively has been questioned$^5, 7, 10, 171$. The calculated changes in activation energy founded on the change in ground state binding and transition state binding show an encouraged Spearman correlation ($r_s=0.786$, $r^2=0.554$) with the experimental activity (Figure 4.7). These calculations indicate that the electrostatic interaction changes appear to be a major role in enzyme efficiency resulting from mutation.

In summary, the Spearman correlation coefficient of 0.8095 means that the predictions of enzyme efficiency from MM/PBSA can give an effective ranking for the studied TthUDGa/DNA complexes. Therefore, transition state locating from DFT calculation combined with MD simulations followed by scaled charging MM/PBSA analysis is an appealing methodology for modeling and ranking enzyme efficiency.
Stability and flexibility of the complexes from MD simulation

To assess the dynamic stability of the whole simulation, we monitored the potential energy and Root Mean Square Deviation (RMSD) relative to the minimized structure along 5 ns MD trajectory. The PE values increased slowly in the initial equilibrium phase, and then the functions were stable after 1.5 ns. Although different patterns are shown for the wild-type Tth UDGa/UDGa and mutants systems, all scenarios demonstrated trivial variation in potential energy during the whole simulation. For convenience, backbone RMSDs of residue E41, F54, N80 and H155, and for each residue at frame 4000 (4.0 ns) are also shown on the top and right of the contour profile (Figure 4.8), respectively. Although all four residues were observed to exhibit a high structural fluctuation deviation compared to
the minimized structure during the initial equilibration phase 0 to 1.0 ns, they maintained an RMSD on the average of 0.75 Å quite well during the entire production run from 2 to 5 ns. The potential energy converged to a very large negative value, and RMSDs converged a value about 0.75 Å in all scenarios, which indicates that the conformations of the Tth UDGa/UDGa complexes achieved good equilibrium. In addition, all-atoms and backbone Root Mean Square Fluctuation (RMSF) relative to the average structure was used to analyze the dynamic behavior of individual residues. We did not observe any considerable change of RMSF even at the active pocket since RMSF in all systems are in the range of 0.25-1.25 Å during the entire 5 ns MD trajectory. The maximum RMSF for both all-atoms and backbone RMSF in all scenarios was limited to 1.5 Å, specifying that the simulations yielded stable trajectories, and therefore afforded an applicable basis for the MM/PBSA free energy analysis.
Figure 4.9. Close-up views of the Tth UDGb-DNA active site interactions in the energy minimized structures. Tube trace of the protein is colored in tan. Uridine (U), inosine (I), and xanthosine (X), as well as amino acids Asp75, Asn120, and His190, are colored by atom type. Water molecules are labeled in purple. Dashed lines indicate inferred hydrogen bonds or water association. (A), modeled Tth UDGb-U interactions. (B), modeled Tth UDGb-I interactions. (C), modeled Tth UDGb-X interactions. (D), chemical illustration of Tth UDGb-U interactions. (E), chemical illustration of Tth UDGb-I interactions. (F), chemical illustration of Tth UDGb-X interactions.

Specificity and catalytic mechanism in family 5 Tth UDGb
To understand the role of Asp75 in the excision of both pyrimidine and purine base damage, we modeled U, I, and X into the crystal structures and carried out molecular dynamic MD simulation analysis of family 5 Tth UDGb.

In the modeled Tth UDGb-U structure, the side chain of Asp75 is in close proximity to the O4 of uracil. A distance averaging 4.9 Å is consistent with a bridging water molecule mediating an interaction between Asp75 and the O4 of uracil. A closer look at the modeled Tth UDGb structures suggests that Asp75 is in close proximity to the 5’ phosphate of the uridine, which can raise the pKa and facilitate protonation of the Asp75 side chain. This will allow the bridging water molecule to form a hydrogen bond with the O4, which facilitates the removal of uracil (Figure 4.9A and D). Mutations at the Asp75 position exhibited more profound effects on HDG and XDG activity. In the modeled UDGb-I and UDGb-X structures, the protonated sidechain carboxylate of Asp75 can interact with the N7 in the hypoxanthine and xanthine through a bridging water molecule in a similar fashion (Figure 4.9, B–E and C–F). MD analysis indicates that the carboxyl side chain is within a distance capable of forming a water-mediated hydrogen bond with the N7 in the purine bases (Fig. 9, B and C). On the other hand, the lack of interaction to the N7 moiety in oxanine may be in part responsible for the lack of ODG activity, although the interaction with the N3 in oxanine appears feasible. Activation of a purine base through hydrogen bonding or protonation has been proposed as a catalytic mechanism for the cleavage of glycosidic bonds in purine nucleotides. Within the UDG superfamily, we identified an Ser23-N7 interaction in the E. coli family 2 MUG enzyme and an Met64 main chain-N7 interaction in the Geobacter metallireducens family 3 SMUG1 enzyme that play important
roles in the excision of xanthine bases. Apparently, N7 interaction or protonation is a common catalytic mechanism for leaving group activation in the hydrolysis of purine deaminated bases.

