A Conserved Motif in EL4b of NSS Transporters May Control Antidepressant Binding
by
Michael LeVine
Class of 2011

A thesis submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the Degree of Bachelor of Arts with Departmental Honors in Chemistry

Middletown, Connecticut April, 2011
# Table of Contents

Introduction.................................................................................................2
I. History of Neurotransmitter Transport.....................................................2
II. Alternating Access Model.......................................................................4
III. Monoamine Transport Pharmacology....................................................6
IV. Structural Classification.........................................................................7
V. A Possible Motif in EL4b.........................................................................11
Methods.....................................................................................................14
I. Molecular Dynamics Theory....................................................................14
II. Free Energy Perturbation Theory............................................................20
III. Principal Component Analysis Theory....................................................22
IV. Creation of LeuT-Desipramine Mutants..................................................23
V. Molecular Dynamics Production..............................................................24
VI. Free Energy Perturbation Calculations....................................................24
VII. Analysis of Molecular Dynamics Trajectories........................................25
Results.......................................................................................................26
I. Mutants are converged and equilibrated....................................................26
II. Mutations result in structural changes in EL4..........................................27
III. Mutations result in only small changes to the RMSF within EL4...............27
IV. Mutations result in changes to desipramine binding affinity......................28
V. Mutations alter how EL4b samples conformational space.........................29
Discussion.................................................................................................31
I. The PxS motif is related to desipramine binding.......................................31
II. Energetic changes and structural changes may not be clearly related.........32
III. EL4b may be a dynamic control center..................................................33
VI. Future directions..................................................................................34
Conclusions................................................................................................37
Acknowledgements.......................................................................................39
Appendix....................................................................................................40
I. Molecular Dynamics Specifications..........................................................40
II. Free Energy Perturbation Specifications..................................................41
III. Wild-type..............................................................................................43
IV. A335P..................................................................................................45
V. A335P/G336L.......................................................................................47
VI. A335P/G337S.....................................................................................49
VII. A335P/G336L/G337S.........................................................................51
VIII. A335S..............................................................................................53
IX. A335S/G336L......................................................................................55
X. A335S/G337S......................................................................................57
XI. A335S/G336L/G337S.........................................................................59
XII. G336L..............................................................................................61
XIII G337S..............................................................................................63
XVI G336L/G337S....................................................................................65
References.................................................................................................67
Introduction

I. The History of Neurotransmitter Transport

The brain is able to receive, interpret, and analyze a wide array of stimuli across many sensory modes through the communication of an elaborate system of interconnected cells called neurons. Neurons communicate with each other through chemical signaling at the synapse, and use various small molecules called neurotransmitters as well as larger peptides called neuropeptides to invoke diverse cellular responses. The presynaptic cell releases neurotransmitters from its axon in response to an action potential, and the neurotransmitters diffuse across the synapse to the postsynaptic dendrite, where they bind to membrane-associated receptor proteins and induce various cellular events. This sophisticated system of neurotransmission uses active transport in order to both add temporal specificity to the chemical signal and to conserve metabolic resources. While glial cells play an important role in the chemical upkeep of synapses, membrane transport proteins exist specifically to reuptake neurotransmitters quickly.

Figure 1 – A model synapse (Adapted from Goodlett, 2001).

To power the active transport of neurotransmitters, many of these reuptake proteins have been shown to be coupled with the Na+ gradient across the neuronal membrane (Kanner, 1983), which is believed to provide the energy required for active transport. Since the 1970s, it has been known that there are specific uptake systems for each neurotransmitter (Iversen and Johnston, 1971). One of the largest families of transport proteins is the Na+/Cl- dependant transporter family, which depend on both Na+ and Cl- for proper function. Within the general family across all organisms exists the neurotransmitter:sodium symporters (NSS). The superfamily is further broken down into four sub-families: GABA transporters, monoamine transporters, amino acid transporters, and orphan transporters. All Na+/Cl- transporters share a similar overall structure, characterized by 12 transmembrane domains. The monoamine transporter (MAT) subfamily includes the dopamine (DAT), serotonin (SERT), and norepinephrine (NET) transporters. The first characterized transport protein, the GABA transporter, was cloned and sequenced in 1990 and led to the beginning of intense pharmacological study to understand its structure and function (Guastella et al., 1990; Nelson, 1990) of the family. Study of the MAT family blossomed when the human norepinephrine transporter was cloned in 1991 and it was shown that all transport properties, such as Na+ dependence, tricyclic antidepressant (TCA) inhibition, and substrate specificity were encoded in the gene (Pacholeczyk et al., 1991) and not in post-translational modifications. That same year both the dopamine and serotonin transporters were successfully cloned and sequenced (Kilty et al., 1991;
Shimada et al., 1991; Blakely et al. 1991). The sequences of NSS transporters, and more specifically of the monoamine transporters, are highly homologous (See Fig. 2) and show even higher homology throughout structural domains shown to be important for function.

Figure 2 – Sequence alignment of the human monoamine transporters. Conserved residues are in blue and similar residues are in magenta.

II. Alternating Access Model

A general mechanism called the alternating access model has been proposed for Na+/Cl- transporters. The original model, called the simple allosteric model for membrane pumps, proposed that a simple Na+ ATPase could be modeled with two events and four conformations (Jardetzky, 1966). The pump would begin with an open binding site on the intracellular side. The pump binds a Na+ ion, causing a conformational change such the pump opens to the extracellular side, exposing the Na+ ion. The pump would then be phosphorylated, triggering the release of the Na+ ion into the extracellular space due to an allosteric change in the affinity of the Na+...
binding site (See Fig. 3). This model was adapted for transporters which rely on electrochemical gradients (Scarborough, 1985). In this model, four significant conformations exist, named in regards to the accessibility of the binding site in relation to extracellular and intracellular space: outward-facing open, outward-facing occluded, inward-facing occluded, and inward-facing open. In this model, both substrate and ions would bind while the transporter is in the outward-facing open conformation. The transporter would then progress through the occluded conformations, and release would occur in the inward-facing open conformation. Many models propose the ion binding causes conformational changes which stabilize substrate binding, and thus the ions bind first. Based on the dynamics of another bacterial transporter, a more specific model has been proposed (Gouaux, 2005, see Fig. 4). While this is a simplistic model, it provides a bare-bones approach to understand the transporter family as a whole as well as a direction for future study.
III. Monoamine Transporter Pharmacology

The monoamine transporters have been studied extensively due to their interactions with several drugs and are often used as medicinal targets for neurotransmitter-related disorders. Changes in ambient levels of monoamine neurotransmitters have been shown play roles in neurological diseases such as serotonin in depression (Cowen, 2011), dopamine in attention-deficit (Hutz, 2010), and norephinephrine in obesity (Lambert, 2010). A diverse family of psychiatric drugs exists and each has a unique binding profile across the MAT family. The tricyclic antidepressants (TCAs) bind to SERT and NET with high affinity, while having little affinity for DAT (Richelson, 1997). Methylphenidate, also known at Ritalin, binds with high affinity to DAT and NET but not SERT (Markowitz, 2006). Furthermore, all monoamine transporters interact with the illicit stimulants amphetamine and cocaine, and other recreational drugs such as methamphetamine make interactions at one or more MAT proteins (Caldwell, 1974). For nearly all inhibitors of the monoamine transporter family, it is believed that inhibitor binding occurs outside of the primary substrate binding (S1) pocket, possibly in a secondary binding (S2) pocket, although there is still unclear if all inhibitors bind in the same pocket or if inhibitors display diverse binding modes in various pockets within the transporter cavity. Cocaine has been shown to bind in such a way as to overlap with primary binding pocket (Gether, 2008), which may indicate that not all inhibitors bind to S2. However, the TCA binding site will be considered S2 for the duration of this paper. The large variability in inhibitor activity indicates that there are both highly conserved residues in S2 which are likely involved in the general mechanism of
inhibitor binding as well as additional non-conserved residues making inhibitor-specific interactions.

