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

Thijs Beuming, Lei Shi, Jonathan A. Javitch, and Harel Weinstein

Department of Physiology and Biophysics, and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, New York (T.B., L.S., H.W.); Center for Molecular Recognition and Departments of Psychiatry and Pharmacology, College of Physicians and Surgeons, Columbia University, New York, New York (J.A.J.).

Received May 4, 2006; accepted July 31, 2006

ABSTRACT

The recently elucidated crystal structure of a prokaryotic member of the neurotransmitter/sodium symporter (NSS) family (Yamashita et al., 2005) is a major advance toward understanding structure-function relationships in this important class of transporters. To aid in the generalization of these results, we present here a comprehensive sequence alignment of all known prokaryotic and eukaryotic NSS proteins, based on the crystal structure of the leucine transporter from Aquifex aeolicus (LeuT). Regions of low sequence identity between prokaryotic and eukaryotic transporters were aligned with the aid of a number of bioinformatics tools, and the resulting alignments were validated by comparison with experimental data. In a number of regions, including the transmembrane segments 4, 5, and 9 as well as extracellular loops 2, 3, and 4, our alignment differs from the one proposed previously [Nature (Lond) 437: 215–223, 2005]. Important similarities and differences among the sequences of NSS proteins in regions likely to determine selectivity in substrate binding and mechanisms of transport regulation are discussed in the context of the LeuT structure and the alignment.

The neurotransmitter/sodium symporter (NSS) family [TC code 2.A.22 (Saier, 1999)] includes transporters that are responsible for the termination of neurotransmission through uptake of various neurotransmitters, including dopamine, norepinephrine, serotonin, GABA, and glycine. This family of proteins is also referred to as the sodium/neurotransmitter symporter family (SNF) in the Uniprot/Swissprot classification system, or sodium- and chloride-dependent neurotransmitter transporter family (SLC6) in the Human Genome Organization (HUGO) Nomenclature system (Chen et al., 2004), but the term NSS will be used here. A subset of this family, the biogenic amine transporters, which include the dopamine transporter (DAT), norepinephrine transporter (NET), and serotonin transporter (SERT), are the molecular targets for psychostimulant drugs such as cocaine and amphetamine and for many antidepressants. Experimentally well characterized members of the NSS family are mostly from eukaryotic organisms, all of which are predicted to have 12 transmembrane segments (TMs). However, genes encoding more than 200 putative transport proteins homologous to these transporters have recently been identified in prokaryote genomes. TnaT of Symbiobacterium thermophilum, has been shown to be an Na\(^{+}\)-dependent tryptophan transporter (Androutsellis-Theotokis et al., 2003). The structure of another prokaryotic member of the NSS-family, a leucine transporter (LeuT) from Aquifex Aelolicus, has been determined to 1.65-Å resolution by X-ray crystallography (Yamashita et al., 2005). It is odd that the majority of homologous prokaryotic transporters are predicted to have 11 transmembrane segments (TMs). However, genes encoding more than 200 putative transport proteins homologous to these transporters have recently been identified in prokaryote genomes. TnaT of Symbiobacterium thermophilum, has been shown to be an Na\(^{+}\)-dependent tryptophan transporter (Androutsellis-Theotokis et al., 2003). The structure of another prokaryotic member of the NSS-family, a leucine transporter (LeuT) from Aquifex Aelolicus, has been determined to 1.65-Å resolution by X-ray crystallography (Yamashita et al., 2005). It is odd that the majority of homologous prokaryotic transporters are predicted to have 11 TMs (Quick et al., 2006). One of these, a tyrosine transporter that is fully functional with only 11 TMs, has been characterized recently (Quick et al., 2006).

The use of LeuT and other prokaryotic NSS proteins to enable structural and functional inferences regarding their...
been improved where necessary by the consideration of ex-
refined using membrane protein-specific algorithms and has
the context of the LeuT structure. This alignment has been
1), including both prokaryotic and eukaryotic members, in
loops is essential for understanding the structural context of
participate directly in binding (Sen et al., 2005). In addition,
are likely to modulate the conformation of the core TMs that
alter the conformation of TM4 in DAT (Hastrup et al., 2003).
other inhibitors. Cocaine binding has also been shown to
have shown to impair surface expression and lead to
mutation of a residue in the poorly conserved TM9 of NET
and TM12 (Gu et al., 2006; Neubauer et al., 2006) have been
shown to play an important role in the binding of cocaine and
in the case of the amino acid transporters (Tyt1, TnaT, and Glyt1),
in TM1 and TM6, as is the case in the LeuT/leucine complex. Thus,
substrates were positioned in the models by matching equivalent
alignment previously published (Yamashita et al., 2005). Based on
with the structure of LeuT as a template. To investigate the effect of
different alignments on the structural interpretation of biochemical
data, different preliminary models of DAT and SERT were made
data, different preliminary models of DAT and SERT were made
based on the alignment presented here (shown in Fig. 1) and the
alignment previously published (Yamashita et al., 2005). Based on
the assumption that the binding pockets in NSS proteins overlap,
substrates were positioned in the models by matching equivalent positions in the LeuT/leucine complex. In this binding mode,
the charged moieties of the substrates (i.e., the amine in dopamine,
serotin, and norepinephrine and the amino acid backbone in ty-
rosine, glycine, and tryptophan) interact with the unwound regions
in TM1 and TM6, as is the case in the LeuT/leucine complex. Thus,
in the case of the amino acid transporters (Tyt1, TnaT, and Glyt1),
the side chain of leucine was effectively replaced by those of tyrosine,
tryptophan, and glycine, respectively, whereas the coordinates of the
backbone atoms were kept the same. For the biogenic amines (dopa-
mie, norepinephrine, and epinephrine), the amine-nitrogen was
superimposed on the nitrogen of the amino acid, whereas the ar-
omatic ring was superimposed on the leucine side chain. In all cases,
the space available in the binding sites was sufficient to accommo-
date the substrate without major structural rearrangement of the TM
do mains. However, a brief energy minimization using CHARMM
(Brooks et al., 1983) was performed to eliminate any localized steric
incompatibilities.

