Revealing the Mechanistic Diversity of the LeuT Fold: A Comparative Analysis of the Leucine Transporter and the Hydantoin Transporter using Electron Paramagnetic Resonance Spectroscopy

By

Kelli Nicole Kazmier

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

in

Chemical and Physical Biology

December, 2013

Nashville, Tennessee

Approved:

Charles E. Cobb, Ph.D.

Randy D. Blakely, Ph.D

D. Borden Lacy, Ph.D.

Jens Meiler, Ph.D.

To my family, Maxine, KaCee, Jayden, Brian, Dee, Joseph, and James

and

Especially to my partner, Luke

Thank you for supporting me through this journey

ACKNOWLEDGEMENTS

This work was funded in part by the National Institutes of Health predoctoral fellowships including a T32 institutional training grant through the National Institute of Neurological Disorders and Stroke and an individual F31 Ruth L. Kirschstein National Research Service Award through the National Institute of Mental Health. Without this financial support, the work reported here would not have been possible.

I would also like to thank the individuals who provided training and mentorship throughout the completion of this work. Specifically, I would like to thank my thesis committee, Dr. Chuck Cobb, Dr. Randy Blakely, Dr. Borden Lacy, and Dr. Jens Meiler, for their guidance and support. I would like to thank Drs. Derek Claxton and Rich Stein for teachin me to conduct and analyze DEER experiments. I am grateful for molecular biology training I received from Dr. Hanane Koteiche. I would like to thank Dr. Ryan Steed for imparting his knowledge of figure making and graphic design and Dr. Nathan Alexander for teaching me computer programming and introducing me to the Rosetta and BCL suite of programs. Special thanks go to Dr. Shruti Sharma for aiding in many of the experiments contained within this thesis.

Finally, I must acknowledge my mentors, Drs. Hassane Mcharouab and Jens Meiler. I am grateful to Jens for imparting his wisdom regarding computational structural biology and for the support and enthusiasm that he has provided me throughout the completion of my thesis. I would like to thank Hassane, my primary mentor, for inspiring my interest and offering me the tools to understand the principles of dynamic protein mechanisms. I am grateful for the independence that he afforded me to pursue the questions that I was interested in, to interpret the experimental results, and to write about and present my work. Finally, I must thank him for all of the guidance he has provided to me both scientifically and professionally. I will always be grateful for having been a member of the Mchaourab Lab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE STRUCTURE AND FUNCTION OF THE LeuT FOLD

Principles of alternating access in transporters

Transporters as cellular gatekeepers

The transfer of physiologically important molecules into and out of cells requires overcoming the physical barrier imposed by the lipid membrane. The lipid membrane is comprised primarily of phospholipids that orient into a bilayer organization in the aqueous conditions typical of cellular environments. The central hydrocarbon-rich region of the membrane is capable of excluding most hydrophilic molecules and functions to protect cellular integrity. To overcome this barrier, organisms have evolved dedicated proteins to specifically shuttle necessary molecules through the membrane in a regulated fashion. These proteins include channels, pores, transporters, and pumps, and operate using diverse mechanisms that facilitate transfer of essentially all physiologically necessary molecules such as ions, nucleic acids, amino acids, osmolytes, lipids, as well as macromolecules such as peptides and proteins.

Among these, transporters are primarily responsible for the transfer of small molecule substrates which underlie physiologically important processes such as regulation of glucose, recycling of nucleobases, neurotransmission, and efflux of toxic compounds. Transporters are functionally diverse, but tend to be grouped into two main categories: primary active transporters and secondary active transporters. Primary active transporters are defined by their direct utilization of chemical energy, typically though ATP-dependent mechanisms, to power conformational changes resulting in substrate transport. Secondary active transporters, on the other hand, employ cellular electrochemical gradients, often created through primary active

processes, to promote transport. While specific ion and substrate dependence varies significantly among secondary active transporters, functionality can be grouped into three general categories: uniport, symport, and antiport. Uniporters allow diffusion of specific chemical substrates through the membrane, relying on the chemical gradient of substrate to impose directionality. In contrast, symporters and antiporters couple the energetically favorable translocation of ions or substrates to energetically uphill transport of substrates against their concentration gradients. Distinction among these groups is defined by the directionality of ion and substrate transfer with antiport comprising opposite directionality and symport encompassing unidirectional mechanisms. Antiport has been further subdivided to include exchangers, which couple the transport of two substrates, often related precursor and product molecules, in opposite directions. Exchangers have been shown to operate in both ion-dependent and -independent modes.

Alternating access defines transport mechanisms

One of the earliest theoretical descriptions for transport detailed a general mechanism of alternating access where centrally located binding sites for ions and substrates were exposed to extracellular and intracellular sides of the membrane through conformational reorientations of the transporter structure[1-4]. The alternating access model stood in contrast to alternative models that described carrier proteins physically traversing the membrane to deposit substrates and ions on either side of the membrane[5]. While the carrier model may be present in ionophores such as valinomycin, it is clear that alternating access is a more appropriate description for transport protein processes as the energy required for the carrier model would be prohibitively high in most cases.

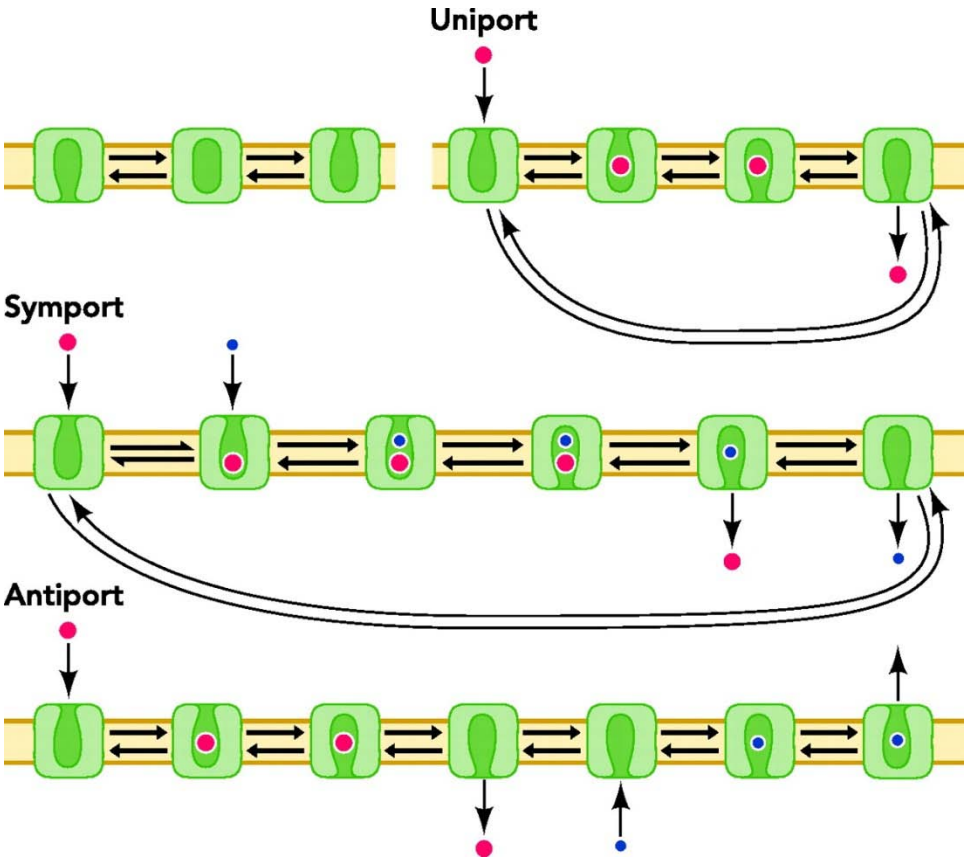


Figure 1.1. Theoretical models of alternating access. Figure reproduced with permission from [6].

Due to differential utilization of electrochemical gradients, uniport, antiport, and symport are expected to exhibit distinct alternating access mechanisms related to the free energy landscape and ligand-dependent sampling of alternative conformational states[7] (Fig. 1.1). Specifically, uniport requires a relatively flat energy landscape wherein apo and substrate-bound conditions are free to stochastically sample relevant conformations including outward-facing and inward-facing conformations. Like uniporters, the transition between inward-facing and outward-facing conformations in symporters would necessarily occur under apo conditions, suggestive of a relatively flat energetic relationship between states. However, the ligand dependence of symport remains poorly defined as it is currently unclear how symporters couple the co-transported solutes, how solute leak is prevented given the similar free energies of the

conformations, and to what degree ligand binding shifts equilibria between states to promote transport. In contrast, conformational transitions between inward-facing and outward-facing conformations are thought to be specifically ligand-dependent in antiporters, suggesting insurmountable energy barriers between states in the absence of energy provided by binding of ligand. Furthermore, antiporters have typically been expected to exhibit competitive relationships between ions and substrates, while this relationship would in theory be neutral or synergistic in symporters. It is currently unclear whether differences between uniporters, antiporters, and symporters entail merely differential free energy considerations for ion and substrate binding to the alternative conformations of the transporter or entirely different conformational sampling. Answers to these questions will require detailed structural and dynamic investigations of representative proteins of each transport class.

With the relatively recent emergence of 3D structures of transport proteins[8-11], theoretical descriptions of transport have begun to be experimentally evaluated and defined at atomic resolution. The leucine transporter (LeuT), a Na^+-coupled symporter from *Aquifex aeolicus* and a member of the neurotransmitter:Na^+ symporter family (NSS) became the founding member of the LeuT Fold structural class with the publication of its first crystal structure in 2005[12]. Subsequently, seven additional transport proteins were reported to display the LeuT Fold architecture[13-21] (Table 1.1). Crystal structures of these proteins have been reported in outward-facing, occluded, and inward-facing conformations defined by solvent accessibility to static, centralized ion and substrate binding sites. These structures support the alternating access hypothesis as the general mechanism of transport for LeuT Fold proteins. Using these structures in conjunction with functional, biophysical, and computational techniques to understand alternating access of the LeuT Fold proteins remains at the forefront of research in the field.

Common Name	Function	Type	Family	Coupled Ions	Known substrates	Oligomerization state
LeuT	Amino Acid Transporter	Symport	NSS	2 Na^+ [12]	L-Leucine[12] L-Alanine[22] L-Methionine[22] L-Tyrosine[22] Glycine[22]	Crystallographic Dimer[12]/ Functional Monomer[23]
vSGLT	Sugar Transporter	Symport	SSS	1 Na^+ [15]	Galactose[15]	Crystallographic Dimer[15]/ Functional Monomer[24-25]
Mhp1	Nucleobase Precursor Transporter	Symport	NCS1	1 Na^+ [14]	5-indolylmethyl-hydantoin[26] 5-benzylhydantoin[26]	Monomer[14]
BetP	Osmolyte Transporter	Symport	BCCT	2 Na^+ [27]	Betaine[27]	Trimer[16]
CaiT	Carnitine/ γ-Butyrobetaine Transporter	Antiport (Exchange)	BCCT	None[28]	L-Carnitine[29] D-Carnitine[29] crotonobetaine[29] γ-Butyrobetaine[29]	Trimer[30]
ApcT	Amino Acid Transporter	Symport	APC	1 H^+ [17]	L-Alanine[17] L-Glutamate[17] L-Serine[17] L-Glutamine[17] L-Methionine[17] L-Phenylalanine[17]	Monomer[17]
AdiC	Virtual Proton Pump	Antiport (Exchange)	APC	None[31-32]	L-Arginine[31-32] Agmatine[31-32]	Dimer[33-34]
GadC	Virtual Proton Pump	Antiport (Exchange)	APC	None[35-36]	L-Glutamate[35-36] γ-aminobutric acid (GABA) [35-36] L-Glutamine[13] L-Methionine[37] L-Leucine[37]	Dimer[37]

Table 1.1. LeuT Fold transporter descriptions and functional properties

Functional diversity and biological significance of LeuT Fold proteins

The emergence of the LeuT Fold as a common structural scaffold for transporter families was unexpected due to the lack of similarity in sequence or function. The LeuT Fold includes Na^+-coupled symporters like LeuT including the benzylhydantoin:Na^+ symporter[14] (Mhp1) of the nucleobase:cation symporter-1 family (NCS1), the betaine:Na^+ symporter[16] (BetP) from the betaine/carnitine/choline transporter family (BCCT), and the galactose:Na^+ symporter[15] (vSGLT) of the solute:Na^+ symporter family (SSS). However, additional functional modes are also

represented in LeuT Fold proteins including H^+-dependent transport in ApcT[17] of the amino acid-polyamine-organocation family (APC), H^+-dependent exchange in AdiC[19-20] and GadC[13] also of the APC family, and ion-independent exchange in CaiT[18,21] of the BCCT family. Each of these proteins catalyze transport of chemically unique substrates, promote distinct physiological processes, and represent diverse families of transporters (Table 1.1). It is a goal of this work, to chart the relationships between structure and function in the LeuT Fold to begin to understand how diverse functionality evident in LeuT Fold members can be accommodated on the same structural scaffold.

LeuT, bacterial homolog of neurotransmitter transporters

LeuT is prokaryotic member of the NSS family of transporters, also known as the solute carrier 6 (SLC6) family[38]. These transporters harness Na^+ and Cl^- electrochemical gradients to power transport of biogenic amines, amino acids, and osmolytes[38]. NSS members including human dopamine, serotonin, and norepinephrine transporters facilitate reuptake of neurotransmitters from the synapse into presynaptic neurons[38]. NSS function is vital for terminating neurochemical signals, maintaining intracellular neurotransmitter concentrations, and priming the cell for subsequent signaling events. Development of therapeutic strategies for a number of psychiatric illnesses such as depression, anxiety, epilepsy, attention deficit hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), as well as chemical dependence to drugs of abuse including cocaine and amphetamine requires a deep understanding of NSS function[39]. Toward this end, structural characterization of hyperthermophilic LeuT has been used to provide context for biochemical and functional data of eukaryotic NSS.

The use of LeuT as a model for NSS has been rationalized by sequence and predicted topological similarity between LeuT and NSS. Overall sequence identity between LeuT and other NSS members is low, in the range of 20-25%[12,40]. However, in certain regions identity can

reach as high as 50%[12,40]. With the publication of the LeuT structure, sequence alignment pinpointed these highly similar stretches of sequence to fall into functionally important regions such as substrate and ion binding sites[12,40]. While wild type LeuT functions in a Cl^--independent manner, it can be engineered to acquire the Cl^- dependence characteristic of eukaryotic transporters with a single point mutation[41-42]. Moreover, its transport activity is inhibited by tricyclic antidepressants[43-44] (TCAs) and selective serotonin reuptake inhibitors[45] (SSRIs), which are also inhibitors of the eukaryotic transporters, though through different interactions. Because of the similarities in sequence and function, LeuT has emerged as a paradigm for understanding the structural basis of alternating access in the NSS family.

LeuT transports small and medium-sized hydrophobic amino acids including Leu, Ala, Gly, Met, and Tyr[22]. The transport of biogenic amines, which in comparison lack the carboxy groups of amino acids, requires subtle differences in substrate coordination evident in different constituent binding site residues[12]. Na^+:substrate stoichiometries vary among members with 1:1 stoichiometries in the norepinephrine transporter (NET) and the serotonin transporter (SERT), 2:1 in the dopamine transporter (DAT), the γ-aminobutyric acid transporter (GAT), and glycine transporters GlyT1b, and 3:1 in GlyT2[46]. With a 2:1 stoichiometry, the functional mechanisms of LeuT may only be applicable to members with similar stoichiometry and the LeuT structure cannot predict the third Na^+ binding site[12]. Eukaryotic members differ from LeuT in a number of other ways. These include the fact that LeuT is not inhibited by classes of NSS antagonists. In addition, LeuT is a functional monomer that lacks C- and N-terminal regulatory domains as well as N-glycosylation sites present in some NSS. While these divergences temper applicability of some interpretations of LeuT with regard to the family as a whole, it is nevertheless most likely informative to NSS function generally.

vSGLT, a sugar uptake system

Sodium solute symporters (SSS or SLC5), of which the galactose transporter vSGLT of *Vibrio paraheamolyticus* is a member, couple the transport of sugars, amino acids, inorganic ions, and vitamins to the inward-oriented cellular Na^+ gradient[47]. Known disease states resulting from mutations in human glucose (SGLT1) and iodide (NIS) transporters have been shown to include glucose-galactose malabsorption (GGM) and iodide transport defect (ITD)[48-49]. Furthermore, these transporters are targets for oral rehydration therapy, type II diabetes, and obesity therapies. vSGLT has significant sequence similarity to other SSS members, with 32% identity to SGLT1 (60% similarity), 19% identity to NIS (58% similarity), and 18% identity to the proline symporter (PutP, 57% similarity)[15].

As the only structurally characterized member of the class, the vSGLT crystal structures[15,50] have become the standard for understanding the structural and mechanistic implications of an extensive array of biochemical and functional datasets. The vSGLT structure exhibits common features of sugar-binding site architecture, with a Y263 residue located in the substrate binding site providing stacking interactions with the pyranose ring of sugar moieties. Furthermore, OH-groups of the galactose ring are hydrogen bonded including coordination of the C4-OH by Y87 and the C2-OH by K294, another common feature of sugar-binding proteins. Mutation of conserved residues between vSGLT and SGLT1, including a known disease causing mutation located in the sugar-binding site, have been shown to ablate galactose transport in vSGLT supporting the conclusion that vSGLT is an appropriate model of mammalian SSS.

Mhp1, a model for nucleobase recycling

The nucleobase-cation-symport-1 (NCS1) family of transporters catalyzes uptake of nucleobases such as uracil, cytosine, and thiamine, as well as related derivatives for use as sources of energy and molecular precursor molecules[51-52]. Present in bacteria, fungi, and plants,

the NCS1 family couples substrate uptake to both Na^+ and H^+ gradients[53-54]. Mhp1 of *Microbacterium liquefaciens* functions as a part of a metabolic salvage pathway, wherein indolylmethyl- and benzyl-hydantoins are imported for conversion to amino acids, Trp and Phe[26]. The hydantoin molecule captured in the substrate-bound crystal structure of Mhp1[14] shows π-stacking interactions with the indole-ring of W117 and hydrogen bonds with Q121 and N318, residues that are almost completely conserved throughout the family suggesting broad applicability of the Mhp1 structure for NSC1 transporters, in particular for understanding substrate binding. Remarkably, the uracil:H^+ symporter UraA, a member of the functionally related NCS2 family, manifests a conformation distinct from Mhp1 and the LeuT Fold, although it also presents an inverted repeat topology[55]. UraA represents an interesting counterexample to the LeuT Fold structure-function relationship, where transporter proteins with similar functionality display distinct structural profiles.

BetP and CaiT: osmolyte transporters

BetP is a Na^+-dependent betaine (or glycine betaine) transporter of *Corynebacterium glutamincum* that participates in bacterial osmotic stress response[56]. Functioning as an osmotic sensor and regulator, it accumulates betaine, an osmolyte, from the extracellular environment to high intracellular concentrations, thereby counteracting osmotic driven water flux and ensuring cellular hydration and intracellular hydrostatic pressure[56]. By comparison, CaiT does not participate in osmoregulation. CaiT is an ion-independent exchanger of carnitine and γ-butyrobetaine, precursor/product related molecules involved in anaerobic growth pathway in bacteria, resulting in coupled import of carnitine and export of γ-butyrobetaine[56]. In humans, CaiT scavenges carnitine for its roles in fatty acid transport in the inner mitochondrial membrane and various metabolic pathways where carnitine deficiency has been shown to result in hypoglycemia, skeletal-muscle myopathy, and cardiomyopathy[57-59]. CaiT was crystallized from

both *Escherichia coli* and *Proteus mirabilis*[18,21]. A resulting comparison of the structure reveals nearly identical architectural features with overlapping carnitine and γ-butyrobetaine binding sites.

Both CaiT and BetP are members of the betaine/choline/carnitine transporter (BCCT) family. A comparative analysis of these protein structures reveals insight into a variety of BCCT functional mechanisms. Betaine and carnitine, like other osmolytes, are highly polar organic compounds that inherently segregate from protein surfaces. Therefore, understanding how these molecules are bound and transported is a relevant question in the field. The crystal structures of BetP[16,60] and CaiT[16,60] inform on this question, showing uniquely hydrophobic substrate permeation pathways, proposed to limit osmolyte repulsion. Both BetP[16,60] and CaiT[16,60] are functional trimers. In BetP the trimeric architecture appears to be intimately related to regulation of transport with inter-protomeric interactions mediated by the C-terminal regulatory domain[61]. However, this domain is not present in CaiT, and the role of oligomerization in regulation of CaiT is unknown. Comparative analysis also illuminates differences in ion-dependent and ion-independent mechanisms. The Na^+ binding sites identified in BetP are very similar to the sites identified in LeuT as Na1 and Na2[16]. In CaiT, these sites are replaced by a positively charged Arg in Na2 and a Met in Na1[21]. The Met mutation is conserved among mammalian organic cation/carnitine transporters[21] (OCTN) and also evident in LeuT Fold member ApcT[17]. Despite these differences, it is likely that other BCCT members also adopt the LeuT Fold topology, exhibit aromatic substrate permeation pathways, and exist in homotrimer oligomerization states.

APC superfamily representatives: ApcT and AdiC/GadC

The amino acid, polyamine, organocation (APC, multiple SLC families) transporters are a large and functionally diverse superfamily including uniporters, antiporters, and symporters[62] with roles in nutrient scavenging, pH regulation, promotion of insulin release, nitric oxide

synthesis, and cell volume homeostasis and are associated with known disease states such as asthma, cancer, cystinuria, and lysinuric protein intolerance[63-68]. ApcT is a broad specificity, H^+ dependent, amino acid transporter from *Methanocaldococcus jannaschii*. ApcT binds most canonical amino acids with a transport preference for intermediate sized moieties[17]. Transport is robust at low external pH and insignificant at neutral pH. Lys 158 is the predicted source of H^+ dependence and is located at a equivalent position to the Na2 site in Na^+ coupled LeuT Fold members[17]. As stated previously, the Na1 site in ApcT has been mutated to a Met as seen in CaiT[17]. Due to its similarity to LeuT and its response to externally orientated pH gradients, it is predicted to function as a symporter[17].

AdiC of *E. coli* and *Salmonella enterica* (95% identical) is another structurally characterized member of the APC family[19-20,69]. It participates in the extreme acid resistance system of enteric pathogens as a mechanism of coping with the low pH of the stomach of hosts[31-32]. Although, AdiC functions to exchange extracellular arginine for intracellular agmatine, this process is most significant for its virtual proton pump chemistry, with the imported arginine undergoing a separate process of decarboxylation to agmatine and agmatine export resulting in a net loss of 1 H^+ per transport cycle[31-32]. The glutamate: γ-aminobutric acid (GABA) antiporter, GadC, also participates in the acid resistance system of *E. coli* using similar virtual proton pump chemistry, i.e. decarboxylation of glutamate and export of reaction product GABA, to maintain intracellular pH[35]. GadC has significant sequence similarity to AdiC, and therefore, they are typically grouped together in descriptions of structure and function[35-36]. However, due to the broad functionality of the APC family, it is unclear how generally applicable descriptions of these proteins are to the family at large.

The emergence of a similar fold among these transporters was unexpected both due to the lack of sequence similarity, but as importantly, to the functional diversity of the represented protein families. It is currently unknown whether topological similarity in these cases translates

to similarities in protein functional mechanisms. Answering these questions will have important consequences for understanding the relationship between structure and function in transporters.

The LeuT Fold architecture

The LeuT structure

As expected from secondary structure prediction, the LeuT structure consists of 12 transmembrane (TM) helices connected by a number of structured loops on both the intracellular and extracellular sides of the protein[12] (Fig. 1.2). Unexpectedly, the structure contained an internal structural repeat with TMs 1-5 and TMs 6-10 related by a 2-fold pseudosymmetry (176.5˚) around an axis parallel to the membrane[12]. The inverted repeat motifs are superimposable to a C_α root mean squared deviation (RMSD) of 5.3 Å[12]. Deviations between repeats are primarily located in the first two helices of each repeat (TMs 1 and 2/6 and 7) and the structure of the remaining helices are highly similar in the overlay[70] (Fig. 1.2). The structural deviation between repeats is the result of the outward-facing nature of the structure and the orientation of this set of helices forms the basis for a number of mechanistic interpretations for the LeuT Fold described below.

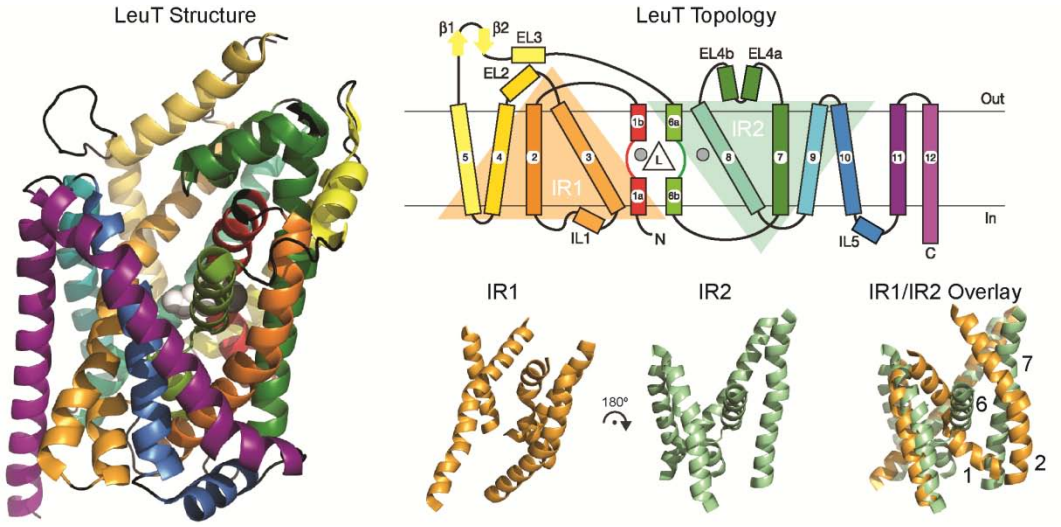


Figure 1.2. LeuT structure, topology, and internal inverted repeat structural symmetry. Figure modified with permission from [12]

The internal inverted repeat symmetry of LeuT was not evident from sequence and, as such, the two inverted repeats were predicted to be evolutionary unrelated[12]. TMs 1-10, that form the inverted repeat motifs, were considered to comprise the functional core of the protein as some LeuT homologues were predicted to have only 10 or 11 TM helices, rendering the additional TMs 11 and 12 non-conserved[71]. An evaluation of the structure further supported this assertion, with TMs 11 and 12 located peripherally to the TM 1-10 core domain[12]. The LeuT crystal structure was captured as a dimer, with the dimer interface manifesting as a 4 helix bundle including TMs 9 and 12 of both protomers as well as interactions with extracellular loop (EL) 2[12]. While LeuT is currently thought to be a functional monomer, non-core helices may serve roles in oligomerization in eukaryotic homologues[72].

Symmetrically related TMs 1 and 6 were identified previously as functionally important helices containing a high number of residues conserved throughout the NSS family[40]. These helices were found in the structure to be discontinuous with central residues V23 and G24 in TM1 and S256-G260 in TM6 exhibiting an extended structure[12]. The discontinuous region faces residues in TM 3 and 8 that are also highly conserved[12]. The original LeuT crystal structure[12]

was crystallized in the presence of Na^+ and Leu, which were bound to ion and substrate binding sites located in the discontinuous region (Fig. 1.3). The carboxy and amino groups of the Leu substrate molecule were found to be coordinated by hydrogen bonds provided by main chain carbonyl oxygens (residues A22, F253, and T254), amide nitrogens (residues L25 and G26) and side chain hydroxyl groups (residues Y108 and S256) as well as the dipole moments of the unwound helices in TMs 1 and 6. Equivalent residues to Y108 were previously implicated in substrate binding in eukaryotic homologues[73-74]. The aliphatic section of Leu was surrounded by a sterically complementary hydrophobic pocket including residues V104, Y108, F253, S256, F259, S355, and I359[12] (Fig. 1.3).

Substrate specificity resulted from both the specific partial charge environment and shape complementarity of the substrate binding site evident in a series of LeuT crystal structures bound to substrates including Gly, Ala, Leu, Met, tyrosine analog L-4-fluorophenylalanine, and nontransportable inhibitor, Trp[22]. A comparative analysis with LeuT homologues points to divergence in residue 24 between amino acid transport and biogenic amine transport. In amino acid transporters, residue 24 is a glycine. However, in biogenic amine transporters, this residue is an aspartic acid, which provides a carboxy group not present in the biogenic amines as compared to amino acids, which participates in Na^+ ion coordination[12]. Shape complementarily was particularly evident in the hydrophobic pockets of homologous proteins. In the glycine transporter, the hydrophobic resides are larger and positioned to create a smaller substrate binding cavity[12]. In contrast, equivalent residues in SERT are smaller and expected to form a larger binding pocket to accommodate the larger serotonin molecules[12]. Conserved residue F259 has been hypothesized based on its position to participate in π-π stacking interactions with aromatic rings of biogenic amine substrates[12].

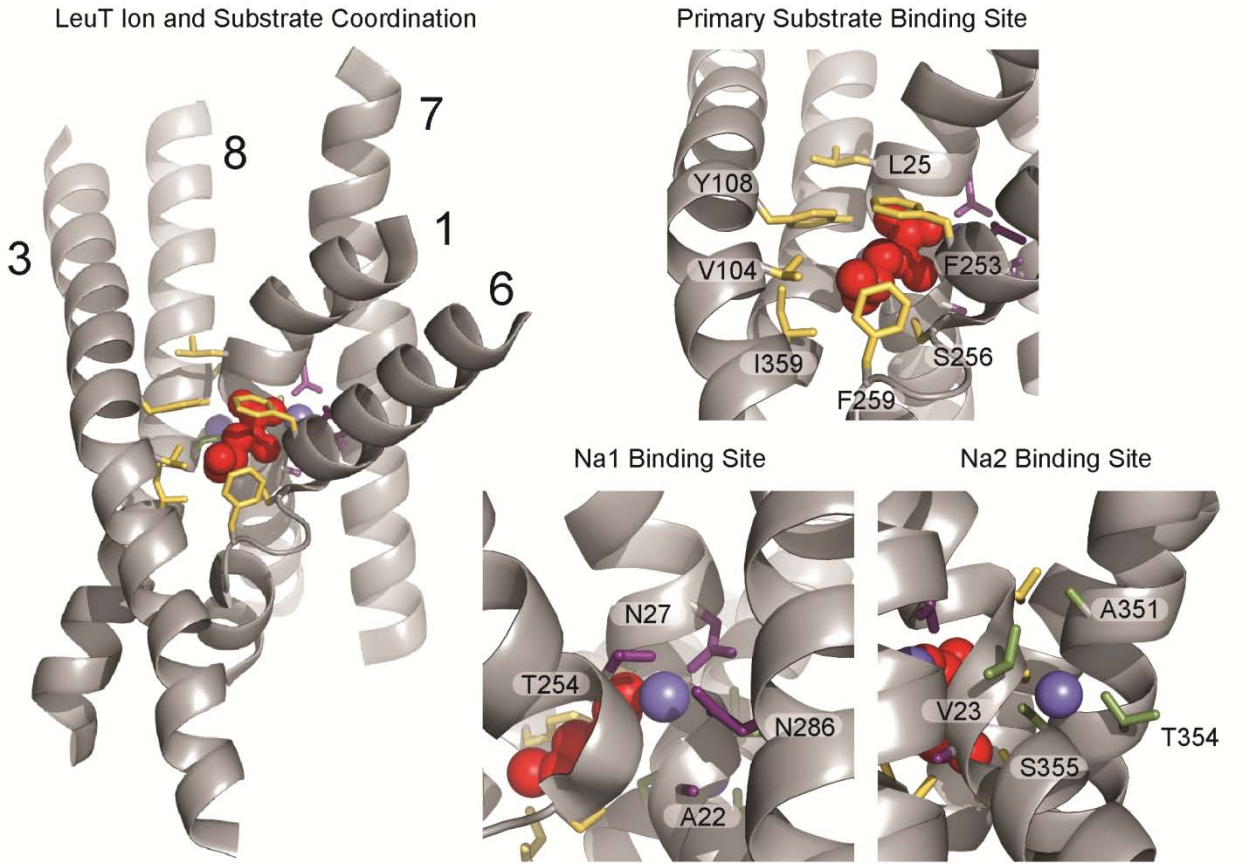


Figure 1.3. LeuT Na^+ and substrate binding sites highlighting participating residues

Two Na^+ ion binding sites were also evident in the LeuT structure, termed Na1 and Na2[12] (Fig. 1.3). Na1 is located at the interface of the discontinuous regions of TMs 1 and 6, coordinated by main chain carboxy and carbonyl oxygens of residues A22 and T254, side chain carbonyls of N27 and N286, and the hydroxyl group provided by T254. Furthermore, Na1 interacts directly with the carboxy group of the Leu molecule. Direct coordination was reasoned to support Na^+-dependent Leu uptake[12]. The Na2 site is situated between TMs 1 and 8, distal to the Na1 and Leu sites by 7.0 Å and 5.9 Å, respectively[12]. Na2 is coordinated by carbonyl oxygens of residues G20, V23, and A351 and hydroxyl oxygens of residues T354 and S355. Na selectivity at these sites was postulated to be the result of steric selection. The region was well defined in the crystal structures and the authors speculated that this was the result of Na

binding stabilizing the otherwise flexible region, providing necessary interactions for formation of the substrate binding site[12]. Both Na1 and substrate sites include highly conserved residues in NSS and are therefore considered to be conserved throughout the family. However, Na2 is less well conserved. This may be related to the varying ion:substrate stoichiometries found in NSS members, with 1:1, 2:1, and 3:1 stoichiometries all existing within the family[46]. Although, Cl^- dependence is an important feature of eukaryotic NSS, LeuT was shown to be Cl^--independent and no Cl^- binding sites were found in the structure[12].

The original LeuT crystal structure represented an outward-facing occluded conformation, with a closed extracellular thin gate and an open extracellular thick gate. The thin extracellular gate was composed of individual aromatic residues Y108 and F253, in positions that shielded the substrate from the extracellular solution. The extracellular thick gate refers to the position of extracellular helices that create the extracellular permeation pathway and its open orientation was defined by the presence of a solvent accessible extracellular vestibule[12] (Fig. 1.4). In contrast, the intracellular side was completely closed with no evident intracellular permeation pathway. Intracellular hydrogen bonding interactions between residue R5, S267, Y268, Q361, and D369 was thought to stabilize the intracellular thick gate in a closed position[12] (Fig. 1.4). This interaction was first identified in eukaryotic homologues and mutation of these residues has been shown to impede transporter function[75]. Extracellular residues R30 and D404, G408, T409 were suggested to provide a similar function for stabilization of the extracellular closed thick gate[12] (Fig. 1.4). Mutations of individual stabilization residues, both extracellular and intracellular have been subsequently shown to shift conformational equilibria using Single Molecule-Fluorescence Resonance Energy Transfer (SM-FRET) spectroscopy[76].

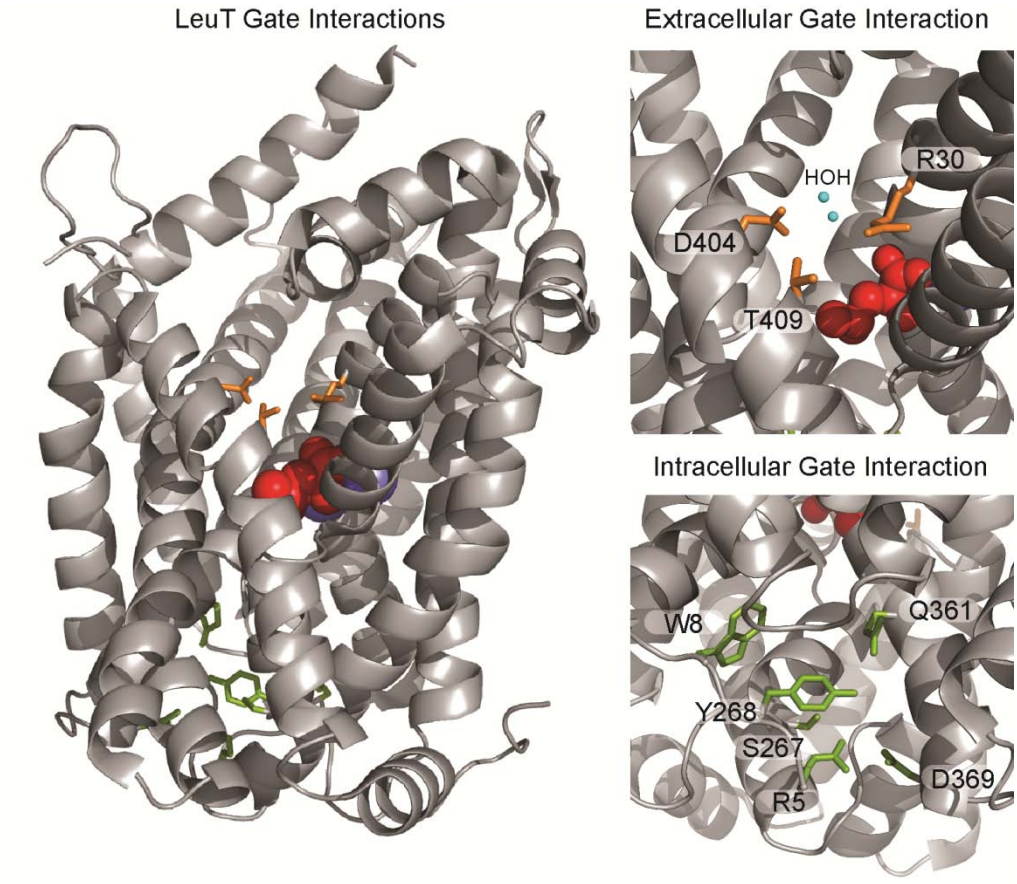


Figure 1.4. LeuT extracellular and intracellular gating interactions highlighting participating residues and water molecules.

Additional functional motifs were suggested based on the LeuT structure[12]. The N terminus displayed an extended loop that formed extensive interactions with the intracellular face of the transporter at sites in TMs 1a, 4, 5, 6b and 8. Conserved residues Y5 and W8 formed specific interactions in this region, with residue Y5 participating in intracellular thick gate stabilization, as described previously. Intracellular loop (IL) 1, a small, helical loop connecting TMs 2 and 3, was identified as reentrant and was a potential source of intracellular occlusion. Extracellular loops EL2 and EL4 were identified as potentially involved in extracellular occlusion with EL4 specifically making interactions with TM1b[12].

Core inverted repeats define LeuT Fold structures

With the publication of additional LeuT Fold structures, a comparative analysis of LeuT Fold structural features became possible and common features could be identified. The five-helix inverted repeat motifs, first reported in the LeuT structure[12], have come to define the LeuT Fold and are therefore necessarily a conserved feature of all LeuT Fold proteins. Internal structural symmetry was not novel at the time of the LeuT structure publication, having been shown in structures of aquaporin[77], a water channel, as well as multi-drug efflux pump AcrB[8], Major Facilitator Superfamily (MFS) transporters GlpT[10] and LacY[11], and Na^+-coupled secondary active transporters GltPh[78] and NhaA[79]. Internal repeat topologies in transporters tend to be the rule rather than the exception, with publication of new structures serving to support established internal repeat folds. The functional relevance of these repeat structures is currently a topic of intense debate, with the eventual conclusion set to have far reaching consequences for structure-function relationships in transporters.

Structural similarity between the inverted repeats within individual LeuT Fold proteins is consistently quite high with RMSD values falling between 2 and 6Å, despite lacking significant sequence similarity in each instance. All inverted repeats in the LeuT Fold are related by a two-fold symmetry around an axis parallel to the membrane. Subsequent analyses of the fold identified 2 four-helix bundles referred to as the bundle (TMs 1, 2, 6, and 7 in LeuT) and scaffold (TMs 3, 4, 8, and 9 in LeuT) motifs, that each include 2 helices from each inverted repeat[70]. Present in each LeuT Fold protein, these helical bundle motifs segregate to opposite sides of the transporter and ion and substrate binding sites are located at their interface. Relative motions between these motifs, predicted from their alternate positions in the various conformations represented by LeuT Fold structures, have been hypothesized to form the structural basis of a unified alternating access mechanism for the LeuT Fold[6] (Fig. 1.5).

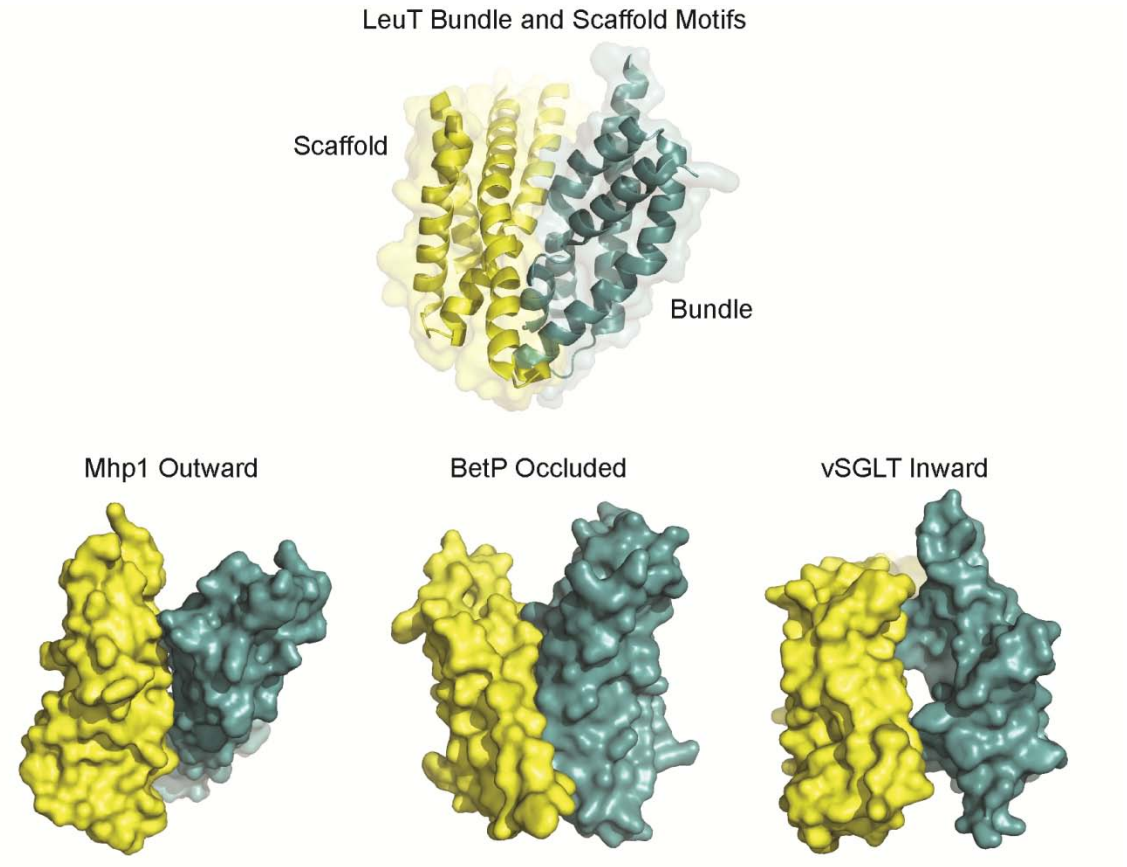


Figure 1.5. Relationship between bundle and scaffold motifs in LeuT Fold proteins

Non-core helices and oligomerization

Structurally characterized members typically present 12 TM total helices, however additional non-core helices can be accommodated as shown in vSGLT, which has 14 TM helices[15]. Inverted repeats, consisting of the 10 core TMs, have thus far been shown to be exclusively sequentially contiguous, with non-core helices appearing N- or C-terminal to the core inverted repeat structure. In the case of vSGLT, two additional non-core helices result in both N-terminal (1 helix) and C-terminal (3 helices) positions[15]. Non-core helices also vary in tertiary position relative to the core among the LeuT Fold proteins. While most LeuT Fold proteins have been shown to be functional as monomers, they often oligomerize as dimers and

trimers. In some cases, like LeuT, dimerization seems to be the result of crystallization conditions[12,15]. In others, such as AdiC, dimerization may be the natural preference of the protein[19]. In the case of BetP, trimerization results in regulatory interactions between protomers[61]. However, in CaiT, the function of trimerization is unknown. In each case, oligomerization interfaces involve non-core helices. Non-core helices also tend to show more sequence variation within families as compared to core residues, thereby suggesting protein-specific functionality that is consistent with differential oligomerization states between proteins.

Discontinuous helices provide ion and substrate binding sites

The importance of discontinuous helices becomes immediately obvious in a comparative analysis of LeuT Fold structures. TM helices equivalent to TM1 in LeuT (henceforth TM1') show a centrally-located discontinuous region in all LeuT Fold members. These typically highly conserved, glycine-rich regions invariably provides backbone carbonyl and amine groups that serve to stabilize substrate, and where applicable, ion binding sites. The structure of TM6 is discontinuous in LeuT and most LeuT Fold members. varies. However, TM6' is bent is CaiT[18,21] and fully continuous in BetP[16]. The discontinuous nature of TMs 1' and 6' further provides dipole moments that participate in this coordination. Discontinuous helices have been shown to be functionally important for a number of transporter families including the GltPh Fold[78] with discontinuous, reentrant helices supporting ion and substrate binding sites.

The primary substrate binding site is conserved

The primary substrate binding site referred to here as S1 is currently completely conserved among LeuT Fold proteins. In each case, S1 is located in a highly conserved region at the interface of the discontinuous TMs 1' and 6' and TMs 3' and 8'. S1 is coordinated by main chain carbonyl and carboxyl oxygens and amide nitrogens and side chain hydroxyls. Another common feature is a hydrophobic pocket, often a Trp box, that provides stabilization for aliphatic

and aromatic substrate regions and solvent exclusion of bound substrate by comprising “thin gates” described above. S1 is also universally stabilized by proximally-bound ions or, in the case of ion-independent members, residue side chain interactions. Substrate specificity has been defined in LeuT by shape complementarity, with the size of the binding pocket and constituent residues correlated with substrate binding affinities and transport rates[12].

Alternative substrate binding sites are protein-specific

Additional substrate binding sites have been proposed for LeuT Fold members. Specifically, a secondary binding site has been putatively identified in LeuT (S2) based on MD simulations, and supported by binding and flux experiments[80]. Though controversial, the S2 site is projected to be located in the extracellular vestibule at the interface of TMs 1, 10 and EL4. S2 has also been identified as a binding site of the tricyclic antidepressant inhibitors[43-44], selective serotonin reuptake inhibitors[45], detergent inhibitor β-OG[81], and of the nontransportable amino acid Trp[22]. The secondary substrate binding site has been purported to function as an allosteric regulator of intracellular gating in the transport mechanism of LeuT[80].

CaiT was crystallized in both substrate-free, γ-butyrobetaine-bound, and L-carnitine-bound conditions[18,21]. In the γ-butyrobetaine-bound structure[21], two substrate binding sites are present, the classic LeuT Fold primary S1 binding site and a secondary site located in an extracellular cavity at the end of the closed extracellular vestibule. This location is distinct from the secondary site predicted for LeuT. Authors speculate that sequence similarity in this region between CaiT and BetP may suggest a conservation of the secondary substrate binding in BetP[21], however this has not been confirmed by any of the BetP crystal structures. The L-carnitine-bound[18] CaiT structure shows four bound substrate molecules. Both the primary S1 site as well as the additional extracellular site of the γ-butyrobetaine-bound structure contain carnitine molecules. In addition, two intracellular sites are also identified. One approximately 6Å

below S1 involves putative gating residues Y327 and Q377 and has been speculated to fulfill a reciprocal, but similar role to S2 in LeuT. The final site, located at the extreme end of the intracellular vestibule is considered transient and non-specific. Overall, these results suggest possible mechanistic roles for alternative binding sites. However, conclusive determination of these roles will require additional experimental testing. It is clear that the substrate permeation pathways provide stable interactions for substrate as it traverses extracellular and intracellular solvent accessible vestibules.

Chemical properties of permeation pathways are substrate-specific

As substrate binding sites feature large degrees of commonality with subtle substrate specific features, so too do substrate permeation pathways. In LeuT, the ability to sequester substrate and guide substrate to and from the binding site, is evident in specific interactions predicted in the extracellular vestibule. In vSGLT, the intracellular permeation pathway is hydrophilic to offer complementary interactions with sugar substrates[15,50]. BetP and CaiT on the other hand, have highly hydrophobic cavities, lined with Trp residues, to minimize repulsion of osmolytes[56]. The negative potential of the AdiC cavity is thought to aid binding of the positively charged substrates[19-20]. In each case, the substrate permeation pathways provide specific chemical environments to recruit and maintain substrate.

Ion stoichiometry and dependence are defined structurally

Among LeuT Fold members, ion dependence and stoichiometry varies. The LeuT structure identified two Na^+ binding sites, Na1 located proximal to the substrate binding site coordinated by discontinuous regions of TMs 1 and 6 and Na2 at the interface of TMs 1 and 8 (Fig. 1.3). Na1 directly coordinates and is coordinated by substrate, while Na2 shows indirect interactions with substrate mediated by intervening residues[12]. For the Na^+-dependent members, Na2 appears to be conserved, showing evidence of bound Na in this location in all

structures. In ApcT, a H^+-dependent member, the Na2 site of LeuT is superimposible with the amine group of K158 residue that when protonated provides a similar positive charge and therefore may perform similar function to a bound Na[17]. In ion-independent CaiT, the Na2 site is characterized by a positively charged Arg residue providing similar interactions with TMs 1' and 8[21].

The Na1 site is somewhat rare within the LeuT Fold, with only LeuT and BetP showing evidence of Na binding in this region[12,16]. In [21] and ApcT[17], the sulfur atom of a Met residue interacts with the carboxyl groups of the substrates in a similar fashion to Na^+ in the Na1 site. Other LeuT Fold members show no obvious interactions in the region of Na1. AdiC[19-20] and GadC[13] show no similarities to either Na binding site. Evidence suggests that Na2 is sufficient for stabilization of substrate binding with Mhp1 showing a 10-fold increase in substrate affinity in the presence of Na^+[14]. MD simulations in LeuT have suggested a conformational role for Na2[82], but this has not been experimentally confirmed. These differences in ion interaction may have interesting applications in functional mechanisms.

Ionic gating interactions are specific to LeuT

The LeuT structure identified interactions between conserved residues R30 and D404, G408, and T409 on the extracellular side and R5 and D369, S267, and Y268 on the intracellular side[12] (Fig. 1.4). These interactions are speculated to form the basis for stabilization of the extracellular and intracellular closed position of the thick gates. Mutation of these residues has been shown to affect conformational sampling in LeuT. Specifically, single molecule Fluorescence Resonance Energy Transfer (smFRET) investigations revealed that mutation of the intracellular gate, including R5A and Y268A mutations, resulted in a shift in conformational equilibria to favor a more open-in/closed-out conformation than sampled in the WT control, as monitored by the N terminus on the intracellular side and EL3 on the extracellular side[76]. The interpretation of these results was that these mutations favored the open-in conformation, but

additional experimentation is necessary to confirm this interpretation. It is clear that the intracellular interaction is likely important for function in the NSS family as this interaction was shown to be conserved in homology modeling of DAT[75]. However, these interactions are not well conserved among the LeuT Fold as no other LeuT Fold member shows evidence of equivalent interactions.

Structured loops may participate in gating

The extracellular loop EL4' (TM7'-8') and intracellular loop IL1' (TM 2'-3') structures are conserved within the LeuT Fold. Both are evident in Mhp1, vSGLT, ApcT, BetP, CaiT, GadC, and possibly AdiC. This level of conservation lends support to hypotheses that these structures are mechanistically relevant. Dynamics investigations in LeuT identified ligand-dependent conformational changes in EL4 and linked these changes to solvent accessibility in the extracellular vestibule[23]. IL1 has been speculated to have a role in intracellular occlusion[12], however, there is no structural or dynamic data that suggests independent movement of IL1.

C-terminal domains provide unique mechanisms of regulation

Not evident in LeuT, but present in some LeuT Fold members are unique C-terminal domains. In BetP, a long, positively-charged C-terminal helix regulates transporter activation by mediating interactions between protomers in response to high osmolarity environments[16]. The proposed mechanism of activation suggests that in response to high intracellular concentrations of K^+, triggered by high osmolarity, a C terminal domain interacts with the negatively charged N' terminal region which is tethered to TM1 and IL3 of the neighboring bundle motif to activate transport[61]. This view has been supported by Ala and Pro scanning and N and C-terminal truncation mutants showing differential activation profiles with mutations of conserved residues abolishing transport[56]. However, this interaction is not conserved among BCCT family members.

GadC presents an even more novel domain, the C terminal plug[13]. Composed of residues 477-511, this folded domain is located in the intracellular permeation pathway, completely blocking the vestibule. At high pH this region is in its closed conformation[13]. Transport cannot occur under these conditions, as shown in a cross-linking experiment[13]. Deletion of the domain, results in successful transport, but with a shifted pH dependence profile that favors transport at higher pH[13]. How these unique structural features regulate transporter mechanisms will continue to be a subject of intense interest in the field.

Models of alternating access in the LeuT Fold

Inverted Repeat Model suggested a Rocking Bundle mechanism

The symmetric nature of the LeuT Fold stimulated conjectures of a symmetric functional mechanism. This hypothesis was formalized computationally in a model of inward-facing conformation of LeuT by Forrest et al[70]. (Fig. 1.6). In this model, the sequences of each inverted repeat were threaded onto the conformations of the alternate repeat. The intellectual foundation of this model relied on the outward-facing nature of the original LeuT structure. Swapping these repeat conformations resulted in inward-facing conformation. A comparison of the outward-facing structure and the inward-facing model identified a 4-helix bundle (TMs 1, 2, 6, and 7) called the bundle as the primary source of conformational rearrangement, undergoing a rotation around an axis near the substrate and ion binding sites. This rotation opened a putative intracellular permeation pathway lined by helices 1, 5, 6, and 8, a finding which was in agreement with solvent accessibility investigations on NSS homologue SERT[70]. This model was named the Rocking Bundle Model of LeuT transport and was distinguished in comparison to subsequent models based on its rigid-body concomitant motions of TMs 1, 2, 6, and 7[6]. It served to explain the molecular symmetry of the LeuT Fold as fundamental to its mechanism of

action, generated testable hypotheses for LeuT function, and offered a potential mechanism for understanding functional mechanisms of symmetric transporters of other structural folds.

