Computational techniques to illuminate secrets of the monoamine transporters

Bonnie A. Merchant
COMPUTATIONAL TECHNIQUES TO ILLUMINATE SECRETS OF
THE MONOAMINE TRANSPORTERS

A Thesis
Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for
the degree Masters of Science

By
Bonnie A. Merchant

December 2012
COMPUTATIONAL TECHNIQUES TO ILLUMINATE SECRETS OF
THE MONOAMINE TRANSPORTERS

By

Bonnie A. Merchant

Approved November 16, 2012

Jeffry D. Madura
Professor of Chemistry and
Biochemistry
(Committee Chair)

Michael Cascio
Associate Professor of Chemistry and
Biochemistry
(Committee Member)

Jeffrey D. Evanseck
Professor of Chemistry and
Biochemistry
(Committee Member)

David W. Seybert
Dean, Bayer School of Natural and
Environmental Sciences
Professor of Chemistry and
Biochemistry

Ralph A. Wheeler
Chair, Department of Chemistry and
Biochemistry
Professor of Chemistry and
Biochemistry
ABSTRACT

COMPUTATIONAL TECHNIQUES TO ILLUMINATE SECRETS OF THE MONOAMINE TRANSPORTERS

By

Bonnie A. Merchant

December 2012

Thesis supervised by Jeffry D. Madura

The solute carrier family regulates the flow of various substances such as drugs, amino acids, sugars and inorganic ions across the cell membrane. In particular, the monoamine transporters (MATs) are responsible for the regulation of synaptic neurotransmitter levels. Their structures and conformational rearrangements associated with transport remain unsolved. We employed novel computational techniques to identify the binding pocket of cocaine in the dopamine transporter (DAT) and two intracellular pathways for substrate egress in LeuT and DAT. We review possible coarse grained molecular dynamics techniques to extend the temporal scale accessible in simulations of the MATs. Finally, we present the first computational study of DAT in the presence of an explicit electrochemical gradient. In these simulations, we identify a water wire in DAT that may suggest that DAT does not follow an alternating access mechanism.
DEDICATION

This thesis would be incomplete without a mention of the support given unto me by my dear friend, Carrie, to whom this thesis is dedicated. Without her lifting me up when this thesis seemed insurmountable, I doubt it would ever have been completed.
ACKNOWLEDGEMENT

It is a pleasure to thank the many people that have made this thesis possible. First and foremost, I would like to thank my advisor and mentor, Jeffry D. Madura. You have pushed me intellectually beyond anything I thought I could accomplish. I thank you for the guidance, encouragement and support and opportunities, while introducing me to the wonders and frustrations of scientific research. Your persistence in having us “sit down and play with it” provided spirited motivation to consistently become an independent thinker and problem-solver.

I also have to thank the members of my committee, Professors Michael Cascio and Jeffry D. Evanseck for their helpful, advice, mentoring and thoughtful discussions. I am thankful for the many experimental and computational collaborators I have had the opportunity to work with and learn from. Thank you to the Professors Wheeler and Evanseck for the joint-computational group meetings over the years. Collaborators from the Lapinsky and Surratt groups, particularly Tammy Nolan and Laura Geffert, consistently provided useful and stimulating scientific considerations, in addition to a shared love of furry felines.

Fellow lab mates, who provided a stimulating and fun environment in which to learn and grow, undoubtedly shaped this journey. I am grateful for the time I have spent with Sankar Manepalli, Kalyan Immadisetty and Eliana Asciutto. Special gratitude is extended to Matt Srnec, for keeping me looking fresh with Friday haircuts. I am forever indebted and grateful for my friendship with Tim Gaborek, my left-hand man. To our undergraduate students, especially Jimmy Brancho, James Thomas, Jon Gibbons and
Emily Litzenberger, thank you for the opportunity to mentor you, and conversely, always teaching me a thing or two. I wish you the best in your future endeavors.

I am especially grateful to (past and present) graduate students Rachelle Palchesko Simko, Becky Wagner, Kristen Kruszewski, Gavin Buckholtz, Linh Nguyen, Courtney Watkins and of course, my boys- Tim Evans and Matt Purzycki for helping me get through the difficult times, and for all the emotional support, camaraderie, entertainment, and caring they provided.

Over the years, I have received priceless assistance from the staff of the Department and the University. Special thanks to the Department’s administrative support, Sandy Russell, Amy Stroyne and Margaret Cowburn for helping the Department to run smoothly and for assisting me in many different ways. I would like the folks in CTS, especially John Astleford and Drew Smorul for working with me on various projects to keep our lab running smoothly and churning out the data. Your assistance and expertise is highly valued. Scott Boesch- you are a life saver!

I wish to thank my parents and family. They bore me, raised me, supported me, and believed in me. I would not be the person I am today without the ceaseless support and love I have received from Carrie McHargue, Ziad Abulleil and the Paris family. I am so thankful for all the times you’ve scraped me off the ground and reassembled me into a human again. You have all contributed irreversibly to the person I have become. I cannot thank you enough.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................ iv
DEDICATION .................................................................................................. v
ACKNOWLEDGEMENT ............................................................................... vi
LIST OF TABLES ......................................................................................... x
LIST OF FIGURES ....................................................................................... xi

MOVING MOLECULES: THE ROLE AND FUNCTION OF MEMBRANE-BOUND TRANSPORTERS ........................................... 1
  1.1 Introduction .............................................................................................. 1
  1.2 Moving Molecules across the Membrane ........................................... 1
  1.3 Classifying Proteins Based on Sequence ........................................... 3
  1.4 Classifying Proteins Based on Structure ........................................... 4
  1.5 Method of Transport .............................................................................. 5
  1.6 The LeuT-Fold ....................................................................................... 11
  1.7 Summary ............................................................................................... 14
  1.8 References ............................................................................................ 15

INSIGHTS FROM MOLECULAR DYNAMICS: THE BINDING SITE OF COCAINE IN THE DOPAMINE TRANSPORTER AND PERMEATION PATHWAYS OF SUBSTRATES IN THE LEUCINE AND DOPAMINE TRANSPORTERS ........ 19
  2.1 Abstract ................................................................................................. 19
  2.2 Significance and Function of the MATs ............................................... 20
  2.3 The Leucine Transporter, LeuT_Aa .................................................... 22
    2.3.1 Structure and Relationship to MATs ........................................... 22
    2.3.2 Primary Binding Pocket ............................................................... 24
    2.3.4 Proposed Mechanism ................................................................. 25
    2.3.5 A Secondary Binding Pocket ...................................................... 26
  2.4 Methods ................................................................................................. 29
    2.4.1 LeuT_Aa Structure for MCTI Simulations ................................... 29
    2.4.2 DAT Structure for MCTI Simulations ......................................... 29
    2.4.3 Solvated Protein and Lipid Bilayer Construction for RAMD Simulations .... 30
    2.4.4 Multi-Configuration Thermodynamic Integration ...................... 31
    2.4.5 Random Acceleration Molecular Dynamics ............................. 34
  2.5 Results and Discussion ......................................................................... 38
    2.5.1 MCTI ......................................................................................... 38
    2.5.2 RAMD ....................................................................................... 45
  2.6 Summary ............................................................................................... 52
    2.6.1 Insights regarding the leucine transporter .................................. 52
    2.6.2 Insights regarding the dopamine transporter ........................... 54
  2.7 Acknowledgments ............................................................................... 55
  2.8 References ............................................................................................ 55
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Selection of Pfam clans and superfamilies</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Selection of transporters and their associated gene names</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Summary of available LeuT structures</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Transporters and their associated PDB codes</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>Summary of RAMD simulation details and parameters</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>Summary of pathways observed in RAMD simulations of the leucine transporter</td>
<td>47</td>
</tr>
<tr>
<td>7</td>
<td>Observed residues interacting via hydrogen bonding for each pathway cluster in DAT</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>Strengths of interactions defined for non-bonded interactions of CG beads in the MARTINI model</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>Force constants and equilibrium bond angles for the angular potential term of the MARTINI force field</td>
<td>69</td>
</tr>
<tr>
<td>10</td>
<td>Density, self-diffusion coefficient, surface tension, isothermal compressibility, free energy of solvation and bead ratio of various CG models</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>Approximate extracellular and intracellular concentrations of potassium, sodium, chloride and calcium in a mammalian cell</td>
<td>92</td>
</tr>
<tr>
<td>12</td>
<td>Calculated potentials for potassium, sodium and chloride ions and overall cell</td>
<td>102</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Types of transport.</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Total and yearly additions of entries to the Protein Data Bank.</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>The alternating access mechanism.</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>Total and year additions of unique folds into the Protein Data Bank.</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Topology of LeuT.</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>The LeuT-Fold is constructed of five inverted repeats.</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Regulation of neurotransmitter levels at the chemical synapse.</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>Crystal structure of LeuT embedded in a lipid bilayer.</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>Annotated sequence alignment of LeuT and DAT.</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>The alternating access mechanism in LeuT.</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>Close-up of the S1 and S2 sites in LeuT.</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>A graphic representation of the LeuT:Leu complex for MCTI simulations.</td>
<td>33</td>
</tr>
<tr>
<td>13</td>
<td>Free energies profiles of the substrate leucine through LeuT.</td>
<td>39</td>
</tr>
<tr>
<td>14</td>
<td>Free energies profiles of the substrate dopamine and inhibitor cocaine in DAT.</td>
<td>41</td>
</tr>
<tr>
<td>15</td>
<td>Residue interactions of DAT with dopamine and cocaine at the primary (S1) pocket.</td>
<td>43</td>
</tr>
<tr>
<td>16</td>
<td>Stop-over sites identified for leucine in LeuT, water sphere environment.</td>
<td>45</td>
</tr>
<tr>
<td>17</td>
<td>Selection of pathways observed in LeuT$_{Aa}$ RAMD simulations.</td>
<td>49</td>
</tr>
<tr>
<td>18</td>
<td>Pathways observed in the DAT:DA RAMD simulations.</td>
<td>52</td>
</tr>
<tr>
<td>19</td>
<td>Schematic of computer simulation methods over varying temporal and spatial scales.</td>
<td>62</td>
</tr>
</tbody>
</table>
Figure 20. Results from Scopus® upon searching for various terms regarding coarse-grained molecular dynamics.................................................................64

Figure 21. Mapping of CG beads onto A) DPPC lipid, B) Cysteine, C) Leucine and D) water molecules in the MARTINI model..........................................................67

Figure 22. Schematic representation of a hybrid adaptive AA/CG simulation system.....76

Figure 23. Proposed system set up for dual-lipid explicit ion gradient simulations. ....79

Figure 24. Representation of DAT embedded in a POPE membrane..............................92

Figure 25. Two-dimensional depiction of periodic boundary conditions. .......................94

Figure 26. Modified simulation cell to achieve the explicit ionic gradient while employing PBC in molecular dynamics simulations.................................................95

Figure 27. Glide-plane boundary conditions for molecular dynamics simulations using the Extended System Program.................................................................96

Figure 28. Snapshot of the CNT dual bilayer system after minimization, temperature equilibration and initial dynamics (5ns).............................................................98

Figure 29. DAT dual bilayer system.............................................................................102

Figure 30. Snapshots from the CNT dual bilayer simulation.........................................104

Figure 31. Selected salt bridge distances observed to fluctuate the most of the collected trajectories.................................................................................................105

Figure 32. Snapshots from the apo and S1 trajectories..................................................107

Figure 33. Isosurface representing water within 3 Å of DAT. .......................................108

Figure 34. Isosurfaces that represent the overall positions of ions during the DAT trajectory.........................................................................................................109
CHAPTER ONE

MOVING MOLECULES: THE ROLE AND FUNCTION OF MEMBRANE-BOUND TRANSPORTERS

1.1 Introduction

Comprised of amino acids, the building blocks of life, proteins are as diverse as they are ubiquitous. With countless permutations of amino acid sequences, various secondary and tertiary structures arise. Functional proteins can vary in length from short polypeptide chains to large, complex compilations of thousands of amino acids. Despite their diversity, proteins can be organized by sequence similarity, common fold or function. By analyzing their structural similarities and differences, key issues involving their functions and mechanisms can be resolved.

1.2 Moving Molecules across the Membrane

The phospholipid plasma membrane of a cell acts as a selective barrier to facilitate the separation of the cytoplasmic and extracellular fluids. There are three main ways for a substance to cross the membrane: passive diffusion, facilitated diffusion and active transport. In passive diffusion, substances will spontaneously move from a region of higher concentration to that of a lower concentration. The diffusion of substances across the membrane is highly affected by the substances characteristics like the size, hydrophobicity and charge on the molecule. Thus, only small, uncharged, hydrophobic molecules like fatty acids and alcohols will freely cross the membrane via passive diffusion.
In contrast, sodium ions which are charged do not freely flow across the membrane from a higher concentration to a lower concentration. The typical concentration of sodium ions in the extracellular fluid is 145 mM and 10 mM in the cytoplasmic fluid. Despite the concentration gradient, sodium ions do not flow freely from extracellular fluid to the intracellular fluid because the ions are too polar to pass through the hydrophobic lipid membrane. In this case, there are membrane proteins which form pores for the ions to pass through, called channels. These channels thus facilitate the diffusion of the substance, with the concentration gradient. Since the substance is still moving with the concentration gradient, facilitated diffusion is also referred to as passive transport or uniport, since a single molecule is moved across the membrane (as shown in the first panel of Figure 1).

![Figure 1. Types of transport. Uniport: where a single substrate (red star) is transported across the membrane Symport: where a substrate is driven against its concentration gradient and the driving molecule (purple circle) moves with its concentration gradient in the same direction across the membrane. Antiport: where a substrate is driven against its concentration gradient while the driving molecule moves with its concentration gradient in the opposite direction across the membrane.](image)

The last type of transport describes the movement of substances against their concentration gradient. There are two types of active transport: primary and secondary. In primary active transport, chemical energy (such as ATP) is used directly to transport the
substance across the membrane. However, in secondary active transport, the membrane proteins rely on the electrochemical gradient (generated by primary active transport) as the energy source for transport.

Since secondary active transporter couple the translocation of two or more substrates, these molecular pumps are generally called co-transporters. The “driving” substrate (often a proton or ion) always moves across the membrane with its' gradient while the “driven” substrate moves against its' gradient. In this way, the driven molecules arrives on the opposite side of the membrane via the energy stored in the electrochemical gradient.³

The direction of movement across the membrane of the substrates indicates either symport (same direction) or antiport (opposite) as indicated in the second and third panels of Figure 1. Simple symport systems are often called “co-transporters””while simple antiport systems are often called “exchangers”. In later sections, we will discuss proposed transport mechanism for transporters.

1.3 Classifying Proteins Based on Sequence

The study of proteins and bioinformatics became increasingly popular in the late 1960s and 1970s, and has become an integral part of understanding biology and the human body today. Significant progress was made in this time, and nomenclature problems began to arise in human genetics. In 1979, the HUGO Gene Nomenclature Committee (HGNC)⁴ presented a full set of nomenclature guidelines at the Edinburgh Human Genome Meeting (HGM). Since then, the HGNC has assigned unique gene symbols and names to over 33,000 human loci, of which around 19,000 are protein coding.⁵
To be assigned to a specific family in the HUGO Gene Nomenclature Committee (HGNC), a protein must display a 20-25% amino acid sequence identity to another member of the family.\(^6\) The largest family classified by HGNC is the G protein-coupled receptors (GPCR) which contains over 800 members. Second to the GPCRs is the Solute Carrier (SLC) family, which contains nearly 400 members.\(^7\) Within each family, there are distinct sub-families which group more related proteins. Accordingly, there are 51 distinct families of SLCs in humans.\(^6\) The structure, function and current knowledge of the SLCs will be the focus of this chapter.

1.4 Classifying Proteins Based on Structure

Not only do we have sequences of proteins available for study, but for many proteins, we have experimentally-derived structures of the proteins. For example, in Figure 2 we highlight the total searchable structures in the Protein Data Bank\(^8\) (PDB: www.pdb.org) as well as the yearly additions to the PDB. These experimentally derived structures go beyond providing the sequence- they provide us a three-dimensional image of how the protein folds together.

![Figure 2](image.png)

Figure 2. Total and yearly additions of entries to the Protein Data Bank. Data obtained from www.pdb.org.\(^8\)
Another classification system used to organize the vast amount of protein data available is Pfam. This system is built on the concept of clans and families which organize homologous proteins, with a focus on functional domain regions of proteins obtained from literature, known structures and different databases. Within the Pfam classification system, families are grouped into clans which are believed to have a common evolutionarily ancestor. In the current release of Pfam (version 24.0), there are 11,912 families, of which 3,131 of those families are organized into roughly 300 clans.

Within the Pfam classification system, the 51 SLC families (according to HGNC nomenclature) primarily fall into three clans, the amino acid polyamine organocation (APC) family, the major facilitator superfamily (MFS) and the cation-proton antiporter/anion transporter (CPA/AT) family. The majority of the SLCs fall into the MFS, followed by the APC superfamily. Table 1 provides a selection of the clans or families that the SLCs belong to, as described in a study by Hoglund, Nordstrom, Schioth and Fredriksson.

Table 1. Selection of Pfam clans and superfamilies which include SLC families. †Indicates only a family name, as no clan is assigned by Pfam. Data obtained from Hoglund, Nordstrom, Schioth and Fredriksson.

<table>
<thead>
<tr>
<th>Clan or Family† Name</th>
<th>SLC Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFS</td>
<td>SLC2, 15, 16, 17, 18, 19, 21, 22, 29, 33, 37, 43, 45, 46</td>
</tr>
<tr>
<td>APC Superfamily</td>
<td>SLC4 5, 7, 12, 23, 26, 32, 36, 38</td>
</tr>
<tr>
<td>CPA/AT transporter superfamily; Ion Channel (VIC) superfamily</td>
<td>SLC9, 10</td>
</tr>
<tr>
<td>SNF Superfamily†</td>
<td>SLC6</td>
</tr>
<tr>
<td>SDF superfamily†</td>
<td>SLC1</td>
</tr>
</tbody>
</table>

1.5 Method of Transport

The SLC family is responsible for moving macromolecules across membranes. SLCs regulate the flow of various substances such as drugs, amino acids, sugars and
inorganic ions across the cell membrane. With the exception of a few members, the SLC families are functionally related in that they rely on an ion gradient as the driving force for transport. Thus, most of the SLCs are co-transporters.

An important feature of the SLCs is that the binding site of the substrate is never accessible from both sides of the membrane concurrently. Since only one side is accessible at a time, the mechanism of transport has long been described as an “alternating access” mechanism. The early mechanistic understanding of this entailed the substrate binding to one side of the transporter and simply migrating to the other side where it was released.