Results

Creating an Initial Alignment. A nonredundant set of
344 NSS protein sequences, including 177 eukaryotic and
167 prokaryotic members, was aligned as described above.
The following five regions of apparent low conservation in the
NSS family were then considered for manual refinement: 1) the
second extracellular loop (EL2), 2) the region encompassing
TM4, IL2, EL3, and TM5, 3) EL4, 4) TM9, and 5) TM12.
These regions are indicated on a secondary structure map of
the LeuT sequence in Fig. 2. The alignment was refined for
the entire family of prokaryotic and eukaryotic NSS proteins;
for clarity, however, we illustrate the specific refinement
steps and the use of experimental data to guide this refine-

Materials and Methods

Sequence Retrieval and Initial Alignment. Sequences of NSS
family members were collected by blasting against the NCBI RefSeq
database using a BioPerl script (Stujich et al., 2002). At the outset,
224 eukaryotic and 231 prokaryotic sequences were identified.
Re-
moving sequences with \( \geq 95\% \) identity resulted in a total of 177
eukaryotic and 167 prokaryotic proteins. Prokaryotic and eukaryotic
transporters were aligned separately using ProbCons (Do et al.,
2005). After manual adjustment, the two subalignments were
merged by profile-profile alignment using T-coffee (Notredame et al.,
2000). The combined alignment was then subjected to further man-
ual adjustment as described under Results.

Topology Prediction. The use of consensus results from a vari-
ety of algorithms in the prediction of TM-helix topology has been
shown to yield reliable results, and we have previously shown that
the combination of five methods (i.e., TOPPRED2 (von Heijne, 1992),
ORTIENTM (Liakopoulos et al., 2001), HMMTOP (Tusnady and Si-
mon, 2001), TMHMM (Krogh et al., 2001) and MEMSAT3 (Jones et
al., 1994)) can predict the center of a TM segment with an average
ear of two residues (Beuming and Weinstein, 2005). Such consen-
sus topology prediction was applied here to validate the alignments
of TM4 and TM12.

Prediction of Interior and Lipid-Exposed Faces of TM Do-
mains. To predict the interior and surface-exposed faces of TM
domains, we used a method that integrates a knowledge-based sur-
face propensity scale with a conservation criterion to yield the prob-
ability that a particular face of a TM faces the interior or is lipid-
exposed (Beuming and Weinstein, 2004, 2005). The prediction of
interior faces was used to facilitate the alignment of TM9. An align-
ment of 64 eukaryotic sequences was used to predict the interior
residues.

Secondary Structure Prediction. The results from several of
the best-performing secondary structure prediction methods (Koh et
al., 2003) (i.e., PROFphp (Rost et al., 2004), Sspro (Pollastri et al.,
2002), and PSIPRED (Jones, 1999)) were used to guide the align-
ment of the second and fourth extracellular loops (EL2 and EL4).

Molecular Modeling. Models of DAT, SERT, NET, Glyt1, Tyt1,
and TnaT were generated using MODELLER (Fiser and Sali, 2003)
with the structure of LeuT as a template. To investigate the effect of
different alignments on the structural interpretation of biochemical
data, different preliminary models of DAT and SERT were made
based on the alignment presented here (shown in Fig. 1) and the
alignment previously published (Yamashita et al., 2005). Based on
the assumption that the binding pockets in NSS proteins overlap,
substrates were positioned in the models by matching equivalent positions in the LeuT/leucine complex. For example, mutation of residues in TM2 (Wu and Gu, 2003)
and TM12 (Gu et al., 2006; Neubauer et al., 2006) have been
shown to play an important role in the binding of cocaine and
orthostatic intolerance in humans (Hahn et al., 2003). Fur-
thermore, many of the loops in NSS proteins are not con-
served, and although their role in the functioning of LeuT is
not yet well understood, they may contribute to the perme-
ation pathway at a certain stage in the translocation cycle,
and hence play an important functional role (e.g., see Loland et al. (2002)).
Therefore, correct alignment of the less conserved TMs and
loops is essential for understanding the structural context of
the functioning of these transporters. To this end, we present
here a comprehensive alignment of the NSS family (see Fig.
1), including both prokaryotic and eukaryotic members, in
the context of the LeuT structure. This alignment has been
refined using membrane protein-specific algorithms and has
been improved where necessary by the consideration of ex-
perimental data.
Fig. 1. Structure-based alignment of prokaryotic and eukaryotic NSS proteins. Secondary structure elements were taken from the structure of LeuT (Yamashita et al., 2005). Putative membrane spanning regions (indicated with blue bars) were obtained from the PDB TM database (Tusnady et al., 2005). The residues chosen for the structure-based generic numbering scheme are indicated with *H11569. The long N and C termini of the eukaryotic NSS proteins in the alignment have been omitted for clarity. Conserved residues have been shaded according to the following color scheme: basic (H, R, K) in blue, acidic (D, E) in red, aromatic (F, Y, W) in purple, hydrophobic (L, I, V, M) in yellow, polar (N, Q) in orange, proline (P) in brown, and small (G, T, S, C, A) in green. The figure was prepared using TEXshade (Beitz, 2000).
ment with the examples of the alignment of LeuT with DAT and/or SERT. A subset of the final refined alignment is shown in Fig. 1. The complete and updated alignments of the prokaryotic and eukaryotic NSS proteins are available at http://icb.med.cornell.edu/trac.