According to the AP site cocrystal structure, Asn120 is involved in coordinating a water molecule that is located on the opposite side of the deoxyribose. Experimental data showed that elimination of the amide group or lengthening of the amide group by one methylene carbon leads to the loss of XDG activity and a substantial reduction of HDG and UDG activity\textsuperscript{172}. Based on the structural information and biochemical analysis, we speculate that Asn120 may perform a functional role similar to Asn18 in E. coli family 2 MUG, in which the Asn helps activate/position a water molecule for initiating a nucleophilic attack on the glycosidic bond. An average distance between OD1 of Asn120 and C1’ carbon of deoxyribose is \( \sim 5.0 \, \text{Å} \). A structural comparison between Asn18 in MUG and Asn120 in Tth UDGb is shown in Figure 4.10. Although differences in the sequence lengths between the Tth UDGb and the E. coli MUG enzyme prevent a perfect alignment, strong structural similarity is noted in the core secondary structural elements including five \( \beta \)-sheets and four \( \alpha \)-helices. The strong structural similarity is also noted in the active sites. The Asn18 in MUG and Asn120 in Tth UDGb hydrogen bond with water through their side chains, positioning the water molecule proximal to the anomeric carbon of the bound nucleotide. Thus, different families in the UDG superfamily have adopted the same amino acid residue in different structural locations to perform a similar function.
Figure 4.10. Comparison of interactions between Tth UDGb-N120 with water and E. coli MUG-N18 with water. A, superimposition of UDGb-AP structure (Protein Data Bank code 2DEM; green) with E. coli MUG structure (Protein Data Bank code 1MUG; silver). The AP site is colored by atom type. The two structures were superimposed using the program VMD. B, close-up view of Tth UDGb-N120-water and E. coli MUG-N18-water interactions using the same coloring scheme described in A. Asn120 and the interacting water in Tth UDGb structure are shown in purple. Asn18 and the interacting water in the MUG structure are shown in orange.
Figure 4.11. Modeling of interactions between glycosylase and uracil. The hydrogen bonds are shown in blue. (A) Interactions between MUG-WT and uracil based on a model built from the solved crystal structure (PDB 1MUG). K68 and uracil are colored by atom type. (B) Interactions between MUG-K68N and uracil. MUG-K68N structure was modeled from PDB MUG-K68N. K68N and uracil are colored by atom type. (C) Interactions between UNG-WT and uracil based on the solved crystal structure (PDB 2EUG). N123 and uracil are colored by atom type. (D) Interactions between UNG-N123A and uracil. UNG-N123A structure was modeled from PDB 2EUG. N123A and uracil are colored by atom type. (E) Interactions between Tth UDGb-WT and uracil based on the solved crystal structure (PDB 2DEM). A111 and uracil are colored by atom type. (F) Interactions between UDGb-A111N and uracil. UDGb-A111N structure was modeled from PDB 2DEM. A111N and uracil are colored by atom type.
MD simulations of family 2 mismatch-specific UDG (MUG)

Family 2 mismatch-specific uracil DNA glycosylase (MUG) from Escherichia coli is known to exhibit glycosylase activity on three mismatched base pairs, T/U, G/U and C/U. To understand the structural effect of the K68N substitution on MUG, we modeled the Asn substitution to the MUG structure. While the K68 in WT MUG is pointing away from the uracil base, K68N allowed the amide side chain to rotate and form bidentate hydrogen bonds with uracil, as seen in E. coli UNG (Figure 4.11A and B). To understand how N123A may alter the interactions with uracil, we modeled UNG-N123A structure. In comparison with UNG-WT, UNG-N123A lost hydrogen bonds with the uracil base (Figure 4.11C and D). Likewise, Tth UDGb-A111N again formed bidentate hydrogen bonds with uracil (Figure 4.11E and F). We also find N123 in E. coli UNG to be profoundly important for UDG activity. Structurally, N123 forms bidentate hydrogen bonds with N3 and O4 of uracil. Evidently, the possibility to form bidentate hydrogen bonds offered by the K68N mutant greatly enhances the productive MUG-uracil binding, which leads to more efficient excision of uracil from all the base pairs\cite{173}.