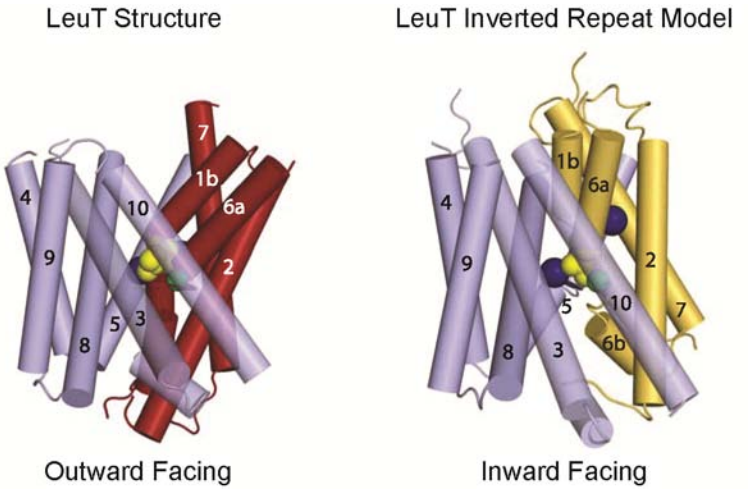


Figure 1.6. Inverted Repeat Model of the LeuT inward-facing conformation. The position of the bundle motif is colored for comparison. Figure modified with permission from [70].

Ligand-dependent conformational transitions in LeuT

Site-directed spin labeling and Electron Paramagnetic Resonance (EPR) spectroscopy were used to monitor the ligand-dependence of conformational transitions of the extracellular side of LeuT[23] (Fig. 1.7). These investigations revealed that the presence of Na^+ increased solvent accessibility of individual spin labeled residues located in the extracellular vestibule. The Na^+-dependent change in accessibility was coupled to increases in distance between EL4 and the protein core as monitored by distance distributions of spin label pairs. Together, these results suggest that Na^+ is responsible for the conformational shift to the outward-facing conformation in LeuT. Oppositely, the binding of Leu, in the presence of Na^+, served to reduce solvent accessibility in the vestibule and decreased distances relating the position of EL4, indicative of a closed extracellular conformation. This result was consistent with the original Na^+- and Leu- bound crystal structure which showed an occluded position of EL4. However, these

results suggested a dehydrated S1 site and extracellular vestibule that in some ways contradicted the solvent filled cavity of the occluded crystal structure. Furthermore, this work revealed equilibria between conformations under all ligand-conditions suggesting that transport was the result of ligand-dependent shifts in these equilibria. Similarly, inhibitors of LeuT like Trp, were shown to shift equilibria toward the outward-facing conformation, providing evidence that conformational selection was the source of non-competitive inhibition in LeuT.

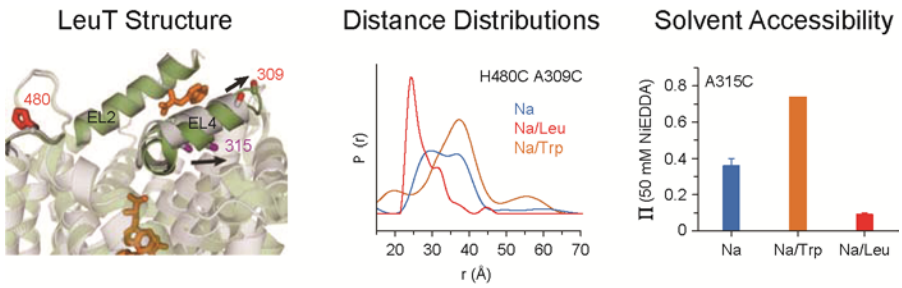


Figure 1.7. EPR investigations of LeuT extracellular conformational dynamics. (Left) LeuT structures comparing Leu-bound and Trp-bound structures. (Middle) DEER distance distributions relating the position of EL4 in Na^+-bound, Leu-bound, and Trp-bound intermediates. (Right) Corresponding measures of solvent accessibility to extracellular permeation pathway. Figure modified with permission from [23].

Single molecule FRET (smFRET) investigations complemented the EPR investigations described above by monitoring the conformational dynamics of the intracellular side of LeuT[76] (Fig. 1.8). Doubly fluorescently labeled LeuT mutants relating the position of the N terminus relative to IL1 revealed differential FRET states consistent with two distinct conformations separated by approximately 13Å in distance. In the apo state, these two conformations were relatively equally populated indicating dynamic sampling of both open-in and closed-in conformations. The binding of Na^+ and Na^+/Leu shifted the relative proportion of these conformations to favor the high FRET state, closed-in conformation. These N terminal conformational changes were projected to extend to TM1a as MD simulations identified similar changes in the position of TM1[76].

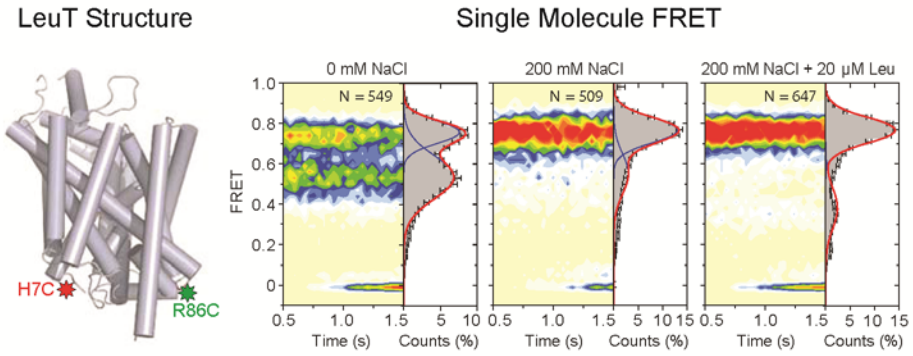


Figure 1.8. Single molecular FRET monitoring LeuT intracellular conformational dynamics. (Left) Positions of the fluorescent probe labeling sites on the LeuT structure. (Right) SM-FRET results for apo, Na^+-bound, and Leu-bound intermediates. Figure modified with permission from [76].

Integration of the descriptions of intracellular and extracellular conformational dynamics provides a generalized mechanism of transport in LeuT. In the apo state, the protein appears to adopt predominantly inward-facing and occluded conformations, while rarely sampling an outward-facing conformation. In the presence of Na^+, the proportion of outward-facing molecules is increased, though the occluded conformation is also heavily populated and it is the inward-facing conformation that is only rarely sampled. With the binding of Leu, the occluded conformation is heavily favored, with only very rare excursions to either the inward- or outward-facing conformations. We can reason then that the rate-limited step in the transport cycle would reside in the transition from the Na^+- and substrate- bound occluded conformation to the inward-facing conformation.

A holistic view of the LeuT transport cycle reveals a number of principles underlying the dynamics of alternating access. It can be concluded that ligand binding results in shifts in conformational equilibria between states rather that discrete changes in conformation. Conformational changes in LeuT do not appear to be the result of rigid body reorientation of functional motifs, as ligand binding results in differential conformational shifts on intracellular and extracellular sides. Furthermore, some conformational transitions are not triggered by

ligand binding, but rather the result of low probability dynamic sampling of alternate conformations.

Mhp1 crystal structures: Experimental support for the Rocking Bundle Model

With the publication of crystal structures of Mhp1 in multiple conformations[14,83], Mhp1 transporter became the first member of the LeuT Fold to be represented by a full complement of states. The three published structures included a Na^+-bound, substrate-free outward-facing conformation[14], a Na^+- and substrate-bound outward-facing occluded conformation[14], and a Na^+- and substrate-free inward-facing conformation[83] (Fig. 1.9). These conformations were captured without structural perturbations such as point mutations of sequence or antibody stabilization. However, the inward-facing conformation showed evidence of an unidentified electron density located in the Na^+ binding site[83]. Purported to be an unknown inhibitor, interpretation of the mechanistic identity of the inward-facing structure is therefore somewhat complicated. The outward-facing occluded conformation[14] and the inward-facing conformation[83] were also relatively low resolution at 4Å and 3.8Å, respectively. Low resolution may complicate comparisons between the structures where differences in motif position often vary within that range.

The mechanistic model proposed from these structures[83] was generally consistent with the Rocking Bundle mechanism hypothesis and similar to the Inverted Repeat Model described above[6]. The motions underlying alternating access in the Mhp1 structures were defined by the relative orientation of the bundle and scaffold motifs. Both the bundle and scaffold motifs were found to be rigid units, with 0.7Å and 0.9Å RMSD for the motifs, respectively, among the three structures[83]. The only deviation from an exclusively Rocking Bundle mechanism was the presence of symmetrically related TM helices 5 and 10 participating in gating interactions[83]. In particular, TM10 was shown in the occluded structure to block substrate release from the extracellular side. MD simulations with these structures revealed that these gating motions were

likely stochastic and not directly related to the motion of the bundle and scaffold motifs[83]. TM9 of the scaffold was also implicated in the MD simulations to move in concert with TM 10 in the gating of the extracellular side[83]. The intracellular cavity of Mhp1 was identified as being lined by TMs 1, 3, 5, 6, and 8[83], remarkably similar to the results of solvent accessibility studies in SERT[70], the LeuT homolog.

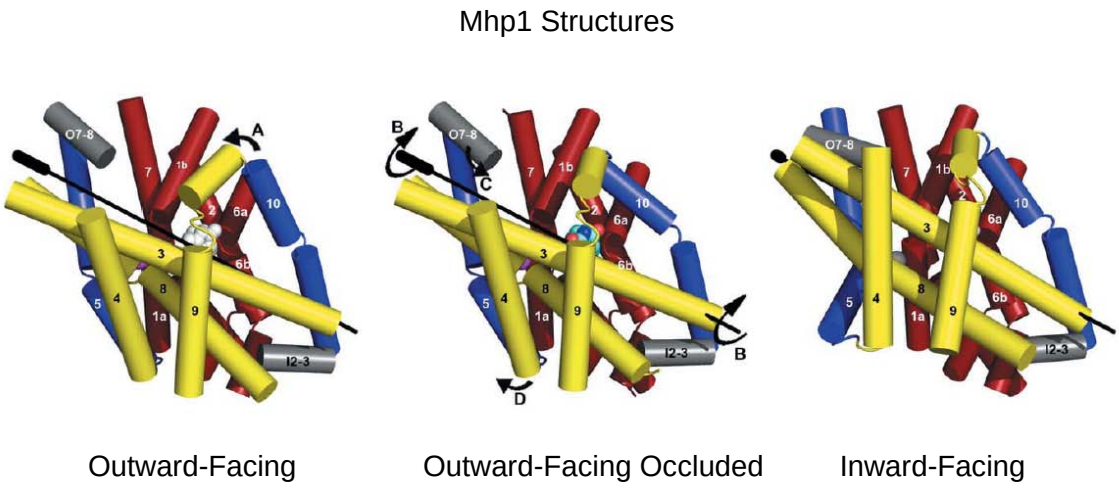


Figure 1.9. Mhp1 crystal structures define the Mhp1 transport cycle. These structures highlight the rotation of the scaffold (yellow) relative to the bundle (red). Figure modified with permission from [83].

These structures were placed into a Mhp1 transport cycle[84] with the binding of Na^+ stabilizing an outward-facing conformation that in the apo state would be stochastically sampled. The binding of substrate would then trigger a conformational rearrangement to an inward-facing conformation. This assertion was likely speculative as it is unclear how this inference could be related to the ligand-dependencies evident in the Mhp1 crystal structures, which show a Na^+- and substrate-bound outward-facing occluded conformation.

LeuT Crystal Structure Model predicts asymmetric transport mechanism

The publication of the outward-facing and inward-facing structures of LeuT proposed an atomic resolution model of the conformational cycle of LeuT transport[85]. To capture these conformations crystallographically, a number of mutations were required and antibodies were used for conformational selection and stabilization of crystal contacts[85]. The mutations were largely designed to promote specific conformational states informed by the ligand-dependent conformational equilibria uncovered in the spectroscopic investigations detailed above[23,76]. Both structures contained a K288A mutation shown to enhance substrate flux[86]. MD simulations have suggested that this mutation, located at the center of TM7, eliminates a hydrophobic mismatch that may cause membrane thinning and water penetration, limiting transport rates in WT LeuT[87]. The outward-facing structure additionally incorporated a Y108F mutation to weaken the primary substrate binding site and promote population of a Na^+-bound, substrate-free state expected to enhance an outward-facing conformation[85]. The inward-facing conformation required a number of additional mutations including T354V and S355A, disrupting the Na2 Na^+ binding site, and a Y268A mutation used to destabilize the ionic intracellular gate and promote an inward-facing conformation[85].

The Na^+-bound outward-facing conformation showed hinge bending motion of TMs 1b, 2a, and 6a around pivot points located at residues V23, G55, and L257, respectively[85] (Fig. 1.10, left). In addition, EL3 and TM11 were shifted outward in position relative to the outward-facing occluded structures (Fig. 1.10, right). The outward-facing structure[85] showed similarities to previously reported Trp-bound inhibited structure[22] with a C_α RMSD of 0.4Å. This led to speculation that the inhibitor-bound structures including the Trp-bound structure as well as TCA- and SSRI-bound structures represented native outward-facing structures[85]. The extracellular ionic gate is indeed broken in the outward-facing conformation, but both Na^+ sites are conserved and electron density presumed to represent Na^+ ions exist in both sites[85]. The

intracellular side displays a closed conformation[85] similar to the outward-facing occluded structure[12].

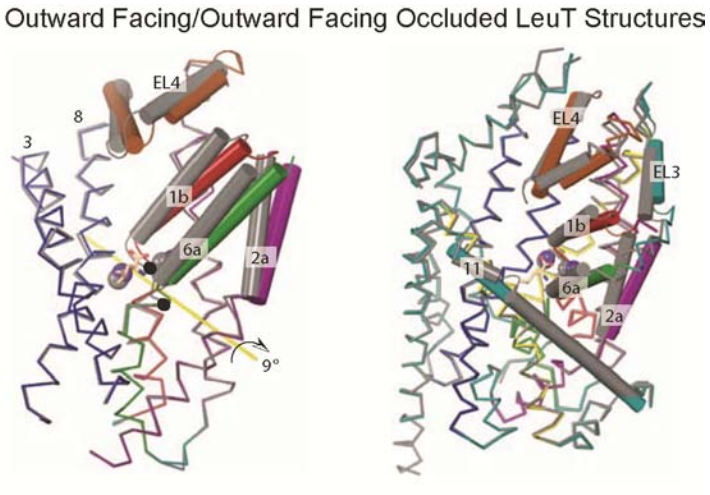


Figure 1.10. Extracellular conformational transitions evident in LeuT crystal structures. Comparison of outward-facing (colored) and outward-facing occluded (gray) structures with participating helices noted. Figure modified with permission from [85].

The inward-facing conformation is characterized by the prominent displacement of intracellular TM1a away from the protein core, tilting 45° into the membrane [85] (Fig. 1.11). The predicted motion of TM1a is coupled to a large scale outward translation of TM5 and a 17° tilt of TM6b away from the central binding site. The resulting intracellular permeation pathway is lined by TMs 1, 5, and 8[85], consistent with the SERT solvent accessibility investigations[70]. As expected, the Y268A mutation disrupted the intracellular ionic gate, with interacting residues separated out of range of hydrogen bonding[85]. In fact, the N terminus, where residues R5 and W8 play important roles in intracellular gating interactions, is completely unresolved suggestive of a substantial increase in dynamic fluctuations consistent with previous spectroscopic evaluations of the Y268A mutation. Furthermore, both of the Na sites are disrupted, Na2 through mutation, but also Na1. With the displacement of TM1 coordinating residues for Na1 were shifted out of position and no longer supported Na^+ binding. The authors suggested that

this may underscore the mechanism of Na^+ release[85]. The substrate binding site, however, is maintained with aromatic residues remaining in position to support substrate binding, forming a putative structural basis for reverse substrate transport[85].

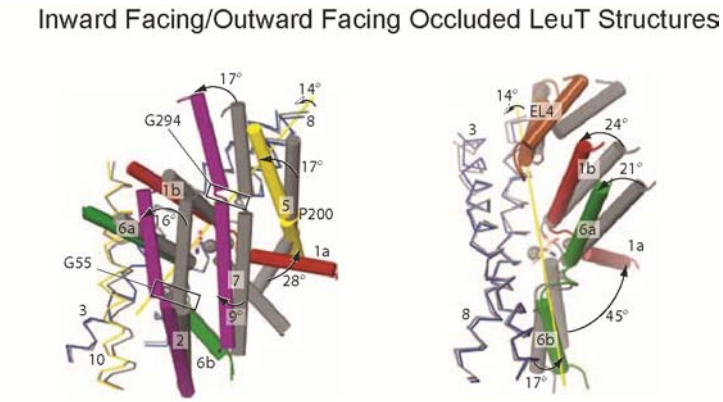


Figure 1.11. LeuT intracellular opening transitions represented in crystal structures. Comparison of inward-facing (colored) and outward-facing occluded (gray) structures with participating helices noted. Figure modified with permission from [85].

On the extracellular side, TMs 1b and 6a and EL4 shift into a more closed conformation than was evident in any of the preceding crystal structures[85] (Fig. 1.11, right). TMs 1b and 6a are each tilted on the order of 20˚ toward the scaffold domain and EL4 dips into the extracellular cavity thereby providing extracellular occlusion. The extracellular ionic gate is locked in this conformation, with the R30 residues participating in a salt bridge with D404 and hydrogen bonding interactions with G408 and T409. Additional interactions were evident between EL4 and TM10 and TM6 and A319. However, an unidentified electron density remained in the extracellular vestibule, bound at the proposed S2 site[85]. Leu was excluded as a possible identity for the molecule based on lack of fit between the electron density and amino acid structure. Rather, the density was suggested to be the result of bound detergent, buffer, or water molecules[85]. Previously, the S2 site was shown to support specific binding of a β-OG

molecule[81], the crystallization detergent as well as possibly lipid molecules[88]. The implications of these interactions have not been fully elucidated, but will be discussed below.

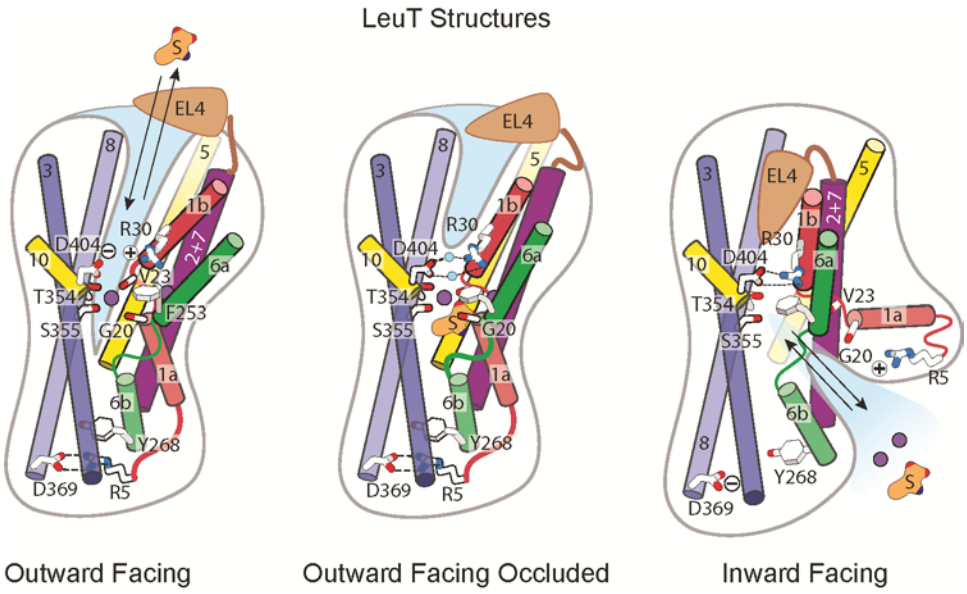


Figure 1.12. Cartoon model of LeuT transport based on the LeuT crystal structures. Figure modified with permission from [85].

What is immediately evident from the LeuT series of structures is the dissimilarity to the Rocking Bundle Model of transport predicted based on LeuT symmetry[6] and supported by the Mhp1 model of transport[83]. On the extracellular side, TMs 1 and 6 of the bundle bend at their discontinuous sections and EL4 translates outward to reveal an accessible substrate permeation pathway[85] (Fig. 1.12, left). With ion and substrate bound, LeuT adopts an occluded conformation that is characterized by subtle shifts in the position of TMs 1 and 6 that align the aromatic residues of the thin gate to shield the substrate from a solvent filled extracellular vestibule[12,85] (Fig. 1.12, middle). This vestibule is blocked from the extracellular solvent by the position of EL4. Therefore, this mechanism proposes regulation specifically through thin gates on the extracellular side. On the intracellular side, TM1a is released from the intracellular ionic gate interaction to adopt a highly flexible and tilted conformation located within the membrane.

With this translation of TM1a, TM5 shifts outward which in theory pressures the extracellular side to close. Therefore, the mechanism of coupling between the intracellular and extracellular cavities is proposed to be dependent upon the relationship between TMs 1 and 5[85].

The transport mechanism described above is asymmetric and specifically relates a non-rigid body break in the bundle motif[85]. The bundle motif aligns only to 3.5Å C_α RMSD between the occluded[12] and inward-facing structures[85], with the deviation predominantly the result of shift in position of TM1a relative to TMs 1b, 2, 6, and 7. TM1a is highly flexible in the crystal structure and it is unclear whether its position is indeed native[85]. The authors argued that the position of TM1a was not a crystallographic artifact as it was not located along crystal or fragment antigen binding (FAB) fragment contacts[85]. Furthermore, the high flexibility and magnitude of conformational change evident in TM1a was suggested in smFRET investigations of the N terminus and MD simulations of TM1a[76]. Symmetry is additionally broken in this mechanism as EL4 is highly involved in extracellular occlusion, but symmetric intracellular equivalent IL1, is static in position[85].

The LeuT S2 site and allosteric regulation of transport

Although alternative substrate binding sites were not evident in any of the LeuT crystal structures, the existence of a secondary substrate binding site was predicted based on steered MD (SMD) simulations[80]. Pulling a substrate molecule from the substrate binding site toward the extracellular vestibule resulted in a favorable interaction at a position approximately 10Å above the primary site. Occupancy of both substrate binding sites was demonstrated with both substrate molecules remaining bound over a 10 ns of simulation time. Like the primary site, the secondary site offered a hydrophobic pocket that interacted with substrate side chain group and ionic interactions with residues R30 and D404 that provide stabilization for Leu's carboxy and amine functional groups, respectively[80]. These residues were previously identified as participating in extracellular gating and inhibitors like Trp[22], TCAs[43-44], and SSRIs[45] were found

to bind in the S2 site (see below), highlighting the importance of this region. There is currently no evidence that the S2 site is conserved among other LeuT Fold members.

The proposal of a secondary substrate binding site was supported experimentally with radiotracer binding and transport assays[80]. The Leu:LeuT stoichiometry was found to be 1.8 in binding assays. This value was reduced to 1.0 and 0.9 when mutations in the secondary site, L400C and I111C respectively, were included in the WT background. These investigations further suggested a mechanistic role for the S2 site as an allosteric regulator of the intracellular gate. When LeuT was reconstituted into outward-out orientations in proteoliposomes (PL), Leu bound to S1 was transported into the PL only when a Leu molecule was bound at the S2 site[80]. In the S2 mutant background, no transport of substrate was observed[80]. Furthermore, smFRET investigations suggested that Ala-induced intracellular opening was decreased by mutation of the S2 site[89].

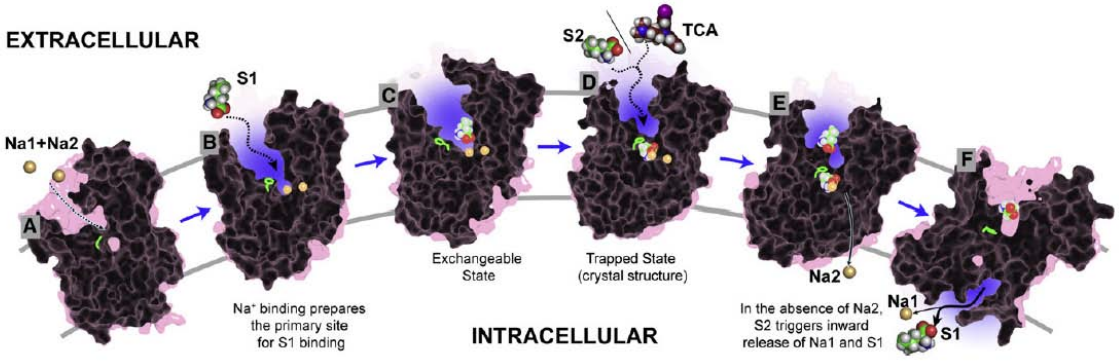


Figure 1.13. Allosteric regulation model of LeuT transport resulting from SMD simulations. Figure reproduced with permission from [80].

The model of transport resulting from these results integrated the functional data and SMD results[80] (Fig. 1.13). The SMD investigations predicted an apo conformation similar to the occluded crystal structure. This analysis misses the inward-facing component of the apo conformation, but is not inconsistent with the spectroscopic results described above[23,76].

Consistent with these investigations, SMD also identified a Na^+-dependent shift to favor an outward-facing conformation[80]. Simultaneous binding of Na^+ and substrate in S1 was proposed to adopt an occluded conformation, with binding in S2 required for allosteric opening of the intracellular gate and translocation of S1 substrate[80].

The secondary binding site hypothesis was challenged by the work of Gouaux and colleagues who conducted a series of experiments probing the Leu:LeuT stoichiometry[86]. Using ITC, equilibrium dialysis, and SPA measurements, a 0.7-0.8 stoichiometry was consistently reported across methodologies and a single high affinity substrate binding site was proposed. The substoichiometric ratio was suggested to be the possible result of LeuT molecules unable to bind substrate possibly because of aggregation. It was also shown that not only was transport possible with one site, but that the transport kinetics of Ala uptake were most consistent with a single site model. A subsequent analysis refuted this report by highlighting a pitfall of the Gouaux laboratory protein preparation method[90]. Centrifugal concentration beyond 10-fold in volume reduction caused a time-dependent, irreversible decrease in stoichiometry from 2 to 1. Under these conditions, detergent molecules including β-OG and β-DDM were able to block the secondary binding site. While this explained the difference in apparent stoichiometries between laboratory preparations and cautioned the field to the potential problems associated with detergent effects, it did not resolve the question of whether substrate binding at S2 was required for substrate transport.

Inhibitors bind at S2 in LeuT

In addition to its role as a putative substrate binding site, S2 of LeuT is the location of known inhibitor interactions with SSRIs[45], TCAs[44-45], and non-transportable amino acids[22]. Structural characterization of interactions of inhibitor molecules with the S2 LeuT site has shown inhibitor-specific interaction residues (Fig. 1.14). In most cases, however, inhibitors interact directly or through water molecules with the R30 and D404 extracellular gate residues. This is

directly related to inhibitor preference for binding and stabilizing the outward-facing LeuT conformation. Interaction of the TCAs with the secondary substrate binding site were confirmed in radioligand binding assays, where substrate binding at the secondary site was reduced in the presence of TCAs[80]. smFRET[76] and EPR[23] also identified a shift in conformation to favor the outward-facing conformation in the presence of inhibitors including the TCAs and Trp, respectively.

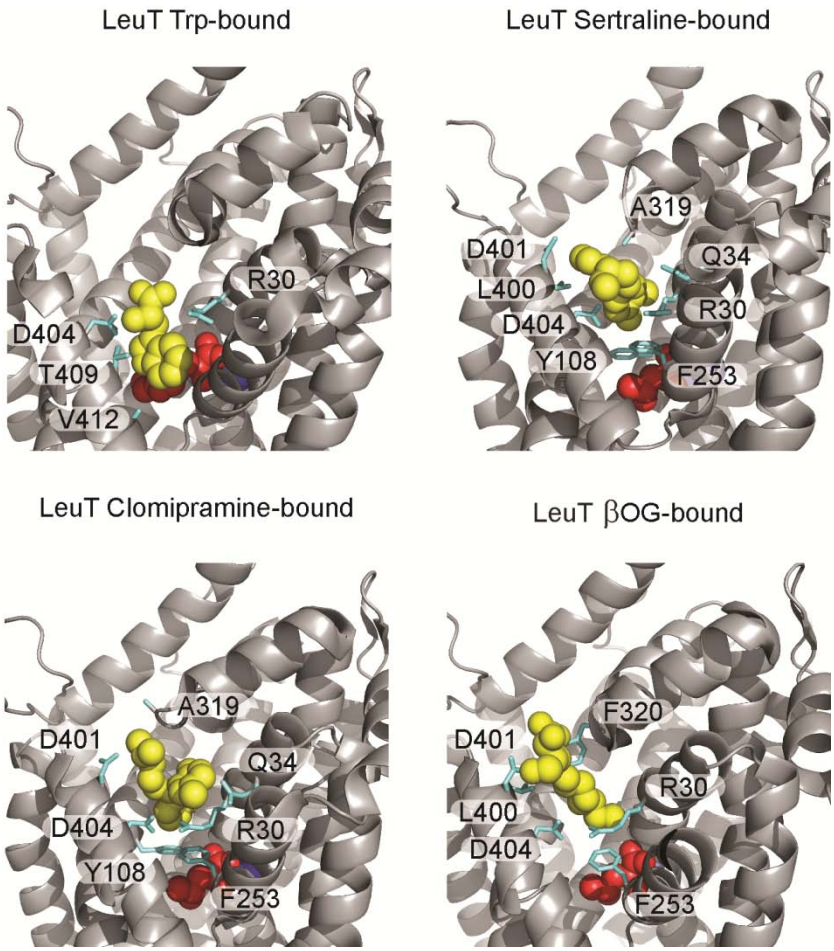


Figure 1.14. Inhibitor binding interactions at the S2 binding site. Inhibitors are represented in yellow, substrates in red and interacting residues in cyan.

β-OG, a short-chain crystallization detergent used in the LeuT structures, acts as a high-affinity inhibitor, binding in the extracellular vestibule at S2 of LeuT, stabilizing the protein in an outward-facing occluded conformation, and rendering LeuT inactive[81]. A β-OG molecule was shown to be clearly resolved in the crystal structure[81]. Furthermore, substrate stoichiometry is reduced to ~1 in the presence of β-OG and Ala transport into proteoliposomes (PLs) is obstructed, similar other traditional inhibitors. It is likely that all detergent solubilized structures to date represent inhibited, functionally blocked forms of the transporter[81].

In response, LeuT was purified in n-dodecyl-β-D-maltoside (C_{12}M) and selenomethionine derivatives of C_{12}M and β-OG and crystallized in DMPC/CHAPSO bicelles[88]. All resulting structures were very similar in conformation (< 1.0 Å RMSD to -OG structure (2A65)). A tightly bound detergent molecule was found in the extracellular vestibule of the *n*-heptyl seleno-β-D-glucoside-bound (β-SeHG) structure, but not in the C_{12}SeM-bound structure. This result was interpreted as a validation of the mechanistic relevance of the outward-facing occluded conformation as well as a dismissal of the potential contribution of β-OG conformational interference[88]. While resolvable detergent molecules were not identified by selenomethionine anomalous diffraction, an unidentified prominent peak was identified in the C_{12}M structure at the exact location as the high affinity site found for β-OG and β-SeHG[88]. An opposite, but compelling interpretation of these results suggests that the unidentified molecule is functioning as an inhibitor, interacting with S2 to promote the outward-facing conformation.

Hybrid Models of alternating access: BetP and vSGLT

Additional crystallographic representations of the BetP[60] and vSGLT[50] revealed a potential LeuT Fold mechanism incorporating small scale relative rotation between bundle and scaffold motifs with independent translation of TM1a'. In BetP[60], a series of intermediate orientation structures were produced using modifications of sequence including an N terminal truncation, surface-residue engineering, and for some structures, a G153D mutation that

increased affinity for Na^+ and specificity for choline. The structures consistently showed an asymmetric trimer organization with each protomer manifesting a distinct orientation of the bundle motif[60]. The orientations were defined based on solvent access and included outward-facing apo, outward-facing occluded apo, occluded apo, occluded substrate-bound, and two inward-facing substrate-bound conformations with either betaine or choline in the substrate binding site. A comparison of these structures revealed features identified individually in the LeuT and Mhp1 models. A relatively modest rotation of the bundle relative to the scaffold of 13° was identified as was a small scale translation of the TM1a' of 5Å with a tilt of 18° relative to the protein core. These small scale transitions were suggested to be the result of restraint provided by interprotomeric interactions. On the intracellular side, the C-terminal domain of one protomer interacts with the N' terminus, TM1a', and IL3 on an adjacent protomer[60]. In addition, EL3 participates in an extracellular network comprising residues from the adjacent TM1b' and EL4[61]. Both of these interactions were thought to limit bundle mobility and result in decreased transport rates in the absence of activation. Subtle rearrangements of TM5' and 10' were predicted to support uncoupled gating of the intracellular and extracellular sides, respectively[60], as in Mhp1[83].

Although only inward-facing[50] and inward-facing occluded[15] conformations have been captured for vSGLT, the conformational transition implied by these structures is quite similar to that described for BetP[60]. Transition from inward-facing occluded to fully inward-facing involved very small scale rotation of the bundle relative to the scaffold of 6° and an independent tilt of TM1a of 13°[50]. The change in position was subtle enough to register a 1.2Å RMSD between the structures. Additional conformational changes are expected to support the vSGLT transport cycle. However, both BetP and vSGLT demonstrate the capacity for small scale transitions to allow solvent access and to break stabilizing interactions of the Na and substrate binding sites as a mechanism for inward-oriented Na and substrate release and therefore directional transport of substrate.

A model of alternating access in exchangers: AdiC/GadC

The crystal structures of AdiC[19-20,69] and GadC[13] provide an opportunity to examine a potential transport mechanism for the exchangers of the LeuT Fold and compare this to the proposed mechanism for symporters as a basis for understanding differential functionalities. Although it is typically unwise to paint different proteins into the same transport cycle, AdiC and GadC provide conditions that make it reasonable to do so. Namely, they share significant sequence and functional similarity. Also at this time, there is no example of an exchanger with a full complement of protein structures characterizing, inward-facing and outward-facing conformations.

AdiC has been captured in two conformations, outward-facing apo and substrate-bound outward-facing occluded, among four structures from two organisms[19-20,69]. Alternatively, GadC[13] was captured in an inward-facing conformation, though this was the result of the C-terminal plug domain, unique to GadC, propping the intracellular vestibule in an open position. A comparison of the two proteins shows an entirely rigid body rotation of the bundle approximately 35˚ relative to the scaffold[13]. This is by far the largest relative change evident for a LeuT Fold transport, which may indicate that the structural comparison between two different proteins may overestimate the difference. It is also possible that the C-terminal plug may be holding GadC in an extreme position that it may relax from upon plug dissociation. In either case, it is interesting to note that no independent motions of TM1 are evident in this putative mechanism.

Unified Mechanism of LeuT Fold Transport

Despite seemingly diverse individual mechanisms, extensive efforts have been put forth to sew constituent structures into a unified mechanism of LeuT transport. The rationale behind these approaches has been, simply, that structure implies function. Early on these collaborative approaches evolved from necessity, as only single structures were available for each protein. Initial reviews inserted conformational examples from different proteins into transport cycles and

were intended to inform where structures were not yet available[46]. In these, the outward facing structure of LeuT, the occluded structure of Mhp1, and the inward-facing structure of vSGLT were taken together to address the conformation changes underlying alternating access.

However, even recently, with multiple structures in alternate conformations available for many of these proteins, efforts are still undertaken to combine all available crystal structures into a unified mechanism with each structure included as a necessary intermediate in the cycle[6,91-92]. As the number of structures increased, models have required increasing levels of complexity to accommodate what seem to be potentially diverse mechanisms[91]. Even when not explicitly placed into a transport cycle, these structures have been used to define average conformational transitions among members[6,92]. Although the LeuT Fold may display some common characteristics among its members, it is also very likely that individual members operate using unique functional properties endowed by distinct sequence, ion utilization, and allosteric sites.

Rigid body rotation of bundle and scaffold motifs are conserved

Among the mechanisms proposed for LeuT Fold members the relationship between the bundle and scaffold motifs features prominently. In each instance, there appears to be a rotation around a central axis of the bundle relative to the scaffold. In some members, like Mhp1[83] and AdiC/GadC[83], bundle rotation may serve as the predominant mechanism of alternating access. In others[50,60], it is likely a small part of the conformational reorientation that accompanies solvent accessibility to the binding sites. The fact that the degree of relative rotation seems to correlate with rigidity of the bundle and scaffold motifs speaks to this dichotomy directly. It is interesting to note that this conserved mechanistic feature was predicted based on the molecular symmetry of the LeuT Fold. Therefore, it is reasonable to conclude that all LeuT Fold members and possibly all symmetric transporters may indeed adopt a Rocking Bundle mechanism to some degree.

Independent motions of TM1a are a common feature of transport

Evident from the Crystal Structure Models of alternating access is the importance of TM1a in creating intracellular permeation pathways. Among members characterized in both outward-facing and inward-facing conformations, TM1a serves as the major source of deviation of rigid body structure in only LeuT[85] and BetP[60]. It is interesting that these members both have the addition Na1 site located at the discontinuous region of TM1. Mhp1[83] and AdiC/GadC[83] show no such deviation in TM1a and also do not have the Na1 site. It is possible that non-rigid body deviation may be related to the Na stoichiometry of individual LeuT Fold members. Structural characterization of additional members will be required to investigate this theory.

Two distinct occlusion mechanisms are present in LeuT Fold members

As noted previously all members have shown thin gates directly coordinating substrate that provide occlusion from solvent at the substrate binding site. However, additional mechanisms of occlusion vary among members. Unique among members, ionic interactions serve to stabilize extracellular and intracellular thick gates in LeuT[85]. In addition, cation-π interactions between protomers of BetP provide a mechanism of transport regulation, rather than gating[61]. In LeuT, EL4 provides extracellular occlusion by physically blocking the vestibule from solvent access[12,23]. Interestingly, the helical loops equivalent to EL4 in BetP[16] and Mhp1[14] displays a single helical segment rather than the broken reentrant helix evident in LeuT and are not thought to substantially contribute to extracellular occlusion. In contrast, symmetric, bent TMs 5 and 10 undergo hinge-like motions that provide additional thin gates on the intracellular and extracellular sides, respectively, in Mhp1 and BetP. TMs 5 and 10 are not bent in LeuT[12] and do not seem to undergo stochastic gating as in Mhp1 and BetP. GadC[13] obviously presents an extreme and unique mechanism of intracellular occlusion with the C-terminal plug completely inhabiting the intracellular cavity. Once again, it appears that there may be subfamilies within the LeuT functional mechanisms that are related to conserved structural features.

Inward-facing conformation often regulated by allosteric mechanisms

Mechanisms of regulation of transport vary substantially among members and may represent individual refinements on the LeuT Fold architecture that allow divergent functions on a common scaffold. For example, LeuT may employ an extracellular secondary substrate binding site to support release of the intracellular gate[80]. BetP takes advantage of its trimeric organization to encourage interprotomer interaction and regulation through its C-terminal helix[16,61]. GadC[13] utilizes its C-terminal plug to differentially regulate pH thresholds for activation of transport. It is possible that less obvious features will become apparent with further investigation of individual LeuT Fold members and may be the key to understanding individual mechanisms of alternating access.

Crystal structure conformational sampling shows divergent ligand-dependence

One curious feature that remains to be addressed is the seemingly divergent ligand-dependence of the crystal structure orientations. We know from investigations with LeuT that ligand-dependence is best understood as a feature of the ensemble[23], as shifts in conformational equilibria are often obscured by conformational selection of the crystal lattice. While crystal structure orientations should not be used to exclusively define ligand-dependence of conformational states, it can be instructive as to where free energy minima may reside for comparison between structures. A comparison of the structures revealed unexpected conformational selection by biochemical state. In LeuT, the apo state, though engineered, displayed an inward-facing conformation and the substrate-bound state favored an outward-facing occluded conformation[85]. Despite the mechanistic differences between the proteins, this pattern held for Mhp1[83]. However, BetP showed an opposite ligand-dependence with the apo structures relating outward-facing, outward-facing occluded, and occluded conformations, while substrate bound structures favored inward-facing and occluded conformations[60]. This is particularly surprising given the similar Na^+ dependence between LeuT and BetP. Furthermore,

vSGLT[50] and CaiT[18,21] suggested inward-oriented structures for both apo and substrate bound conditions. Oppositely, AdiC[19-20,69] displayed outward-facing conformations for both apo and substrate-bound conditions. And GadC[13], its functional homolog, favored an inward-facing, albeit plug-blocked, apo state. These differences likely highlight the conformational pressure imposed by crystal lattice forces and also the problem of using crystal structures to define transport cycles. Despite essentially identical transport directionalities among LeuT members, underlying transport mechanisms may be more complicated than expected based on the structural similarity of members.

Significance of EPR investigations in LeuT and Mhp1

The work presented in Chapters 4 and 5 has sought to address outstanding questions introduced above related to the conformational sampling and ligand-dependent conformational equilibria in LeuT and Mhp1 to begin to understand the transport in the LeuT Fold. Specifically, this work has been interested in defining what conformations represent stable intermediates in the transport cycle. This question is both general and specific in nature. Theoretically, alternating access descriptions of symport require sampling of outward-facing and inward-facing conformations that provide solvent accessible cavities to the central ion and substrate binding sites. However, crystallographic descriptions of protein structure often relate occluded conformations where access to binding sites are blocked from both sides of the membrane. One aspect of this research seeks to define whether occluded conformations represent discrete conformational intermediates and what role these occluded conformations play in the transport cycles of LeuT and Mhp1. In addition to general descriptions of conformational sampling, this work aims to define the dynamic motifs that transition to form the alternate conformations of the transport cycle. While the EPR distance measurements described here are necessarily low resolution, we have compiled a dense data set that is capable of defining the magnitude and

directionality of conformational transitions for each dynamic motif in Mhp1 and LeuT. Using the principles of architectural organization provided by the crystallographic representations of these proteins, we describe in detail the conformations that are sampled in LeuT and Mhp1 transport cycles. Furthermore, due to the ensemble nature of EPR measurements, all conformations that are sampled above a minimum probability threshold can be described, including novel conformations not yet captured crystallographically.

The ensemble measurements conducted in this work provide insight into the conformational equiliria between sampled intermediate states, information that can only be indirectly inferred in the static representations offered by crystallography. Here, questions relating to the energetic landscape of conformational transitions and their relationship to ligand binding can be probed. The energetic relationships between transport intermediates define how transport is accomplished. Conformational transitions can be triggered by the binding of ligand or can occur stochastically. This work defines the equilibria between intermediates as a function of their population within the ensemble to describe how LeuT and Mhp1 cycle through conformational intermediates during transport.

Na^+-coupled symport couples the energetically favorable transfer of Na^+ to the energetically unfavorable transfer of substrate. Coupling can occur in two ways, through direct stabilization of substrate binding by Na^+ and through stabilization of conformational intermediates as a result of Na^+ binding that allow formation of the substrate binding site or more efficient substrate binding. In the following Chapters, the question of how ligand transport is coupled and how conformational transitions are powered by ligand binding is addressed. For Mhp1 and LeuT, ion coupling mechanisms are related to their different Na^+ stoichiometries. Both Mhp1 and LeuT utilize a conserved Na^+ binding site referred to as the Na2 site. LeuT has an additional Na^+ binding site, Na1. Through a comparative analysis of the conformational stabilization due to Na^+ binding, this work will probe the individual roles played by Na^+ binding at

each of these sites and relate the mechanistic implications of divergent Na^+ stoichiometry in LeuT and Mhp1.

EPR investigations of structural dynamics complement atomic resolution investigations by casting a critical eye on the mechanistic identity of crystallographically captured conformations. Due to the solution nature of EPR measurements, potential biases associated with the crystallization process can be evaluated and a consensus description of conformational intermediates can be reached. In LeuT, where conserved residues have been mutated, this work will also evaluate the potential contributions of these mutations to the resulting conformational representations. Through independently evaluating LeuT and Mhp1 crystal structures, this work can confirm or augment descriptions of conformational intermediates and their associated models of transport mechanisms.

Of the LeuT Fold members represented with crystal structures, several commonalities in conformational sampling including the rotation of the bundle motif relative to the scaffold motif and the independent translation of TM1a are evident as would be expected for proteins exhibiting a common structural architecture. However, the diversity in degree to which these translations manifest in each LeuT Fold member as well as the presence of unique structural features and protein-specific conformational intermediates gives pause to interpretations of a unified LeuT Fold transport mechanism. The comparative analysis of transport mechanisms of LeuT and Mhp1 provided here will begin to address the similarities and differences among transport mechanisms in the LeuT Fold and whether a unified mechanism of transport can be established for LeuT Fold proteins. In doing so, the relationships between structure and function in the LeuT Fold will be addressed.

References

1. Mitchell, P. A general theory of membrane transport from studies of bacteria. *Nature* **180**, 134-136 (1957).
2. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969-970 (1966).
3. Vidaver, G. A. Inhibition of parallel flux and augmentation of counter flux shown by transport models not involving a mobile carrier. *J Theor Biol* **10**, 301-306 (1966).
4. Patlak, C. S. Contributions to the theory of active transport: Ii. The gate type non-carrier mechanism and generalizations concerning tracer flow, efficiency, and measurement of energy expenditure. *Bull. Math. Biophys.* **19**, 209-235 (1957).
5. Crane, R. K., Forstner, G. & Eichholz, A. Studies on the mechanism of the intestinal absorption of sugars. X. An effect of Na+ concentration on the apparent michaelis constants for intestinal sugar transport, in vitro. *Biochim Biophys Acta* **109**, 467-477 (1965).
6. Forrest, L. R. & Rudnick, G. The rocking bundle: A mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)* **24**, 377-386 (2009).
7. Jencks, W. P. Utilization of binding energy and coupling rules for active transport and other coupled vectorial processes. *Methods Enzymol* **171**, 145-164 (1989).
8. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587-593 (2002).
9. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a CLC chloride channel at 3.0 a reveals the molecular basis of anion selectivity. *Nature* **415**, 287-294 (2002).
10. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from escherichia coli. *Science* **301**, 616-620 (2003).
11. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. Structure and mechanism of the lactose permease of escherichia coli. *Science* **301**, 610-615 (2003).
12. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature* **437**, 215-223 (2005).
13. Ma, D., Lu, P., Yan, C., Fan, C., Yin, P., Wang, J. & Shi, Y. Structure and mechanism of a glutamate-GABA antiporter. *Nature* **483**, 632-636 (2012).
14. Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J. & Cameron, A. D. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709-713 (2008).
15. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. *Science* **321**, 810-814 (2008).
16. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* **458**, 47-52 (2009).
17. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a Na+-independent amino acid transporter. *Science* **325**, 1010-1014 (2009).

18. Tang, L., Bai, L., Wang, W. H. & Jiang, T. Crystal structure of the carnitine transporter and insights into the antiport mechanism. *Nat Struct Mol Biol* **17**, 492-496 (2010).
19. Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y. & Miller, C. Structure of a prokaryotic virtual proton pump at 3.2 a resolution. *Nature* **460**, 1040-1043 (2009).
20. Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., Wang, J. & Shi, Y. Structure and mechanism of an amino acid antiporter. *Science* **324**, 1565-1568 (2009).
21. Schulze, S., Koster, S., Geldmacher, U., Terwisscha van Scheltinga, A. C. & Kuhlbrandt, W. Structural basis of Na(+)-independent and cooperative substrate/product antiport in CaiT. *Nature* **467**, 233-236 (2010).
22. Singh, S. K., Piscitelli, C. L., Yamashita, A. & Gouaux, E. A competitive inhibitor traps leut in an open-to-out conformation. *Science* **322**, 1655-1661 (2008).
23. Claxton, D. P., Quick, M., Shi, L., de Carvalho, F. D., Weinstein, H., Javitch, J. A. & Mchaourab, H. S. Ion/substrate-dependent conformational dynamics of a bacterial homolog of Neurotransmitter:Sodium Symporters. *Nat Struct Mol Biol* **17**, 822-829 (2010).
24. Turk, E., Kim, O., le Coutre, J., Whitelegge, J. P., Eskandari, S., Lam, J. T., Kreman, M., Zampighi, G., Faull, K. F. & Wright, E. M. Molecular characterization of vibrio parahaemolyticus vsglt: A model for sodium-coupled sugar cotransporters. *J Biol Chem* **275**, 25711-25716 (2000).
25. Eskandari, S., Wright, E. M., Kreman, M., Starace, D. M. & Zampighi, G. A. Structural analysis of cloned plasma membrane proteins by freeze-fracture electron microscopy. *Proc Natl Acad Sci U S A* **95**, 11235-11240 (1998).
26. Suzuki, S. & Henderson, P. J. The hydantoin transport protein from microbacterium liquefaciens. *J Bacteriol* **188**, 3329-3336 (2006).
27. Farwick, M., Siewe, R. M. & Kramer, R. Glycine betaine uptake after hyperosmotic shift in corynebacterium glutamicum. *J Bacteriol* **177**, 4690-4695 (1995).
28. Jung, H., Buchholz, M., Clausen, J., Nietschke, M., Revermann, A., Schmid, R. & Jung, K. CaiT of escherichia coli, a new transporter catalyzing l-carnitine/gamma - butyrobetaine exchange. *J Biol Chem* **277**, 39251-39258 (2002).
29. Ujwal, M. L., Jung, H., Bibi, E., Manoil, C., Altenbach, C., Hubbell, W. L. & Kaback, H. R. Membrane topology of helices vii and xi in the lactose permease of escherichia coli studied by LacY-PhoA fusion analysis and site-directed spectroscopy. *Biochemistry* **34**, 14909-14917 (1995).
30. Vinothkumar, K. R., Raunser, S., Jung, H. & Kuhlbrandt, W. Oligomeric structure of the carnitine transporter CaiT from escherichia coli. *J Biol Chem* **281**, 4795-4801 (2006).
31. Iyer, R., Williams, C. & Miller, C. Arginine-agmatine antiporter in extreme acid resistance in escherichia coli. *J Bacteriol* **185**, 6556-6561 (2003).
32. Gong, S., Richard, H. & Foster, J. W. YjdE (AdiC) is the Arginine:Agmatine antiporter essential for arginine-dependent acid resistance in escherichia coli. *J Bacteriol* **185**, 4402-4409 (2003).
33. Fang, Y., Kolmakova-Partensky, L. & Miller, C. A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. *J Biol Chem* **282**, 176-182 (2007).
34. Casagrande, F., Ratera, M., Schenk, A. D., Chami, M., Valencia, E., Lopez, J. M., Torrents, D., Engel, A., Palacin, M. & Fotiadis, D. Projection structure of a member of the amino acid/polyamine/organocation transporter superfamily. *J Biol Chem* **283**, 33240-33248 (2008).
35. Hersh, B. M., Farooq, F. T., Barstad, D. N., Blankenhorn, D. L. & Slonczewski, J. L. A glutamate-dependent acid resistance gene in escherichia coli. *J Bacteriol* **178**, 3978-3981 (1996).

36. Castanie-Cornet, M. P., Penfound, T. A., Smith, D., Elliott, J. F. & Foster, J. W. Control of acid resistance in escherichia coli. *J Bacteriol* **181**, 3525-3535 (1999).
37. Jao, C. C., Hegde, B. G., Gallop, J. L., Hegde, P. B., McMahon, H. T., Haworth, I. S. & Langen, R. Roles of amphipathic helices and the bar domain of endophilin in membrane curvature generation. *J Biol Chem*.
38. Masson, J., Sagne, C., Hamon, M. & El Mestikawy, S. Neurotransmitter transporters in the central nervous system. *Pharmacol Rev* **51**, 439-464 (1999).
39. Singh, S. K. LeuT: A prokaryotic stepping stone on the way to a eukaryotic neurotransmitter transporter structure. *Channels (Austin)* **2** (2008).
40. Beuming, T., Shi, L., Javitch, J. A. & Weinstein, H. A comprehensive structure-based alignment of prokaryotic and eukaryotic Neurotransmitter/Na+ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Mol Pharmacol* **70**, 1630-1642 (2006).
41. Forrest, L. R., Tavoulari, S., Zhang, Y. W., Rudnick, G. & Honig, B. Identification of a chloride ion binding site in na+/cl -dependent transporters. *Proc Natl Acad Sci U S A* **104**, 12761-12766 (2007).
42. Zomot, E., Bendahan, A., Quick, M., Zhao, Y., Javitch, J. A. & Kanner, B. I. Mechanism of chloride interaction with neurotransmitter:Sodium symporters. *Nature* **449**, 726-730 (2007).
43. Zhou, Z., Zhen, J., Karpowich, N. K., Goetz, R. M., Law, C. J., Reith, M. E. & Wang, D. N. Leut-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317**, 1390-1393 (2007).
44. Singh, S. K., Yamashita, A. & Gouaux, E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448**, 952-956 (2007).
45. Zhou, Z., Zhen, J., Karpowich, N. K., Law, C. J., Reith, M. E. & Wang, D. N. Antidepressant specificity of serotonin transporter suggested by three leut-ssri structures. *Nat Struct Mol Biol* **16**, 652-657 (2009).
46. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347-355 (2009).
47. Wright, E. M., Loo, D. D. & Hirayama, B. A. Biology of human sodium glucose transporters. *Physiol Rev* **91**, 733-794 (2011).
48. Reed-Tsur, M. D., De la Vieja, A., Ginter, C. S. & Carrasco, N. Molecular characterization of v59e nis, a na+/i- symporter mutant that causes congenital i- transport defect. *Endocrinology* **149**, 3077-3084 (2008).
49. Wright, E. M., Hirayama, B. A. & Loo, D. F. Active sugar transport in health and disease. *J Intern Med* **261**, 32-43 (2007).
50. Watanabe, A., Choe, S., Chaptal, V., Rosenberg, J. M., Wright, E. M., Grabe, M. & Abramson, J. The mechanism of sodium and substrate release from the binding pocket of vsglt. *Nature* **468**, 988-991 (2010).
51. Pantazopoulou, A. & Diallinas, G. Fungal nucleobase transporters. *FEMS Microbiol Rev* **31**, 657-675 (2007).
52. Ren, Q., Chen, K. & Paulsen, I. T. Transportdb: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res* **35**, D274-279 (2007).
53. Saier, M. H., Jr. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. *Adv Microb Physiol* **40**, 81-136 (1998).
54. Saier, M. H., Jr., Tran, C. V. & Barabote, R. D. Tcdb: The transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**, D181-186 (2006).

55. Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J. & Yan, N. Structure and mechanism of the uracil transporter uraa. *Nature* **472**, 243-246 (2011).
56. Ziegler, C., Bremer, E. & Kramer, R. The bcct family of carriers: From physiology to crystal structure. *Mol Microbiol* **78**, 13-34 (2010).
57. Fritz, I. B. & Yue, K. T. Effects of carnitine on acetyl-coa oxidation by heart muscle mitochondria. *Am J Physiol* **206**, 531-535 (1964).
58. Treem, W. R., Stanley, C. A., Finegold, D. N., Hale, D. E. & Coates, P. M. Primary carnitine deficiency due to a failure of carnitine transport in kidney, muscle, and fibroblasts. *N Engl J Med* **319**, 1331-1336 (1988).
59. Nalecz, K. A., Miecz, D., Berezowski, V. & Cecchelli, R. Carnitine: Transport and physiological functions in the brain. *Mol Aspects Med* **25**, 551-567 (2004).
60. Perez, C., Koshy, C., Yildiz, O. & Ziegler, C. Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter betp. *Nature* **490**, 126-130 (2012).
61. Gartner, R. M., Perez, C., Koshy, C. & Ziegler, C. Role of bundle helices in a regulatory crosstalk in the trimeric betaine transporter betp. *J Mol Biol* **414**, 327-336 (2011).
62. Jack, D. L., Paulsen, I. T. & Saier, M. H. The amino acid/polyamine/organocation (apc) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* **146 (Pt 8)**, 1797-1814 (2000).
63. Zimmermann, N., King, N. E., Laporte, J., Yang, M., Mishra, A., Pope, S. M., Muntel, E. E., Witte, D. P., Pegg, A. A., Foster, P. S., Hamid, Q. & Rothenberg, M. E. Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J Clin Invest* **111**, 1863-1874 (2003).
64. Yanagida, O., Kanai, Y., Chairoungdua, A., Kim, D. K., Segawa, H., Nii, T., Cha, S. H., Matsuo, H., Fukushima, J., Fukasawa, Y., Tani, Y., Taketani, Y., Uchino, H., Kim, J. Y., Inatomi, J., Okayasu, I., Miyamoto, K., Takeda, E., Goya, T. & Endou, H. Human l-type amino acid transporter 1 (lat1): Characterization of function and expression in tumor cell lines. *Biochim Biophys Acta* **1514**, 291-302 (2001).
65. Calonge, M. J., Gasparini, P., Chillaron, J., Chillon, M., Gallucci, M., Rousaud, F., Zelante, L., Testar, X., Dallapiccola, B., Di Silverio, F. & et al. Cystinuria caused by mutations in rbat, a gene involved in the transport of cystine. *Nat Genet* **6**, 420-425 (1994).
66. Torrents, D., Mykkanen, J., Pineda, M., Feliubadalo, L., Estevez, R., de Cid, R., Sanjurjo, P., Zorzano, A., Nunes, V., Huoponen, K., Reinikainen, A., Simell, O., Savontaus, M. L., Aula, P. & Palacin, M. Identification of slc7a7, encoding y+lat-1, as the lysinuric protein intolerance gene. *Nat Genet* **21**, 293-296 (1999).
67. Borsani, G., Bassi, M. T., Sperandeo, M. P., De Grandi, A., Buoninconti, A., Riboni, M., Manzoni, M., Incerti, B., Pepe, A., Andria, G., Ballabio, A. & Sebastio, G. Slc7a7, encoding a putative permease-related protein, is mutated in patients with lysinuric protein intolerance. *Nat Genet* **21**, 297-301 (1999).
68. Harada, N., Nagasaki, A., Hata, H., Matsuzaki, H., Matsuno, F. & Mitsuya, H. Down-regulation of cd98 in melphalan-resistant myeloma cells with reduced drug uptake. *Acta Haematol* **103**, 144-151 (2000).
69. Gao, X., Zhou, L., Jiao, X., Lu, F., Yan, C., Zeng, X., Wang, J. & Shi, Y. Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463**, 828-832 (2010).
70. Forrest, L. R., Zhang, Y. W., Jacobs, M. T., Gesmonde, J., Xie, L., Honig, B. H. & Rudnick, G. Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci U S A* **105**, 10338-10343 (2008).

71. Androutsellis-Theotokis, A., Goldberg, N. R., Ueda, K., Beppu, T., Beckman, M. L., Das, S., Javitch, J. A. & Rudnick, G. Characterization of a functional bacterial homologue of sodium-dependent neurotransmitter transporters. *J Biol Chem* **278**, 12703-12709 (2003).
72. Just, H., Sitte, H. H., Schmid, J. A., Freissmuth, M. & Kudlacek, O. Identification of an additional interaction domain in transmembrane domains 11 and 12 that supports oligomer formation in the human serotonin transporter. *J Biol Chem* **279**, 6650-6657 (2004).
73. Chen, J. G., Liu-Chen, S. & Rudnick, G. Determination of external loop topology in the serotonin transporter by site-directed chemical labeling. *J Biol Chem* **273**, 12675-12681 (1998).
74. Bismuth, Y., Kavanaugh, M. P. & Kanner, B. I. Tyrosine 140 of the gamma-aminobutyric acid transporter gat-1 plays a critical role in neurotransmitter recognition. *J Biol Chem* **272**, 16096-16102 (1997).
75. Kniazeff, J., Shi, L., Loland, C. J., Javitch, J. A., Weinstein, H. & Gether, U. An intracellular interaction network regulates conformational transitions in the dopamine transporter. *J Biol Chem* **283**, 17691-17701 (2008).
76. Zhao, Y., Terry, D., Shi, L., Weinstein, H., Blanchard, S. C. & Javitch, J. A. Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188-193 (2010).
77. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A. & Fujiyoshi, Y. Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599-605 (2000).
78. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from pyrococcus horikoshii. *Nature* **431**, 811-818 (2004).
79. Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E. & Michel, H. Structure of a na+/h+ antiporter and insights into mechanism of action and regulation by ph. *Nature* **435**, 1197-1202 (2005).
80. Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J. A. The mechanism of a neurotransmitter:Sodium symporter--inward release of na+ and substrate is triggered by substrate in a second binding site. *Mol Cell* **30**, 667-677 (2008).
81. Quick, M., Winther, A. M., Shi, L., Nissen, P., Weinstein, H. & 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 U S A* **106**, 5563-5568 (2009).
82. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the leut and apct transporters are modulated by common mechanisms. *Biophys J* **99**, L103-105 (2010).
83. Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter mhp1. *Science* **328**, 470-473 (2010).
84. Weyand, S., Shimamura, T., Beckstein, O., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter mhp1. *J Synchrotron Radiat* **18**, 20-23 (2011).
85. Krishnamurthy, H. & Gouaux, E. X-ray structures of leut in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469-474 (2012).
86. Piscitelli, C. L., Krishnamurthy, H. & Gouaux, E. Neurotransmitter/sodium symporter orthologue leut has a single high-affinity substrate site. *Nature* **468**, 1129-1132 (2010).
87. Mondal, S., Khelashvili, G., Shi, L. & Weinstein, H. The cost of living in the membrane: A case study of hydrophobic mismatch for the multi-segment protein leut. *Chem Phys Lipids* **169**, 27-38 (2013).

88. Wang, H., Elferich, J. & Gouaux, E. Structures of leut in bicelles define conformation and substrate binding in a membrane-like context. *Nat Struct Mol Biol* **19**, 212-219 (2012).
89. Zhao, Y., Terry, D. S., Shi, L., Quick, M., Weinstein, H., Blanchard, S. C. & Javitch, J. A. Substrate-modulated gating dynamics in a na+-coupled neurotransmitter transporter homologue. *Nature* **474**, 109-113 (2011).
90. Quick, M., Shi, L., Zehnpfennig, B., Weinstein, H. & Javitch, J. A. Experimental conditions can obscure the second high-affinity site in leut. *Nat Struct Mol Biol* **19**, 207-211 (2012).
91. Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta* **1807**, 167-188 (2011).
92. Jeschke, G. A comparative study of structures and structural transitions of secondary transporters with the leut fold. *Eur Biophys J* **42**, 181-197 (2013).

CHAPTER 2

ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY

Emergence of EPR spectroscopy as a structural biology methodology

Structural biology of membrane proteins

Structural biology represents an interdisciplinary approach to understanding biological questions through investigating the molecular structure and dynamics of macromolecules such as proteins and DNA. This discipline has emerged as a critical component in describing protein functional mechanisms and understanding the processes that support life and create disease. Serving as a central node, structural biology connects various disciplines through rationalizing genetic polymorphisms, providing context for biochemical signaling events, and generating templates for protein engineering and rational drug design. After decades of slow progress, completion of genome sequencing projects, advances in protein expression and purification, and methodological innovations have begun to overcome long-standing barriers and bottlenecks spurring a spectacular acceleration in the pace of protein structure determination and more recently, membrane protein structure determination.

Membrane proteins are key control points in cell communication, in movement of molecules across membrane barriers, in the flow and use of energy, as well as in triggering the initiation of numerous signaling pathways. The 3D structures of these high value drug targets elucidate the architectural principles that define classes of membrane proteins, exposing motifs that determine their stability and enable them to inhabit the lipid bilayer[1], and unlocking secrets of ion channel selectivity, transporter specificity[2], receptor/ligand interactions[3] and catalysis in the membrane[4]. Atomic resolution protein structures have revolutionized our understanding of

these processes and set the stage for the next frontier of structural biology: the translation of static structures into protein mechanisms.

Using protein crystal structures to define conformational intermediates

While crystallographic snapshots frame biochemical and functional data in a structural context, achieving a mechanistic description of biological function requires an understanding of dynamics, the fourth dimension of protein structure. The function of channels, transporters, and receptors is intimately associated with their ability to execute movements that enable opening of a gate, alternate access of a substrate binding pocket to different sides of the membrane, or expose signaling sequences. Excursions between conformers can be thermally activated; a view in stark contrast to the static picture communicated by crystal structures. In some cases, models of conformational changes can be inferred from a "patchwork of different homologs fortuitously crystallized in different states"[5], but the caveat is that the observed distribution of structures may reflect "the idiosyncrasies of the different homologs"[6] rather than different intermediates in the functional cycle.

Even in instances where multiple conformations of a protein have been captured crystallographically, interpretation of these structures in a mechanistic context must be cautious as representations of protein dynamics and conformational sampling can be altered by the crystallization process. Crystal contacts can act as a conformational selectivity filter distorting highly flexible but functionally critical segments and/or stabilizing conformations that may be sparsely populated in solution. Moreover, membrane proteins' natural milieu is the lipid bilayer, which differs in its physico-chemical properties from detergent micelles, the preferred crystallography solvent. Accentuating this concern, detergent selection criteria often emphasize crystal and diffraction qualities at the expense of functional considerations thus dictating the use of harsh detergents.

Similarly, the natural conformational polydispersity of dynamic proteins presents a challenge to the crystallization process. Therefore, highly flexible proteins are significantly underrepresented in the Protein Data Base. Methods used to limit polydispersity make interpretation of structures potentially problematic. Manipulations of sequence including thermal stabilization, surface engineering, addition of non-native structural domains, and deletion of dynamic motifs that intentionally alter protein conformational sampling have been used extensively to broaden the base of protein classes amenable to crystallographic structure determination. Conformational selection of intermediate states, often the goal of structural investigations of dynamic proteins, has been attempted through mutation of conserved residues and/or by binding the protein to "conformationally selective" antibodies resulting in ambiguity over the mechanistic identity of ensuing structures. Together these and other factors conspire to cloud the interpretation of crystal structures into a mechanistic context[7]. Therefore, a detailed understanding of membrane protein functional cycles requires a description of the nature, amplitude and time scale of conformational equilibria and/or triggered conformational changes in a native-like environment.

Spectroscopic approaches to describe dynamic protein structure and mechanism

Dynamics is the realm of spectroscopy by excellence. Spectroscopic approaches such as nuclear magnetic resonance (NMR)[8] and electron paramagnetic resonance (EPR)[9-10] have been successfully used to directly detect many facets of protein dynamics including molecular tumbling, domain movements, backbone fluctuations, and side chain isomerizations. As solution-based approaches, these methods accommodate polydispersity allowing proteins to sample equilibrium dynamic modes or undergo triggered conformational changes and these features are monitored as an ensemble of states. Furthermore, these experiments are conducted under conditions more closely resembling their native environment with only minor modifications to sequence. Despite its potential, the use of NMR has been limited by mediocre

sensitivity, the need for isotopic labeling, and molecular mass limitations that exclude the vast majority of membrane proteins. In contrast, sensitivity and size are not limiting for probe-based spectroscopic approaches like EPR that interpret spectral properties of site-specifically incorporated probes to deduce local structural features. For these reasons, EPR has become as a necessary complementary approach for evaluating dynamic transport mechanisms of membrane proteins.

EPR theory

Physical principles underlying the EPR signal

EPR spectroscopy, also referred to as electron spin resonance (ESR) spectroscopy, reports microwave radiation-induced transitions in discrete energy levels of unpaired electrons in a magnetic field[11-12]. Unpaired electrons possess net spin and orbital angular momentum which are neutralized when electrons are paired. Therefore, EPR spectroscopy requires native or exogenous unpaired electrons within the molecule of interest to produce an EPR signal. Electronic transitions of free electrons within the system are sensitive to the properties of their molecular environment and it is this sensitivity that is exploited in the various experiments available to the EPR spectroscopist including local mobility and solvent accessibility measurements as well as distance measurements between multiple spins. These measurements can be interpreted to describe molecular motions and interactions and can be used as restraints for computational modeling of protein structure and dynamics.

The physical basis for the EPR signal derives from the permanent magnetic moments of the electrons, conferred by their charge and quantized spin angular momentum[11-12]. One of the advantages of EPR over NMR is the significantly greater magnitude of magnetic moments of electron spins compared to protons. Practically, this results in higher signal sensitivity, which

lowers sample concentration requirements and allows longer distance spin-spin interactions and faster molecular motions to be monitored.

In an external magnetic field (H_o), the magnetic moments of electrons align into discrete energy levels corresponding to parallel and antiparallel orientations relative to the magnetic field[11-12]. The energetic description of the interaction between the electron and the H_o can be described along the z axis by the Hamiltonian

$$\mathcal{H} = -\mu_z H_o$$

where μ_z describes the magnetic moment of the electron. This function can be written to include the z axis electron spin quantum operator, S_z, as

$$\mathcal{H} = gBH_oS_z$$

where g is the electron g factor and B is the electron Bohr magneton, both known constants for a free electron system. For the electron, there are two possible eigenvalues (M_s) of S_z, ±½. Therefore two energy levels referred to as α and β that are defined in Zeeman interaction energies as $-½gBH_o$ and $+½gBH_o$, respectively. In paired electron systems, electrons must possess different M_s based on the Pauli Exclusion Principle. In these cases, spin angular momentum cancels and the magnetic moment of the system is zero. It is for this reason, that paired electrons are EPR silent and unpaired electrons are required for EPR analysis.

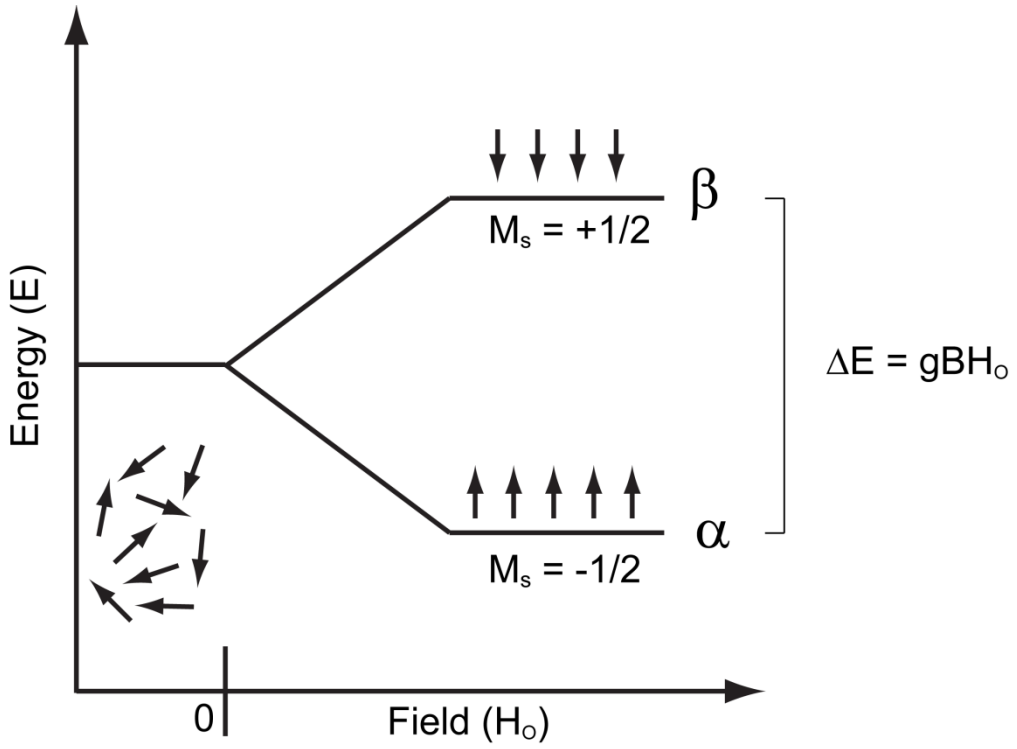


Figure 2.1. Population of spins into discrete energy levels in the presence of a magnetic field

In an ensemble of molecules possessing unpaired electrons, electron spins in a magnetic field occupy either the α or β energy states, with the α state corresponding to the lower energy, parallel orientation and the β state corresponding to the higher energy, antiparallel orientation relative to the direction of the magnetic field[11-12] (Fig. 2.1). The difference in energy (ΔE) between these states is defined by the strength of the magnetic field.

$$\Delta E = gBH_o$$

In the absence of a magnetic field there is no energetic difference between these states. As field strength increases, the difference in energy between the states increases as well. At thermal equilibrium, the relative population of spins in each of the energy levels is defined by the Boltzmann distribution

$$\frac{N_\alpha}{N_\beta} = e^{(\Delta E/kT)}$$

where k is the Boltzmann constant, T is the absolute temperature, and N_α and N_β correspond to the populations of states α and β, respectively. For example, the ratio of N_α/N_β at 293K in a 1T magnetic field is ~1.0046, slightly favoring the lower energy α state.

The EPR experiment measures net transitions between these two states[11-12]. According to Planck's law, electron spins can be induced to transition between energy levels through application and absorbance of a frequency of the electromagnetic radiation, v, that corresponds to the energy difference between the states, modified by the Planck constant, h.

$$\Delta E = gBH_o = hv$$

This resonance condition results in electron spin transitions from both α and β state populations with equal probability. Therefore, the net transition occurs from the α state to the β state based on the small population differences between the states dictated by the Boltzmann distribution. It is the net transition that is measured by the EPR experiment, which manifests as an absorbance of electromagnetic radiation at given frequency and at a given magnetic field strength. EPR experiments can be conducted by scanning either the frequency or magnetic field regimes to induce this absorbance. As the signal strength is dependent upon the Boltzmann distribution, it is therefore dependent upon the magnetic field strength and temperature. This relationship results in increased signal strength at lower experimental temperatures and higher magnetic field strengths.

Introducing free electrons into proteins with Site-Directed Spin Labeling

To conduct EPR experiments, an unpaired electron must be present in the macromolecular system of interest. Due to the rarity of unpaired electrons in nature, applications of EPR have historically been limited to biological systems and processes that naturally incorporate EPR active transition metals, such as the photosynthetic reaction centers[13], and organic radicals including biradical and triplet state molecules[11] and oxidation/reduction reactions[14]. This rarity, however, provides EPR with significantly increased sensitivity over NMR

where NMR-active protons and carbon isotopes are naturally ubiquitous. For these reasons, development of EPR methodologies that allow site-specific incorporation of unpaired electrons into protein systems was a highly sought advancement in the field. Progress on this front came with two major methodological achievements: the development of site-directed mutagenesis and recombinant protein expression and the synthesis of stable, paramagnetic spin probes. Together these methods became site-directed spin labeling[10] (SDSL, Fig. 2.2).

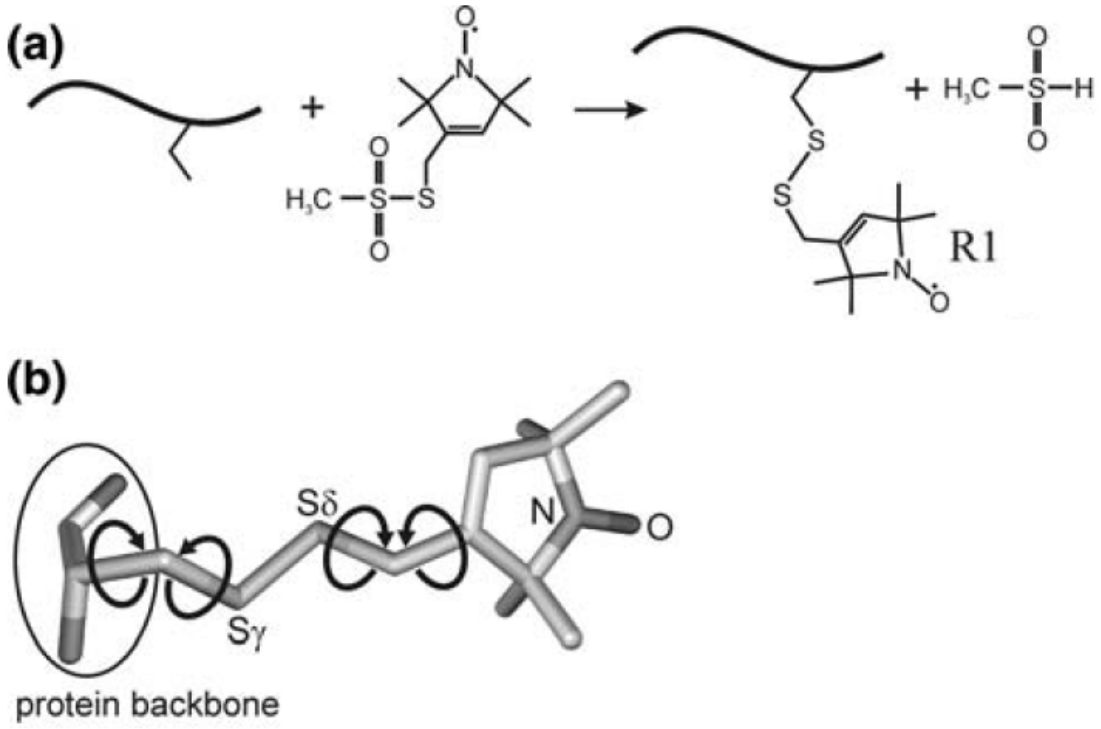


Figure 2.2. Site-Directed Spin Labeling. (a) Incorporation of the spin label into the protein. (b) Structure and rotameric freedom of MTSSL. Figure modified with permission from [15].

Specifically, SDSL[10] involves mutagenesis of protein sequence to engineer cysteine residues only at selected sites of interest in the protein sequence. Sites for cysteine replacement are selected to avoid structural perturbation, typically located on the protein surface at non-conserved residues. A study measuring the structural and functional effects of

spin labeling in T4L found that native residue substitution and spin labeling at these residues did not significantly alter protein structural and functional properties[16]. Recominant DNA is introduced into protein expression vectors such as E. coli and resulting protein is purified from whole cell extract using affinity, ion exchange, and/or size exclusion chromatography. Reducing agents such as DTT are used throughout the purification process to shield exposed cysteine residues. Isolated protein undergoes spin labeling, wherein spin probes containing stable radical species such as the nitroxide radical are covalently attached to engineered cysteine residues through thiol reactive functional groups including methanethiolsulfonate, maleimide, and iodoacetamide moieties[15] (Fig. 2.2a). This methodology allowed selective placement of paramagnetic centers into diamagnetic protein systems and initiated the widespread use of EPR as a technique for understanding protein structure and dynamics.

The most commonly used spin label for protein investigations is the methanethiosulfonate nitroxide label[15] (MTSSL, Fig. 2.2b). In this molecule, the radical species is located in a π-like orbital along the N-O bond, located nearest to the N nucleus. The radical is stable, even in the presence of biological reductants, due to steric shielding provided by the proximal set of dimethyl groups of the pyrrole ring. MTSSL is readily attached to cysteine residues through its highly reactive thiol functional group and can be easily cleaved with reducing agents for control experiments. Furthermore, MTSSL is theoretically well characterized[16] and provides a balance in molecular flexibility, allowing high efficiency labeling at most sites without limiting molecular motion interpretations with an overabundance of rotameric freedom.

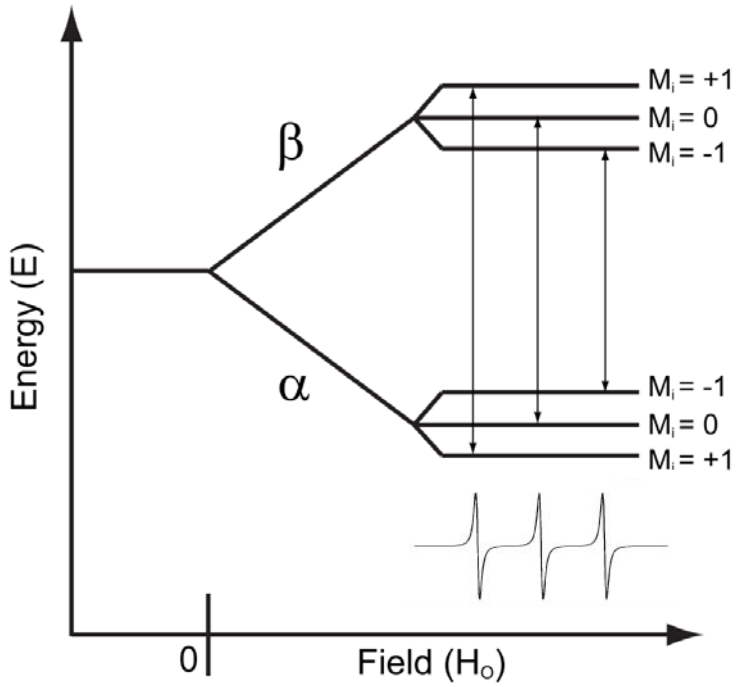


Figure 2.3. Interaction between electron and nuclear spins produce additional energy levels. Allowed transitions between energy levels result in the three absorbance peaks characteristic of the MTSSL EPR spectra.

The hyperfine interaction defines the EPR signal in spin labeled systems

The EPR signal reported by this molecule is influenced by spin-spin interactions between the unpaired electron and the N nucleus, referred to as the hyperfine interaction[11-12]. Like the electron, the ^{14}N nucleus possesses a permanent magnetic moment as a result of its nuclear spin and angular momentum properties. The spin quantum operator for the ^{14}N nucleus is associated with three eigenvalues (M_I), ±1 and 0. Therefore, there exist three energy states for the ^{14}N nuclear spin described by their Zeeman interaction energies as $\pm 1 g_n B_n H_o$ and $0 g_n B_n H_o$, where g_n and B_n represent the nuclear g factor and Bohr magneton, respectively. The interaction between the electron spin and the adjacent nuclear spin of the N introduces additional energy states for the electron, based on the combination of the two electron energy states and the three nuclear energy states, for a total of six energy states (Fig. 2.3). Of these six potential electronic transitions, only three are quantum mechanically allowed, as the nuclear

spins are not affected by microwave radiation and therefore their spins states remain static during electronic transitions. These three allowed transitions of the nitroxide radical form the basis for the three characteristic absorbance peaks of the EPR signal.

Orientation-dependence of the EPR signal allows mobility investigations

The observed EPR signal is dependent upon the orientation of the nitroxide spin label relative to the magnetic field and this dependence gives rise to the spin label anisotropy measurements that characterize local spin label mobility investigations. Orientation dependence is the result of interactions between electron spin angular momentum and orbital angular momentum. For radicals like the nitroxide radical that inhabit non-spherical orbitals, the local magnetic field experienced by the electron will vary based on orientation relative to the external magnetic field. For this reason, the electron and nuclear g factors and hyperfine coupling constants that define energy states and resonance frequencies are also orientation-dependent. Therefore, EPR spectroscopy can be used to monitor the orientation of the spin label.

In solution, the EPR signal reports the time-averaged ensemble of spin labeled molecules[12]. If the spin labels positions are static relative to the molecule, all possible spin orientations and their associated resonance conditions contribute to the EPR spectrum resulting in a broadened “rigid limit” line shape (Fig. 2.4). Alternatively, if the spin label has complete freedom of motion, as is the case for free, unbound label, then the orientation-dependent differences in local magnetic field are averaged out over time and the resulting spectrum yields the narrow “fast motion” line shape, characteristic of orientation-independent spectra. EPR spectra can be used to define spin label anisotropy in a continuum between these extremes as rates of reorientation and can report relative population of differently mobile subpopulations within the ensemble[17] (Fig. 2.4). Because spin label anisotropy is related to local environmental features, these mobility analyses can be used to describe surface exposure, define secondary

structure and topology, and monitor ligand and protein interactions as an equilibrium feature of the molecular ensemble.

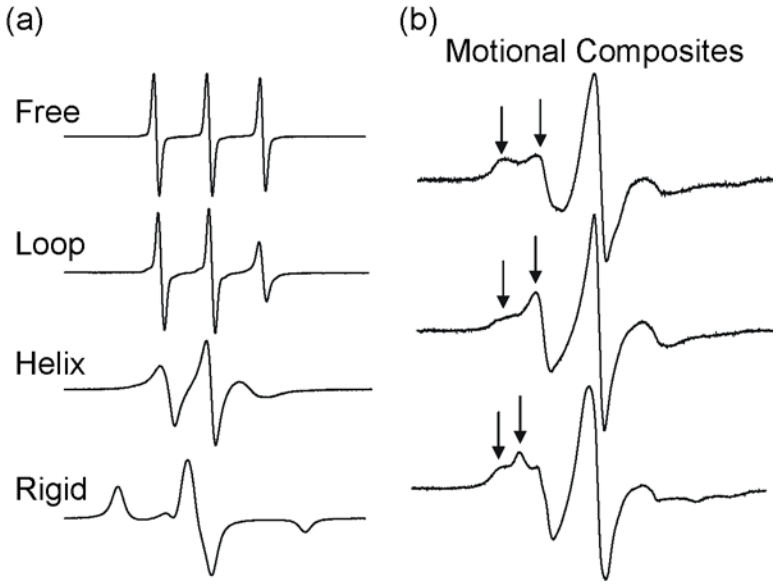


Figure 2.4. EPR spectra of differently mobile spin labels. (Left) Spin label mobility reports local environmental features such as secondary structure and tertiary contacts. (Right) Multiple mobility profiles within the ensemble will be reflected as composite EPR spectra. Figure modified with permission from [17].

Electron spin relaxation processes

As stated previously, EPR measures the net transitions of electron spins between discrete energy states[11-12]. The processes described thus far have discussed monitoring of absorbance of electromagnetic radiation and electron spin excitation. The rate of electron spin relaxation to equilibrium can also be monitored by EPR. As relaxation processes are also dependent upon environmental properties, these measurements can be used to describe such features as solvent accessibility and are used for distance measurements[11-12].

In an external magnetic field, the net population difference between the parallel α and antiparallel β energy states results in a bulk magnetization vector, M, located on the +z axis, parallel to external magnetic field[11-12]. The magnitude of the vector is proportional to the

difference in population between the two states. In the presence of applied microwave radiation, electron spins transition between states changing the population frequencies of each state. If the microwave radiation induces transitions to the point that the populations equalize, M along the z axis becomes zero. In pulsed methods, this duration of microwave radiation is referred to as a $\pi/2$ pulse, as M rotates 90° into the x -y plane to report a zero value along the z axis[18]. When the microwave radiation induces electronic transitions such that the net population difference completely favors the higher energy β state, M aligns along the –z axis, antiparallel to the magnetic field[18]. This duration of microwave radiation is referred to as a π pulse wherein M rotates 180°. These microwave pulses impose non -equilibrium conditions and at their termination, thermal equilibrium must be restored.

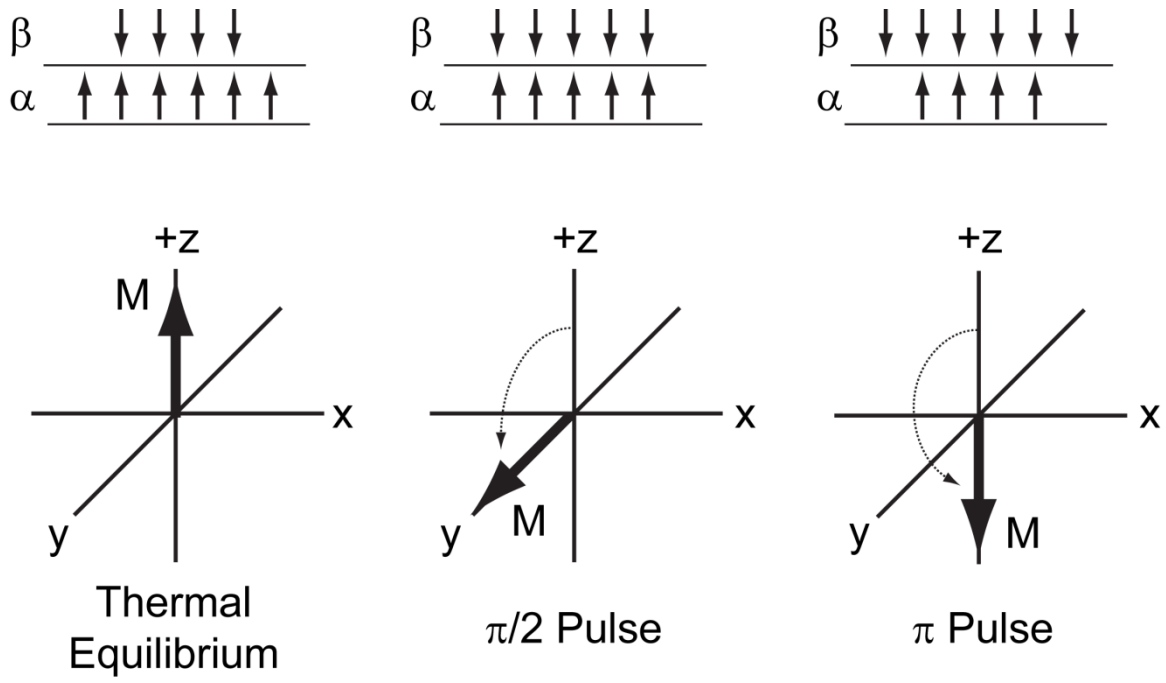


Figure 2.5. Changes in directionality of the bulk magnetization vector due to microwave pulses. Relaxation to thermal equilibrium along the z axis is spin-lattice relaxation.

The process by which M returns to +z axis is referred to as spin-lattice relaxation[11]. The rate constant that describes the recovery of thermal equilibrium along the z axis is T_1, the spin-lattice relaxation time[11]. Spin-lattice relaxation is the result of energy transfer between excited spins and the molecular and intermolecular environment or lattice through non-radiative mechanisms such as molecular rotations or vibrations. The rate of this recovery, T_1, is related to the degree of coupling between electron spins and the environment and is specific to the protein and solution conditions. With this energy transfer, spins are allowed to return to the lower energy states and thermal equilibrium is recovered. The frequency of these energy transfers naturally decay exponentially and this is the basis for the exponential decay component of EPR spectra. The relationship between M and T_1 as a function of time is related in the Bloch equation

$$\frac{dM_z}{dt} = \frac{M_o - M_z}{T_1}$$

where M_z is the initial bulk magnetization along the z axis and M_o is the bulk magnetization at thermal equilibrium after time T_1.

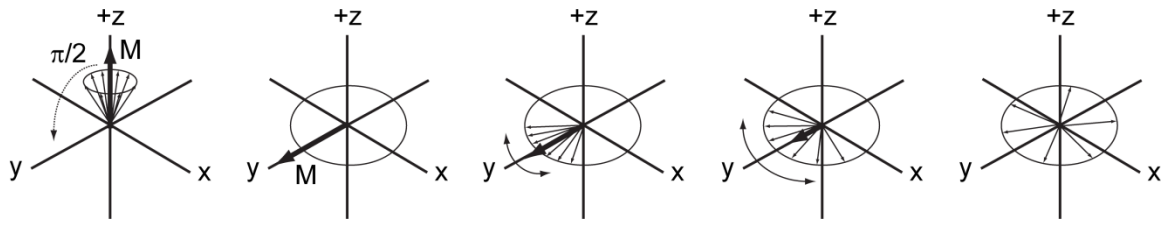


Figure 2.6. Spin-spin relaxation causes spin dephasing and signal attenuation in the x-y plane

As described previously, M reports the net orientation of electron spins between parallel and antiparallel orientations in the presence of an external magnetic field[11-12]. At thermal equilibrium, M lies along the +z axis relative the external magnetic field. M, as an average orientation in this case, can be misleading as it implies electron spins are aligning in the z axis.

In reality, individual electron spins in both orientations precess around z axis, into the x-y plane (Fig. 2.6). This is due to the interaction between electron spin angular momentum and the magnetic field resulting in torque being applied to the electron spin magnetic moment by the magnetic field[11-12]. The frequency of this precession is referred to as the Larmor frequency. In absence of relaxation, electron spins precess in phase and therefore bulk magnetization in the x-y plane will align. However, dipolar and collisional exchange interactions between spins can cause changes in spin angular momentum, which result in a dephasing of electron spin precession and a decrease in the bulk magnetization in the x-y plane[11] (Fig. 2.6). This is referred to as spin-spin relaxation and the rate at which the magnetization in the x-y plane exponentially decays to zero, is T_2[11]. The relationship between bulk magnetization in the x-y plane, given by M_x and M_y, and the spin-spin relaxation time, T_2, is given by the Bloch equations

$$\frac{dM_x}{dt} = \gamma H_o M_y - \frac{M_x}{T_2}$$

$$\frac{dM_y}{dt} = \gamma H_o M_x - \frac{M_y}{T_2}$$

where γ is the electron gyromagnetic ratio, equal to $2\pi gB/h$. Bulk magnetization monitored in the x or y planes contributes to the EPR signal as an exponentially decreasing oscillation as in the classic free induction decay (FID).

Relaxation mechanism measurements monitor label accessibility to solvent

Monitoring these relaxation processes can provide valuable information about the molecular system and form the basis for measures of solvent accessibility and distance measurement between spins[10,19]. In the solvent accessibility measurement[20], net electron spin excitation is saturated by application of microwave radiation and subsequent recovery rate, T_1, is measured in the presence or absence of water soluble paramagnetic relaxation agents (PRA)

such as NiEDDA or molecular oxygen (O_2) in solution. Collision between PRAs and radical species results in spin exchange, which decreases electron spin saturation and increases monitored T_1 relative to the absence of the PRA. The increase in T_1 is correlated to the rate of collision frequency between the PRA and the spin label and therefore reports relative accessibility of the spin label to the solution environment. These measurements can be used to define surface exposure as well as conformational transitions resulting in changes in solvent accessibility[10], an important feature of molecular biochemistry. When monitored at consecutive residues, solvent accessibility measurements can define secondary structure due to the characteristic pattern of changes in accessibility[10] (Fig. 2.7). Furthermore, membrane environments possess gradient concentrations of PRAs that allow accessibility measurements to define membrane depth of spin labels and orientation of membrane imbedded, spin labeled secondary structural elements[21-22].

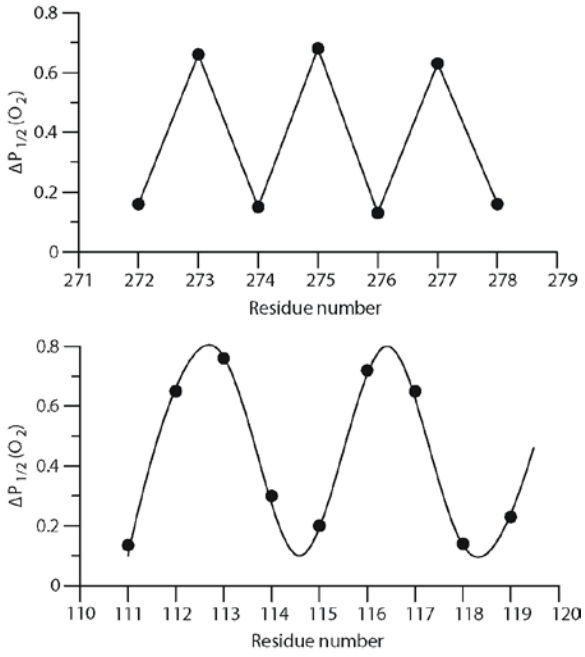


Figure 2.7. Solvent accessibility measurements can define secondary structure. (Top) The characteristic pattern of solvent accessibility for a β strand. (Bottom) The solvent accessibility pattern for a helix. Figure modified with permission from [17].

Distance measurements result from dipolar coupling between spins

Distance measurements rely on the distance-dependence of dipolar coupling interactions between electron spins. In EPR, dipolar coupling occurs when electron spins are in close enough in proximity that they experience each other's magnetic fields, which modulates their effective local magnetic field and causes changes in resonance conditions and reported absorbance peaks in the EPR line shapes[11]. Relative dipolar coupling between spins can be measured through a variety of EPR experiments and distance information extracted from these results can provide importance information about molecular architecture and can inform on protein dynamics.

Spin-spin interactions can occur in two ways, through bonds as in J coupling and through space as in dipolar coupling[11]. The Hamiltonian describing a two nitroxide spin system highlights these potential interactions

$$\mathcal{H} = g^1 B H_o S^1 + g^2 B H_o S^2 - \omega_n (I_z^1 + I_z^2) + \gamma I^1 A^1 S^1 + \gamma I^2 A^2 S^2 + D S^1 S^2 + J S^1 S^2$$

where ω_n is the Larmor frequency of the nitrogen nucleus, I is the spin quantum operator of the nitrogen nucleus, A is the isotropic hyperfine coupling constant, D is the dipolar coupling tensor, and J is the spin-spin exchange factor. The numeric superscripts identify variables for each electron spin. EPR experiments are typically conducted in doubly spin labeled molecular systems or in two interacting singly labeled molecular systems. Due to the length of the spin label tether, through bond J coupling interactions do not significantly contribute to spin-spin interactions in these systems and are disregarded in this analysis[23]. Through dipolar coupling mechanisms, distances can be monitored to when two spins are in relative proximity, in a range between 8 and 60Å in distance[23].

Without contribution from dipolar and J coupling, the Hamiltonian reduces to describe the sum of two non-interacting spins. In the presence of dipolar coupling, the energy of the

interaction is inversely proportional to the cube of the distance (r^3) and is orientation-dependent[11]. Thus, the energetic relationship between the interacting dipoles (μ) is expressed in terms of distance, r, and orientation of the interaction relative to the magnetic field, θ (Fig. 2.8, left).

$$E = \frac{\mu_1 \mu_2}{r^3} (3\,cos^2\,\theta - 1)$$

The angular dependence of the dipolar interaction is illustrated in the Pake pattern[23] (Fig. 2.8, right). The Pake pattern represents the dipolar frequency pattern averaged over all angles from parallel (θ = 0˚) to perpendicular (θ = 90˚) relative the magnetic field. The frequency of separation between the prominent peaks of the Pake pattern is inversely proportional to the distance between the interacting spins at θ = 90˚. This relationship represents the theoretical basis of distance measurement in EPR.

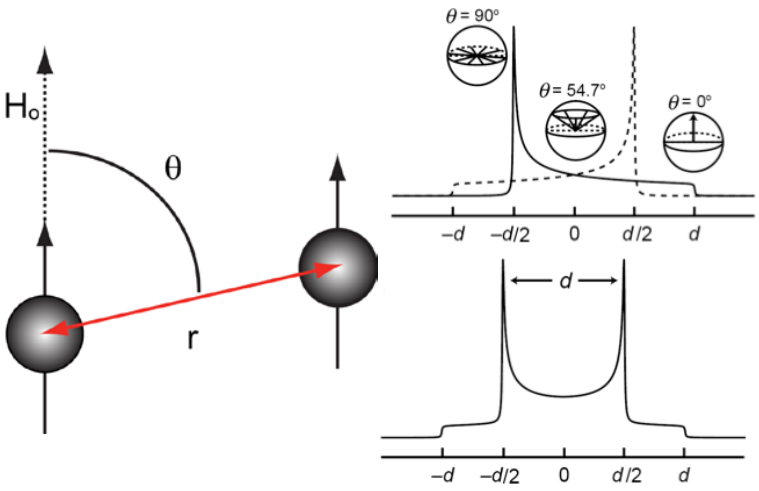


Figure 2.8. Orientation-dependence of the dipolar interaction and Pake patterns. Figure reproduced from tutorial materials associated with [24].

For interactions less than 20Å in distance, distance measurements can be achieved without relaxation measurements through broadening of the EPR line shape in continuous wave (CW) measurements like the mobility analyses described above (Fig. 2.9). Broadening effects

are distance-dependent with shorter interspin distances reporting increased broadening effects[19]. The interspin distance can be extracted from the EPR lineshape through convolution/deconvolution analyses that compare the sum of the singly labeled spectra to the doubly labeled spectra isolating the dipolar contribution[23] (Fig. 2.9). In these analyses, the Pake broadening function is modeled using variable distance values until it converges to fit the experimental doubly labeled spectra. The distance parameters resulting in the fit of the spectra are reported as a Gaussian distribution representing their relative probabilities.

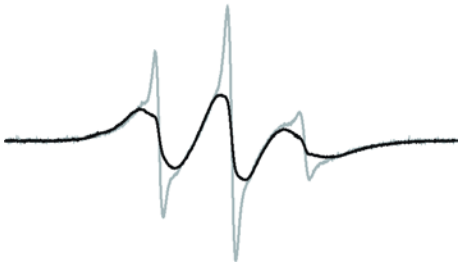


Figure 2.9. Spin-spin broadening. (Gray) The sum of single EPR spectra. (Black) The spin-spin broadened spectra. Figure reproduced with permission from [17].

The Double Electron-Electron Resonance (DEER) experiment

DEER measures long-range spin-spin distances

For distance greater than 20Å, broadening due to dipolar coupling can no longer be monitored due to limitations imposed by the intrinsic width of the EPR line shape. Pulsed EPR methods extend the measurable distance between two electron spins (Figure 1A) up to 60 Å, and in favorable cases to 80 Å, by separating the dipolar term in the spin Hamiltonian for exclusive detection[25-26]. The extended range of pulsed EPR distance measurements makes them more suitable for many molecular applications than the CW experiments described previously. Although appropriate pulse sequences have long been developed[27-28], the

widespread application of dipolar EPR spectroscopy was spurred by commercialization of high sensitivity pulsed EPR spectrometers and the model-free analysis of dipolar interactions to calculate the distance distribution between two electron spins[29-30]. Freed and coworkers[25] developed pulse sequences to detect double quantum coherence between spins which promises an order of magnitude in increased sensitivity. However, double electron-electron resonance (DEER), or pulsed electron double resonance (PELDOR), is the most commonly used method for distance measurements between spin labels[18].

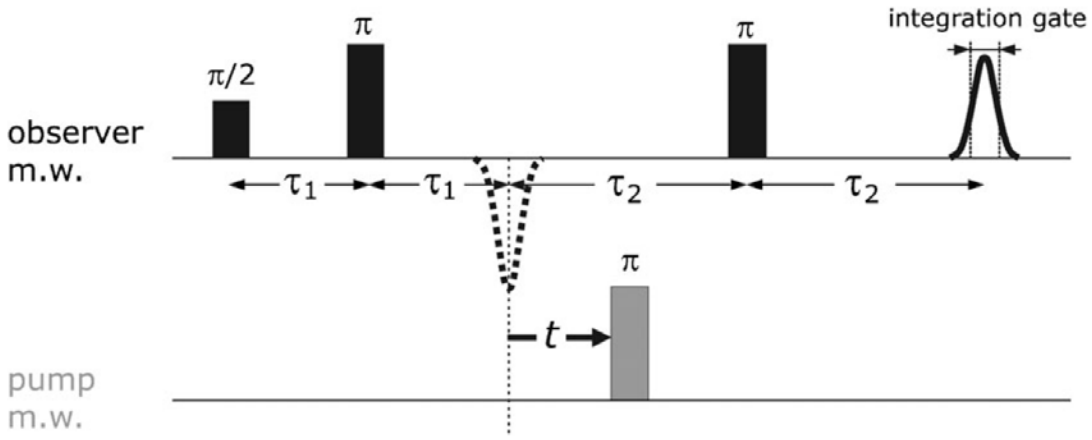


Figure 2.10. DEER pulse sequence and resulting spin echoes. Figure reproduced with permission from [26]

As described previously, dipolar coupling occurs when spins experience the dipolar magnetic field of other spins[11]. Dipolar coupling manifests in the EPR signal as spin-spin relaxation in the x-y plane due to the contribution of the dipolar interaction to the angular rate of precession of dipolar coupled spins as a function of the strength of the dipolar interaction which is distance-dependent[18]. Therefore, by monitoring the degree to which electron spin angular precession rates are affected by the dipolar interaction, distance distributions can be extracted.

The traditional DEER experiment isolates the dipolar contribution to the angular precession rate through the use of a four pulse sequence[27-28] that “observes” the angular

precession rates of spins resonating at one frequency, v_o, while spins resonating at another non-overlapping frequency, $v_{p,}$ are inverted with a “pump” pulse (Fig. 2.10). To begin a DEER experiment[18], the bulk magnetization of the observe spins is rotated into the x-y plane by a π/2 microwave pulse. Due to local field inhomogeneities including dipolar interactions, spins begin to dephase in the x-y plane. A π pulse is then applied to the observe spins, flipping them 180° which reverses the dephasing and results in a refocused Hahn spin echo. Refocusing pulses are a fundamental practical feature of pulsed EPR experiments, as they allow signal detection outside of instrument dead time[31]. To isolate the dipolar frequency contribution to the total magnetic field experienced by the observe spins, a π pulse is applied to the pump spins resulting in a rotation of the magnetization vector of the pump spins into the –z axis[18]. This pulse is applied at varying time points relative to the Hahn echo. The inversion of the pump spins reverses the net angular rate contribution of dipolar interaction from positive to negative. This creates a dipolar interaction-specific phase lag such that when a second π pulse is applied to the observe spins, dipolar coupled spins are not refocused and the dipolar effect manifests as a modulation of the integrated intensity of the spin echo[18] (Fig. 2.11). The degree of coherence loss is a function of the strength of the dipolar interaction and the time at which the pump π pulse was applied. The modulated refocused echo intensity is recorded for each of the varying pump π pulse time points over the course of several DEER experiments. This separates the dipolar contribution from the pump pulse time-dependence. As a function of the pump pulse delay, the echo intensity of the experimental DEER signal will oscillate at the frequency of the dipolar interaction[18].

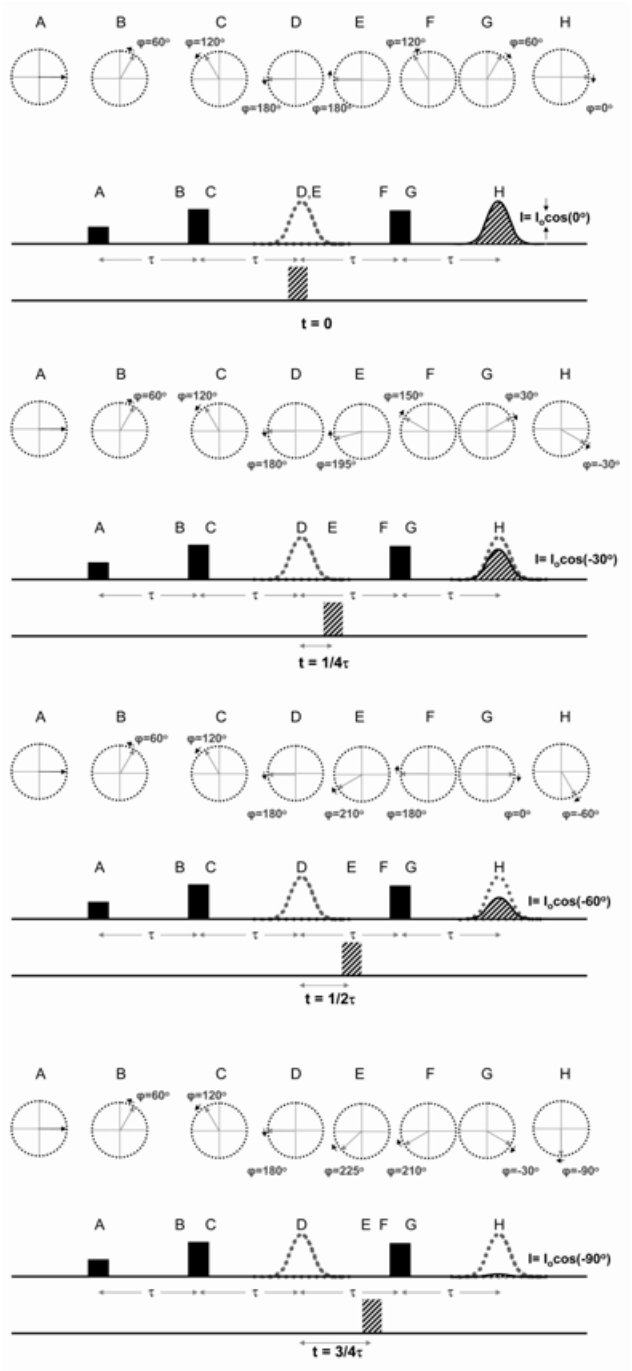


Figure 2.11. The DEER experiment monitors phase lag of dipolar coupled spins as an attenuation of spin echo intensity. Figure reproduced with permission from [18].

As described previously, the dipolar frequency is inversely proportion to the cube of the interspin distance[11]. Practically, this means shorter distances will result in faster oscillations of the DEER signal and longer distances will result in slower oscillations[18] (Fig. 2.12). Longer

distances require longer spin evolution times to define their oscillations. Therefore, the upper limit for distance measurements is defined by the phase memory time because once the magnetization in the x-y plane as completely dephased, spins cannot be refocused. The phase memory time is primarily a function of T_2 and can be increased by decreasing experimental temperatures. To accommodate this, pulsed EPR experiments are conducted at temperatures between 50 and 80K. The result of this practical necessity is that pulsed EPR experiments represent solid state measurements and static descriptions of the conformational ensemble. It is noted that the solid state nature of the measurement implies that distance distributions contain contributions from all protein dynamic modes regardless of their time scales. At cryogenic temperature, the limit for accurate measurement of distance distributions is typically reached between 60 and 80Å.

