Molecular Pharmacology Fast Forward. Published on July 31, 2006 as doi: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

MOL#26120

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, NY 10021 (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 10032 (J.A.J.).

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

MOL#26120

Running title: Structure-based alignment of neurotransmitter/Na^+ symporters

Corresponding Author:

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

Tel: 212-746 6358, Fax: 212-746 8690, E-Mail: haw2002@med.cornell.edu

Number of text pages: 16, number of tables: 2, number of figures: 9, number of references: 51, number of words in abstract: 168, number of words in introduction: 727 number of words in discussion: 1485

Abbreviations: NSS, neurotransmitter:sodium symporter; SNF, sodium:neurotransmitter symporter family; SLC6, sodium- and chloride-dependent neurotransmitter transporter family; HUGO, Human Genome Organization (HUGO); TM, transmembrane segment; SASA, solvent accessible surface area; SCAM, substituted-cysteine accessibility method; MTS, methanethiosulfonate; MTSEA, aminoethylmethanethiosulfonate; MTSET, [2-(trimethylammonium) ethyl] methanethiosulfonate bromide; MTSES, sodium (2-sulfonatoethyl)-methanethiosulfonate; EL, extracellular loop; IL, intracellular loop; DAT, dopamine transporter; NET, norepinephrine transporter; SERT, serotonin transporter; BET, betaine transporter; GAT, GABA transporter; TAUT, taurin transporter; CREAT, creatine transporter; GLYT, glycine transporter; PROT, proline transporter; NAAT, neutral amino acid transporter; NCAAT, neutral and cationic amino acid transporter.

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 towards 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 and 4, our alignment differs from the one proposed by Yamashita et al. (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.

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Introduction:

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 prokaryotes genomes. TnaT of *Symbiobacterium thermophilum*, has been shown to be a Na^+-dependent tryptophan transporter (Androutsellis-Theotokis et al., 2003). Recently, the structure of another prokaryotic member of the NSS-family, a leucine transporter (LeuT) from *Aquifex Aeolicus*, has been determined to 1.65 Å resolution by X-ray crystallography (Yamashita et al., 2005). Curiously, the majority of homologous prokaryotic transporters are predicted to contain only 11 TMs (Quick et al., submitted). One of these, a tyrosine transporter that is fully

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

functional with only 11 TMs, has been characterized recently (Quick et al., submitted).

The use of LeuT and other prokaryotic NSS-proteins to enable structural and functional inferences regarding their eukaryotic homologs is predicated on an accurate alignment. However, the sequence identity of LeuT with clinically important eukaryotic transporters such as DAT, SERT or NET is only 20, 21 and 24%, respectively. Alignment of proteins at this low level of similarity is problematic (Rost, 1999), and the additional application of specialized algorithms, as well as the consideration of biochemical and pharmacological data, are necessary for validation and refinement. Despite low overall sequence conservation between LeuT and other NSS-proteins, several regions are highly conserved throughout the family. These include the TMs surrounding the leucine binding site in LeuT, namely TMs 1, 3, 6, and 8. The existence of such high conservation in functionally important regions of the protein (see below) suggests that these proteins likely share a common mechanism of transport, and that these regions can be modeled with relatively high accuracy in other NSS family members.

A number of regions outside of the direct substrate contacts, however, are poorly conserved and are difficult to align. Nonetheless, many of these are important for function, as evidenced by the many residues in distal TMs, the mutation of which has been shown to disrupt binding and/or transport properties (Chen and Reith, 2000; Goldberg et al., 2003; Norregaard and Gether, 2001; Volz and Schenk, 2005). For example, mutation of residues in TM2 (Wu and Gu, 2003) and TM12 (Gu et al., 2006; Neubauer et al., 2006) have been proposed to play an important role in the binding of cocaine and other inhibitors. Cocaine binding has also been shown to alter the conformation of TM4 in DAT (Hastrup et al., 2003). In the LeuT structure, these TMs do

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

not line the binding site and are located on the periphery of the structure, where they likely modulate the conformation of the core TMs that participate directly in binding (Sen et al., 2005). In addition, mutation of a residue in the poorly conserved TM9 of NET has been shown to impair surface expression and lead to orthostatic intolerance in humans (Hahn et al., 2003). Furthermore, many of the loops in NSS proteins are not conserved, and while their role in the functioning of LeuT is not yet well understood, they may contribute to the permeation pathway at a certain stage in the translocation cycle, and hence play an important functional role (e.g. see (Loland et al., 2002)).

Consequently, 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 Figure 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 refined where necessary by the consideration of experimental data.

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Methods:

Sequence retrieval and initial alignment: Sequences of NSS family members were collected by blasting against the NCBI RefSeq database using a BioPerl script (Stajich et al., 2002). Initially, 224 eukaryotic and 231 prokaryotic sequences were identified. Removing sequences with >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 sub-alignments were merged by profile-profile alignment using T-coffee (Notredame et al., 2000). The combined alignment was then subjected to further manual adjustment as described in Results.

Topology prediction: The use of consensus results from a variety 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 5 methods (i.e. TOPPRED2 (von Heijne, 1992), ORIENTM (Liakopoulos et al., 2001), HMMTOP (Tusnady and Simon, 2001), TMHMM (Krogh et al., 2001) and MEMSAT (Jones et al., 1994)) can predict the center of a TM segment with an average error of two residues (Beuming and Weinstein, 2005). Such consensus topology prediction was applied here to validate the alignment of TM4 and TM12.

Prediction of interior and lipid exposed faces of TM domains: To predict the interior and surface-exposed faces of TM domains, we used a method that integrates a knowledge-

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

based surface propensity scale with a conservation criterion to yield the probability that a particular face of a TM faces the interior or is lipid-exposed (Beuming and Weinstein, 2004; Beuming and Weinstein, 2005). The prediction of interior faces was used to facilitate the alignment of TM9. An alignment 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. PROFphd (Rost et al., 2004), Sspro (Pollastri et al., 2002), and PSIPRED (Jones, 1999), were used to guide the alignment 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 Figure 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, serotonin and norepinephrine, and the amino acid backbone in tyrosine, 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

those of tyrosine, tryptophan, and glycine, respectively, while the coordinates of the backbone atoms were kept the same. For the biogenic amines (dopamine, norepinephrine, and epinephrine) the amine-nitrogen was superimposed on the nitrogen of the amino acid, while the aromatic ring was superimposed on the leucine side-chain. In all cases, the space available in the binding sites was sufficient to accommodate the substrate without major structural rearrangement of the TM domains. However, a brief energy minimization using CHARMM (Brooks et al., 1983) was performed to eliminate any localized steric incompatibilities.

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Results:

Creating an initial alignment: A non-redundant 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, and TM5, 3) EL4, 4) TM9, and 5) TM12. These regions are indicated on a secondary structure map of the LeuT sequence in Figure 2. The alignment was refined for the entire family of prokaryotic and eukaryotic NSS-proteins but, for clarity, we illustrate the specific refinement steps and the use of experimental data to guide this refinement with the examples of the alignment of LeuT with DAT and/or SERT. A subset of the final refined alignment is shown in Figure 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 Figure 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. Similarly, 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 Figure 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 appears 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 Figure 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

evidence comes from the use of the sequence of a prokaryotic tryptophan transporter TnaT (Androutsellis-Theotokis et al., 2003), as a comparative intermediate between LeuT and 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% sequence identity) and between TnaT and DAT (30% sequence identity), if TnaT is aligned with DAT as proposed here. If the alignment of Yamashita et al. (2005) is followed, the resulting sequence identity between DAT and TnaT falls to only 4%. Several pairwise homologous regions are indicated in black in Figure 3. These regions also have significant homology within the individual prokaryotic 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 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 alignment presented here, Ile248 in DAT and Met176 in LeuT (shaded white in Figure 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 substituted cysteines in IL2 and TM5 in SERT (Zhang and Rudnick, 2005) provides further justification for the alignment chosen here. Thus, in a membrane preparation, cysteines substituted for SERT residues Trp271,

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 towards the interior of the protein where it presumably lines part of the transport pathway (see Figure 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 appears to have the same length as well as a 100% conserved aromatic residue (Phe324 in LeuT and Phe391 in DAT) (see Figure 1). In contrast, EL4a is highly variable in length and ranges between 3 and 80 residues. Interestingly, the EL4a segments of DAT, SERT, NET, GAT and the glycine transporters appear to be similar in length to that of LeuT and are predicted to have an α-helical component.