In 1990, Peter Mitchell proposed the transport process should be thought as a moving barrier. He described a central fixed binding site in the transporter with alternating permeability barriers to either aqueous phase. This mechanism proposed two main states of the protein, an outward-facing state and an inward facing state. As the mechanism likely cannot be so simply described by two states, there may exist several intermediates, for example: the outward-facing open (OF-o), outward-facing occluded (OF-occ), inward-facing occluded (IF-occ) and the inward-facing open (IF-o) as shown in Figure 3.

Here, the OF-o conformation would be ready to bind a substrate (and driving ions) to the central binding site, while the OF-occ conformation would represent an outward-facing conformation with the bound substrates, closed off from the extracellular medium by gates. Similarly, a conformational shift in the protein would allow for an inward-facing conformation with the bound substrate, yet still closed off from the intracellular medium by gates (IF-occ). Finally, another conformational shift would allow
for the substrate to be released to the cytoplasm once those gates are open in the IF-o conformation.

Figure 3. The alternating access mechanism proposes a central binding site that is alternately accessible to the extracellular and intracellular mediums. Prokaryotic transporters crystallized in specific conformations of the transport cycle are highlighted for each conformation.

The alternating access mechanism is generally widely accepted for many prokaryotic homologues of human SLCs, including the sodium coupled symporters: the leucine transporter (LeuT\textsuperscript{15}), the benzylhydantoin transporter (Mhp1\textsuperscript{16,17}), the betaine transporer (BetP\textsuperscript{18}) and the galactose transporter (vSLGT\textsuperscript{19}). Additionally, the alternating access mechanism has been accepted for the proton-coupled amino acid symporter
(ApcT\textsuperscript{20}) the Na\textsuperscript{+}-independent carnitine-butyrobetaine antiporter (CaIT\textsuperscript{21}) the arginine-agmatine antiporter (AdiC\textsuperscript{22,23}) and lactose permease (LacY\textsuperscript{24}), a proton/sugar symporter. However, despite the wide acceptance of the alternating access mechanism for these transporters (which include symporters and antiporters), the exact mechanistic details of substrate binding, release and conformational rearrangements related to each protein remain relatively elusive.

Importantly, those prokaryotic structures mentioned above have been crystallized in recent years and have provided templates in order to study and explore the structure and function of their eukaryotic homologues (see Table 2). For example, homology models of the human Na\textsuperscript{+}-glucose cotransporter hSGLT1 in the outward-open and inward-occluded states have been constructed based on crystal structures of Mhp1 and vSGLT, respectively.\textsuperscript{25}

Table 2. Selection of transporters and their associated gene names, examples and prokaryotic templates for homology modeling.

<table>
<thead>
<tr>
<th>Pfam Clan or Family</th>
<th>Gene Family Name</th>
<th>Example(s)</th>
<th>Prokaryotic Template(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>SLC5: Sodium-glucose co-transporter family</td>
<td>SGLT1, SGLT2</td>
<td>vSGLT, Mhp1</td>
</tr>
<tr>
<td>SNF</td>
<td>SLC: Sodium and chloride dependent neurotransmitter transporter family</td>
<td>DAT, NET, SERT</td>
<td>LeuT</td>
</tr>
<tr>
<td>MFS</td>
<td>SLC19: Folate/thiamine transporter family</td>
<td>RFC, ThTr1</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>SLC38: Sodium-coupled neutral amino acid transporter family</td>
<td>SNAT2, SNAT4</td>
<td>Mhp1, LeuT</td>
</tr>
<tr>
<td>MFS</td>
<td>SLC26: Folate transporter family</td>
<td>PCFT</td>
<td></td>
</tr>
</tbody>
</table>

Crystallization of LeuT\textsuperscript{15} from \textit{Aquifex Aeolicus} in 2005 provided was a landmark event, in that it provided a prokaryotic structure template for human neurotransmitters transporters of the SLC6 family, which include the dopamine (DAT), serotonin (SERT) and norepinephrine (NET) transporters. At present, there are over 30 structures available
for LeuT (see Table 3). Despite many structures for LeuT, the vast majority are of an outward facing states—either occluded or open, with or without a variety of substrates bound to the transporter.

In fact, for LeuT there is only one structure that reflects an inward-open state. The 3TT3 structure was obtained by the Gouaux group in 2012 via a complex of a mutated LeuT with conformation-specific antibody fragments. In this way, the Gouaux group was able to capture the elusive inward-facing state of LeuT, despite at relatively low resolution compared to many of the other structures. Their structure suggested that large-scale conformational rearrangements were necessary to reach in the inward-facing state, particularly of TMs 1, 2, 5, 6 and 7. They observed distinct hinge-bending movements of the partially unwound TM1a (by about 45° from the closed state), and the closure of the extracellular vestibule by extracellular loop 4.

The insights garnered from this structure should be interpreted carefully. Firstly, they report that electron density for residues 1-10 was not observed and electron density for residue 11-24 was weak. Importantly, residues 10-24 comprise TM1a, the helix that is observed to move nearly 45° from its position in the closed state. TM6b, by contrast only rotates away from the binding site by 17°. In contrast, TMs 1b and 6a tilt together in more of a coupled fashion by 24° and 21°, respectively. This may bring into question the fact that the two TM pairs, 1a/6b and 1b/6a actually do participate in coupled movements which require roughly a 20° rotation. The exaggerated movement of TM1a may in fact be an artifact of the complex with the antibody fragment.
Table 3. Summary of available LeuT structures, organized by decreasing resolution. Relevant details such as the bound substrate, mutation or environmental condition and conformation is noted for each entry.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Substrate, Mutation or Environmental Condition</th>
<th>Conformation</th>
<th>Resolution (Å)</th>
<th>Year</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A65</td>
<td>Leucine</td>
<td>O-occ</td>
<td>1.65</td>
<td>2005</td>
<td>16</td>
</tr>
<tr>
<td>2Q72</td>
<td>Impramine</td>
<td>O-occ</td>
<td>1.70</td>
<td>2007</td>
<td>26</td>
</tr>
<tr>
<td>3F3E</td>
<td>Leucine</td>
<td>O-o</td>
<td>1.80</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>2Q6H</td>
<td>Clomimpraine</td>
<td>O-occ</td>
<td>1.85</td>
<td>2007</td>
<td>26</td>
</tr>
<tr>
<td>2QE1</td>
<td>Alanine and clomimpraine</td>
<td>O-occ</td>
<td>1.85</td>
<td>2007</td>
<td>26</td>
</tr>
<tr>
<td>2QB4</td>
<td>Desipramine</td>
<td>O-occ</td>
<td>1.90</td>
<td>2007</td>
<td>26</td>
</tr>
<tr>
<td>3F48</td>
<td>Alanine</td>
<td>O-o</td>
<td>1.90</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3F4I</td>
<td>Selenomethione</td>
<td>O-o</td>
<td>1.95</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3GJD</td>
<td>OG (detergent)</td>
<td>O-occ</td>
<td>2.00</td>
<td>2009</td>
<td>28</td>
</tr>
<tr>
<td>3F3A</td>
<td>Tryptophan</td>
<td>O-o</td>
<td>2.00</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3F3C</td>
<td>4-fluro-phenylalanine</td>
<td>O-occ</td>
<td>2.10</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3USP</td>
<td>Heptyl-B-D-Selenglucose</td>
<td>O-occ</td>
<td>2.10</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3GWU</td>
<td>Sertraline</td>
<td>O-occ</td>
<td>2.14</td>
<td>2009</td>
<td>30</td>
</tr>
<tr>
<td>3F4J</td>
<td>Glycine</td>
<td>O-occ</td>
<td>2.15</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3MPN</td>
<td>F177R mutant</td>
<td></td>
<td>2.25</td>
<td>2010</td>
<td>31</td>
</tr>
<tr>
<td>3MPQ</td>
<td>I204R mutant</td>
<td></td>
<td>2.25</td>
<td>2010</td>
<td>31</td>
</tr>
<tr>
<td>3F3D</td>
<td>Methionine</td>
<td>O-occ</td>
<td>2.30</td>
<td>2008</td>
<td>27</td>
</tr>
<tr>
<td>3GWV</td>
<td>r-fluoxetine</td>
<td>O-occ</td>
<td>2.35</td>
<td>2009</td>
<td>30</td>
</tr>
<tr>
<td>3GWW</td>
<td>s-fluoxetine</td>
<td>O-occ</td>
<td>2.46</td>
<td>2009</td>
<td>30</td>
</tr>
<tr>
<td>3USG</td>
<td>Leucine (bicelle: C2)</td>
<td>O-occ</td>
<td>2.50</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3Q55</td>
<td>Tryptophan (I359Q)</td>
<td></td>
<td>2.60</td>
<td>2012</td>
<td>32</td>
</tr>
<tr>
<td>3Q54</td>
<td>Tryptophan (F259V)</td>
<td></td>
<td>2.63</td>
<td>2012</td>
<td>32</td>
</tr>
<tr>
<td>3USL</td>
<td>Selenomethione (bicelle: C2)</td>
<td>O-occ</td>
<td>2.71</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3Q56</td>
<td>Tryptophan (F259V, I359Q)</td>
<td>O-occ</td>
<td>2.80</td>
<td>2012</td>
<td>32</td>
</tr>
<tr>
<td>3GJC</td>
<td>OG (E290S)</td>
<td>O-occ</td>
<td>2.80</td>
<td>2009</td>
<td>28</td>
</tr>
<tr>
<td>2QJU</td>
<td>Desipramine</td>
<td></td>
<td>2.90</td>
<td>2007</td>
<td>33</td>
</tr>
<tr>
<td>3TU0</td>
<td>Alanine (T355V, S357A, K258A)</td>
<td></td>
<td>2.99</td>
<td>2012</td>
<td>32</td>
</tr>
<tr>
<td>3USM</td>
<td>Selenomethione (bicelle: C2)</td>
<td>O-occ</td>
<td>3.01</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3TT1</td>
<td>Fab</td>
<td>O-o</td>
<td>3.10</td>
<td>2012</td>
<td>34</td>
</tr>
<tr>
<td>3USI</td>
<td>Leucine (bicelle: P2)</td>
<td>O-occ</td>
<td>3.11</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3TT3</td>
<td>Fab</td>
<td>O-o</td>
<td>3.22</td>
<td>2012</td>
<td>34</td>
</tr>
<tr>
<td>3USJ</td>
<td>Leucine (bicelle: P21)</td>
<td>O-occ</td>
<td>3.50</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3USK</td>
<td>Leucine (bicelle: P21)</td>
<td>O-occ</td>
<td>4.50</td>
<td>2012</td>
<td>29</td>
</tr>
<tr>
<td>3USO</td>
<td>Selenomethione (bicelle: P2/P212)</td>
<td>O-occ</td>
<td>4.50</td>
<td>2012</td>
<td>29</td>
</tr>
</tbody>
</table>
1.6 The LeuT-Fold

As mentioned in Section 1.4, the number of available structures for proteins in the PDB is ever-increasing. Interestingly, the number of novel folds identified each year has recently apparently plateaued, as shown in Figure 4. Since 2009, no new unique folds have been identified. However, the structures of proteins that have been experimentally determined have been assigned to previously identified folds.

![Figure 4. Total and yearly additions of unique folds into the Protein Data Bank.](image)

Prior to the experimental isolation of the inward-facing state of LeuT, it was found that other prokaryotic homologues of the SLCs also shared a fold similar to LeuT. While the overall structure of LeuT is asymmetric and contains 12 TMs as shown in Figure 5, there are two similar structural motifs present, each containing five TMs. Repeat A is comprised of TMs 1-5, while TMs 1-6 represent Repeat B. The two repeats share little to no sequence similarity, but their structural conservation allows for a superposition of one another yielding a RMSD of 3.9Å. The two repeats are related by a twofold symmetry around an axis through the center of the membrane plane and named the "LeuT-Fold". The unwound TM1 (TM1a and TM1b) is complemented by the first TM of Repeat B, TM 6 (TMs 6a and 6b). Together, these unwound TMs form the binding site for sodium and leucine in LeuT.
Figure 5. Topology of LeuT: Repeat A (TMs 1-5) are highlighted in blue, while Repeat B (TMs 6-10) are highlighted in red. The binding site is shown with a purple star, surrounded by the two sodium binding sites shown in yellow. TMs 11 and 12 are peripheral and are not part of the inverted repeat scheme.

In the three-dimensional crystal structures, we find that the repeats of LeuT are interlaced which gives rise to a group of seven central TMs which line the binding site and putative permeation pathways: TMs 1, 2, 3, 6, 7, 8 and 10. The individual repeats and overall structure of LeuT is shown in Figure 6. By comparing just the first two TMs of each repeat (TMs 1/2 and TMs 6/7) reveal similar arrangements three-dimensionally. Additionally, a comparison of the last three TMs of each repeats shown a greater than expected similarity, than a comparison based wholly on the repeats. These insights led Forrest and Rudnick to identify TMs 1, 2, 6 and 7 as a "bundle" that would likely move together during conformational changes of transport. The remaining six TMs (3, 4, 5, 8, 9 and 10) thus comprise the "scaffold". Thus, to satisfy the alternating access mechanism, the transporter could simply change the tilt of the bundle, relative to the scaffold to achieve an outward facing and inward facing state.13
Figure 6. The LeuT-Fold is constructed of five inverted repeats. TMs 1-5 are shown in red, orange, yellow, green and purple tubes, while TMs 6-10 are shown in the same color and a ribbon representation. The top panel of images shows the repeats from a top-down view from the extracellular side of LeuT. The bottom panel of images shows the repeats from a side view.

The elegant symmetry of the LeuT-fold (or inverted repeats, or the 5 + 5 inverted repeats) has been identified in many secondary active transporters that share the alternating access mechanism as the working model of transport. Consequently, it is an emerging superfamily of sodium-dependent co-transporters of high interest. The family members may have a variable number of helices ranging from 10 to 12, yet they all share the five TM helix repeat motif as observed in LeuT. Their substrates bind near to the center of the transporter, at the unwound regions of TMs 1 and 6. Their sequence are hardly similar, yet superposition of the core helices yield a RMSD of less than 4.0 Å.

In recent years, transporters that share the LeuT-Fold (namely: Mph1, BetP, ApcT, CaiT, AdiC and vSLGT) have been isolated in a variety of conformations.
detailed in Table 4. With experimentally determined structure of conformations that
reflect each conformation through the transport cycle, insights can be gained into how the
molecular pumps utilize the sodium gradient to pump the substrate from the extracellular
side to the cytoplasm.

Table 4. Transporters and their associated PDB codes, which reflect different
conformations along the transport cycle for the alternating access mechanism.

<table>
<thead>
<tr>
<th>Outward</th>
<th>Apo- Open</th>
<th>Mhp1 (2JLN), AdiC (3LRB, 3NCY)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Substrate Bound</td>
<td>AdiC (3OB6)</td>
</tr>
<tr>
<td></td>
<td>Bound- Occluded</td>
<td>LeuT (2A65), Mhp1 (2JLO), AdiC (3L1L)</td>
</tr>
<tr>
<td>Intermediate-</td>
<td>Fully Occluded</td>
<td>BetP (2WIT)</td>
</tr>
<tr>
<td>Bound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inward</td>
<td>Bound- Occluded</td>
<td>vSLGT (3DH4)</td>
</tr>
<tr>
<td></td>
<td>Substrate Bound</td>
<td>CaiT (3HFX, 2WSX) BetP (3PO3)</td>
</tr>
<tr>
<td></td>
<td>Apo- Open</td>
<td>Mhp1 (2X79), CaiT (2WSW)</td>
</tr>
<tr>
<td>Intermediate-</td>
<td>Fully Occluded</td>
<td>ApcT (3GIA)</td>
</tr>
<tr>
<td>Apo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1.7 Summary

While crystallography has provided invaluable starting structures and snapshots
throughout the transport cycle, there still remain open questions regarding the dynamics
of the process of transport. For example, in sodium-dependent co-transporters, in what
order do the ions and substrate bind to the outward-facing state? What are the structural
changes (generally, and in each transporter) that occur in response to substrate and ion
binding? What atomic-level events must be triggered to allow for conformational change
in the transporter? In order to answer these questions, a focus must be placed on solving
the structure of a single transporter in each of the steps of the transport cycle. These
starting structures can then spawn further biochemical and computational studies to
bridge the gaps between each individual state, gathering critical dynamic data.
Furthermore, those same questions can be applied to human secondary active transporters. A disadvantage in studying human secondary active transporters is a lack of crystallographic data for their structures. Prokaryotic templates have been used in order to study the eukaryotic homologues. However, how do the principles and insights gained from studying prokaryotic secondary active transporters translate to their eukaryotic homologues? Indeed, in order to fully understand and appreciate the biological and pharmacological properties unique to human secondary transporters, we will need to solve the crystal structures of eukaryotic homologues.

1.8 References


CHAPTER TWO

INSIGHTS FROM MOLECULAR DYNAMICS: THE BINDING SITE OF COCAINE IN THE DOPAMINE TRANSPORTER AND PERMEATION PATHWAYS OF SUBSTRATES IN THE LEUCINE AND DOPAMINE TRANSPORTERS

2.1 Abstract

The dopamine transporter (DAT) facilitates the regulation of synaptic neurotransmitter levels. As a target for therapeutic and illicit psycho-stimulant drugs like antidepressants and cocaine, DAT has been studied intensively. Despite a wealth of mutational and physiological data regarding DAT, the structure remains unsolved and details of the transport mechanism, binding sites and conformational changes remain debated. A bacterial homologue of DAT, the leucine transporter (LeuT<sub>Aa</sub>) has been used as a template and framework for modeling and understanding DAT. Free energy profiles obtained from Multi-Configuration Thermodynamic Integration allowed us to correctly identify the primary and secondary binding pockets of LeuT<sub>Aa</sub>. A comparison of free energy profiles for dopamine and cocaine in DAT suggests that the binding site of cocaine is located in a secondary pocket, not the primary substrate site. Two recurring primary pathways for intracellular substrate release from the primary pocket are identified in both transporters using the Random Acceleration Molecular Dynamics method. One pathway appears to follow transmembranes (TMs) 1a and 6b while the other pathway follows along TMs 6b and 8. Interestingly, we observe that a single sodium ion is co-transported with leucine during both simulations types.
2.2 Significance and Function of the MATs

In order for an electrical message to pass from one neuron to another, small neurotransmitter molecules are responsible for getting the message across the chemical synapse. Figure 7 illustrates the chemical synapse. Neurotransmitters such as dopamine, serotonin and norepinephrine are released from the presynaptic nerve via vesicles. Upon release, the neurotransmitters diffuse into the synaptic cleft and are free to bind to their respective receptors on the postsynaptic nerve. Integral membrane proteins located on the

![Figure 7. Regulation of neurotransmitter levels at the chemical synapse by monoamine transporters.](image)
presynaptic nerve are responsible for regulating the levels of neurotransmitters in the synaptic cleft. The transporters shuttle their respective neurotransmitter molecules from the synaptic cleft for reuse or destruction into the presynaptic nerve.