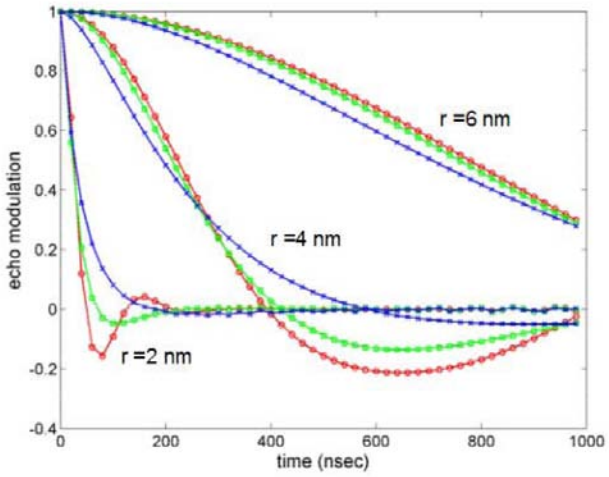


Figure 2.12. DEER spectra at different spin-spin distances (r) and Gaussian distribution widths (Δr = 0.5nm (red), 2nm (green), and 4nm (blue)). Figure reproduced with permission from [18]

DEER distributions reflect all distance components within the ensemble

The distance relationship between spin labels will be reflected in the signal coherence of the DEER spectra (Fig. 2.12). If a distance relationship is narrowly defined, meaning that the

inter-spin distance is at or near a single distance value, the signal coherence will be high and the spectra will reflect a single oscillation frequency of maximal amplitude. The extent to which spin labels possess different distance relationships within the molecular ensemble is reflected in the DEER signal as a composite of multiple frequencies each with oscillation amplitude related to the relative population of the associated distance component within the ensemble. A broad, continuous distance distribution between spins results in the featureless DEER signal as the various frequency components destructively interfere with one another and cancel. In a doubly labeled protein system, the DEER signal is composed of both intramolecular and intermolecular dipolar interactions. The intermolecular interactions represent randomly oriented intermolecular relationships and therefore manifest as a featureless exponential decay within the DEER signal. This component is subtracted out as a background contribution leaving the intramolecular oscillating decay for distance analysis.

DEER distance analysis

To quantitatively extract distance information from DEER spectra, Fourier transformation analysis can be used to derive the frequency domain and obtain the corresponding Pake pattern[23], as described previously. However, it is more common to directly fit the DEER spectra using the parameterized Gaussian model or Tikhonov regularization. Tikhonov regularization[29] simulates the time-domain data to fit the experimental data by identifying a regularization parameter that best satisfies the need for a unique solution without over fitting the data. The ability to identify a reasonable regularization parameter is related to the signal to noise ratio. Insufficient signal to noise may result in selection of regularization parameters that introduce artificial peaks in the distance distribution. Independent of the distance analysis methodology, the resulting distance distribution will reflect the probability of a distance between the two spins characterized by the weighted average distance, r_{av}, and a standard deviation, σ. These values can be used to describe molecular features such as protein structure, equilibrium fluctuation

dynamics, and ligand-triggered conformational changes.

DEER distributions are composed of spin label and protein dynamics

Equilibrium fluctuation dynamics refers to thermally driven protein motion occurring on multiple time scales with different amplitudes[8]. It reflects protein excursions between local energy minima and is manifested by dynamic modes of side chain isomerization on the ps-ns scale, ns excursion of flexible loops, and all the way to movement of secondary structures or domains in the μs-s regime. In the solid state, this conformational sampling results in static disorder, provided the freezing process does not trap fluctuating structural elements in a single energy minimum.

Identifying the contribution of spin label motion to the distribution width

In the absence of protein fluctuations, the intrinsic width of the distance distribution arises from the flexibility of the spin label side chain. The most commonly used spin label, MTSSL, allows rotations around four internal bonds linking the nitroxide ring to the protein backbone[16] (Fig. 2.2). Crystal structures of spin labeled T4 lysozyme defined a subset of spin label rotamers some of which are resolved to the nitroxide rings[32]. Transition between these rotamers can change the distance between the labels thus contributing to the width of the distance distribution. The structures also reveal the potential for direct interaction between the ring and neighboring side chain and main chain atoms potentially biasing the rotamer population and making prediction of the intrinsic width more complex.

Therefore, interpretation of the width of the distance distribution requires untangling the intrinsic contribution from that of protein dynamics. In principle, it is possible to use atomistic molecular dynamics (MD) simulations to sample the rotamer distribution for each label in a pair and obtain the distribution width in the absence of protein dynamics. Despite a number of successful reports[33-35], long computation times, particularly when considered for multiple label

pairs, in conjunction with potential imprecision in spin label parameterization hinder routine application of this approach. Moreover, trajectories as long as 100 ns may not be sufficient to efficiently sample the spin label rotamer space [36].

To overcome this problem, Jeschke and co-workers[37] created a spin label rotamer library from a long MD trajectory thereby circumventing the need for repeated MD simulations for each pair. Using a computational modeling program called MMM, the rotamers are inserted into the protein of interest at experimental spin labeling sites and are evaluated for their relative energies calculated from a modified Lennard-Jones potential. A simulated distance distribution is thus generated from the pair-wise distances between rotamers weighted by their relative population. This approach has been successfully applied to determine the dimer arrangement of the Na^+/H^+ exchanger[38] and a transmembrane segment in the proline symporter PutP[39]. However, extensive benchmarking is needed to assess whether the rotamer library provides a complete representation of the spin label conformational space. Using a similar approach, including crystallographically sampled rotamers with MD simulated rotamers, a MTSSL rotamer library was recently introduced into the protein modeling program Rosetta[40]. This program, RosettaEPR, allows full-atom spin label modeling within the Rosetta suite of protein modeling tools including EPR data-driven computational structure prediction (see below).

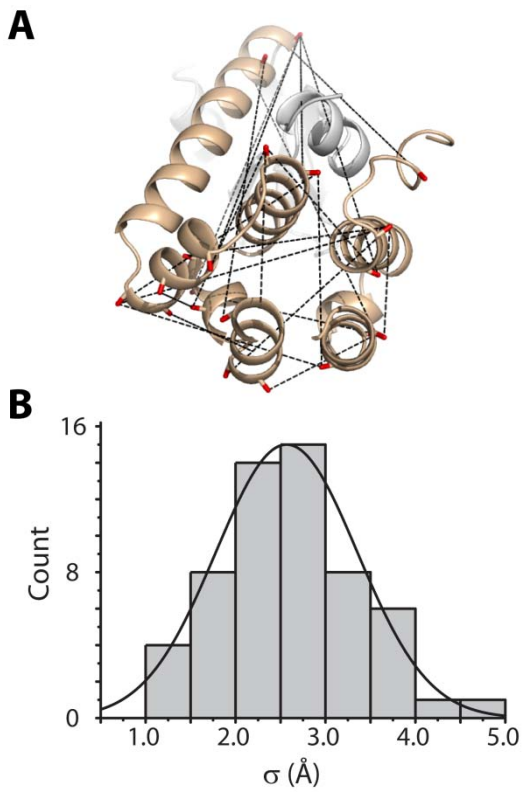


Figure 2.13. Empirically-determined intrinsic width of distance distributions. (A) Structure of T4L highlighting representative pairs used for distance measurements between spin labels. (B) Sigma (σ) values calculated from experimental distance distributions from T4L are shown as a histogram binned at intervals of 0.5Å.

In addition to these approaches, we have adopted an empirical approach to obtain an estimate of the intrinsic distribution width (Fig. 2.13). For this purpose, ~60 pairs of spin labels were introduced at surface sites in T4 lysozyme (T4L) focusing on the helical C-terminal domain and avoiding regions of the protein affected by the hinge bending motion in solution (data from Chapter 3 and unpublished data). Each of the resulting distance distributions was parameterized by r_{av} and σ. The histogram displays the frequency of a given standard deviation binned every 0.5 Å. Although the sites were selected to be solvent-exposed, the distance distributions are generally narrow, consistent with previous models of limited-amplitude motion of the MTSSL spin label at such sites[16,41]. Ideally, a similar benchmarking exercise would establish the intrinsic distribution width for a membrane protein model system. However, we

note that the spin label mobility is not altered at lipid-facing exposed sites strongly suggesting that the intrinsic distribution width for membrane proteins will have a similar shape to Figure 2.13[42-43].

DEER readout of triggered conformational changes

Membrane protein functional cycles require the interconversion between distinct conformations or shifts in preexisting conformational equilibria. Typically, segments of the protein undergo defined motions in response to energy input such as changes in transmembrane voltage, binding or hydrolysis of ATP or binding of ligand or substrates. To the extent that these movements result in changes in residue environment, they also alter the mobility and accessibility of spin labels. In most cases however, these parameters cannot be quantitatively interpreted to reveal the nature and magnitude of the underlying structural changes.

In contrast, protein motions are directly manifested by changes in the average distance and/or the shape and width of the distribution (Fig. 2.14). The former reports the amplitude of movement between two most probable conformations of the protein while the latter reflects changes in the underlying conformational ensemble as illustrated below. The simplest interpretation of the DEER data in terms of structural changes requires that the set of spin label rotamers remains unchanged between the different protein conformations. Repacking of the label can lead to changes in r_{av} and/or affect the width of the distance distribution. In general, judicious selection of unconstrained, exposed sites for spin labeling circumvents this confounding factor. The room temperature spin label mobility can be used to confirm the lack of spin label repacking as a result of conformational changes. DEER distance measurements have been extensively used to describe the molecular motions of a number of membrane protein systems including the ATP-binding cassette transporter MsbA[43-45], the Major Facilitator Superfamily transporter LacY[46-48], the G protein coupled receptors Rhodopsin[49-50], and the

potassium channel KcsA[51-52]. These experiments served to successfully evaluate crystallographic representations of protein conformational sampling and define aspects of dynamic protein mechanisms. Chapters 4 and 5 will extensively utilize measurement of conformational dynamics by DEER in membrane proteins, LeuT and Mhp1.

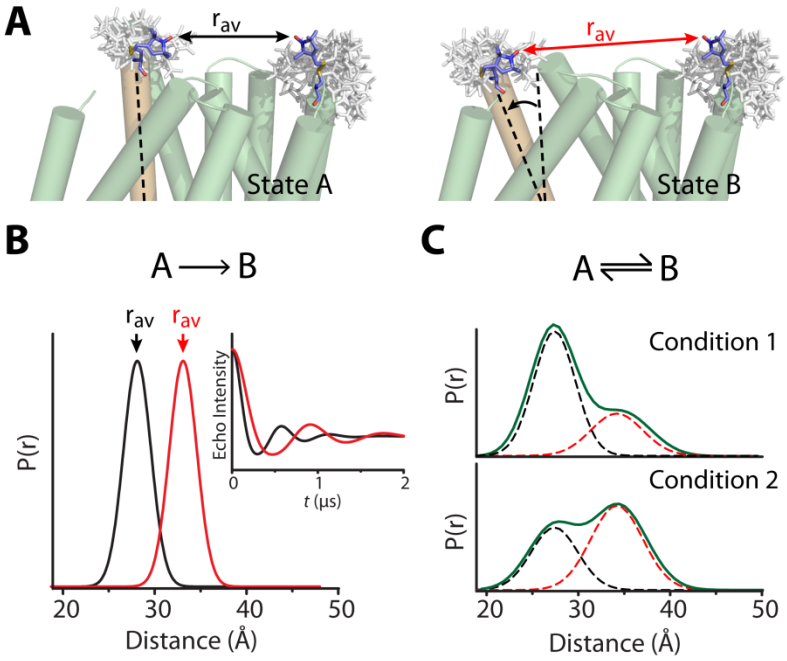


Figure 2.14. DEER detection of triggered conformational changes. (A) Hypothetical motion of a transmembrane helix (orange) during the transition from State A to State B alters the average distance (r_{av}, arrows) between spin labels. (B) If states A and B are distinct conformers of different energies, the conformational shift will manifest primarily as a change in r_{av}, evident as an increased period of the spin echo decay (inset). (C) If states A and B represent two conformations present in equilibrium, altering the biochemical conditions will alter the contribution of each distinct conformation (dashed curves) to the distance distribution (green curve).

Modeling protein structure and conformational dynamics from EPR restraints

Development of DEER as a method of structure determination and evaluation

Despite the success of spin labeling in identifying and mapping conformational changes, evidenced by work highlighted above, transformation of EPR distances between spin labels to corresponding restraints between C_α carbons is remains challenging. For spin labeling EPR to

become a platform for discovery, computational methods for structural and dynamic interpretation of EPR parameters need to be developed. As extrinsic probes, spin labels shape the methodology and interpretation of EPR in fundamental ways. Not only is there the potential for functional and structural perturbation but the spin label linking arm introduces intrinsic uncertainties to models constructed from EPR restraints. In contrast to the determination of EPR parameters, which is firmly established in rigorous treatment of the spin Hamiltonian, structural interpretation of the data necessitates a model of the spin label relative to the backbone, an internally consistent transfer function that links spectral and structural parameters. An additional consequence of using reporter groups is the sparseness of EPR data sets. Limited by experimental throughput, the number of EPR restraints per residue is typically many fewer than that used in NMR structure determination. Importantly, as discussed below, in the absence of a crystal structure, the restraints are not necessarily optimal or of uniform value for modeling structure and dynamics.

EPR restraints are capable of modeling protein structure to high resolution

Rosetta is a knowledge-based, *de novo* protein structure prediction program shown to be among the most effective algorithms available[53]. Recently, Rosetta was extended to model membrane proteins[54]. Similar to Rosetta, Rosetta-Membrane structures are built with the fragment assembly method. However, the energy function and conformational sampling algorithm have been modified to reflect the membrane environment. For proteins under 150 residues, Rosetta-membrane has consistently produced models under 4Å RMSD to the crystal structure[55]. Rosetta has also been used to model proteins on the order of LeuT, such as rhodopsin (7 helices, 278 residues) to an RMSD of 9.2 Å and H^+/Cl^- exchange transporter (7 helices, 203 residues) to 12.4 Å RMSD[54]. Furthermore, advances in the Rosetta membrane protocol have allowed 4 Å RMSD of native structure models of membrane proteins (up to 300 residues) to be achieved with a single restraint[56]. Although a benchmarked strategy for EPR-

based modeling of structures is not yet available, NMR-based approaches[57-59] serve to reinforce potential development in this area.

The more general question of whether EPR restraints restrict conformational space to a set of convergent models of acceptable resolution or enable detailed description of structural rearrangements starting from a high resolution structure has only been recently addressed. Alexander *et al.*[60] carried out a systematic feasibility analysis of *de novo* protein structure determination from EPR restraints in T4L. This study also aimed to directly define the information content of EPR restraints and the impact of the sparse data on model quality. The distance between spin labels was converted into a distance range between β-carbons using a simple "motion on a cone" model, treating the spin label as an average vector relative to the β-carbon. Because of an assumed isotropic distribution of the label in this model, the function relating the distance between the two spin labels to that between the corresponding β-carbon was relatively broad, *i.e.* the derived restraint has large uncertainty.

This study made two novel contributions. First, it heralded the use of the Rosetta folding algorithm[53] as an alternative computation platform to MD simulations. Second, it demonstrated that a detailed model of the spin label conformations at each site may not be required. Even with a simple boundary function to interpret the restraints, twenty-five EPR restraints were sufficient to generate models with the correct fold. Subsequent high resolution refinement yielded structures that are within 1 Å RMSD from the crystal structure. This remarkable outcome was rationalized by the robust Rosetta knowledge-based energy function, which captures the principle of protein assembly encoded in known structures, compensating for the sparseness of EPR restraints. In turn, the EPR restraints efficiently restrict conformational space enabling Rosetta to find the global energy minimum.

The RosettaEPR approach to EPR-based structure determination

The ultimate goal of the RosettaEPR project[40,61] is to establish a suite of algorithms that guide experimentalists in the selection of labeling sites and provide a platform for structural interpretation of the data. An algorithm for the optimal selection of EPR restraints is described in Chapter 3. A basic experimental throughput for this EPR-based computational structure determination approach (Fig. 2.15) would include optimized restraint selection based on primarily predicted secondary structural information or available experimentally-determined definitions. Distance measurements would be conducted for a minimum set of spin labeled mutants. Distance measurements would be converted to probabilistic representation of the C_α distances either directly within the computational modeling software or as an independent mathematical transformation. Models of protein structure would be generated using the Rosetta protein structure modeling algorithm. High scoring models would be selected for both their relative free energy and their agreement with EPR restraints. These models would go through a series of all-atom structural refinements to yield a model or ensemble of models that simultaneously reflect the information provided by both EPR and the knowledge-base of the PDB. Upon extensive benchmarking, such an approach could represent a methodology for atomic resolution structure determination from EPR distance measurements. Furthermore, it presents a novel approach for modeling conformational intermediate states without the caveats associated with crystallographic conformational selection.

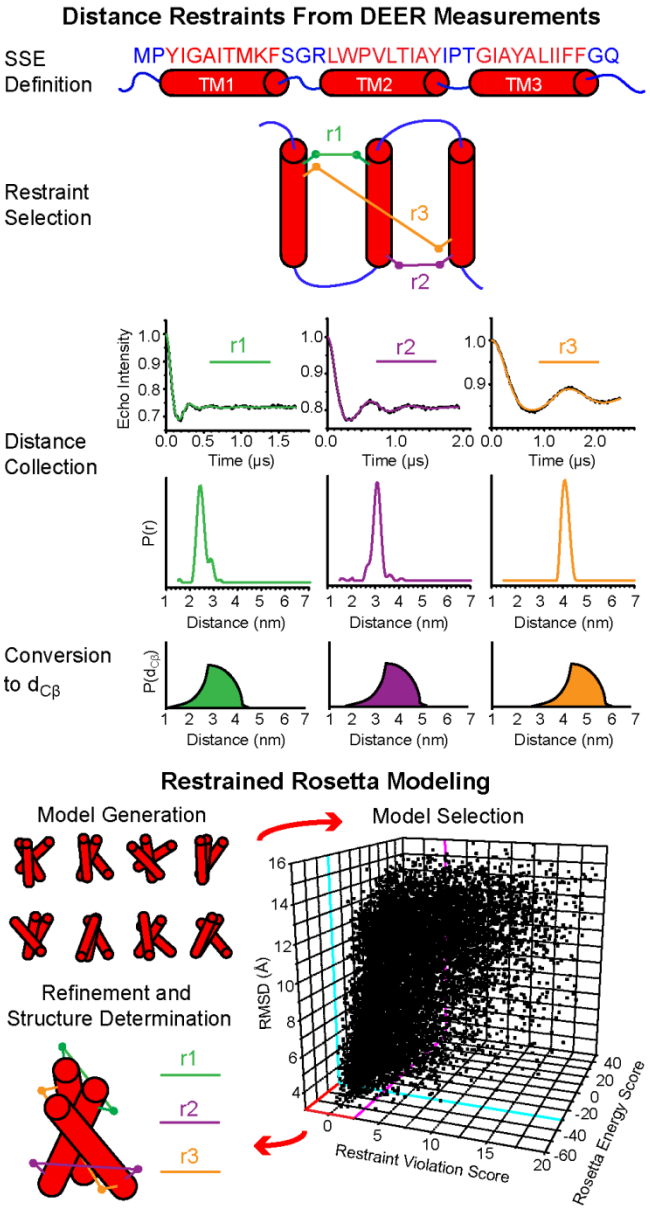


Figure. 2.15. Structure determination by EPR and Rosetta. Overview of hypothetical *de novo* modeling of a polytopic membrane protein guided by EPR restraints. Three restraints are highlighted for simplicity but a larger number is required even for a small 3-helix protein. In this scheme, secondary structural element (SSE) definitions inform optimized selection of label pairs for restraints. Analysis of DEER measurements returns distance distributions, which are transformed into probabilistic boundary functions to describe the distance between β-carbons (dC_β) of the label pairs. Restraint violation scores measure model agreement with these functions and guide Monte Carlo modeling trajectories. Selecting for models with both low energy and low restraint violations have been shown to effectively limit model pools to low RMSD models (as shown in the 3D plot). These models proceed to all-atom, high resolution refinement with explicit modeling of the restraints, resulting in a best model.

References

1. Bowie, J. U. Stabilizing membrane proteins. *Curr Opin Struct Biol* **11**, 397-402 (2001).
2. Gouaux, E. & Mackinnon, R. Principles of selective ion transport in channels and pumps. *Science* **310**, 1461-1465 (2005).
3. Kobilka, B. & Schertler, G. F. New g-protein-coupled receptor crystal structures: Insights and limitations. *Trends Pharmacol Sci* **29**, 79-83 (2008).
4. Ha, Y. Structural principles of intramembrane proteases. *Curr Opin Struct Biol* **17**, 405-411 (2007).
5. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347-355 (2009).
6. Rees, D. C., Johnson, E. & Lewinson, O. Abc transporters: The power to change. *Nat Rev Mol Cell Biol* **10**, 218-227 (2009).
7. Cross, T. A., Sharma, M., Yi, M. & Zhou, H. X. Influence of solubilizing environments on membrane protein structures. *Trends Biochem Sci* **36**, 117-125 (2011).
8. Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using nmr. *Trends Biochem Sci* **34**, 601-611 (2009).
9. Ogawa, S. & McConnell, H. M. Spin-label study of hemoglobin conformations in solution. *Proc Natl Acad Sci U S A* **58**, 19-26 (1967).
10. Hubbell, W. L., Mchaourab, H. S., Altenbach, C. & Lietzow, M. A. Watching proteins move using site-directed spin labeling. *Structure* **4**, 779-783 (1996).
11. Weil, J. A. & Bolton, J. R. *Electron paramagnetic resonance : Elementary theory and practical applications*. 2nd edn, (Wiley-Interscience, 2007).
12. Carrington, A. & McLachlan, A. D. *Introduction to magnetic resonance : With applications to chemistry and chemical physics*. (Chapman and Hall ; Wiley, 1979).
13. Calvo, R., Passeggi, M. C., Isaacson, R. A., Okamura, M. Y. & Feher, G. Electron paramagnetic resonance investigation of photosynthetic reaction centers from rhodobacter sphaeroides r-26 in which fe2+ was replaced by cu2+. Determination of hyperfine interactions and exchange and dipole-dipole interactions between cu2+ and qa. *Biophys J* **58**, 149-165 (1990).
14. Rifkind, J. M., Ramasamy, S., Manoharan, P. T., Nagababu, E. & Mohanty, J. G. Redox reactions of hemoglobin. *Antioxid Redox Signal* **6**, 657-666 (2004).
15. Klare, J. P. & Steinhoff, H. J. Spin labeling epr. *Photosynth Res* **102**, 377-390 (2009).
16. Mchaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of spin-labeled side chains in t4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692-7704 (1996).
17. Klug, C. S. & Feix, J. B. Methods and applications of site-directed spin labeling epr spectroscopy. *Methods Cell Biol* **84**, 617-658 (2008).
18. Fajer, P., Brown, L. & Song, L. in *Esr spectroscopy in membrane biophysics* Vol. 27 *Biological magnetic resonance* Ch. 4, 95-128 (Springer US, 2007).
19. Hubbell, W. L., Cafiso, D. S. & Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nature Structural Biology.* **7**, 735-739 (2000).
20. Altenbach, C., Froncisz, W., Hemker, R., Mchaourab, H. & Hubbell, W. L. Accessibility of nitroxide side chains: Absolute heisenberg exchange rates from power saturation epr. *Biophys J* **89**, 2103-2112 (2005).
21. Altenbach, C., Greenhalgh, D. A., Khorana, H. G. & Hubbell, W. L. A collision gradient method to determine the immersion depth of nitroxides in lipid bilayers: Application to spin-labeled mutants of bacteriorhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1667-1671 (1994).

22. Hubbell, W. L., Gross, A., Langen, R. & Lietzow, M. A. Recent advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology* **8**, 649-656 (1998).
23. Rabenstein, M. D. & Shin, Y. K. Determination of the distance between two spin labels attached to a macromolecule. *Proceedings of the National Academy of Sciences of the United States of America.* **92**, 8239-8243 (1995).
24. Boldt, K. Simulating spin dynamics in nmr with a new computer program intended for education: Insensitive. *Concepts in Magnetic Resonance Part A* **38A**, 17-24 (2011).
25. Borbat, P. P., Mchaourab, H. S. & Freed, J. H. Protein structure determination using long-distance constraints from double-quantum coherence esr: Study of t4 lysozyme. *J Am Chem Soc* **124**, 5304-5314 (2002).
26. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **9**, 1895-1910 (2007).
27. Milov, A. D., Ponomarev, A. B. & Tsvetkov, Y. D. Electron electron double-resonance in electron-spin echo - model biradical systems and the sensitized photolysis of decalin. *Chemical Physics Letters* **110**, 67-72 (1984).
28. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *J Magn Reson* **142**, 331-340 (2000).
29. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed esr using tikhonov regularization. *Journal of Magnetic Resonance* **172**, 279-295 (2005).
30. Jeschke, G., Wegener, C., Nietschke, M., Jung, H. & Steinhoff, H. J. Interresidual distance determination by four-pulse double electron-electron resonance in an integral membrane protein: The na+/proline transporter putp of escherichia coli. *Biophysical Journal* **86**, 2551-2557 (2004).
31. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *Journal of Magnetic Resonance.* **142**, 331-340 (2000).
32. Langen, R., Oh, K. J., Cascio, D. & Hubbell, W. L. Crystal structures of spin labeled t4 lysozyme mutants: Implications for the interpretation of epr spectra in terms of structure. *Biochemistry.* **39**, 8396-8405 (2000).
33. Sompornpisut, P., Roux, B. & Perozo, E. Structural refinement of membrane proteins by restrained molecular dynamics and solvent accessibility data. *Biophys J* **95**, 5349-5361 (2008).
34. Fajer, M. I., Li, H., Yang, W. & Fajer, P. G. Mapping electron paramagnetic resonance spin label conformations by the simulated scaling method. *Journal of the American Chemical Society* **129**, 13840-13846 (2007).
35. Borovykh, I. V., Ceola, S., Gajula, P., Gast, P., Steinhoff, H. J. & Huber, M. Distance between a native cofactor and a spin label in the reaction centre of rhodobacter sphaeroides by a two-frequency pulsed electron paramagnetic resonance method and molecular dynamics simulations. *J Magn Reson* **180**, 178-185 (2006).
36. Sezer, D., Freed, J. H. & Roux, B. Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyalanine alpha-helix. *J Phys Chem B* **112**, 5755-5767 (2008).
37. Polyhach, Y., Bordignon, E. & Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys Chem Chem Phys* **13**, 2356-2366 (2011).
38. Hilger, D., Polyhach, Y., Padan, E., Jung, H. & Jeschke, G. High-resolution structure of a na+/h+ antiporter dimer obtained by pulsed electron paramagnetic resonance distance measurements. *Biophys J* **93**, 3675-3683 (2007).

39. Hilger, D., Polyhach, Y., Jung, H. & Jeschke, G. Backbone structure of transmembrane domain ix of the na+/proline transporter putp of escherichia coli. *Biophys J* **96**, 217-225 (2009).
40. Alexander, N. S., Stein, R. A., Koteiche, H. A., Kaufmann, K. W., Mchaourab, H. S. & Meiler, J. Rosettaepr: Rotamer library for spin label structure and dynamics. *PLoS One* **8**, e72851 (2013).
41. Columbus, L., Kalai, T., Jeko, J., Hideg, K. & Hubbell, W. L. Molecular motion of spin labeled side chains in alpha-helices: Analysis by variation of side chain structure. *Biochemistry* **40**, 3828-3846 (2001).
42. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: Spin labeling of bacteriorhodopsin mutants. *Science* **248**, 1088-1092 (1990).
43. Dong, J., Yang, G. & Mchaourab, H. S. Structural basis of energy transduction in the transport cycle of msba. *Science* **308**, 1023-1028 (2005).
44. Zou, P. & Mchaourab, H. S. Alternating access of the putative substrate-binding chamber in the abc transporter msba. *J Mol Biol* **393**, 574-585 (2009).
45. Zou, P., Bortolus, M. & Mchaourab, H. S. Conformational cycle of the abc transporter msba in liposomes: Detailed analysis using double electron-electron resonance spectroscopy. *J Mol Biol* **393**, 586-597 (2009).
46. Ujwal, M. L., Jung, H., Bibi, E., Manoil, C., Altenbach, C., Hubbell, W. L. & Kaback, H. R. Membrane topology of helices vii and xi in the lactose permease of escherichia coli studied by lacy-phoa fusion analysis and site-directed spectroscopy. *Biochemistry* **34**, 14909-14917 (1995).
47. Wu, J., Voss, J., Hubbell, W. L. & Kaback, H. R. Site-directed spin labeling and chemical crosslinking demonstrate that helix v is close to helices vii and viii in the lactose permease of escherichia coli. *Proc Natl Acad Sci U S A* **93**, 10123-10127 (1996).
48. Voss, J., Wu, J., Hubbell, W. L., Jacques, V., Meares, C. F. & Kaback, H. R. Helix packing in the lactose permease of escherichia coli: Distances between site-directed nitroxides and a lanthanide. *Biochemistry* **40**, 3184-3188 (2001).
49. Hubbell, W. L., Altenbach, C., Hubbell, C. M. & Khorana, H. G. Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Adv Protein Chem* **63**, 243-290 (2003).
50. Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P. & Hubbell, W. L. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc Natl Acad Sci U S A* **105**, 7439-7444 (2008).
51. Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying k+-channel activation gating. *Science* **285**, 73-78 (1999).
52. Cordero-Morales, J. F., Jogini, V., Lewis, A., Vasquez, V., Cortes, D. M., Roux, B. & Perozo, E. Molecular driving forces determining potassium channel slow inactivation. *Nat Struct Mol Biol* **14**, 1062-1069 (2007).
53. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**, 363-382 (2008).
54. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using rosetta. *Proteins* **62**, 1010-1025 (2006).
55. Barth, P., Schonbrun, J. & Baker, D. Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A* **104**, 15682-15687 (2007).
56. Barth, P., Wallner, B. & Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci U S A* **106**, 1409-1414 (2009).
57. Rohl, C. A. Protein structure estimation from minimal restraints using rosetta. *Methods Enzymol* **394**, 244-260 (2005).

58. Weiner, B. E., Alexander, N., Akin, L. R., Woetzel, N., Karakas, M. & Meiler, J. Bcl::Fold-protein topology determination from limited nmr restraints. *Proteins* (2013).
59. Ganguly, S., Weiner, B. E. & Meiler, J. Membrane protein structure determination using paramagnetic tags. *Structure* **19**, 441-443 (2011).
60. Alexander, N., Bortolus, M., Al-Mestarihi, A., Mchaourab, H. & Meiler, J. De novo high-resolution protein structure determination from sparse spin-labeling epr data. *Structure* **16**, 181-195 (2008).
61. Hirst, S. J., Alexander, N., Mchaourab, H. S. & Meiler, J. Rosettaepr: An integrated tool for protein structure determination from sparse epr data. *J Struct Biol* **173**, 506-514 (2011).

CHAPTER 3

SELECTION OF SPIN LABELING SITES FOR COMPUTATIONAL STRUCTURE DETERMINATION

Abstract

A hybrid protein structure determination approach combining sparse Electron Paramagnetic Resonance (EPR) distance restraints and Rosetta *de novo* protein folding has been previously demonstrated to yield high quality models[1]. However, widespread application of this methodology to proteins of unknown structures is hindered by the lack of a general strategy to place spin label pairs in the primary sequence. In this work, we report the development of an algorithm that optimally selects spin labeling positions for the purpose of distance measurements by EPR. For the α-helical subdomain of T4 lysozyme (T4L), simulated restraints that maximize sequence separation between the two spin labels while simultaneously ensuring pairwise connectivity of secondary structure elements yielded vastly improved models by Rosetta folding. 50% of all these models have the correct fold compared to only 21% and 8% correctly folded models when randomly placed restraints or no restraints are used, respectively. Moreover, the improvements in model quality require a limited number of optimized restraints, the number of which is determined by the pairwise connectivities of T4L α-helices. The predicted improvement in Rosetta model quality was verified by experimental determination of distances between spin labels pairs selected by the algorithm. Overall, our results reinforce the rationale for the combined use of sparse EPR distance restraints and *de novo* folding. By alleviating the experimental bottleneck associated with restraint selection, this algorithm sets the stage for extending computational structure determination to larger, traditionally elusive protein topologies of critical structural and biochemical importance.

Introduction

Decades into the structural biology revolution, tens of thousands of structures have been deposited in the Protein Data Bank (PDB) cataloging protein folds, defining motifs of catalysis, and revealing architectures of protein complexes. The overarching goal of delineating the biochemical and physiological circuitry that interconnect to form cells and organisms requires further progress on two fronts. The sampling of structure space has been uneven; primarily skewed towards classes of proteins amenable to analysis by the leading structural methods. Undersampled protein structure space includes proteins of high functional and pharmacological significance such as multispan membrane proteins[2] and large, conformationally heterogeneous soluble proteins[3]. In addition, protein function often involves the transitions between conformational states or shifts in the equilibrium between such states. Static crystallographic snapshots represent a limited and sometimes biased view of the conformational space of dynamic proteins. Structures trapped in the confines of the crystal lattice may not be defined mechanistically or may be distorted by non-native environments such as detergent solubilization or osmotically active molecules [4].

These two challenges motivated the development of both theoretical and experimental methods to accelerate the speed of structure determination and to describe protein dynamic dimensions. EPR spectroscopy in conjunction with site-directed spin labeling (SDSL)[5-6] has been extensively applied to map conformational changes in soluble[7-8] and membrane proteins[9-17] and to probe the structure of dynamic oligomers[18-19] and amyloids[20-21]. Combining residue-specific measures of solvent accessibility and local dynamics with global geometric distance restraints describing packing of secondary structures and domains, this approach provides enough restraints for modeling protein structures and their rearrangements[22-25]. High sensitivity, absence of size limits and restriction on environment and/or solvent enables the evaluation of

crystallographic structures and comparative models under native-like, well defined biochemical conditions.

However this approach is intrinsically limited by the need for incorporation of spin labels into protein sequences. Compared to other restraint-based approaches such as Nuclear Magnetic Resonance (NMR) Spectroscopy, this reduces the experimental throughput effectively reducing the practical number of obtainable restraints. Moreover, the linking arm of the spin label tethering it to the protein introduces uncertainty in the interpretation of EPR parameters in terms of backbone structure. In the case of distance measurements, the translation of a precisely measured distance between spin labels to a restraint between corresponding β-carbons (C_β) is model dependent. Models derived from molecular dynamic simulations[26-29], crystallographic rotamer libraries[10], or based on simple geometric considerations[1] have been used to rationalize the experimental EPR distances.

A general approach for protein structure determination from EPR restraints was developed by Alexander et al.[1]. It capitalizes on the *de novo* protein structure prediction algorithm, Rosetta[30-38], to overcome the sparseness of EPR experimental restraints. The premise of this work was that restriction of conformational space by the EPR restraints increases Rosetta's efficiency in finding native folds. That a limited number of distances between pairs of spin labels significantly improved the quality of models put to rest concerns regarding the value of EPR distances as restraints for modeling. Experimental EPR distances were translated into C_β-C_β restraints using a simple cone model with virtually no restriction of spin label rotameric states.

The limited throughput of EPR methods and the ensuing restraint sparseness encourages a rational approach in the selection of spin labeled sites. Alexander et al.[1] demonstrated the importance of high information content (defined as the ratio between sequence separation and Euclidean distance) as a criterion for restraint quality. The improvement in model quality was attributed to a third of the restraints with the highest

information content. However, for proteins of unknown structures where the Euclidian distance is not known, using the numerator (i.e. sequence separation) as a proxy for information content will cluster restraints between the ends of the primary sequence.

This paper reports the development and experimental application of a general algorithm for selection of optimized distance restraint patterns for protein structure determination. Starting from sequence information, an iterative computational approach validated by Rosetta *de novo* folding yielded the best scoring scheme for restraint selection. Using the α-helical domain of T4 lysozyme (T4L) as a model system[39], we demonstrate that restraints selected to simultaneously optimize sequence separation and pairwise connectivity of secondary structures led to high quality models. To test the robustness of the algorithm, distances were experimentally measured between pairs of spin labels at residue positions selected by the algorithm. Rosetta folding using these distances yielded high quality models as predicted.

Methods and Materials

Algorithm Development

Input parameters of secondary structure and solvent exposure predictions of the C-terminal 107 amino acids of T4L were obtained using psipred[40] and NetSurfP[41] analyses, respectively. The ideal secondary structure definitions were obtained directly from the crystal structure of T4L (PDB ID: 2LZM). The ideal solvent exposure definitions were generated from the T4L crystal structure (2LZM) using a Rosetta neighbor count protocol. A neighbor count threshold of smaller than or equal to 9 defines solvent exposed residues[42].

The Monte Carlo protocol is initiated with a random distribution of spin label pairs that yield a total score for the distribution terms being tested. Each iteration of the Monte Carlo optimization involves random reassignment of label positions for a single pair. New label positions that improve or equal the best previous score are accepted. A typical optimization

included 10,000 iteration steps and 10 optimization trajectories after which scores converged. Restraint patterns were generated on local clusters using a perl script.

Sequence Separation Term

The Sequence Separation score (S_{SS}) is calculated by taking the natural log of the number of amino acids separating the two spin labels in each restraint pair (d_i), averaging over all restraint pairs (r), and normalizing to the natural log of the sequence length (g) to yield a value between 0 and 1.

$$S_{SS} = (\sum_{i=1}^{r} \ln d_i)/(r \times \ln g),$$

Thus, the sequence separation term effectively applies a penalty function for pairs separated by a small number of amino acids. This penalty logarithmically decreases with increased label separation. The logarithmic scaling is a modification of the original information content measure[1]. We found that the improvement in model quality measures becomes less dependent on sequence separation as d_i increases (data not shown).

Secondary Structure Term

The Secondary Structure term distributes the spin labels evenly among the secondary structural elements (SSE). First, an ideal number of spin labels per SSE (Q) is calculated by dividing the number of spin labels (l) (twice the number of restraints) by the number of SSEs (s). We define $Q' = div(l,s)$ and $Q'' = Q' + 1$. Note that the floor Q' and ceiling Q'' are acceptable integer values for Q. Further, we define remainder of l/s as $R = mod(l,s)$. An optimal spin label distribution will have Q'' labels in R SSEs, and Q' labels in all the others.

The Secondary Structure score (S_{SSE}) has two equally weighted components, $S_{SSE(L)}$ and $S_{SSE(S)}$. The first component, $S_{SSE(L)}$, is the average percentage of labels positioned in each SSE up to the ideal value, Q". Thus,

$$S_{SSE(L)} = \frac{1}{l}\sum_{i=1}^{s} \min\,(l_i\,, Q")\ ,$$

where l_i = number of labels in the i[th] SSE. As defined, this component favors placement of labels into SSE during the optimization trajectory. The second component of the score, $S_{SSE(S)}$, is derived from the fraction of SSEs that contain exactly the ideal number of spin labels:

$$S_{SSE(S)} = \frac{1}{s}\{min[E', (s-R)] + min[E", R]\},$$

where E' is the number of SSEs with Q' labels and $E"$ is the number of SSEs with $Q"$ labels. While $S_{SSE(L)}$ determines progress in achieving an optimal spin label placement during the Monte Carlo optimization, $S_{SSE(S)}$ is needed to arrive at precisely the correct number of spin labels for every SSE (data not shown). The two scores are averaged to yield the total S_{SSE} term with values between 0 and 1.

Element Connection Term

Element Connection (S_{EC}) favors patterns that connect each pair of SSEs with restraints. The ideal number of connections for each SSE pair (C) is defined by the ratio between the number of restraints (r) and the number of SSE pairs (p), $p = \big(s(s-1)\big)/2$, where s = number of SSEs. We define C' = $div(r,p)$ and $C"$ = $C' + 1$. In this term, floor C' and ceiling $C"$ are acceptable integer values for C. In addition, we define remainder of r/p as M = $mod(r,p)$. An optimal restraint distribution will have $C"$ restraints in M SSE pairs, and C' restraints in all the others. Like the Secondary Structure term, S_{EC} is a composite of two equally weighted component scores, $S_{EC(R)}$ and $S_{EC(C)}$. $S_{EC(R)}$ is the average percentage of restraints in each SSE pair up to the ideal value, $C"$. Thus,

$$S_{EC(R)} = \frac{1}{r}\sum_{i=1}^{p} \min\,(r_i\,, C")\ ,$$

where r_i = number of labels in the i[th] SSE pair. This component favors placement of restraints into SSE pairs during the optimization trajectory.

The second component of this term, $S_{EC(C)}$, is derived from the fraction of SSE pairs that contain exactly the ideal number of restraints:

$$S_{EC(C)} = \frac{1}{p}\{min[F', (p - M)] + min[F'', M]\},$$

where F' is the number of SSE pairs with C' restraints and F'' is the number of SSEs with C'' restraints. As in the Secondary Structure term, the composite scores of this term are complementary with $S_{EC(R)}$ measuring progress toward the optimal restraint placement and $S_{EC(C)}$ determining the correct number of restraints for every SSE pair. The two scores are averaged to yield the total S_{EC} term with values between 0 and 1.

Label Density Term

The Label Density score, S_{LD}, imposes equal distribution of spin labels along the sequence. For this purpose, spin label positions are treated as a vector $(a_o, a_1, \ldots, a_l, a_{l+1})$, where a_0 is the N-terminus and a_{l+1} is the C-terminus and $a_1, \ldots, a_l$ are the positions of the spin labels and l = number of spin labels. An optimal interval between spin labels (I) is the divisor of the ratio of the sequence length (g) to the number of intervals (n), where $n = l + 1$: $I = div(g, n)$. The score utilizes a harmonic penalty function. A normalization function, $f(x) = (x + 1)^{-1}$, is applied to rescale values between 0 and 1. Thus the term is defined as:

$$S_{LD} = \frac{1}{n}\sum_{i=0}^{n}\{[(a_{i+1} - a_i) - I]^2 + 1\}^{-1}.$$

Restraint assisted Rosetta Folding simulations and EPR distance interpretation

Rosetta simulations were performed in Rosetta++[30-33]. Specific standard Rosetta procedures were used that are described in details elsewhere[34]. In these course-grained simulations, residues side chains are regarded as centroid superatoms[30]. All T4L homologs were excluded from the fragment database prior to modeling in order to simulate structure determination of a novel protein fold as closely as possible. Models were obtained in

independent simulations on a cluster in Vanderbilt University's Advanced Computing Center for Research & Education (ACCRE). For each simulation, 1,000 models were created using the restraints selected by the algorithm for the α-helical subdomain of T4L (residues 58-164). In the algorithm optimization phase, C_α root mean squared deviation (RMSD) distributions and model quality measures for residues 70-164 were reported for all 1,000 models resulting from Rosetta folding. Residues 58-69 were excluded from RMSD analysis as these residues link the α-helical subdomain to the excluded β-strand subdomain and tended to vary in our models due to the absence of the β-strand domain. C_α RMSDs were used due to the course-grained nature of the modeling. In the experimental implementation phase, models were additionally filtered by lowest energy and restraint violation scores.

EPR distance restraints were implemented in Rosetta in a RosettaNMR[43-44] protocol as described previously[1,45]. Briefly, distance restraints are used as an additional penalty in the Rosetta energy function. This penalty is zero if the C_β-C_β distance (d_{CB}) of the restraint residues fall within the range specified. If this distance falls outside this range, a quadratic penalty function is applied. The boundary range used was based on the motion-on-a-cone model developed by Alexander et al.[1]. This model yielded a function describing the relationship between the experimentally measured spin label distance (d_{SL}) and the $d_{C\beta}$. The d_{SL} defines the range allowed for $d_{C\beta}$ (d_{SL}-12.5 Å to d_{SL}+2.5 Å) which corresponds to the most probable relative spin label orientations. For simulated restraints, the crystallographic $d_{C\beta}$ is used as the experimental distance (i.e. $d_{SL} - d_{C\beta} = 0$ or a parallel spin label orientation).

Recombinant expression and purification of T4L mutants

Cysteine residues were systematically introduced into a cysless T4L construct through double point mutations at restraint positions identified by the algorithm using *QuikChange*™ Site-Directed Mutagenesis Kit (Stratagene) as previously described[9]. Sample preparation has been described elsewhere[39,46]. Briefly, T4L mutants were sequenced, transformed into K38

cells, and expressed in Luria Broth (LB). All mutants were purified using cation exchange chromatography, labeled with a 5 fold excess of MTSSL (*S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl Methanethiosulfonate spin label, Toronto Research Chemicals) at room temperature for 2 hours, desalted and concentrated. A total of 21 double mutants resulted in the restraints used for the current analysis.

EPR distance measurements

Of the 21 restraints, 19 distances were found to be within the distance range appropriate for double electron-electron resonance (DEER) distance measurement[47-49]. DEER measurements were performed on a Bruker 580 pulsed EPR spectrometer operating at X-band (10 GHz) using a standard four-pulse protocol[49]. Experiments were performed at 83 K. Sample concentrations were 150 μM in a MOPS/Tris buffer (9 mM MOPS, 6 mM Tris, 50 mM NaCl, 0.02%(w/v) Sodium Azide, 0.1 mM EDTA) with 20%(v/v) glycerol as a cryoprotectant and a sample volume of 50 μl. Spin echo decays were baseline-corrected and analyzed by Tikhonov regularization[50-51] to determine average distances and distributions in distance (Appendix B). For all data, the selected regularization parameter corresponds to the elbow of the L-curve[50].

For the 2 pairs with distances too short for DEER analysis, distance distributions were determined from the continuous wave (CW) EPR spectra using the CWdipfit program developed by Peter Fajer and colleagues (http://www.sb.fsu.edu/~fajer/Programs/CWdipFit/cwdipfit)[52]. For each pair, fully labeled and underlabeled samples were prepared. Fully labeled samples were prepared as described above. Preparation of the underlabeled samples included incubation with 0.5x MTSSL for 1 hour at room temperature followed by addition of 20-fold excess of a diamagnetic MTSSL analog, (1-Acetyl-2,2,5,5,-tetramethyl-Δ3-pyrroline-3-methyl) Methanethiosulfonate (Toronto Research Chemicals). The fully labeled samples display distance-dependant dipolar coupling, while the underlabeled samples represent the EPR spectrum in its absence. CWdipFit assumes Gaussian-shaped distance distributions between

spin labels and utilizes Monte Carlo/SIMPLEX algorithm to fit dipolar coupled spectra using the underlabeled spectra as a proxy for the sum of singles[52-53]. The dipolar coupled spectra and fits are shown in Appendix B.

Results

Methodology

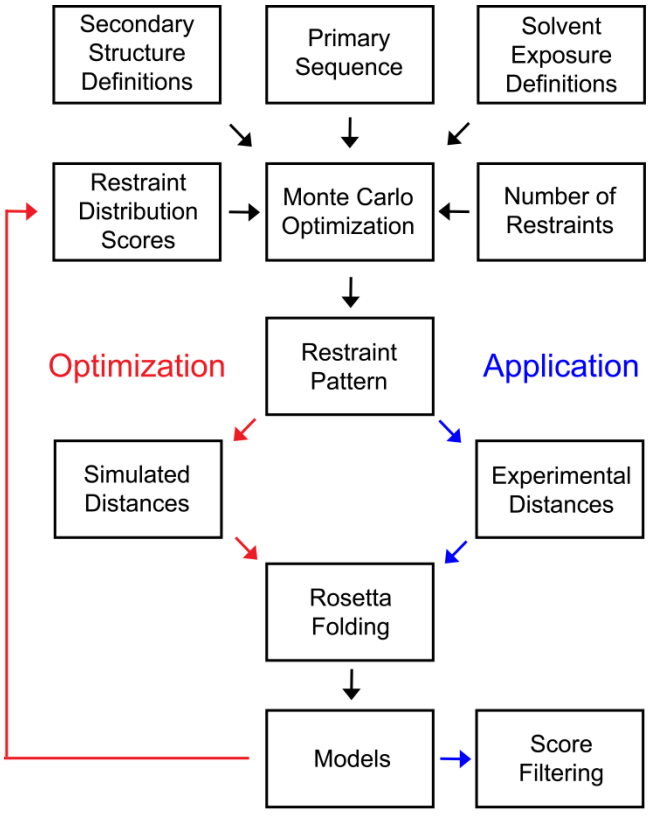


Figure 3.1. Methodological flowchart. Using sequence definitions as input, a Monte Carlo approach was applied to iteratively optimize the restraint distribution scores and number of restraints. Simulated restraints were calculated to guide Rosetta folding and modeling outcomes served as indicators of optimal restraint patterns (red arrow). To test the applicability of the algorithm for restraint selection, experimental distances were measured for an optimized restraint pattern and incorporated into the Rosetta folding algorithm (blue arrows). The resulting models were filtered by energy and restraint violation scores to exclusively yield high quality models.

The overall strategy, illustrated by the flowchart in Figure 3.1, uses the primary sequence, secondary structure, and solvent exposure definitions as input parameters. For secondary structure and solvent exposure, predicted and ideal (defined by the crystal structure) definitions were compared to assess the impact on model quality. The algorithm relies on a Monte Carlo search to optimize the restraint distribution terms that place pairs of spin labels along the sequence (Supp. Fig. 3.1). Briefly, a Sequence separation term, defined as the number of intervening amino acids between two spin labels in a pair, was included as an approximation for information content. To balance its tendency to cluster spin labels at the N- and C-termini, three terms favoring uniform sequence coverage were investigated. A secondary structure element (SSE) connection term (Element Connection) evenly connects all pairs of secondary structures, in this case 7 α-helices, with restraints effectively introducing a triangulation strategy. Alternatively, a Label Density term which distributes spin labels along the sequence at equal and regular intervals was included. Finally, we tested the efficacy of a Secondary Structure term that confines spin labels to segments of secondary structures avoiding loops and termini. Term combinations and weight ratios were evaluated for their effectiveness in selecting informative restraints for Rosetta folding (Supp. Fig. 3.2 and 3.3). The combination of Sequence Separation and Element Connection terms at a 1:1 weight ratio consistently yielded restraint patterns that resulted in the highest quality models by Rosetta folding. Figure 3.2 illustrates how an initial random distribution of labels is shuffled to maximize the Sequence Separation and Element Connections scores.

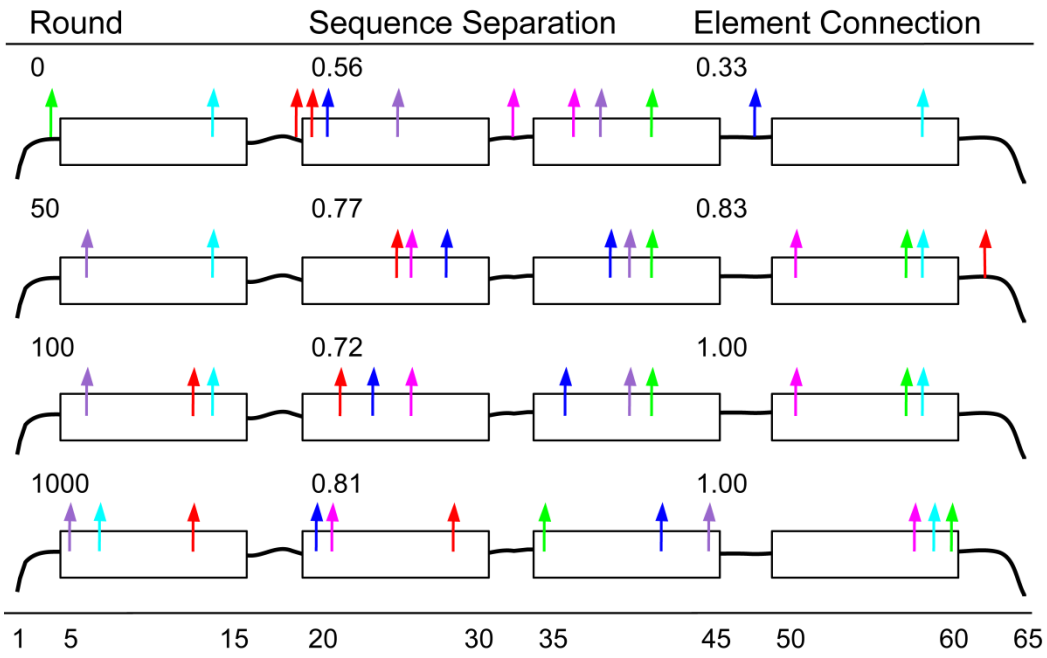


Figure 3.2. An example optimization trajectory. In our simplified 4 helix, 6-restraint system, each arrow represents a spin label and arrows of the same color correspond to spin label pairs. Starting from an initial random distribution, the algorithm repositions the spin labels to maximize Sequence Separation and Element Connection scores resulting in an optimized pattern.

In the algorithm development phase described above, the term combination and relative weight were determined using simulated EPR distances. For this purpose, the distance between the β-carbon of each pair of residues, $d_{C\beta}$, was obtained from the crystal structure (2LZM) and used as an experimental restraint. To simulate the uncertainty associated with interpretation of distances between spin labels, the corresponding restraint was allowed a range of $d_{C\beta}$ -12.5 Å to $d_{C\beta}$ +2.5 Å based on the motion on a cone model described previously[1] and in the Methods. Models with $d_{C\beta}$ distances outside this range are penalized in the Rosetta Energy score.

The output of the restraint-assisted Rosetta folding consisted of 1,000 models. Quality measures defined by the models C_α RMSD to the crystal structure were used as indicators of improvement in the Rosetta sampling of conformational space. To avoid perturbation due to spin label incorporation, the algorithm excluded residues predicted to be buried. This did not affect the quality of models generated by Rosetta (Supp. Fig. 3.4). In contrast, the use of

predicted secondary structure resulted in a significant decrease in model quality (Supp. Fig. 3.4). Therefore, for the purposes of evaluating the effectiveness of the algorithm, secondary structure definitions were based on the crystal structure.

The α-helical subdomain of T4L (residues 58-164) was selected as a model system for this analysis. T4L has been extensively investigated by spin labeling[54-55] and was the target of a previous study to assess the potential of EPR restraints to increase the efficiency of conformational space sampling by Rosetta[1]. The 107 amino acid target region is well within the size limit for efficient structure prediction by Rosetta *de novo* folding[35]. For the analysis presented here, we excluded structures homologous to T4L from the fragment library to mimic protein structure prediction of novel protein folds. Under these conditions, Rosetta folding in the absence of restraints yields consistently about 8% correctly folded models leaving sufficient dynamic range to evaluate the impact of EPR restraints.