**EL2.** TM3 and TM4 are connected by EL2, which is the longest loop in NSS proteins, containing 41 residues in LeuT, 65 residues in DAT, and 60 residues in SERT. The LeuT EL2 consists of an extended stretch of 12 residues (Gly125–Asp136) followed by a 16-residue α-helical segment (Pro137–Ile152) and a C-terminal coiled region of 13 residues (Gly153–Ser165) that forms a large turn and connects the loop to TM4.

The results from secondary structure prediction indicate that part of EL2 in the eukaryotic NSS proteins is helical as well, and this region in DAT is predicted to comprise Pro212 to His223. Indeed, this is the only region in EL2 in which both eukaryotic and prokaryotic transporters have some degree of conservation, indicating that the α-helix in EL2 is likely to be a conserved feature of all NSS proteins. The conserved PxxE[Y/F] motif in LeuT and many of the eukaryotic transporters can be used to align the α-helical segments (see Fig. 1). In the case of DAT, this particular alignment of the central α-helix requires that 21 residues be inserted in the coiled region N-terminal to the helix. According to the LeuT-based molecular model, this insertion can be accommodated between the residues aligned to positions Asn133 and Ala134 in LeuT without clashes with the other extracellular loops. Likewise, three additional residues in DAT need to be inserted in the C-terminal part of the loop, and these residues can be placed between the residues aligned to positions Gly157 and Asp158 in LeuT.

**TM4, IL2, and TM5.** Two conserved residues, a Gly in the second intracellular loop (IL2), and an Arg/Lys in the N terminus of TM5 (indicated with an arrow in Fig. 3), can be used to align TMs 4 and 5. In the alignment of Yamashita et al. (2005) residues Gly190 and Lys196 in LeuT were aligned with these two conserved positions, whereas we have chosen Gly186 and Arg193 instead. This alternative requires the introduction of a single residue gap in IL2 for the eukaryotic sequences, and generates a four residue insertion in EL3 for LeuT. This four-residue β-strand insertion in EL3 seems to be a unique feature of LeuT, as the insertion is absent in 453 of 455 prokaryotic and eukaryotic NSS proteins, with LeuT and Q8U1F4 being the only exceptions (see Fig. 3).

The validity of the alignment presented here for TM4-TM5 is supported by a number of observations from sequence analysis and structure/function studies. The first line of evidence comes from the use of the sequence of a prokaryotic tryptophan transporter TntA (Androutsellis-Theotokis et al., 2003), as a comparative intermediate between LeuT and

Fig. 2. Secondary structure map of the entire LeuT sequence. Regions for which the alignment has been validated in this study are indicated in color: light blue for EL2, yellow for TMs 4 and 5, green for EL4, red for TM9, and orange for TM12. Residues in the LeuT binding site (see Table 2) are colored dark blue. Of these binding site residues, those further highlighted with cyan borders are conserved, whereas those with white borders are nonconserved residues in the binding pocket. The conserved residues chosen for reference (n.50) in the generic numbering scheme (see Table 1) have black borders. Regions predicted to be located within the membrane core are demarcated by straight blue lines. For simplicity, TM segments incorporating nonhelical stretches (i.e., TMs 1 and 6) are shown as continuous helices.
DAT. Whereas the sequences of LeuT and DAT are very
dissimilar in the TM4/5 region (defined as Leu233 to
Gly289 in DAT) with only 11% sequence identity, there is
considerable similarity between LeuT and TnaT (34% se-
quence identity) and between TnaT and DAT (30% se-
quence identity), if TnaT is aligned with DAT as proposed
here. If the alignment of Yamashita et al. (2005) is fol-
lowed, the resulting sequence identity between DAT and
TnaT falls to only 4%. Several pairwise homologous re-
gions are indicated in black in Fig. 3. These regions also
have significant homology within the individual prokary-
otic and eukaryotic transporter alignments.

Further support for the alternative alignment we propose
here comes from the analysis of topology prediction for TM4.
The residue in DAT that is predicted (see Materials and
Methods) to be located at the center of TM4 is Ile248. The
central residue of the TM segment in LeuT is Met176 (as
defined in the PDB_TM (Tusnady et al., 2005)). In the align-
ment presented here, Ile248 in DAT and Met176 in LeuT
(shaded white in Fig. 3) are indeed aligned, whereas in the
previously proposed alignment (Yamashita et al., 2005),
Met176 in LeuT is aligned with Leu244 in DAT. Note that
the two possible alignments for TM4-TM5 are shifted by four
residues, or by approximately one helical turn. Therefore, the
method for predicting the most probable interior helical face
(see the application to TM9) cannot distinguish between the
two alignments.