VI. Structural Classification

Membrane protein structural characterization is often difficult due to the inability to crystallize purified protein using conventional methods. Protein crystallization relies greatly on polar contacts between molecules to form a crystal lattice, and the majority of the surface area of many membrane proteins is significantly hydrophobic. Also, their lack of solubility in polar solvents without denaturing requires the addition of detergents, which need to be kept at minimal concentrations (Prive, 2007). To avoid crystallization issues, structural determination of membrane proteins can be performed using NMR. This can also be difficult due to size and solubility limitations, but the field of membrane protein NMR has made great strides as of late (Wang, 2008). Due to these limitations, there are relatively few crystal structures for membrane proteins and no structures for mammalian Na+/Cl-transporters. However, the structure of a bacterial homologue, the leucine transporter (LeuT), was solved using x-ray crystallography with the leucine substrate bound (Gouaux, 2005), as well as in complex with several of the TCAs (Gouaux, 2007 at 1.9 A and Wang, 2007 at 2.9 A). LeuT is only moderately homologous to the monoamine transporters in terms of sequence, but is believed to share high structural similarity (See Fig. 5). Many labs have begun to use the LeuT structure as the basis for computational homology models for DAT (Madura, 2008), NET (Lewis, 2010), and SERT (Schiott, 2008), and have used the LeuT-inhibitor complex as a model for ligand binding in S2. Hopefully, homology models of the MATs will lead to
structural insight in regards to substrate and inhibitor binding specificity, as well as the design of new drugs with tailored binding profiles.

Figure 5 – Sequence alignment of LeuT with the human monoamine transporters. Conserved residues are in blue and similar residues are in magenta.

There are many issues in regards to using LeuT as a model for the MAT family. LeuT is a bacterial protein whereas the MAT family is mammalian. The membrane context in which LeuT exists may be very different than that in which a MAT exists. Also, whereas the MAT family regulates the levels of neurotransmitters in synapses, LeuT transports an amino acid into the bacteria, likely for use in protein synthesis. These two functions serve largely different purposes, and thus may have evolved largely different kinetics and thermodynamics, as well as regulation. However, it is very difficult to investigate these differences experimentally before any prediction about the nature of these differences has been made.
Based on crystal structures of LeuT bound to the TCA desipramine, it has been proposed that several residues are involved in direct contacts with desipramine. Specifically, it has been suggested that two residues within the extra-cellular loop 4 (EL4) hairpin, F320 and A319, play a role in desipramine binding (Wang, 2007). EL4 is made up of two helices, EL4a and EL4b, which form a hairpin and this region is often considered the mouth of the substrate pore. A comparison of the occluded LeuT structure with the desipramine-bound structure indicates that upon binding desipramine, the EL4 hairpin as well as F320 and A319 are pushed out towards the extracellular space. F320 and A319 make contacts with the tricyclic ring, as shown in Fig. 6. When desipramine was docked into MAT homology models, it appears that EL4a residues may crowd the binding site in DAT and SERT, while it appears to accommodate desipramine in LeuT and NET, which provides a simple explanation for desipramine’s activity profile. Interestingly, the residues in EL4 hairpin which make contacts with desipramine differ for each of the transporters, indicating that EL4 structure may play a large role in inhibitor specificity. The desipramine-sensitive NET contains an...
alanine and glycine in the EL4 hairpin, and it is possible that these interactions may not be significant, and that the presence of phenylalanine in LeuT and proline in the other transporters may crowd the desipramine binding site rather than provide stabilizing interactions. S2 appears to be strongly conserved, but the lack of sequence conservation in the EL4 binding residues may be responsible for the diversity of MAT affinities seen in TCAs and psychostimulants.

However, it has been shown that the detergent which LeuT is crystallized in, n-octyl-beta-d-glucopyranoside (BOG), inhibits transport (Javitch, 2009). All LeuT structures which are not bound to an inhibitor have BOG present in S2, and thus it is unclear if the “unbound” structure is in actuality an inhibited or occluded conformation, and it is also unclear if the conformational difference between the two is characteristic of desipramine binding or simply a difference between the binding of desipramine and the binding of BOG. Thus, there is truly very little known in regards to the mechanism by which inhibitors bind and block substrate transport.

The molecular dynamics of both a substrate-bound LeuT crystal structure (2A65) and a DAT homology model in implicit membrane and solvent (Madura, 2008) have been simulated. While the DAT homology model was structurally similar to LeuT throughout the trajectory, local unwinding of EL4a occurred in DAT. The
conformation of EL4 in inhibitor-bound crystal structures varies depending on inhibitor, indicating that affecting the ability of EL4 to adopt these conformations could influence the ability of a transporter to bind some inhibitors. Throughout the trajectory, LeuT sampled conformations which were in close relation to the desipramine-inhibited structure, and as the largest difference between the occluded and inhibited states reside in EL4, it is likely that dynamics of EL4 are essential to proper equilibrium sampling and inhibitor binding. Furthermore, it may be reasonable to believe that conformational diversity in EL4, as opposed to sequence diversity, may be related to the inhibitor binding selectivity seen within the MAT family, but there is little structural data in regards to EL4 in any of the MAT due to the lack of crystal structures.

V. A Possible Motif in EL4b

A sequence alignment NSS transporters revealed that EL4, and especially EL4b, is highly conserved. Specifically, there is a conserved three residue motif in turn following EL4b. The motif, PxS, where x is a hydrophobic residue, is strongly conserved among the mammalian NSS transporters; however LeuT contains the residues AGG in the place of this motif. The replacement of the rigid proline with flexible glycines may accompany large structural changes throughout EL4 and thus affect EL4 dynamics and inhibitor interactions with EL4 hairpin residues. The effect of mutation at this motif has been studied in mammalian NET, which contains only a partially conserved motif, SGS. NET shows significant inhibition by desipramine, while DAT, which contains PLS, shows little inhibition. The NET double mutant S399P/G400L, which introduces the full motif, shows a 1000-fold decrease in
desipramine potency, while each mutation individually induces little change (Giros, 2001). This indicates the PxS motif may play a role in determining the conformation of S2 and its binding properties for at least NET. While a mechanism is unclear, the PxS motif may limit the conformational flexibility of EL4, and thus the ability of the EL4 hairpin to accommodate desipramine. Using a DAT homology model, these mutations were investigated, and local changes to EL4 structure as well as free energy of binding of desipramine were observed (Madura, unpublished data), yet DAT has been shown to only weakly be inhibited by desipramine, and thus may inherently be a poor model for investigating the role of the motif in inhibitor binding.

Figure 8 – Sequence alignment of the human NSS transporters. Conserved residues are in blue, and similar residues are in magenta.