Analysis of the entire EL4 segment in SERT (residues 386 to 423) with the substituted-cysteine accessibility method (SCAM) supported a helix-coil-helix motif for EL4, in the absence of any other structural information (Mitchell et al., 2004). Residues 386-399 displayed α-helical periodicity of sensitivity to MTS reagents, whereas residues 400-408 all appeared to be accessible. Residues 409-421 were noted to have a α-helical periodicity of residue property, with acidic residues clustered on one face. Moreover, the

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

corresponding substituted cysteine mutants had the fastest reaction rates with MTS reagents. This information is consistent with related, but more limited, data from GAT (Zomot and Kanner, 2003) and has been used to guide the alignment of SERT with LeuT, by threading the EL4a residues of SERT through the structure of LeuT. This structure-based alignment was carried out such that positions that point towards the solvent are occupied by the residues most reactive with the MTS reagents, whereas inaccessible or less reactive residues are at positions that are more buried (see Figure 5). The resulting alignment is shown in Figure 1. Notably, this alignment is in good agreement with the demonstration of an endogenous zinc binding site in EL4 of DAT (Norregaard et al., 2000), SERT (Mitchell et al., 2004) and GAT (MacAulay et al., 2001). In DAT, the zinc binding site is formed by residues His375 (EL4a) and Glu396 (EL4b), together with a residue in EL2, His193. According to the alignment shown in Figure 1, the residues that correspond to Glu396 and His375 in LeuT are Asn310 and Ala329, respectively. The aligned positions in SERT are Arg390 in EL4a and Glu412 in EL4b (Mitchell et al., 2004), as shown in Figure 5. The Cβ-Cβ distance between these two residues is approximately 11 Å in our SERT model (Figure 5, left panel), 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) (Figure 5).

TM9: This is another region where there is no similarity between LeuT and eukaryotic NSS-proteins. Unfortunately, very little 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 Figure 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 panel B. 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. Notably, Figure 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 panel B 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 sodium-dependent transporters (Quick et al., submitted). Consensus topology prediction of DAT indicates that the central residue of TM12 is Ser568 (see Figure 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 Figure 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 (Figure 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.

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 as 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_{84}1.50. 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 1-4, 6-8, and 11 this numbering scheme remains valid in the present alignment. However, in the structure of LeuT, the reference residues proposed as X.50

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 an * in Figure 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 due to 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 Figure 9). Since most of the NSS substrates identified to date are amino acids (see below for a discussion of the exceptions, the biogenic amines) and since sodium appears 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, submitted). This exquisite specificity likely results in substantial part from interactions with these non-conserved residues that line the more intracellular portion of the binding site (white in Figure 9).

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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). Consequently, 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, due in part to LeuT being a non-representative NSS-protein, even among the prokaryotes. Indeed, the average sequence identity of LeuT to eukaryotic NSS-proteins is ~18%, which is amongst the lowest values of the 231 bacterial and archaeal NSS-proteins. Many other prokaryotic family members have substantially higher similarity to eukaryotic NSS; for example, the tryptophan transporter TnaT 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

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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., submitted). 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, as 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 Figure 9). In many cases, the side chains of

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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 out 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 and 1.43 are 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. noted that the residues that coordinate Na2 are less well conserved than those that contact Na1 (Yamashita et al., 2005). However, of the 5 residues that contact Na2, 4 are highly conserved in eukaryotic NSS-proteins (Table 2). In 3 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 TnaT and Tyt1 (and in 21% of prokaryotic NSS), which also appear to have a stoichiometry of 2 Na^+: 1 substrate (Androutsellis-Theotokis et al., 2003, Quick et al., submitted). 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^+: 1 substrate for several eukaryotic NSS with Asp (i.e. GAT-1 or DAT (Krause and Schwarz, 2005;

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Krueger, 1990)) 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 1 Na^+: 1 substrate 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., submitted) 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 non-conserved 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 6.56 and 6.59 with the bulk of the substrate side-chain in the model of the complex shown in Figure 9. In these initial models obtained by simple mapping of the substrate into the homology model (see Methods for details), the smallest

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

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, tyrosine and tryptophan. To the extent that the binding sites are similar in different transporters, the substrate mapping also indicates in what appears to be a different orientation of the substrates in the biogenic amine transporters (Figure 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 appears 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 emphasize the preliminary nature of results obtained from mere mapping of ligands across protein binding sites. Moreover, the LeuT structure is proposed to be a closed state in which the bound substrate is occluded from both the extracellular and the intracellular sides of the membrane (Yamashita et al., 2005) so that the use of experimental data to support a structural context for other functional states is complex. While we are now pursuing more rigorous docking procedures in a dynamic context of the protein, 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:

Androutsellis-Theotokis A, Goldberg NR, Ueda K, Beppu T, Beckman ML, Das S, Javitch JA and Rudnick G (2003) Characterization of a functional bacterial homologue of sodium-dependent neurotransmitter transporters. *J Biol Chem* **278**:12703-12709.

Ballesteros JA and Weinstein H (1995) Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods Neuroscience* **25**:366-428.

Beitz E (2000) TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **16**:135-139.

Beuming T and Weinstein H (2004) A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. *Bioinformatics* **20**:1822-1835.

Beuming T and Weinstein H (2005) Modeling membrane proteins based on low-resolution electron microscopy maps: a template for the TM domains of the oxalate transporter OxlT. *Protein Eng Des Sel* **18**:119-125.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**:187-217.

Chen N and Reith ME (2000) Structure and function of the dopamine transporter. *Eur J Pharmacol* **405**:329-339.

Chen NH, Reith ME and Quick MW (2004) Synaptic uptake and beyond: the sodium- and chloride-dependent neurotransmitter transporter family SLC6. *Pflugers Arch* **447**:519-531.

Delano WL (2002) *The PyMOL User's Manual*. DeLano Scientific, San Carlos, CA, USA.

Do CB, Mahabhashyam MS, Brudno M and Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* **15**:330-340.

Fiser A and Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**:461-491.

Goldberg NR, Beuming T, Soyer OS, Goldstein RA, Weinstein H and Javitch JA (2003) Probing conformational changes in neurotransmitter transporters: a structural context. *Eur J Pharmacol* **479**:3-12.

Gu H, Caplan MJ and Rudnick G (1998) Cloned catecholamine transporters expressed in polarized epithelial cells: sorting, drug sensitivity, and ion-coupling stoichiometry. *Adv Pharmacol* **42**:175-179.

Gu HH, Wu X and Han DD (2006) Conserved serine residues in serotonin transporter contribute to high-affinity cocaine binding. *Biochem Biophys Res Commun.*

Hahn MK, Robertson D and Blakely RD (2003) A mutation in the human norepinephrine transporter gene (SLC6A2) associated with orthostatic intolerance disrupts surface expression of mutant and wild-type transporters. *J Neurosci* **23**:4470-4478.

Hastrup H, Sen N and Javitch JA (2003) The human dopamine transporter forms a tetramer in the plasma membrane: cross-linking of a cysteine in the fourth transmembrane segment is sensitive to cocaine analogs. *J Biol Chem* **278**:45045-45048.