The neurotransmitters dopamine, serotonin and norepinephrine are known as monoamines. Their respective transporters, DAT, SERT and NET are thus classified as monoamine transporters (MATs). Dysfunction of the MATs can have devastating effects on motor function, memory and mood. Diseases and disorders such as: depression, obsessive-compulsive disorder (OCD), epilepsy, autism, attention-deficit hyperactivity disorder (ADHD), Parkinson’s and Alzheimer’s diseases are associated with the dysfunction of the MATs. Additionally, the MATs are targets for the therapeutic drugs used to treat these diseases and disorders, as well as illicit drugs such as cocaine and amphetamine. In DAT, it is believed that cocaine blocks the re-uptake of dopamine from the synaptic cleft, thereby allowing dopamine neurotransmitters to accumulate in the nucleus accumbens, the “pleasure center” of the brain.

DAT (SLC6A3), NET (SLC6A2) and SERT (SLC6A4) belong to the solute carrier 6 (SLC6) gene family. The membrane-bound MATs are also classified as neurotransmitter:sodium symporters (NSS) which rely on the naturally occurring neuronal Na$^+$ gradient to drive the transport of substrates against their concentration gradient. Despite a lack of three-dimensional crystallographic data for the MATs, they are believed to contain 12 transmembrane (TM) helices. Hypotheses regarding their structure have recently been bolstered by the publication of a related transporter, addressed in the next section.
2.3 The Leucine Transporter, LeuT\textsubscript{Aa}

2.3.1 Structure and Relationship to MATs

In 2005, Yamashita \textit{et al.}\textsuperscript{12} published the crystal structure of bacterial MAT homologue, the leucine transporter (LeuT\textsubscript{Aa}) shown in Figure 8. The crystal structure presented a 12 transmembrane (TM) integral membrane protein. LeuT\textsubscript{Aa} exhibits 20-25\% sequence identity and 40-45\% sequence similarity to the MATs.\textsuperscript{13} Despite a relatively low sequence identity, the disparities in the sequences of MATs and LeuT\textsubscript{Aa} are mainly located in the loop regions, while the binding regions (TMs 1 and 6) display the highest degrees of sequence identity and similarity.\textsuperscript{14} Identical and similar residues of LeuT\textsubscript{Aa} and DAT (a focus of this work) are highlighted in Figure 9.

The publication of the LeuT\textsubscript{Aa} crystal structure\textsuperscript{12} provided an invigorating breakthrough for the study of MATs as there is no direct crystallographic data of the MATs. With the structure of LeuT\textsubscript{Aa} now available, computational techniques can and have been applied to the MATs by using LeuT\textsubscript{Aa} as a template. There have been many studies\textsuperscript{11,14-29} that set out to investigate the molecular principles behind the transport mechanism, conformational changes, pathways, and sites for the permeation of ligands through the MATs and LeuT\textsubscript{Aa}. Despite this, the fundamental mechanism of transport has not been clearly elucidated for the MATs or their homologue, LeuT\textsubscript{Aa}.  


Figure 8. Crystal structure of LeuT\textsubscript{Aa} (PDB 2A65) embedded in a lipid bilayer. The 12 transmembrane helices are colored as ribbons: TM1, red; TM2, dark orange; TM3, light orange; TM4, yellow; TM5, lime-green; TM6, sea-foam green; TM7, green-cyan; TM8, cyan; TM9, navy-blue; TM10, violet; TM11, purple; TM12, mauve.
2.3.2 Primary Binding Pocket

In the 2A65 structure, leucine was found bound midway through the transporter at what is believed to be the primary binding pocket. Bound with leucine are two sodium atoms, Na1 and Na2. Na1 is directly coordinated to the leucine substrate, in addition to 1A22, 1N27, 6T254 and 7N286. (This nomenclature identifies the TM helix first, then the residue name and number. For example, 1G20 indicates glycine 20 on TM 1.) Na2, roughly 7.0Å away from Na1, is coordinated through interactions with 1G20, 1V23,
Based on the original crystal structure, Yamashita et al. proposed that the sodium ions are required for transport as they serve an organizational role for leucine in the binding site.

Further studies have attempted to tease out the exact roles of Na1 and Na2. One study suggested that the Na2 binding site is a structural site that plays a role in substrate binding and local pocket formation. In another study, in which an inward-facing state of LeuT\textsubscript{Aa} was generated, Na2 was observed to be released first (leucine and Na1 release were not observed) to the intracellular milieu after conformational rearrangement (through targeted molecular dynamics). Yet another study showed that binding of Na\textsuperscript{+} in the Na2 site is required to enhance the ion selectivity of Na\textsuperscript{+} in the Na1 site. Another study suggested that the binding and unbinding of sodium at the Na2 site allows the substrate to shift the conformational equilibrium needed for transport by altering the energetic requirements. In Chapter Three, we will discuss plausible roles for Na1 and Na2 in terms of the transport mechanism of LeuT\textsubscript{Aa}.

2.3.4 Proposed Mechanism

To transport substrates across the membrane, it has been proposed that LeuT\textsubscript{Aa} follows an alternating access mechanism. This model proposes that the substrate binds to one side of the membrane protein, while some energy input drives a conformational change in the protein allowing the substrate to be transported to the other milieu. This transport is proposed to occur if and only if all co-transported components are bound, and not when the primary pocket is partially filled. LeuT\textsubscript{Aa} thus alternates between the outward-facing-open (OF-o) and inward-facing-open (IF-o) states, where
2.3.5 A Secondary Binding Pocket

The transport picture of LeuT_{Aa}, and by extension, the MATs, is further clouded by a disparity in results regarding the number of substrates required for transport. Recently, two independent studies have reported on the transport of leucine in LeuT_{Aa}. One study indicated that only one substrate is necessary for transport, while the other study concluded that two substrates are needed for transport. The disparity is detailed in a recent review on the topic.\textsuperscript{41}
The hypothesis of a two-substrate transport mechanism arises from the fact that two pockets have been identified in LeuT<sub>As</sub>. There exists the primary substrate pocket buried midway through the transporter. A second pocket often referred to throughout the literature as the extracellular vestibule, secondary pocket, inhibitor pocket or (the notation we will use here) the S2 pocket is located roughly 10Å toward the extracellular milieu. (By this naming convention, the primary pocket becomes the S1 pocket.) The aromatic lid formed by 3Y108 and 6F253 as well as the extracellular salt bridge formed...

Figure 11. Close-up of the S1 and S2 sites in LeuT<sub>As</sub>. The crystal structure of 2Q6H is shown, which is LeuT<sub>As</sub> with leucine bound in the S1 pocket and clomipramine bound in the S2 pocket. The S1 pocket is highlighted with an red molecular surface and the S2 pocket is highlighted with an orange molecular surface. The ligands are displayed in gray ball-and-stick models. The aromatic lid formed by 3Y108 and 6F253 as well as the extracellular salt bridge formed by 10D404 and 1R30 are shown in yellow and green stick models, respectively. LeuT<sub>As</sub> is shown in blue ribbons.
by 1R30 and 10D404 delineates these two pockets, highlighted in Figure 11. In DAT, the location of the S1 and S2 pockets are anticipated to be analogous separated by a corresponding aromatic lid (3Y156/6F319) and salt bridge (1R85/10D475).

In terms of the S1 and S2 sites of DAT, the binding mode and pocket of cocaine has long been studied computationally and experimentally.$^{11,15,28,33,42-49}$ Two recent, independent studies present data regarding the binding location of cocaine in DAT.$^{47,49}$ Beuming $et$ $al.$$^{47}$ performed binding assays and observed cocaine binding to be competitive with dopamine binding. Based on a molecular model and site-directed mutagenesis, they propose that the binding site of cocaine in DAT overlaps with the substrate binding pocket of dopamine. An independent study by Huang $et$ $al.$$^{49}$ involving molecular modeling and dynamics simulations suggests that the cocaine binding site is close to, but does not overlap with, the dopamine binding site of DAT. Huang $et$ $al.$ identify the binding pocket of cocaine to be in the “substrate-entry tunnel” (analogous to the S2 site) from the extracellular milieu composed of residues from TMs 1, 3, 6, 8 and 10. They suggest this binding pocket to be, at minimum, the initial binding site for cocaine in DAT. They point out that DAT can naturally accommodate cocaine in this pocket. DAT would need to undergo significant conformational rearrangements in order to expand the substrate pocket to accommodate for the large size of cocaine, compared to dopamine.$^{49}$
2.4 Methods

2.4.1 LeuT_Aa Structure for MCTI Simulations

The coordinates of LeuT_Aa were obtained from the Protein Data Bank entry 2A65 (www.rcsb.org, MMDB accession no. 34395) corresponding to the bacterial (Aquifex aeolicus) leucine transporter. This particular structure is missing residues el2N133 and el2A134, in addition to four residues at each N- and C termini. el2N133 and el2A134 were reintroduced into the structure using the model builder module in Molecular Operating Environment 2008.10. Two structures were generated for use in MCTI simulations. The first was only the protein in medium (in vacuum environment); the second structure included the protein in a water sphere with a 25Å radius. The total system sizes were 8,321 atoms and 36,816 atoms, respectively. These environment choices were made in order to reduce system size and subsequent computational cost of MD simulations.

2.4.2 DAT Structure for MCTI Simulations

Construction of the molecular model for DAT has been described in detail elsewhere. Briefly, the FASTA sequence of the rat DAT protein (SwissProt locus SC6A3-RAT; accession number P23977; NCBI accession number AAB21099) was used as the query in three independent sequence alignments with LeuT_Aa: Robetta server, 3D-JIGSAW server and the alignment of Yamashita et al. The model from the Robetta server was chosen as it produced the best docking energies for dopamine and amphetamine in the software program MOE. The AMBER99 all-atom force field was used to add hydrogen atoms and assign partial charges. The newly added hydrogen
atoms were relaxed via several cycles of energy minimization using a conjugated
gradient/truncated Newton optimization algorithm. The convergence criteria for final
sidechain refinement was carried out with AMBER99: 0.05 kcal/mol and a dielectric
constant $\varepsilon = 3$. Since the 2A65 structure was used as a template, the final DAT model
represents and outward-facing conformation. Additional dynamics of the DAT molecular
model were collected,\textsuperscript{11} which totaled 35 ns of production dynamics of DAT embedded
in a bilayer (described below). The coordinates of DAT were obtained from the 32 ns
structure and isolated for use in the MCTI simulations, in vacuum.

2.4.3 Solvated Protein and Lipid Bilayer Construction for RAMD

Simulations

Construction of the embedded DAT and LeuT\textsubscript{Aa} systems has been describe
elsewhere.\textsuperscript{11} Briefly, the LeuT\textsubscript{Aa} crystal structure and DAT outward-facing model were
embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE)
membrane using the membrane builder plugin of Visual Molecular Dynamics (VMD).\textsuperscript{53}
The protein and POPE membrane were solvated with approximately 32,000 TIP3 waters.
Sodium and chloride ions were added to the system, yielding an overall neutral system at
approximately 0.2 M NaCl. The 32 ns structure (also used for MCTI simulations) if the
DAT molecular model was obtained from\textsuperscript{11} and used for all RAMD simulations
involving the dopamine transporter.

One LeuT\textsubscript{Aa} structure was obtained from\textsuperscript{54} which reflects the initial 2A65
structure after 250 ns of accelerated molecular dynamics (aMD).\textsuperscript{55} The final protein
structure from the aMD simulation was used in this study for RAMD simulations.
Coordinates of two inward-facing LeuT\textsubscript{Aa} models were obtained from the Supplemental
These will be referred to as M1 and M2. Each of these models were embedded in a POPE membrane and solvated as described above. Structures were minimized, equilibrated and put through a production run of 1 ns using a 2 fs timestep at a constant temperature (310K) and pressure (1 atm) before using them for RAMD simulations.

2.4.4 Multi-Configuration Thermodynamic Integration

To study the proposed permeation pathway of LeuT<sub>AA</sub> and DAT, the Multi-Configuration Thermodynamic Integration (MCTI) method<sup>56</sup> was utilized. MCTI has been used to compute the potential of mean force (PMF) around the φ and ψ angles of alanine dipeptide.<sup>57</sup> Additionally, this method has been used to calculate the free energy barrier for the OH transfer step of the hydroxylation reaction of p-hydroxybenzoate in the enzyme p-hydroxybenzoate hydroxylase (PHBH).<sup>58</sup> The method has also been used to investigate the free energy surfaces of alanine (Ala) and R-methylalanine (Aib) homopeptides.<sup>59</sup>

The MCTI method uses MD simulations to generate configurations. The method then uses these configurations to smoothly transform one state to another while calculating the change in free energy. The Gibbs free energy difference between the two states is described by the Hamiltonians H(λ=0) (i.e. ligand in S1/S2) and H(λ=1) (i.e. ligand at extracellular/intracellular milieu) and can be obtained from the integral:

$$\Delta G = \int_0^1 \left( \frac{\delta H}{\delta \lambda} \right) d\lambda$$

Equation 1.
where $\Delta G$ is the change in Gibbs energy, $\lambda$ is a coupling parameter and $\langle \frac{\delta H}{\delta \lambda} \rangle$ is the ensemble average of the partial derivative of the Hamiltonian with respect to the partial derivatives of $\lambda$. By introducing the coupling parameter $\lambda$, the Hamiltonian is thus $\lambda$-dependent thereby connecting $\lambda$ to the position and momenta of the particles. Since molecular dynamics are done in discrete steps, the integral is evaluated as a sum of the ensemble averages:

$$\Delta G \approx \sum_i \langle \frac{\delta H}{\delta \lambda} \rangle \Delta \lambda_i$$

Equation 2.

where $i$ counts over the values of $\lambda$ and $\Delta \lambda_i$ is the difference between each value of $\lambda$.

This change in $\lambda$ is carried out at discrete steps called windows. At each window, a molecular dynamics simulation is performed which consists of an equilibration period, followed by a data accumulation stage in order to calculate the free energy.$^{56}$

Transport of ligands through both the leucine and dopamine transporters was simulated. For $\text{LeuT}_{\text{Aa}}$;Leu and $\text{DAT:DA}$ complexes, the substrate was bound in the primary (S1) pocket of the protein. However, in the $\text{DAT:COC}$ complex, cocaine was docked into the S2 pocket for the starting structure. For each complex, the structure was equilibrated at a constant temperature (310K) and pressure (1 atm) for 500 ps prior to use in MCTI simulations.

For each complex, a permeation pathway for the substrate was selected from a pocket (S1 or S2) to either the extracellular or intracellular milieu. Two vectors were defined for each complex. The first vector was defined from the geometric center of the substrate through the midpoint of the extracellular gates ($1R30/10D404$ in $\text{LeuT}_{\text{Aa}}$,}
The second vector was defined from the geometric center of the substrate through the midpoints of the intracellular gates (1R5/8D369 in LeuT<sub>Aa</sub>, 1R60/8D435 in DAT). These fixed vectors were determined through the use of VMD 1.8.7. Geometric centers of the gate residues were determined in VMD, and the midpoint of the two gate residues were manually calculated. The lengths of the vectors were extended at least 7Å past the protein. This is depicted in Figure 12. The LeuT<sub>Aa</sub>:Leu complex for MCTI simulations.

Figure 12. A graphic representation of the LeuT<sub>Aa</sub>:Leu complex for MCTI simulations. The substrate, leucine is shown in the primary pocket in VDW spheres. Intracellular (1R5/8D369) and extracellular (1R30/10D404) salt bridges are highlighted in a black CPK representation. The proposed permeation pathway from the pocket to each milieu is shown with a blue vector. The 12 transmembrane helices are colored as ribbons: TM1, red; TM2, dark orange; TM3, light orange; TM4, yellow; TM5, lime-green; TM6, seafoam green; TM7, green-cyan; TM8, cyan; TM9, navy-blue; TM10, violet; TM11, purple; TM12, mauve. This coloring scheme for TM helices will be used throughout the figures, unless otherwise noted.
complex was simulated in both a vacuum and water sphere environment. The DAT:DA and DAT:COC complexes were simulated in vacuum only.

MCTI simulations were carried out using the implementation found in NAMD 2.6. The z-distance option was used to incrementally move the substrates (leucine and dopamine) from the primary (S1) pockets of LeuT\(_{Aa}\) and DAT, respectively. In the DAT:COC complex, cocaine was incrementally moved from the S2. For each permeation pathway, the vector was split into 250 windows. This resulted in the substrate being moved out of the pocket in 0.18Å increments on average, through the respective gate and exiting either milieu. An accumulation time of 60 ps was used, with 20 ps of relaxation used between each window.

Each MCTI simulation from the pocket to either milieu was 40 ns. For each setup, a total of twenty MCTI simulations were carried out; ten simulations where the substrate was incrementally moved from the pocket to the extracellular milieu and ten simulations where the substrate was incrementally moved from the pocket to the intracellular milieu. In total, 1.6 \(\mu\)s of MCTI MD trajectory data was collected for each transporter. Multiple replicates of the simulations were executed such that the resulting free energy profiles could be averaged and overall noise in the data could be reduced. Simulations were carried out through TeraGrid resources provided by the Pittsburgh Supercomputing Center and National Institute for Computational Sciences.

2.4.5 Random Acceleration Molecular Dynamics

The Random Acceleration Molecular Dynamics (RAMD)\(^{61,62}\) method, implemented in NAMD\(^{60}\) was used to investigate the unbiased entrance and exit
pathways for a variety of ligands and structures relating to the MATs and LeuT. 

Previously, the RAMD method has been used to identify egress routes for ligands from buried active sites in Cytochrome P450cam, retinoic acid receptor, liver fatty acid binding protein, the vitamin D receptor and the β2-adrenergic receptor.

In the RAMD method, a force is applied to the ligand in order to explore pathways. The force constant, $k$ is chosen by the user, and the unit vector, $\vec{r}_0$ is obtained from a random number generator. The force on the ligand is then given by:

$$\vec{F} = k\vec{r}_0$$

Equation 3.

The force, $\vec{F}$ is maintained on the ligand for a specified number of steps, $m$, so long as the ligand maintains a threshold velocity, given by:

$$v_{min} = \frac{r_{min}}{\Delta t}$$

Equation 4.