As both EL4b and the PxS motif are highly conserved, it is important to understand how it may affect the structure and pharmacology of NSS transporters. LeuT is an excellent model system for studying the motif as it not only shows significant desipramine potency, but its structure in complex with it has been solved to high-resolution. However, there is currently no pharmacological data in regards to mutations within EL4b in LeuT, and pure pharmacological studies do not take advantage of the rich body of biophysical and structural work on LeuT. To investigate the structural and binding changes resulting from mutations at the PxS motif region of
EL4b, computational methods can be used to model structural changes, analyze how the bound complex samples conformational space, and calculate changes in the free energy of binding.

To investigate the effects of the introduction of the PxS motif to LeuT structure and inhibitor binding, our group created computational models of several LeuT mutants bound to desipramine and studied their dynamics and binding energies using molecular dynamics and free energy perturbation calculations. If the direct interaction between LeuT and desipramine is changed due to the introduction of the PxS motif, it would be expected that there would be an observed change in the free energy of binding, and that there changes may accompany conformational changes within S2 or EL4. However, if the motif is involved in the dynamics of the transporter, such as how often the transporter samples a conformation which is interacting with desipramine, changes in the free energy of binding may not be clearly linked to large structural changes. To investigate changes in dynamics, principle component analysis (PCA) and root mean square fluctuation (RMSF) calculations were used to investigate the flexibility and conformational diversity in EL4, giving rise to a better understand as to how LeuT samples its bound ensemble and how changes in sampling are related to binding energy.
Methods

I. Molecular Dynamics Theory

Chemists have long sought to calculate the behavior of chemical systems with respect to time such that they may provide accurate deterministic predictions of theoretical problems in order to guide the rational design of experiments. Both quantum mechanics and statistical mechanics have explained many physical phenomena as well as made predictions that have been tested and confirmed using experimental methods. Using quantum mechanics, predictions can be made for small system over an extremely short amount of time, but these systems are mostly limited to gas phase at 0 K due to the computational limitations of current computer technology. Thus, investigating problems with biological implications, such as calculating the conformational properties of a protein during important events like folding and ligand binding is far past the computational limit for quantum methods. However, for systems as large as even a small hydrated protein, the Born Correspondence Principle states that their behavior approaches classical and can be accurately predicted using classical Newtonian physics. Thus, to computationally study biological systems, electron-implicit ball-and-spring models can be used, which greatly reduces the computational demand to the point where these calculations become practical and useful to chemists and biophysicists (Wainwright, 1959).

The classical potential energy of a system can be described as the summation of its elastic potential energy, electrostatic potential energy, and van der Waals potential energy. Elastic potentials can be used to describe the bonds between two
atoms A and B and the angle formed by three atoms A, B, and C by modeling the bond or angle as a spring between the exterior atoms and using Hooke’s Law. In order to model each bonded term as an elastic potential, there must be an associated force constant for each combination of atoms that is in a bonded relationship. To do this, force constants are determined both experimentally and by quantum mechanical calculation.

\[ U_{bond, \ angle} = - \int Fdx = \frac{kx^2}{2} \]

The dihedral angle, the angle between the planes formed by A-B-C and B-C-D, is described using a Fourier function, where \( k \) is the multiplicative constant.

\[ U_{dihedral} = \begin{cases} 
  k(1 + \cos(n\varphi + \phi)), & n > 0 \\
  k(\varphi - \phi)^2, & n = 0 
\end{cases} \]

Electrostatic potential for non-bonded atom interactions is described using the classical equation for potential energy between two point charges, Coulomb’s Law. The partial charge on any given atom is a function of the atoms that it is bound to, and thus as there is a force constant for each elastic potential, there are charge parameters which vary based on what the atom is bound to.

\[ U_{electrostatic} = \frac{q_a q_b}{r^2} \]

The potential energy contribution from van der Waals interactions is estimated using the Lennard-Jones potential for neutral particles, and as in the elastic and electrostatic potential terms, there are parameters for the Lennard-Jones constant for each interaction. One parameter, epsilon, is the well-depth and represents the strength
of interaction between the two atoms. The other parameter, sigma, is the distance the two atoms can approach each other. The Lennard-Jones equation is

\[ U_{\text{Lennard-Jones}} = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right] \]

where \( r^{12} \) is the repulsive term and \( r^{6} \) is the attractive term.

The entire set of force constants, charges, and Lennard-Jones parameters are known as a force field, and many have been optimized for specific problems and are constantly being updated with the newest and most accurate experimental and theoretical calculations. A force-field which is atom-explicit is called an all-atoms force field, whereas those who merge the collection of atoms into single entities are known as united-atom or coarse-grain force fields.

Using a force field built for biological molecules, classical mechanics calculations can be achieved using Newton’s equations for force and motion.

\[ F = ma = m \frac{dv}{dt} = m \frac{d^2x}{dt^2} \]

For conservative forces one can write

\[ F = -\nabla U \]

The displacement of an atom as the system evolves in respect to time can be calculated from the mass, acceleration, and initial position, in combination with the potential energy functions described above.

To begin a molecular dynamics calculation, there must be a set of initial positions and velocities for each atom in the system. However, the calculation of a biological system would not be accurate if it did not account for the presence of ambient kinetic energy, as no ambient kinetic energy would correspond to a system at
0 K, and biological systems do not exist at 0 K in nature. In order to study the
dynamics of a biological system, the system must be set to the proper temperature.
Temperature is simply a measure of the average kinetic energy of the system, so in
order to add an initial temperature parameter, an initial distribution of velocities can
be used to describe the distribution of kinetic energy in the initial system. To
distribute these velocities, a Maxwell-Boltzmann distribution for the kinetic energy of
gasses is used.

\[ f_p(p_x, p_y, p_z) = \left( \frac{1}{2\pi m k_b T} \right)^{\frac{3}{2}} e^{-\frac{p_x^2 + p_y^2 + p_z^2}{2m k_b T}} \]

At any point in time during a trajectory, the temperature of the system can be
calculated.

\[ T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{|p_i|}{2m_i} \]

In a biological system, the temperature is relatively constant. To keep the temperature
constant, Langevin dynamics are often used. Using Langevin dynamics, the velocities
of all atoms are constantly adjusted such that the temperature is kept constant. This
type of temperature control simulates a temperature bath in which that bulk water can
absorb and dissipate any energetic changes. Thus, the system modeled is known as a
closed system or a canonical ensemble, as energy can be exchanged but matter
cannot.

If an initial system begins at a high potential energy, it is possible that
introducing kinetic energy into the system will cause it to become unstable and leave
its local energy minima. If one wishes to understand how a protein samples in a
specific energy well, such as the case in short simulations, one must start within that well and relax the system by minimizing the potential energy. Once the system has settled into the bottom of a local potential energy well, kinetic energy can be added. By keeping temperature constant throughout the experiment, any conversation of potential energy to kinetic energy will result in global temperature damping, thus lowering the total energy of the system. This ensures the system will reach its equilibrium potential energy well, but allows the system to move between its thermally available conformations in that well.

The partition function describes the ensemble of states which makes up the equilibrium ensemble at a given temperature, while the Boltzmann distribution describes the probability the system is in one of the available substates. The Ergodic Hypothesis states that as time of sampling approaches infinity, the time average of the system becomes equivalent to the ensemble average of the system.

\[
\langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}
\]

Thus, by calculating the behavior of the system over an infinite amount of time, all conformations as well as their weight within the equilibrium ensemble can be determined, and the average behavior of the system over the time should correspond with the average behavior of the ensemble. However finite time periods should capture the most significant and likely most important system dynamics, and as macrostates with very little thermodynamic probability are essential irrelevant to overall system behavior, finite time calculations should reveal the ensemble behavior. In short simulations, conformational shifts through microstates in a single macrostate can be observed, such as the torsion of a side-chain or the exchange of a solvent.
molecule. In longer simulations, jumps between states which are separated by large energy barriers could be seen, such as between an open and closed state of a binding protein.