Henry LK, Defelice LJ and Blakely RD (2006) Getting the message across: a recent transporter structure shows the way. *Neuron* **49**:791-796.

Humphrey W, Dalke A and Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**:33-38, 27-38.

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**:195-202.

Jones DT, Taylor WR and Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**:3038-3049.

Koh IY, Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Eswar N, Grana O, Pazos F, Valencia A, Sali A and Rost B (2003) EVA: Evaluation of protein structure prediction servers. *Nucleic Acids Res* **31**:3311-3315.

Krause S and Schwarz W (2005) Identification and selective inhibition of the channel mode of the neuronal GABA transporter 1. *Mol Pharmacol* **68**:1728-1735.

Krogh A, Larsson B, von Heijne G and Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**:567-580.

Krueger BK (1990) Kinetics and block of dopamine uptake in synaptosomes from rat caudate nucleus. *J Neurochem* **55**:260-267.

Liakopoulos TD, Pasquier C and Hamodrakas SJ (2001) A novel tool for the prediction of transmembrane protein topology based on a statistical analysis of the SwissProt database: the OrienTM algorithm. *Protein Eng* **14**:387-390.

Loland CJ, Norregaard L, Litman T and Gether U (2002) Generation of an activating Zn(2+) switch in the dopamine transporter: mutation of an intracellular tyrosine constitutively alters the conformational equilibrium of the transport cycle. *Proc Natl Acad Sci U S A* **99**:1683-1688.

MacAulay N, Bendahan A, Loland CJ, Zeuthen T, Kanner BI and Gether U (2001) Engineered Zn(2+) switches in the gamma-aminobutyric acid (GABA) transporter-1. Differential effects on GABA uptake and currents. *J Biol Chem* **276**:40476-40485.

Mitchell SM, Lee E, Garcia ML and Stephan MM (2004) Structure and function of extracellular loop 4 of the serotonin transporter as revealed by cysteine-scanning mutagenesis. *J Biol Chem* **279**:24089-24099.

Neubauer HA, Hansen CG and Wiborg O (2006) Dissection of an allosteric mechanism on the serotonin transporter: a cross-species study. *Mol Pharmacol* **69**:1242-1250.

Norregaard L and Gether U (2001) The monoamine neurotransmitter transporters: structure, conformational changes and molecular gating. *Curr Opin Drug Discov Devel* **4**:591-601.

Norregaard L, Visiers I, Loland CJ, Ballesteros J, Weinstein H and Gether U (2000) Structural probing of a microdomain in the dopamine transporter by engineering of artificial Zn2+ binding sites. *Biochemistry* **39**:15836-15846.

Notredame C, Higgins DG and Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**:205-217.

Pollastri G, Przybylski D, Rost B and Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* **47**:228-235.

Rost B (1999) Twilight zone of protein sequence alignments. *Protein Eng* **12**:85-94.

Rost B, Yachdav G and Liu J (2004) The PredictProtein server. *Nucleic Acids Res* **32**:W321-326.

Roux MJ and Supplisson S (2000) Neuronal and glial glycine transporters have different stoichiometries. *Neuron* **25**:373-383.

Rudnick G (1998) Ion-coupled neurotransmitter transport: thermodynamic vs. kinetic determinations of stoichiometry. *Methods Enzymol* **296**:233-247.

Saier MH, Jr. (1999) A functional-phylogenetic system for the classification of transport proteins. *J Cell Biochem* **Suppl 32-33**:84-94.

Sen N, Shi L, Beuming T, Weinstein H and Javitch JA (2005) A pincer-like configuration of TM2 in the human dopamine transporter is responsible for indirect effects on cocaine binding. *Neuropharmacology* **49**:780-790.

Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD and Birney E (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12**:1611-1618.

Tusnady GE, Dosztanyi Z and Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res* **33 Database Issue**:D275-278.

Tusnady GE and Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**:849-850.

Vardy E, Arkin IT, Gottschalk KE, Kaback HR and Schuldiner S (2004) Structural conservation in the major facilitator superfamily as revealed by comparative modeling. *Protein Sci* **13**:1832-1840.

Volz TJ and Schenk JO (2005) A comprehensive atlas of the topography of functional groups of the dopamine transporter. *Synapse* **58**:72-94.

von Heijne G (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**:487-494.

Wu X and Gu HH (2003) Cocaine affinity decreased by mutations of aromatic residue phenylalanine 105 in the transmembrane domain 2 of dopamine transporter. *Mol Pharmacol* **63**:653-658.

Yamashita A, Singh SK, Kawate T, Jin Y and Gouaux E (2005) Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature* **437**:215-223.

Zhang YW and Rudnick G (2005) Cysteine-scanning mutagenesis of serotonin transporter intracellular loop 2 suggests an alpha-helical conformation. *J Biol Chem* **280**:30807-30813.

Zomot E and Kanner BI (2003) The interaction of the gamma-aminobutyric acid transporter GAT-1 with the neurotransmitter is selectively impaired by sulfhydryl

modification of a conformationally sensitive cysteine residue engineered into extracellular loop IV. *J Biol Chem* **278**:42950-42958.

Footnotes:

This work was supported by National Institutes of Health grants P01 DA12408, R01s DA11495 and DA17293, K02 MH57324 and K05 DA00060.

*These authors contributed equally.

#These authors contributed equally.

Legends for figures:

Figure 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 an *. 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).

Figure 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 non-conserved 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 non-helical stretches (i.e. TMs 1 and 6) are shown as continuous helices.

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 3: Detailed alignment of TM4, TM5 and EL3. A non-redundant 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 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 figure was prepared using TEXshade (Beitz, 2000).

Figure 4: MTSEA-accessible residues in TM5 of SERT. Models of SERT were based on the LeuT structure and the alignment presented here (top) or by Yamashita et al., (2005) (bottom). The TM bundle is viewed from the side (left) or from the top (right). TM4 and TM5 are colored magenta. Loops have been omitted for clarity. Side-chains of residues that are sensitive to MTSEA when mutated to cysteine are indicated as spheres. The alignment presented here places all MTSEA-accessible residues on the helical face of TM5 that points towards the interior of the protein (all residues shown in yellow). In the alignment of Yamashita et al., (2005), residues Val280 (5.28) and Thr284 (5.32) (shown in cyan) are located on the lipid-exposed face of TM5, where they would not be expected

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

to react with MTSEA and lead to inhibition of ligand binding. The figure was prepared using PyMOL (Delano, 2002).

Figure 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 MTSET or MTSES 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 SCAM results as the accessible residues in EL4a and El4b face different direction and the most reactive residues in EL4a are buried towards the interior. The two residues with side-chains rendered in sticks are part an 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 C□-atoms of the residues involved in Zn^{2+}-binding is ~11Å (Zn^{2+} shown in green); in contrast, the model based on the alignment of Yamashita et al. (2005) places them much further apart (~18Å), in orientations that are unlikely to enable Zn^{2+}-binding. The figure was prepared using PyMOL (Delano, 2002).

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 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 T449, T456 and S460 on the lipid-facing surface of TM9.

Figure 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 (T449 (9.31), T456 (9.38) and S460 (9.42), shown in cyan) are located on the lipid-facing surface of TM9. The figure was prepared using PyMOL (Delano, 2002).

Figure 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 an *) are buried in the interior of the protein, and the centrally predicted residue in DAT (Ser568 (12.45), indicated with an #) is kept as close as possible to the center of TM12 in LeuT (Leu493 (12.45), indicated with an #). The figure was prepared using TEXshade (Beitz, 2000).