The threshold velocity, $v_{min}$ has units of Å/fs, where the $r_{min}$ value is determined by the user, and $\Delta t$ indicates the timestep used in the simulation.

In the event the velocity of the ligand falls below the threshold velocity, $v_{min}$, the direction in which the force is applied will be changed. A new (random) unit vector will be chosen by the random number generator and applied to Equation 3. This process repeats until the ligand escapes from the complex. In comparison to the MCTI method where a single, fixed pathway was study, this method allows for an unbiased search for escape routes of the ligand from the complex. Furthermore, the RAMD method enables nanosecond timescales for ligand exit as dissociation kinetics are accelerated in a given RAMD simulation.
RAMD simulations of the following LeuT<sub>Aa</sub> and DAT were carried out. Several LeuT<sub>Aa</sub> structures were chosen for this study. In R1, the substrate leucine (Leu) was placed only in the S1 pocket. In R2, Leu was docked into both the S1 and S2 pockets. In R3 and R4, Leu was placed only in the S1 pocket. Lastly, R5 represents the M1 inward-facing model with Leu bound at S1 and S2. In all simulations involving LeuT<sub>Aa</sub>, the force in RAMD simulations was applied to the Leu in S1. These simulations were carried out in the lipid bilayer and solvent system described earlier.

Table 5 summarizes the various simulation setups. In R1 and R2, the LeuT<sub>Aa</sub> structure after 250 ns of accelerated MD was used. In R1, the substrate leucine (Leu) was placed only in the S1 pocket. In R2, Leu was docked into both the S1 and S2 pockets. In R3 and R4, Leu was placed only in the S1 pocket. Lastly, R5 represents the M1 inward-facing model with Leu bound at S1 and S2. In all simulations involving LeuT<sub>Aa</sub>, the force in RAMD simulations was applied to the Leu in S1. These simulations were carried out in the lipid bilayer and solvent system described earlier.

For the dopamine transporter, RAMD simulations were carried such that the inhibitors amphetamine (AMA) and cocaine (COC) were docked into the S2 pocket, while the S1 pocket remained empty (R7 and R8, respectively). Dopamine (DA) was docked in S1 pocket for setup R6. For simulations R9-11, dopamine was docked into the S1 pocket for each, while dopamine, amphetamine or cocaine were docked into the S2 pocket. As in the case for the LeuT<sub>Aa</sub>:Leu:Leu complexes, the substrate in the S1 pocket was the chosen substrate for the force to be applied.
Table 5. Summary of RAMD simulation details and parameters. All RAMD simulations were carried out with an acceleration of 0.1 kcal/mol Å. Occupancy of the S1 and S2 pockets is noted. In all cases (with the exception of R7 and R8) the force in the RAMD simulations was applied to the substrate found in S1. In R7 and R8, no substrate was docked in S1, and thus the force was applied to amphetamine and cocaine in the S2 pockets.

<table>
<thead>
<tr>
<th>Setup</th>
<th>Transporter</th>
<th>S1</th>
<th>S2</th>
<th># of Trajectories</th>
<th># Extra. Egresses</th>
<th># Intra. Egresses</th>
<th># Failed Escapes</th>
<th>Simulation Length Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>LeuT&lt;sub&gt;aa&lt;/sub&gt;</td>
<td>Leu</td>
<td>-</td>
<td>40</td>
<td>7 (17.5%)</td>
<td>31 (77.5%)</td>
<td>2 (5%)</td>
<td>0.1 - 0.7 ns</td>
</tr>
<tr>
<td>R2</td>
<td>LeuT&lt;sub&gt;aa&lt;/sub&gt;</td>
<td>Leu</td>
<td>Leu</td>
<td>50</td>
<td>0</td>
<td>27 (54%)</td>
<td>31 (6%)</td>
<td>23 ps - 0.3 ns</td>
</tr>
<tr>
<td>R3</td>
<td>Leu&lt;sub&gt;aa&lt;/sub&gt; M1</td>
<td>Leu</td>
<td>-</td>
<td>49</td>
<td>0</td>
<td>46 (94%)</td>
<td>3 (6%)</td>
<td>23 ps - 0.3 ns</td>
</tr>
<tr>
<td>R4</td>
<td>Leu&lt;sub&gt;aa&lt;/sub&gt; M2</td>
<td>Leu</td>
<td>-</td>
<td>50</td>
<td>0</td>
<td>46 (92%)</td>
<td>4 (8%)</td>
<td>28 ps - 0.3 ns</td>
</tr>
<tr>
<td>R5</td>
<td>Leu&lt;sub&gt;aa&lt;/sub&gt; M1</td>
<td>Leu</td>
<td>Leu</td>
<td>50</td>
<td>0</td>
<td>49 (98%)</td>
<td>1 (2%)</td>
<td>34 ps - 1 ns</td>
</tr>
<tr>
<td>R6</td>
<td>DAT</td>
<td>DA</td>
<td>-</td>
<td>50</td>
<td>18 (36%)</td>
<td>28 (56%)</td>
<td>4 (8%)</td>
<td>0.1 - 0.7 ns</td>
</tr>
<tr>
<td>R7</td>
<td>DAT</td>
<td>-</td>
<td>AMA</td>
<td>50</td>
<td>44 (88%)</td>
<td>3 (6%)</td>
<td>3 (6%)</td>
<td>46 ps - 1 ns</td>
</tr>
<tr>
<td>R8</td>
<td>DAT</td>
<td>-</td>
<td>COC</td>
<td>50</td>
<td>50 (100%)</td>
<td>0</td>
<td>0</td>
<td>24 - 31 ps</td>
</tr>
<tr>
<td>R9</td>
<td>DAT</td>
<td>DA</td>
<td>DA</td>
<td>49</td>
<td>24 (49%)</td>
<td>20 (41%)</td>
<td>5 (10%)</td>
<td>80 ps - 0.9 ns</td>
</tr>
<tr>
<td>R10</td>
<td>DAT</td>
<td>DA</td>
<td>AMA</td>
<td>50</td>
<td>32 (64%)</td>
<td>10 (20%)</td>
<td>8 (16%)</td>
<td>80 ps - 0.7 ns</td>
</tr>
<tr>
<td>R11</td>
<td>DAT</td>
<td>DA</td>
<td>COC</td>
<td>49</td>
<td>22 (45%)</td>
<td>11 (22%)</td>
<td>16 (33%)</td>
<td>6 ps - 0.1 ns</td>
</tr>
</tbody>
</table>

Between 40 and 50 replicates were executed for each setup, as indicated in Table 5. As with the MCTI simulations, the structure of each complex was equilibrated at a constant temperature (310K) and pressure (1 atm) for 500 ps before performing the RAMD simulations. All RAMD simulations were also carried out at a constant temperature of 310 K, constant pressure of 1 atm, using periodic boundary conditions for the system and timestep of 2 fs. A force constant of $k=0.1$ kcal/mol Å was used for all simulations. This force was maintained for $m=50$ steps and the $r_{min}$ distance was set to 0.02Å. Thus, the threshold velocity, as given by Equation 4, is equal to 0.01 Å/fs. The escape criterion for the simulations was set to a distance of 40 Å from the initial center of mass for the ligand. This distance is chosen based on the overall size and shape of the transporter. Thus, when the ligand reached a distance greater than 40 Å from the initial starting position, the RAMD simulation would be completed. Simulations were carried out using NAMD 2.7b2<sup>60</sup> through TeraGrid resources provided by the National Institute for Computational Sciences.
2.5 Results and Discussion

2.5.1 MCTI

The average free energy score as substrates are moved along the proposed permeation pathway are presented in Figure 13 and Figure 14 for the following complexes: LeuT$_{Aa}$:Leu, DAT:DA and DAT:COC. Note that we are not reporting quantitative numbers; rather, we use this as a qualitative approach to identify stop-over sites and sites of interest in the transporters. The timescale of the transport process is estimated to be on the order of 1 ms$^{12}$ and each individual MCTI simulation was only 40 ns in length. These simulations are informative; however they do not allow for significant conformational changes in the protein. The sites labeled A-H in Figure 13 and Figure 14 were isolated from the MCTI trajectory files as sites of interest and analyzed in MOE 2009.10$^{51}$ to identify the interacting residues of the transporter with the substrate at that site. MOE 2009.10 was used to generate all 2D interaction maps for the substrate at the sites.

**Low Free Energy Scores Correspond to S1 and S2 in LeuT$_{Aa}$**

In both environments studied for the LeuT$_{Aa}$:Leu complex, relative low free energy scores are observed for the primary (S1) and secondary (S2) substrate pockets, as indicated by site C (S2) and site E (S1) in Figure 13. An increase in free energy score is observed in both environments as Leu is moved from the S2 pocket to the S1 pocket, which can be attributed to breaking of the aromatic lid (3Y108/6F253) that seals off the S1 site from the extracellular milieu. Substrate movement from S1 toward the
Figure 13. Free energies profiles of the substrate leucine through LeuT$_{Aa}$, as calculated via the MCTI method in both environments (vacuum and water sphere). Identified stopover sites (points of interest) are labeled A-H with horizontal lines depicting their general position in their respective transporter.

Intracellular milieu (site F) results in an increase in free energy score, attributed to disruption of binding pocket residue interactions. An increase and relative positive free energy score is also observed in both environments (although it is more evident in the vacuum environment) when leucine passes through intracellular gate (1R5/8D369) as represented by stopover site H.

Investigation of interacting residues, in conjunction with the free energy profiles, confirms that the S2 pocket identified in this study is in agreement with work done by
Quick et al.\textsuperscript{22} While there are difference in the water sphere and vacuum simulations for the LeuT\textsubscript{Aa}:Leu complex, the distinct similarities, particularly those of the S1 and S2 sites, gives us confidence that this method can be effectively used to investigate complexes of dopamine and cocaine with a previously validated DAT molecular model.\textsuperscript{15}

**Profiles of dopamine and cocaine in DAT differ at S1**

Similar features are observed in the DAT:DA free energy profile as in the LeuT\textsubscript{Aa}:Leu profiles. Low free energy scores are observed for dopamine at both S1 (site E) and S2 (site D), as is expected for the natural substrate of DAT. For DA, the magnitudes of those free energy scores are similar. A free energy score barrier is observed between those sites, indicating the breakage of the aromatic lid (3Y156/6F319) separating S1 and S2 in DAT.

An important feature of the DAT:COC profile is the striking difference as the inhibitor is moved toward the intracellular milieu from the S2 pocket. Here, a positive free energy score for cocaine in the intracellular half of DAT is expected as cocaine is an inhibitor that is not transported. A relative low free energy score at the S2 site is observed (site D) for cocaine in DAT. However, a higher free energy score is observed for cocaine at the S1 pocket of DAT. Recalling that free energy scores for DA at S1 and S2 were low, and similar, these results suggest that S2 is a more favorable pocket for cocaine in DAT, compared to the S1 pocket.
Figure 14. Free energies profiles of the substrate dopamine and inhibitor cocaine in DAT, as calculated via the MCTI method in vacuum. Identified stopover sites (points of interest) are labeled A-H with horizontal lines depicting their general position in their respective transporter.

*Cocaine is not stabilized by S1 residues in DAT*

Further investigation of the DAT:DA and DAT:COC trajectories and coordinates reveal that cocaine does not participate in side-chain interactions with DAT beyond site C labeled on the DAT:COC free energy profile in Figure 14. Coordinates of dopamine and cocaine in S1 of DAT were extracted from trajectory frames corresponding to the local minima present at stopover site E which reflects the S1 site in DAT. The extracted coordinates were then used for a short MD simulation (1ns) to allow for the macromolecule and amino acid side-chains to relax.
The final structures were then loaded into MOE 2009.10, where 2D ligand interactions plots were generated, see Figure 15. At the S1 site, dopamine is stabilized by interactions with 1F76, 1D79, 6V327 and 8S421 and fits comfortably in the binding cavity. In comparison, cocaine at the S1 site is not stabilized by any amino acid side-chain interactions and appears to be quite large for the given cavity size. These results further suggest that the binding site of cocaine is located above the aromatic lid (3Y156/6F319) of DAT in what Huang et al. refer to as the “substrate-entry tunnel”, or what we refer to here as the S2 site.

**Pocket volumes of S1 and S2 in DAT**

The volumes of the S1 and S2 pockets in the presence of dopamine and cocaine were analyzed from extracted coordinates of the ligands at S1 and S2 pockets. Ligand and amino acid side-chains of the four structures were allowed to relax for 1 ns. After MD, the ligands were removed and only the protein structure were used for analysis using POcket Volume MEasurer (POVME). In this method, a volume-grid file is generated by defining an inclusion sphere centered about a defined region that encompasses the binding pocket. Points from the volume-grid file that are close to protein atoms are deleted, leaving points that define the binding pocket. Prior to deleting the ligands, their geometric centers were obtained and used as the centers of the inclusion spheres of size 10 Å. A grid-spacing of 1.0 Å was used while a padding of 1.09 Å was used.
Figure 15. Residue interactions of DAT with dopamine (A) and cocaine (B) at the primary (S1) pocket. Coordinates of dopamine and cocaine in S1 of DAT were extracted from trajectory frames corresponding to the local minima present at stopover site E which corresponds to the S1 site. The apparent lack of residue interactions with cocaine at S1 suggests that cocaine does not favorably bind at the same pocket as dopamine in DAT. 2D interaction maps were generated using MOE 2009.10.

The S1 pocket of the DAT:DA complex occupies a volume of 140 Å³. In order for cocaine to bind in the same pocket, S1 must expand to nearly twice the volume, 261 Å³. Analysis of the S2 pocket provided an interesting insight. For the DAT:DA complex, the S2 pocket volume is comparable to the S1 pocket: 141 Å³. The S2 pocket of DAT allowed DA to nestle into the S2 pocket with a similar pocket volume as in S1. By comparison, the pocket volume of S2 in the DAT:COC complex is far greater: 305 Å³. The general flexibility of the “substrate-entry tunnel” and subsequent S2 pocket allows for a bulky inhibitor such as cocaine to bind to DAT, while the interior of the protein at S1 does not provide the flexibility necessary to accommodate and coordinate cocaine.
Co-transport of Na1 with leucine

During all LeuT_{Aa}-Leu MCTI simulations, we observed an interesting escorting behavior of the Na\(^+\) of Na1 site. Movement of sodium in the Na2 binding pocket was not observed. In Figure 16, we present snapshots of the simulated transport of leucine in LeuT_{Aa}. Leucine at each of the sites of interest (A-H) from Figure 13 is shown. Key interacting residues are labeled in each of the insets in addition to the positions of Na1 and Na2. Throughout the transport process, Na1 remains close to leucine while leucine was moved from the binding pocket to either milieu. It is important to note that the MCTI method was not used to move either sodium, and that indeed Na1 was co-transported with leucine during simulated transport.

Interestingly, in the case of DAT MCTI simulations, co-transport of sodium with neither dopamine nor cocaine was noted. This observation is consistent with observations that dopamine binds to the DAT in the absence of Na\(^+\).\(^{44,45,68}\)

2.5.2 RAMD

Table 5 provides an overview of the systems studied, generalized egress points and the range of simulation lengths. Initially, pathways were clusters on very general criteria: extracellular escapes, intracellular escapes or failed escapes. A “failed escape” refers to either 1) a simulation in which the substrate does not find an escape pathway during the simulation (e.g. substrate does not leave the pocket) or 2) a simulation in which the escape pathway sampled for the substrate is neither extracellular or intracellular; rather the substrate finds an escape pathway from the pocket out the middle of the protein and in to membrane.
Figure 16. MCTI stop-over sites identified for leucine in LeuT, water sphere environment. Stop-over sites are labeled A-H, represented by a surface representation of the substrate at that particular site, with a 3.5 Å probe radius. Leucine is presented in a ball and stick representation, with the surface representation changed to transparent for clarity. (In some insets, nearby pockets are changed to transparent for added clarity and to highlight proximity and overlaps of the stop-over sites.) Key residues at each site are identified and labeled, drawn in a stick model. Note, ``E'' is the S1 site while ``C'' represents the S2 site.

For a given setup, all of the trajectory files from the simulations were analyzed using the Volmap Plugin in VMD 1.8.7\textsuperscript{53} to visualize the escape route of the molecule of interest. The center of mass for the substrate was selected and volume files indicating the substrate occupancy were generated. Individual pathways of a given setup were then clustered based on visual inspection of the volume files and sodium movement to determine one or a few main pathways. Those trajectories that fall into a main pathway
category were then analyzed using the Hydrogen Bonds Plugin of VMD 1.8.7. Hydrogen bonds between the substrate and the protein were calculated for each trajectory of a given pathway cluster. In this way, we categorize the pathways observed from 500+ RAMD simulations.

**General findings: LeuT<sub>As</sub>**

In RAMD simulations using the LeuT<sub>As</sub> crystal structure after dynamics with a single substrate bound in S1 (Setup R1), we observe that the majority (77.5%) (see Table 5) of escape routes sampled were via an intracellular egress point. Noting that this structure has been through roughly 250 ns of accelerated MD, a large majority of intracellular escapes may indicate that the structure is approaching an inward-facing state. As anticipated, RAMD simulations using either inward-facing model (R3-R5) from Shaikh and Tajkhorshid<sup>24</sup> resulted in primarily (92%+) intracellular escapes, with leucine bound only in S1 or leucine bound in both S1 and S2. In the dual-bound 2A65 structure after dynamics (R2), we observe the only successful egresses to be to the intracellular milieu. We do note that 46% of those simulations resulted in failed escapes. This too may suggest that the structure<sup>54</sup> is approaching an inward-facing conformation.

After initial clustering, it became apparent that a single intracellular escape pathway was not being sampled. Thus, we classify to primary intracellular escape pathways found in RAMD simulations of all LeuT<sub>As</sub> models. Table 6 and Table 7 highlight the further classification of intracellular and extracellular escape pathways. Several of these pathways are shown in Figure 17 and Figure 18. After clustering, hydrogen bonding of leucine to LeuT<sub>As</sub> was analyzed. Residues not part of the substrate
pocket are listed in Table 6 and Table 7.

**Pathways: LeuT<sub>Aa</sub>**

The main extracellular pathway observed in RAMD simulations of LeuT<sub>Aa</sub> is shown in Figure 17A. Aside from interacting with residues in the defined binding pocket of LeuT<sub>Aa</sub>, hydrogen bonding was only observed between leucine and T254 of TM6. 6T254 is one of the four residues (in addition to 1A22, 1N27 and 7N286 and the substrate, leucine) that coordinate the sodium of Na1. In fact, across most of the pathways sampled for LeuT<sub>Aa</sub>, 6T254 has been identified as an interacting residue. Based on these observations, it is clear that the coordination of leucine, Na1 and 6T254 play a pivotal role in the release of substrate from the primary pocket.