If one were to simply calculate the system coordinates for an initial and final time point, the motions that occur within that unit of time are lost. In order to provide a description of the systems behavior over time, the calculation must be broken into small units of times called time steps. Ideally, the time step should be infinitesimally small. However, this is largely unnecessary for a system which has physical limits on the rate of relevant change. Due the numeral method used to integrate Newton’s equations of motion, the standard time step in a molecular dynamics calculation is on the order of 1 – 2 femtoseconds. While this is far more efficient than infinitesimally small time step, it still creates a computational limit on the length of a molecular dynamics trajectory, as hundreds of thousands of pair-wise calculations must be repeats hundreds of thousands of times. For a large system, such as a small hydrated protein, the traditional limit is on the order of microseconds, but trajectories up to 1 millisecond have recently been achieved using supercomputers built for molecular dynamics computing (Shaw, 2010). However when only studying one conformational state and its local dynamics and properties, short simulations on the order of 1 to 10 ns can provide valuable information and accurate predictions.

NAMD is a molecular dynamics simulation package written using the Charm++ parallel processing model and was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign (Phillips, 2005). The
package was developed specifically to model large systems, with the ability to compute systems in the millions of atoms. Thus, NAMD is an ideal simulation package for large biological systems, and the use of NAMD to model biological systems has been extensively verified in both aqueous and membrane biomolecules, such as DNAse-1 (Reisler, 2010) and LeuT (Madura, 2008). The CHARMM general force field (CGFF) is an ideal force field for studying drug-like molecules in biological system (MacKerell, 2010), and has been shown to be compatible with the NAMD package.

II. Free Energy Perturbation Theory

To calculate absolute binding free energy, molecular dynamic calculations may be used in theory. However, traditional molecular dynamics calculations are inadequate due to the immense about of conformational space that the ligand would likely need to sample before a binding event would occur. Even in the smallest system, this could take far too long. However, the binding energy be calculated indirectly using a thermodynamic cycle (McCammon, 1984). In the cycle shown in Fig. 9, there are two reactions which produce distinct states in which the binding free energy can be calculated from: the annihilation of the ligand from the bound protein system, and the annihilation of the ligand from the bulk water. These reactions describe distinct states where there is little conformational

![Figure 9 - The thermodynamic cycle for desipramine dissociation.](image-url)
sampling possible, and thus their free energies can be determined accurately in a computationally-practical manner using this double annihilation method. However, simply calculating the difference between the bound and unbound states is problematic because in order for the calculation to converge properly, the energy change must be small, and the ensemble of the initial and final states most overlap. Free energy calculations can be achieved by the use of a coupling parameter $\lambda$, where $\lambda = 1$ is the bound state and $\lambda = 0$ is the unbound state. In order to calculate the difference in energy between the bound and unbound states, one must simply solve the Zwanzig equation:

$$
\Delta G(A \rightarrow B) = G_B - G_A = -k_B T \ln e^{\frac{U_B - U_A}{k_B T}} \geq_A
$$

Instead of calculating the difference between $\lambda = 0$ and $\lambda = 1$, smaller windows can be used, such as calculating the change for $\lambda = 0$ to $\lambda = .02$. In order to produce states where $\lambda$ is between 0 and 1, the atoms corresponding to the ligand can be gradually annihilated throughout the course of the calculation by treating the non-bonded terms as functions of $\lambda$, where as $\lambda$ approaches 0, the non-bonded terms contribution approach 0. To calculate the free energy change for a window, the energy of the initial state is calculated, and then the final state is created by increasing the $\lambda$ parameter, equilibrating the new system, and then calculating the energy change. Each window calculation produces the initial state for the following window, allowing for a continuous and efficient calculation of the free energy change over time (Beveridge, 1989). Assuming that the unbound ligand is in bulk solution and
there are no interactions between the protein and ligand, the annihilation of the ligand can be performed in water without the protein present.

The NAMD package can be used to calculate free energy perturbation. The free energy perturbation method and the double annihilation technique have been used to calculate the energy of binding with high accuracy (MacCammon, 1997). Furthermore, the approached has been verified to be accurate in membrane protein models.

III. Principle Component Analysis Theory

In order to study conformational substates within a molecular dynamics trajectory, the statistical method of principal component analysis (PCA) is often used. PCA is an orthogonal linear transformation of a set of correlated data into a new coordinate system such that the largest variance in the data is in the first coordinate and the second largest variance in the second coordinate, and so forth. PCA is used to transform a set of possibly correlated observable into a set of uncorrelated observables called principal components. When analyzing a molecular dynamics trajectory using PCA, each data point is the set of all atomic coordinates at a point in time. Thus, PCA can be used to create a new coordinate system which best differentiates configurations of the system. By creating many principal components, many conformational substates can be differentiated in a hierarchical manner. However, two-dimensional and three-dimensional PCA are generally sufficient to differentiate major conformations and transitions within a trajectory.

Root-mean square fluctuation (RMSF) describes how much an atom fluctuates from its average coordinate on average, and is the computational equivalent to B-
factor; a large RMSF represents a flexible region of the backbone, whereas a low RMSF represents a rigid region. RMSF is useful in determining robust changes in conformational sampling due to perturbation.

Bio3D is a package within R, a statistical environment, used for the analysis of protein structure and dynamics. Bio3D was developed by Barry Grant at the McCammon lab at the University of California at San Diego (McCammon, 2006). Bio3D can perform several statistical analyses, such as PCA, root mean square deviation and fluctuation, and covariance matrices. Bio3D was used to create 2-dimensional PCA analysis as well as to calculate RMSF.

VI. Creation of LeuT-Desipramine Mutants

Twelve mutants were designed for every combination of the A335P, G336L, and G337S mutations. To create LeuT mutants already bound to desipramine, mutations were applied to the desipramine-bound LeuT crystal structure (2QB4) after detergents and a secondary desipramine, bound in the transmembrane region, were removed. Waters found in the crystal structure were not removed, as they may be involved in binding site interactions, and the structure was then placed in a water sphere of radius 54 containing 50551 total water molecules. In order to apply mutations, the Mutator plugin of VMD was used. The Mutator plugin replaces the atoms of a single amino acid with the atoms of the new amino acid, using the coordinates found in the standard CHARMM topology and fitting them into the existing backbone. While membrane protein systems are traditionally modeled in implicit membranes, free energy perturbation calculations become far too computationally expensive due to the large sampling required. Because the most
significant conformational change should occur in EL4, harmonic constrains were applied to all a-carbons outside of EL4 in order to hold the crystal structure orientation of the transmembrane domains while allowing for rearrangement related to the mutations. The previous crystal structures of LeuT indicate is that there is little difference between the transmembrane helices of inhibitor-bound and unbound LeuT, and thus it is unexpected that the constraints on the transmembrane helices will affect the binding energy. These constraints were used for both the molecular dynamics as well as free energy perturbation calculations.