Figure 9. The substrate binding site in NSS-proteins. The left panel shows the structure of LeuT (Yamashita et al., 2005). Only TMs that contribute to the binding site of leucine (yellow) and the two sodium ions (blue) are shown (TMs 1,3,6,7 and 8). The panels on the right show models of DAT, NET and SERT (top) and Glyt1, Tyt1 and TnaT (bottom). Models were generated with Modeller (Fiser and Sali, 2003), using the LeuT structure as a template and following our proposed alignment. Ligands were docked based on superposition, followed by a brief minimization (1000 steps) using the CHARMM molecular dynamics package (Brooks et al., 1983). Residues conserved among NSS-proteins are shown in cyan, whereas residues that are variable are shown in white. See Table 2 for a more detailed description of this classification. G24 (1.45) is shown in green – this residue is conserved in all NSS-proteins except in the biogenic amine transporters, where it is an Asp. As discussed in the text, residues that interact with the side-chains and aromatic rings of the substrate are variable, whereas residues that interact with the amine or amino-acid groups of the substrate, or with the two sodium ions, are conserved. The figure was prepared using VMD (Humphrey et al., 1996).

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Tables:

Table 1: Structure-based Generic Numbering of NSS-proteins. Index residues indicated with a * differ from the previously proposed numbering system (Goldberg et al. 2003).

TM	hDAT	LeuT	Conservation in eukaryotic transporters	Conservation in prokaryotic transporters
1	W84	L29	95%	97%
2	P112	P57	100%	79%
3	Y156	Y108	92%	100%
4	C243	V171	77%	Not conserved
5*	P273	L202	95%	Not conserved
6	Q317	Q250	95%	89%
7	F365	S298	95%	83%
8	F412	F345	90%	70%
9*	F457	F387	86% (F/Y)	55% (F/Y)
10*	F478	W406	93% (F/Y)	25% (F/Y)
11	P529	P457	97%	81%
12*	P573	T498	95%	Not conserved

Molecular Pharmacology Fast Forward. Published on July 31, 2006 as DOI: 10.1124/mol.106.026120
This article has not been copyedited and formatted. The final version may differ from this version.

Table 2: Conservation of binding-site residues in NSS-proteins. The generic numbers and the residue numbers in LeuT are shown in column 1. Columns 2 to 6 list the various contacts of the two sodium ions and the leucine substrate, as observed in the LeuT structure (Yamashita et al., 2005). 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 2.5Å distance cut-off was used to determine all Na-coordinating residues. # is used to indicate whether these residues contact leucine or sodium via backbone atoms, side-chain atoms, or both. The moiety of the substrate that is involved in the interactions is indicated with 'b' (backbone), 's' (side-chain), or 'bs' (both). Residues from prokaryotic NSS-family members with known substrates (LeuT, TnaT and Tyt1) are shown in columns 7-9, while column 10 and 11 list the variability in 167 prokaryotic NSS-proteins. Residues from eukaryotic NSS-proteins with known substrates are shown in columns 12 to 22. These include transporters for dopamine (DAT), norepinephrine (NET), serotonin (SERT), betaine (BET), GABA (GAT), taurine (TAUT), creatine (CREAT), glycine (GLYT), proline (PROT), neutral amino acids (NAAT), and neutral and cationic amino acids (NCAAT). Finally, columns 23 and 24 show the variability in 177 eukaryotic NSS-proteins. Conserved residues shown in cyan in Figure 8 are indicated in bold font. Contact positions classified as conserved were those in all NSS-proteins with known substrate 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 biogenic amine transporters where it is an Asp, and is shown in italic font.

	Contacts					Prokaryotic					Eukaryotic												
Generic Nr. (LeuT Nr.)	Side-chain	Back-bone	Leu	Na1	Na2	LeuT	TnaT	Tyt1	Types		DAT	NET	SERT	BET	GAT	TAUT	CREAT	GLYT	PROT	NAAT	NCAAT	Types	
1.41 (20)		#			**x**	**G**	**G**	**G**	**G(98%)**	**4**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G(83%),S(11%),A(5%)**	**4**
1.42 (21)	#	#	bs			N	N	S	S(70%),A(21%)	8	F	F	Y/F	E	E/Y/L	G	F	Y	Y	F	Y	Y(38%),F(31%),E(8%), L(7%),G(6%)	10
1.43 (22)		#	b	x		A	A	A	A(89%),T(8%)	5	A	A	A	I	I/A	F	A	A	C	C	A	A(55%),S(21%), I(8%),C(5%)	8
1.44 (23)		#			**x**	**V**	**V**	**V**	**V(64%),I(34%)**	**6**	**V**	**V**	**V**	**I**	**I**	**V**	**V**	**V**	**V**	**V**	**V**	**V(76%),I(19%)**	**5**
1.45 (24)		#	*b*			*G*	*G*	*G*	*G(99%)*	*2*	*D*	*D*	*D*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G(75%),D(17%)*	*5*
1.46 (25)		#	**b**			**L**	**L**	**M**	**L(81%),F(7%),I(5%)**	**7**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L(94%)**	**4**
1.47 (26)		#	**b**			**G**	**G**	**G**	**G(98%)**	**2**	**A**	**A**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G(85%),A(9%)**	**4**
1.48 (27)	**#**			**x**		**N**	**N**	**N**	**N(67%),A(27%)**	**5**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N(98%)**	**2**
3.46 (104)	#		s			V	V	I	I(63%),L(20%),V(15%)	5	V	V	I/V	L	L/T	L	C	I	V	V	V	V(45%),L(21%),I(17%)	9
3.50 (108)	**#**		**bs**			**Y**	**Y**	**Y**	**F(59%),Y(38%)**	**3**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y(91%),F(5%)**	**3**
6.53 (253)	**#**	**#**	**bs**			**F**	**F**	**F**	**F(96%)**	**3**	**F**	**F/Y**	**F**	**F**	**F**	**F**	**F**	**F/Y**	**Y**	**F/Y**	**Y**	**F(71%),Y(25%)**	**4**
6.54 (254)	**#**	**#**	**b**	**x**		**T**	**S**	**S**	**S(57%),T(26%)**	**6**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S(88%),A(10%)**	**4**
6.56 (256)	#		bs			S	T	S	S(81%),T(10%),G(5%)	5	G	G	G	A	G/A	A	A	S/G	G	S	S	G(58%),S(24%),A(19%)	3
6.59 (259)	#		s			F	V	G	F(27%),V(20%),M(18%), I(10%),G(9%),A(7%)	11	F	F	F	Q	L	L	L	W	F	F	W	F(48%),L(19%),W(15%)	11
6.61 (261)	#		s			A	V	G	I(31%),V(26%),A(10%), S(9%),T(7%),G(7%),C(7%)	9	V	V	V	C	C/S/A	A	A	G	G	G	G	G(38%),V(20%),A(11%),S(10%), C(8%),T(5%)	10
7.38 (286)	**#**			**x**		**N**	**N**	**D**	**N(50%),D(40%),T(8%)**	**6**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N(88%),D(9%)**	**3**
8.56 (351)		**#**			**x**	**A**	**A**	**A**	**A(98%)**	**5**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L(92%)**	**4**
8.59 (354)	#				x	T	S	S	T(77%),S(21%)	5	D	D	D	D	D	D	D	D/G	D	S	D	D(56%),G(25%),S(11%)	7
8.60 (355)	**#**		**s**		**x**	**S**	**S**	**S**	**S(90%),T(8%)**	**3**	**S**	**S**	**S/F**	**S**	**S**	**S**	**S**	**T**	**S**	**S**	**S**	**S(83%),T(15%)**	**3**
8.63 (358)	#		s			A	S	N	S(80%),A(6%), T(5%),N(5%)	7	G	G	A/G	V	V/C	V	V	A/C	A	G	A	A(23%),C(13%), G(39%),V(22%)	6
8.64 (359)	#		s			I	Q	M	M(38%),L(36%),I(21%)	7	G	G	G	C	T/C	E	G	L/T	F	N	S	G(23%),T(19%),I(14%), L(8%),M(7%),C(7%),N(6%)	12