Finally, experimental data on the accessibility of substi-
tuted cysteines in IL2 and TM5 in SERT (Zhang and Rud-
nick, 2005) provides further justification for the alignment
chosen here. Thus, in a membrane preparation, cysteines
substituted for SERT residues Trp271, Gly273, Val274,
Ser277, Val280, Val281 and Thr284 react readily with the
sulfhydryl reagent methanethiosulfinate ethylammonium
(MTSEA). These data are entirely consistent with the present
alignment, which places all these residues on the helical face
of TM5 that points toward the interior of the protein where it
presumably lines part of the transport pathway (see Fig. 4).
In contrast, following the alignment proposed by Yamashita
et al. (2005) would place residues Val280 and Thr284 at the
lipid-exposed face of TM5 where they would be expected to be
much less reactive with MTSEA.

**EL4.** Whereas EL2 and EL3 both contain a central helix
flanked by two coiled regions, EL4 in LeuT adopts a helix-
coil-helix-coil secondary structure pattern, with both helical
elements (EL4a and EL4b) arranged in an approximately
perpendicular orientation. EL4b probably exists in all NSS
proteins, and it seems to have the same length as well as a
100% conserved aromatic residue (Phe324 in LeuT and
Phe391 in DAT) (see Fig. 1). In contrast, EL4a is highly

---

**Fig. 3.** Detailed alignment of TM4, TM5 and EL3. A nonredundant but
representative sequence alignment was generated by randomly selecting
sequences from the complete alignment so that no sequence had >30%
identity to any other sequence. The two conserved residues in IL2 that
were used to align the flanking TMs 4 and 5 are indicated with an upward
arrow. Red bars indicate the predicted TM domains in DAT (see Materials
and Methods), and blue bars the observed TM domains in LeuT. Black
residue letters on a white background indicate the predicted central
residue in TM4 of DAT (Ile248), and the observed central residue in TM4
of LeuT (Met176). White residue letters on a black background indicate
regions that are similar between either DAT and TnaT (left and right) or
between TnaT and LeuT. The common name of each sequence is followed
by the species name and a Uniprot accession number. The figure was
prepared using TEXshade (Beitz, 2000).
The Cβ–Cβ distance between these two residues is approximately 11 Å in our SERT model (Fig. 5, left), which is in excellent agreement with the constraints from the zinc-site. In contrast, this distance is 18 Å in the model based on the alignment by Yamashita et al. (2005) (Fig. 5).

**TM9.** This is another region where there is no similarity between LeuT and eukaryotic NSS proteins. Unfortunately, very few experimental data exist to differentiate alternative alignments. However, the existence of some conservation among eukaryotic transporters enables the identification of buried and lipid-exposed faces, using a previously published method (Beuming and Weinstein, 2004; Beuming and Weinstein, 2005). According to this method, the TM helix was partitioned into 7 contiguous faces (see Fig. 6) AEB, EBF, BFC, FCG, CGD, GDA and DAE. An algorithm (Beuming and Weinstein, 2004) was used to predict the face with the highest average probability of facing the interior of the protein. For DAT, the BFC face in TM9 has the highest average interior probability score. This face includes small and polar residues Glu446, Thr449, Thr456, Ser460, Cys463 and Gly467. If TM9 in DAT and LeuT are aligned to maximize the overlap between the residues on the predicted interior face of DAT and the observed interior face of LeuT, the resulting alignment is that shown in Fig. 6B. Residues on the BFC face of DAT, and buried residues in LeuT, are shaded in black. This alignment is different from that of Yamashita et al. (2005), in which the buried face of LeuT is aligned with the CGD face of DAT. It is noteworthy that Fig. 7 shows that the previous alignment also places the polar residues Thr449, Thr456, and Ser460 on the lipid-facing surface of TM9. For these reasons, we recommend our alignment in Fig. 6B as being more consistent with the polar conservation of TM segments.

**TM12.** TM12 is among the least conserved regions in NSS transporters, and it is even absent in the majority of prokaryotic NSS proteins, despite their preserved function as sodi-
Fig. 5. A comparison of accessibility patterns and Zn\(^{2+}\)-binding residues in the fourth extracellular loop (EL4) of NSS proteins in two different models of SERT. The two models were based on the alignment presented here (left) or on the alignment of Yamashita et al. (2005) (right). The backbone of EL4 is shown as a purple ribbon, and the rest of the transporter is shown in a gray surface representation. Positions at which substituted cysteine mutants reacted with [2-(trimethylammonium)ethyl]methanethiosulfonate bromide or sodium (2-sulfonatoethyl)-methanethiosulfonate with a rate greater than 1000 M\(^{-1}\) min\(^{-1}\) are colored yellow (Mitchell et al., 2004). In the model based on the alignment presented here, the accessible positions in EL4a face the accessible positions in EL4b and are in close proximity. In the model based on the alignment of Yamashita et al. (2005), it is more difficult to accommodate the substituted-cysteine accessibility method results because the accessible residues in EL4a and EL4b face different directions, and the most reactive residues in EL4a are buried toward the interior. The two residues with side chains rendered in sticks are part of endogenous Zn\(^{2+}\)-binding site that has been characterized in DAT (Norregaard et al., 2000), SERT (Mitchell et al., 2004), and GAT (MacAulay et al., 2001). In our model, the distance between the Cβ atoms of the residues involved in Zn\(^{2+}\) binding is ~11 Å (Zn\(^{2+}\) shown in green); in contrast, the model based on the alignment of Yamashita et al. (2005) places them much further apart (~18 Å), in orientations that are unlikely to enable Zn\(^{2+}\)-binding. The figure was prepared using PyMOL (Delano, 2002).