Optimized restraints increase the fraction of correct topology models

Following selection of the terms and their relative weights described above, we assessed the degree to which optimized restraint patterns improve the quality of T4L models predicted by Rosetta. For this purpose, 10 sets of 21 restraints were used in conjunction with Rosetta to generate 1,000 T4L models. An equivalent number of models was generated by folding without restraints as well as in the presence of 21 randomly selected restraints. Consistently, models obtained using optimized restraints had vastly better quality measures (Fig. 3.3). A left shift in the RMSD distribution reflects the presence of a major population of models with RMSD below 7.5 Å (Fig. 3.3a). It is generally accepted that 7.5 Å is the RMSD at which models have the correct overall fold as the native structure[56]. Thus using optimized restraints, 54.4% of Rosetta models achieve the general fold compared to 21.0% and 8.0% of models if randomized or no restraints are used, respectively (Fig. 3.3b).

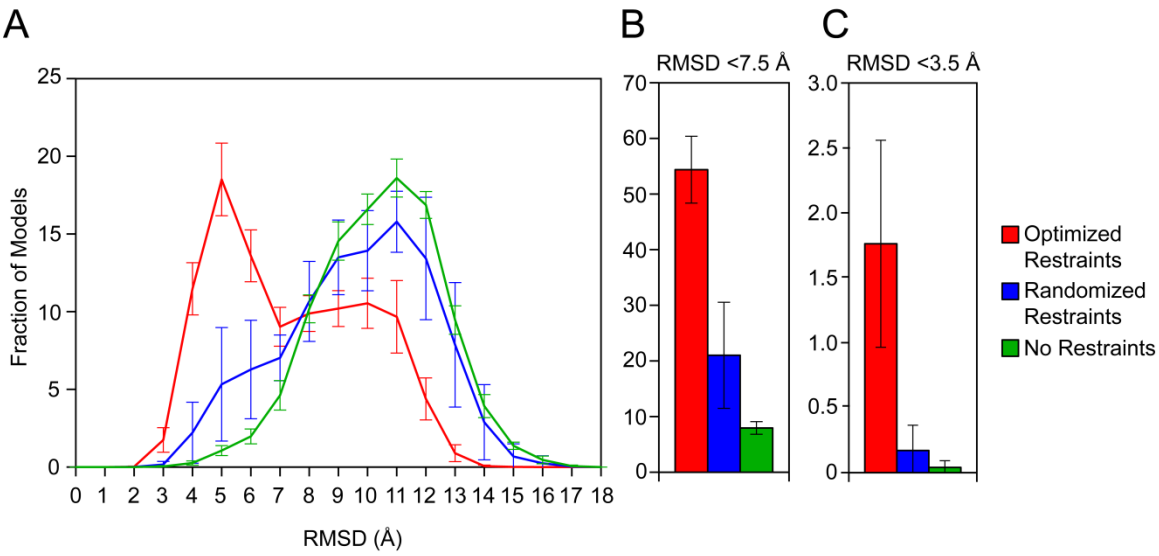


Figure 3.3. Effects of incorporation of restraints into Rosetta Folding. (A) C_α RMSD distributions of models generated with optimized restraints, randomized restraints, and no restraints. The distinct left shift in the distribution with optimized restraints is indicative of a more efficient restriction of the conformational search space. Fraction of models fulfilling two measures of model quality, C_α RMSD < 7.5 Å (B) and C_α-RMSD <3.5 Å (C) display even more pronounced improvements. Models with RMSD < 7.5 Å to the crystal structure typically have the correct general fold, while models with RMSD < 3.5 Å are considered candidates for high resolution refinement.

Optimized restraints also lead to a significant increase in the percentage of models with $C_{\square}$-RMSDs below 3.5 Å reflecting more effective sampling of conformational space by Rosetta (Fig. 3.3c). These models being closest to the native structure are ideal candidates for subsequent high resolution refinement[1]. Using an RMSD cutoff of 3.5 Å as a criterion, 1.7% of models generated by incorporation of optimized restraints are considered high quality. To achieve 1Å resolution, a starting set of at least 2,000 such models are needed[35], which is within a computational reasonable time frame. In contrast, only 0.2% of models generated using randomized restraints fulfilled the 3.5 Å RMSD criterion. Thus, to achieve high resolution, one million models are needed which requires substantially more computational resources. If no restraints are used, the computational cost becomes prohibitive, as only 0.04% percent of models have less than 3.5 Å RMSD, therefore requiring tens of millions of models. Furthermore, EPR restraints allow selection of correct topology models for refinement[1].

Improvement of model quality requires a limited number of restraints

The choice of 21 restraints above was dictated by detailed analysis of the dependence of model quality on restraint number. For this purpose, the Rosetta folding protocol of Fig. 3.1 was applied successively increasing the number of restraints followed by assessment of model quality. Note that 21 restraints are required to fulfill all pairwise connections between the 7 helices of T4L C-terminal domain. Therefore for restraint numbers larger than 21, the algorithm was modified to ensure that the additional restraints duplicating existing secondary structure connectivities are evenly distributed.

Fig. 3.4a demonstrates that increasing the number of restraints leads to a rapid increase in the percentage of models having the correct fold (C_α RMSD below 7.5 Å). This effect is pronounced with as few as 5-10 restraints. The trend levels off in the region of 20-22 restraints suggesting that redundant connections between secondary structures add little information (Fig. 3.4a). In contrast, a more stringent quality measure, the percentage of models with RMSD below 3.5 Å, hardly improves until the number of restraints is well above 10 (Fig. 3.4b). This lag reflects the significantly lower probability that these models are sampled in the absence of restriction on the search space. Indeed, this number remains rather unaffected by the introduction of additional random restraints.

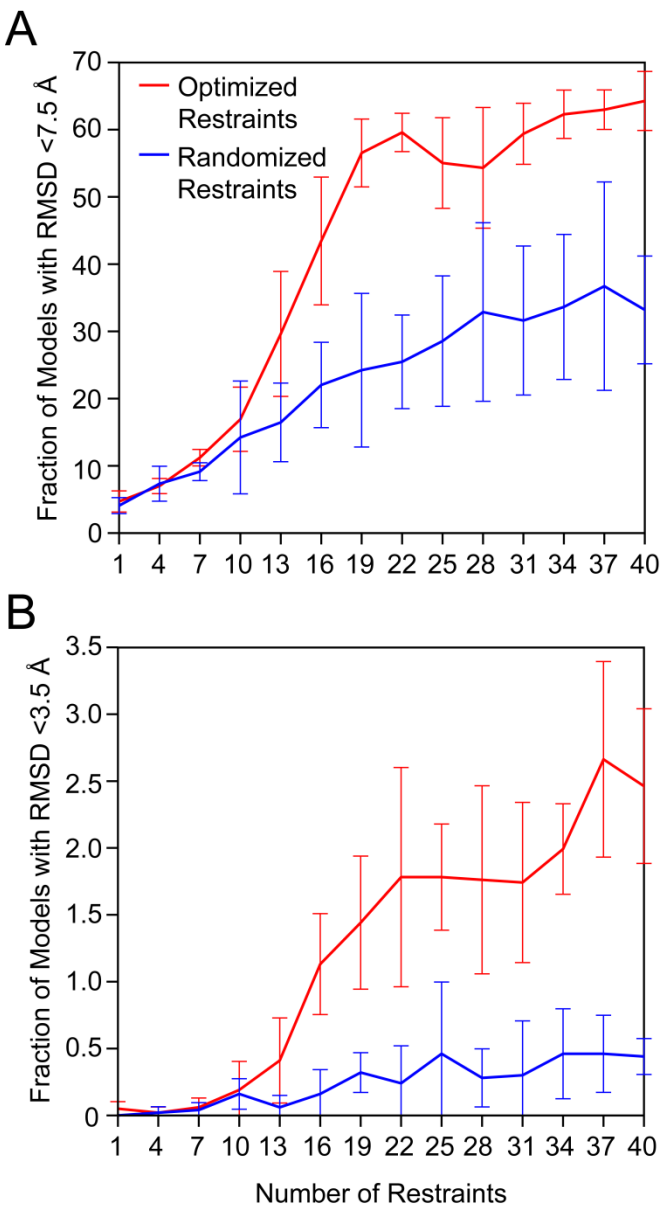


Figure 3.4. Impact of restraint number on model quality. Model quality measures, C_α RMSD <7.5 Å (A) and C_α RMSD <3.5 Å (B) improve with increasing number of restraints. The improvement levels off at a threshold number where the α-helical pairwise connectivities are fulfilled (20-22 restraints for the 7 helices of T4L C-terminal domain).

The percentage of models with the correct fold plateaus at approximately 60 percent. The rest typically fulfill the restraints but have incorrect folds. This is not surprising given the soft interpretation of the restraints within a wide error margin (15 Å) by the cone model. It is likely that this limitation also accounts for the relatively limited percentage of high quality models, i.e.

with RMSD below 3.5 Å. Interestingly, the incorrectly folded models score worse in Rosetta's knowledge-based potential (see below) allowing for selection of correctly folded models by energy score. In addition, the improved overall quality with few restraints (Fig. 3.4) provides a plausible explanation for the surprisingly good performance of random restraints in Figure 3.3.

Rosetta folding of T4L using optimized experimental restraints

The optimization of the algorithm used simulated distances between residue pairs. As described above, this approximation centers the distribution at the $d_{C\beta}$ while the experimental distribution is centered on the distance separating the two spin labels. The offset between these two values is determined by the relative orientation of the labels and represents the major source of uncertainty in interpretation of EPR distance restraints. To assess the consequences of this approximation and validate the optimization strategy, we carried out Rosetta folding of T4L using experimentally determined distances for a set of spin label pairs selected by the algorithm described above. Double cysteine mutants were constructed and the corresponding proteins purified and spin labeled as described in the Methods. Most pairs, except two, were in the distance range suitable for DEER analysis[49]. Spin echo decays were baseline-corrected and analyzed by Tikhonov regularization[50] to yield distance distributions as described in the Methods and illustrated in Appendix B. For the short range pairs (86C/112C and 127C/155C), spectral simulation was used to extract a Gaussian distribution of distance from the CW-EPR spectra (Appendix B).

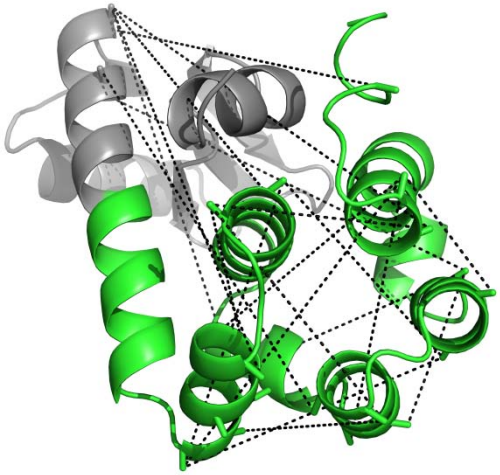


Figure 3.5. Location of experimentally measured EPR restraints in the T4L crystal structure. T4L is colored to highlight the subdomains included in (green) and excluded from (silver) RMSD calculations. The dotted lines represent the pairs of spin labels in the restraints.

The position of the pairs is mapped onto the T4L crystal structure in Fig. 3.5. Table 3.1 reports the average distance between the spin label pairs as well as the width of the distance distribution. Compared to the $d_{C\beta}$, the deviations show the expected pattern of larger spin label distances. The distributions are predominantly narrow despite the surface exposed location of the spin labeled sites. Thus, even though most spin labels are mobile as evidenced by the EPR lineshapes (data not shown), it appears that the sampled rotameric states are restricted.

Mutants	Average distance ± distribution width (Å)
59/159	41.9 ± 2.7
60/90	37.8 ± 4.5
60/94	25.5 ± 3.1
60/109	35.2 ± 2.6
60/154	34.1 ± 2.0
62/134	41.1 ± 1.5
64/122	34.1 ± 2.5
82/94	30.7 ± 3.3
82/132	26.3 ± 3.5
82/155	35.8 ± 2.5
83/123	20.5 ± 3.4
93/112	26.1 ± 1.5
93/123	24.8 ± 2.3
93/154	25.1 ± 2.4
94/132	31.7 ± 1.3
108/155	35.2 ± 2.3
109/134	30.6 ± 2.8
115/155	28.2 ± 2.4
116/134	20.2 ± 1.5
86/112[a]	13.0 ± 5.1
127/155[a]	12.1 ± 3.4

[a] Determined by simulation of CW spectra with CWdipFit. All other distances were obtained from Tikhonov analysis of DEER signals.

Table 3.1. Average distance and distribution width for experimentally measured restraints.

Figure 3.6 demonstrates that Rosetta folding of 10,000 models using the experimental distances leads to improvement in model quality measures that follow the same trends of Figure 3.3. These include a left shift in the C_α RMSD distribution, an increase in the fraction of models with the correct folding topology (RMSD < 7.5 Å), and more importantly of the percentage of high quality models (RMSD < 3.5 Å). However, these improvements underperform those expected from simulated distances. The origin of this underperformance can be rationalized by comparing the upper bound of the simulated and experimental restraints. Experimentally determined distances tend to be larger than the $d_{C\beta}$ thereby increasing the upper bound. Thus, conformational space is less restrained leading to a reduced model quality.

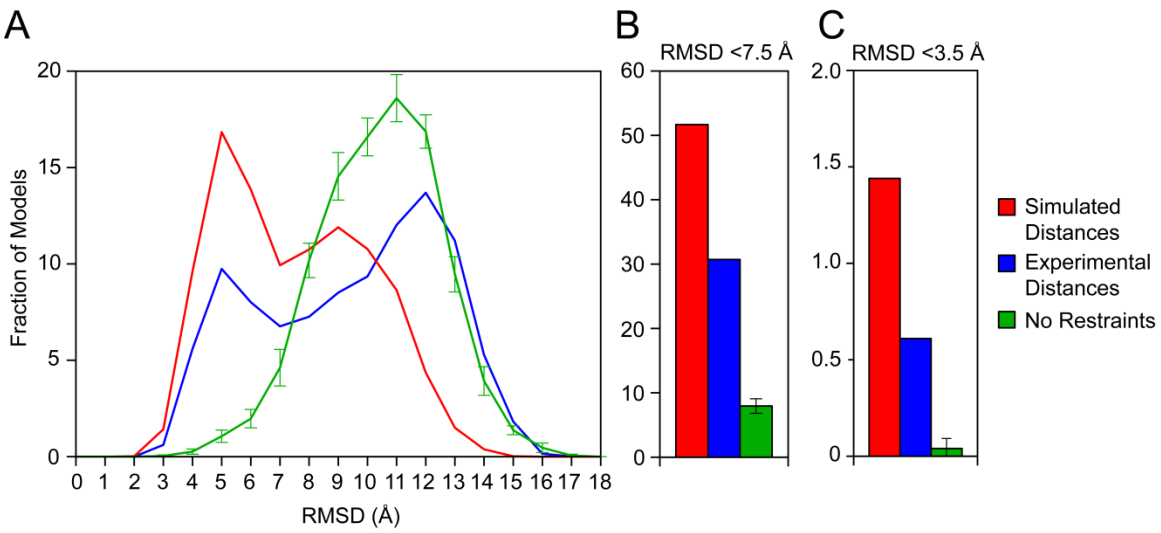


Figure 3.6. Comparison of models generated using experimental and simulated restraints. Simulated and experimentally measured distances (Table 1) for the optimized pattern shown in Fig. 5 were used to fold T4L using Rosetta. The resulting $C_{\square}$-RMSD distributions (A) and quality measures (B, C) were derived from 10,000 models for each set of distances (simulated and experimental). The no restraints curve represents 10 Rosetta trajectories each of 1,000 models. The difference between the simulated and experimental results highlights the effect of the typically larger upper bound associated with experimentally determined distances.

The models generated by incorporation of experimental restraints into Rosetta folding were sorted based on their Rosetta energy and restraint violation scores. While models of vastly different RMSDs have similar Rosetta energy or restraint violation scores, only models with low RMSDs have low scores in both criteria. Figure 3.7 demonstrates the improvement in model quality when a Rosetta energy score threshold of below 30 and a cumulative restraint violation score threshold of less than 2.5 Å were applied. This resulted in an enrichment factor of 7.2 for models with RMSDs below 3.5 Å, retaining 44 of the 61 original models. Thus the combination of these two scores can identify the subset of models with topologies closest to the native structure.

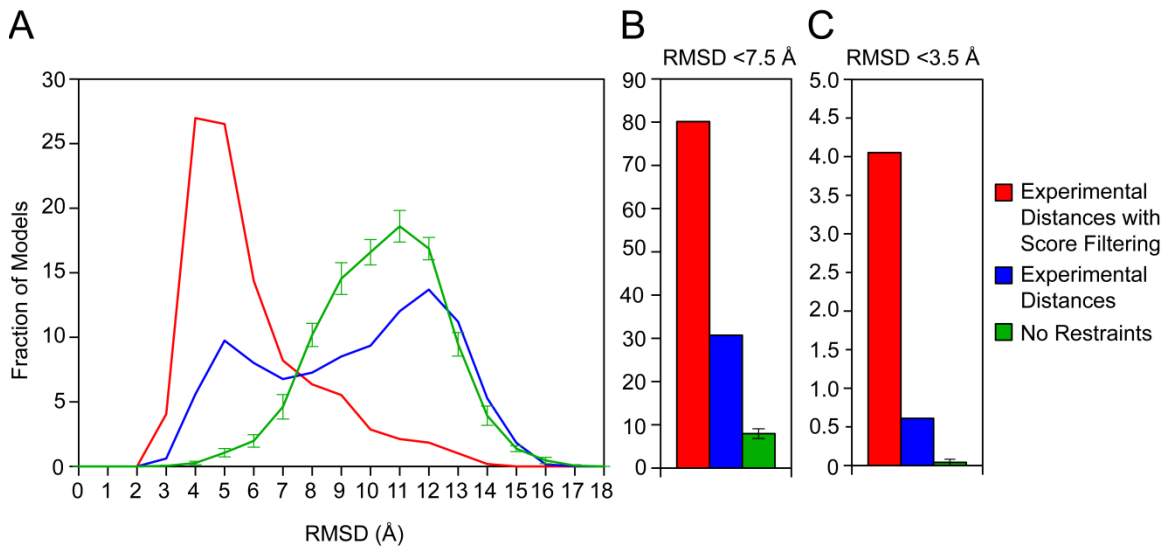


Figure 3.7. Effects of filtering by Rosetta energy and restraint violation scores on model quality. (A) RMSD distributions and (B, C) model quality measures shown in Figure 3.6 after score filtering. Filtering by Rosetta energy score and restraint violation score excludes almost exclusively low quality models.

Discussion

The requirement for incorporation of spin labels into protein sequences shapes the methodology of spin labeling in two fundamental ways. First, the experimental throughput is limited leading to sparse restraints. Second, the arm linking the spin label to the protein backbone introduces an uncertainty in the interpretation of these restraints. The algorithm presented in this paper advances the methodological blueprint of spin labeling and EPR spectroscopy by optimizing the information content of EPR distance restraints and consequently alleviating the experimental bottleneck.

The experimental implementation of this strategy presented here charts a roadmap for future improvements. As expected, using the cone model of Alexander et al. [1]for interpretation of the EPR distances significantly compromises the quality of the experimental data. Narrow distance distributions at a number of sites imply a tighter limit on the distance range than the 15 Å assumed in the cone model. Furthermore, the shape of the distribution (Appendix A) is in

stark contrast to the flat scoring potential implemented in the Rosetta protocol. The consequence of these approximations is that topologies with C_α-RMSDs as large as 12 Å fulfill the EPR restraints. We are developing a probability function to describe the offset distribution between the d_{SL} and $d_{C\beta}$[57]. Furthermore, explicit modeling of the spin label should limit the uncertainty associated with unknown spin label orientation to the backbone. It has been demonstrated that molecular dynamics simulations can reproduce average distances between spin labels[28-29]. Though more computationally intensive, these approaches will enhance Rosetta models quality specifically increasing the fraction of those below 3.5 Å RMSD.

The performance of the algorithm is also degraded when prediction rather than actual secondary structures are used. The origin of this effect is the inaccurate prediction of the number of secondary structures which for a fixed number of restraints alter the required pairwise connectivities. In the context of the application of this approach to a protein of unknown structure, the location and length of secondary structure can be experimentally determined and/or verified through nitroxide scanning experiments[58-60].

That the many approximations did not hinder the identification of the correct fold by Rosetta reflects the robustness of its energy function. Similarly, a few EPR restraints lead to a measurable improvement in the quality of the folds highlighting the critical role of these restraints in reducing Rosetta's conformational search space. These findings reinforce the rationale of using de novo folding to balance the sparseness of the EPR restraints and their intrinsically lower quality.

Although the algorithm developed in this paper is general, our ultimate goal is to develop a suite of tools to determine structure of membrane proteins. While Rosetta has been successfully used to generate constrained models of membrane proteins[61-63], it is likely to be less robust given the limited number of folds and topologies in the protein data bank. Though this may be partially mitigated by the restricted diversity of membrane protein fold imposed by the membrane environment, the number of EPR restraints needed to obtain high quality models

is likely to be larger. Furthermore, the rule of one restraint per pair of secondary structures may have to be modified for the longer helices found in these proteins. In this context, redundant restraints may prove important for longer helices common in transmembrane proteins. We expect that additional algorithm terms to optimize the distribution of redundant restraints will be developed. Nevertheless, this algorithm represents a first step in this direction.

References

1. Alexander, N., Bortolus, M., Al-Mestarihi, A. & Mchaourab, H. De novo high-resolution protein structure determination from sparse spin-labeling epr data. *Structure* **16**, 181-195 (2008).
2. Tusnady, G., Dosztanyi, Z. & Simon, I. Transmembrane proteins in the protein data bank: Identification and classification. *Bioinformatics* **20**, 2964-2972 (2004).
3. Haley, D. A., Bova, M. P., Huang, Q. L., Mchaourab, H. S. & Stewart, P. L. Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. *Journal of Molecular Biology* **298**, 261-272 (2000).
4. Fanucci, G. E., Coggshall, K. A., Cadieux, N., Kim, M., Kadner, R. J. & Cafiso, D. S. Substrate-induced conformational changes of the periplasmic n-terminus of an outer-membrane transporter by site-directed spin labeling. *Biochemistry* **42**, 1391-1400 (2003).
5. Hubbell, W. L., Gross, A., Langen, R. & Lietzow, M. A. Recent advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology* **8**, 649-656 (1998).
6. Hubbell, W. L., Mchaourab, H. S., Altenbach, C. & Lietzow, M. A. Watching proteins move using site-directed spin labeling. *Structure* **4**, 779-783 (1996).
7. Persson, M., Harbridge, J. R., Hammarstrom, P., Mitri, R., Martensson, L. G., Carlsson, U., Eaton, G. R. & Eaton, S. S. Comparison of electron paramagnetic resonance methods to determine distances between spin labels on human carbonic anhydrase ii. *Biophys J* **80**, 2886-2897 (2001).
8. Claxton, D. P., Zou, P. & Mchaourab, H. S. Structure and orientation of t4 lysozyme bound to the small heat shock protein alpha-crystallin. *J Mol Biol* **375**, 1026-1039 (2008).
9. Dong, J., Yang, G. & Mchaourab, H. S. Structural basis of energy transduction in the transport cycle of msba. *Science* **308**, 1023-1028 (2005).
10. Altenbach, C., Kusnetzow, A. K., Ernst, O. P., Hofmann, K. P. & Hubbell, W. L. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc Natl Acad Sci U S A* **105**, 7439-7444 (2008).
11. Borbat, P. P. & Freed, J. H. Measuring distances by pulsed dipolar esr spectroscopy: Spin-labeled histidine kinases. *Methods in Enzymology* **423**, 52-116 (2007).
12. Smirnova, I., Kasho, V., Choe, J. Y., Altenbach, C., Hubbell, W. L. & Kaback, H. R. Sugar binding induces an outward facing conformation of lacy. *Proc Natl Acad Sci U S A* **104**, 16504-16509 (2007).
13. Zou, P., Bortolus, M. & Mchaourab, H. S. Conformational cycle of the abc transporter msba in liposomes: Detailed analysis using double electron-electron resonance spectroscopy. *J Mol Biol* **393**, 586-597 (2009).
14. Zou, P. & Mchaourab, H. S. Alternating access of the putative substrate-binding chamber in the abc transporter msba. *J Mol Biol* **393**, 574-585 (2009).
15. Wegener, C., Tebbe, S., Steinhoff, H. J. & Jung, H. Spin labeling analysis of structure and dynamics of the na(+)/proline transporter of escherichia coli. *Biochemistry* **39**, 4831-4837 (2000).
16. Padmavathi, P. V. & Steinhoff, H. J. Conformation of the closed channel state of colicin a in proteoliposomes: An umbrella model. *J Mol Biol* **378**, 204-214 (2008).
17. Hanelt, I., Wunnicke, D., Muller-Trimbusch, M., Vor der Bruggen, M., Kraus, I., Bakker, E. P. & Steinhoff, H. J. Membrane region m2c2 in subunit ktrb of the k+ uptake system ktrab from vibrio alginolyticus forms a flexible gate controlling k+ flux: An electron paramagnetic resonance study. *J Biol Chem* **285**, 28210-28219 (2010).
18. Koteiche, H. A. & Mchaourab, H. S. The determinants of the oligomeric structure in hsp16.5 are encoded in the alpha-crystallin domain. *FEBS Lett* **519**, 16-22 (2002).

19. Mchaourab, H. S., Godar, J. A. & Stewart, P. L. Structure and mechanism of protein stability sensors: Chaperone activity of small heat shock proteins. *Biochemistry* **48**, 3828-3837 (2009).
20. Jao, C. C., Hegde, B. G., Chen, J., Haworth, I. S. & Langen, R. Structure of membrane-bound alpha-synuclein from site-directed spin labeling and computational refinement. *Proc Natl Acad Sci U S A* **105**, 19666-19671 (2008).
21. Margittai, M. & Langen, R. Spin labeling analysis of amyloids and other protein aggregates. *Methods Enzymol* **413**, 122-139 (2006).
22. Jeschke, G., Wegener, C., Nietschke, M., Jung, H. & Steinhoff, H. J. Interresidual distance determination by four-pulse double electron-electron resonance in an integral membrane protein: The na+/proline transporter putp of escherichia coli. *Biophysical Journal* **86**, 2551-2557 (2004).
23. Hilger, D., Polyhach, Y., Padan, E., Jung, H. & Jeschke, G. High-resolution structure of a na+/h+ antiporter dimer obtained by pulsed electron paramagnetic resonance distance measurements. *Biophys J* **93**, 3675-3683 (2007).
24. Hilger, D., Polyhach, Y., Jung, H. & Jeschke, G. Backbone structure of transmembrane domain ix of the na+/proline transporter putp of escherichia coli. *Biophys J* **96**, 217-225 (2009).
25. Bhatnagar, J., Borbat, P. P., Pollard, A. M., Bilwes, A. M., Freed, J. H. & Crane, B. R. Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the chea histidine kinase, and the coupling protein chew as determined by pulsed dipolar esr spectroscopy. *Biochemistry* **49**, 3824-3841 (2010).
26. Fajer, M. I., Li, H., Yang, W. & Fajer, P. G. Mapping electron paramagnetic resonance spin label conformations by the simulated scaling method. *Journal of the American Chemical Society* **129**, 13840-13846 (2007).
27. Sale, K., Song, L., Liu, Y. S., Perozo, E. & Fajer, P. Explicit treatment of spin labels in modeling of distance constraints from dipolar epr and deer. *Journal of the American Chemical Society* **127**, 9334-9335 (2005).
28. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **9**, 1895-1910 (2007).
29. Borovykh, I. V., Ceola, S., Gajula, P., Gast, P., Steinhoff, H. J. & Huber, M. Distance between a native cofactor and a spin label in the reaction centre of rhodobacter sphaeroides by a two-frequency pulsed electron paramagnetic resonance method and molecular dynamics simulations. *J Magn Reson* **180**, 178-185 (2006).
30. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J Mol Biol* **268**, 209-225 (1997).
31. Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C. & Baker, D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**, 82-95 (1999).
32. Bonneau, R., Strauss, C. E. & Baker, D. Improving the performance of rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. *Proteins* **43**, 1-11 (2001).
33. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using rosetta. *Methods Enzymol* **383**, 66-93 (2004).
34. Bradley, P., Malmstrom, L., Qian, B., Schonbrun, J., Chivian, D., Kim, D. E., Meiler, J., Misura, K. M. & Baker, D. Free modeling with rosetta in casp6. *Proteins* **61 Suppl 7**, 128-134 (2005).
35. Bradley, P., Misura, K. & Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-1871 (2005).

36. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using rosetta. *Proteins* **62**, 1010-1025 (2006).
37. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**, 363-382 (2008).
38. Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B. H., Das, R., Grishin, N. V. & Baker, D. Structure prediction for casp8 with all-atom refinement using rosetta. *Proteins* **77 Suppl 9**, 89-99 (2009).
39. Matthews, B. W. Structural and genetic analysis of the folding and function of t4 lysozyme. *FASEB Journal* **10**, 35-41 (1996).
40. McGuffin, L. J., Bryson, K. & Jones, D. T. The psipred protein structure prediction server. *Bioinformatics* **16**, 404-405 (2000).
41. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. & Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* **9**, 51 (2009).
42. Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R. & Meiler, J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J Mol Model* **15**, 1093-1108 (2009).
43. Bowers, P., Strauss, C. & Baker, D. De novo protein structure determination using sparse nmr data. *Journal of Biomolecular NMR* **18**, 311-318 (2000).
44. Rohl, C. A. & Baker, D. De novo determination of protein backbone structure from residual dipolar couplings using rosetta. *Journal of the American Chemical Society* **124**, 2723-2729 (2002).
45. Alexander, N., Bortolus, M., Al-Mestarihi, A., Mchaourab, H. & Meiler, J. De novo high-resolution protein structure determination from sparse spin-labeling epr data. *Structure* **16**, 181-195 (2008).
46. Koteiche, H. A. & Mchaourab, H. S. Folding pattern of the alpha-crystallin domain in alphaa-crystallin determined by site-directed spin labeling. *Journal of Molecular Biology.* **294**, 561-577 (1999).
47. Borbat, P. P., Davis, J. H., Butcher, S. E. & Freed, J. H. Measurement of large distances in biomolecules using double-quantum filtered refocused electron spin-echoes. *Journal of the American Chemical Society* **126**, 7746-7747 (2004).
48. Borbat, P. P., Mchaourab, H. S. & Freed, J. H. Protein structure determination using long-distance constraints from double-quantum coherence esr: Study of t4 lysozyme. *Journal of the American Chemical Society.* **124**, 5304-5314 (2002).
49. Jeschke, G. Distance measurements in the nanometer range by pulse epr. *Chemphyschem* **3**, 927-932 (2002).
50. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed esr using tikhonov regularization. *Journal of Magnetic Resonance* **172**, 279-295 (2005).
51. Jeschke, G., Koch, A., Jonas, U. & Godt, A. Direct conversion of epr dipolar time evolution data to distance distributions. *J Magn Reson* **155**, 72-82 (2002).
52. Sen, K. I., Logan, T. M. & Fajer, P. G. Protein dynamics and monomer-monomer interactions in antr activation by electron paramagnetic resonance and double electron-electron resonance. *Biochemistry* **46**, 11639-11649 (2007).
53. Sugata, K., Song, L., Nakamura, M., Ueki, S., Fajer, P. G. & Arata, T. Nucleotide-induced flexibility change in neck linkers of dimeric kinesin as detected by distance measurements using spin-labeling epr. *J Mol Biol* **386**, 626-636 (2009).
54. Mchaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of spin-labeled side chains in t4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692-7704 (1996).

55. Mchaourab, H. S., Oh, K. J., Fang, C. J. & Hubbell, W. L. Conformation of t4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. *Biochemistry.* **36**, 307-316 (1997).
56. Zemla, A., Venclovas, C., Moult, J. & Fidelis, K. Processing and analysis of casp3 protein structure predictions. *Proteins* **Suppl 3**, 22-29 (1999).
57. Hirst, S. J., Alexander, N., Mchaourab, H. S. & Meiler, J. Rosettaepr: An integrated tool for protein structure determination from sparse epr data. *J Struct Biol* **173**, 506-514 (2011).
58. Kaplan, R. S., Mayor, J. A., Kotaria, R., Walters, D. E. & Mchaourab, H. S. The yeast mitochondrial citrate transport protein: Determination of secondary structure and solvent accessibility of transmembrane domain iv using site-directed spin labeling. *Biochemistry* **39**, 9157-9163 (2000).
59. Berengian, A. R., Bova, M. P. & Mchaourab, H. S. Structure and function of the conserved domain in alphaa-crystallin. Site-directed spin labeling identifies a beta-strand located near a subunit interface. *Biochemistry* **36**, 9951-9957 (1997).
60. Mansoor, S. E., Mchaourab, H. S. & Farrens, D. L. Determination of protein secondary structure and solvent accessibility using site-directed fluorescence labeling. Studies of t4 lysozyme using the fluorescent probe monobromobimane. *Biochemistry* **38**, 16383-16393 (1999).
61. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using rosetta. *Proteins* **62**, 1010-1025 (2006).
62. Barth, P., Wallner, B. & Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci USA* **106**, 1409-1414 (2009).
63. McAllister, S. R. & Floudas, C. A. Alpha-helical topology prediction and generation of distance restraints in membrane proteins. *Biophys J* **95**, 5281-5295 (2008).

CHAPTER 4

CONFORMATIONAL DYNAMICS OF LIGAND-DEPENDENT ALTERNATING ACCESS IN LeuT

Abstract

The leucine transporter (LeuT) is a bacterial homolog of neurotransmitter:sodium symporters (NSS) that catalyze reuptake of neurotransmitters at the synapse. Crystal structures of wild type and mutant LeuT have been interpreted as conformational states in the coupled transport cycle. However, the mechanistic identities of these structures have not been validated and the ligand-dependent conformational equilibrium of LeuT has not been defined. Here, we utilized distance measurements between spin label pairs to elucidate Na^+- and leucine-dependent conformational changes on the intracellular and extracellular sides of the transporter. The results identify structural motifs that underlie the Na^+ and leucine driven isomerization of LeuT between outward-, inward-facing and occluded states. The conformational changes reported here present a dynamic picture of the alternating access mechanism of LeuT and NSS that is different to the inferences reached from currently available crystal structures.

Introduction

Secondary active transporters harness the energy of ion gradients to power the uphill movement of solutes across membranes. Mitchell proposed[1] and others elaborated[2-4] “alternating access” mechanisms wherein the transporter transitions between two conformational states that alternately expose the substrate binding site to the two sides of the membrane. Uncovering these conformational states and the mechanism by which the energy of

ion or substrate gradients drives the transition between these states is fundamental to understanding active transport. Ion-coupled transporters of the LeuT Fold support both symport and antiport modes that couple translocation of chemically diverse substrates to transmembrane ion gradients[5]. The recurrent LeuT Fold consists of two sets of five transmembrane helices related by two-fold symmetry around an axis nearly parallel to the membrane[6]. Ions and substrates are bound near the middle of the membrane, stabilized by electrostatic interactions with unwound regions of transmembrane helices (TM) 1 and 6. The leucine transporter (LeuT), the founding member of this structural class, is a Na^+-coupled amino acid transporter from *Aquifex aeolicus* that shows sequence similarity to neurotransmitter:sodium symporters (NSS)[6]. The NSS family includes biogenic amine transporters that terminate chemical neurotransmission by the active removal of neurotransmitters from the synaptic cleft. While wild type (WT) LeuT functions as a H^+-dependent Na^+ symporter, it also has been engineered to acquire the Cl^- dependence characteristic of the mechanisms of eukaryotic transporters[7-9]. Moreover, its transport activity is inhibited by tricyclic antidepressants[10-11], which are also inhibitors of the eukaryotic transporters. Because of these similarities in sequence and function, as well as its suitability for biochemical and biophysical study, LeuT has emerged as a paradigm for modeling the structure of NSS transporters and their complexes with inhibitors, and to probe conformational changes that may form the basis of alternating access in this family.

Rapid progress in structure determination of LeuT Fold transporters has defined their molecular architectures and revealed the principles of ion and substrate binding[12-17], but it remains difficult to extrapolate these static snapshots to a suite of conformational steps underlying alternating access. Crystal structures have been classified in relation to the assumed mechanism as inward-facing, outward-facing or substrate-occluded states, interpolated to infer plausible pathways of substrate binding and release, and cast as intermediates in the transport cycle[13-16,18]. However, development of transport models is critically dependent on validation of

the mechanistic identities of the available structures and their incorporation into a dynamic framework of ion- and substrate-dependent equilibria[19-21].

For LeuT, three different conformations observed crystallographically have been classified as outward-facing (PDB ID: 3TT1)[13], inward-facing (PDB ID: 3TT3)[13], and substrate-occluded (PDB ID: 2A65)[6] based on the accessibility of the binding site. Capturing the first two conformations necessitated multiple mutations of highly conserved residues, and subsequent conformational selection by antibodies[13]. The inference that the resulting structures represent actual intermediates in the alternating access cycle has not been verified. Furthermore, the occluded structure, bound to Na^+ and Leu as well as other transported amino acids, was determined in the WT background[6], but the relation of this conformation to the transport mechanism was challenged on the basis of the inhibitory effects of *n*-octyl-β-D-glucopyranoside (β-OG)[22], the detergent in which the structure was determined. Thus even with the results of previous spectroscopic[23-24], computational[25-26] and functional[25,27] analyses seeking to clarify specific aspects of LeuT conformational dynamics, a global perspective on its Na^+- and Leu-dependent conformational changes is still lacking. Here, we utilized Site-Directed Spin-Labeling (SDSL)[28] and Double Electron-Electron Resonance (DEER) spectroscopy[29] to measure distance probabilities between spin label pairs in LeuT in order to (i)- define the ligand-dependent conformational equilibrium of LeuT, (ii)- identify the structural elements that mediate alternating access, and (iii)- investigate whether the LeuT conformational cycle involves isomerization between the crystal structures. Interpretation of these results suggests a novel mechanism of LeuT alternating access that is at variance with the current model stimulated by the crystal structures[13].

Methods

Mutagenesis, expression, purification and labeling of LeuT

All LeuT mutations were introduced into the recombinant LeuT construct containing an N-terminal decahistidine tag[25] using PCR-based site-directed mutagenesis and confirmed by DNA sequencing. Mutant LeuT was expressed in *Escherichia coli* BL21(DE3) as previously described[25]. LeuT was washed three times in 200mM Tris-MES, pH 7.5, 20% (v/v) glycerol to remove bound leucine and subsequently extracted from native membranes with 40 mM (2% w/v) *n*-dodecyl-β-maltoside (β-DDM, Anatrace). LeuT was purified by Ni^{2+} affinity chromatography and spin-labeled with 0.35 mM *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanethiosulfonate spin label (MTSSL, Enzo Life Sciences) for 2 hours at room temperature and 4 °C overnight. Spin labeled LeuT was separated from free spin labels and aggregated protein using size exclusion chromatography performed on a Shodex KW-803 column in a buffer consisting of 200 mM Tris-MES buffer, 0.05% (w/v) β-DDM and 20% (v/v) glycerol at pH 7.2. LeuT was concentrated with Amicon Ultra columns (100 kDa, Millipore). For *n*-octyl-β-D-glucoside (β-OG, Anatrace) samples, Na^+ and excess Leu were added to Ni-affinity purified LeuT, the mixture purified by size exclusion chromatography in β-OG buffer (200mM Tris-MES pH7.2, 200mM NaCl, 1 μM leucine, 20% glycerol (v/v), 40mM (1.2% w/v) β-OG), and subsequently concentrated with Amicon Ultra columns (50 kDa, Millipore). Protein concentration was determined using an extinction coefficient of 1.91 cm^{-1} mg^{-1} at 280 nm for all mutants. All DEER samples were prepared in the 50 – 200 μM protein concentration range. A final concentration of glycerol of 30% (w/v) was used in all samples as a cryoprotectant. The Na^+ state was obtained by addition of 200mM NaCl. The Na^+/Leu state was obtained by adding 4-fold molar excess of Leu to protein, in addition to 200mM NaCl.

LeuT functional analysis

Equilibrium binding of ^{3}H-leucine (140 Ci/mmol; American Radiolabeled Chemicals, Inc.) at the indicated concentrations and specific radioactivities was performed with the scintillation proximity assay (SPA)[30]. 0.8 pmol of unconcentrated purified and spin-labeled LeuT were bound to 250 µg copper-coated YSi-SPA beads (Perkin Elmer) in 100 µL assay buffer (150 mM Tris, Mes, pH 7.5, 50 mM NaCl, 1 mM TCEP, 20% (v/v) glycerol, 0.1% (w/v) β-DDM) for 16 h at 4 °C prior to measuring the samples in a Wallac photomultiplier tube MicroBeta^{TM} microplate counter in the SPA mode. To determine the non-proximity background signal, samples were incubated in the presence of 800 mM imidazole, which competes with the His-tagged protein for binding to the copper-coated SPA beads. The non-proximity signal (in counts per minute, cpm), was subtracted from the total cpm (in the absence of imidazole) to obtain the specific cpm. Data points show the mean ± the SEM of triplicate determinations normalized as a percentage of WT. Experiments were conducted using 100, 1000, or 1500 nM ^{3}H-leucine. Saturation binding curves were constructed for a subset of representative LeuT mutants by varying ^{3}H-leucine concentration between 10 nM and 5 µM and are normalized as a percentage of WT. Curves were fitted using the non-linear curve fit, one site binding function in Origin 8 (OriginLab).

DEER spectroscopy

Distance measurements were conducted on a Bruker 580 pulsed electron paramagnetic resonance (EPR) spectrometer operating at Q-band frequency (33.9 GHz) using a standard four-pulse DEER sequence as previously described[31]. All DEER experiments were performed at 83 K. The frequency difference between pump and observed was typically 63 MHz. Dipolar evolution times were designed to allow identification of background slopes, when possible. Echo decays were shortened by 500ns to remove the baseline step that results from overlap between pump and observe pulse as previously described[32]. background-corrected and fit with the DEER Analysis 2011 program[33] using Tikhonov regularization[34] to obtain distance distributions.

Aggregated protein, resulting from concentration and validated by gel electrophoresis, appears in some samples as a non-specific distribution peak near 50 Å. This peak shifts depending on the decay time of the echo suggesting a broad distribution. This assignment was confirmed by re-analysis of the sample following DEER measurements by size-exclusion chromatography.

Rotamer simulation

Distance distributions for each mutant were simulated for LeuT crystal structures (PDB ID: 2A65, 3TT1, 3TT3) using the rotamer library-based prediction software MMM 2011[35-36]. Rotamer library calculations were conducted at 83K.

Results

To define the equilibrium of LeuT between conformational states and investigate its modulation by Na^+ and substrate, we measured distance distributions between spin label pairs under ligand conditions expected to promote transitions between transport intermediates[23-24]. An extensive set of spin label pairs was designed to present a comprehensive view of the extracellular and intracellular sides of the transporter. This set included pairs in TMs identified as focal points of structural rearrangements in LeuT crystal structures[6,13] (highlighted in Figs. 4.1a and 4.2a). We verified that all spin-labeled LeuT mutants introduced in a WT background bind leucine (Supp. Fig. 4.1) in a Na^+-dependent manner. We observed changes in the level of binding relative to the WT (dashed line, Supp. Fig. 4.1) for mutants in TM6 as well as for spin label pairs introduced in a Y268A background or an R5A background (red bar graph, Supp. Fig. 4.1). These two background mutations were constructed to partially mimic the disruption of the intracellular gate in the inward-facing LeuT crystal structure[24,37]. Analysis of binding isotherms for selected mutants demonstrates that the lower level of binding reflects reduced affinity but similar stoichiometry relative to the WT (Supp. Fig. 4.2).

LeuT conformational equilibrium: modulation by Na^+ and substrate

Model-free analysis of DEER decays by Tikhonov regularization[34,38] yields distributions that describe the distance probabilities between a pair of spin labels. In addition to reporting the average distance, the width of these distance distributions reflects protein dynamic modes that modulate the distance between the two spin labels[19,39-40]. Dynamics at room temperature is manifested as static disorder in the solid state conditions under which the DEER data are collected. Thus, the broad and multi-component distance distributions between pairs of labeled residues on the extracellular side of LeuT (Fig. 4.1) are consistent with a highly dynamic transporter undergoing fluctuations between multiple conformations in equilibrium. To assign distance components, we used different ligand conditions to enhance the populations of transporters in particular conformations. In the presence of Na^+/Leu (red traces, Fig. 4.1), the distribution would be expected to favor the state captured by the crystal structure of Na^+/Leu bound LeuT (PDB ID: 2A65)[6] classified as substrate-occluded. In the presence of Na^+ (blue traces, Fig. 4.1b), the transporter would be expected to favor an outward-facing conformation[23] poised to bind substrate. In the absence of ion and substrate (apo condition), LeuT is expected to sample inward-facing, outward-facing, and occluded conformations. These predictions are supported by previous EPR solvent accessibility measurements as well as SM-FRET experiments described in Chapter 1.

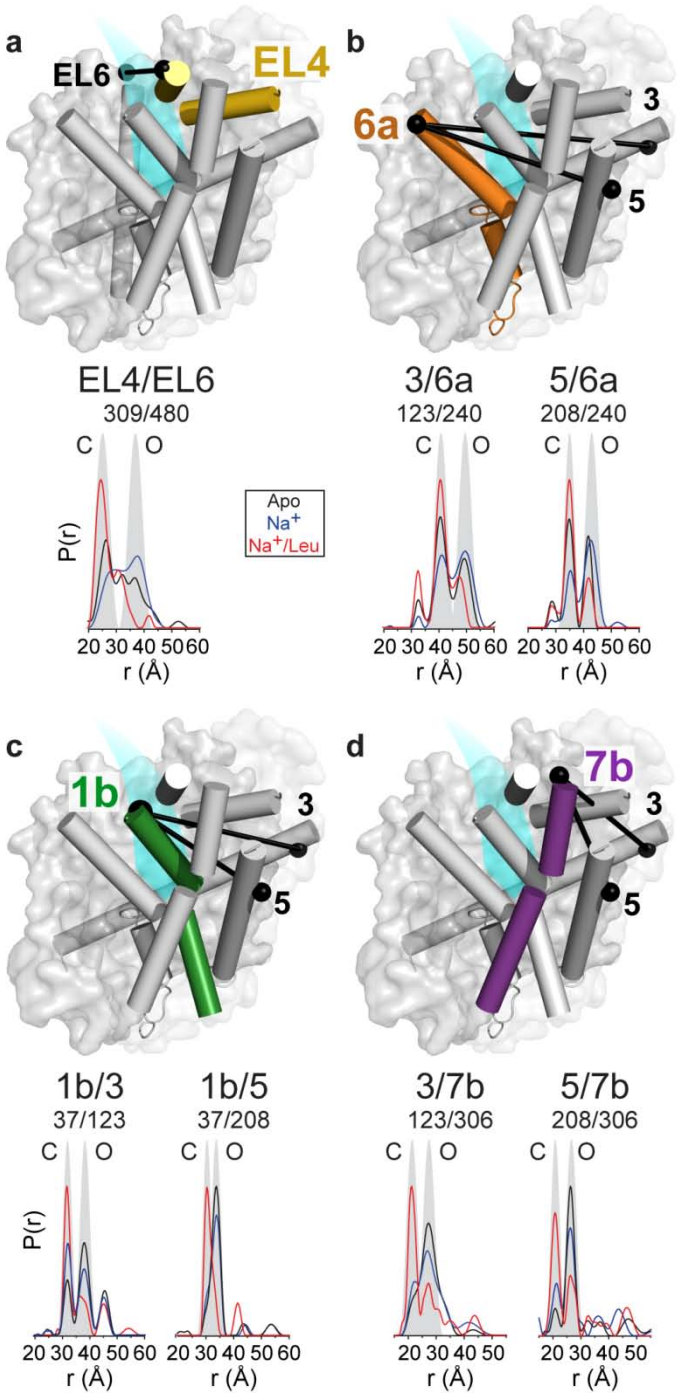


Figure 4.1. Na^+-induced opening and Na^+/substrate-induced closing of the LeuT extracellular side. a-d, Sites for distances measurements are shown for each colored helix on the 2A65 structure with distance measurements represented by a solid line. Distance distributions for each pair were obtained in the apo, Na^+-bound (Na^+), and Na^+- and Leu-bound (Na^+/Leu) intermediates. The multi-component distributions reflect multiple conformations of LeuT in equilibrium. For illustration, we simulated the distance components corresponding to the outward-open (O) and outward-closed (C) conformations using the average distance and width of each component. The resulting Gaussians are superimposed in gray. The shift in the conformational equilibrium of EL4, TM6a, TM1b, and TM7b relative to static reference points are shown in a, b, c, and d respectively.

Patterns of changes in distance distributions (Fig. 4.1 a-d) point to movements associated with Na^+- and Leu-dependent opening and closing of the extracellular side of LeuT. The analysis identifies two structural motifs, one consisting of EL4 and TM6a (Fig. 4.1, a and b), and the other consisting of TMs 1b and 7b (Fig. 4.1, c and d). The first motif responds to Na^+ binding by an "outward-opening" that increases the population of the longer distance component in the distance distributions relative to apo (Fig. 4.1). This component represents conformation(s) wherein TM6a moves away from reference points in TMs 3 and 5 while EL4 moves away from EL6. Notably, the second motif is relatively less sensitive to Na^+, suggesting that even under *apo* conditions it already favors an outward-open conformation (Fig. 4.1). Thus, Na^+ binding biases LeuT towards an outward-open conformation(s) in which both of the interrogated structural motifs are in their open positions. As previously reported[23], transition to an outward-open conformation involves movements of extracellular loops, including EL2, EL3 and EL4, and is accompanied by increased water accessibility in the permeation pathway that leads to the binding site.

We find that leucine binding reverses the Na^+-induced shift in equilibrium, consistent with a closing at the extracellular face of the transporter. Distance distributions report concurrent shifts in the equilibria of TMs 1, 6, and 7 and EL4 to favor the distance component associated with their closed positions (Fig. 4.1). In this presumably substrate-occluded conformation, TMs 1b, 6a and 7b move closer to reference points in TMs 3 and 5. For TMs 1b, 6a, 7b, and EL4, the magnitude of changes in the average distance between components corresponding to the outward-facing and substrate-occluded conformations is consistently larger than that predicted by comparing crystal structures of the states defined as outward-facing and substrate-occluded states (see below).

Similar interrogation of the dynamics at the intracellular side with a network of spin label pairs identifies TMs 6b and 7a and the N-terminal segment as undergoing the most substantial Na^+ and Leu-dependent movements. TM7a distance distributions are distinctly bimodal in the

apo conditions, reflecting the equilibrium between inward-open and inward-closed conformations (Fig. 4.2b). The distances between TM7 and reference points in TM4 and IL1 decrease with inward-opening. Na^+ and Na^+/Leu binding shift the equilibrium in the same direction, *i.e.* favor the same distance component, consistent with both conditions stabilizing an inward-closed conformation (Fig. 4.2b). The nature of the TM6b movement was more challenging to define because its buried environment hindered spin label incorporation at non-destabilizing, exposed sites (Supp. Figs. 4.1, 4.2, and 4.3). Nevertheless, spin label pairs monitoring distances from the C-terminal loop of TM6 (IL3) to TM4 show ligand-dependent changes in average distance and distribution width (Fig. 4.2a). The N-terminal segment (residues 1-10) displays a Na^+- and Leu-dependent shift between two populations that represent transition between the inward-open (longer distance component) and the inward-closed positions (shorter distance component) (Fig. 4.2c). Presumably, this movement, which was predicted by MD simulations[24] and previously reported by single-molecule FRET[24], is associated with release of a putative intracellular gate consisting of a network of charge interactions involving the N-terminus and IL1 and stabilized by Tyr268 in IL3[37] (Supp. Fig. 4.3a).

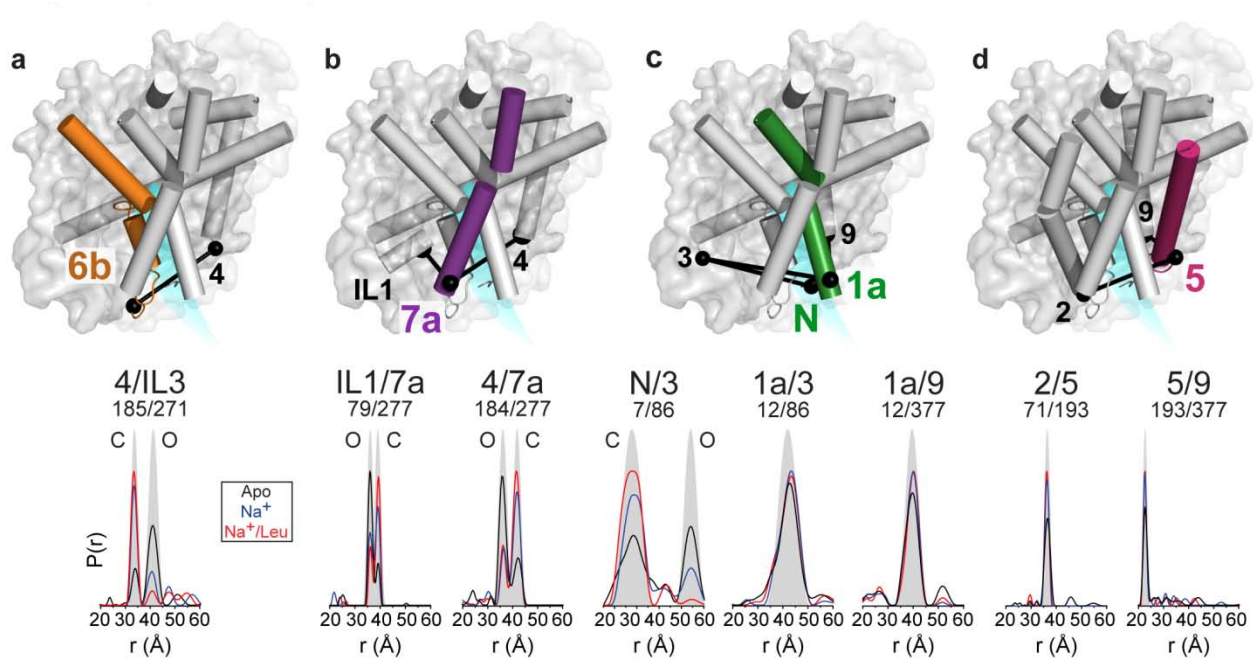


Figure 4.2.TMs 6, 7 and the N-terminal segment mediate the opening of the intracellular side. a-d, Sites for distances measurements are shown for each colored helix on the 2A65 structure with distance measurements represented by a solid line. Distance distributions for each pair were obtained under three conditions as in Fig. 4.1. a-b, IL3 and TM7 distributions indicate an equilibrium between two conformations that is modulated by Na^+ and substrate binding. c, In contrast, TM1 distributions are broad suggesting a dynamic helix but we did not detect a component consistent with the large conformational change suggested by the crystal structure. d, TM5 distributions are narrow and do not indicate ligand-dependent conformational changes.