Table 6. Summary of pathways observed in RAMD simulations of the leucine transporter. Setups correspond to those listed in Table 5. Pathway clusters are identified as intracellular (I) or extracellular (E), followed by a number denoting a specific pathway. After analysis, each pathway was assigned to one of two recurring pathways observed: along TMs 6b/8 or along TMs 1a/6b. Observation of sodium movement (Na1 in all cases) with leucine is indicated, as well as how many paths belong to the cluster. Observed residues interacting via hydrogen bonding for each pathway cluster are listed. Note that the residue list omits binding pocket residues (as described in Yamashita et al.: N21, A22, L25, G26, V104, Y108, F252, F253, S256, F259, S355, I359) for clarity.

<table>
<thead>
<tr>
<th>Setup</th>
<th>Path</th>
<th>TM Cluster</th>
<th>Na?</th>
<th># Times</th>
<th>Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>E1</td>
<td>–</td>
<td>No</td>
<td>7</td>
<td>6T254</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>6b/8</td>
<td>Yes</td>
<td>14</td>
<td>i1T77, 6T254, 6G260, 6A261, 6T264, 8A358, 12L500</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>1α/6b</td>
<td>No</td>
<td>14</td>
<td>1R5, 6T254, 6G260, 6T264, 6S267, 6Y268, 8A358, 8Q361</td>
</tr>
<tr>
<td>R2</td>
<td>I1</td>
<td>1α/6b</td>
<td>Yes</td>
<td>8</td>
<td>2E62, 2W63, 6T254, 6L255, 7D274</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>1α/6b</td>
<td>No</td>
<td>9</td>
<td>2W63, 6T254, 6L255, 7Q273</td>
</tr>
<tr>
<td>R3</td>
<td>I1</td>
<td>6b/8</td>
<td>Yes</td>
<td>31</td>
<td>1G24, 1N27, i11G73, i11H74, 6T254, 6A261, 6T264, 6Y265, 6Y268</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>1α/6b</td>
<td>No</td>
<td>23</td>
<td>i11G73, i11G75, i11T76, 6T254, 6T264, 6Y265, 8E368</td>
</tr>
<tr>
<td>R4</td>
<td>I1</td>
<td>1α/6b</td>
<td>Yes</td>
<td>9</td>
<td>6T254, 6T264</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>1α/6b</td>
<td>No</td>
<td>18</td>
<td>6T254, 6R270, 6T264, 6Y265, 8E360</td>
</tr>
<tr>
<td>R5</td>
<td>I1</td>
<td>1α/6b</td>
<td>Yes</td>
<td>24</td>
<td>2W63, 6T254, 6R270, 6T264, 6Y265, 8E360</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>1α/6b</td>
<td>No</td>
<td>18</td>
<td>6T254, 6R270, 7Q273</td>
</tr>
</tbody>
</table>
Two primary intracellular pathways were observed across all simulations of LeuT\textsubscript{Aa}. These simulations appear in panels B, C and D in Figure 17. Panels B and D, which can be described by following along TMs 1a and 6b, represent the first recurring path. The other pathway is represented by panel C, which is best described as following along TMs 6b and 8. In R1, both recurring intracellular pathways were sampled. Interestingly, the 6b/8 pathway (R1I1) was sampled when Na\textsubscript{1} was co-transported; yet the 1a/6b pathway (R1I2) was sampled when Na\textsubscript{1} was not co-transported. In simulations of the dual-bound LeuT\textsubscript{Aa} structure (R2), only the 1a/6b pathway was sampled. It is important to note that this pathway was sampled in both cases, i.e. with and without Na\textsubscript{1} co-transport.

The TM 6b/8 pathway was also sampled heavily in simulations using inward-facing model #1 from Shaikh and Tajkhorshid’s,\textsuperscript{24} represented by setup R3. In this setup, 46 of 49 RAMD jobs resulted in an intracellular egress. Of those 46 pathways, 31 resulted in intracellular egress routes along TMs 6b and 8. Inward-facing M2 from Shaikh and Tajkhorshid’s\textsuperscript{24} was also used (R4) and we observed only the TM 1a/6b pathway to be sampled. Finally, we studied the inward-facing M1, with leucine bound at both the S1 and S2 sites. The resulting pathways, R5I1 and R5I2, are equivalent in terms of pathway sampled, only differing in if Na\textsubscript{1} was co-transported or not.
Figure 17. Selection of pathways observed in LeuT$_{Aa}$ RAMD simulations. (Top Left) R1E1: Main extracellular pathway. (Top Right) R2I1/R2I2: Main intracellular sub-pathway 1a/6b. (Bottom Left) R3I1: Main intracellular sub-pathway 6b/8. (Bottom Right) R5I2: Main intracellular sub-pathway 1a/6b. In each panel, LeuT$_{Aa}$ TMs 1, 3, 6 and 8 are colored as red, light orange, sea-foam green and cyan ribbons respectively. The remainder of the TMs are shown as transparent grey tubes, for clarity. Residues lining the pathway are highlighted in purple stick representations. Observed pathways are represented as isosurfaces; each color of pathway indicates a separate run.
Pathways: DAT

RAMD of the DAT:DA complex resulted in observations three main pathways for ligand egress. The primary extracellular escape pathway observed is similar to that observed in LeuT$_{Aa}$ (R1E1). Similar to LeuT$_{Aa}$, two main intracellular pathways were observed for intracellular escape of dopamine from DAT- R6I1 and R6I2. These pathways are shown in Figure 18. Panels B and C of Figure 18 highlight these two intracellular egress points. The first, R6I1 tends to follow a pathway along TMs 6b/8, while R6I2 tends to follow a pathway along TMs 1a/6b. Path R6I1 was sampled twice as often as R6I2 over the 50 simulations of DAT:DA. RAMD simulations of psycho-stimulant drugs, (amphetamine and cocaine) bound in the S2 pocket of DAT were also carried out. In both case (R7 and R8), only extracellular egresses were observed.

The set of simulations of R9-R11 set out to determine the permeation pathway of dopamine in the presence of dopamine, amphetamine or cocaine (R9, R10, R11, respectively) in the S2 pocket. While it should be intuitive that dopamine should not be able to escape to the extracellular milieu from the S1 pocket with a bound substrate or inhibitor in S2, we do observe extracellular egresses in each case of DAT:DA:DA (R9E1), DAT:DA:AMA (R10E1 and R10E2) and DAT:DA:COC (R11E1). We note that the difference in the R10E1 and R10E2 pathways for DAT:DA:AMA differ only slightly regarding maneuvering about amphetamine in S2 (data not shown). The extracellular pathways observed in R9-R11 are consistent with the observed extracellular egresses of DAT:DA (R6E1) and are overall similar to that observed in LeuT$_{Aa}$ (R1E1).
Table 7. Observed residues interacting via hydrogen bonding for each pathway cluster in DAT. We denote the transmembrane (TM) helix the residue belongs to i.e., F76 specifies the F76 residue, belonging to TM1. Clusters are identified as intracellular (I) or extracellular (E), followed by a number denoting a specific pathway. Observation of sodium movement which was NA1 in all cases with the substrate is also indicated, as well as how many paths belong to the cluster. †In 16/50 DAT:COC RAMD runs, no hydrogen bonding between the protein and cocaine was observed. Note that this list omits binding pocket residues (as described in Indarte et al.: A77, D79, V152, F319, S320, V327, S421 and A422) for clarity. A complete list can be found in supporting information.

<table>
<thead>
<tr>
<th>Setup</th>
<th>Path</th>
<th># Times</th>
<th>Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>R6</td>
<td>E1</td>
<td>14</td>
<td>1F76, 1R85, 1I48, 3S149, 3Y156, 6L321, 6G322, 6F325, el4A382, 10D475</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>15</td>
<td>1F76, 1I1G127, 1I1A128, 3S149, 3Y156, 6L321, 6G322, 6F325, 8G425, 10E490</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>7</td>
<td>1S72, 1F76, 3S149, 3Y156, 6L321, 6G322, 6F325, 8S428</td>
</tr>
<tr>
<td>R7</td>
<td>E1</td>
<td>44</td>
<td>1W84, 3Y156, el4A382, el4T383, el4D384, el4G385, 10F471, 10T472, 10D475</td>
</tr>
<tr>
<td>R8</td>
<td>E1</td>
<td>34†</td>
<td>1W84, 1R85, el4D384, el4G385, el4P386, 10D475</td>
</tr>
<tr>
<td>R9</td>
<td>E1</td>
<td>16</td>
<td>1F76, 3I148, 3S149, 3Y156, 6L321, 6F325, el4T383, 8D420, 8G425, 10T472, 10D475</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>3</td>
<td>1F76, 3S149, 3Y156, 6L321</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>3</td>
<td>1F76, 3S149, 3Y156, 6L321, 7N339, 7D344, 7T348, 8G425,</td>
</tr>
<tr>
<td></td>
<td>I3</td>
<td>3</td>
<td>1F76, 3S149, 6F325, 8G425</td>
</tr>
<tr>
<td></td>
<td>I4</td>
<td>3</td>
<td>1F76, 3S149, 3Y156, 6F325, 8E436</td>
</tr>
<tr>
<td>R10</td>
<td>E1</td>
<td>9</td>
<td>1F76, 1R85, 3I148, 3S149, 3Y151, 3Y156, el4T383, 8S421, 9A442, 10D475, 10A478</td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>12</td>
<td>1F76, 1R85, 3I148, 3S149, 3Y156, 6L321, 6F325, el4A382, 10D475</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>8</td>
<td>1D68, 1S72, 1F76, 3S149, 3Y156, 5L254, 6L321, 7N325, 8D420, 8G424, 8G425, 8S428</td>
</tr>
<tr>
<td>R11</td>
<td>E1</td>
<td>17</td>
<td>1F76, 1R85, 3I148, 3S149, 3Y156, 6L321, 6F325</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>4</td>
<td>1F76, 3S149, 6L321</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>7</td>
<td>1F76, 1I1A128, 1I1W132, 3S149, 3Y156, 6L321, 6F325, 8G425, 8S428, 8E436</td>
</tr>
</tbody>
</table>

In R9, no general consensus was reached for dopamine release to the intracellular milieu from the DAT:DA:DA complex, as evidence of four sub-pathways (R9I1-R9I4) with no strong clustering. Especially compared to such clustered escapes for DA from DAT:DA, this may suggest that DAT does not follow a two-substrate alternating access transport mechanism. In R10, we observe one primary intracellular egress for DA from the DAT:DA:AMA complex. The intracellular pathway, R10I1 is visually equivalent to that of R6I2 of DAT, which follows a pathway that closely follows TM 1a/6b. It should be noted that this pathway was only observed in 8 of 50 RAMD simulations of the DAT:DA:AMA complex.
Figure 18. Pathways observed in the DAT:DA RAMD simulations. (A) Three pathways were observed for the DAT:DA complex: One primary extracellular pathway (R6E1), and two main intracellular pathways (R6I1 and R6I2). (B) Snapshot from a trajectory of an R6I1 egress. Dopamine (shown in VDW spheres) follows a pathway that closely follows TM 6b/8. (C) Snapshot from a trajectory of an R6I2 egress. Dopamine follows a pathway that closely follows TM 1a/6b. Pathways of R6I1 can be seen on the left-hand side of this image. In (A), TMs 1, 3, 6 and 8 are colored as red, light orange, sea-foam green and ribbons, respectively; while in (B) and (C) the TMs are colored the same but represented by surfaces with a 1.4 Å probe radius. The remainder of the TMs are shown are transparent grey tubes or surfaces, for clarity. VDW spheres represent dopamine. In all panels, observed pathways are represented as isosurfaces; each color of pathway indicates a separate run.

2.6 Summary

2.6.1 Insights regarding the leucine transporter

At the S1 and S2 site of LeuT<sub>Aa</sub>, low free energy scores were observed during MCTI simulations. The S2 site identified in this work is in agreement with work done by Quick et al.<sup>22</sup> regarding the S2 site of LeuT<sub>Aa</sub>. This success gave us confidence in using the MCTI method for further investigations of the S1 and S2 sites of our<sup>15</sup> DAT homology model.
While both types of simulations (MCTI and RAMD) were not designed to investigate sodium transport, we made interesting observations regarding sodium co-transport in numerous simulations. In all MCTI LeuTₐₐ simulations, Na₁ was observed to be co-transported. Additionally, in cases when sodium was co-transported in RAMD simulations, the sodium observed to be co-transported was always Na₁. Based on the results from these studies and others, we suggest the following steps regarding the transport mechanism of leucine in LeuTₐₐ.

Na₂ first binds to an outward-facing state of LeuTₐₐ. The binding of Na₂ would play a structural role for local rearrangement of the unwound region of TMs 1 and 6 as suggested by Celik et al. Additionally, Celik et al. suggest that binding of Na₂ provides enhanced selectivity for Na⁺ at the Na₁ site. Upon Na₂ binding and structural rearrangements, Na₁ and leucine approach the interior of LeuTₐₐ from the extracellular milieu together. Following this, release of Na₂ from the S1 region triggers the inward-facing state, as observed in TMD simulations by Shaikh and Tajkhorshid. With an inward-facing state then available, Na₁ and leucine (together) exit the S1 site and make their way to the intracellular milieu.

From both the RAMD and MCTI simulations, the results from this study suggest that Na₁ plays an escorting role for leucine transport. It should be noted that Na₂ release to the intracellular milieu was not observed, as was in a previous study. We maintain that Na₂ is likely released before Na₁ and leucine. The release of Na₂ to trigger transport however, may not have necessary in these simulations as the leucine was being forced out of the S1 site—either along some biased z-vector (MCTI simulations) or in some random direction (RAMD simulations).
During RAMD simulations, two main intracellular pathways were observed in LeuT<sub>Aa</sub>. One pathway appears to follow down TMs 1/6 (R2I1, R2I2, R5I2) and one that follows along TMs 6b/8 (R3I1). Currently, it is difficult to say which of these intracellular pathways is the true intracellular egress pathway. In a movie provided in the supplemental information of Shaikh and Tajkhorshid’s<sup>24</sup> inward-facing model of LeuT<sub>Aa</sub>, Na<sub>2</sub> appears to be released (from their Model 1) along a pathway similar to the pathway observed in this study that follows TMs 6b/8. It should be noted that due to the shorter timescale of the RAMD simulations, it is likely that the discussed pathways are “ligand-induced” egress routes as opposed to “conformationally-selected” egress routes.

2.6.2 Insights regarding the dopamine transporter

At the S1 site of DAT, we observe dopamine to be stabilized by interactions with 1F76, 1D79, 6V327 and 8S421 and fits comfortably in the binding cavity. In comparison, cocaine at the S1 site is not stabilized by any amino acid side-chain interactions and appears to be quite large for the given cavity size. This observation, in addition to the observed differences in the free energy profiles obtained from MCTI simulations, suggest that the binding site of cocaine does not overlap with that of dopamine as has been previously suggested.<sup>47</sup> Rather, these results fall in accord with the observation of Huang <i>et al.</i><sup>49</sup> that the binding site of cocaine is located above the aromatic lid (Y156-F319) of DAT in what Huang <i>et al.</i> refer to as the “substrate-entry tunnel”, or what we refer to here as the S2 site.

The intracellular pathways observed for dopamine unbinding from DAT were visually similar to the two main intracellular pathways that were observed for the LeuT<sub>Aa</sub>. Fewer, and less clustered intracellular pathways were sampled for dopamine from the S1
when a substrate or an inhibitor was present during RAMD simulations. While this is intuitive for cases of inhibitors (amphetamine and cocaine), this observation for the DAT:DA:DA case may suggest that DAT does not require a second substrate to be bound in the S2 pocket for transport. In fact, any bound ligand in S2 may produce an inhibitory effect. Finally, sodium transport was never observed for DAT, in either MCTI or RAMD simulations, which is consistent with observations that dopamine binds to the DAT in the absence of Na⁺.⁴⁴,⁴⁵,⁶⁸

2.7 Acknowledgments

This work is supported by the National Institutes of Health, National Science Foundation, Department of Defense, and The U.S. Department of Education under award numbers 5R01DA27806-2, CHE-0723109 (MRI), and P116Z080180. This research was also supported in part by the NSF through TeraGrid resources provided by National Institute for Computational Sciences and the Pittsburgh Supercomputing Center under grant number MCB060059P.

2.8 References

5. I. MacDougall, R. Griffith, Pharmacophore design and database searching for


25. L. Shi, H. Weinstein, Conformational rearrangements to the intracellular open states of the leut and apct transporters are modulated by common mechanisms. Biophys. J. 99 (2010).


37. B. M. Kronke, P. S. Horanyi, L. Columbus, Structural origins of nitroxide side chain dynamics on membrane protein α-helical sites, Biochemistry 49 (2010) 10045–
41. N. Reyes, S. Tavoulari, To be, or not to be two sites: that is the question about leuT substrate binding, Journal of General Physiology 138 (2011) 467.
50. Chemical Computing Group, Molecular Operating Environment (MOE), 2008.10, 1255 University Street, Suite 1600, Montreal, Quebec, Canada.
51. Chemical Computing Group, Molecular Operating Environment (MOE), 2009.10, 1255 University Street, Suite 1600, Montreal, Quebec, Canada.
54. J. R. Thomas, P. C. Gedeon, B. J. Grant, J. D. Madura, LeuT conformational sampling utilizing accelerated molecular dynamics and principal component analysis, Biophysical Letters Accepted with minor revisions (2012).
CHAPTER THREE

A REVIEW OF COARSE-GRAINED MOLECULAR DYNAMICS

TECHNIQUES TO ACCESS EXTENDED SPATIAL AND TEMPORAL

SCALES IN BIOMOLECULAR SIMULATIONS

3.1 Abstract

The study of very large biomolecular systems over extended spatial and temporal scales is being accomplished through the use of coarse-grained (CG) models. The successes of several CG models, including MARTINI and CMM-CG, and their implementation in current MD software are discussed. In addition, we discuss multiscale CG and mixed all-atom/coarse-grained models for MD simulations. The overall appeal of CG methods is the accessibility of broader temporal and spatial scales for biomolecular and polymer systems. However, care must be taken to fully understand the advantages, assumptions and limitations of each method. Not all CG models are created equally, and care must be taken when preparing the system and choosing the method of coarse-graining. It is the aim of this review to provide a background and guide to readers as they think about coarse graining. An outlook to the future of CG simulations is presented in the summary as well as a specific example of how we plan on using coarse-grained techniques to explore extended dynamics of membrane bound transporter proteins.
3.2 Introduction

3.2.1 Rationale and Interest

Scientific computing in biomolecular simulations has made great strides in the past decade. These achievements include advancements in the hardware, software and forcefields available to model biomolecular and polymer systems. Figure 19 illustrates the simulation techniques that are used across the various temporal and spatial scales of biomolecular and polymer systems. For example, quantum calculations access the microscopic scales in space and time to study small molecules and clusters, while atomic-resolution molecular dynamics classically study systems on the nanometer and nanosecond scales—such as proteins and lipid bilayers. In order to push the temporal scale of atomic-resolution molecular dynamics, algorithms that enhance conformational sampling of the system have been employed. Additionally, the use of GPGPUs is extending the temporal scale of QM and MD calculations. For example, a 10-100ps MD simulation is possible using QM on systems containing 1000-2000 atoms with GPGPUs. While these techniques have been successful in expanding the accessible temporal scale, the calculation of atomic-resolution potentials is not practical for larger biomolecular systems such as lipid rafts and viruses.