Figure 1

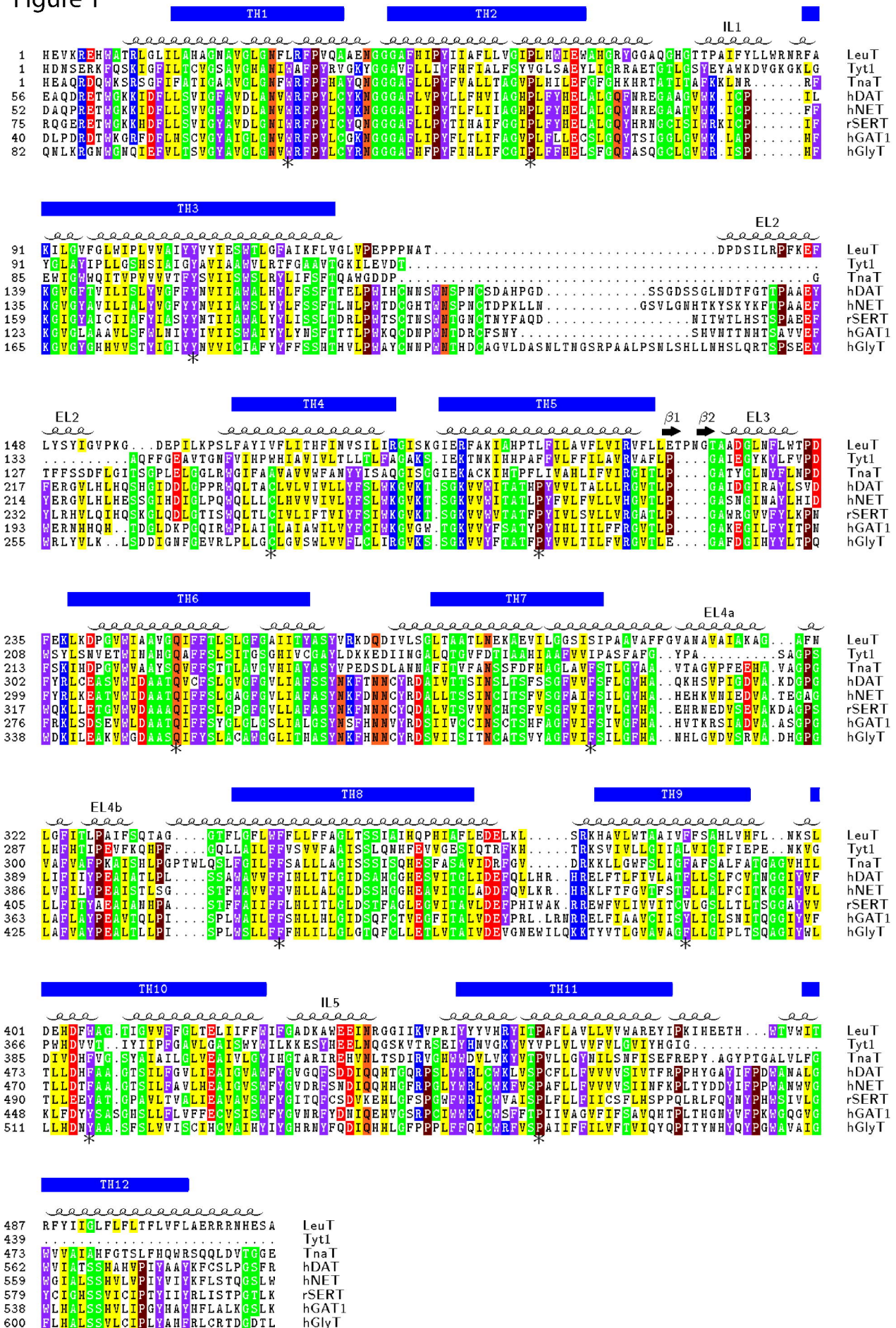

TH1
TH2
IL1
TH3
EL2
TH4
TH5
EL2
β1
β2
EL3
TH6
TH7
EL4a
EL4b
TH8
TH9
TH10
IL5
TH11
TH12
LeuT
Tyt1
TnaT
hDAT
hNET
rSERT
hGAT1
hGlyT

Figure 2

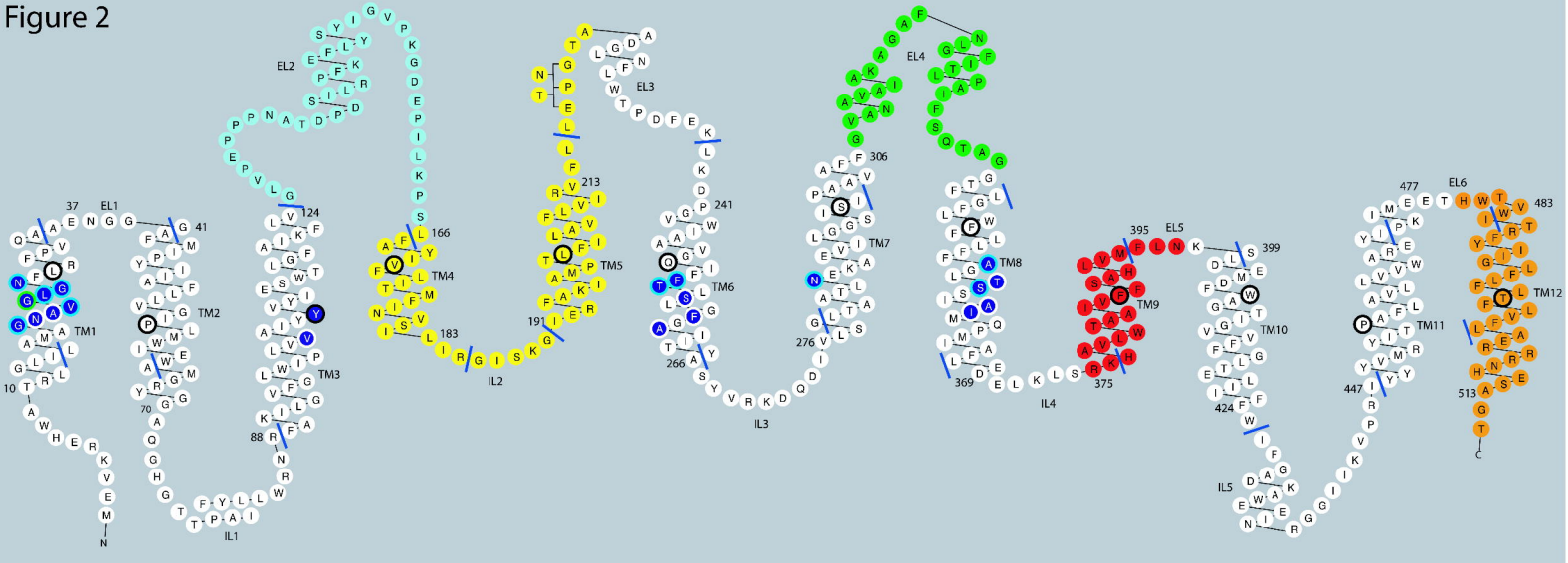
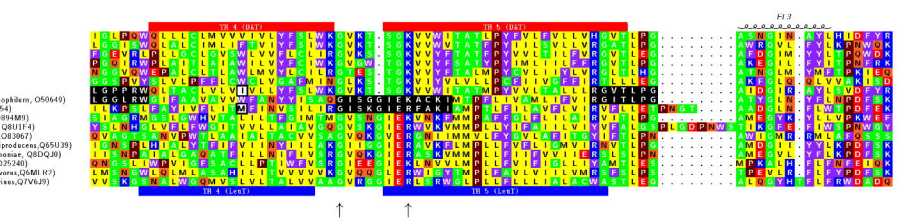