Fig. 6. Prediction of the interior face of TM9 in DAT. A, the BFC face in TM9 has the highest average interior probability score (Beuming and Weinstein, 2004). This face includes small and polar residues Glu446, Thr449, Thr456, Ser460, Cys463, and Gly467. B, alignment of TM9 in DAT and LeuT in which the overlap is maximized between the residues on the predicted interior face of DAT and on the observed interior face of LeuT. Residues on the interior-predicted BFC face of DAT and interior residues in LeuT are shown in white on a black background (top). In the alignment of Yamashita et al. (2005) (bottom), the interior residues of LeuT are aligned with the CGD face of DAT, which places polar residues Thr449, Thr456, and Ser460 on the lipid-facing surface of TM9.
Consensus topology prediction of DAT indicates that the central residue of TM12 is Ser568 (see Fig. 8). Six residues (Gly561, Trp562, Ser567, Ser568, Val572 and Pro573) are relatively conserved in the putative TM12 in eukaryotic NSS proteins, but unlike TM9, no unambiguous interior and exterior faces can be identified for TM12. We have chosen to align TM12 in DAT and LeuT so that residues Gly561, Trp562, Ser568 and Val572 are buried in the interior of the protein, while the centrally predicted residue in DAT (Ser568) is kept as close to the center of TM12 in LeuT (Phe494) as possible. These criteria result in the alignment shown in Fig. 8, which is consistent with that proposed by Yamashita et al. (2005). Note, however, that our small change in the position of the gap between TM11 and TM12 from that of Yamashita et al. (2005) extends the alignment to Trp484 (Fig. 8). This alignment places Pro573 on the lipid-exposed surface of TM12, and it is conceivable that this residue is conserved because it produces a structurally important kink at the C-terminal end of TM12.

**Fig. 7.** Conserved small and polar residues in TM9. Models of DAT were based on the alignment presented here (top) or on the alignment by Yamashita et al. (2005) (bottom). The TM bundle is viewed from the side (left) or from the extracellular side (right). Loops have been omitted for clarity. TM9 is shown in magenta. Residues shown as spheres are predicted to face the interior in eukaryotic transporters. In the alignment of Yamashita et al. (2005), these conserved polar residues [Thr449 (9.31), Thr456 (9.38), and Ser460 (9.42), shown in cyan] are located on the lipid-facing surface of TM9. The figure was prepared using PyMOL (Delano, 2002).

**Fig. 8.** Detailed alignment of TM12. To indicate conservation within the eukaryotic family, all known sequences with >35% identity to DAT are shown. The top red bar indicates the consensus prediction of the location of TM12 in DAT. The blue bar indicates the observed location of TM12 in LeuT. The height of the orange bars on the top indicates the level of conservation for the eukaryotic sequences, and the red bars on the bottom indicate the buried residues in LeuT. DAT and LeuT are aligned so that DAT residues Gly561 (12.38), Trp562 (12.39), Ser568 (12.45), and Val572 (12.49) (all indicated with *) are buried in the interior of the protein, and the centrally predicted residue in DAT [Ser568 (12.45), indicated with #] is kept as close as possible to the center of TM12 in LeuT [Leu493 (12.45), indicated with #]. The figure was prepared using TExshade (Beitz, 2000).
Generic Structure-Based Numbering Scheme. We previously proposed a residue numbering scheme to facilitate comparison of the sequences of different NSS proteins (Goldberg et al., 2003). In this scheme, the most conserved position in the sequence alignment is chosen for each TM segment, and this position is assigned the number 50. Other positions are numbered relative to this reference position; e.g., positions directly N- and C-terminal are designated 49 and 51, respectively. For example, the most conserved residue in TM1 is a tryptophan (Trp84 in DAT), and its generic number is 1.50. In DAT, this tryptophan is then referred to as Trp<sub>1.50</sub>. A similar numbering scheme was initially developed for G-protein-coupled receptors (Ballesteros and Weinstein, 1995). The numbering proposed in Goldberg et al. (2003) was based on a sequence alignment analysis performed before the structure of LeuT was available. For TM helices 1 to 4, 6 to 8, and 11 this numbering scheme remains valid in the present alignment. However, in the structure of LeuT, the reference residues proposed as X.50 for TMs 5, 9, 10, and 12 are located in loop regions, where, despite their high conservation, their use as index positions may be complicated in various NSS proteins by the presence of different insertions and deletions. Therefore, new reference residues are proposed here for these 4 TMs, based on 1) the structurally informed alignment, 2) the extent of conservation in the large alignment combining prokaryotic and eukaryotic transporter sequences, and 3) an additional criterion that the index residue be located within the putative TM segment. Furthermore, we have selected the use of residues such as Pro and aromatics that are less common in the TM helices of membrane proteins (Beuming and Weinstein, 2004) over hydrophobic residues such as Val or Leu, which are more abundant in TM domains. For segments with no common conserved residue in both prokaryotic and eukaryotic transporters, we choose the conserved residue in eukaryotic transporters. The 12 amino acids chosen as reference residues in the Generic Structure-Based Numbering Scheme are shown in Table 1, and indicated with ∗ in Fig. 1; eight of these residues are unchanged from our previously proposed system (Goldberg et al., 2003). A tool for calculating the generic numbering can be found at http://icb.med.cornell.edu/trac.