In order to push the limits of accessible temporal and spatial scales of biomolecular systems, coarse-grained (CG) models can be (and have been) utilized. To coarse-grain a system is to create a simplified, lower-resolution model of the system. This is achieved by grouping clusters of atoms into a new, simplified CG beads (or pseudo-atoms). The level of coarse-graining depends on how many atoms are represented by a CG bead. Increasing the atom-to-bead ratio increases the degree of coarse-graining,
yielding a lower resolution model. Collapsing the representation of atoms into CG beads effectively lowers the total number of degrees of freedom represented in the system. In addition, the CG beads interact with one another through more computationally efficient potentials. Together, these features yield a substantial increase in accessibility of time and space of biomolecular simulations. For example, Arkhipov et al. carried out CG simulations of the reovirus core, one of the largest viral structures available.\textsuperscript{6} With a diameter of nearly 75nm, the reovirus core consists of 480 protein units. A comparable all-atom representation of the reovirus would require almost 60 million atoms for 1.5 $\mu$s. Their CG representation of the reovirus allowed for roughly 1500 times fewer degrees of freedom.
freedom and a 200-fold increase in the integration time step. Using a 48-processor cluster, Arkhipov et al. collected the CG simulation data over several days of run time.\textsuperscript{6}

The essence of CG modeling is not new; in fact the first work on simplified representations of polymers was done by Flory in the 1950s and 1960s.\textsuperscript{7,8} In the mid-1970s Levitt presented his pioneering work on a simplified representation of protein conformations.\textsuperscript{9,10} Since then, there has been a renewed interest in CG modeling, particularly in the past few decades. To highlight the increasing interest in CG modeling, Scopus\textsuperscript{®} literature searches were performed including the words “coarse grained” and “dynamics” as the base search. Note that we included the word “dynamics” in our base search, as to eliminate citations that make reference to coarse-grained rocks or minerals, for example. Starting from the base search of “coarse grained” and “dynamics”, a series of subset searches were executed including the following keywords: protein, lipid, polymer, carbohydrate, micelle, multiscale and hybrid. As Figure 20 shows, the interest in and popularity of CG dynamics techniques has steadily increased since the mid-1990s from roughly 100 citations in 1996 to nearly 700 citations in 2010. It is also evident that within the CG literature, there is significantly more citations including work on proteins and polymers, compared to carbohydrates and lipids.

3.2.2 Fundamental Basis

Many CG models have been proposed over the years.\textsuperscript{11-37} Not all CG models are created equally, and care must be taken when preparing the system and choosing the method of coarse-graining. There are three distinct approaches in coarse-graining a system: energy-based,\textsuperscript{24,30} force-matching method\textsuperscript{22} and structure based.\textsuperscript{15,31} In energy-based CG, the interaction potentials of the CG beads are derived and parameterized such
Figure 20. Results from Scopus® upon searching for various terms regarding coarse-grained molecular dynamics. Data represents the years 1990-2010. Preliminary results from 2011 are omitted.

that the free energies or energies of the all-atom system are reproduced. In the force-
matching method, the sums of the atomistic forces are mapped onto the corresponding
CG beads. Lastly, the structure-based CG method relies on reproducing interactions
obtained from atomistic simulations, often in the form of radial distribution functions.

Despite the variety of approaches to coarse grain biomolecular and polymer
systems, each approach is phenomenological in nature. That is to say, CG models aim to
relate empirical observations of particular phenomena (i.e. energy, force, structure) in
such a way that the CG model provides findings consistent with all-atom (AA) molecular
dynamics or experimental results. The aim of all CG models is thus to provide the most
computationally efficient model for the system while still maintaining an adequate degree
of detail. Although fundamentally the reduction in atomic resolution should be simple
and straightforward there are hidden factors that must be considered in order to carry out the process correctly.

Several groups have investigated the fundamental basis behind the development of good coarse-grained potentials.\textsuperscript{38-43} In a paper by Mullinax and Noid\textsuperscript{42} they discuss the use of a generalized Yvon-Born-Green theory to determine coarse-grained potentials. The important point Mullinax and Noid make is to systematically identify approximations to the many-body potential of mean force (PMF) to sufficiently be accurate and ensure reasonable consistency with the underlying atomistic model. These many-body potentials should be practical to calculate, represent, and employ in a simulation. Mullinax and Noid point out that in condensed phase systems the many-body PMF would be poor if the potential energy function was derived from a sum of pair potentials obtained by neglecting many-body correlations.

Bolhuis, Louis and Hansen\textsuperscript{38} provide an excellent overview of the importance of many-body interactions and correlations when developing coarse-grained potentials for polymer solutions. They show that a state-dependent effective pair potential can be developed for polymers. They also demonstrate that for a system with many-body interactions, the two- and three-body correlations cannot be simultaneously reproduced from an effective pair potential. Once one has an appropriate potential function properly parameterized the next step is to use the correct equation of motion. In a paper by Kinjo and Hyodo\textsuperscript{41} a coarse-grained equation of motion is derived that retains the explicit relationship between coarse-grained and microscopic conditions. This is an important consideration in order not to introduce implicit correlations between particles that could influence the overall behavior of the simulation. The projector operator derived equation
of motion by Kinjo and Hyodo differs from the Brownian dynamics and dissipative particle dynamics by the model of the random force.

As CG modeling is an ever-growing and evolving field of study, it is the aim of this review to provide a background and guide to readers as they explore the wide variety of methods and implementations for carrying out coarse-grained molecular dynamics simulations of biomolecular and polymer systems. We focus on the energy-based and force-matching approaches for coarse-graining. We also discuss current advancements in mixed resolution dynamics. For other reviews concerning CG modeling, see references\textsuperscript{1,31,44-61} in addition to Issue 11 of PCCP 2009, “Coarse-grained modeling of soft condensed matter” and the book “Coarse-Graining of Condensed Phase and Biomolecular Systems” published by the CRC press, edited by Gregory A. Voth.

3.3 Energy-Based Approach to Coarse-Graining

3.3.1 MARTINI Model

Perhaps one of the most widely used CG models is the MARTINI\textsuperscript{19,24} model, which was originally parameterized to study lipid bilayers and is mainly implemented in the MD program, GROMACS.\textsuperscript{62} Over the years, it has been extended to proteins\textsuperscript{30}, carbohydrates\textsuperscript{63} and polarizable water\textsuperscript{64} by Marrink et al. The MARTINI force field has been used to study lipid rafts,\textsuperscript{65-67} protein self-assembly,\textsuperscript{68-70} protein gating,\textsuperscript{71-74} lipid bilayers\textsuperscript{75-80} and lipoprotein particles.\textsuperscript{81,82} Thermodynamic (bilayer phase temperature,\textsuperscript{75} lipid desorption free energy\textsuperscript{24}), dynamic (lipid lateral diffusion rates and timescale of lipid aggregation\textsuperscript{19}) and structural (lipid densities,\textsuperscript{19} area per lipid,\textsuperscript{19} lipid conformation,\textsuperscript{83} tilt angle of transmembrane helices\textsuperscript{30}) properties have been semi-
Figure 21. Mapping of CG beads onto A) DPPC lipid, B) Cysteine, C) Leucine and D) water molecules in the MARTINI model.

quantitatively reproduced using the methods of Marrink et al.

The MARTINI 2.0 model,\textsuperscript{24} henceforth referred to simply as “MARTINI”, uses a four-to-one mapping scheme; on average four heavy atoms are represented by one interaction site or bead. Examples of this can be seen in Figure 21 for a DPPC lipid, examples of amino acids and water. A standard mass of each bead is assigned to 72 amu (corresponding to four water molecules). The use of standard masses yields a more computationally efficient model, but may yield underrepresented dynamics of the system.\textsuperscript{24} For this reason, masses can be manually adjusted for beads. Four main types of interactions sites are considered in MARTINI: polar, nonpolar, apolar and charged. For each of these main types, there are subtypes that describe the hydrogen-bonding capabilities or denote the degree of polarity of the bead. The combination of main types and subtypes of beads yields 18 different bead types. This diversity allows for a more accurate description of the structure and chemical nature of the system compared to the 9
different types in the first version of MARTINI.\textsuperscript{24}

In the MARTINI model, non-bonded interactions are described by a shifted Lennard-Jones 12-6 potential of the form:

\[ U_{LJ}(r) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right] \]

Equation 5.

where \(\varepsilon_{ij}\) is the strength of the interaction between two beads and \(\sigma_{ij}\) represents the closest distance between two beads. The effective size of \(\sigma_{ij} = 4.7 \text{ nm}\) is assumed for each bead except the specialized classes for antifreeze particles and ring structures.\textsuperscript{24} The strength of the non-bonded interactions are discretized into 10 levels in the MARTINI model as seen in Table 8.

Table 8. Strengths of interactions defined for non-bonded interactions of CG beads in the MARTINI model.\textsuperscript{24} *Denotes \(\sigma = 0.62\text{nm}\) for special charged particle interactions.

<table>
<thead>
<tr>
<th>Level</th>
<th>O</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varepsilon) (kJ \cdot mol(^{-1}))</td>
<td>5.6</td>
<td>5.0</td>
<td>4.5</td>
<td>4.0</td>
<td>3.5</td>
<td>3.1</td>
<td>2.7</td>
<td>2.3</td>
<td>2.0</td>
<td>2.0*</td>
</tr>
</tbody>
</table>

Charged particles in the MARTINI force field bear a full charge \(q_{ij}\), which interacts via a shifted Coulombic potential of the form:

\[ U_{el}(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r} \]

Equation 6.

with a relative dielectric constant, \(\varepsilon_r = 15\). Bonded interactions between CG beads are described by a weak harmonic potential:

\[ V_{bond}(R) = \frac{1}{2}K_{bond}(R - R_{bond})^2 \]

Equation 7.
where the equilibrium bond distance is held at \( R_{\text{bond}} = \sigma_{ij} = 4.7 \, \text{nm} \) and the force constant is given by \( K_{\text{bond}} = 1250 \, \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2} \) for all beads. Chain stiffness is achieved through a weak harmonic potential for the angles:

\[
V_{\text{angle}}(\theta) = \frac{1}{2} K_{\text{angle}} (\cos \theta - \cos \theta_0)^2
\]

Equation 8.

where values for \( \theta_0 \) and \( K_{\text{angle}} \) are found in Table 9. For ring particles in the MARTINI model, an improper dihedral angle potential is used to prevent out of plane distortions of more complicated geometries:

\[
V_{\text{ld}}(\theta) = K_{\text{ld}} (\theta - \theta_{\text{id}})^2
\]

Equation 9.

Table 9. Force constants and equilibrium bond angles for for the angular potential term of the MARTINI force field.²⁴

<table>
<thead>
<tr>
<th>Type</th>
<th>( \theta_0 )</th>
<th>( K_{\text{angle}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliphatic Chains</td>
<td>180°</td>
<td>25 , \text{kJ} \cdot \text{mol}^{-1}</td>
</tr>
<tr>
<td>Cis Double Bond</td>
<td>120°</td>
<td>45 , \text{kJ} \cdot \text{mol}^{-1}</td>
</tr>
<tr>
<td>Trans-unsaturated Bond</td>
<td>180°</td>
<td>45 , \text{kJ} \cdot \text{mol}^{-1}</td>
</tr>
</tbody>
</table>

Simplified potentials for the interaction sites (beads) provide a reduction in the degrees of freedom, which enables the use of longer time steps for the integration of Newton’s equations of motion. Traditional values for \( \Delta t \) in all-atom molecular dynamics fall in the range of 0.5 to 2.0 fs,⁸⁴ while Marrink et al. suggest using a time step of 50 fs²⁴,⁸⁵ for systems using the MARTINI model. CG dynamics are generally faster than AA dynamics because the CG interactions are smoother than the AA interactions. This leads to the distinction of “effective time” of a CG simulation. For MARTINI, it is generally
assumed that the “effective time” of a simulation is 4 times that of the actual simulation time. As there has not been extensive testing across the variety of system types, this speed-up factor and the timescale of CG simulations should be interpreted very carefully.\(^85\)

While the use of such a large integration time step greatly increases the time scale accessible by CG simulations, there has been some contention regarding the use of such a large time step. van Gunsteren et al. contend that using a time step larger than 10 fs results in poor integration of the equations of motion, thus resulting in average properties (temperature) that differ greatly from the correct values.\(^86\) For further discussion on the use of time steps in CG simulations, refer to Marrink et al.’s comment\(^87\) as well and van Gunsteren et al.’s reply\(^88\) to the discussion of appropriate time steps in CG simulations.

3.3.2 Residue-based and shape-based coarse-graining

While the MARTINI model is primarily implemented in the molecular dynamics program, GROMACS,\(^62\) the Theoretical and Computational Biophysics Group from the University of Illinois at Urbana-Champaign has developed two coarse-graining methods implemented in NAMD\(^99\) and VMD\(^90\) that address a myriad of scales in biomolecular simulations, one of which is an extension of the MARTINI model. The two methods are residue-based\(^21,91\) (RBCG)- used to model lipid-protein systems and shape-based\(^6,92\) (SBCG)- used to model large macromolecular systems. The RBCG model uses an extension of the MARTINI model to describe the system of interest. Here, a four-to-one mapping scheme is used, as in the MARTINI model. Thus, two beads- one for the backbone and one for the side-chain represent an amino acid. Four water molecules are represented with a bead; while building blocks of lipids are reduced to single beads. A
single bead that includes their hydration shell represents ions in the RBCG model. These CG beads then interact through similar potentials as outlined previously for the MARTINI model.\textsuperscript{21,91}

Shih \textit{et al.} used the RBCG method to study high-density lipoproteins, known as “good cholesterol.”\textsuperscript{21} The resulting RBCG model decreased the system size by a factor of ten compared to the AA representation. Using a 48-processor cluster, Shih \textit{et al.} reached a performance level of 150 ns/day for the 30,000 CG particle system. In comparison, the same system, represented in a fully atomistic manner achieved on 0.1ns/day on the same cluster. Thus, even though the system size was only decreased by ten, a speed-up factor of 1500 is noted.\textsuperscript{21}

The SBCG model was developed in a much more general form, and thus can be applied to any macromolecular system.\textsuperscript{93} With this method, a SBCG model is generated from an all-atom model of the protein (e.g. crystal structure), in which a topology-generating network algorithm\textsuperscript{94} is used for the bead mapping. Interaction potentials of this method take the form of the CHARMM\textsuperscript{95} force field, in which bonded interactions are described by harmonic bond and angle potentials only. Non-bonded interactions here are defined by the 6-12 Lennard-Jones and Coulomb terms. In contrast to the MARTINI and RBCG models, these interactions are derived from all-atom molecular dynamics of the system on the order of \~10 ns. Lennard-Jones, bond and angle parameters are then extracted from the all-atom simulation and applied to the CG system in order to carry out the CG simulation,\textsuperscript{93,96} which yields a system-specific model.

For example, Arkhipov \textit{et al.} carried out CG simulations of the reovirus core, one of the largest viral structures available\textsuperscript{6} using the SBCG method. With a diameter of
nearly 75 nm, the reovirus core consists of 480 protein units. A comparable all-atom representation of the reovirus would require almost 60 million atoms for 1.5\(\mu\)s. Their CG representation of the reovirus allowed for roughly 1500 times fewer degrees of freedom and a 200-fold increase in the integration time step. Using a 48-processor cluster, Arkhipov et al. collected the CG simulation data over several days of run time.\(^{6}\)

3.3.3 CMM-CG Model

Development of the CMM-CG model began at the Center for Molecular Modeling at the University of Pennsylvania by Klein et al. Now at Temple University, Klein and his group continues to develop their CG model which includes parameters for non-ionic liquids (such as alcohol, alkane and ether based molecules),\(^ {25,97}\) water,\(^ {25}\) lipids,\(^ {98}\) amino acids (still in the development stage)\(^ {99}\) and phenyls (with the application to fullerenes).\(^ {100-102}\) This model boasts a parameterization method that utilizes extensive experimental thermodynamic data- downplaying the dependence on all-atom MD simulations, a flaw of other CG models.\(^ {100}\)

The CMM-CG model has successfully reproduced properties such as: toluene dimerization free energy,\(^ {100}\) \(C_{60}\) aggregation behavior\(^ {101}\) and \(C_{60}\) dimerization free energy.\(^ {101}\) Srinivas et al. has studied the phase behavior of polyethyleneoxide-polyethylene (PEO-P EE),\(^ {103-106}\) and the following properties were reproduced in agreement with available experimental or AA results: the area elastic modulus, the scaling of the hydrophobic core thickness, and the lateral chain mobility of bilayers. It should be noted that this model has been primarily used to study polymer systems.

As the development of the CG amino acid model\(^ {98}\) continues, we can expect to see reproducible properties of proteins. In contrast to the MARTINI model, this model
uses two types of LJ potentials to describe the non-bonded interactions, instead of the LJ 12-6 potential:

\[ U_{12-4}(r) = \frac{3\sqrt{3}}{2} \varepsilon \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{4} \]
\[ U_{9-6}(r) = \frac{27}{4} \varepsilon \left( \frac{\sigma}{r} \right)^{9} - \left( \frac{\sigma}{r} \right)^{6} \]

Equation 10.