EL1
EL2
EL3
EL4
EL5
EL6
IL1
IL2
IL3
IL4
IL5
TM1
TM2
TM3
TM4
TM5
TM6
TM7
TM8
TM9
TM10
TM11
TM12
N
C

Figure 3


NET (Human,P23975)
SERT (Human,P31645)
GLYT1 (Human,P48067.2956)
GAT1 (Human,P30531)
Orphan (Rat,Q64093)
Orphan (Human,Q9GZN6)
DAT (Human,Q01959)
Taa1 (Symbiobacterium termophilum, O50649)
LeuT (Aquifex aeolicus, O67854)
Orphan (Clostridium tetani, Q894M9)
Orphan (Pyrococcus furiosus, Q8U1F4)
Orphan (Treponema pallidum,O83867)
Orphan (Mannheimia succiniciproducens,Q65U39)
Orphan (Streptococcus pneumoniae, Q8DQJ0)
Orphan (Helicobacter pylori, O25240)
Orphan (Idiellovibrio bacteriovorus,Q6MI K7)
Orphan (Prochlorococcus marinus,Q7V629)
TM 4 (LeuT)
TM 5 (LeuT)
TM 3 (LeuT)

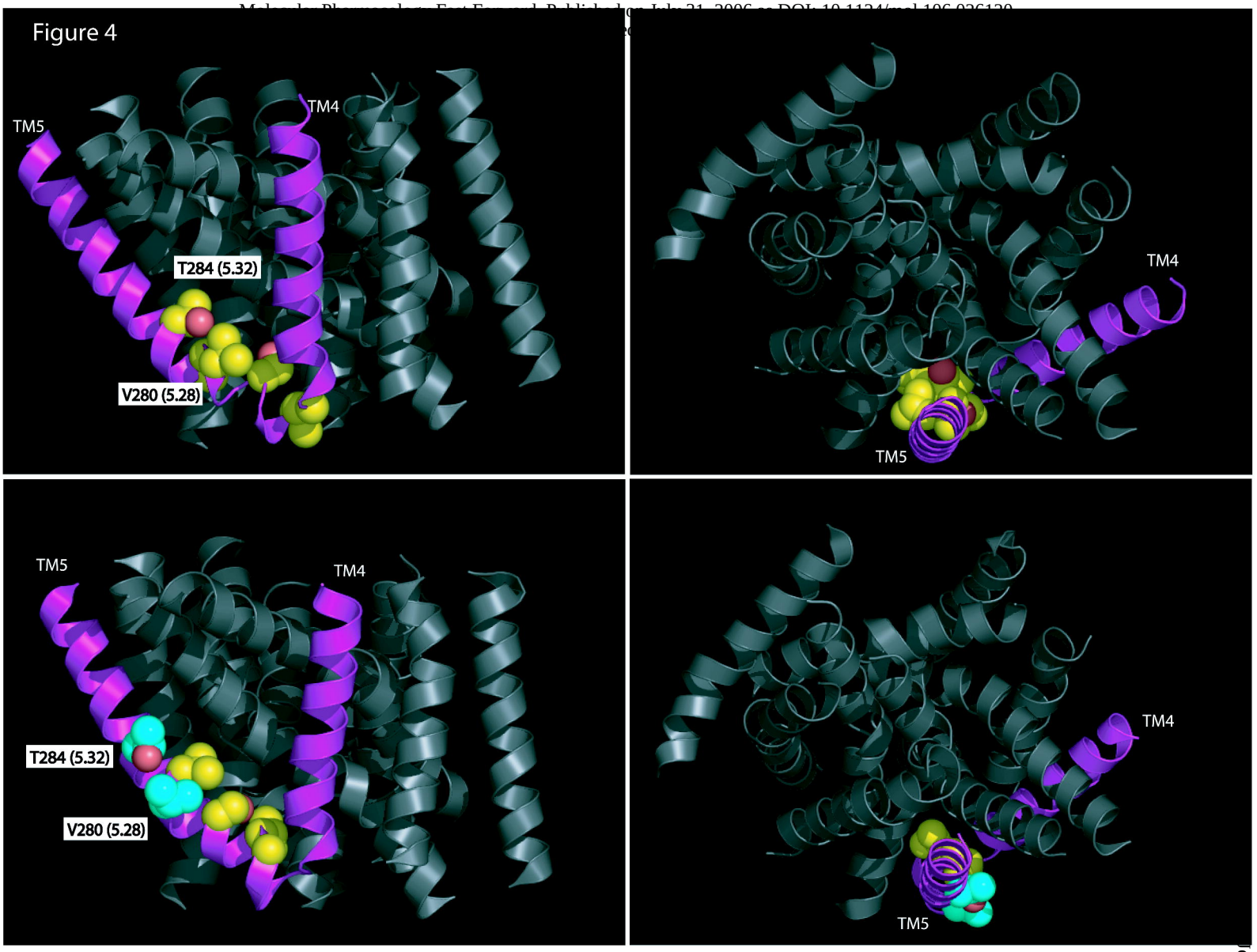


Figure 5

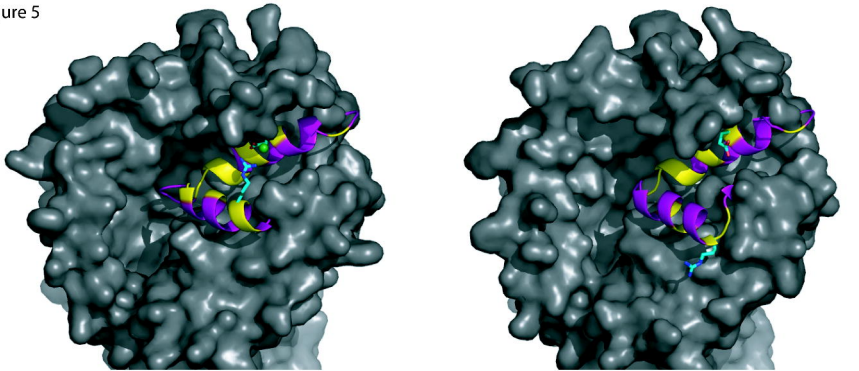
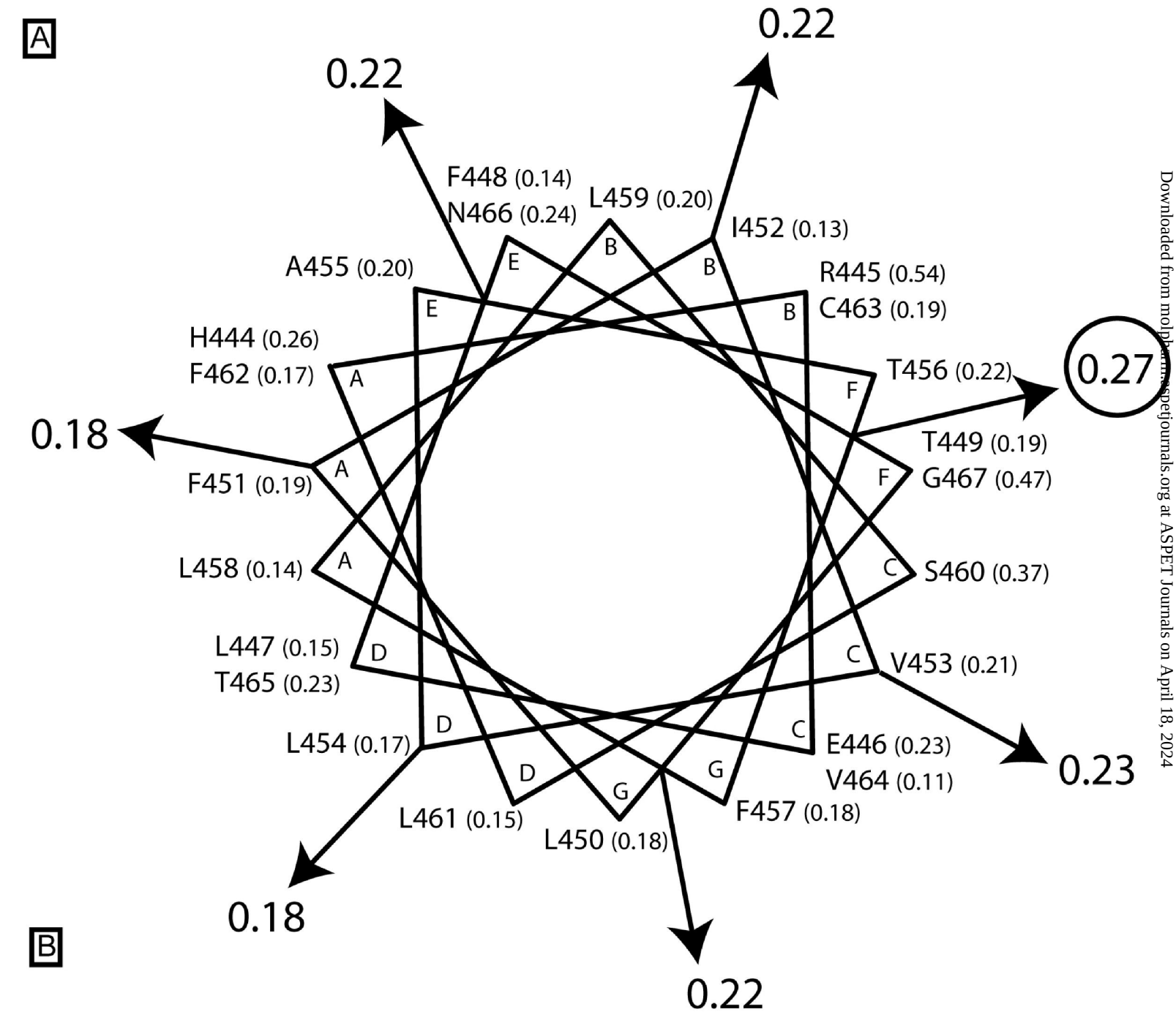