**CONCLUSION**

In this study, a combination of DFT, MD, mutagenesis, and structure-activity relationships provides insights about the evolution of Tth UDGa/DNA complexes. The partial atomic charges of ground state in Charmm had been scaled in three different ways, the resulted
partial atomic charges were then assigned to original atoms in MD especially MM/PBSA calculation. Transition state had been successfully located from DFT frequency and IRC calculations based on the B3LYP/6-31+G(d) level. Tth UDGa bound reactant in a typical 3’-exo sugar ring conformation that favors oxacarbenium ion through nucleotide backbone distortion. Classical MD simulation and MM/PBSA analysis including per residue decomposition and per atom decomposition were applied to investigate the binding affinity and catalytic activity toward the Tth UDGa/DNA complexes. Three different Charmm partial atomic charge methods gave quite similar results about the catalytic affinity of Tth UDGa/DNA complexes. We found that Tth UDGa/DNA complexes recognizes and removes uracil specifically from double-stranded DNA through a mechanism analogous to that of the human family 1 UDG. In the hydrolysis process, the participation of H155 in glycosidic bond scission to form oxacarbenium ion was the rate determining step in $D_N^*A_N$ mechanism while both N80 and F54 did not contribute to lowering the activation energy. A detailed analysis of the transition state conformation of the glycosidic bond scission, binding affinity between Tth UDGa protein and flipping out uridine, and catalytic activity of Tth UDGa/DNA complexes suggests that transition state stabilization could be contributing factors in the evolutionary optimization of Tth UDGa/DNA complexes. A reasonable ranking order was achieved comparing to the experimental results. The structural and quantitative viewpoint obtained from this study provides valuable information in understanding the repair of deaminated DNA damage.
APPENDIX A

Table A.1. Residues of Atm1 ABC exporter form H-bond interaction with GSSG (GSH) in the MD simulation