The LJ 12-4 functional form is used to describe the 3-site water bead, while the LJ 9-6 functional form is used to describe the other beads of the model. Note that this model limits the number of water molecules per CG bead to three, in contrast to the four-site model used in MARTINI. The use of a three-site model by Klein et al. is a direct result of the parameterization of the model, and the desire to maintain a liquid state across 0-100 °C as well as interfacial properties such as surface tension and density.\(^{25}\) The LJ 9-6 functional form is also crucial to prevent the unrealistic freezing of water above 0 °C,\(^{25}\) which has can be problematic for CG models.\(^{24, 86}\) (Recall that Marrink et al. overcame this problem by introducing an anti-freeze particle.) Simple harmonic potentials are utilized for the bond stretching and angle bending, as in the MARTINI model.\(^{100, 107}\)

The CMM-CG model finds its main implementation in LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator),\(^{108}\) with additional implementation in HOOMDBlue (Highly Optimized Objected-oriented Many-particle Dynamics- Blue Edition),\(^{109}\) a fairly new MD software package that runs strictly on GPUs. For CG simulations performed in LAMMPS using this model, Klein et al. have used a two level RESPA multi-time step integrator.\(^{110}\) For bond and angle potentials, a time step of 2 fs was used, and non-bonded interactions were evaluated with a time step of 10 fs.\(^{25, 97-99}\) In a section contributed by Shinoda, Devane and Klein,\(^{107}\) the authors report that it is
possible to simulate systems of 1 million CG particles, which is equivalent to approximately 10 million atoms with current supercomputing resources. Furthermore, they report that with 2,000 processors, it is possible to generate trajectory data on the order of 100 ns/day of such CG systems.

3.4 Force-Matching Approach to Coarse-Graining

3.4.1 MS-CG Model

The multiscale coarse-graining (MS-CG) method was developed by Izvekov and Voth, with the goal of providing a systematic multiscale approach to CG simulations. This method provides a rigorous bridge between atomistic and CG models by utilizing force data from atomistic MD simulations in a systematic variational procedure. The MS-CG method has been used to generate CG models to study peptides, bilayers, carbohydrates, fluids, ionic liquids, transferable models of ionic liquids and mixed resolution models of proteins. The MS-CG method also allows for the use of implicit or explicit solvents.

The authors claim that through a formal statistical mechanics framework, the MS-CG method provides a solid foundation for deriving a consistent CG model. It has been shown for this method, that if no approximations are introduced, the MS-CG variational principle determines the exact many-body CG potential of mean force (PMF). With the exact many-body PMF, the fluctuation of exact atomically described forces from the CG PMF can be examined. Furthermore, the authors suggest that correlation functions of the fluctuations can be calculated and used to formulate a generalized Langevin equation for dynamics. Izvekov and Voth showed this provides a higher
degree of dynamical accuracy than simply applying Newtonian molecular dynamics to the CG PMF.

The resulting MS-CG models can be simulated to collect CG trajectory data using a variety of programs, including DL POLY\textsuperscript{124} as in\textsuperscript{28} and GROMACS\textsuperscript{62} as in\textsuperscript{125,126} Voth, Andersen \textit{et al.} have produced a series of articles, “The Multiscale Coarse-Graining Method. I-VI”\textsuperscript{27,28,127–130} which highlight a variety of questions, applications and implementations of the MS-CG method and are recommended for those interested in the MS-CG method.

3.5 Mixed Resolution Dynamics

3.5.1 Mixed AA-CG systems

Some researchers may find that they wish to retain the all-atom detail in one part of the system, but wish to treat the remainder of the system in a CG manner. One may imagine the case where atomic-detail is desirable at the active site of a protein, while a CG representation of the lipid membrane and aqueous solution surrounding the membrane is acceptable. For situations as such, one may consider generating a mixed AA-CG system. Work on mixed AA/CG system has generally consisted of fixed sets of atoms to be treated at the different levels of resolution\textsuperscript{49,123,131–134}. While these methods and models are useful, they do not allow for any particle exchange between the different resolution levels. An interesting approach to addressing the particle exchange problem in mixed AA/CG systems is the use of adaptive hybrid AA/CG simulations\textsuperscript{135–137}. In each of these methods, the AA and CG regions are connected via an interface, or healing region. It is in this region where particles are smoothly adapted to their respective AA or
CG representation. These methods have been utilized to study molecules where there is a one molecule represented per CG bead; as in the case for water, methane, and a tetrahedral toy molecule. Of these three adaptive hybrid methods, we will focus on the AdResS method.

3.5.2 Adaptive Resolution Scheme

In 2005, Abrams presented an adaptive Monte-Carlo approach to dynamics, while the work of Praprotnik et al. focused on the first adaptive resolution scheme for MD to be presented, AdResS. Figure 22 depicts how this models takes form. There exists three regions: a strictly coarse-grained region, a strictly all-atom region, and a healing or interface region that connects the two. In contrast to other mixed AA/CG

Figure 22. Schematic representation of a hybrid adaptive AA/CG simulation system.
models, the AdResS model allows for particle exchange between these different resolution regions. In the AdResS method, the basis of the healing region is derived from a force-scaling method to couple the AA and CG regions. The force is given by:

\[
f_{\alpha\beta} = w(R_\alpha)w(R_\beta) \sum_{i\alpha} \sum_{j\beta} f_{ij}^{AA} + [1 - w(R_\alpha)w(R_\beta)]f_{\alpha\beta}^{CG}
\]

Equation 11.

where the forces are a blend of the coarse-grained forces, \( f_{\alpha\beta}^{CG} \) on beads \( \alpha \) and \( \beta \) and the atomistic (all-atom) forces, \( f_{ij}^{AA} \) on atoms \( i \) and \( j \). Here, the atoms \( i \) belong to bead \( \alpha \) and atoms \( j \) belong to the CG bead \( \beta \) while \( w \) is a scaling function.\(^{138}\) The scheme was constructed such that it obeys Newton’s third law and maintains the stability of the particles in the interface region. The kinetic energy, temperature and degrees of freedom in the system are governed by an analog of the equipartition theorem:

\[
\langle K_\alpha \rangle = \frac{w(R_\alpha)k_BT}{2}
\]

Equation 12.

where \( \langle K_\alpha \rangle \) is the average kinetic energy per fractional degree of freedom- a unique feature of the AdResS method. The use of fractional degrees of freedom allows interactions and kinetic energy of the system to be scaled as a particle switches from an AA to a CG representation. Kinetic energy is thus controlled via a position dependent thermostat coupled to each fractional degree of freedom.

One of the challenges for the applicability of this method is the mapping and reverse mapping of CG waters. Recalling that one CG water bead in this method represents one water molecule, while in most other CG force fields, a typical CG water
bead represents three or four water molecules. Using the AdResS method, one may not attain the extended temporal and spatial scales to the extent of other methods. However, this method does provide a way to investigate applications such as proteins and receptors, where atomic detail at the active sites is desirable and timescales of conformational changes are on the order of milliseconds and greater. Without a doubt, the ongoing development of hybrid AA/CG dynamics will be intriguing to follow.

3.6 Prospective Utilization

Recently, neurotransmitter transporters have attracted much attention due to their crucial role in regulating neurotransmission within the central nervous system. The monoamine transporters (MATs), which include the dopamine (DAT), serotonin (SERT), and norepinephrine (NET) transporters, are of particular interest due to their association with motor function, memory, mood and addiction in the brain. These transporters act as regulators for their respective neurotransmitters. Dysfunction of the transporters is associated with a wide variety of diseases and disorders, including depression and cocaine addiction. These transporters rely on a Na$^+$ electrochemical gradient for the transport of their substrates.$^{144}$

In order to establish a sodium gradient, the periodicity of the traditional single bilayer system must be modified, as seen in Figure 23. The system will contain two lipid bilayers, and three salt-water baths. Compartments A and C will contain low concentrations (now connected periodically), while compartment B will contain a high concentration of ions. Furthermore, a transporter will be embedded into each of the bilayers. Both transporters will be positioned such that the extracellular region experiences the high ion concentration contained in Compartment B.
By building a dual-bilayer system to attain physiologically relevant conditions, we in effect doubled the size of the system to nearly 500,000 atoms. The timescale of the transport process is estimated to be on the order of 1 ms. Thus, reaching relevant timescales to study the conformational changes of these transporters with atomic resolution is impractical. Here, CG techniques can be utilized to access extended temporal and spatial scales necessary to study the MATs. Ideally, we desire to maintain atomic detail at the active site and translocation pathway to gain an understanding of interacting residues involved in the conformational changes of transport. Furthermore, we are only primarily interested in adding the phenomenon of the presence of a sodium gradient. Thus we envision the use of a mixed AA/CG or an adaptive hybrid representation. Therefore, we gather atomic detail at the active site and translocation pathway; while the lipid bilayers and salt baths are represented in a CG fashion—providing the phenomenon at a reduced computational cost.
3.7 Outlooks and Summary

CG modeling has made a significant impact on scientific computing and the understanding of biomolecular and polymer systems. Even as computing power continues to increase, we expect that the use of CG techniques will continue to tackle cases of extended spatial and temporal scales. However, it is our perspective that in order to properly move forward with the theory, application and implementation of CG simulations, the basic fundamental principles of CG models must be addressed. A clear development and thorough discussion of the fundamental basis of coarse-graining, while discussed by some,\textsuperscript{38-43} will be integral as new models are introduced and existing models continue to be developed.

With a clear fundamental basis, it should be more straightforward to understand how to best balance predictive power, accuracy and efficiency in a CG model. As we move forward, openness and transparency in reporting results is also important; reporting the successes and the failures and critically analyzing both cases is crucial. A robust comparative study of available CG models would be useful for scientists that are intrigued by the benefits of using CG modeling. Table 10 presents a comparison of density $\rho$, self-diffusion coefficient $D$, surface tension $\gamma$, isothermal compressibility $\kappa_T$, free energy of solvation $\Delta G_s$ and bead ratio values of several CG water models currently available. CG models are often developed to reproduce very specific properties, and those properties are of course what authors disclose regarding their model. As an example, the density of water is reported for all the models, while free energy of solvation and heat capacity and sparingly reported. Indeed, an investigation of thermodynamic and structural properties of varying CG models would be useful in determining the choice of CG model.
Table 10. Density $\rho$, self-diffusion coefficient $D$, surface tension $\gamma$, isothermal compressibility $\kappa_T$, free energy of solvation $\Delta G_s$ and bead ratio of various CG models. Bead Ratio is defined as the number of water molecules represented in a CG bead. Thus, four water molecules are represented by one CG bead in the MARTINI model (4-to-1).

‡ Data for the CMM model\textsuperscript{14} was performed by\textsuperscript{150}; they refer to this model as SSRBK, the initials of the authors.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\rho$ / g·cm\textsuperscript{-3}</th>
<th>$D$ / $10^{-5}$ cm\textsuperscript{2}·s\textsuperscript{-1}</th>
<th>$\gamma$ / mN·m\textsuperscript{-1}</th>
<th>$\kappa_T$ / $10^{-5}$ bar\textsuperscript{-1}</th>
<th>$\Delta G_s$ / kJ·mol\textsuperscript{-1}</th>
<th>Bead Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>0.997</td>
<td>2.3</td>
<td>71.6</td>
<td>4.52</td>
<td>-26.5</td>
<td>...</td>
</tr>
<tr>
<td>TIP3P [147–149]</td>
<td>0.985</td>
<td>5.93</td>
<td>52.3</td>
<td>5.76</td>
<td>-25.52</td>
<td>...</td>
</tr>
<tr>
<td>MARTINI [150]</td>
<td>1.005</td>
<td>1.6</td>
<td>32</td>
<td>9</td>
<td>-18</td>
<td>4-to-1</td>
</tr>
<tr>
<td>Pol. MARTINI [64]</td>
<td>1.043</td>
<td>2.5</td>
<td>30.5</td>
<td>...</td>
<td>-18.7</td>
<td>4-to-3</td>
</tr>
<tr>
<td>CMM\textsuperscript{5} [150]</td>
<td>0.993</td>
<td>6.6</td>
<td>71</td>
<td>15</td>
<td>-19</td>
<td>3-to-1</td>
</tr>
<tr>
<td>MS-CG [40]</td>
<td>0.997</td>
<td>9.7</td>
<td>...</td>
<td>14.5</td>
<td>...</td>
<td>1-to-1</td>
</tr>
<tr>
<td>GROMOS CG [151]</td>
<td>0.995</td>
<td>6.9</td>
<td>51.2</td>
<td>8.4</td>
<td>...</td>
<td>5-to-2</td>
</tr>
<tr>
<td>BMW [152]</td>
<td>1.047</td>
<td>...</td>
<td>77</td>
<td>3.3</td>
<td>...</td>
<td>4-to-3</td>
</tr>
<tr>
<td>CSJ [150]</td>
<td>0.998</td>
<td>4.3</td>
<td>71</td>
<td>17</td>
<td>-28</td>
<td>4-to-1</td>
</tr>
</tbody>
</table>

In light of the growing interest in CG modeling, we have presented a summary of several of the current CG techniques available to extend the accessible spatial and temporal scales of biomolecular and polymer simulations. We have provided the reader with an overview of the basis, applications and implementations of several models including the MARTINI, CMM-CG, RBCG, SBCG, MS-CG and mixed AA/CG models. These CG models aim to provide a computationally efficient model while still maintaining an adequate degree of detail. The manner in which these models balance predictive power, accuracy and efficiency is vastly diverse. We note that are different approaches (i.e. energy, force or structure) and assumptions used in the parameterization of the CG model. With this in mind, we stress that care must be taken to fully understand the advantages, assumptions and limitations of the CG methods.
3.8 References

20. E. S. Boek, T. J. Padding, W. K. den Otter, W. J. Briels, Mechanical properties of surfactant bilayer membranes from atomistic and coarse-grained molecular dynamics


149. C. Vega, E. de Miguel, Shirts05, Surface tension of the most popular models of water by using the test-area simulation method 126 (2007) 154707.
CHAPTER FOUR

BUILDING A MORE PHYSIOLOGICALLY RELEVANT SIMULATION SYSTEM: CREATING THE SODIUM GRADIENT FOR SIMULATIONS OF THE MONOAMINE TRANSPORTERS

4.1 Abstract

To develop a comprehensive understanding of the structure-function relationship of the monoamine transporters (MATs), this project sets out to simulate the dopamine transporter (DAT) in the presence of an explicit sodium ion gradient. The research plan is designed to test the following hypothesis: the mechanism of transport for dopamine is that of a single-substrate alternating access mechanism. That is, the dopamine transporter alternates between an outward-facing and an inward facing state in order to shuttle a single substrate across the membrane, thus regulating levels of the neurotransmitters at the synapse. By including (for the first time in any computational MAT study) the energy source used for conformational change and transport of substrates, we anticipate to observe substrate transport and conformational changes over the collected molecular dynamics trajectories.

4.2 Introduction

Ion transport through cell membranes is a fundamental mechanism that plays a key role in many physiological processes, including nerve signaling, energy transduction and muscle contractions. The MATs (discussed in Chapter 2) rely on the co-transport of sodium ions in order to cycle through conformational states of transport. At the
Table 11. Approximate extracellular and intracellular concentrations of potassium, sodium, chloride and calcium in a mammalian cell.\textsuperscript{5}

<table>
<thead>
<tr>
<th>Ion</th>
<th>Extracellular Concentration</th>
<th>Intracellular Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>K\textsuperscript{+}</td>
<td>5 mM</td>
<td>140 mM</td>
</tr>
<tr>
<td>Na\textsuperscript{+}</td>
<td>145 mM</td>
<td>5-15 mM</td>
</tr>
<tr>
<td>Cl\textsuperscript{-}</td>
<td>110 mM</td>
<td>4 mM</td>
</tr>
<tr>
<td>Ca\textsuperscript{2+}</td>
<td>2.5-5 mM</td>
<td>0.0001 mM</td>
</tr>
</tbody>
</table>

membrane interface, there is an overall positive charge in the extracellular salt bath while the intracellular salt bath is of negative charge. The resting potential, or voltage gradient ranges from 60-100 mV in physiological conditions.\textsuperscript{1,3,4} The resting potential is directly linked to the concentrations of ions in the extracellular and intracellular salt baths. Thus, a change in the concentrations will affect the voltage gradient. The approximate physiological concentrations of ions are summarized Table 11.\textsuperscript{5} Figure 24 highlights the dopamine transporter (DAT), embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) membrane and surrounded by approximate physiological ion concentrations.

Figure 24. Representation of DAT (orange ribbons) embedded in a POPE membrane (black CPK representation). Ions are colored as yellow (Na\textsuperscript{+}), blue (K\textsuperscript{+}), green (Cl\textsuperscript{-}) and red (A\textsuperscript{-}, other anions).
4.3 The Trouble with Periodic Boundary Conditions

Despite evidence\(^2\) that suggests that the MATs rely on the electrochemical gradient, there have been no computational studies of the MATs carried out that reproduce the electrochemical gradient. An obstacle in the explicit modeling an electrochemical gradient is the use of periodic boundary conditions (PBC) in modern simulations. PBC, depicted in Figure 25, is a common technique to enable the simulation of a small number of particles, while allowing the particles to experience forces as if they were in bulk solution.\(^6-9\) In the two-dimensional example, the simulation cell would be surrounded by eight neighbors, and in three dimensions, there would be 26 nearest neighbors.\(^10\) If a particle exits the simulation cell from the top of the box, it will re-enter the simulation cell from the bottom of the box, maintaining a constant number of particles in the simulation box. Standard periodic boundary conditions for computer simulations result in the water bath above and below the membrane to be periodically connected.

In terms of a transporter embedded in a membrane, the simulation cell would be constructed with single bilayer and salt-water baths on each side of the bilayer. Accounting for the physiological ionic concentrations discussed earlier, each salt bath would have different concentrations of ions. However, using a standard simulation cell and PBC, these baths would be connected such that particles are free to exchange between the two baths. In this way, the sodium gradient cannot be imposed with PBC. To achieve separated salt baths and consequently impose an explicit ionic gradient over a molecular dynamics simulation, the periodicity of the simulation cell must be altered.
Figure 25. Two-dimensional depiction of periodic boundary conditions. The simulation cell is highlighted in blue, while the periodic images surround the simulations cell, outlined in black.

Modified periodicity of the system can be achieved through constructing a non-traditional simulation cell. A dual bilayer system could be constructed, as shown in Figure 26. The system will contain two lipid bilayers, which creates one primary salt bath in the middle and the two outer baths. The outer baths will represent the intracellular medium (now connected periodically), while the middle bath will represent the extracellular medium. Thus, physiological ion concentrations can be placed and maintained in these separated salt baths. It is this method that we will use to impose the explicit sodium ion gradient in the presence of the dopamine transporter.
Figure 26. Modified simulation cell to achieve the explicit ionic gradient while employing PBC in molecular dynamics simulations.