The Binding Site. In LeuT, residues from TMs 1, 3, 6, 7, and 8 interact directly with leucine and/or the two sodium ions (Yamashita et al., 2005). As noted above, these TMs can be aligned readily because of their high overall conservation. Analysis of the conservation of the binding site residues in the context of the complete alignment leads to a striking dissociation of the contact residues into those that are conserved from those that are not conserved between subgroups (Table 2). For the most part, residues that contact the two sodium ions and the carboxyl or amino moieties of leucine are highly conserved and are arranged at the extracellular side of the binding site (shown in cyan in Fig. 9). Because most of the NSS substrates identified to date are amino acids (see below for a discussion of the exceptions, the biogenic amines) and because sodium seems to provide the driving force in all NSS proteins, the conservation of these contacts bespeaks a shared mechanism of substrate recognition. In contrast to the high degree of conservation at these positions, the residues that interact with the aliphatic side chain of leucine at the intracellular part of the binding site are not conserved. The side chains of other amino acid NSS substrates are quite diverse, and include aromatic rings (tryptophan/tyrosine), branched or unbranched chains (GABA/creatine), or even the absence of a side chain (glycine). The prokaryotic NSS proteins we have studied show quite dramatic substrate specificity. TnaT transports tryptophan with an apparent affinity of less than 1 μM, whereas other aromatic amino acids do not affect transport at 100-fold higher concentrations (Androutsellis-Theotokis et al., 2003). Likewise, Tyt1 transports tyrosine with high apparent affinity, whereas phenylalanine and tryptophan do not inhibit transport at vastly higher concentrations (Quick et al., 2006). This exquisite specificity is likely to result for the most part from interactions with these nonconserved residues that line the more intracellular portion of the binding site (white in Fig. 9).

Discussion

As demonstrated previously for other membrane proteins, the availability of the LeuT structure will enable new interpretation of biochemical and pharmacological data on NSS proteins in a structural context. In addition, the structure will enable the development and testing of more sophisticated hypotheses regarding the structural basis of transport. Models of clinically important NSS proteins such as DAT (Sen et al., 2005) and SERT (Henry et al., 2006) are already being built based on the structure of LeuT. For such efforts to succeed, the correct alignment of the sequences of LeuT and other NSS proteins is critical. Thus, an accurate alignment is the most important single component in building a homology model (Fiser and Sali, 2003). Therefore, we have developed a comprehensive sequence alignment of prokaryotic and eukaryotic NSS proteins based on the structure of LeuT and guided by the results of specific bioinformatics algorithms and by a large body of experimental data, such as side-chain accessibility data and metal ion binding site constraints.

Not surprisingly, the alignment of LeuT with eukaryotic NSS family members is complex, partly because LeuT is a nonrepresentative NSS protein, even among the prokaryotes. Indeed, the average sequence identity of LeuT to eukaryotic NSS proteins is ~18%, which is among the lowest values of the 231 bacterial and archaeal NSS proteins. Many other prokaryotic family members have substantially higher simi-

---

**TABLE 1**

Structure-based generic numbering of NSS proteins

<table>
<thead>
<tr>
<th>TM</th>
<th>hDAT</th>
<th>LeuT</th>
<th>Conservation in Eukaryotic Transporters</th>
<th>Conservation in Prokaryotic Transporters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Trp84</td>
<td>Leu29</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>Pro112</td>
<td>Pro57</td>
<td>100</td>
<td>79</td>
</tr>
<tr>
<td>3</td>
<td>Tyr156</td>
<td>Tyr108</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Cys243</td>
<td>Val171</td>
<td>77</td>
<td>Not conserved</td>
</tr>
<tr>
<td>5*</td>
<td>Pro273</td>
<td>Leu202</td>
<td>95</td>
<td>Not conserved</td>
</tr>
<tr>
<td>6</td>
<td>Gln317</td>
<td>Gln250</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>7</td>
<td>Phe365</td>
<td>Ser288</td>
<td>95</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>Phe412</td>
<td>Phe345</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>9*</td>
<td>Phe457</td>
<td>Phe387</td>
<td>86 (F/Y)</td>
<td>55 (F/Y)</td>
</tr>
<tr>
<td>10*</td>
<td>Phe478</td>
<td>Trp406</td>
<td>93 (F/Y)</td>
<td>25 (F/Y)</td>
</tr>
<tr>
<td>11</td>
<td>Pro529</td>
<td>Pro457</td>
<td>97</td>
<td>81</td>
</tr>
<tr>
<td>12</td>
<td>Pro573</td>
<td>Thr498</td>
<td>95</td>
<td>Not conserved</td>
</tr>
</tbody>
</table>

* Different from the previously proposed numbering system (Goldberg et al., 2003).
### Table 2
Conservation of binding-site residues in NSS proteins

The generic numbers and the residue numbers in LeuT are shown in the first column. The next five columns list the various contacts of the two sodium ions and the leucine substrate, as observed in the LeuT structure. The contact residues for leucine in LeuT were defined as those residues with different solvent-accessible surface areas in the presence and absence of bound substrate. A 2.5 Å distance cut-off was used to determine all Na-coordinating residues. Residues from prokaryotic NSS-family members with known substrates (LeuT, TnaT and TtyT) are shown in the next three columns, and the two columns after that list the variability in 167 eukaryotic NSS-proteins. Residues from eukaryotic NSS-proteins with known substrates are shown in the next 11 columns. These include transporters for DAT, NET, SERT, betaine (BET), GABA (GAT), taurine (TAUT), creatine (CREAT), glycine (GLYT), proline (PROT), neutral amino acids (NCAAT), and neutral and cationic amino acids (NCAAT). Finally, the two columns farthest to the right show the variability in 17 eukaryotic NSS-proteins. Conserved residues shown in **cyan** in Figure 8 are indicated in bold. Contact positions classified as conserved were those in all NSS-proteins with known substrates that featured very similar residue types (Ser and Thr, or Asn and Asp, etc.). Position 1.45 (shown in green in Figure 8), is a Gly in all NSS proteins, except in the homocysteine transporters, where it is an Asp, and is shown in italic. Amino acids are referred to by the one-letter codes.