<table>
<thead>
<tr>
<th>H-bond donor (acceptor) of GSSG (GSH)</th>
<th>Interaction and occupancy in GSSG1</th>
<th>Interaction and occupancy in GSSG2</th>
<th>Interaction and occupancy in GSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>N161_OD1 (0.1%), N269_O (0.1%), N99_OD1 (0.6%), D316_O (0.1%), D316_OD1 (0.2%), D316_OD2 (0.2%), N99_ND2 (8.79%), R101_NH2 (42.86%), Y156_OH (2.4%), R313_NH1 (0.1%), N161_ND2 (0.1%), Y156_OH (0.5%),</td>
<td>Y156_OH (0.6%), M320_SD (0.1%), R210_NH1 (0.1%), T153_OG1 (0.3%), S150_OG (0.1%), K149_O (0.3%), D152_OD2 (0.4%), D152_OD1 (0.5%), Y254_OH (1.2%), T324_OG1 (0.7%), M320_O (0.8%), R146_NH1 (0.2%), R206_NH2 (0.3%), R210_NH1 (1.7%), T324_OG1 (0.1%), Y254_OH (0.1%)</td>
<td>R210_NH1 (0.1%), N102_ND2 (0.1%), N269_OD1 (0.5%), N99_OD1 (0.2%), D152_OD1 (0.1%), D316_O (1.5%), Q272_OE1 (0.1%), Y156_OH (0.4%), N269_ND2 (0.7%), G319_N (0.1%), N102_ND2 (0.1%), R313_NH2 (0.1%), R210_NH1 (0.1%), R210_NH2 (0.2%), M317_O (0.7%), M320_SD (0.1%), P314_O (18.08%), T153_OG1 (0.2%), T324_OG1 (0.2%), Y156_OH (0.4%)</td>
</tr>
<tr>
<td>N2</td>
<td>D316_O (0.2%), T153_OG1 (3.4%)</td>
<td>N161_OD1 (8.59%), D316_O (1.5%), M317_SD (0.20%), M320_SD (0.1%), Y156_OH (6.89%)</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>E262_OE2 (3.0%), E262_OE1 (4.3%)</td>
<td>N161_OD1 (9.19%), D152_OD1 (0.1%), D316_O (2.5%), D316_OD1 (0.3%), D98_OD2 (0.1%), E262_OE1 (1.0%), E262_OE2 (0.8%), M317_O (16.88%),</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>O1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N99_ND2 (0.1%), R101_NH2 (12.09%), N161_ND2 (5.59%), N269_ND2 (35.26%), R313_NH1 (1.0%), R313_NH2 (1.3%), Y156_OH (3.3%), M317_N (0.6%), D316_N (1.5%),</td>
<td>R206_NE (3.1%), R206_NH2 (3.8%), R210_NH1 (0.1%), R210_NH2 (0.1%), K149_NZ (0.2%), T153_OG1 (0.8%)</td>
<td>Q272_NE (0.1%), G319_N (1.8%), N269_ND2 (1.6%), N99_ND2 (0.2%), M320_N (0.9%), Y156_OH (0.2%), N102_ND2 (4.8%), N161_ND2 (0.2%), R313_NE (0.1%), R313_NH2 (0.1%), L318_N (0.1%)</td>
</tr>
<tr>
<td><strong>N4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R313_O (0.1%), N161_OD1 (2.5%), N161_ND2 (0.1%), M317_SD (0.9%), T153_OG1 (0.1%), T324_OG1 (0.2%), Y156_OH (0.6%), M320_SD (0.2%)</td>
<td>T153_OG1 (0.1%), T324_OG1 (0.2%), Y156_OH (0.6%), M320_SD (0.2%)</td>
<td>F160_O (1.8%), T153_OG1 (0.5%), Y156_OH (2.0%)</td>
</tr>
<tr>
<td><strong>N5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y156_OH (18.18%)</td>
<td>T153_OG1 (4.0%)</td>
<td></td>
</tr>
<tr>
<td><strong>N6</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N102_ND2 (5.99%), N102_OD1 (7.89%), N269_ND2 (50.45%), N269_OD1 (2.8%), N99_OD1 (4.2%), D316_O (5.09%), D316_OD1 (2.2%), D316_OD2 (0.4%), N269_ND2 (2.1%), N99_ND2 (2.7%), R101_NH2 (0.3%), N102_ND2 (9.49%), R101_NH1 (3.8%), Q272_OE1 (0.1%), L265_O (0.4%), M317_SD (0.1%), M320_SD (0.3%), F160_O (1.0%), Y156_OH (0.3%)</td>
<td>Y156_OH (2.3%), N161_OD1 (0.9%), D316_O (0.2%), E262_OE2 (0.1%), T153_OG1 (1.1%), T324_OG1 (0.7%), K149_O (0.1%), R210_NH1 (0.1%), R210_NH2 (0.1%), R323_NH1 (0.2%), T324_OG1 (0.1%), Y156_OH (0.1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T153_OG1 (2.1%)</td>
<td>O2</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>O10</td>
<td>N369_ND2 (0.6%), Y156_OH (0.1%), R101_NH2 (0.5%), R313_NH2 (1.5%)</td>
<td>R313_O (0.1%), N161_OD1 (0.1%), N269_ORD1 (0.5%), D316_O (1.6%), D316_OD2 (2.5%), D98_OD1 (0.1%), D98_OD2 (2.5%), G319_N (3.5%), Q272_NE2 (0.1%), N99_ND2 (0.1%), N269_ND2 (1.0%), R101_NH2 (0.2%), M320_N (0.3%), D316_N (0.2%), T153_OG1 (0.3%), N161_ND2 (1.1%), Y156_OH (0.1%), Q272_OE1 (0.3%), F160_O (0.4%), Y156_OH (0.3%),</td>