The pattern of distance changes identifies novel LeuT conformations

Surprisingly, distance distributions monitoring the intracellular side did not report large ligand-dependent changes in the positions of TMs 1a and 5 (Fig. 4.2c,d). This is in stark contrast to inferences, on the basis of the inward-facing crystal structure[13], of a large displacement of TM1a away from TM3 and TM9 that lifts it upward towards the middle of the membrane, and of a sizeable translation of TM5 relative to a scaffold of helices. TM1a distance distributions did not show components that would correspond to such a large amplitude movement (Fig. 4.2c), although changes in distribution widths indicated small scale adjustments of TM1's position relative to TMs 3 and 9. TM5 distributions were consistently narrow and the relative distance to other helices did not change as a result of Na^+ or Na^+/Leu binding (Fig.

4.2d). Thus, the structural rearrangements implied by the inward-open crystal structure[13] do not appear to be sampled to a detectable extent under equilibrium in the WT background.

Together, the apparent discrepancies we detect between the amplitude of movements on the extracellular side relative to inferences based on comparison of the crystal structures, and in the identity of the mobile structural elements on the intracellular side, are puzzling in light of the mechanistic interpretation of the existing crystal structures. To quantitatively compare distance distributions measured here with the distance distributions expected in the various crystals structures, we employed a benchmarked spin label rotamer library[35-36]. In this approach, spin label rotamers, computed from MD simulations, are modeled at the sites of interest. The probability of each rotamer is weighted by a Boltzmann distribution using van-der-Waals interactions between the rotamer and the protein as the energy term. Distance distributions are then calculated from distances between all rotamer pairs (Supp. Fig. 4.4). Comparison of the predicted (from analysis of the crystallographic data) and experimental distance distributions obtained here reveals a clear pattern of inconsistencies on the extracellular side. Specifically for distributions between either TMs 1 or 6 and TMs 3 and 5, the distance component assigned to the occluded conformation (*i.e.* solid red trace favored by Na^+/Leu, Fig. 4.1b,c) is consistently shorter than any component in the predicted distributions based on the 2A65 structure (Supp. Fig. 4). Thus, we infer that this crystal structure underestimates the closing of TMs 1b and 6a induced by Na^+ and Leu binding in solution. Furthermore, the movement of TM7b, implied by the change in the distance distribution observed in the presence of Na^+/Leu (Fig. 4.1d), is not predicted from comparison of the predicted distributions between TM7 and TMs 3 and 5 in the crystal structures of outward-open[13] and substrate-occluded[6] states. Finally, while comparison of these crystal structures suggests movements of TM2 relative to TMs 3 and 5, we observed a tight distribution for these distances that is similar in the apo, Na^+-bound and Na^+/Leu- bound conditions (Supp. Fig. 4.4).

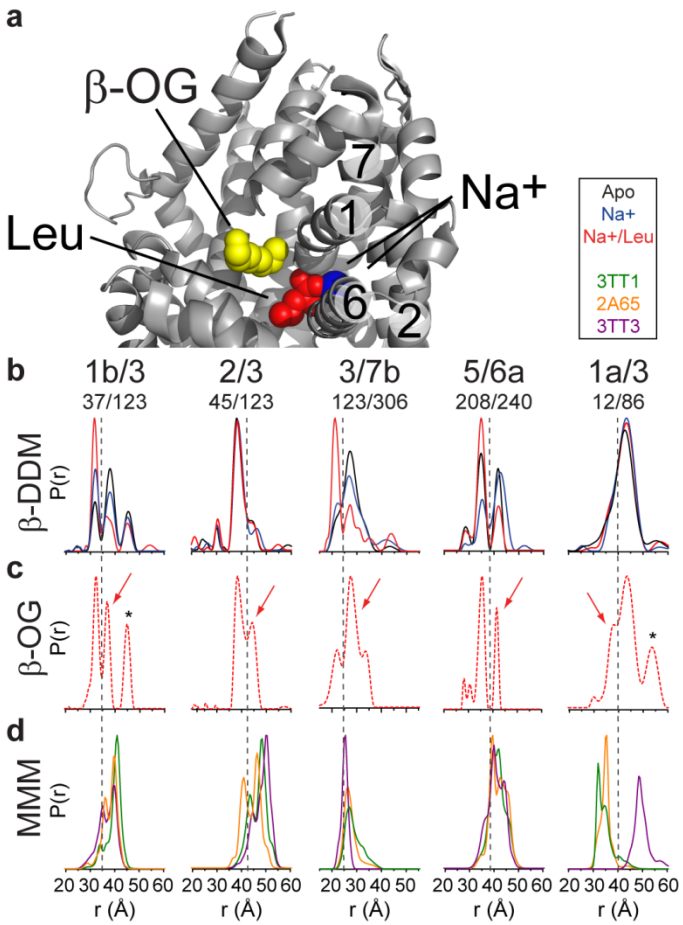


Figure 4.3. β-OG stabilizes the outward-facing conformation of LeuT in the presence of Na^+ and leucine. a, Close-up view of LeuT extracellular vestibule showing the simultaneous binding of leucine (red), Na^+ (blue) and β-OG (yellow) (PDB ID: 3GJD). Comparison of distance distributions in β-DDM (b) and β-OG. Dotted lines allow simultaneous comparison. (c) demonstrate that the latter stabilizes an outward-facing conformation on the extracellular side and a closed conformation on the intracellular side (TM1a/TM3). The corresponding distance component is indicated by an arrow. The component labeled * arises from aggregated protein during the concentration process. This was verified by sized exclusion chromatography where the aggregated protein lead to an asymmetric elution peak (d) Predicted distance distributions from three LeuT crystal structures (3TT1: outward-facing, 2A65: substrate-occluded, 3TT3: inward-facing) using MMM.

We reasoned that β-OG, the detergent used in most crystallization conditions[6,13] might be responsible for the discrepancies between our measurements of extracellular occlusion and those predicted by the crystal structures. Indeed, we found that exchange of LeuT from DDM into β-OG (Fig. 4.3a) shifts the experimental distance distributions for selected TMs 1, 2, 6 and 7 (Fig. 4.3b) towards the distance component assigned to the outward-open conformation (red arrows, Fig. 4.3c), thereby partly or fully resolving the quantitative discrepancy with predicted

distributions (Fig. 4.3d). This is consistent with the notion that a more outward-open conformation is favored in β-OG, presumably as a result of the binding of a β-OG molecule in the extracellular vestibule (Fig. 4.3a)[22]. Together, these findings are consistent with 2A65 representing an outward-facing "occluded" state and not a fully occluded configuration that we observe in solution.

Previous studies suggested an inhibitory effect of β-OG on substrate binding and flux[22,25]. In an attempt to rule out the possibility that β-OG traps LeuT in an inhibited conformation, crystal structures of Na^+/Leu-bound LeuT, which was never exposed to β-OG, have been determined in bicelles and found to adopt similar structures (1Å RMSD) to LeuT purified in β-OG[41]. However, in many of these structures a density was still observed in the extracellular vestibule and considered to be an alkyl chain of lipids or detergents. Thus, it is possible that a general mechanism of inhibition involves the binding of bulky molecules, rather than substrate, in the extracellular vestibule. On the other hand, the tendency to trap an outward-facing conformation may possibly reflect crystal lattice contacts which were nearly identical in the unit cells of the two bicelle structures[41].

Comparing predicted and experimental distributions on the intracellular side reveals a pattern of inconsistencies essentially opposite to that on the extracellular side. Thus, the components in the distance distributions corresponding to the Na^+/Leu state for TMs 1 and 2 tend to be larger than those predicted from the substrate-occluded structure (2A65)[6] (Supp. Fig. 4.5). The direction of the deviation and the effects of Na^+/Leu suggest that TM1 favors a more open conformation relative to that observed in the outward-facing[13] or substrate-bound[6] crystal structures. In contrast to the extracellular side, the shift in the distance distributions upon exchange into β-OG favors the shorter distance component (red arrow, Fig. 4.3b,c), which overlaps with the distribution predicted based on the occluded crystal structure determined in the presence of Na^+/Leu (Fig. 4.3d). Thus the LeuT conformation in β-OG is more outward-open/inward-closed - both in the 2A65 structure and in our DEER measurements - than is the

case for our measurements of the substrate-bound state in the absence of β-OG. This suggests that the actual substrate-occluded state is more closed at the extracellular side and more open at the intracellular side than the 2A65 structure, as is particularly notable for the TM1a-TM3 distance distribution.

The Y268A mutation uncouples helical movements from Na^+ and Leu binding

The capture of LeuT in the inward-open crystal structure required multiple mutations to disrupt the intracellular gate and to weaken the Na2 site[13]. The former was achieved through the substitution of a highly conserved tyrosine (Y268) in TM6b with an alanine. Therefore, we monitored the distance distributions of TM1a and TM5 on the intracellular side in a Y268A background (Fig. 4.4). The pattern of distance changes in this mutant background recapitulates many aspects of the inward-facing[13] crystal structure (Fig. 4.4a). TM1 undergoes a 15 Å change in distance relative to TM9 (dashed lines, Fig. 4.4b,c) that lifts it towards the middle of the membrane leading to a short distance component relative to the extracellular side of TM8 (Supp. Fig. 4.6). The marked displacement of TM5 in the inward-facing crystal structure (Fig. 4.4a) is also captured by the distance distribution to TM9 (Fig. 4.4b,c). However, this "opening" movement of TM5 is not reversed by the binding of Na^+ and Leu (Fig. 4.4c, red dashed traces), despite direct biochemical evidence that these mutants bind Leu (Supp. Figs. 4.1 and 4.2). Thus, in the mutant background, TM5 does not reset to its closed position. Another mutation that was shown to disrupt the intracellular gate, R5A (Fig. 4.4a)[24,37], yields similar distance changes (Fig. 4.4b,c). Thus, the mutations used to generate the inward-open[13] crystal structure shift a major population of the transporters to a conformation that is not readily observed in the ensemble of states sampled in the WT background.

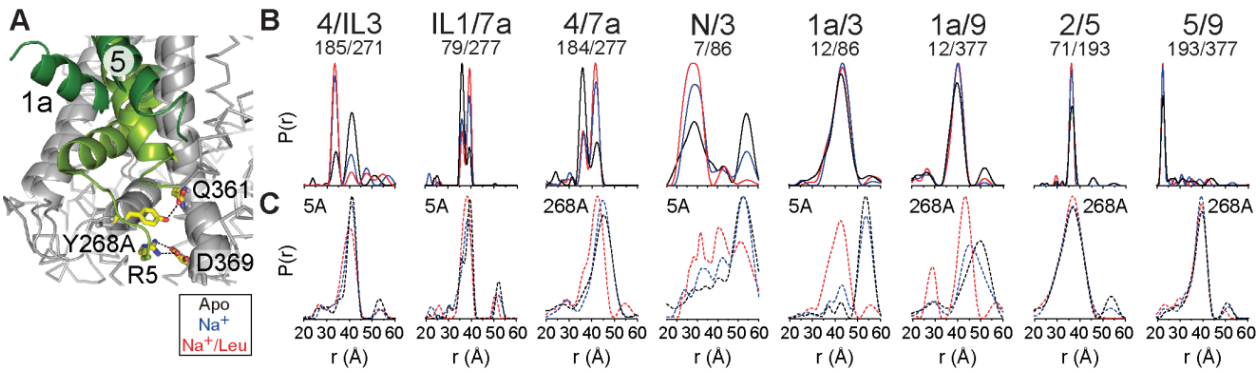


Figure 4.4. The Y268A or R5A mutations induce structural rearrangements in LeuT. a, Close up view of the putative intracellular gate showing the network of charge interactions stabilized by Y268 and involving R5. TMs 1, 5, 6 and 7 are shown in green. b, the mutations Y268A and R5A lead to the appearance of new distance components in the distributions of TMs 1 and 5 (dashed lines) under apo conditions. These components are consistent with the direction and relative amplitude of the structural changes observed in the inward-facing crystal structure (PDB ID: 3TT3). Note that many of these movements are partially reset by Na^+ and leucine.

Discussion

Structural motifs underlying alternating access of LeuT

The pattern of distance changes described above identifies movements of motifs at the extracellular (1b/7b and 6a/EL4) and intracellular ends (6b/7a, N-term) of LeuT. The ligand-dependence of the distance changes reveals how binding of Na^+ and substrate shifts the opening/closing equilibria of these motifs, thereby driving conformational transitions between outward-facing, inward-facing and substrate-occluded states (Fig. 4.5). In the absence of ligands, the motifs sample open and closed positions, thus enabling apo-LeuT to isomerize between the three canonical conformations as would be expected for a symporter. Binding of Na^+, which engages the discontinuous segments of TMs 1 and 6, concomitantly shifts the equilibria of the extracellular and intracellular motifs in opposite directions, effectively favoring an outward-open/inward-closed conformation of LeuT. Together, these observations suggest that the rearrangements of these motifs are mechanistically important for the alternating access mechanism in LeuT.

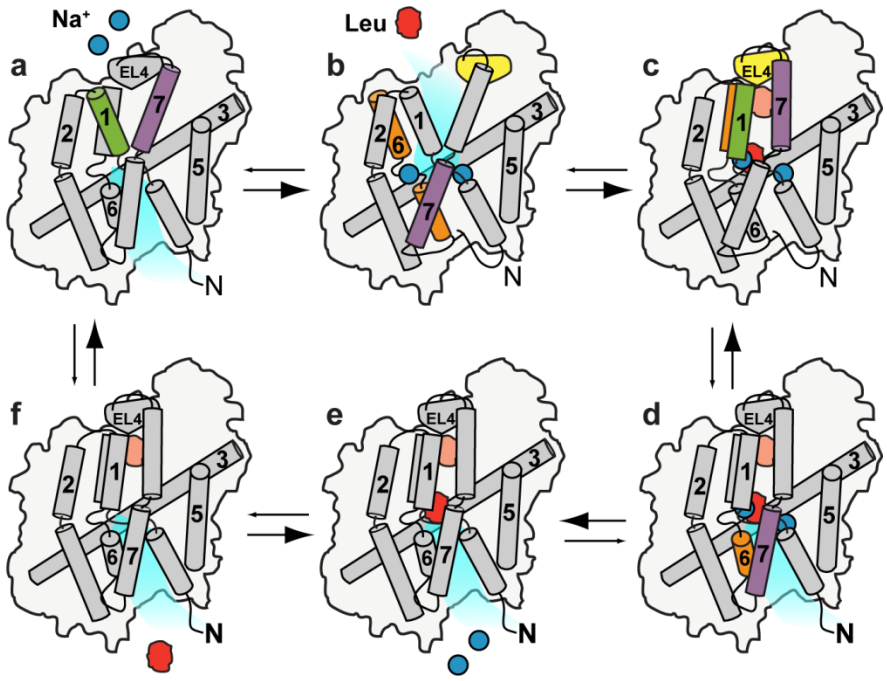


Figure 4.5. EPR-derived model of LeuT transport. The cycle begins following release of ion and substrate to the intracellular side. Apo-LeuT samples inward- and outward-facing conformations. Na^+ binding induces opening of the extracellular side through shifts in the equilibrium of the extracellular motifs. Coupled closing of the intracellular side involves closing of the intracellular motif, which stabilizes the intracellular gate. Leu binding at the S1 site, and presumably at the S2 site as well, causes a large scale closure of the extracellular side leading to an occluded state. Fluctuations on the intracellular side, facilitated by the unwound region of TM6 and a kink at Gly294 of TM7 mediate the opening of the intracellular side. Driven by its concentration gradient, Na^+ dissociates to the intracellular solution. In the absence of bound Na^+, leucine affinity to LeuT is reduced driving it dissociation to the intracellular side. The cycle continues through the isomerization from inward-facing to outward-facing.

One of the novel findings of this work is a ligand-dependent coupling of the extracellular side of the transporter with the intracellular side that involves TMs 6 and 7 (Fig. 4.5). The coupled but opposite shifts in the equilibria of TM6a and TM7a ensure that the intracellular motif and the extracellular motifs are not concurrently open. Because the distance between TMs 1b and 6a does not change upon ligand binding (Supp. Fig. 4.4), we propose that the movement of TM6a to its open position involves pivoting around TM1b, the equilibrium of which remains relatively unchanged in the transition between apo and Na^+-bound LeuT (Fig. 4.5). The coupled movement on the intracellular side results from TM7a favoring its closed position while TM7b on the extracellular side maintains its open position under apo conditions.

Substrate, which is primarily coordinated by the unwound regions of TMs 1 and 6 in the primary binding site (S1), resets these two TMs to their closed positions (Fig. 4.5). Because occupancy of the extracellular vestibule was shown to affect TM7 conformation (Fig. 4.3), we surmise that binding of Leu in a secondary substrate binding site (S2), which was previously proposed from functional and computational studies[25], may forge interactions with EL4 and by extension affect TM7's equilibrium. Consistent with this notion, substrate in the S2 site has been modeled as contacting TMs 1, 10 and EL4.

The structural scaffold, shown in gray half-cylinder in Figure 4.5 and in the C structure in Supplementary Figure 4.7a, supports the movement of the motifs we identified. Small ion/substrate dependent movements are noted for TMs 1, 10 and 11 on the intracellular side; the latter two most likely serve to facilitate the movement of the intracellular motif that they cradle (Supp. Fig. 4.7b). TM1's more open position relative to the crystal structure is consistent with MD simulations[42] that invoke TM1 movement in facilitating substrate release. Notably, these simulations were initiated from the Na^+/Leu crystal structure; so that the dynamic nature of TM1 in the simulations may reflect both the relaxation to its position in the occluded conformation in the absence of β-OG and the response to the progress of substrate towards the intracellular release site. Indeed, the broad distributions of TM1 suggest that this helix is relatively dynamic and with TM1a favoring a more open position relative to the 2A65 structure, the movement of the intracellular motif (Fig. 4.5) enables substrate exit on the intracellular side of the transporter.

The data presented here reinterpret the model of alternating access deduced from the crystal structures, but the proposed motif movements are well-grounded in aspects of TM flexibility that have been inferred from these structures. Specifically, the pivot points observed crystallographically in TMs 1, 6, and 7 rationalize the hinge-like bending of these helices inferred from our results (Fig. 4.5). While our data demonstrate that the crystal structure initially classified as substrate-occluded[6] is more outward-facing and inward-closed than is the

substrate-occluded conformation we measure in solution, the small scale structural changes relating it to the crystallographic outward-facing structure involve bending of TMs 1, 6 and movement of EL4. Similarly, bending of TM7 at the conserved glycine 294 has been observed in the inward-facing[13] crystal structure.

Mechanism of LeuT transport

Coupled transport requires the binding of ions and substrates on one side of the membrane, followed by conformational changes to reorient the binding site that enable subsequent ion and substrate dissociation at the other side of the membrane. To power the uphill movement of substrates, the presence or absence of the symported ions could stabilize and destabilize substrate binding in the outward- and inward-facing conformations, respectively, conferring directionality to the transport process by their concentration gradients. Thus, low Na^+ occupancy during the release step is promoted by the low concentration of Na^+ in the cytoplasm. Additionally, ion binding could alter the energetics of the equilibrium among conformations in the ensemble of conformations, increasing the equilibrium population of those conformations where the substrate site is open to the extracellular milieu. In LeuT, Na^+ fulfills both roles: its binding biases the equilibrium to favor an outward-facing conformation and it subsequently directly coordinates the substrate at the unwound regions of TMs 1 and 6.

We and others[23-24] find that substrate binding induces occlusion of the binding site on both sides of the transporter (Fig. 4.5). Thus, the energetically unfavorable release of Leu, has to be facilitated by low probability equilibrium fluctuations of the intracellular motif (TMs, 6, 7 and N-terminus) to its open position. We surmise that it is these fluctuations that must be coupled to the favorable movement of Na^+ down its concentration gradient, a dissociation event that facilitates formation of the inward-facing conformation. Previous studies proposed that the dissociation of Na2 in the presence of substrate in both the S1 and S2 sites enhances inward opening and facilitates inward release of the Na^+ in Na1[27]. The loss of this Na^+ reduces the

affinity of substrate, which enhances its dissociation to the cytoplasm. Krishnamurthy et al. hypothesized[13] that in the absence of bound Na^+, it is more energetically favorable for TM1 to lift toward the middle of the membrane, dragging along with it the unstructured N-terminus. This now seems unlikely given the absence of this state under apo conditions in solution and its presence only in the Y268A mutant background.

Within the outward-facing or inward-facing ensembles of conformational states investigated and redefined from the results presented here, there are likely to exist yet other sub-states in which side chain rearrangements stabilize and coordinate ligand or induce their release[14-15]. These substates are not differentiated by large amplitude distance changes, and would be obscured by the inherent limitations of probe-based methods such as EPR. Similarly, high energy states that involve major distortions in the structure (such as transitions states) are likely to be fleetingly populated, resulting in an equilibrium population below the data noise level. However, the new insight offered by the data presented here suggests a novel set of conformational changes, which were not observed in the available crystal structures but serve well in reinterpreting the structural information in describing the classic alternating access model.

References

1. Mitchell, P. A general theory of membrane transport from studies of bacteria. *Nature* **180**, 134-136 (1957).
2. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969-970 (1966).
3. Patlak, C. S. Contributions to the theory of active transport: Ii. The gate type non-carrier mechanism and generalizations concerning tracer flow, efficiency, and measurement of energy expenditure. *Bull. Math. Biophys.* **19**, 209-235 (1957).
4. Vidaver, G. A. Inhibition of parallel flux and augmentation of counter flux shown by transport models not involving a mobile carrier. *J Theor Biol* **10**, 301-306 (1966).
5. Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta* **1807**, 167-188 (2011).
6. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of na+/cl--dependent neurotransmitter transporters. *Nature* **437**, 215-223 (2005).
7. Zomot, E., Bendahan, A., Quick, M., Zhao, Y., Javitch, J. A. & Kanner, B. I. Mechanism of chloride interaction with neurotransmitter:Sodium symporters. *Nature* **449**, 726-730 (2007).
8. Forrest, L. R., Tavoulari, S., Zhang, Y. W., Rudnick, G. & Honig, B. Identification of a chloride ion binding site in na+/cl -dependent transporters. *Proc Natl Acad Sci U S A* **104**, 12761-12766 (2007).
9. Kantcheva, A. K., Quick, M., Shi, L., Winther, A. M., Stolzenberg, S., Weinstein, H., Javitch, J. A. & Nissen, P. Chloride binding site of neurotransmitter sodium symporters. *Proc Natl Acad Sci U S A* **110**, 8489-8494 (2013).
10. Singh, S. K., Yamashita, A. & Gouaux, E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448**, 952-956 (2007).
11. Zhou, Z., Zhen, J., Karpowich, N. K., Goetz, R. M., Law, C. J., Reith, M. E. & Wang, D. N. Leut-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317**, 1390-1393 (2007).
12. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347-355 (2009).
13. Krishnamurthy, H. & Gouaux, E. X-ray structures of leut in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469-474 (2012).
14. Watanabe, A., Choe, S., Chaptal, V., Rosenberg, J. M., Wright, E. M., Grabe, M. & Abramson, J. The mechanism of sodium and substrate release from the binding pocket of vsglt. *Nature* **468**, 988-991 (2010).
15. Perez, C., Koshy, C., Yildiz, O. & Ziegler, C. Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter betp. *Nature* **490**, 126-130 (2012).
16. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a na+-independent amino acid transporter. *Science* **325**, 1010-1014 (2009).
17. Weyand, S., Shimamura, T., Beckstein, O., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter mhp1. *J Synchrotron Radiat* **18**, 20-23 (2011).
18. Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. Molecular basis of

alternating access membrane transport by the sodium-hydantoin transporter mhp1. *Science* **328**, 470-473 (2010).
19. Mchaourab, H. S., Steed, P. R. & Kazmier, K. Toward the fourth dimension of membrane protein structure: Insight into dynamics from spin-labeling epr spectroscopy. *Structure* **19**, 1549-1561 (2011).
20. Freed, D. M., Horanyi, P. S., Wiener, M. C. & Cafiso, D. S. Conformational exchange in a membrane transport protein is altered in protein crystals. *Biophys J* **99**, 1604-1610 (2010).
21. Cross, T. A., Sharma, M., Yi, M. & Zhou, H. X. Influence of solubilizing environments on membrane protein structures. *Trends Biochem Sci* **36**, 117-125 (2011).
22. Quick, M., Winther, A. M., Shi, L., Nissen, P., Weinstein, H. & 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 U S A* **106**, 5563-5568 (2009).
23. Claxton, D. P., Quick, M., Shi, L., de Carvalho, F. D., Weinstein, H., Javitch, J. A. & Mchaourab, H. S. Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:Sodium symporters. *Nat Struct Mol Biol* **17**, 822-829 (2010).
24. Zhao, Y., Terry, D., Shi, L., Weinstein, H., Blanchard, S. C. & Javitch, J. A. Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188-193 (2010).
25. Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J. A. The mechanism of a neurotransmitter:Sodium symporter--inward release of na+ and substrate is triggered by substrate in a second binding site. *Mol Cell* **30**, 667-677 (2008).
26. Forrest, L. R., Zhang, Y. W., Jacobs, M. T., Gesmonde, J., Xie, L., Honig, B. H. & Rudnick, G. Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci U S A* **105**, 10338-10343 (2008).
27. Zhao, Y., Terry, D. S., Shi, L., Quick, M., Weinstein, H., Blanchard, S. C. & Javitch, J. A. Substrate-modulated gating dynamics in a na+-coupled neurotransmitter transporter homologue. *Nature* **474**, 109-113 (2011).
28. Hubbell, W. L., Mchaourab, H. S., Altenbach, C. & Lietzow, M. A. Watching proteins move using site-directed spin labeling. *Structure* **4**, 779-783 (1996).
29. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **9**, 1895-1910 (2007).
30. Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proc Natl Acad Sci U S A* **104**, 3603-3608 (2007).
31. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *Journal of Magnetic Resonance.* **142**, 331-340 (2000).
32. Jeschke, G. Deer distance measurements on proteins. *Annu Rev Phys Chem* **63**, 419-446 (2012).
33. Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J., Timmel, C. R., Hilger, D. & Jung, H. Deeranalysis2006—a comprehensive software package for analyzing pulsed eldor data. *Applied Magnetic Resonance* **30**, 473-498 (2006).
34. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed esr using tikhonov regularization. *Journal of Magnetic Resonance* **172**, 279-295 (2005).
35. Polyhach, Y., Bordignon, E. & Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys Chem Chem Phys* **13**, 2356-2366 (2011).

36. Polyhach, Y., Godt, A., Bauer, C. & Jeschke, G. Spin pair geometry revealed by high-field deer in the presence of conformational distributions. *J Magn Reson* **185**, 118-129 (2007).
37. Kniazeff, J., Shi, L., Loland, C. J., Javitch, J. A., Weinstein, H. & Gether, U. An intracellular interaction network regulates conformational transitions in the dopamine transporter. *J Biol Chem* **283**, 17691-17701 (2008).
38. Jeschke, G., Koch, A., Jonas, U. & Godt, A. Direct conversion of epr dipolar time evolution data to distance distributions. *J Magn Reson* **155**, 72-82 (2002).
39. Georgieva, E. R., Borbat, P. P., Ginter, C., Freed, J. H. & Boudker, O. Conformational ensemble of the sodium-coupled aspartate transporter. *Nat Struct Mol Biol* **20**, 215-221 (2013).
40. Hanelt, I., Wunnicke, D., Bordignon, E., Steinhoff, H. J. & Slotboom, D. J. Conformational heterogeneity of the aspartate transporter glt(ph). *Nat Struct Mol Biol* **20**, 210-214 (2013).
41. Wang, H., Elferich, J. & Gouaux, E. Structures of leut in bicelles define conformation and substrate binding in a membrane-like context. *Nat Struct Mol Biol* **19**, 212-219 (2012).
42. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the leut and apct transporters are modulated by common mechanisms. *Biophys J* **99**, L103-105 (2010).

CHAPTER 5

Mhp1 CONFORMATIONAL EQUILIBRIA REVEAL DIVERSITY OF LeuT FOLD PROTEINS

Abstract

Ion-coupled transporters of the LeuT Fold class couple the energetically unfavorable transfer of physiologically essential molecules such as amino acids, sugars, nucleobase precursors, and osmolytes to transmembrane ion gradients. This structural class is defined by a conserved 5-helix inverted repeat that encodes common principles of ion and substrate binding. Representative members such as Mhp1, LeuT, and BetP are structurally well characterized, with published crystal structures outlining outward-facing, occluded, and inward-facing conformations. However, fundamental questions relating to the dynamics of transport in this class remain unanswered. Specifically, to what degree do the captured crystal structures represent stable intermediates of the transport cycle? How is transport of substrate coupled to the binding and release of co-transported ion? Do LeuT Fold proteins operate through a unified transport mechanism? To begin to address these questions, we have used distance measurements between pairs of spin labels to define the conformational dynamics of the Na^+-coupled symporter, Mhp1. Our results support the assertion that the inward-facing and outward-facing Mhp1 crystal structures represent sampled intermediate states in solution. Here, we provide mechanistic context for these structures, mapping them into a novel transport cycle based on ion- and substrate-dependent conformational equilibria. In contrast to LeuT, our results suggest an absence of Na^+-dependent conformational change in Mhp1. A comparative analysis suggests that the conserved Na2 site stabilizes substrate binding providing a direct mechanism of coupling, while Na1 site, present only in LeuT and BetP, promotes conformational transitions that effect efficient substrate binding. We postulate that these

differential ion coupling mechanisms may define functional subfamilies within the LeuT Fold class arguing against a unified mechanism.

Introduction

The LeuT Fold has emerged as the structural scaffold for an increasing number of sequence-unrelated transporters (Fig. 5.1). The fold includes Na^+-coupled symporters such as the leucine transporter[1] (LeuT) from the neurotransmitter:Na^+ symporter (NSS) family, the galactose transporter[2] (vSGLT) from the Na^+:solute symporter (SSS) family, the betaine transporter[3] (BetP) from the betaine/carnitine/choline transporter (BCCT) family, and the benzylhydantoin transporter[4] (Mhp1) from the nucleobase:cation symporter 1 (NCS1) family, as well as the H^+-coupled amino acid transporter ApcT[5], the H^+-dependent carnitine exchanger CaiT[6-7], and ion-independent exchangers AdiC[8-9] and GadC[10]. These families participate in numerous physiologically important processes including neurotransmitter reuptake[11], glucose regulation[12], osmotic stress response[13], nucleobase recycling[14-15], and acid resistance. Due to this functional diversity, understanding the mechanistic implications of the LeuT Fold is relevant to an array of neurological[16], metabolic[17-21], and infectious diseases[13].

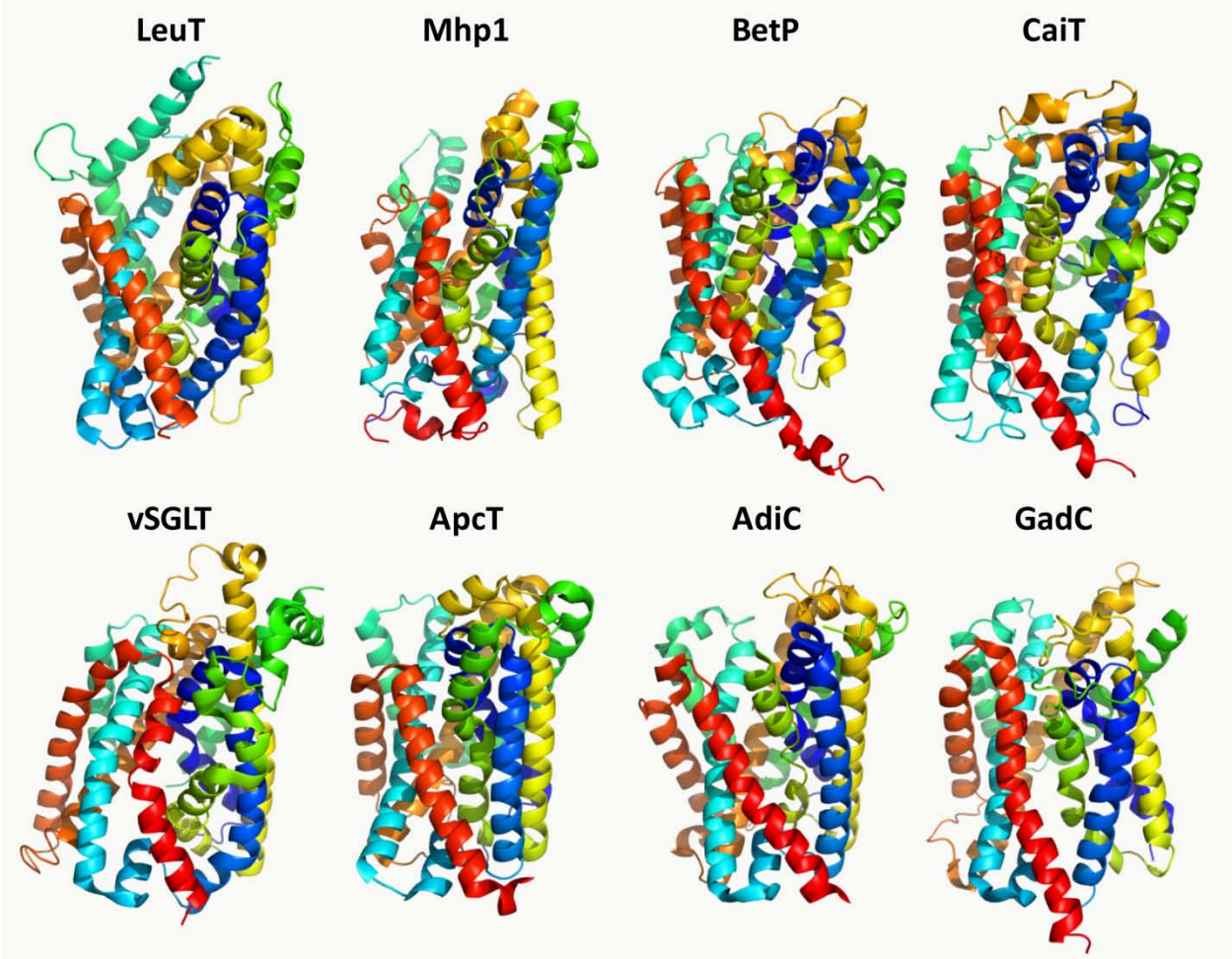


Figure 5.1. Structural comparison of core helices of LeuT Fold members. Core helices correspond to TMs 1-10 in LeuT.

The structural similarity of these diverse proteins can be rationalized by similarities in mechanisms of ion and substrate coordination inherent to the LeuT Fold architecture. The LeuT Fold is defined by two sets of five transmembrane helices (TMs 1-5 and 6-10, numbering defined henceforth as in LeuT) related by two-fold symmetry around an axis parallel to the membrane[1]. Discontinuous regions of symmetrically related TM helices 1 and 6 provide conserved interactions for ion and substrate binding sites[1]. The primary substrate binding site (S1) is located at the interface of these helices, in nearly identical positions in all LeuT Fold members. Although ion-dependence and stoichiometry varies among members, this variation

can be explained by deviation in key residues. The Na2 Na^+ binding site, coordinated by discontinuous TM1 and TM8, is conserved among all Na^+-coupled members, as well as the H^+-coupled ApcT and ion-independent CaiT, where this site is characterized by the presence of positively charged Lys and Arg residues[5,7]. This site has been shown to support substrate binding in Mhp1 evidenced by a 10-fold increase in substrate binding affinity in the presence of Na^+ [4]. It has also been suggested based on steered molecular dynamics simulations (MD) in LeuT that this site may be responsible for Na^+-dependent conformational change[22], although this assertion has not yet been supported experimentally. The Na1 site, located proximal to S1, is less well conserved structurally, identified in only LeuT[1] and BetP[3]. However, in CaiT[7] and ApcT[5], the sulfur atom of a Met residue provides similar interactions as the Na1 site. In each of these instances, Na^+ or Met at this site directly coordinates the carboxyl groups of substrates bound in S1. Despite the wealth of structural information, the mechanisms of ion coupling in the LeuT Fold class are not currently well understood and implications of the divergence in Na^+ binding sites in the class have not yet been established.

Common structural features of the LeuT Fold have also stimulated conjectures of a unified mechanism of alternating access[23-25]. Broadly defined, alternating access describes protein conformational transitions that allow solvent access to central ion and substrate binding sites alternatively from extracellular and intracellular sides of the membrane, thereby promoting ion and substrate translocation[26-29]. Mhp1, a Na^+-coupled symporter of *Microbacterium liquefaciens*, was the first LeuT Fold member to be characterized by a complement of canonical states representing outward-facing, inward-facing, and outward-facing occluded conformations purported to represent an alternating access cycle[30]. This series of structures identified a 4-helix bundle (TMs 1, 2, 6, and 7), referred to as the bundle motif, as the functional subdomain, the orientation of which defined directionality of solvent access. The model mechanism resulting from the Mhp1 structures has been referred to as a Rocking Bundle mechanism and was predicted based on the inherent inverted repeat symmetry of the LeuT Fold[31]. Due to the

pervasiveness of inverted repeat symmetry, the Rocking Bundle mechanism was projected to underlie alternating access for LeuT Fold members[31], as well as structurally dissimilar classes of symmetric transporters[24,32-34]. However, the publication of crystal structures representing various conformations of individual LeuT Fold members has complicated this interpretation. The diversity of the structural rearrangements implicit in these structures is seemingly inconsistent with a conserved conformational cycle. While reasonable attempts have been made to accommodate each of the existing crystal structure conformations into a unified mechanism of transport[24-25], it remains unclear whether a single mechanism should be assumed *a priori* for the LeuT Fold class.

Recent examples have also called into question the sufficiency of crystal structures to alone describe dynamic transport mechanisms. In particular, the series of LeuT crystal structures implied an entirely distinct, asymmetric transport mechanism characterized by limited extracellular conformational changes and a prominent displacement of intracellular TM1a away from the protein core into the membrane[35]. To define the mechanistic context of this LeuT suite of crystal structures, we recently investigated the ion- and substrate-dependent conformational equilibria of LeuT (Chapter 4). Our results suggested that the LeuT crystal structures were influenced by crystallographic conditions and by disruptive mutations introduced to promote conformational selection, which limited sampling of intermediates and resulted in alternate conformations not populated in the WT background. We proposed a novel model of transport that described previously unidentified inward-facing and substrate-occluded conformations as well as revealed specific shifts in conformational equilibria associated with Na^+ and substrate binding, which form the basis of the transport cycle. This work supported the mechanistic divergence of the LeuT Fold as well as underscored the importance of identifying the mechanistic identities of crystal structures and conformational equilibria between states to accurately define transport mechanisms.

Here, we demonstrate a similar analysis to that employed to define the conformational cycle of LeuT. Namely, we utilized site-directed spin labeling (SDSL)[36] and Double Electron-Electron Resonance (DEER) spectroscopy[37] to elucidate the conformational sampling, ion coupling, and ligand-dependent transport cycle underlying alternating access in Mhp1. We demonstrate that Mhp1 indeed samples the captured crystallographic conformations during transport, thereby generally supporting the Rocking Bundle mechanism in Mhp1. However, the model of Mhp1 transport that we propose depends on low probability transitions between states rather than triggered conformational changes. Specifically, our results indicate that Mhp1 operates without the Na^+-dependent conformational transitions described in LeuT. The results described here further support the conclusion that LeuT and Mhp1 operate using distinct conformational cycles and begins to elucidate the nature of this divergence as the result of the differential mechanisms of coupling to the Na^+ gradient as defined by Na^+:substrate stoichiometries.

Methods

Mutagenesis, expression, purification and labeling of Mhp1

The Mhp1 construct was engineered to be cysless (C69A, C234A and C327A) with a C-terminal decahistidine tag using gene synthesis (Genescript) and cloned into a pqo18 vector. Cysteine residues were introduced into the cysless construct using site-directed mutagenesis[36] and confirmed by DNA sequencing. Mhp1 mutants were expressed in *E. coli* C43(DE3) which was grown to absorbance A600 of 1.0 before induction with 0.5mM IPTG and shaken at 25 °C for 16 hours. Mhp1 was extracted from native membranes in 40 mM (2% w/v) β-DDM in 50 mM Tris-HCl, 200mM NaCl, and 20% glycerol (v/v) at pH 7.5 before purification with Ni^{2+} affinity chromatography. The protein was spin-labeled and purified using the same protocol as LeuT (Chapter 4, Methods) in a buffer consisting of 50mM Tris-MES, 0.05% (w/v) β-DDM and 20%

(v/v) glycerol at pH 7.2. Protein concentration was determined using an extinction coefficient of 1.84 cm^2 mg^{-1} at 280 nm. Purified Mhp1 were concentrated with Amicon Ultra columns (100 kDa, Millipore). Deer samples were prepared in the 50 – 200 μM protein concentration range. A final concentration of glycerol of 30% (w/v) was used in all samples as a cryoprotectant. The Na^+ state was obtained by addition of 200mM NaCl. The Na^+/BH state was obtained through addition of 5mM 5-benzyl hydantoin (BH, Toronto Research Chemicals) and 200mM NaCl.

Mhp1 functional analysis

Binding of BH to purified and spin-labeled mutant Mhp1 protein was monitored using a Trp fluorescence quenching assay[4]. Measurements were conducted using 2.5 μM Mhp1 and 2 mM BH in 50 mM Tris-MES, pH 7.2, 15 mM NaCl, 0.05% (w/v) β-DDM and 20% (v/v) glycerol containing buffer at room temperature. Samples were excited at 285 nm and fluorescence intensity was collected at 348 nm before and after addition of BH. The decrease in fluorescence is expressed as a percentage of peak height. Complete binding curves for WT and the 136/278 mutant were also obtained using varying of BH concentrations between 0.1 and 5 mM. Curves were fitted and K_d determined using the non-linear curve fit, one site binding function with no weights in Origin 8 (OriginLab).

DEER spectroscopy

Distance measurements were conducted on a Bruker 580 pulsed electron paramagnetic resonance (EPR) spectrometer operating at Q-band frequency (33.9 GHz) using a standard four-pulse DEER sequence as previously described[38]. All DEER experiments were performed at 83 K. Dipolar evolution times were designed allow identify background slope, when possible. Echo decays were background-corrected and fit with the DEER Analysis 2011 program[39] using Tikhonov regularization[40] to obtain distance distributions. Aggregated protein, resulting from

concentration and validated by gel electrophoresis, appears in some samples as a non-specific peak near 50 Å.

Results and Discussion

Mhp1 conformational transitions support the Rocking Bundle mechanism

Crystal structures have captured WT Mhp1 in inward-facing, outward-facing occluded and outward-facing conformations[4,30]. The primary geometric transformation relating outward-facing and inward-facing structures, involves the rigid body rotation of the bundle motif relative to the scaffold motif, termed the Rocking Bundle mechanism (Fig. 1.9). The outward-facing and outward-facing occluded structures differ only in their position of extracellular TM10, a putative gating helix, that physically blocks the extracellular vestibule in the occluded structure. The structurally equivalent intracellular TM5 deviates in position between the inward-facing and outward-facing structures and has also been suggested to participate in the gating mechanisms.

The magnitude and directionality of the crystallographic transition is shown in a RMSD comparison between the outward-facing occluded and inward-facing structures (Fig. 5.2). In this comparison, the width of the ribbon backbone connotes the deviation between the structures in Å for each residue position. To evaluate this Crystal Structure Model and describe Mhp1 transport cycle, a series of distance distributions were measured using DEER to reveal ligand-dependent shifts in conformational equilibria at various sites in Mhp1. These sites were specifically selected to limit functional perturbation and promote information content for future computational modeling investigations. All spin labeled Mhp1 mutants reported here bind the substrate benzylhydantoin (BH, Supp. Fig. 5.1).

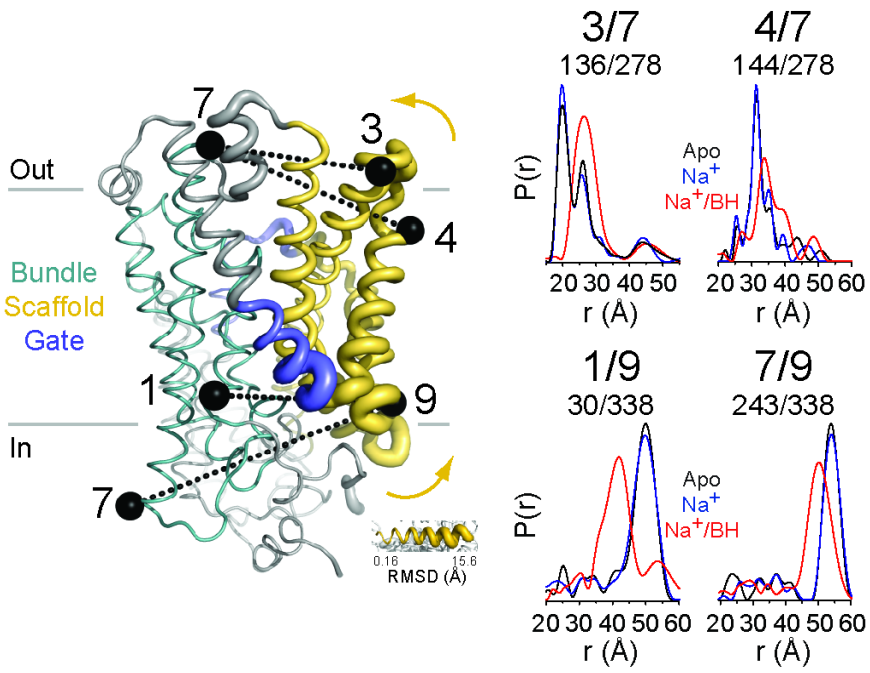


Figure 5.2. Conformational relationship between the bundle and scaffold motifs. (Left) RMSD between outward-facing occluded (PDB ID: 2JLO) and inward-facing (PDB ID: 2X79) Mhp1 crystal structures mapped onto the occluded structure. The bundle is shown in green, the scaffold in yellow, and putative gate helices in blue. The yellow arrows indicate the direction of the movement inferred from comparison of the crystal structures. The locations of representative spin label pairs are highlighted on the occluded structure by black spheres connected by a dotted line. Distance distributions for each pair were obtained in the apo, Na^+-bound (Na^+), and Na^+-and BH-bound (Na^+/BH) intermediates. The multi-component distributions reflect multiple conformations of Mhp1 in equilibrium between inward-facing and outward-facing conformations.

The distance distributions monitoring the bundle relative to scaffold, profile a symporter in equilibrium between inward- and outward-facing conformations (Fig. 5.2). On the intracellular side, distance distributions report two populations consistent with an equilibrium between distinct conformations. The longer component, favored under apo and Na^+ conditions, is compatible with an inward-facing structure. The presence of Na^+ and BH shifts the equilibrium to favor the shorter component associated with an outward-facing and/or outward-facing occluded conformation. These conformational changes are further described by the extracellular distributions, where the apo and Na^+-bound states favor the shorter distance component (inward-facing conformation) and binding Na^+ and BH favors the longer component (outward-facing conformation). These measurements agree with a rigid body conformational transition

between bundle and scaffold motifs. Furthermore, the magnitude of conformational change is remarkably consistent with that predicted by comparing the Mhp1 crystal structures supporting the interpretation that crystal structure conformations are indeed sampled in solution. The broad and overlapping distance distributions of the intracellular side (*e.g.* TMs 7 and 9 in Fig. 5.2) and the distinctly bimodal distributions of the extracellular side (*e.g.* TMs 3 and 7 in Fig. 5.2) suggest fluctuations occur between inward-facing and outward-facing conformations under all biochemical conditions.

Crystallographically identified gate motifs are highly dynamic

The gating transitions identified in the Mhp1 crystal structures for TMs 5 and 10 were also monitored using DEER. Relative to the bundle motif, intracellular TM5 favors an open conformation in the apo and Na^+-bound states and a closed conformation in the Na^+- and BH-bound states (Fig. 5.3). These distance distributions are significantly broader than those relating the bundle and scaffold motifs, indicative of a higher degree of flexibility, especially in the inward-facing conformation. The magnitudes of conformational transition are generally consistent with the Mhp1 crystal structures for intracellular TM5. However, the similarity between the conformational changes evident in TM5 and the relationship between the bundle and scaffold lead to the conclusion that TM5 may operate as a functional element of the scaffold rather than an independent gate.

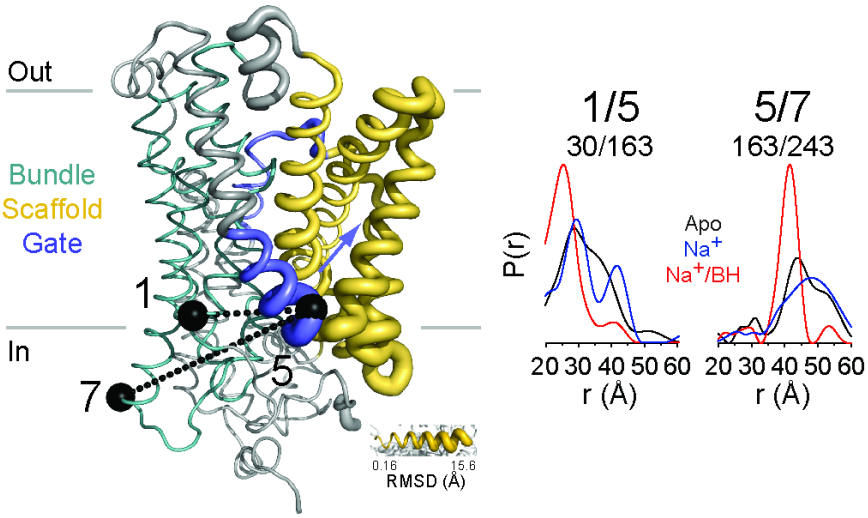


Figure 5.3. Dynamics of TM5 relative to the bundle on the intracellular side of Mhp1. (Left) As in Fig. 5.2, RMSD between outward-facing occluded (PDB ID: 2JLO) and inward-facing (PDB ID: 2X79) Mhp1 crystal structures mapped onto the occluded structure. The blue arrow indicates the direction of the movement inferred from comparison of the crystal structures. (Right) The distributions reflect Mhp1 equilibrium between inward-facing and outward-facing conformations as in the bundle and scaffold comparison.

The crystal structures also identify extracellular TM10 as undergoing a dramatic shift in position between the outward-facing occluded and outward-facing structures (Fig. 5.4). This shift in position led to the conclusion that Mhp1 samples a distinct occluded intermediate during transport. Deviating from crystallographic representations, distance distributions for extracellular TM10 show little evidence of discrete conformational changes predicted by the crystallographically captured conformations (Fig. 5.4). In the presence of the Na^+ and BH, the width of the distribution increases suggesting an increase in flexibility in the outward-facing conformation, similar in principle to that shown in TM5 (Fig. 5.3). However, based on the TM10 distance distributions, we conclude that the crystallographic outward-facing occluded conformation overestimates the dynamic flexibility of TM 10 and that the Mhp1 transport cycle may not necessitate an occluded intermediate state defined by the position of TM10.

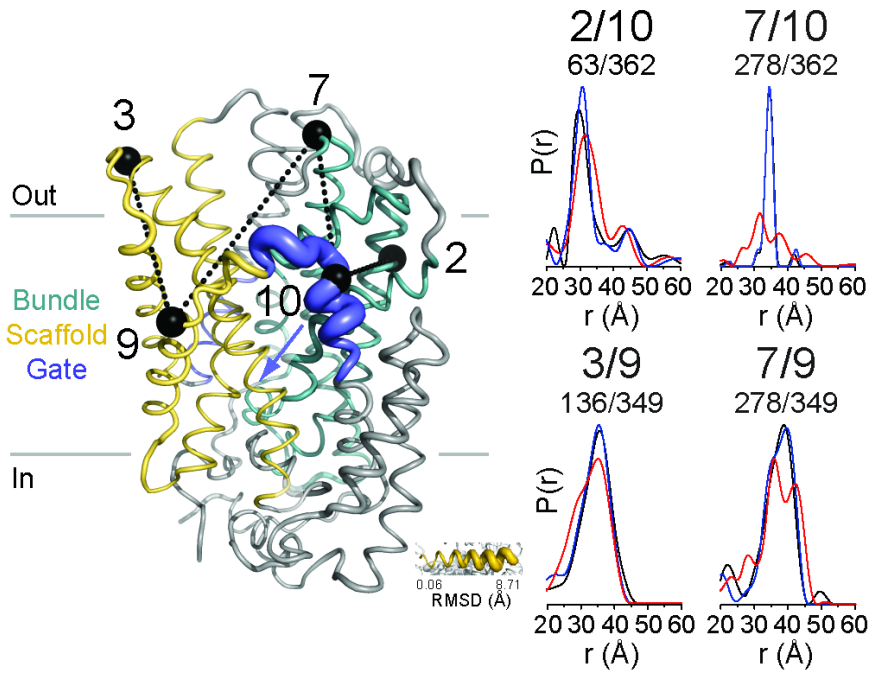


Figure 5.4. Dynamics of TMs 9 and 10 on the extracellular side of Mhp1 (Left) RMSD between outward-facing occluded (PDB ID: 2JLO) and outward-facing (PDB ID: 2JLN) Mhp1 crystal structures mapped onto the occluded structure. The blue arrow indicates the direction of the movement inferred from comparison of the crystal structures for TM10. (Right) The distributions show limited conformational changes in this region as an increase in flexibility in the Na^+/BH intermediate.

We identified a very similar dynamic profile in extracellular TM9, directly adjacent to TM10 (Fig. 5.4). While independent dynamic fluctuations in this region were not evident from comparison of the Mhp1 crystal structures, previous MD simulations[30] predicted TM9 dynamics operating in conjunction with TM10 fluctuations. Distance distributions relating the position of TM9 show that TM9 deviates from rigid body motion. In the presence of Na^+ and BH, the breadth of the TM9 distributions increases indicating an increase in dynamic range, but does not sample discrete alternative conformations as would be expected of a scaffold helix relative to the bundle motif. This increase in flexibility is independent of scaffold motions as is evident in the intra-scaffold measurements. The similarity between the TM9 and TM10 profiles leads us to conclude that the motions of TM9 and 10 are related and functionally similar.