V. Molecular Dynamics Production

Molecular dynamics were calculated at 310 K with time steps of 2 femtoseconds. Each mutant was initially minimized for 20 ps, and then molecular dynamics were produced for 500 ps of equilibration. The Combined CHARMM All-Hydrogen Parameter File for CHARMM22 Proteins and CHARMM27 Lipids in combination with desipramine parameters which were developed by Kalyan Chakravarthi in the Madura Group based on the Topology for the CHARMM General Force Field v. 2a5 for Small Molecule Drug Design. The specifications of the molecular dynamics productions are presented section I of the Appendix.

IV. Free Energy Perturbation Calculations

Free energy perturbations were calculated with windows of $d\lambda = 0.04$. In each of the simulations, the desipramine atoms were annihilated as $\lambda$ approached 1. For each window, the system was equilibrated with the new $\lambda$ parameter for 50,000 steps, with one femtosecond time steps, and then sampled for 350,000 steps, at which
energy calculations are collected. The specifications of the free energy perturbation calculations are presented in section II of the Appendix.

V. Analysis of Molecular Dynamics Trajectories

Molecular dynamics trajectories were analyzed using both VMD and the Bio3D package in R. Structural changes were analyzed using the RMSD calculator within VMD. Root mean square fluctuation (RMSF) and PCA calculations were achieved using Bio3D. For both RMSF and PCA, the minimization and first 50 picoseconds of the equilibration were excluded in order to capture only equilibrium dynamics. In order to differentiate the difference between sampling in each mutant, the trajectories of all 12 systems were fit to the 2QB4 structure and then combined for PCA analysis.
Results

The results for each system can be found in its corresponding profile within the Appendix.

I. Mutants are converged and equilibrated

The molecular dynamics simulations of each mutant displayed convergence to a potential energy minimum early within the trajectory of 500 picoseconds. The potential energy over time is plotted for each mutant in Fig. 1 of their profile in the Appendix. All mutants display potential energy convergence within less than 100 picoseconds. Fig. 1 in each profile also displays temperature vs. time. As temperature is being held constant in the simulation, it is important to display convergence to the target temperature as well as that the temperature is maintained properly throughout the calculation, as is seen in each simulation. In addition, as a protein converges to a minimum in the free energy landscape, one would expect that it would begin to sample less conformational space as the total energy available to the system decreases and high energy conformations are no longer accessible. The RMSD for each mutant from its starting conformation converges within 100 ps, as seen in Fig 2a of each profile. All together, the data suggest that each simulation resulted in a system equilibrated within a local energy minimum.

The introductions of harmonic constraints conserved the crystal structure coordinates within 0.7 Å for the wild-type, and there was very little variance in the overall structure of each mutant from the wild-type. While these constraints do not serve as an absolute substitute for an implicit membrane, it appears they were able to
maintain the crystal structure TM coordinates. Most importantly, there did not appear to be substantial change to the orientation of the domains around S2, so while there was some slight rearrangement, it is unlikely it will affect the free energy perturbation calculations.

II. Mutations result in structural changes in EL4

All mutations resulted in structural changes local to the point of mutation, with all mutants, excluding A335S, G336L, and A335P displaying greater than 1.0 Å RMSD from the wild-type structure in residues 335 to 337 (See Table 1 of each profile in Appendix). It is likely that the A335S mutation results in little local change because the serine and alanine residues have comparable molecular volume, and thus no steric clash is introduced. The G336L and A335P mutants may also display little structural change because the loss of flexibility is mitigated by the neighboring glycine, and indeed mutants containing the A335P mutation with either one or both of the glycine mutations all resulted in substantial local change. Most mutations also resulted in structural changes throughout EL4, as defined as residues 292 to 337, but not as significantly as seen locally (See Fig. 3a of each profile in the Appendix). Interestingly, there was very little change at F320 and A319 in any of the mutants. Fig. 3b of each profile in the Appendix displays S2 for that mutant aligned with the wild-type. While all S2 residues other than F320 and A319 were constrained, there is still little observed change in the packing of S2.

III. Mutations result in only small changes to the RMSF within EL4

There were only small changes in RMSF in EL4 for each mutant in comparison to the wild-type (See Fig. 2b of each profile in the Appendix).
Surprisingly, many mutations led to increased RMSF at the point of mutation. This was unexpected as the A335P, G336L, and G337S mutations would all be expected to limit the local conformational flexibility by restricting backbone torsions. It is clear that in each mutant, the changes in flexibility tend to occur local to the mutations or at the hairpin tip of EL4a. This may indicate that mutations have most significantly altered the dynamics at these two regions. However, the magnitude of change in RMSF is quite low. Limited flexibility is expected in a bound complex, and as the RMSF for any given residue in EL4 never exceeds 1 Å, the differences between mutants may be insignificant and reveals little as to the effect of the mutation on desipramine binding.

IV. Mutations result in changes to desipramine binding affinity

Free energy perturbation calculations revealed that all but two mutations had large negative effects on binding affinity. The mutations can be easily clustered into two groups: conserved affinity and no affinity. A335P/G336L/G337S and A335S/G336L/G337S both maintained wild-type affinity or greater (-16.324 and -18.867 kcal/mol, respectively). Each Appendix profile displays the calculated free energy of binding for each mutant. Interestingly, the A335S/G336L/G337S mutant showed greater affinity for desipramine than the wild-type and A335P/G336L/G337S. In hDAT, the P401S mutation has been shown to increase both despramine and nortriptyline potency (Giros, 2001), and it is possible these two mutations change desipramine affinity through comparable modes. The remaining mutants displayed positive binding energy, indicating they would not bind to desipramine.

A335S/G337S, which resembles NET, showed a complete loss of affinity. While this
is unexpected, the P401S/L402G hDAT mutant has also been shown to reduce TCA potency (Giros, 2001). None of the mutants which displayed positive binding affinity displayed the spontaneous dissociation of desipramine during their equilibration or free energy perturbation calculations. This may indicate the activation energy associated with the desipramine association is too high to be accessed in a short simulation.

V. Mutations alter how EL4b samples conformational space

In order to maximize the variation which could be described by PCA, residues 319 to 337 were chosen for analysis. This selection includes both the site of mutation as well as the residues in the EL4 hairpin that should be interacting with desipramine, and these two sites would not only be expected to vary greatly throughout the mutants, but also should play a large role in the free energy of binding. The first three eigenvalues accounted for 36.4%, 25.7% and 9.5% of the variance respectively, for a total of 71.6% of the variance in the data (See Fig. 10). This selection of residues not only maximized the variance which could be described with a 2-dimensional PCA, but also maximized the difference between the second and third eigenvalues.

In order to investigate general changes in EL4 dynamics introduced by each mutation, each time point in the trajectory of each mutant was plotted with the corresponding time points in the trajectory of the wild-type in a plot of PC1 vs PC2.
Using just the first two eigenvalues to describe the conformational space being sampled, it is clear that the G337S and A335P mutations change the dynamics of EL4 significantly, as there is little overlap between the mutants and the wild-type. However, there are minimal changes from the A335S, A335S/G336L, A335P/G336L, and G336L mutations. Also, as both triple mutants show very little overlap with the wild-type, there does not appear to be a clear connection between the conformational space being described using PCA and desipramine binding energy. Nonetheless, it is clear that many of the mutants sample significantly different conformational space.
Discussion

I. The PxS motif is related to desipramine binding

Many mutations to the PxS motif resulted in large changes to desipramine free energy of binding. Interestingly, the A335P/G336L/G337S mutant, which is the full insertion of the PxS motif, resulted in no change in binding energy, while the addition of this motif in hNET experimentally resulted in drastic changes in desipramine potency. Unexpectedly, the A335S/G337S mutant, corresponding to NET, showed a loss of affinity while NET shows much greater inhibition by desipramine than LeuT. Clearly, the PxS motif is somehow related to the binding of desipramine, and in a way robust enough such that affinity can be completely abolished. Only two of the mutations which resulted in a positive desipramine binding energy, A335S/G337S and A335P, can be found in the NSS transporters. Conversely, the remaining mutants do not appear to occur within the mammalian family, and the strong bias against the positive binding energy mutations indicates that while the mechanism by which these mutations affect binding energy may not be clear, this region of EL4b is essential to controlling S2 binding in a way that evolution indicates is important.