<td>R210_NH2 (0.2%), T215_OG1 (0.4%), Y156_OH (0.9%)</td>
</tr>
<tr>
<td></td>
<td>R313_NH1 (1.9%), R313_NH2 (0.2%), M317_N (0.2%), R313_NE (1.8%), T324_OG1 (1.5%), T154_OG1 (0.3%),</td>
<td>Y156_OH (0.4%)</td>
<td>R313_NH1 (0.1%), R313_NE (1.0%), Y156_OH (1.5%), R313_NH2 (0.3%), M317_N (0.1%), K142_NZ (0.1%), L138_N (0.9%), N269_ND2 (0.2%), G319_N (0.7%)</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------</td>
<td>----------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>O6</td>
<td>R313_NH2 (5.99%), N161_ND2 (0.4%), R313_NH1 (0.1%),</td>
<td>R210_NH1 (0.1%), R210_NH2 (0.4%), T153_OG1 (6.89%)</td>
<td></td>
</tr>
<tr>
<td>O7</td>
<td>Y156_OH (1.5%), N161_ND2 (0.4%),</td>
<td>T153_OG1 (0.2%), Y156_OH (0.2%)</td>
<td></td>
</tr>
<tr>
<td>O8</td>
<td>M320_N (3.0%), N269_ND2 (0.5%), Y156_OH (2.0%), N99_ND2 (0.1%), N102_ND2 (0.3%), R101_NH2 (1.9%), N161_ND2 (0.2%),</td>
<td>R146_NH1 (3.1%), R210_NH1 (0.1%), R210_NH2 (0.5%), R323_NE (0.1%), R323_NH2 (0.2%), N269_ND2 (4.1%), M320_N (4.1%), T153_OG1 (0.2%), T324_OG1 (0.1%), Y156_OH (0.1%)</td>
<td></td>
</tr>
<tr>
<td>O9</td>
<td>N102_OD1 (0.6%), N269_OD1 (0.1%), N99_OD1 (0.4%), D316_O (0.5%), D316_OD1 (1.5%), D316_OD2 (0.2%), D98_O (0.3%), D98_OD2 (0.3%), G319_N (0.1%), Q272_NE2 (0.3%), M320_N (0.3%), N269_ND2 (0.3%), Y156_OH (0.1%), N102_ND2 (0.4%), N99_ND2 (0.1%), N161_ND2 (0.1%), R101_NH2 (0.3%), Q262_OE1 (0.1%), Q272_OE2 (4.7%),</td>
<td>Y156_OH (0.5%), N161_OD1 (0.1%), N269_OD1 (0.3%), D316_O (0.7%), E262_OE1 (0.1%), T153_OG1 (3.3%), T324_OG1 (3.4%), K149_O (0.2%), D152_OD2 (0.7%), R210_NH2 (0.1%), N269_ND2 (1.5%), M320_N (1.5%), Y156_OH (0.1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M320_SD (0.1%), Y156_O (2.2%), Y156_OH (0.7%)</td>
<td>T324_OG1 (7.89%), T153_OG1 (5.09%), M320_SD (0.2%), D152_OD1 (1.0%), E262_OE1 (8.19%), E106_OE1 (3.2%), E106_OE2 (6.69%), E262_OE2 (10.79%), R206_NH2 (0.5%)</td>
<td>R313_O (2.2%), N161_OD1 (2.3%), N269_OD1 (0.1%), D152_OD1 (1.0%), D152_OD2 (1.0%), D316_O (2.4%), D316_OD1 (3.2%), D316_OD2 (0.6%), Q272_OE1 (4.1%), E106_OE2 (0.2%), E145_O (0.3%), E145_OE1 (0.1%), E145_OE2 (0.2%), E168_OE1 (1.5%), E168_OE2 (3.1%), E262_OE1 (2.7%), E262_OE2 (2.9%), Y156_OH (0.2%), T165_OG1 (0.1%), N161_ND2 (0.5%), L318_N (0.2%), N269_ND2 (1.1%), G319_N (0.3%), L315_O (0.5%), M317_O (0.1%), F160_O (0.1%), P314_O (0.4%), T153_OG1 (0.1%), T165_OG1 (0.5%), T324_OG1 (0.7%), Y156_OH (1.5%)</td>
</tr>
<tr>
<td>--</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>OXT1</td>
<td>R313_O (0.1%), N102_OD1 (0.1%), N161_OD1 (0.2%), N161_ND2 (0.1%), M317_N (0.1%), T324_OG1 (0.1%), T153_OG1 (0.4%), P314_O (0.2%), T153_OG1 (0.1%), T324_OG1 (0.2%), Y156_OH (0.1%)</td>
<td>T153_OG1 (2.8%), T153_O (0.3%), D152_OD2 (0.1%), D152_OD1 (0.1%), T324_OG1 (3.3%), K149_NZ (0.1%), T153_OG1 (0.3%), Y156_OH (0.1%)</td>
<td>R313_O (0.1%), R313_NH2 (0.3%), N161_OD1 (0.2%), D316_O (1.8%), D316_OD1 (0.2%), D98_OD2 (0.1%), E145_O (0.1%), N161_ND2 (0.3%), M317_N (0.1%), M320_N (0.2%),</td>
</tr>
<tr>
<td>OXT2</td>
<td>R313_O (0.1%), D316_OD1 (0.1%), D316_OD2 (0.2%), Y156_OH (0.1%), N161_ND2 (0.3%), P314_O (0.1%)</td>
<td>T153_OG1 (2.8%), T153_O (0.3%), D152_OD2 (0.1%), D152_OD1 (0.1%), T324_OG1 (3.3%), K149_NZ (0.1%), T153_OG1 (0.3%), Y156_OH (0.1%)</td>
<td>R313_O (0.1%), R313_NH2 (0.3%), N161_OD1 (0.2%), D316_O (1.8%), D316_OD1 (0.2%), D98_OD2 (0.1%), E145_O (0.1%), N161_ND2 (0.3%), M317_N (0.1%), M320_N (0.2%),</td>
</tr>
<tr>
<td>S</td>
<td>Y156_OH (0.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M317_O (0.2%), Y156_OH (0.4%)</td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


156. N. Guex, M. C. Peitsch, ELECTROPHORESIS 1997, 18, 2714-2723.