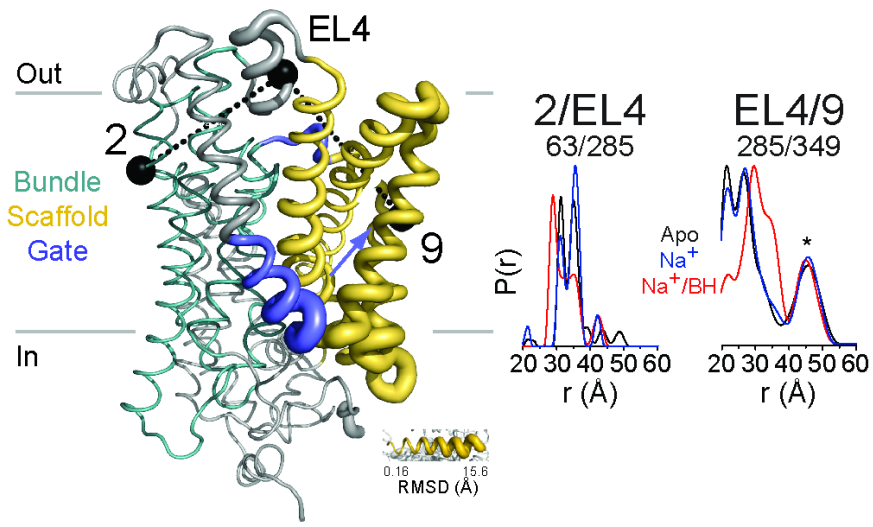


Figure 5.5. Motions of EL4 relative to bundle and scaffold motifs. (Left) RMSD between outward-facing occluded (PDB ID: 2JLO) and inward-facing (PDB ID: 2X79) Mhp1 crystal structures mapped onto the occluded structure. (Right) The distributions show distinct conformational changes in this region potentially indicative of a extracellular gating mechanism.

EL4, the extracellular loop between TMs 7 and 8, was not identified in the previous crystal structure analysis as a gate helix, though a small scale change in its position was evident in a comparison of the outward-facing/outward-facing occluded and inward-facing structures (Fig. 5.5). Distance distributions relating the dynamics of the of this motif display a uniquely trimodal profile (Fig. 5.5). Relative to TM2 of the bundle, two distance populations are favored in the apo and Na^+-bound states. In the presence of Na^+ and BH, EL4 shifts to a new position nearer to the bundle. This distribution profile is confirmed in distance distributions relating EL4 to the scaffold, with opposite directionality of motion, though this distribution is somewhat complicated by the inclusion of independent motions of TM9 evident in the Na^+- and BH-bound distribution described previously (red traces, Fig. 5.4). The evident motions of EL4, independent of both bundle and scaffold motifs, suggest that EL4 may participate in independent regulation of extracellular occlusion and substrate access. We speculate that the bimodal distributions of EL4 (apo and Na^+-bound) may correspond to the bimodal distributions shown in Figure 5.2

describing the inward-facing and outward-facing conformations. In the presence of Na^+ and BH, EL4 adopts a previously unsampled position, which may represent a distinct occluded conformation.

Mhp1 equilibrium is not coupled to Na^+ binding

Previous representations of the Mhp1 transport cycle have implied that Na^+ triggers a conformational change in Mhp1[41]. This was primarily due to comparison with the LeuT transport cycle, where such effects have been reported[42] as well as MD simulations in LeuT that have pointed to the conserved Na2 site as the source of these conformational transitions[22]. Furthermore, MD simulations[30] in Mhp1 suggested that Na^+ binding to the Na2 site indeed shifts the equilibrium between states. Contrary to these expectations, the energetics of Mhp1 conformational equilibrium are not shifted by Na^+ binding at the Na2 site. Almost superimposable distributions for all mutants in the Na^+-bound and apo states (Figs. 5.2-5.5) suggest little change in the relative population of inward- versus outward-facing conformations. In contrast, it is the binding of the substrate benzylhydantoin in the presence of Na^+ that dramatically shifts the equilibrium to favor the outward-facing conformation. It is from these inferences that we can begin to conclude that LeuT and Mhp1 do indeed operate using distinct mechanisms of transport and define the structural basis for mechanistic divergence.

The distinct role of Na^+ in Mhp1 and LeuT alternating access

Symport requires the binding of ions and substrates on one side of the membrane, conformational changes to reorient the binding site, followed by a dissociation step to the other side of the membrane. To enforce directional movement of substrates, co-transported ions can stabilize and destabilize substrate binding in the outward- and inward-facing conformations conferring directionality by their concentration gradients. Additionally, ion binding can directly alter the energetics of the conformational equilibrium to favor a particular conformation. A

comparative analysis of the Mhp1 and LeuT Na^+-dependent conformational changes illustrates an example of these two mechanisms and allows definition of the individual roles of the Na1 and Na2 sites.

In LeuT, which possesses both Na1 and Na2, we previously described Na^+-dependent conformational transitions wherein the apo state favors an inward-facing conformation and the Na^+-bound state favors an outward-facing conformation, but were unable to specifically identify the Na site responsible for this transition (Chapter 4). It has been previously reported that Na2 in Mhp1 provides stabilization to substrate binding, evident in decreased K_D[4]. This coupled with the lack of shifts in conformational equilibria in the presence of Na^+ in Mhp1 reported here, leads us to conclude that the conserved Na2 site serves to couple transport of substrate to the Na^+ gradient through direct stabilization of substrate binding. In contrast, the additional mechanism of conformational coupling to the Na^+ gradient evident in LeuT, we propose is the result of Na^+ binding at the non-conserved Na1 site. Without this site, Mhp1 symport is achieved via thermodynamically coupled binding and release of Na^+ and benzylhydantoin and equilibrium fluctuations between conformational states. This role of Na^+ as a conformational trigger may represent a critical mechanistic divergence between two classes of LeuT Fold transporters with Na^+/substrate stoichiometry of 1 versus 2.

Mhp1 transport cycle dependent on low probability transitions

By framing Mhp1 crystal structures in a conformational equilibrium and characterizing its response to the ion and substrate effects, we derive a plausible model of how isomerization of the transporter between inward-facing and outward-facing conformations mediates transport (Fig. 5.6). On the basis of mutual stabilization[4], we propose that Na^+ and substrate concurrently bind to the minor population of outward-facing apo-Mhp1 and shift the equilibrium to favor the outward-facing conformation. Previous MD simulations[30] demonstrated that the presence of BH blocks the pathway of Na^+ dissociation to the extracellular side. Extracellular occlusion may also

be aided by a change in the position of EL4 (grey cylinder, Fig. 5.6). The overlapping distance distributions in the apo/Na^+ and substrate-bound states (Fig. 5.2) indicate that Na^+/BH-bound Mhp1 can fluctuate to an open-in conformation, which would enable Na^+ to dissociate to the intracellular side in the presence of a gradient. Na^+ release lowers substrate affinity thereby triggering substrate release. However, the low probability of this event, demonstrated in the low population probability in the distance distribution, defines the rate limiting step of transport. The oscillation of the apo intermediate between open-in and open-out conformations initiates a new cycle of transport. This model is contrary to previously expectations[41], as the lack of triggered conformational changes throughout the transport cycle, is counterintuitive. However, this demonstrates the importance of supporting crystallographic representations of structures with information describing ligand-dependent conformational equilibria.

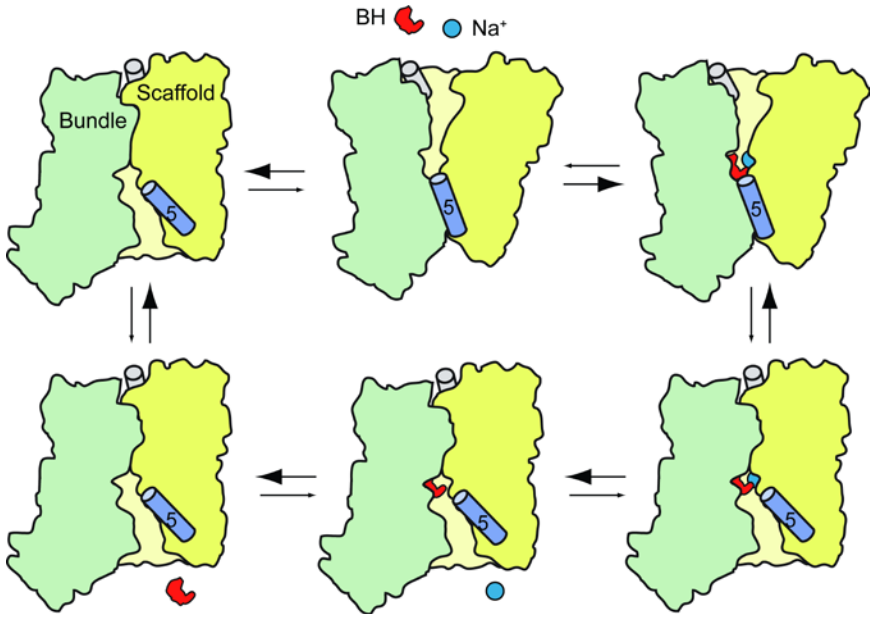


Figure 5.6. EPR-derived model of Mhp1 transport. Apo-Mhp1, through low probability transitions, samples outward-facing conformations, allowing likely simultaneous binding of the Na^+ and substrate. Binding of Na^+ and substrate results in a stabilization of the outward-facing conformation. Low probability fluctuations allow sampling of the inward-facing conformation. Driven by its concentration gradient, Na^+ dissociates to the intracellular solution. In the absence of bound Na^+, BH affinity to Mhp1 is reduced which drives simultaneous dissociation of BH to the intracellular side. The cycle continues through the isomerization from apo-inward-facing to apo-outward-facing.

The divergent transport mechanisms of Mhp1 and LeuT

In Chapter 4, the conformational cycle of LeuT was described using a similar analysis to that applied to Mhp1 here. We demonstrated that the opening/closing of the two extracellular motifs, TM1/TM6 and TM7/EL4, and the intracellular motif TM6/7 mediate LeuT alternating access, with TMs 6 and 7 providing direct coupling of conformational equilibria to Na^+ and substrate binding. This stands in stark contrast to the Mhp1 conformational cycle outlined above. Specifically, Mhp1 operates through a symmetric, essentially rigid body Rocking Bundle mechanism that couples to the Na^+ gradient through mutual stabilization of Na^+ and substrate. LeuT functions asymmetrically, additionally coupled to the Na^+ gradient through triggered conformational changes applied at TMs 6 and 7, as well as through mutual stabilization of substrate binding by the Na2 site. LeuT samples a global occluded conformation, with all dynamic motifs sampling closed conformations simultaneously, while in Mhp1 occlusion occurs through gates which are dynamic and, only in the case of EL4, independent of Rocking Bundle motions. It total, this comparison unequivocally demonstrates that LeuT and Mhp1 operate using completely distinct transport mechanisms arguing against a unified mechanism for the LeuT fold class.

The novel perspective that emerges from this comparative analysis is that the functional diversity of the conserved LeuT fold, manifested by diverse substrates, coupling modes, number and identity of the co-transported ion, *requires* distinct structural schemes of alternating access. The work illuminates a critical mechanistic element which has been missing from the analysis thus far, namely, how ion binding shapes the energy landscape of conformations. While we speculate, based on our results, that the role of the conserved Na2 site is to stabilize substrate binding without conformational selection, analysis of other Na^+-coupled transporters is needed to test this conjecture. Such analysis will also test a similarly tantalizing notion that subclasses of ion-coupled LeuT Fold transporters, defined by their transport modes and/or type and number of symported ions, share commonalities in their structural mechanics of alternating access.

References

1. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of na+/cl--dependent neurotransmitter transporters. *Nature* **437**, 215-223 (2005).
2. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J. The crystal structure of a sodium galactose transporter reveals mechanistic insights into na+/sugar symport. *Science* **321**, 810-814 (2008).
3. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the na(+)/betaine symporter betp. *Nature* **458**, 47-52 (2009).
4. Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J. & Cameron, A. D. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709-713 (2008).
5. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a na+-independent amino acid transporter. *Science* **325**, 1010-1014 (2009).
6. Tang, L., Bai, L., Wang, W. H. & Jiang, T. Crystal structure of the carnitine transporter and insights into the antiport mechanism. *Nat Struct Mol Biol* **17**, 492-496 (2010).
7. Schulze, S., Koster, S., Geldmacher, U., Terwisscha van Scheltinga, A. C. & Kuhlbrandt, W. Structural basis of na(+)-independent and cooperative substrate/product antiport in cait. *Nature* **467**, 233-236 (2010).
8. Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y. & Miller, C. Structure of a prokaryotic virtual proton pump at 3.2 a resolution. *Nature* **460**, 1040-1043 (2009).
9. Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., Wang, J. & Shi, Y. Structure and mechanism of an amino acid antiporter. *Science* **324**, 1565-1568 (2009).
10. Ma, D., Lu, P., Yan, C., Fan, C., Yin, P., Wang, J. & Shi, Y. Structure and mechanism of a glutamate-gaba antiporter. *Nature* **483**, 632-636 (2012).
11. Masson, J., Sagne, C., Hamon, M. & El Mestikawy, S. Neurotransmitter transporters in the central nervous system. *Pharmacol Rev* **51**, 439-464 (1999).
12. Wright, E. M., Loo, D. D. & Hirayama, B. A. Biology of human sodium glucose transporters. *Physiol Rev* **91**, 733-794 (2011).
13. Ziegler, C., Bremer, E. & Kramer, R. The bcct family of carriers: From physiology to crystal structure. *Mol Microbiol* **78**, 13-34 (2010).
14. Pantazopoulou, A. & Diallinas, G. Fungal nucleobase transporters. *FEMS Microbiol Rev* **31**, 657-675 (2007).
15. Ren, Q., Chen, K. & Paulsen, I. T. Transportdb: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res* **35**, D274-279 (2007).
16. Singh, S. K. Leut: A prokaryotic stepping stone on the way to a eukaryotic neurotransmitter transporter structure. *Channels (Austin)* **2** (2008).
17. Reed-Tsur, M. D., De la Vieja, A., Ginter, C. S. & Carrasco, N. Molecular characterization of v59e nis, a na+/i- symporter mutant that causes congenital i- transport defect. *Endocrinology* **149**, 3077-3084 (2008).
18. Wright, E. M., Hirayama, B. A. & Loo, D. F. Active sugar transport in health and disease. *J Intern Med* **261**, 32-43 (2007).

19. Calonge, M. J., Gasparini, P., Chillaron, J., Chillon, M., Gallucci, M., Rousaud, F., Zelante, L., Testar, X., Dallapiccola, B., Di Silverio, F. & et al. Cystinuria caused by mutations in rbat, a gene involved in the transport of cystine. *Nat Genet* **6**, 420-425 (1994).
20. Torrents, D., Mykkanen, J., Pineda, M., Feliubadalo, L., Estevez, R., de Cid, R., Sanjurjo, P., Zorzano, A., Nunes, V., Huoponen, K., Reinikainen, A., Simell, O., Savontaus, M. L., Aula, P. & Palacin, M. Identification of slc7a7, encoding y+lat-1, as the lysinuric protein intolerance gene. *Nat Genet* **21**, 293-296 (1999).
21. Borsani, G., Bassi, M. T., Sperandeo, M. P., De Grandi, A., Buoninconti, A., Riboni, M., Manzoni, M., Incerti, B., Pepe, A., Andria, G., Ballabio, A. & Sebastio, G. Slc7a7, encoding a putative permease-related protein, is mutated in patients with lysinuric protein intolerance. *Nat Genet* **21**, 297-301 (1999).
22. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the leut and apct transporters are modulated by common mechanisms. *Biophys J* **99**, L103-105 (2010).
23. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347-355 (2009).
24. Forrest, L. R., Kramer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta* **1807**, 167-188 (2011).
25. Jeschke, G. A comparative study of structures and structural transitions of secondary transporters with the leut fold. *Eur Biophys J* **42**, 181-197 (2013).
26. Mitchell, P. A general theory of membrane transport from studies of bacteria. *Nature* **180**, 134-136 (1957).
27. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969-970 (1966).
28. Patlak, C. S. Contributions to the theory of active transport: Ii. The gate type non-carrier mechanism and generalizations concerning tracer flow, efficiency, and measurement of energy expenditure. *Bull. Math. Biophys.* **19**, 209-235 (1957).
29. Vidaver, G. A. Inhibition of parallel flux and augmentation of counter flux shown by transport models not involving a mobile carrier. *J Theor Biol* **10**, 301-306 (1966).
30. Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter mhp1. *Science* **328**, 470-473 (2010).
31. Forrest, L. R. & Rudnick, G. The rocking bundle: A mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)* **24**, 377-386 (2009).
32. Crisman, T. J., Qu, S., Kanner, B. I. & Forrest, L. R. Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc Natl Acad Sci U S A* **106**, 20752-20757 (2009).
33. Radestock, S. & Forrest, L. R. The alternating-access mechanism of mfs transporters arises from inverted-topology repeats. *J Mol Biol* **407**, 698-715 (2011).
34. Schushan, M., Rimon, A., Haliloglu, T., Forrest, L. R., Padan, E. & Ben-Tal, N. A model-structure of a periplasm-facing state of the nhaa antiporter suggests the molecular underpinnings of ph-induced conformational changes. *J Biol Chem* **287**, 18249-18261 (2012).
35. Krishnamurthy, H. & Gouaux, E. X-ray structures of leut in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469-474 (2012).
36. Hubbell, W. L., Mchaourab, H. S., Altenbach, C. & Lietzow, M. A. Watching proteins move using site-directed spin labeling. *Structure* **4**, 779-783 (1996).

37. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys Chem Chem Phys* **9**, 1895-1910 (2007).
38. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *Journal of Magnetic Resonance.* **142**, 331-340 (2000).
39. Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J., Timmel, C. R., Hilger, D. & Jung, H. Deeranalysis2006—a comprehensive software package for analyzing pulsed eldor data. *Applied Magnetic Resonance* **30**, 473-498 (2006).
40. Chiang, Y. W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed esr using tikhonov regularization. *Journal of Magnetic Resonance* **172**, 279-295 (2005).
41. Weyand, S., Shimamura, T., Beckstein, O., Sansom, M. S., Iwata, S., Henderson, P. J. & Cameron, A. D. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter mhp1. *J Synchrotron Radiat* **18**, 20-23 (2011).
42. Claxton, D. P., Quick, M., Shi, L., de Carvalho, F. D., Weinstein, H., Javitch, J. A. & Mchaourab, H. S. Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:Sodium symporters. *Nat Struct Mol Biol* **17**, 822-829 (2010).

CHAPTER 6

PERSPECTIVES ON FUTURE DIRECTIONS

Validating EPR results in the presence of the lipid membrane

The distance measurements conducted for LeuT and Mhp1 in this thesis were collected in a detergent environment. While this approach has experimental advantages in throughput and in signal-to-noise ratios over measurements conducted in a liposome environment, caveats outlined in Chapter 2 also apply to these investigations. Specifically, the lateral pressure imposed by the membrane environment may provide necessary stability for protein structure and/or specific interactions with lipid molecules can alter protein function. Thus, it is important for investigations such as these to include experiments in the presence of lipids to ensure that the detergent environment is not affecting protein function. Specifically, in LeuT, the presence of a lipid molecule bound in the extracellular vestibule[1] poses the question of whether lipid interactions directly regulate the function of LeuT. While initial investigations have shown similarity between detergent-solubilized and lipid-reconstituted LeuT conformational sampling[2], we are interested in validating the LeuT and Mhp1 measurements described in Chapters 4 and 5 in a lipid environment.

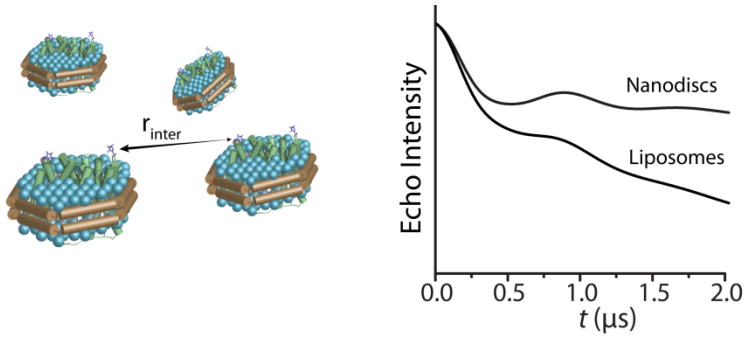


Figure. 6.1. Nanodisc reconstitution of membrane proteins for EPR investigations. In addition to providing a more native-like lipid environment, nanodisc reconstitution of membrane proteins, offers increased intermolecular distances between spins (left), which results in a decrease in background signal (right) compared to proteoliposome prepared samples.

As mentioned above, protein reconstitution into liposomes results in a decrease in signal to noise compared to detergent-solubilized samples. This is due to an increase in the intermolecular contribution to the EPR signal associated with the high effective concentrations in the two-dimensional environment of a proteoliposome that accentuate the background contribution. This contribution imposes severe limits on sensitivity, distance range, and experimental throughput in proteoliposome samples. To overcome these limitations, Zou and Mchaourab[3] reported a general methodology which relies on reconstitution of spin labeled membrane proteins into Nanodiscs (also referred to as nanoscale bilayers). These bilayers are a class of soluble nanoscale assemblies of lipids surrounded by a belt of amphipathic protein derived from apolipoprotein A1[4] (Fig. 6.1, left). By careful manipulation of the molar ratios between the three components, it is possible to reconstitute a single membrane protein per bilayer disk resetting the dimensionality of the DEER background factor to three (compared to approximately 2 in proteoliposomes) (Fig. 6.1, right). The use of Nanodiscs is facilitated by an order of magnitude increase in DEER sensitivity achieved at Q-band frequency[5] relative to the commonly used X-band frequency. The synergistic convergence of these two technologies overcomes the bottlenecks for widespread application of DEER to sample-limited membrane proteins. Using this approach on a select number of LeuT and Mhp1 mutants, we could validate

the conformational equilibria associated with detergent-solubilized forms and probe the hypothesized lipid interactions in LeuT.

Define the intracellular permeation pathway with solvent accessibility in LeuT

One of the questions that this work was intended to address related to the alternating access cycles of the LeuT and Mhp1. Specifically, what conformational rearrangements were necessary for the central binding site to transition from being accessible from the extracellular side to the intracellular side and vice versa. While a great success of the work presented here has been to define the conformational rearrangements associated with the LeuT conformational cycle, one disadvantage of this approach is its inability to correlate the conformational changes seen in the DEER distributions with changes in solvent accessibility at the ion and substrate binding sites. Verification that the conformational changes that are visualized using distance measurements are sufficient to allow access of ion and substrate to the central binding sites solvent is a necessary next step for this research. Similar to an analysis conducted on the extracellular side of LeuT[2] (Fig 6.2), a systematic solvent accessibility investigation of the intracellular permeation pathway proposed by this work, would validate the propsed location of this pathway and verify that the conformational changes described above result in changes in patterns of solvent accessibility extending to the ion and substrate binding sites. Given the inconsistencies between the EPR results on the intracellular side of LeuT and the LeuT crystal structures, definition of the location of the intracellular permeation pathway would provide necessary support for the mechanistic conclusions proposed in this thesis. Furthermore, defining the specific location of the intracellular vestibule would allow rational design of novel therapeutics to target this region.

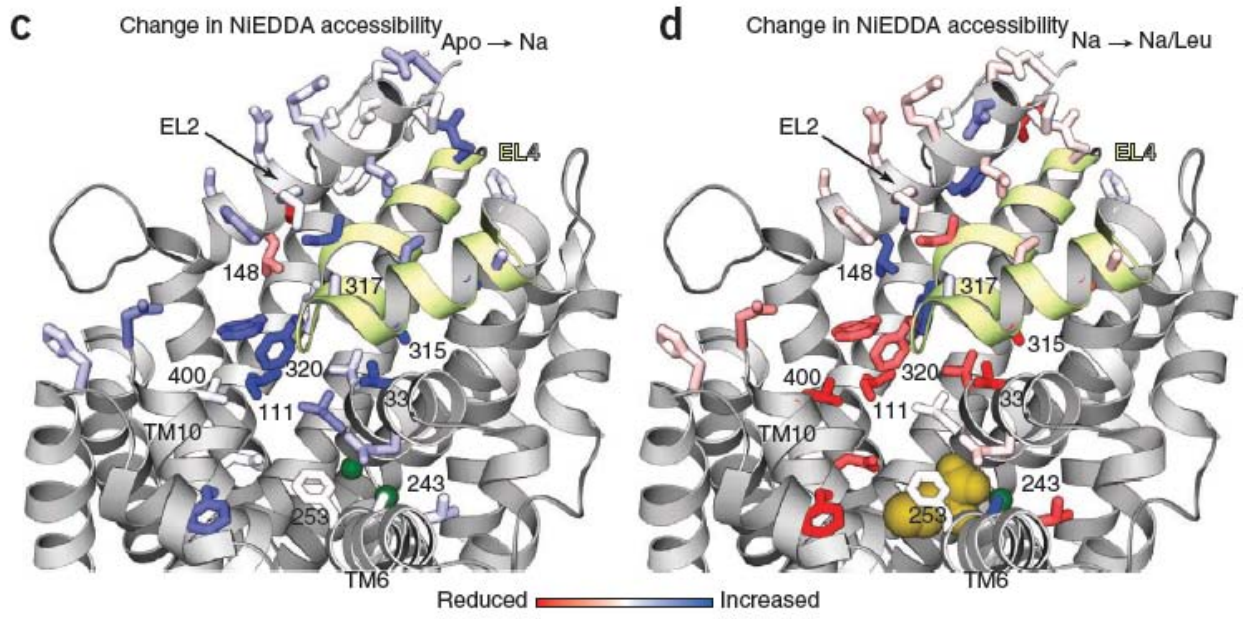


Figure. 6.2. Solvent accessibility measurements conducted on the extracellular side of LeuT. Measurements monitored the changes in accessibility to the PRA NiEDDA in apo, Na^+-bound and Leu-bound intermediates at residues in the extracellular vestibule noted by number and colored to reflect relative change in accessibility between intermediate states. This analysis defined the Na^+-bound intermediate as significantly more solvent accessible that either the apo or the Leu-bound intermediates. This trend extended all the way to the substrate binding site. Figure reused with permission from [2].

Conformational dynamics of DAT

While this thesis was being compiled, a crystal structure of a dopamine transporter (DAT) from *Drosophila melanogaster* was released[6]. The structure represented an inhibited form, bound to the TCA nortriptyline at S2 and cholesterol at the interface of intracellular TMs 1, 5, and 7 (Fig. 6.3). This ortholog has 50% sequence identity to mammalian NSS making it a significantly closer cousin than LeuT. The DAT structure served to confirm the importance of the LeuT investigations described here, as the structural core was virtually identical to the outward-facing orientations of LeuT, with only minor deviations, such as a kink in TM12 centered around the conserved proline at residue 572. A comparison of the outward-facing LeuT structure and the DAT structure resulted in an RMSD of ~2Å. This structural similarity highlights the

evolutionary structural conservation of the LeuT Fold among NSS members and reinforces that importance of understanding LeuT Fold mechanisms as a means of understanding human disease. However, the results described in this thesis underscore the necessity of investigations to confirm mechanistic similarity of structurally similar proteins. Future work in this area, should evaluate the conformational sampling and transport mechanism of this ortholog as a comparison among LeuT Fold members as well as a closer approximation of the function of human NSS.

A number of mutations were used to thermostabilize the DAT structure, as WT-DAT was unable to bind substrate, and was therefore structurally compromised, upon extraction from cellular membranes[6]. While it is possible that an extraction and reconstitution procedure could be developed to maintain or reintroduce structural integrity, it is likely that an EPR analysis would require a thermostabilized construct as well, potentially complicating interpretation of the results. In this scenario, the effects of thermostability mutants on conformational equilibria would not be directly investigated and comparisons with LeuT would be confounded by this additional variable. However, the insights gained by an EPR investigation of the mutant DAT mechanism would undoubtedly provide valuable insights into the function of NSS and the LeuT Fold. Furthermore, it is arguable that this construct may provide a more compelling target for understanding human NSS than LeuT given its greater sequence similarity even in the thermostabilized mutant.

Therefore, I propose a similar investigation to that described in Chapter 4 to monitor the conformational dynamics of this DAT construct using EPR. This analysis would compare the transport cycle described for LeuT to measurements conducted in DAT. Of particular importance, measurements describing the dynamic fluctuations of the bundle helices to specifically identify potential ligand-dependent translations of TM1a. As a key source of divergence between the model of LeuT transport described here and the proposed Crystal Structure Model of LeuT transport, investigations in the region would provide an independent

evaluation of these models. Additional measurements would probe the overall architecture of conformational intermediates as well as ligand-dependent conformational equilibria between states. This investigation would inform on the role of increased flexibility of the DAT ortholog on the conformational energetic landscape and dynamic fluctuations between conformational intermediates. How the likely divergent energetics of DAT affects the DAT transport cycle and how Na^+ and substrate binding are used to stabilize conformational intermediates of DAT are questions that could be addressed that would improve our understanding of human NSS mechanisms and eukaryotic membrane protein function.

Furthermore, these experiments would investigate the conformational effects of the divergences between the structures. In particular, the kinked TM12 would be monitored to define differential dynamic modes potentially resulting from the conserved proline residue. In LeuT, TM 12 is thought to be static in position throughout the transport cycle. Conservation of P572 among human NSS as well as its location in a putative oligomerization-mediating helix in human NSS may indicating an additional functional role for TM12 in these transporters. EPR investigations would monitor the conformational dynamics of this helix seeking the presence of discrete alternative conformations within the ensemble. The effect of ligand binding on TM12 dynamics would be investigated. Moreover, the conformational equilibria of TM12, as well as other dynamic motifs, could be monitored as a function of oligomerization to describe the conformational effects of oligomerization and potential interprotomer regulation of transport activity.

A foundational conclusion of the work presented in this thesis is that the divergence in transport cycles between LeuT and Mhp1 is based in part on differing ion stoichiometries. Like LeuT, DAT functions with a 2:1 Na^+:substrate stoichiometry[6]. Through investigation of DAT Na^+-dependent conformational dynamics, it would be possible to test the hypothesis that functional subclasses within the LeuT Fold are defined by their Na^+ stoichiometry. In these experiments, distance measurements between equivalent residues in LeuT and DAT would be investigated

for Na^+-dependent conformational equilibria. If DAT manifested similar shifts in equilibria in the presence of Na^+ as compared to LeuT, this result would serve to support our interpretation that Na1 is a conformational trigger in the LeuT Fold. Furthermore, DAT maintains a Cl^- binding site that can be reproduced in LeuT with a single residue substitution[7-8]. The conformational effect of the Cl^- binding could be similarly probed as a function of global structure in DAT and the Cl^- binding site mutant could be used into investigate the mechanistic effects of the Cl^- binding site divergence between LeuT and DAT. From these investigations, we could begin to describe the mechanistic roles played by the co-transported ions in NSS.

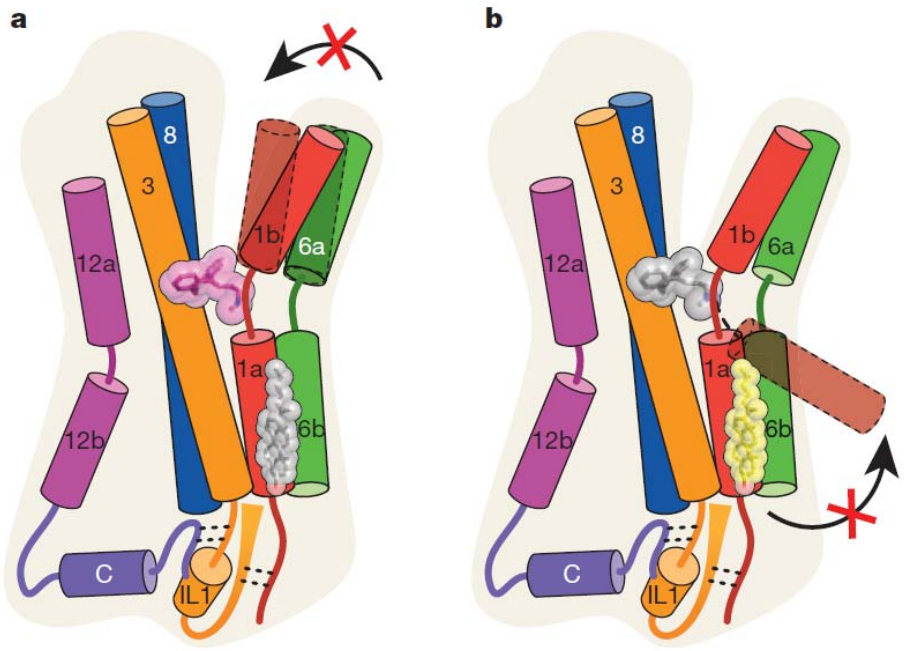


Figure. 6.3. Cartoon model of DAT and effects of antidepressants and cholesterol highlighting expected motions based on the LeuT crystal structures. Figure reused with permission from [6]

A cholesterol molecule was resolved in the structure located at the interface of TMs 1, 5, and 7 on the intracellular side[6]. The presence of this molecule was not necessarily surprising as cholesterol is known to exhibit effects on binding properties of cocaine[9]. The authors suggested

that cholesterol bound in this region served as “glue” holding TM1 in the closed position, stabilizing the outward-facing conformation. In light of the results presented in this thesis, it is interesting that the location of this molecule also lies along the interface of the intracellular dynamic TM6-7 motif proposed in the EPR investigations. Future investigations with DAT, would test the hypothesized role of cholesterol, specifically that cholesterol stabilizes the outward-facing conformation in DAT. Mutants monitoring the extracellular and intracellular sides, would determine whether cholesterol binding shifts the conformational equilibria of DAT and whether dynamic motions of the intracellular side, i.e. TMs 6 and 7, were impeded by the presence of cholesterol. These investigations would seek to define the regulatory properties of cholesterol on the DAT transport mechanism.

To capture the DAT structure, residue deletions were made in the N- and C-termini (ΔN1-20; ΔC602-631) as well as EL2 (Δ164-206). These deletions may deleteriously impact DAT protein functionality as these areas have been identified in regulation of transport activity[10], plasma membrane expression[11], glycosylation[12] and phosphorylation[13] regulation, and functionally important protein interactions with autoreceptors[14], syntaxin[15], and $Ca2^{+}$/calmodulin kinase II (CaMKII)[16]. Due to the density of functional properties in these regions, it is unclear whether the DAT structure can usefully inform on these processes. Taken into an EPR experiment, some of these regions could be reintroduced into the construct and the functional properties of these dynamic structural features could be monitored to provide a more complete picture of the function of DAT. Furthermore, the effects of important protein interactions with DAT could be specifically probed to describe regulation of DAT in the context of protein conformational equilibria.

With the publication of the DAT structure, new avenues of research have been opened. First, this likely heralds a new era in structural investigation of eukaryotic and eventually mammalian NSS. While the caveats associated with the study of labile membrane proteins will continue to pose challenges, technological advances will likely improve current preparation

techniques to accommodate them. It is my view that EPR, will play a key role in the future of structural investigations of NSS due to the inherent methodological advantages associated with EPR investigations of dynamic proteins and the unique view EPR offers on conformational equilibria in the molecular ensemble. The long term goal of EPR investigations will be in investigating the effects of known disease causing mutations and novel therapeutics on protein conformational sampling. This is an exciting time to be an EPR spectroscopist.

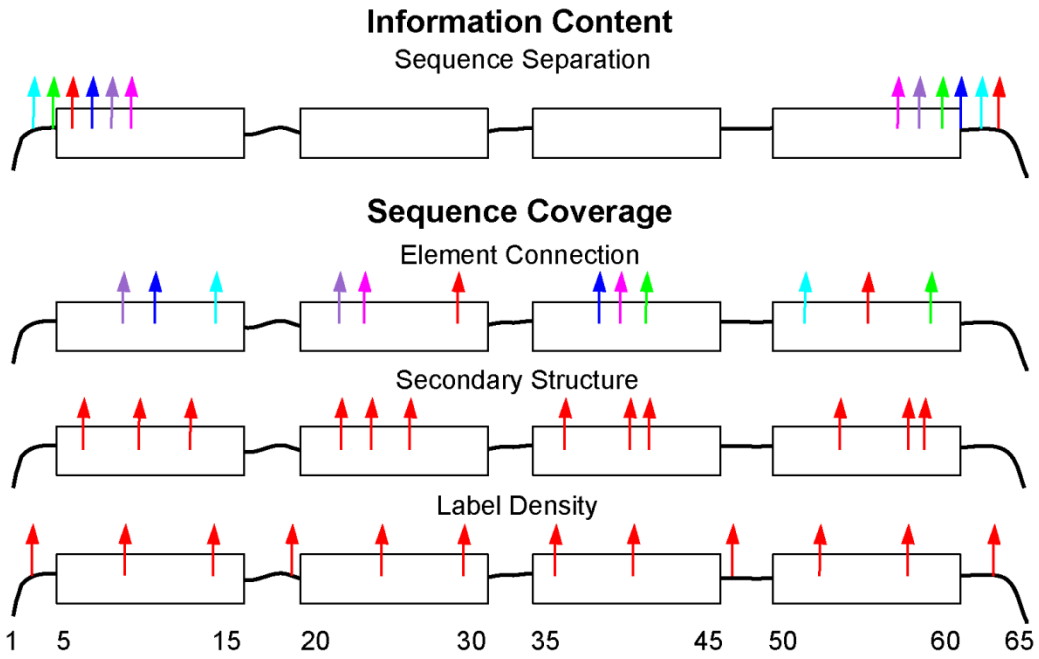
References

1. Wang, H., Elferich, J. & Gouaux, E. Structures of leut in bicelles define conformation and substrate binding in a membrane-like context. *Nat Struct Mol Biol* **19**, 212-219 (2012).
2. Claxton, D. P., Quick, M., Shi, L., de Carvalho, F. D., Weinstein, H., Javitch, J. A. & Mchaourab, H. S. Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:Sodium symporters. *Nat Struct Mol Biol* **17**, 822-829 (2010).
3. Zou, P. & Mchaourab, H. S. Increased sensitivity and extended range of distance measurements in spin-labeled membrane proteins: Q-band double electron-electron resonance and nanoscale bilayers. *Biophys J* **98**, L18-20 (2010).
4. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into nanodiscs. *FEBS Lett* **584**, 1721-1727 (2010).
5. Ghimire, H., McCarrick, R. M., Budil, D. E. & Lorigan, G. A. Significantly improved sensitivity of q-band peldor/deer experiments relative to x-band is observed in measuring the intercoil distance of a leucine zipper motif peptide (gcn4-lz). *Biochemistry* **48**, 5782-5784 (2009).
6. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* (2013).
7. Forrest, L. R., Tavoulari, S., Zhang, Y. W., Rudnick, G. & Honig, B. Identification of a chloride ion binding site in na+/cl -dependent transporters. *Proc Natl Acad Sci U S A* **104**, 12761-12766 (2007).
8. Zomot, E., Bendahan, A., Quick, M., Zhao, Y., Javitch, J. A. & Kanner, B. I. Mechanism of chloride interaction with neurotransmitter:Sodium symporters. *Nature* **449**, 726-730 (2007).
9. Hong, W. C. & Amara, S. G. Membrane cholesterol modulates the outward facing conformation of the dopamine transporter and alters cocaine binding. *J Biol Chem* **285**, 32616-32626 (2010).
10. Chen, J. G., Liu-Chen, S. & Rudnick, G. External cysteine residues in the serotonin transporter. *Biochemistry* **36**, 1479-1486 (1997).
11. Wang, J. B., Moriwaki, A. & Uhl, G. R. Dopamine transporter cysteine mutants: Second extracellular loop cysteines are required for transporter expression. *J Neurochem* **64**, 1416-1419 (1995).
12. Nunez, E. & Aragon, C. Structural analysis and functional role of the carbohydrate component of glycine transporter. *J Biol Chem* **269**, 16920-16924 (1994).
13. Vaughan, R. A., Huff, R. A., Uhl, G. R. & Kuhar, M. J. Protein kinase c-mediated phosphorylation and functional regulation of dopamine transporters in striatal synaptosomes. *J Biol Chem* **272**, 15541-15546 (1997).
14. Lee, F. J., Pei, L., Moszczynska, A., Vukusic, B., Fletcher, P. J. & Liu, F. Dopamine transporter cell surface localization facilitated by a direct interaction with the dopamine d2 receptor. *EMBO J* **26**, 2127-2136 (2007).
15. Sung, U., Apparsundaram, S., Galli, A., Kahlig, K. M., Savchenko, V., Schroeter, S., Quick, M. W. & Blakely, R. D. A regulated interaction of syntaxin 1a with the antidepressant-sensitive norepinephrine transporter establishes catecholamine clearance capacity. *J Neurosci* **23**, 1697-1709 (2003).
16. Fog, J. U., Khoshbouei, H., Holy, M., Owens, W. A., Vaegter, C. B., Sen, N., Nikandrova, Y., Bowton, E., McMahon, D. G., Colbran, R. J., Daws, L. C., Sitte, H. H., Javitch, J. A., Galli, A. & Gether, U. Calmodulin kinase ii interacts with the dopamine transporter c

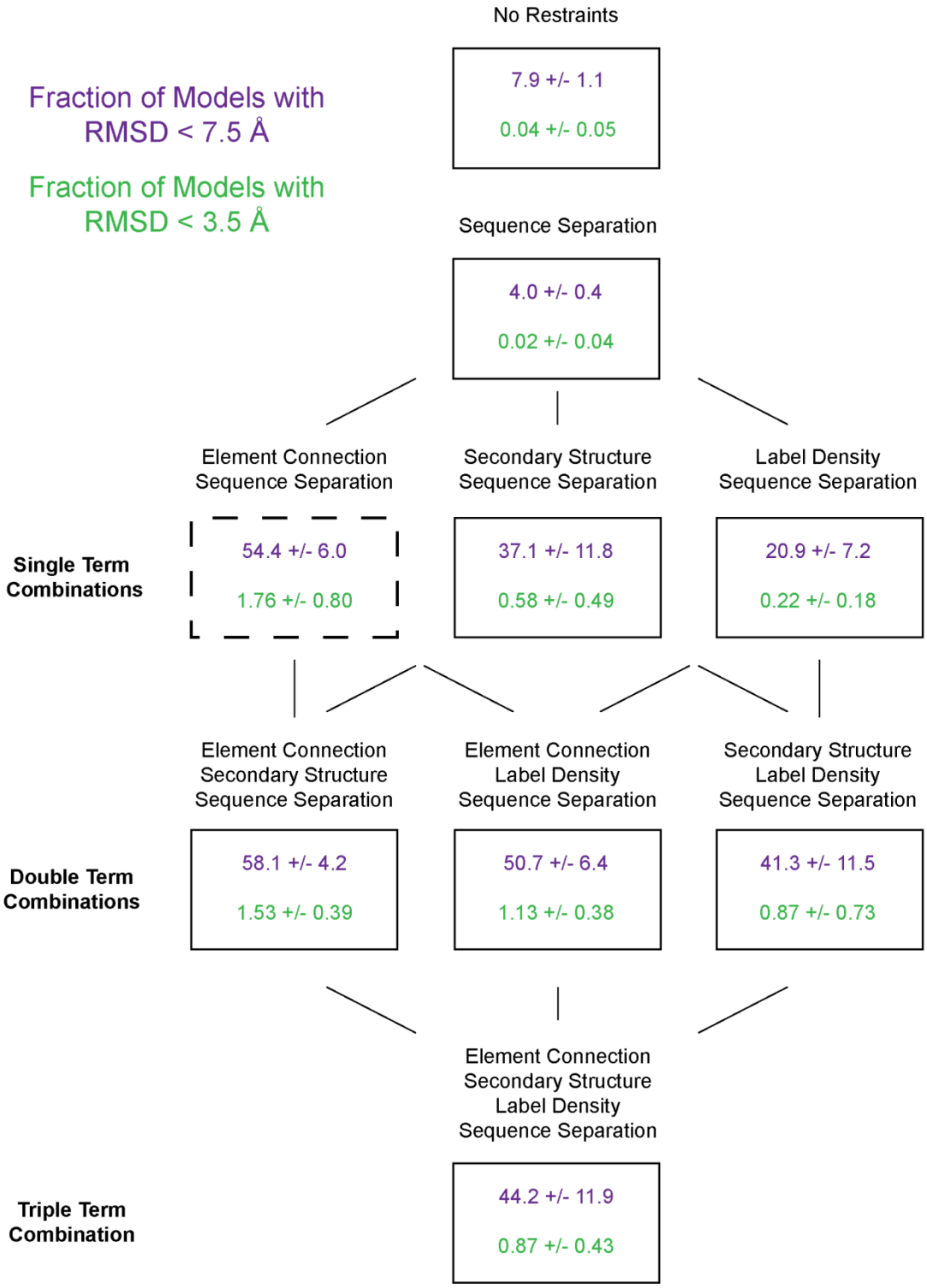
terminus to regulate amphetamine-induced reverse transport. *Neuron* **51**, 417-429 (2006).

APPENDIX A

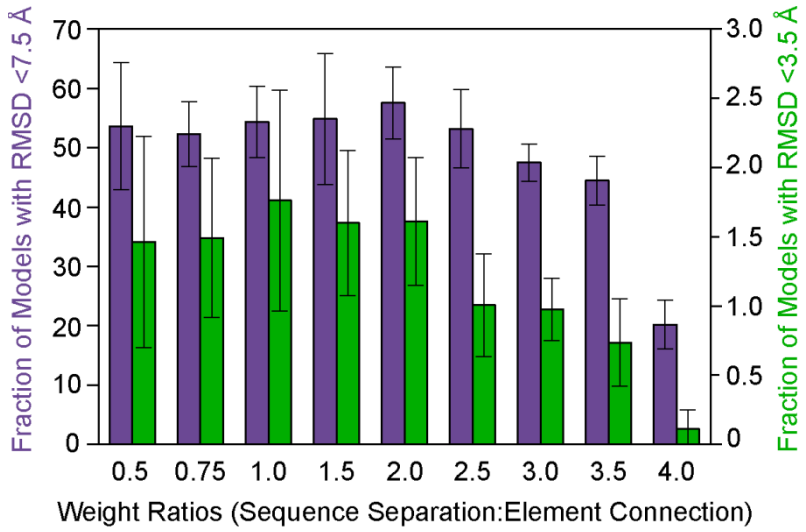
SUPPLEMENTAL FIGURES



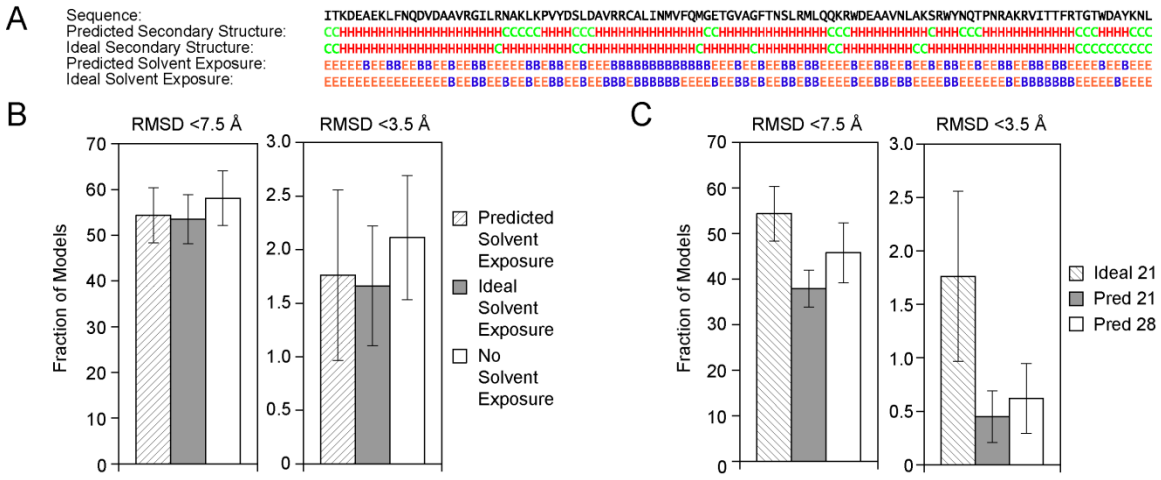
Supplementary Figure 3.1. Restraint distribution scores. In our simplified 4 helix, 6-restraint system, each arrow represents a spin label and arrows of the same color correspond to the same spin label pair. Sequence Separation, which maximizes the number of amino acids between spin labels in a pair, is a proxy for information content. Three sequence coverage terms were tested: Element Connection, Secondary Structure, and Label Density as defined in the methods section.



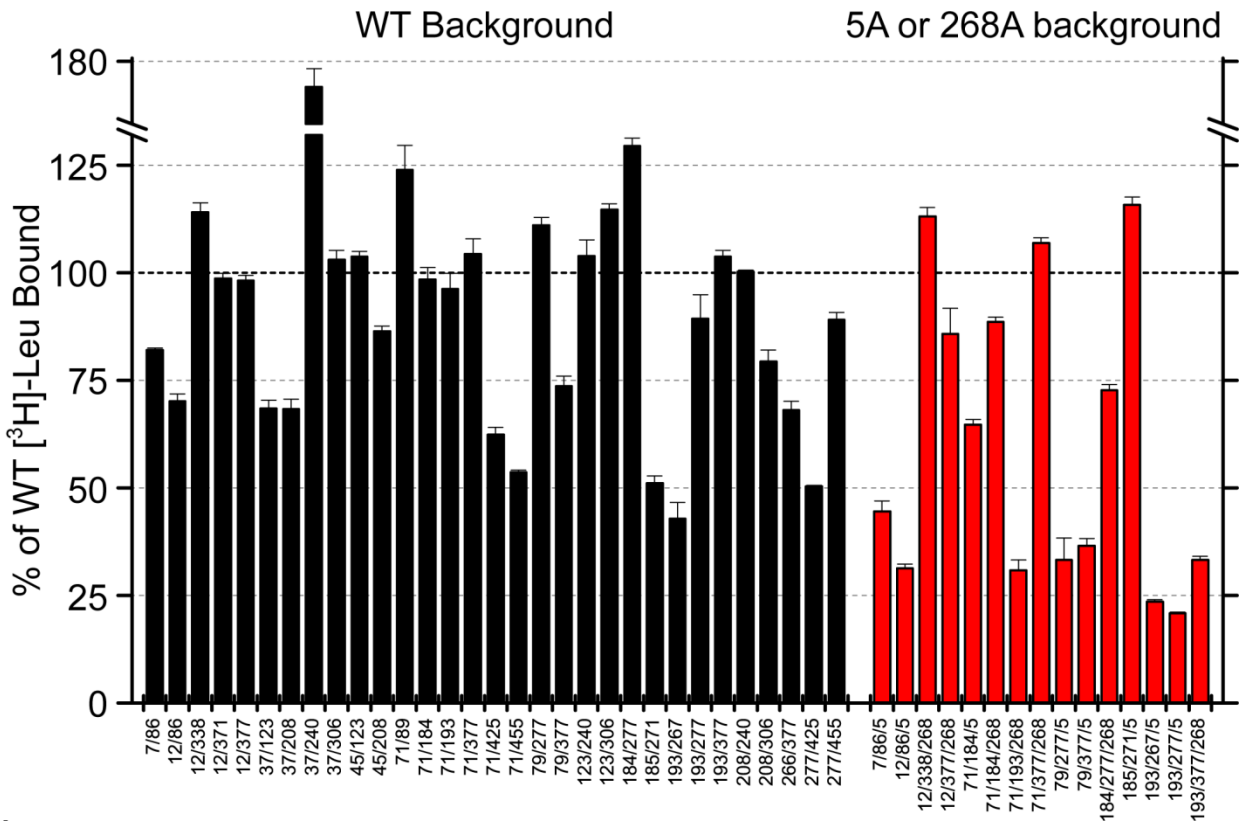
Supplementary Figure 3.2. Effects of sequence coverage terms on Rosetta model quality measures. Sequence coverage terms were combined with Sequence Separation to generate simulated restraint patterns. These were then incorporated into Rosetta to fold T4L C-terminal domain. The outcomes were evaluated by model quality measures, $C_{\square}$-RMSD <7.5 Å (purple) and $C_{\square}$-RMSD <3.5 Å (green). The best combination was found to be Element Connection and Sequence Separation (black dotted box).



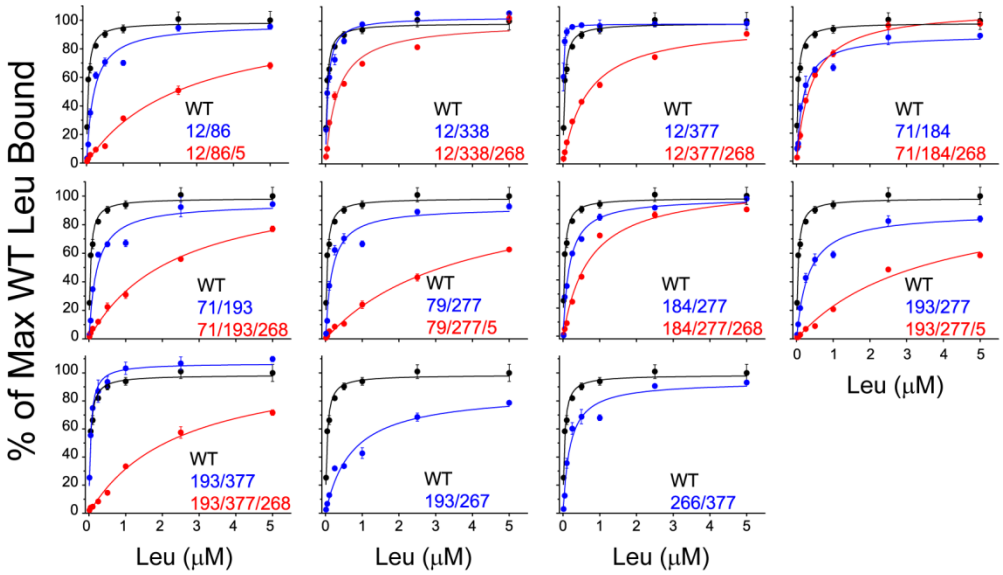
Supplementary Figure 3.3. Element connection and sequence separation weight ratios. Effect of varying Sequence Separation weight relative to Element Connection on model quality measures, C_α-RMSD <7.5 Å (purple) and C_α-RMSD <3.5 Å (green). The best weighting ratio was 1:1, denoted with an asterisk.