While the existence of a secondary substrate which binds in S2 is still arguable, the apparent evolutionary drive against those motifs which disrupt binding of inhibitors in S2 may indicate that some binding event in S2 is critical to function. While it is possible that S2 plays a role in the initial sequestering of substrate, aiding the transport from the extracellular space to S1, it could also be possible that S2 is a site used for endogenous regulation of transporter function. If the secondary substrate
is the same as the primary substrate, which has been proposed for both LeuT and DAT, it is possible that transporters have evolved S2 to act as a secondary concentration-dependant transport control. It is also possible that under certain circumstances, it would be beneficial to prolong a neurotransmitter signal by slowing reuptake, and thus some sort of endogenous inhibitor may exist. The salt-bridge formation with D402 with desipramine in the LeuT-desipramine structure, which is believed to stop the release of substrate, may be evidence of an existing endogenous mechanism to slow reuptake. Regardless, the existence of S2 and the possibility that a highly conserved region of EL4b modulates its affinity for inhibitors indicates there may be some interaction occurring at S2 in vivo.

II. Energetic changes and structural changes may not be clearly related

It is often assumed that function can be clearly derived from structure, and thus one would expect that changes in binding energy should be accompanied by obvious structural changes. Yet, none of the mutants displayed large changes in the structure of S2 or in the orientation of F320 and A319, while nearly all the mutations resulted in positive binding affinity. On the other hand, the A335P/G336L/G337S and A335S/G336L/G337S mutants displayed larger structural change locally and in EL4, but both resulted in wild-type affinity. These results indicate that there is not always a clear relationship between the average structure of a protein, such as that seen in an x-ray structure, and the binding energy. The binding energy calculated using experimental or computational methods is actually an ensemble average over each state sampled within the ensemble. While the average structure may present a clear picture of the most highly populated conformation, it is possible that conformations
which are only lowly populated could play a significant role in binding. As the first
two principle components did not seem to indicate that conformational sampling was
disrupted by many of the mutations, it must be the case that the binding energy was
significantly modified by very subtle changes in loop dynamics and the gain or loss of
various less-sampled microstates. If so, crystal structures of these LeuT mutants may
provide very little information as to how the mutant effects function. It may be
possible that LeuT mutants would be better characterized by use of NMR and single-
molecule FRET, which could capture changes in dynamics.

III. EL4b may be a dynamic control center

The PCA data reveals that the G337S and A335P mutations, several double
mutants, and both triple mutants appear to dramatically change the conformational
sampling of EL4. As both the proline and serine are strongly conserved in
mammalian NSS transporters, this change in conformational sampling may account
for a major difference in inhibitor binding between the bacterial LeuT and the
homologous mammalian NSS transporters. However, the A335S and G336L
mutations appear to have little change to the conformational sampling as described by
PCA, and this may indicate that these mutations subtly affect the conformational
sampling in EL4, and may not have substantial effects on overall EL4 dynamics. The
sequence in EL4b may be highly conserved in order to maintain the proper sampling
of NSS transporters, playing the role of dynamic control center by modulating both
major and minor dynamics. As a binding site must contain the proper residues to
interact with its ligand, there is little room for dynamic control within the binding site
itself without sacrificing enthalpically favorable interactions. Thus, remote dynamic
control centers like what is seen in EL4b may be efficient and common methods of fine-tuning binding. Curiously, the SLS sequence is not seen in the mammalian NSS transporters despite its affinity for desipramine. As its sampling does not overlap with the wild-type or A335P/G336L/G337S, the serine could induce some change in the sampling of EL4 which is crucial to another interaction the transporters are involved in. It is possible that the PxS sequence is highly conserved because it is able to modulate the dynamics of EL4 such to maximize its ability to perform each of its functions, while other sequences which still lead to desipramine binding lead to loss of function in other interactions.

VI. Future directions

In the preceding simulations, the assumption that harmonic constraints across the protein backbone would hold it to a conformation similar to the membrane-explicit structure may have biased the free energy calculations, and as one would not expect such drastic changes in binding energy based on existing pharmacology in NET and DAT, the existence of bias is probable. In order to validate the accuracy of the data, these simulations should be repeated using a more accurate membrane-explicit method, or an all-atom membrane-implicit system. It could be possible that the membrane-protein interactions involving residues around S2 could play a much more significant role in the free energy perturbation calculations. It is also possible that conformational change accompanying the binding of ligand are much greater than expected from comparing the 2A65 “occluded” and 2QB4 “inhibited” crystal structures. These aspects could be investigated much more accurately in an all-atom-explicit system.
Furthermore, to investigate whether these effects can be generalized to the binding of all S2 inhibitors, the calculations should be repeated for both clomipramine and imipramine, which have been crystallized in complex with LeuT. Also, as various inhibitors have been indicated to interact in S2 using mutational studies, those inhibitors could be studied using docking methods. In addition, in order to determine if the effects observed in LeuT are representative of a general trend in the family, these calculations should also be repeated on homology models of the NSS transporters.

As there has yet to be a fully comprehensive force field which provides experimentally accurate results habitually, computational studies of protein mutants only provide a prediction as to the altered structural and functional properties of the mutant, and even all-atoms simulations may provide extremely inaccurate results. Therefore, all computational predictions should be experimentally verified using in vivo and in vitro methods. To confirm these computational findings, the LeuT mutants presented could be studied structurally using X-ray crystallography, both in complex with inhibitors and without, as well as using kinetic and inhibitor binding studies.

Here, a novel dynamic control site was located by consolidating research from several fields within neurobiology and biochemistry. The discovery and understanding of these sites could not only play a significant role in the growing understanding of protein mechanism, but also the engineering of efficient synthetic proteins. Proteins are exceptionally complicated molecular machines, and it is very difficult to characterize their mechanism using pair-wise interactions and
descriptions; however as the field begins to discover more of the mechanism nature
has used to design life, the more likely it will be that some day we can use these
designs to benefit society. Thus, the combination of bio-informatics with biophysical
and biochemical methods to characterize more dynamic control sites is crucial to
development of all fields of the life sciences.
Conclusions

The bacterial leucine transporter LeuT has been the model for the structure of the NSS and monoamine transporters since its crystal structure was solved in 2005. While the LeuT sequence and structure is clearly homologous to that of the NSS and monoamine transporters, possession of similar structural characteristics does not lead to identical functional and mechanical characteristics. A man’s toolbox may contain many tools, and while there may be numerous hammers, the owner would surely claim that while all hammers have a similar form, they are not the same. The subtle differences between a carpenter’s hammer and an upholstery hammer may be missed by the untrained eye, yet it is the subtle differences between the two which define their existence as unique tools for unique purposes and justify the inclusion of each hammer in the toolbox. The designs of these two hammers were certainly not independent; the basic hammer structure provided a scaffold for which engineers could build a more specific tool. In the same sense, nature would not have evolved several different transport proteins if each did not perform its function better than the others.