<table>
<thead>
<tr>
<th>Generic No. (LeuT No.)</th>
<th>Contacts</th>
<th>Prokaryotic</th>
<th>Eukaryotic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Side Chain</td>
<td>LeuNa1Na2LeuT</td>
<td>TnaTTtyTL</td>
</tr>
<tr>
<td>1.41 (20)</td>
<td>/</td>
<td>G GG G G G (98%)</td>
<td>4</td>
</tr>
<tr>
<td>1.42 (21)</td>
<td>/ bs</td>
<td>N N NS S S (70%), A(21%)</td>
<td>8</td>
</tr>
<tr>
<td>1.43 (22)</td>
<td>/ b</td>
<td>A A A A A (89%), T(8%)</td>
<td>5</td>
</tr>
<tr>
<td>1.44 (23)</td>
<td>/ bs</td>
<td>V V V V V F V I, I G V V V V V V V</td>
<td>6</td>
</tr>
<tr>
<td>1.45 (24)</td>
<td>/ b</td>
<td>L L L M L(81%), F(7%),</td>
<td>7</td>
</tr>
<tr>
<td>1.46 (25)</td>
<td>/ b</td>
<td>G GG G G G (96%)</td>
<td>2</td>
</tr>
<tr>
<td>1.47 (26)</td>
<td>/ b</td>
<td>G GG G G G (96%)</td>
<td>2</td>
</tr>
<tr>
<td>1.48 (27)</td>
<td>/ s</td>
<td>N N N N N(67%), A(27%)</td>
<td>5</td>
</tr>
<tr>
<td>3.46 (104)</td>
<td>/ s</td>
<td>V V I I I(83%), V(20%), V(15%)</td>
<td>5</td>
</tr>
<tr>
<td>3.50 (108)</td>
<td>/ bs</td>
<td>Y Y Y F(59%), Y(38%)</td>
<td>3</td>
</tr>
<tr>
<td>6.53 (253)</td>
<td>/ bs</td>
<td>F FF F F F(96%)</td>
<td>3</td>
</tr>
<tr>
<td>6.54 (254)</td>
<td>/ bs</td>
<td>T SS S S S(57%), T(26%)</td>
<td>6</td>
</tr>
<tr>
<td>6.56 (256)</td>
<td>/ bs</td>
<td>S S S S S(81%), T(10%), G(5%)</td>
<td>5</td>
</tr>
<tr>
<td>6.59 (259)</td>
<td>/ s</td>
<td>F V G(27%), V(20%), M(18%), E(16%), D(9%), G(9%), A(7%)</td>
<td>11</td>
</tr>
<tr>
<td>6.61 (261)</td>
<td>/ s</td>
<td>A V G I(31%), V(26%), A(10%), S(9%), T(7%), G(7%), C(7%)</td>
<td>9</td>
</tr>
<tr>
<td>7.38 (286)</td>
<td>/</td>
<td>N N N N N N D(50%), D(40%), T(8%)</td>
<td>6</td>
</tr>
<tr>
<td>8.56 (351)</td>
<td>/</td>
<td>A A A A A(98%)</td>
<td>5</td>
</tr>
<tr>
<td>8.59 (354)</td>
<td>/</td>
<td>T S S S(77%), S(21%)</td>
<td>5</td>
</tr>
<tr>
<td>8.60 (355)</td>
<td>/ s</td>
<td>S S S S S(90%), T(8%)</td>
<td>3</td>
</tr>
<tr>
<td>8.63 (358)</td>
<td>/ s</td>
<td>A S N N S S(90%), A(6%), T(5%), N(5%)</td>
<td>7</td>
</tr>
<tr>
<td>8.64 (359)</td>
<td>/ s</td>
<td>I Q M M(38%), L(36%), I(21%)</td>
<td>7</td>
</tr>
</tbody>
</table>
larity to eukaryotic NSS; for example, the tryptophan transporter TnαT has an average similarity of 28% to the eukaryotic NSS proteins. Because we expect the overall folds of these proteins to be extremely similar, we have attempted to validate our alignment between LeuT and the eukaryotic NSS proteins throughout all parts of the protein. One major structural difference is the apparent deviation in the C-terminal end of the protein, considering the majority of NSS that only contain 11 TMs (Quick et al., 2006). Other differences in sequence are likely to lead to differences in local structural features (i.e., kinks, bulges, or other distortions), which are likely to be most pronounced in the less well conserved regions of the proteins (Vardy et al., 2004).

The loop regions connecting the TM helices are most variable in sequence, with numerous insertions and deletions, and the N and C termini of DAT, SERT, and other eukaryotic NSS proteins are much longer than those of LeuT (or of most other prokaryotic NSS proteins). Therefore, the structure of LeuT cannot be used for straightforward homology modeling of these parts of the proteins. As for other membrane proteins, additional structural information in combination with molecular modeling and simulations will be required to understand the structural implications of these local differences and their functional consequences.