Figure 6

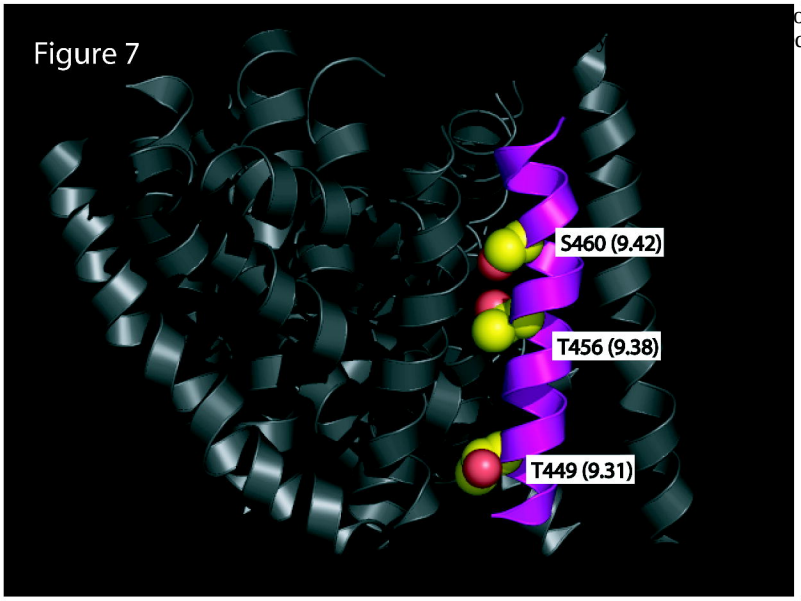


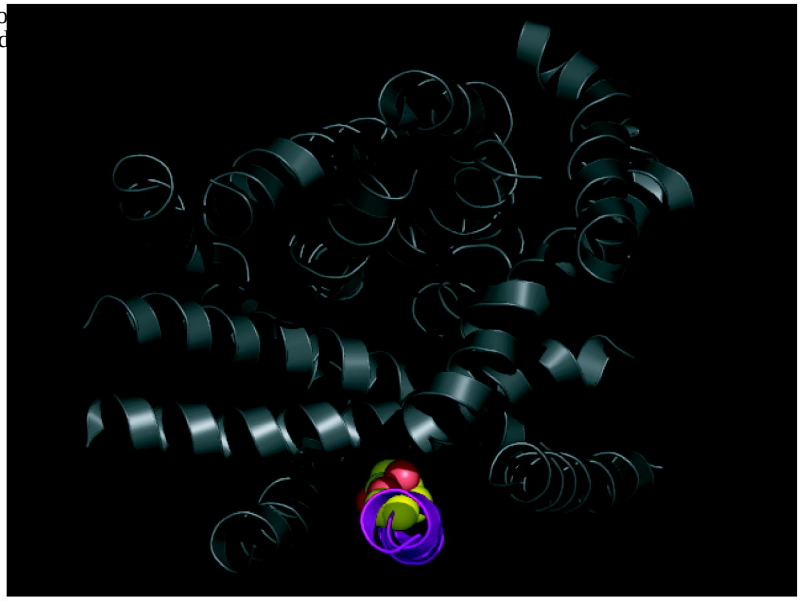
hDAT	444	H	**R**	**E**	L	F	**T**	L	F	**I**	V	L	A	**T**	F	L	**L**	**S**	L	F	**C**	**V**	T	N	**G**	467
LeuT	474	**S**	R	K	H	**A**	**V**	L	W	**T**	A	A	I	**V**	F	F	**S**	**A**	H	L	**V**	**M**	F	L	N	497

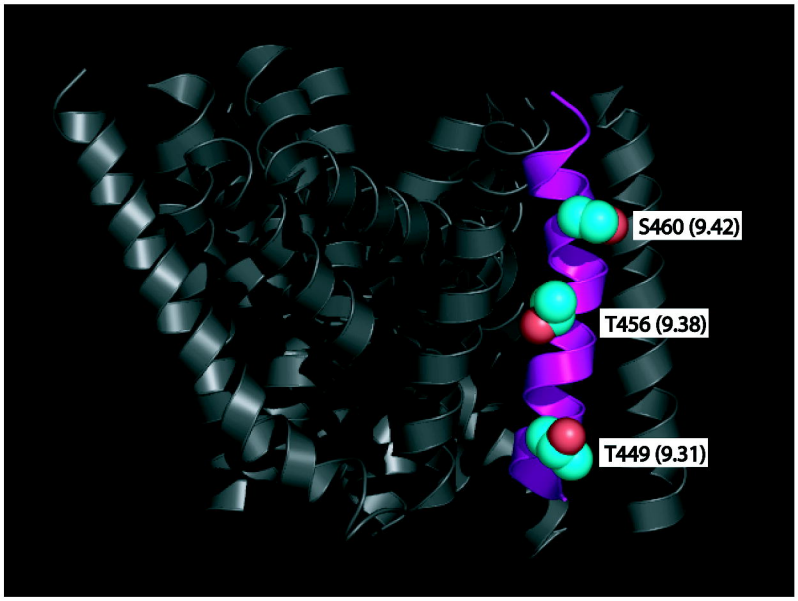
Yamashita et al. (2005)

hDAT	443	R	H	**R**	**E**	L	F	**T**	L	F	**I**	V	L	A	**T**	F	L	**L**	**S**	L	F	**C**	**V**	T	N	466
LeuT	474	**S**	R	K	H	**A**	**V**	L	W	**T**	A	A	I	**V**	F	F	**S**	**A**	H	L	**V**	**M**	F	L	N	497