4.3.1 Alternatives to PBC

Implemented in the Extended System Program\textsuperscript{11} (ESP), a parallel MD code developed in the research laboratory of Professor B.M. Pettitt, glide-plane boundary conditions (GBC\textsuperscript{12}) provide a unique variation on modified boundary conditions. Briefly, a traditional simulation cell would be constructed with a single bilayer and salt baths on each side of the bilayer. Thus the system has dimensions of $L_x$, $L_y$ and $L_z$. GBC, depicted in Figure 27 are achieved via a reflection of the simulation cell across the $xy$-plane and a translation across the $x$- and $y$-axes. Therefore, when a particle leaves through a side plane, it re-enters via the opposite side with unchanged velocity. Furthermore, if a particle exits the box through the top or bottom, it re-enters the box through the same plane. The point of re-entry is offset from the point of exit by $L_x/2$ and $L_y/2$ and the velocity of the particle is reversed in the $z$-direction.

The use of GBC would allow the user to maintain a reasonable simulation cell size since only a single bilayer is required. As computational cost is directly linked to the size of the simulation system, minimizing the simulation cell size is desirable. While the use of GBC may sound appealing for this reason, there are drawbacks to ESP and GBC.
The ESP program is a home-grown package which is sparsely used (outside of the group it was developed by) and is not very well documented.

Figure 27. Glide-plane boundary conditions for molecular dynamics simulations using the Extended System Program. A concentration gradient can be maintained with these boundary conditions with a traditional (single bilayer) simulation cell setup.

Preliminary molecular dynamics simulations carried out using the ESP program, with both PBC and GBC yielded less than optimal timings, even for a small test case of less than 10,000 atoms. Thus, the focus of this project will be on the use of a dual-bilayer system to impose the explicit ion concentration. Perhaps in the future the ESP program could either be optimized or the GBC be implemented in a larger, more widely-used MD engine like NAMD. Certainly, with increasing interest in transporters, ion channels and
attaining physiological relevant simulation conditions, the GBC could be applied to a variety of biophysical modeling problems.

4.4 Carbon Nanotubes: A Test Case

To assess the viability of using a dual bilayer simulation cell to model an electrochemical gradient for the MATs, a smaller test system was first constructed and studied. Nanopores and carbon nanotubes (CNTs) have been studied\textsuperscript{14-17} and shown to transport ions and thus have implications in desalination technologies. Thus, because of the ion transport properties, a 47 Å long CNT, with a diameter of 13 Å in the arm-chair configuration was generated using the Carbon Nanostructure Builder Plugin, Version 1.4 of VMD 1.9.1.\textsuperscript{18} The CNT was then embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane, generated using the Membrane Plugin, Version 1.1 of VMD.\textsuperscript{18} 13 Å water slabs were added to each side of the membrane. This single bilayer system was then minimized for 2 ps, at 1 fs time-steps. During the equilibration stage, the temperature of the system was increased in steps of 10 K from 0 K to 300 K with 2,000 steps at each temperature. A 2 fs time-step was used for temperature equilibration.

The final structure from temperature equilibration was extracted. The Autoionize Plugin, Version 1.3 of VMD\textsuperscript{18} was used to add 5 Na\textsuperscript{+}, 5 K\textsuperscript{+} and 10 Cl\textsuperscript{−} ions to the single bilayer system. This system (CNT, membrane, water and ions) was then duplicated to create the other half of the dual bilayer CNT system. The duplicate was reflected then translated the z-length of the original box and merged with the original. Since the Autoionize Plugin randomly places the ions in the system, coordinates of the ions in the newly created dual-bilayer system were exchanged to maintain all Cl\textsuperscript{−} ions in the middle
water bath and Na\(^+\)/K\(^+\) in the outer (connected) water baths. The final system had dimensions of 82 Å x 82 Å x 175 Å, consisting of 97,634 atoms.

Figure 28. Snapshot of the CNT dual bilayer system after minimization, temperature equilibration and initial dynamics (5ns). Green, ochre and yellow VDW spheres represent chloride, potassium and sodium ions, respectively.

The dual-bilayer CNT system was then re-minimized and re-equilibrated, according to the previous protocol. The final production run is to consist of 400 ns of trajectory data acquisition, with a 2 fs time-step, starting from the last structure of temperature equilibration. Trajectory frames were saved out every 10,000 steps, thus each frame represents a snapshot of the system every 20 ps. Due to the hydrophobic nature of the nanotube, the coordinates of the nanotube atoms were harmonically restrained during the simulation to maintain the nanotube’s position in the bilayer.

All minimization, equilibration and MD runs of the single and dual-bilayer system were executed using NAMD 2.8\(^{13}\) on Kraken at the National Institute for Computational Science at Oak Ridge National Laboratory, via TeraGrid\(^{19}\) allocations. Results from
simulations of the dual-bilayer test system will later be compared to data collected from simulations using ESP and GBC with a single-bilayer system.

4.5 DAT System Preparation

4.5.1 DAT Model Construction

Construction of the molecular model for DAT has been described in detail elsewhere. Briefly, the FASTA sequence of the rat DAT protein (SwissProt locus SC6A3-RAT; accession number P23977; NCBI accession number AAB21099) was used as the query in three independent sequence alignments with LeuT_Aa: Robetta server, 3D-JIGSAW server and the alignment of Yamashita et al. The model from the Robetta server was chosen as it produced the best docking energies for dopamine and amphetamine in the software program MOE. The AMBER99 all-atom force field was used to add hydrogen atoms and assign partial charges. The newly added hydrogen atoms were relaxed via several cycles of energy minimization using a conjugated gradient/truncated Newton optimization algorithm. The convergence criteria for final side chain refinement were carried out with AMBER99: 0.05 kcal/mol and a dielectric constant \( \varepsilon = 3 \). Since the 2A65 structure of LeuT_Aa was used as a template, the final DAT model represents an outward-facing conformation. Additional dynamics of the DAT molecular model were collected, which totaled 35 ns of production dynamics of DAT embedded in a bilayer (described below).

4.5.2 Embedded DAT System

Construction of the embedded DAT system has been describe elsewhere. Briefly, the DAT outward-facing model was embedded in a 1-palmitoyl-2-oleoyl-sn-
glycero-3-phosphoethanolamine (POPE) membrane using the Membrane Plugin, Version 1.1 of VMD. The protein and POPE membrane were solvated with approximately 32,000 TIP3 waters. Sodium and chloride ions were added to the system, yielding an overall neutral system at approximately 0.2 M NaCl. The 32 ns structure of the DAT molecular model was obtained from and used as the starting structure for the present study.

4.5.3 Dual Bilayer Construction

This structure was duplicated and translated as before with the CNT system. The water bath volumes were calculated to be $1.3 \times 10^{-21}$ L and $8.8 \times 10^{-22}$ L for the extracellular and intracellular water baths, respectively. In order to attain a physiologically relevant system, mimicking ion concentrations in the mammalian cell is desirable. The number of Na$^+$, K$^+$ and Cl$^-$ ions were calculated based on the bath volumes and the given concentrations of the ions in a mammalian cell in Table 11. The final system, with ions added to each bath is shown in Figure 29.

In these simulations, Ca$^{2+}$ was not incorporated into the system as the physiological concentrations are so minute and the volume of the baths is miniscule. Additional Cl$^-$ ions were added to the system to account for the additional negative charges that are present in the surrounding mediums. Addition of the extra Cl$^-$ ions also served to provide an overall neutral system for the MD dynamics simulations, which is desirable for PME calculations. The intracellular bath carried an overall charge of -9e while the extracellular bath carried an overall charge of +11e. The final dual-bilayer system size was 118 Å x 118 Å x 227 Å, with 274,292 total atoms. One of the dopamine transporters was left fully bound (one dopamine molecule and two sodium atoms at the
primary binding pocket) and the other represented an apo state. Two dopamine molecules were also added to the extracellular (middle) salt bath.

4.5.4 Calculation of the Equilibrium Potential

The resting potential of a cell, or voltage gradient ranges from 60-100 mV in physiological conditions. Also of interest is the equilibrium potential of the individual ions present at the chemical synapse. The individual equilibrium potentials of ions can be calculated using the Nernst equation. For example, to calculate the equilibrium potential of sodium we would have:

\[ E_{eq,Na^+} = \frac{RT}{zF} \ln \frac{[Na^+]_E}{[Na^+]_I} \]

Equation 13.

where \([Na^+]_E\) denotes the extracellular concentration of sodium and \([Na^+]_I\) denotes the intracellular concentration of sodium. The equilibrium potential of the cell can then be calculated based on the potentials of the ions. Potassium is ~100 times more permeable through the membrane than either sodium or chloride. Thus it follows the equilibrium cell potential can be calculated via Equation 14.

\[ E_{cell} = \frac{RT}{zF} \cdot \ln \left( \frac{100[K^+]_E + [Na^+]_E + [Cl^-]_E}{100[K^+]_I + [Na^+]_I + [Cl^-]_I} \right) \]

Equation 14.

Calculated potentials are summarized in Table 12. Based on the concentrations set out in the dual-bilayer system, the calculated value of \(E_{cell}\) is 74mV which falls in the range of 60-100mV.
Figure 29. DAT dual bilayer system. The extracellular bath is located between the two membranes. The intracellular bath is represented by the two outer baths, which are connected periodically.

Table 12. Calculated potentials for potassium, sodium and chloride ions and overall cell. $E_A^-$ represents the potential contributions from the additional chloride ions that were added the system. Thus, $E_{Cl^-} + E_A^-$ represents the calculated potential value for all chloride ions in the system.

<table>
<thead>
<tr>
<th>$E_{K^+}$</th>
<th>$E_{Na^+}$</th>
<th>$E_{Cl^-}$</th>
<th>$E_A^-$</th>
<th>$E_{Cl^-} + E_A^-$</th>
<th>$E_{Cell}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-84 mV</td>
<td>67 mV</td>
<td>-83 mV</td>
<td>46 mV</td>
<td>-5.0 mV</td>
<td>-74 mV</td>
</tr>
</tbody>
</table>
4.6 Results and Discussion

4.6.1 CNT System

**Permeation of K\(^+\), Na\(^+\) and Cl\(^-\)**

To date, a total of 74 ns of trajectory data of the dual bilayer CNT system has been collected. During the temperature equilibration phase (less than 1 ns), some ions began to permeate through the membrane and nanotube, as shown in the first panel of Figure 30. The permeation of K\(^+\) was observed during the trajectory and is depicted throughout the panels Figure 30. From nearly the midway point of the membrane, K\(^+\) permeates the membrane within 360 ps. Meanwhile, Na\(^+\) requires nearly 2.5 ns to permeate the membrane. In the carbon nanotube, we observed a single Cl\(^-\) ion to cross from one salt bath to the other. From nearly the midway point of the nanotube, Cl\(^-\) crossed to the other side within 1.6 ns. As trajectory data continues to be collected, a focus will be put on further permeation rates for each ion.

4.6.2 DAT System

**Analysis of salt bridges**

Over the collected trajectory (50 ns), salt bridges in DAT were analyzed utilizing the Salt Bridge plugin of VMD 1.9.1. A total of 16 salt bridge pairs were identified in both configurations (fully bound and apo) of DAT. An additional 6 salt bridge pairs were identified in either configuration, but not both. The 16 salt bridge pairs found in both configurations are: D68/R65, D230/R227, D300/R303, D344/K66, D380/R379, D435/R60, D435/R444, D475/R85, D505/K508, E61/K336, E214/R218, E217/R218,
E306/R303, E427/R60, E427/R444 and E436/K133. Their fluctuations over the trajectory are summarized in Figure 31.

Figure 30. Snapshots from the CNT dual bilayer simulation. The nanotube and ions are represented in CPK while the $K^+$, $Na^+$ and $Cl^-$ ions observed to permeate are highlighted in larger VDW spheres are highlighted with a violet circle.
Figure 31. Selected salt bridge distances observed to fluctuate the most of the collected trajectories. Salt bridges are highlighted in DAT in the lower right panel. The following residues are colored as: D68/R65: dark blue, D230/R227: red, D300/E306/R303: orange, D344/K66: yellow, D380/R379: green, D435/E427/R60/R444: pink, D475/R85: black, D505/K508: light blue, E61/K336: violet, E214/E217/R218: ochre, and E436/K133: silver.
Formation of a water wire in the apo configuration

Water, sodium and potassium were observed to permeate toward the primary binding pocket of DAT in both configurations of DAT, to differing degrees. Snapshots throughout the simulation are presented in Figure 32 which highlights this for the bound and apo configurations of DAT. Further analysis resulted in the observation of a water wire formed from the extracellular medium to the intracellular medium in the apo configuration of DAT, but not in the bound configuration of DAT. To further highlight this, an isosurface of water within 3Å of the protein was created using the Volmap Tool of VMD 1.9.1. and is presented in Figure 33.

In the apo configuration, the binding pocket of S1 becomes filled with water and the S1 site is accessible by water to both the intracellular and extracellular mediums. The water channel that connects the S1 site of DAT to the intracellular milieu is defined by TMs 3,8 and 10. There also appears to be a second water wire that is beginning to form along TMs 1 and 6 to the intracellular milieu.

In the case of DAT with dopamine bound at the S1 site, it is apparent that a water wire is not formed. Water is observed to permeate just past the D475/R85 (extracellular gate) salt bridge, but is halted by the aromatic lid just above the primary binding site. On the intracellular side, water is beginning to permeate toward the S1 site, along TMs 3, 8 and 10 but does not yet begin to hydrate the primary binding site. This is also supported by the salt bridge analysis presented in Figure 31. The distance separation of D475/R85 is ~3 Å in the apo configuration and ~6 Å in the bound configuration, indicating that the gate is closed in the apo configuration but open in the bound configuration.
Figure 32. Snapshots from the apo (left) and S1 (right) trajectories. DAT is represented in ribbons with TMs 1,3,6 and 8 highlighted in red, orange, green and blue respectively. Salt bridges are highlighted as in Figure 31. Ions (sodium: yellow, potassium: brown and chlorine: cyan) within a 3Å radius of the protein are shown in spheres with water surrounding those ions in a 4Å radius. The substrate, dopamine is shown in small VDW spheres, where applicable. Finally, water within and close to the protein is highlighted in small VDW spheres.
Figure 33. Isosurface (isovalue = 0.6) representing water within 3 Å of DAT. A close-up of each configuration is shown. To the left is the apo configuration of DAT while the right is the configuration with the bound sodium and substrate.

Importantly, the formation of a water wire or water channel in DAT may suggest that DAT does not follow an alternating access mechanism,\textsuperscript{27} which has been proposed\textsuperscript{28,29} as the mechanism for the MAT bacterial homologue, LeuT\textsubscript{Aa}. The alternating access mechanism requires the binding site to be alternately accessible to the extracellular and intracellular milieus. Therefore, the substrate binds to one side of the membrane protein, while some energy input drives a conformational change in the protein allowing the substrate to be transported to the other milieu. This transport is proposed to occur if and only if all co-transported components are bound, and not when the primary pocket is partially filled.\textsuperscript{30} Thus, with the observation of a water wire, we will be interested in the continuation of these trajectories to potentially observe the release of substrate from the S1 site to the intracellular milieu and what conformational changes are associated with that process.
**Presence of potassium in the intracellular opening**

Since potassium was observed to penetrate DAT from the intracellular milieu, the Volmap Tool was also used to visualize the regions of DAT where ions are located throughout the trajectory. Figure 34 shows isosurfaces of sodium, potassium and chloride over the duration of the trajectory. Opaque surfaces represent isovalue of 1.5, 2.0 and 2.65 for sodium, potassium and chloride, respectively. Transparent surfaces represent isovalue of 2.5, 2.75 and 2.8 for sodium, potassium and chloride, respectively.

![Figure 34](image)

Figure 34. Yellow, brown and cyan isosurfaces represent the overall positions of ions (sodium, potassium and chloride, respectively) during the trajectory for the apo (left) and S1 (right) configurations.

These surfaces aim to represent regions where the ions spent a large majority of time during the trajectory. In both configurations, it is observed that potassium is coordinating with residues along the intracellular entrance and even into the interior of
DAT. In the apo configuration, potassium enters DAT nearly to the binding site from the intracellular milieu. This also indicates that the binding pocket of the apo configuration is accessible from the intracellular solvent.

4.7 Future Directions

4.7.1 Data Analysis

A rigorous trajectory analysis will be performed on the data collected for DAT in physiological conditions. Correlated motions of groups of atoms can be obtained using principal components analysis (PCA) which attempts to extract pattern information from a set of data by first reducing the dimensionality of a data set, without loss of the information.\textsuperscript{31-33} This is accomplished through analysis of the eigenvectors and eigenvalues from the diagonalization of the covariance matrix and re-expressing the data in terms of selected (important) eigenvectors. This results in a new data set composed of a smaller number of variables, or principal components. These components can then be used as a predictor for subsequent analysis.\textsuperscript{32} For example, recent PCA work with other long time-scale dynamics in our group revealed that the majority of the variance in conformations of LeuT\textsubscript{Aa} could be accounted for with TM 1a and 6b. Thus, different conformations of LeuT\textsubscript{Aa} can be identified through tracking the changes in TM 1a and 6b over a simulation.

The resulting trajectories will also be analyzed with POVME- POcket Volume MEasurer,\textsuperscript{34} to track the changes of the pocket (S1 and S2) volumes over the duration of the simulation. The dxTuber\textsuperscript{35} or HOLE\textsuperscript{36} programs will be used to identify, visualize and characterize the water and substrate pathways through the transporter. Interactions of
dopamine with the transporter, as well as distance fluctuation of key salt-bridges and gates will continue to be tracked and characterized over the duration of the trajectory.

Key residues of observed conformational changes will be submitted to the Surratt group for experimental validation. Dopamine transporter with the recommended mutations will be expressed and used in uptake assays to determine the effect of the mutation on transport. Residues determined to be accessible from the intracellular milieu will also be submitted to the Surratt group for accessibility studies. These techniques will be used to further validate the structure of the DAT model and resulting conformational states and changes.

Since the systems in this study include an electrochemical gradient, an emphasis will also be placed on the electrostatics, potential and ion and water leakage over the simulation. These results can be compared to current experimental work in the Amara group at the University of Pittsburgh. Indeed, with such long timescale dynamics, there will be a heavy emphasis on the analysis of these trajectories to illuminate as much as possible regarding the mechanistic behavior of the dopamine transporter with regard to the alternating access mechanism as well as substrate release from and entrance to the binding pocket.

4.8 Acknowledgements

This research was supported in part by the National Science Foundation through TeraGrid resources provided by the National Institute for Computational Sciences.
4.9 References

11. B.M. Pettitt. ESP: extended system program. Developed in the Research Laboratory of Prof. B. M. Pettitt at the University of Houston. Houston, TX: 1996.


22. Chemical Computing Group, Molecular Operating Environment (MOE), 2009.10, 1255 University Street, Suite 1600, Montreal, Quebec, Canada.