Despite these expected differences, the high conservation among NSS proteins in the TMs surrounding the binding site (i.e., TMs 1, 3, 6, and 8) indicate the appropriateness of LeuT as a structural template for molecular modeling. Although conservation in these TMs is sufficient to allow unambiguous alignment, we would expect to see substantial differences in the nature of the residues that form the substrate binding site, because these residues are important for the specificity of these transporters. That said, it is remarkable that many of the residues in LeuT that interact either with the amino or carboxylate groups of the substrate or with the two sodium ions are conserved in both eukaryotic and prokaryotic NSS proteins (see above and Fig. 9). In many cases, the side chains of these conserved residues form essential contacts with leucine in LeuT and presumably with the substrate in homologous transporters as well (see Table 2). However, at other positions, particularly in TM1 and TM6, contact with leucine or the sodium ions are mediated by main chain atoms and not by the side chains. Nonetheless, the majority (8 of 11) of these residues are highly conserved as well. These eight conserved positions included residues in TM1 (positions 1.41, 1.44, 1.45, 1.46, and 1.47), TM6 (6.53 and 6.54), and TM8 (8.56). Positions 1.42, 1.43, and 6.56 are not conserved. We presume that the extensive conservation of the side chains in main chain contact positions serves to shape the binding site by maintaining the distortions that are necessary to coordinate substrate and the sodium ions.

Yamashita et al. (2005) noted that the residues that coordinate Na2 (the second sodium ion) are less well conserved than those that contact Na1. However, of the five residues that contact Na2, four are highly conserved in eukaryotic NSS proteins (Table 2). In three of these, the main chain contacts Na2 (Table 2). Of the side chains that contact Na2, Ser8.60 is also highly conserved in prokaryotic and eukaryotic NSS. The other side chain contact, residue 8.59, is a Thr in LeuT (and in 77% of prokaryotic NSS-members) and a Ser in TnαT and Tyt1 (and in 21% of prokaryotic NSS), which also seem to have a stoichiometry of two sodium ions: one substrate molecule (Androutsellis-Theotokis et al., 2003; Quick et al., 2006). In eukaryotic NSS proteins, Asp, Gly, and Ser are present at 8.59 (56, 25, and 11%, respectively). Thus, in the prokaryotic NSS proteins, Ser can substitute for Thr in...
the interaction with Na2, and it is likely that Ser and Asp could substitute in the eukaryotic NSS as well. Indeed, there is experimental evidence for a stoichiometry of 2 Na+; one substrate for several eukaryotic NSS with Asp [i.e., GAT-1 or DAT (Krueger, 1990; Krause and Schwarz, 2005)] or Gly (Roux and Supplisson, 2000) at 8.59, which suggests to us a conserved function for the Na2 binding site. Given the apparent conservation of both sodium sites in eukaryotic NSS proteins, the structural basis for the observed stoichiometry of one sodium ion: one substrate molecule in SERT (Rudnick, 1998) and NET (Gu et al., 1998), which also contain an Asp at position 8.59, is not clear.

A notable exception to the conservation of contact residues is position 1.45, which is an Asp in the entire biogenic amine NSS subfamily, and a Gly in the amino acid NSS subfamily. In the LeuT/leucine complex, the carboxylate group of the substrate interacts with the backbone carbonyl at position 1.45 (Yamashita et al., 2005), and it can be expected that in the biogenic amine transporters, the negatively charged side chain of Asp1.45 interacts with the protonated amine of the substrates. Mutation of Gly1.45 to Asp in TnaT (Androutsellis-Theotokis et al., 2003) and Tyt1 (Quick et al., 2006) abolished tryptophan and tyrosine transport, respectively, but in neither case did this mutation lead to transport of the corresponding biogenic amine (tryptamine or tyramine), indicating that the conversion of an amino acid transporter to a biogenic amine transporter is more complex than this single residue change.

Our analysis of conservation patterns suggests a critical role of the nonconserved residues deeper in the binding site in determining specificity. This is indicated by an apparent complementarity between these residues and the nature of the moiety representing the “side chain” in the substrates. For example, note the correlation of the bulk of the side chains at positions 1.64 and 8.64 with the bulk of the substrate side chain in the model of the complex shown in Fig. 9. In these initial models obtained by simple mapping of the substrate into the homology model (see Materials and Methods for details), the smallest substrate, glycine, is paired with Trp6.59 in Glyt1, whereas 6.59 is a glycine in Tyt1 and a valine in TnaT, which bind the larger substrates, tryptophan and tryptophan. To the extent that the binding sites are similar in different transporters, the substrate mapping also results in what seems to be a different orientation of the substrates in the biogenic amine transporters (Fig. 9) that contain glycine at 8.64, in contrast to the substantially bulkier residues in most NSS members at this position. Thus, the more spacious cavity created by glycine at 8.64 seems to permit a shift in the positioning of the substrates, thereby allowing for a relatively bulky Phe6.59 that might not have been accommodated in Tyt1 and TnaT, where the side chains at 8.64 are much bulkier.

Although initial glimpses of the binding sites are now enabled by our comprehensive alignment, it is essential to support a structural context for other functional states is complex. Although we are now pursuing more rigorous docking procedures in a dynamic context of the protoin, and are testing these models experimentally in a variety of NSS proteins, these first clues suggest that armed with the LeuT structure and appropriate alignments, substantial insights can be gleaned into the molecular bases of NSS function. Much more work will be required to identify and validate the determinants of specificity and of the dynamics of the transport process, the next great challenges in understanding the function of the clinically important transporters.

References


Address correspondence to: Harel Weinstein. Department of Physiology and Biophysics, Weill Medical College of Cornell University. Box 75, Room E-509, 1300 York Ave., New York, NY 10021. E-mail: haw2002@med.cornell.edu