Figure 7







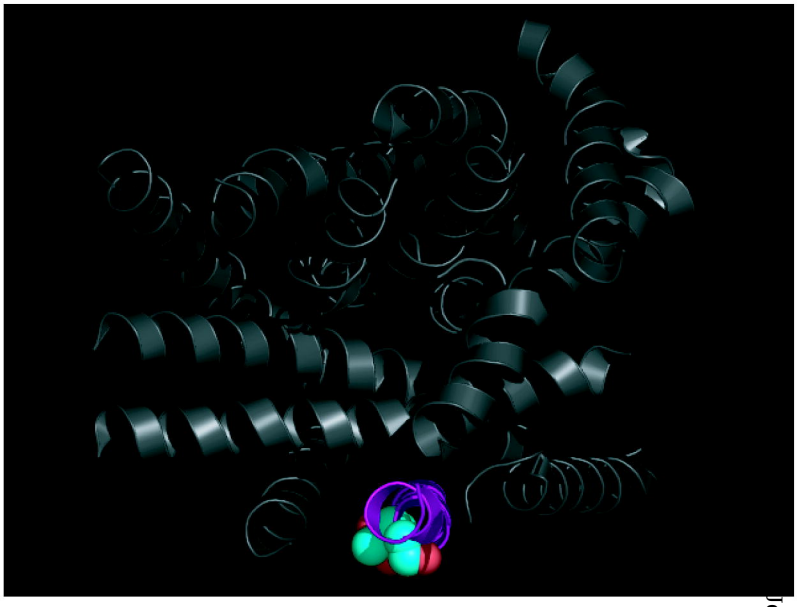
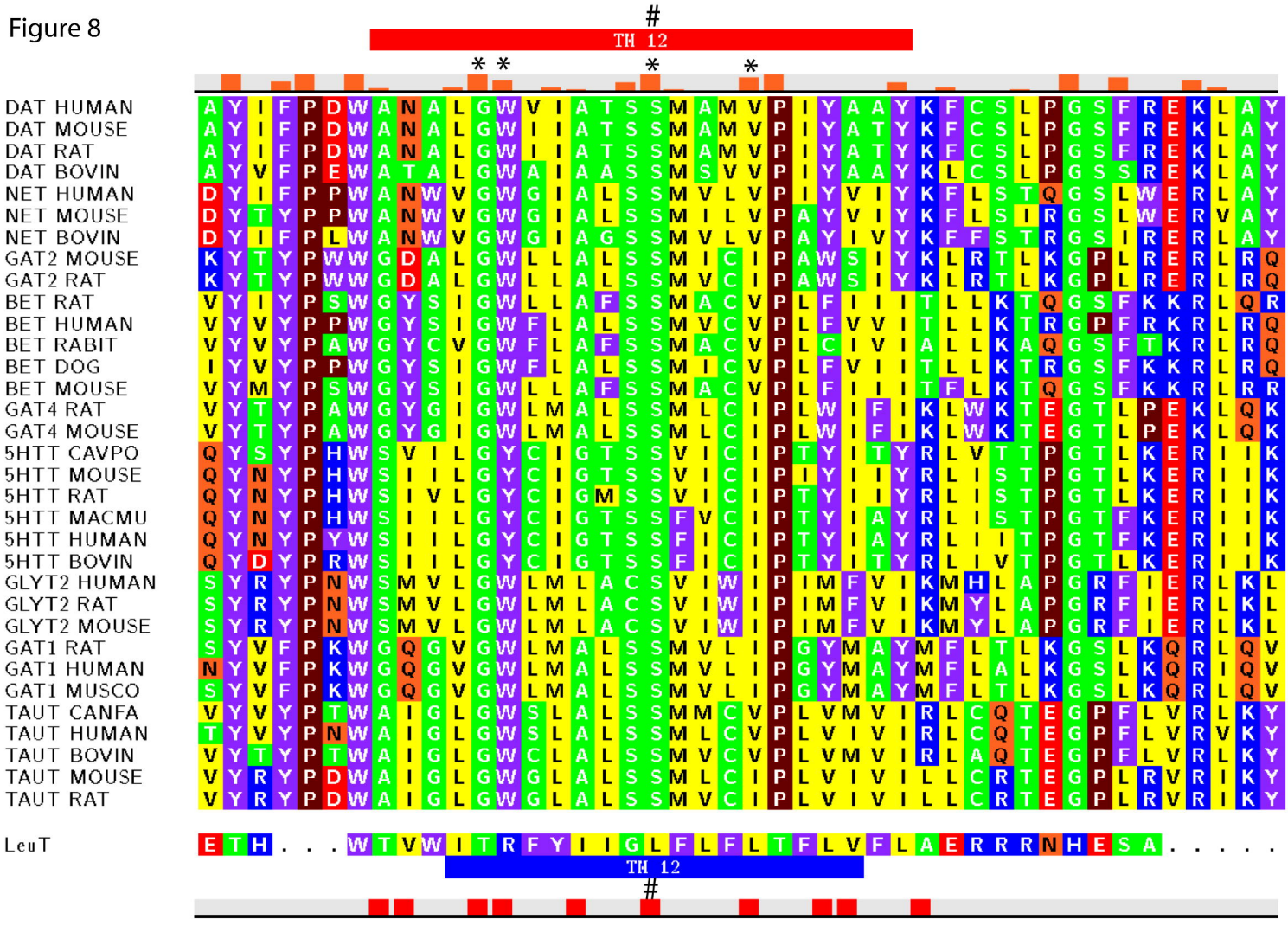


Figure 8



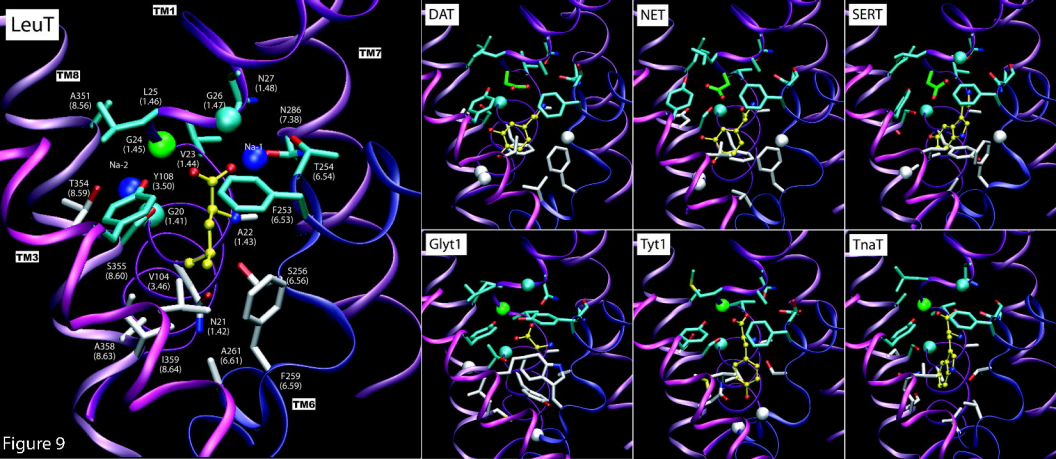

LeuT
TM1
TM7
TM8
TM3
TM6
A351 (8.56)
L25 (1.46)
G26 (1.47)
N27 (1.48)
N286 (7.38)
G24 (1.45)
Na-2
V23 (1.44)
Na-1
T254 (6.54)
T354 (8.59)
Y108 (3.50)
G20 (1.41)
F253 (6.53)
A22 (1.43)
S355 (8.60)
V104 (3.46)
S256 (6.56)
N21 (1.42)
A358 (8.63)
I359 (8.64)
A261 (6.61)
F259 (6.59)
DAT
NET
SERT
Glyt1
Tyt1
TnaT

Figure 9