Here, it is shown that subtle differences in EL4b can have drastic changes on how LeuT binds inhibitors in S2. It may be no coincidence that EL4b is weakly conserved between LeuT and the NSS transporters, as it is the overall fold which is the transporter scaffold, not the individual residues that make it up. While LeuT may be inhibited by several TCAs, its own evolution may have led it to bind these drugs in a very different way than the monoamine transporters, and thus using the LeuT-TCA
crystal structures as models for TCA binding may be like trying to upholster using a carpenter’s hammer.

It is clear that events such as ligand binding are highly coordinated and the residues essential for ligand binding may lie far from the immediate binding site. A protein at equilibrium is not a static structure; dynamic motions indicate a constant flow of energy through the protein, and residues far outside the site of interest may have evolved very carefully to direct and perfect essential motions. If the role of LeuT is to sequester essential cellular building blocks, it has likely evolved to do so with distinct kinetics and regulation, and thus there are subtle, yet essential, aspects of its structure which make it an efficient leucine uptake mechanism. These aspects may not be conserved in the mammalian counterparts who must uptake neurotransmitters in a manner which temporally coordinates action potentials while retaining specificity and a high level of regulation, and may have been lost in lieu of the proper function-specific characteristics.

For millennia, evolution has created sophisticated, efficient machines from a limited number of parts; it is unlikely that the force that led to a structure as intricate and well-regulated as the human body would not also perfect each element for its specific contribution to the overall system. In order to forward our understanding of the living organism, it is of great importance that these considerations be taken when practicing science, such that we do not over generalize or under appreciate the beautiful engineering that has created the wonderful world around us.
Acknowledgements

I’d like to thank Prof. Jeffry Madura for the opportunity to work in the Madura group at Duquesne University and the funding to complete my research. I would also like to thank Bonnie Merchant for help with molecular dynamics, James Branco for help with free energy perturbation, and James Thomas help with principal component analysis. I would also like to thank Prof. Erika Taylor for her constant advising and guidance as I wrote my thesis and Prof. David Beveridge for helping me get involved with the Madura group.

This work is supported by the National Institutes of Health, National Science Foundation/Department of Defense, The U.S. Department of Education and the Teragrid under award numbers R01DA027806, CHE-1005145 (REU/ASSURE), CHE-0723109 (MRI), P116Z080180 and TG MCB060061N.
Appendix

I. Molecular Dynamics Specifications

```
set molname       NAME
structure       ${molname}.psf
coordinates    ${molname}.pdb
set temperature 310
set outputname  ${molname}harel4
firsttimestep    0

# Input
paraTypeCharmm on
parameters
k_par_all27_prot_lipid.prm
temperature $temperature

# Force-Field Parameters
exclude scaled1-4
1-4scaling    1.0
cutoff       12.0
switching on
switchdist    10.0
pairlistdist  13.5

# Integrator Parameters
timestep       2.0
rigidBonds all
nonbondedFreq   1
fullElectFrequency  2
steppercyle  10

# Constant Temperature Control
langevin on
langevinDamping 5
langevinTemp $temperature
langevinHydrogen off

# Output
outputName     $outputname
restartfreq    500
dcdfreq        250
outputEnergies 100
outputPressure 100

# Spherical boundary conditions
sphericalBC on
sphericalBCcenter 26.314733505249023
30.784503936767578
21.567747116088867
sphericalBCr1 54.0
sphericalBCk1 10
sphericalBCexp1 2

# Harmonic Constraints
constraints on
consexp 2
constraintScaling 1.0
consRef
${molname}el4_ca.pdb
consKFile
${molname}.el4_ca.pdb
consKCol B

# Minimization
minimize 10000
reinitvels $temperature
run 250000
```
II. Free Energy Perturbation Specifications

# FLEXIBLE CELL
useflexiblecell no

# FORCE FIELD
parameters
k_par_all27_prot_lipid.prm
paraTypeCharmm on

# INPUT
set molname 2QB4
structure ${molname}_ws.psf
coordinates
${molname}_harel4.pdb

# SPHERICAL BOUNDARY CONDITIONS
sphericalBC on
sphericalBCcenter 27.665475845336914
21.22206115722656
22.3729646911621
sphericalBCr1 54
sphericalBCk1 10
sphericalBCexp1 2

# Harmonic Constraints
constraints on
consexp 2
constraintScaling 1.0
consRef
${molname}el4_ca.pdb
consKFile
${molname}el4_ca.pdb
consKCol B

# OUTPUT
outputenergies 500
outputtiming 500
restartfreq 500
outputname
${molname}_harel4_fep
restartname
${molname}_harel4_fep
binaryrestart yes
binaryoutput no

# PME
PME no

# SPACE PARTITIONING
splitpatch hydrogen
hgroupcutoff 2.8
stepspercycle 20
margin 1.0

# CUT-OFFS
switching on
switchdist 8.0
cutoff 12.0
pairlistdist 13.0

# RESPA
timestep 1.0
fullElectFrequency 2
nonbondedFreq 1

# 1-4 NON-BONDED
exclude scaled1-4
1-4scaling 1.0

# COM
commotion no

# SHAKE
rigidbonds all
rigidtolerance 0.000001
rigiditerations 400
# DCD
DCDfreq 1000
DCDfile ${molname}_harel4_fep.dcd

# COLVARS SECTION
colvars off

# FEP PARAMETERS
source fep.tcl
alch on
alchType FEP
alchFile ${molname}_harel4.fep
alchCol B
alchOutFreq 10
alchOutFile ${molname}_harel4.fepout

alchElecLambdaStart 0.4
alchVdwLambdaEnd 1.0
alchVdwShiftCoeff 5.0
alchdecouple on
alchEquilSteps 50000
alchLambda 0.0
alchLambda2 0.0
set dLambda 0.04
set Lambda0 0.0

# MINIMIZE
minimize 5000

# FEP RUN
while {Lambda0 < 0.28} {
    alchLambda $Lambda0
    set Lambda0 [expr $Lambda0 + dLambda]
    alchLambda2 $Lambda0
    run 400000
}
III. Wild Type (2QB4)

Fig. 1 – Potential Energy and Temperature during the 2QB4 simulation

Fig. 2a - RMSD during the 2QB4 simulation
Fig. 2b – Structural alignment of 2QB4 equilibrated structure and the crystal structure.