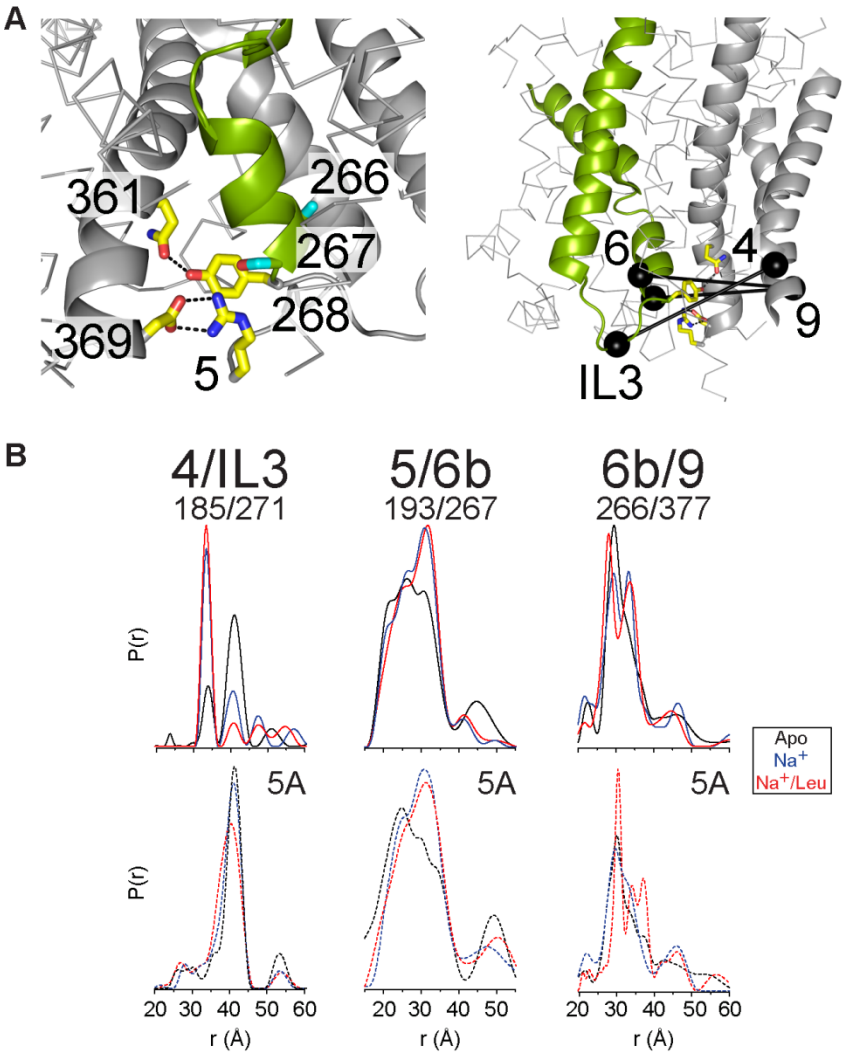
Supplementary Figure 3.4. Effects of solvent exposure and secondary structure prediction on model quality measures. (A) Comparison of predictions to crystal structure definitions for secondary structure and solvent exposure. (B) Effects of excluding buried residues, predicted and ideal. (C) Comparison of predicted (Pred) and Ideal secondary structures for 21 restraints. Predicted secondary structure yielded 8 helices rather than the 7 found in the crystal structure. Therefore to account for the increase in connectivities for 8 helices, we included the effects of using 28 restraints.



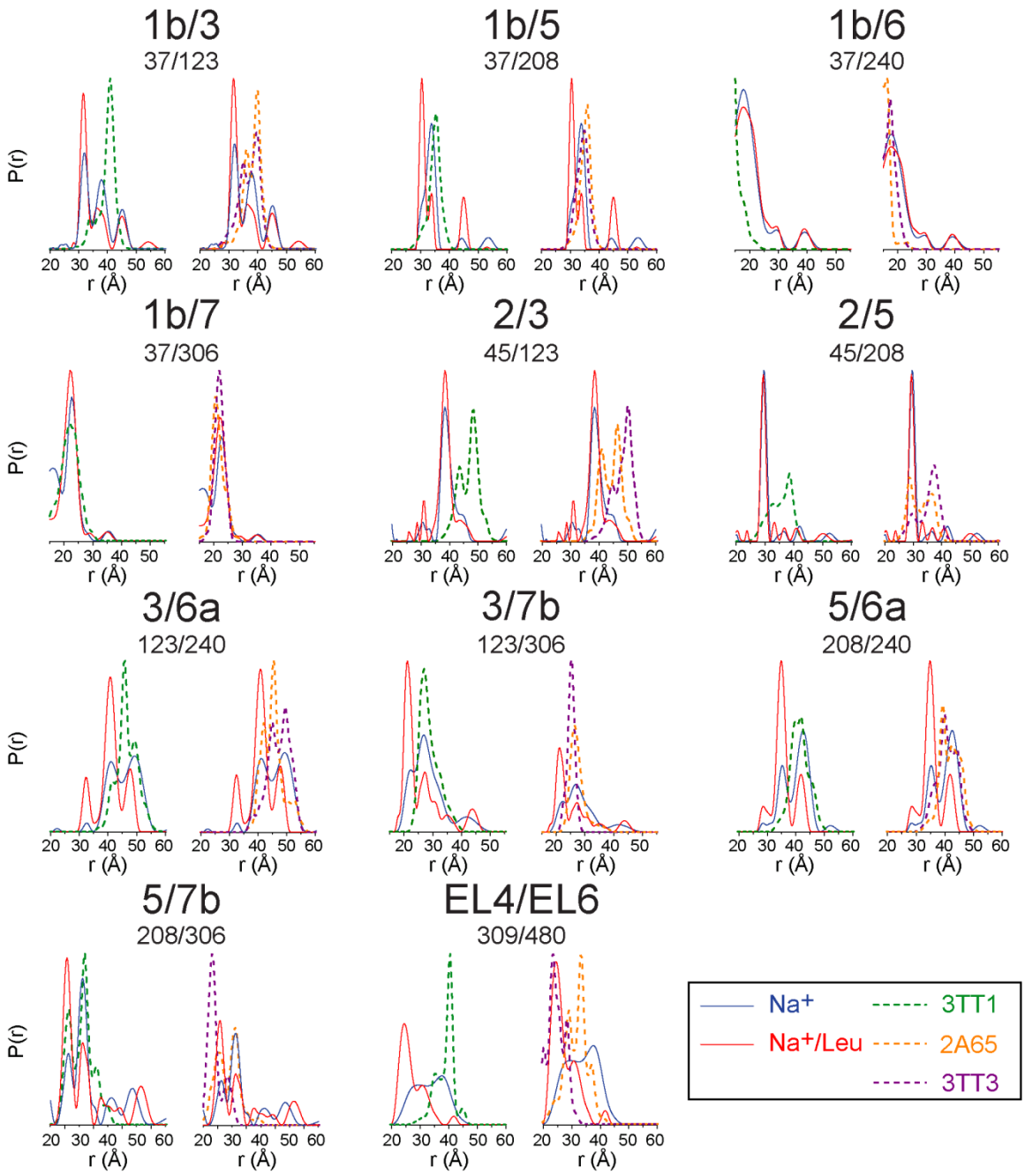
Supplementary Figure 4.1. Na^+-dependent leucine binding to spin-labeled LeuT variants. Leu binding in the presence of Na^+ to spin labeled LeuT was determined by means of the scintillation proximity assay (SPA) as described in the Methods. Data were normalized with respect to the activity of LeuT-WT ($n \geq 3$). Data are shown as the mean ± S.E.M. The data was normalized to the level of leucine binding to the WT, determined concurrently. Most mutants excluding those in TM6 or in the Y268A/R5A mutant backgrounds bind Leu with similar affinity and stoichiometry as the WT.



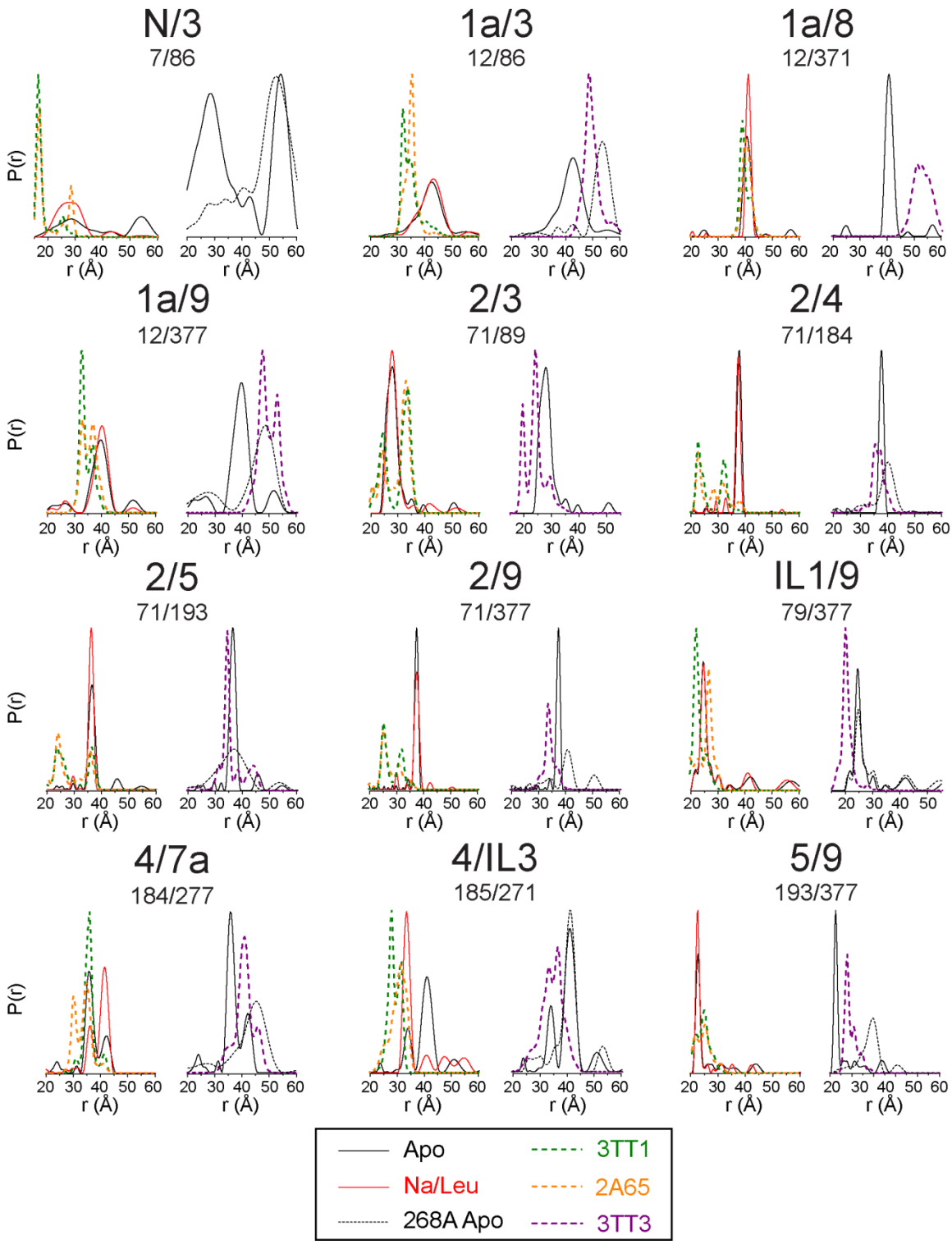
Supplementary Figure 4.2. Leu binding isotherms of destabilized LeuT mutants. For samples that showed low binding levels in Supplementary Figure 1, we constructed binding isotherms to assess the change in affinity. The y axis is the amount of bound leucine normalized to the WT. The right shift of the curve indicates that these mutants, destabilized either by the Y268A/R5A substitutions or spin labeling of TM6, have lower affinity to leucine. Nevertheless, they bind leucine to the same level as WT at high leucine concentrations. DEER distributions are determined at 50-200 μM LeuT concentration in the presence of excess Leu which ensures near complete occupancy.



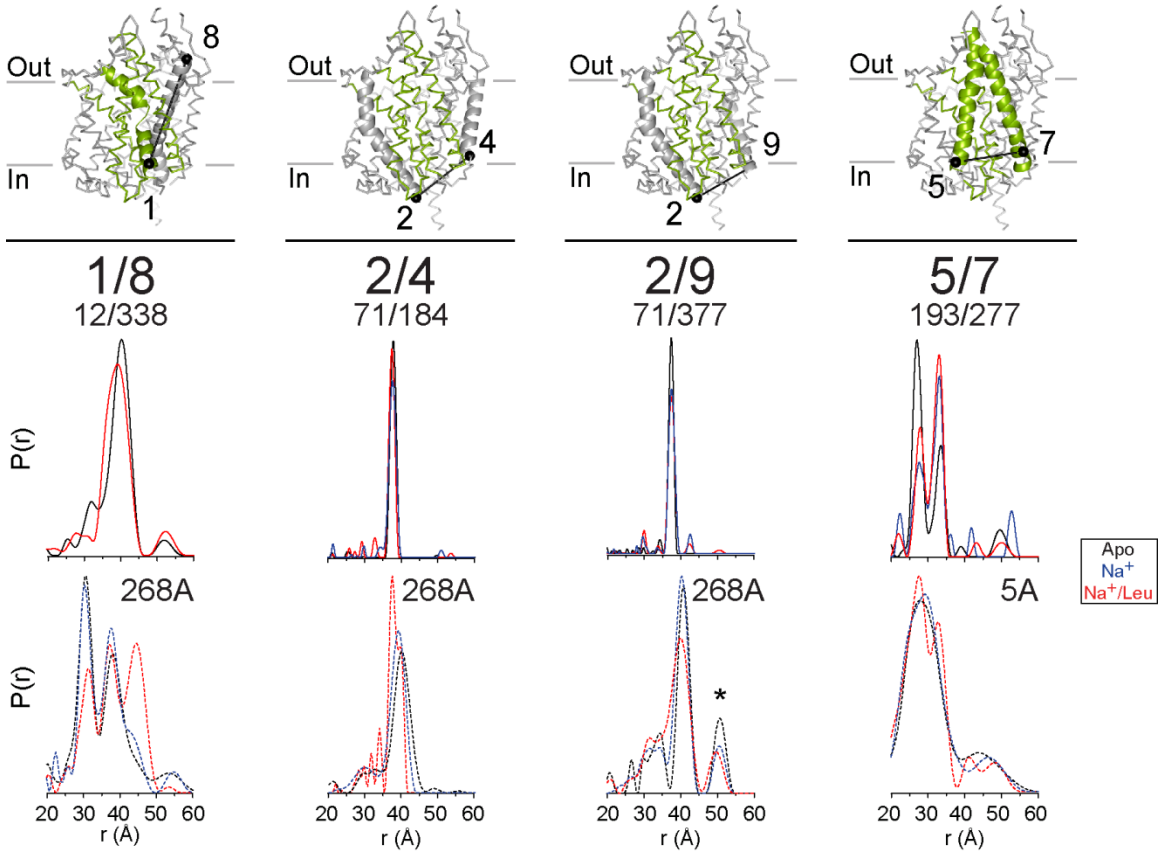
Supplementary Figure 4.3. Spin labeling of TM6 residues. a) (Left) The intracellular network of hydrogen bonding and electrostatic interactions stabilized by Y268A of IL3. (Right) The network of spin label pairs monitoring the movement of TM6b and the loop connecting to TM7 is shown by black spheres connected by solid lines. b) Na^+- and Leu-dependent changes in the distance distributions demonstrating the movement of TM6b in the WT (solid lines) and the 5A (dashed lines) backgrounds. The broad distributions hinder analysis of the magnitude of distance changes but are consistent with the ligand-dependent equilibrium of this TM between multiple conformations



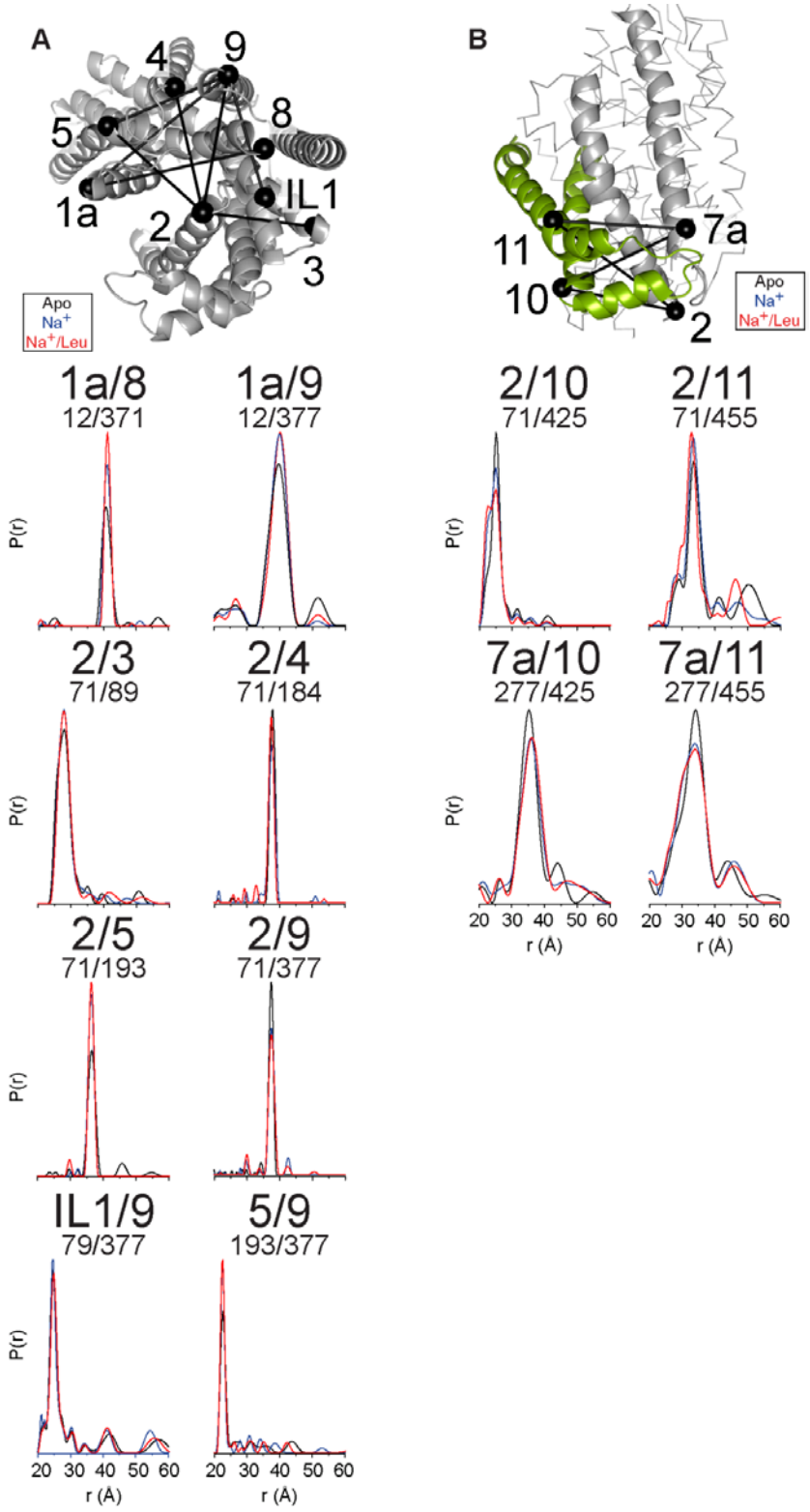
Supplementary Figure 4.4. Crystal structures of LeuT underestimate the closing of the extracellular vestibule. Comparison of experimental distance distributions (solid lines) with predicted distributions based on the "inward-facing" (3TT3), "outward-facing" (3TT1) and "substrate-occluded" (2A65) crystal structures. The MMM package was used to generate predicted distance distributions on the extracellular side. This comparison demonstrates that the experimental distances in the Na^+/Leu-bound state fall outside the predicted distributions regardless of the crystal structure. Because MMM typically overestimates the distribution width, we interpret the systematic deviations between calculated and experimental distributions as evidence that the substrate-occluded conformation we observe in solution is not represented in the crystallographic record. Hence the crystal structures underestimate the closing of the extracellular side upon Na^+/Leu binding. The only exception to this seems to be EL4, for which the predicated distribution from the inward-facing structure overlaps the experimental distribution obtained in the presence of Na^+/Leu.



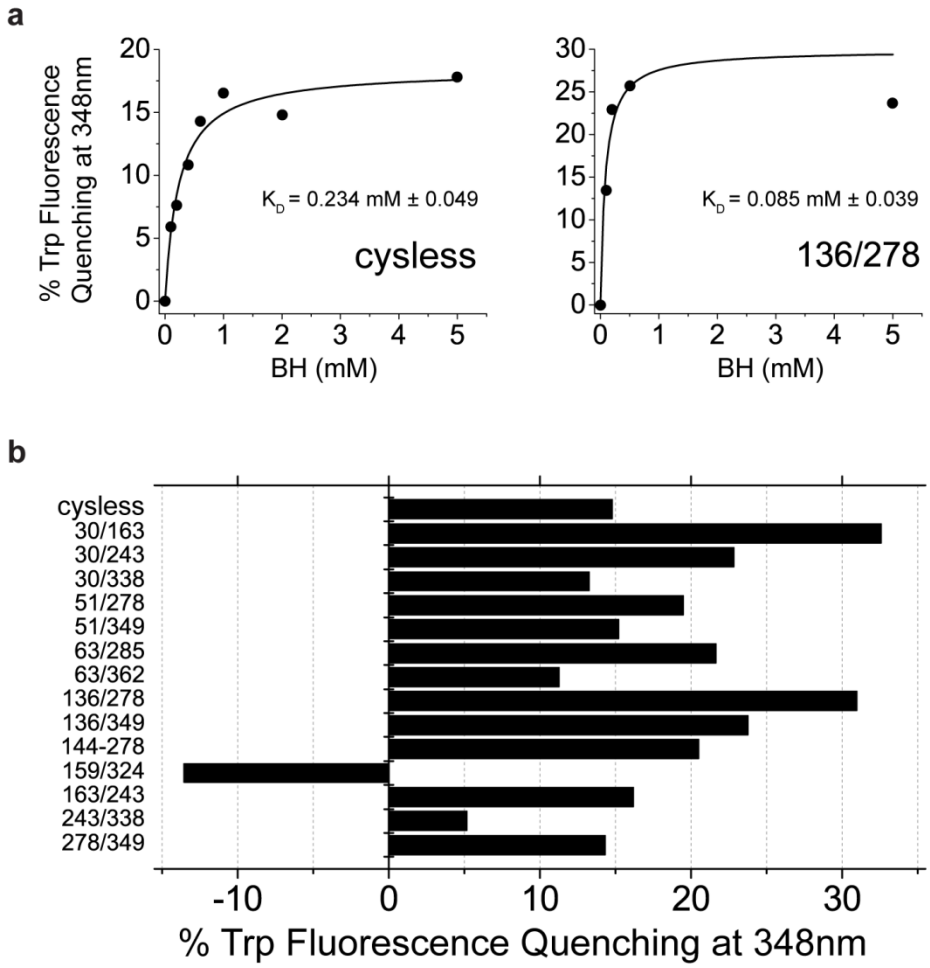
Supplementary Figure 4.5. The "substrate-occluded" crystal structure overestimates the closing of TMs 1 and 2 on the intracellular side. Comparison between experimental and MMM-predicted distances shows systematic deviations in the distributions of TMs 1 and 2. In contrast to the extracellular side, here the average distances are larger than those predicted by the occluded structure.



Supplementary Figure 4.6. Comparison of distance distributions at selected sites in the WT (solid line) and the Y268A and R5A (dashed lines) backgrounds. The Y268A mutation induces movement of TMs 1a and 5 in the direction expected based on the inward-facing crystal structure. The (*) indicates components arising from aggregation during concentration of the mutants after gel filtration. The addition of Na^+/Leu does not reset the distributions back to WT-like. This suggests a loss of conformational coupling. Areas known to be a part of the rigid C structure (grey helices) also show changes in the distance distributions in the Y268A background. We interpret these changes as indicative of the global destabilization effects of these mutations.



Supplementary Figure 4.7. The LeuT scaffold. a) The static scaffold of LeuT. Narrow distance distributions between TMs 2, 3, 4, 5, 8, and 9 suggest a predominantly rigid scaffold. TM1a may undergo small scale movement as indicated by the width change in the TM1a/TM9 distribution. b) Evidence of small scale movements of TMs 10 and 11 relative to TM2. This movement appears to be coordinated with that of TM7a as their pairwise distributions do not show Na^+- and/or Na^+/Leu- dependent changes in average distance or width.

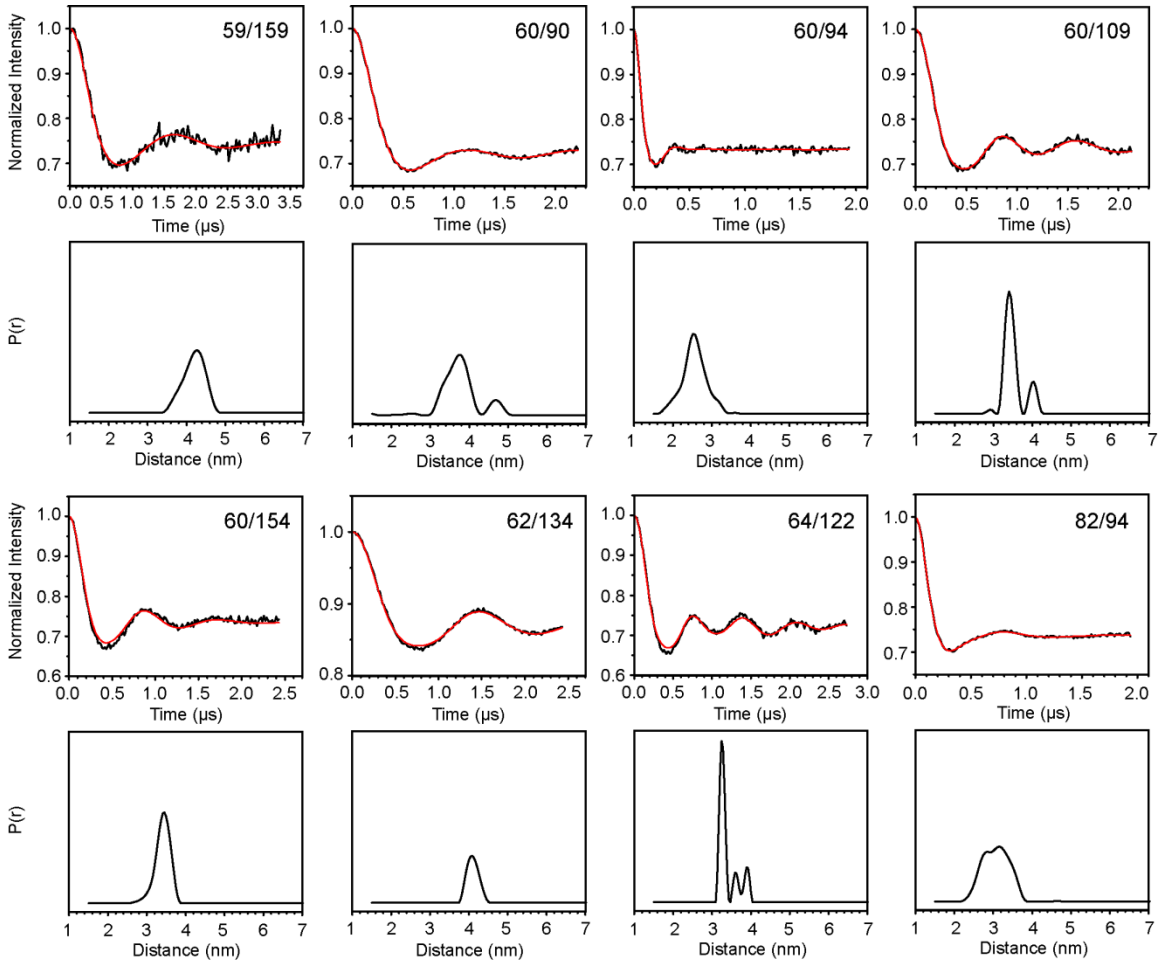


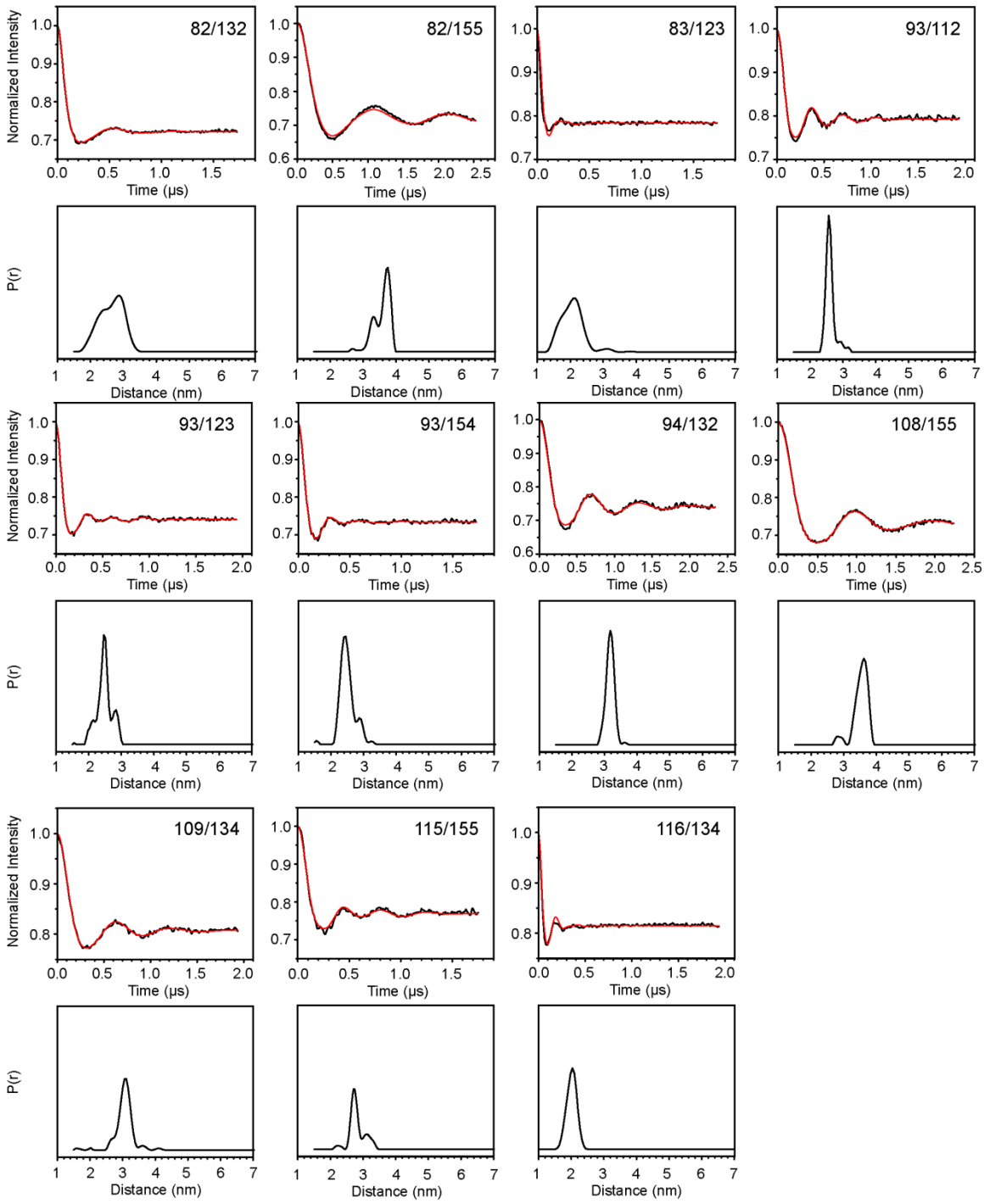
Supplementary Figure 5.1. BH binding to spin-labeled Mhp1 variants. BH binding in the presence of Na^+ to spin labeled Mhp1 was determined by means of the Trp Fluorescence Quenching assay as described in the Methods. a, Complete binding curves for cysless and mutant (136/278) Mhp1. Resulting K_D values are comparable with WT. b, Quenching values for all mutants at 2.5μM Mhp1 and 2mM BH compared to cysless construct.

APPENDIX B

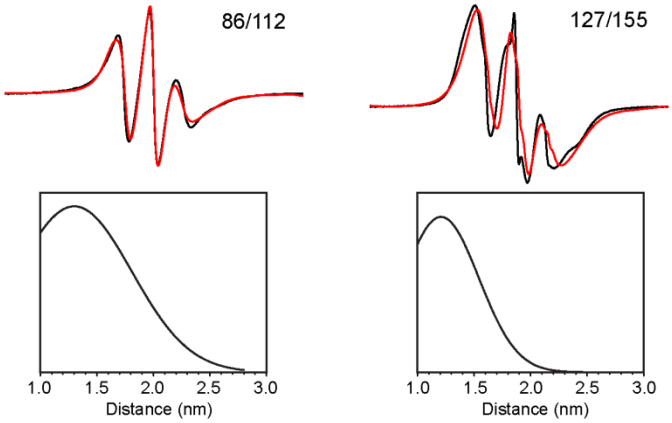
EPR DATA SETS BY PROTEIN AND MUTANT

T4 Lysozyme



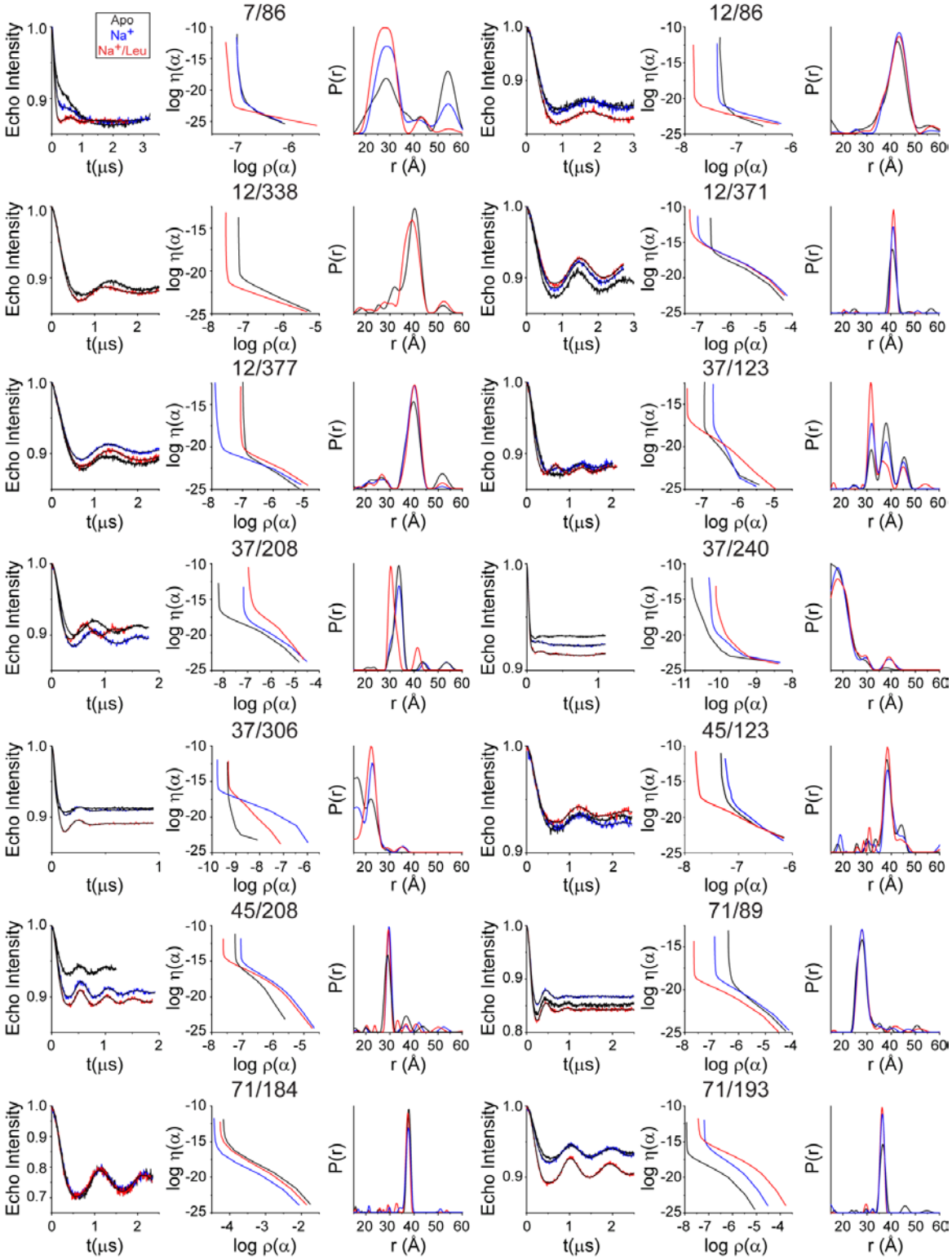


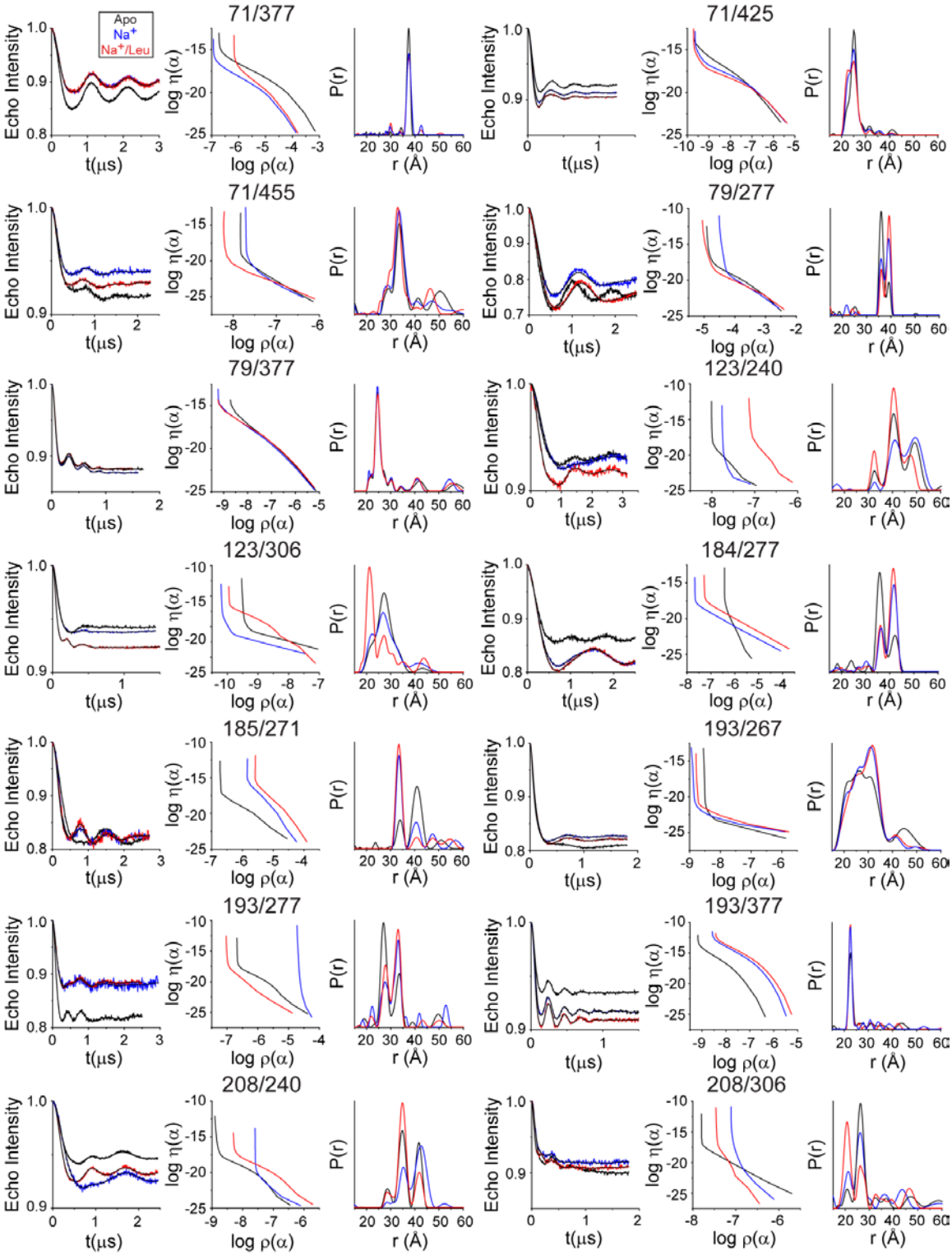
T4L mutants. (Top) Background-corrected decays and (Bottom) normalized distance distribution.

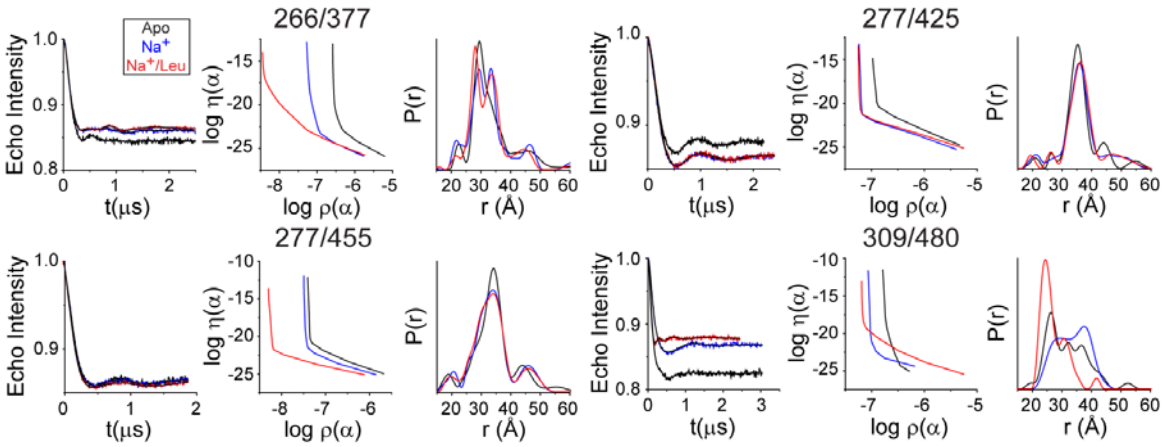


T4L mutants. (Top) Baseline-corrected spin echoes or CW spectra (86/112 and 127/155) along with corresponding distance distributions (Bottom). The experimental data is shown in black, the fits are shown in red.

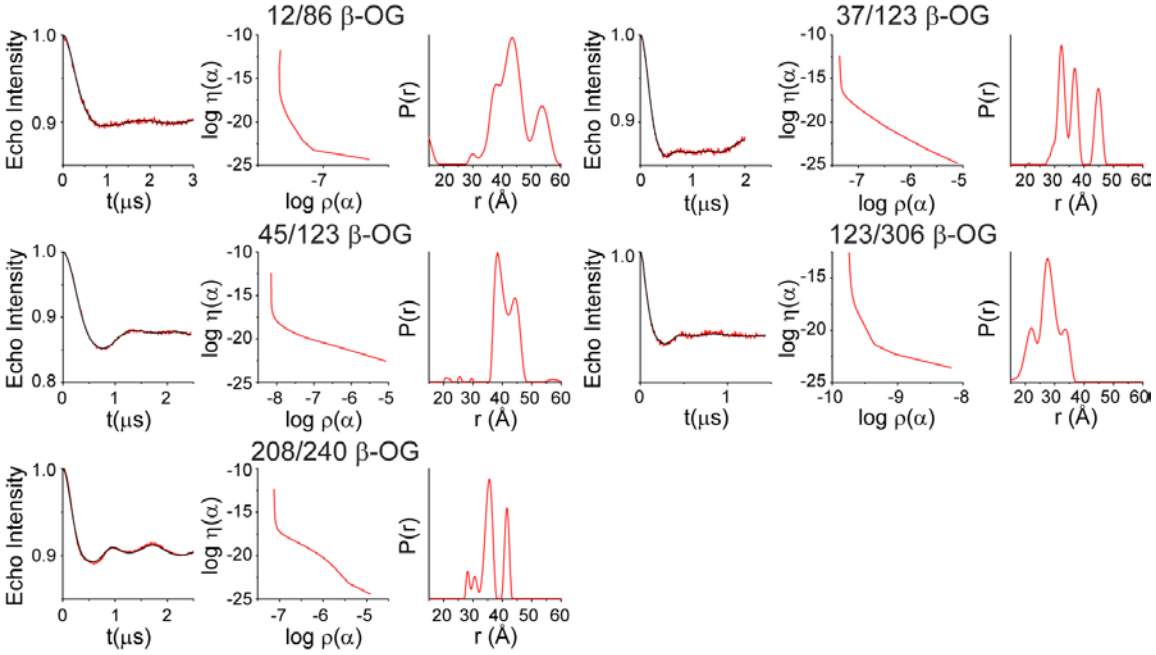
LeuT


Apo
Na+
Na+/Leu
7/86
12/86
12/338
12/371
12/377
37/123
37/208
37/240
37/306
45/123
45/208
71/89
71/184
71/193
Echo Intensity
t(μs)
log η(α)
log ρ(α)
P(r)
r (Å)

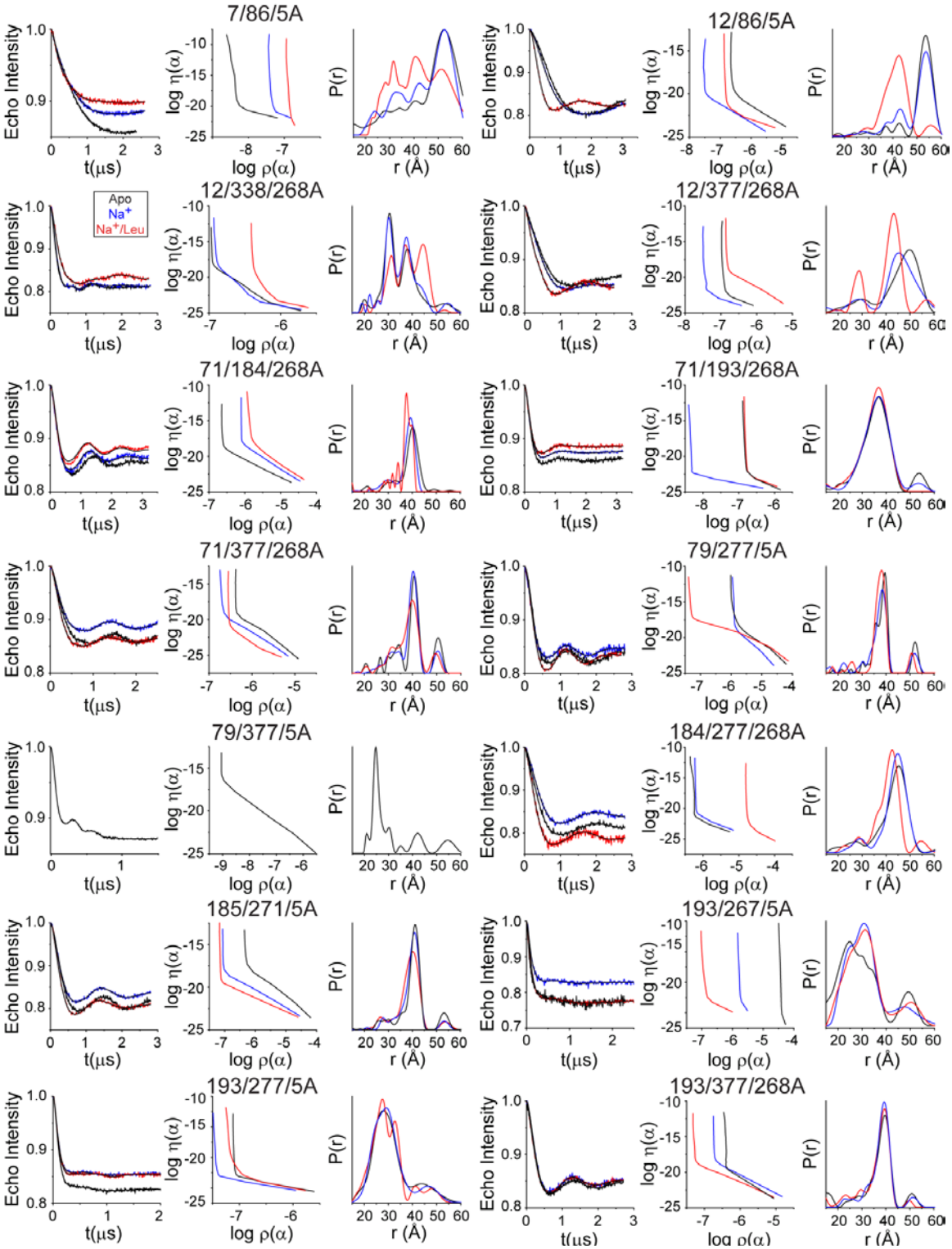

Apo
Na+
Na+/Leu
71/377
71/425
71/455
79/277
79/377
123/240
123/306
184/277
185/271
193/267
193/277
193/377
208/240
208/306
Echo Intensity
t(μs)
log η(α)
log ρ(α)
P(r)
r (Å)

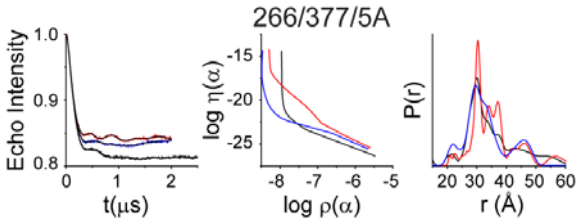


LeuT mutants in WT background. (Left) Background-corrected decays, (Middle) L curves, α parameter selected at elbow, and (Right) normalized distance distributions.



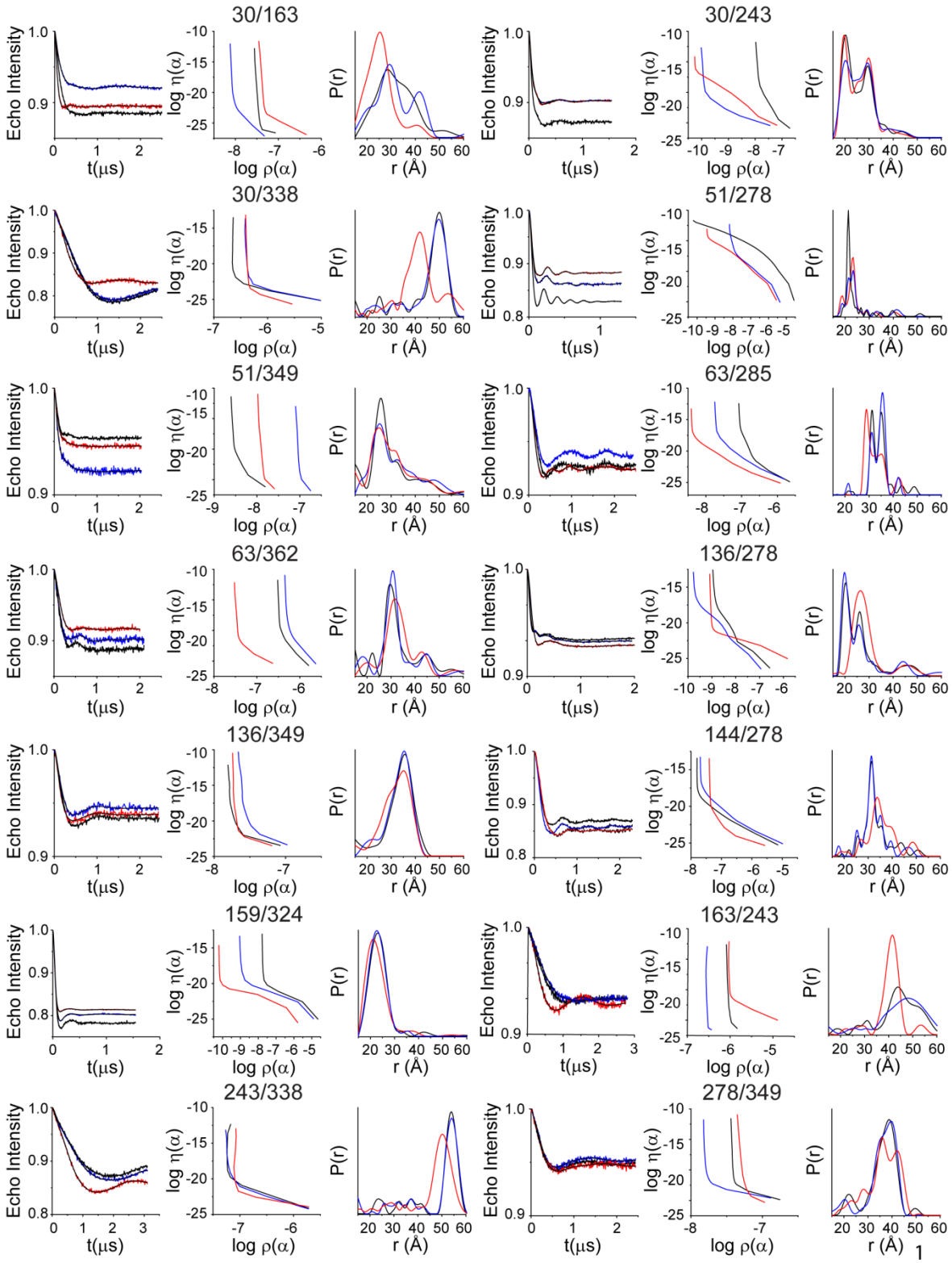
LeuT mutants in WT background in presence of β-OG. (Left) Background-corrected decays, (Middle) L curves, α parameter selected at elbow, and (Right) normalized distance distributions.


7/86/5A
12/86/5A
12/338/268A
12/377/268A
71/184/268A
71/193/268A
71/377/268A
79/277/5A
79/377/5A
184/277/268A
185/271/5A
193/267/5A
193/277/5A
193/377/268A
Apo
Na+
Na+/Leu
Echo Intensity
t(μs)
log η(α)
log ρ(α)
P(r)
r (Å)



LeuT mutants in R5A and Y268A backgrounds. (Left) Background-corrected decays, (Middle) L curves, α parameter selected at elbow, and (Right) normalized distance distributions.

Mhp1



Mhp1 mutants. (Left) Background-corrected decays, (Middle) L curves, α parameter selected at elbow, and (Right) normalized distance distributions.