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD (Å) from x-ray</td>
<td>0.379</td>
<td>0.591</td>
<td>0.545</td>
<td>0.671</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of 2QB4 model (wild-type) from the 2QB4 x-ray structure.
Fig. 3 – S2

Fig. 4 – 2QB4 RMSF across EL4

Fig. 5 – PCA of 2QB4

$\Delta G_{binding} = -16.545 \text{ kcal/mol}$
VI. A335P

Fig. 1 – Potential Energy and Temperature during the A335P simulation

Fig. 2a - RMSD during the A335P simulation
Fig. 2b – Change in RMSF of A335P compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD (Å) from 2QB4</td>
<td>0.409</td>
<td>0.722</td>
<td>0.831</td>
<td>0.602</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335P from the wild-type
Fig. 3a – EL4 structural alignment of A335P (red) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335P (red) and the wild-type (cyan)

Figure 4 – PCA of A335P dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 64.421 \text{ kcal/mol} \]
V. A335P/G336L

Fig. 1 – Potential Energy and Temperature during the A335P/G336L simulation

Fig. 2a - RMSD across the A335P/G336L simulation
Fig. 2b – Change in RMSF of A335P/G336L compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.436</td>
<td>0.841</td>
<td>1.003</td>
<td>0.597</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335P/G336L from the wild-type
Fig. 3a – EL4 structural alignment of A335P/G336L (silver) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335P/G336L (silver) and the wild-type (cyan)

**A335P/G336L PCA**

Figure 4 – PCA of A335P/G336L dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 18.599 \text{ kcal/mol} \]
VI. A335P/G337S

Fig. 1 – Potential Energy and Temperature during the A335P/G337S simulation

Fig. 2a – RMSD during the A335P/G337S simulation
Fig. 2b – Change in RMSF of A335P/G337S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.394</td>
<td>0.635</td>
<td>1.399</td>
<td>0.474</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335P/G337S from wild-type
Fig. 3a – EL4 structural alignment of A335P/G337S (orange) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335P/G337S (orange) and the wild-type (cyan)

Figure 4 – PCA of A335P/G337S dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 20.406 \text{ kcal/mol} \]
VII. A335P/G336L/G337S

Fig. 1 – Potential Energy and Temperature during the A335P/G336L/G337S simulation

Fig. 2a - RMSD during the A335P/G336L/G337S simulation
Fig. 2b – Change in RMSF of A335P/G336L/G337S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.432</td>
<td>0.824</td>
<td>2.016</td>
<td>0.497</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335P/G336L/G337S from wild-type
Fig. 3a – EL4 structural alignment of A335P/G336L/G337S (yellow) and the wild-type (cyan)

Fig. 3b – S2 structural alignment of A335P/G336L/G337S (yellow) and the wild-type (cyan)

Figure 4 – PCA of A335P/G336L/G337S dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = -16.324 \text{ kcal/mol} \]
VIII. A335S

Fig. 1 – Potential Energy and Temperature during the A335S simulation

Fig. 2a - RMSD during the A335S simulation
Fig. 2b – Change in RMSF of A335S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from</td>
<td>0.385</td>
<td>0.549</td>
<td>0.708</td>
<td>0.369</td>
</tr>
<tr>
<td>2QB4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335P/G336L from the wild-type
Fig. 3a – EL4 structural alignment of A335S (silver) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335S (silver) and the wild-type (cyan)

Figure 4 – PCA of A335S dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 61.461 \text{ kcal/mol} \]
IX. A335S/G336L

Fig. 1 – Potential Energy and Temperature during the A335S/G336L simulation

Fig. 2a - RMSD during the A335S/G336L simulation
Fig. 2b – Change in RMSF of A335S/G336L compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.405</td>
<td>0.654</td>
<td>1.34</td>
<td>0.593</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335S/G336L from wild-type
Fig. 3a – EL4 structural alignment of A335S/G336L (green) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335S/G336L (green) and the wild-type (cyan)

A335S/G336L PCA

Figure 4 – PCA of A335S/G336L dynamics (red) compared to 2QB4 (black)

$\Delta G_{\text{binding}} = 20.735$ kcal/mol
X. A335S/G337S

Fig. 1 – Potential Energy and Temperature during the A335S/G337S simulation

Fig. 2a - RMSD during the A335S/G337S simulation
Fig. 2b – Change in RMSF of A335S/G337S compared to wild-type

Table 1 – RMSD of A335S/G337S from the wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.458</td>
<td>0.955</td>
<td>2.146</td>
<td>0.693</td>
</tr>
</tbody>
</table>
Fig. 3a – EL4 structural alignment of A335S/G336L (white) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335S/G336L (white) and the wild-type (cyan)

Figure 4 – PCA of A335S/G337S dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 22.042 \text{ kcal/mol} \]
XII. A335S/G336L/G337S

Fig. 1 – Potential Energy and Temperature over time in the A335S/G336L/G337S simulation

Fig. 2a - RMSD during the A335S/G336L/G337S simulation
Fig. 2b – Change in RMSF of A335S/G336L/G337S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.48</td>
<td>1.085</td>
<td>2.705</td>
<td>0.335</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of A335S/G336L/G337S from the wild-type
Fig. 3a – EL4 structural alignment of A335S/G336L/G337S (pink) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of A335S/G336L/G337S (pink) and the wild-type (cyan)

Figure 4 – PCA of A335S/G336L/G337S dynamics (red) compared to 2QB4 (black)

$$\Delta G_{\text{binding}} = -18.867 \text{ kcal/mol}$$
XIV. G336L

Fig. 1 – Potential Energy and Temperature over time in the G336L simulation.

Fig. 2a - RMSD across the G336L simulation
Fig. 2b – Change in RMSF of G336L compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.393</td>
<td>0.609</td>
<td>0.805</td>
<td>0.448</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of G336L from the wild-type
Fig. 3a – EL4 structural alignment of G336L (purple) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of G336L (purple) and the wild-type (cyan)

Figure 4 – PCA of G336L dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 57.851 \text{ kcal/mol} \]
XV. G337S

Fig. 1 – Potential Energy and Temperature during the G337S simulation

Fig. 2a - RMSD across the G337S simulation
Fig. 2b – Change in RMSF of G337S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.425</td>
<td>0.797</td>
<td>1.867</td>
<td>0.295</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of G336L from wild-type.
Fig. 3a – EL4 structural alignment of G337S (green) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of G337S (green) and the wild-type (cyan)

Figure 4 – PCA of G337S dynamics (red) compared to 2QB4 (black)

\[ \Delta G_{\text{binding}} = 16.994 \text{ kcal/mol} \]
XVI. G336L/G337S

Fig. 1 – Potential Energy and Temperature during the G336L/G337S simulation

Fig. 2a - RMSD during the G336L/G337S simulation
Fig. 2b – Change in RMSF of G336L/G337S compared to wild-type

<table>
<thead>
<tr>
<th>Residues</th>
<th>All</th>
<th>292-337</th>
<th>335-337</th>
<th>319-320</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD(Å) from 2QB4</td>
<td>0.434</td>
<td>0.872</td>
<td>1.909</td>
<td>0.353</td>
</tr>
</tbody>
</table>

Table 1 – RMSD of G336L/G337S from wild-type
Fig. 3a – EL4 structural alignment of G336L/G337S (purple) and the wild-type (cyan)
Fig. 3b – S2 structural alignment of G336L/G337S (purple) and the wild-type (cyan)

\[ \Delta G_{\text{binding}} = 59.957 \text{ kcal/mol} \]

Figure 4 – PCA of G336L/G337S dynamics (red) compared to 2QB4 (black)
References


Shimada, S. Cloning and expression of a cocaine-senesitive dopamine transporter complementary DNA. Science 254(5031), 1195 (1991)


Cowen, P.J. and Sharp, T. 5-HT and depression: is the glass half-full? *Curr Opin Pharmacol.* **11**(1), 45-51 (2011)


Madura, J.D., and Surratt, C.K. Dopamine transporter comparative molecular modeling and binding site prediction using the LeuT(Aa) leucine transporter as a template. *Proteins* **70**(3), 1033-46


Javitch, J.A. Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc Natl Acad Sci USA* **106**(14), 5563-8 (2009)


